ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MOLECULAR DOCKING STUDY ON THE DNA BINDING DOMAIN OF MASTER REGULATOR OF THE MAMMALIAN HSR PROTEIN

M.Sc. THESIS

Gülbahar BOZAN

Department of Chemistry

Chemistry Programme

JULY 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

MOLECULAR DOCKING STUDY ON THE DNA BINDING DOMAIN OF MASTER REGULATOR OF THE MAMMALIAN HSR PROTEIN

M.Sc. THESIS

Gülbahar BOZAN
(509171231)

Department of Chemistry

Chemistry Programme

Thesis Advisor: Prof. Dr. Nurcan TÜZÜN
Thesis Co-Advisor: Prof. Dr. Fethiye Aylin SUNGUR

JULY 2020

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

BAŞ DÜZENLEYİCİ MEMELİ HSR PROTEİNİNİN DNA BAĞLANMA BÖLGESİ ÜZERİNE MOLEKÜLER KENETLEME ÇALIŞMASI

YÜKSEK LİSANS TEZİ

Gülbahar BOZAN
(509171231)

Kimya Anabilim Dalı

Kimya Programı

Tez Danışmanı: Prof. Dr. Nurcan TÜZÜN
Eş Danışmanş Prof. Dr. Fethiye Aylin SUNGUR

TEMMUZ 2020

Gülbahar Bozan, a M.Sc. student of ITU Graduate School of Science Engineering and Technology student ID 509171231, successfully defended the thesis/dissertation entitled "MOLECULAR DOCKING STUDY ON THE DNA BINDING DOMAIN OF MASTER REGULATOR OF THE MAMMALIAN HSR PROTEIN", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Nurcan TÜZÜN** Istanbul Technical University
Co-advisor :	**Prof. Dr. Fethiye Aylin SUNGUR** Istanbul Technical University
Jury Members :	**Assoc. Prof. Ayşe Özge Kürkçüoğlu LEVİTAS** Istanbul Technical University
	Assoc. Prof. Özge ŞENSOY Istanbul Medipol University
	Assoc. Prof. Arzu HATİPOĞLU Yıldız Technical University

Date of Submission : 15 June 2020
Date of Defense : 14 July 2020

To my family,

FOREWORD

First and foremost, it is with genuine pleasure that I would like to thank my thesis advisors, Prof. Nurcan Tüzün and Prof. Fethiye Aylin Sungur, from the Istanbul Technical University. They have consistently allowed this dissertation to be my work but steered me in the right direction whenever necessary. I gained lots of experiences from Prof. Nurcan Tüzün, and she supported me not only in science but also in life since my bachelor times.

It is my privilege to thank Prof. Mati Karelson from the University of Tartu for letting me study under his supervision, and sharing with me other perspectives in the fields of Chemistry.

I want to express my gratitude to PhD students Larisa Ivanova and Selma Zukic from the University of Tartu, for showing me what I can learn in a research team and in so little time. They were always there when I needed moral support and guidance.

A special thanks to the most thoughtful person in the world, Filipe Sena, who supported me unconditionally and made sure I had all the motivation and tools necessary whenever I lacked them.

It is with great pleasure that I thank my family, for their the fantastic help and availability to support me through all hardships I've been through to complete this work. This thesis would not have been possible without them. They supported me always when I needed their help and companion.

July 2020

Gülbahar BOZAN
(Chemist)

TABLE OF CONTENTS

Page

ABBREVIATIONS

ADME/Tox **:** Absorption, Distribution, Metabolism, Excretion- Toxicity
ATF1 **:** Cyclic AMP-dependent transcription factor ATF-1
ATP **:** Adenozin 3'-trifosfat
CLA : Cantharidin
DBD **:** DNA binding domain
DNA **:** Deoxyribonucleic acid
HSE **:** Heat Shock Element
HSF1 **:** Heat Shock Factor 1
HSP **:** Heat Shock Protein
HSR **:** Heat Shock Response
HTVS **:** High-Throughput Virtual Screening
LZ : Leucine zipper
MM-GBSA : Molecular Mechanics Generalized Born Surface Area
mRNA : messenger RNA
NMR : Nuclear Magnetic Resonance
PTM : Post Translational Modifications
RB : Rotatable Bonds
RD : Regulatory Domain
RHT : Rohinitib
RMSD : Root mean square deviation
RMSF : Root mean square fluctuation
TAD **:** Trans-activation domain
QSAR : Quantitative structure-activity relationship method

SYMBOLS

ALA **:** Alanine amino acid
ASN **:** Asparagine amino acid
ARG **:** Arginine amino acid
GLN **:** Glutamine amino acid
GLY **:** Glycine amino acid
PHE **:** Phenylalanine amino acid
HBA **:** Hydrogen Bond Acceptor
HBD **:** Hydrogen Bond Donor
HIS : Histidine amino acid
IC_{50} : The half maximal inhibitory concentration
kDa **:** kilo dalton
LYS **:** Lysine amino acid
MD **:** Molecular Dynamic
MW **:** Molecular Weight
ns **:** nanosecond
P **:** pressure
PRO **:** Proline amino acid
SER **:** Serine amino acid
T **:** Temperature
TYR **:** Tyrosine amino acid
VAL **:** Valine amino acid
ΔG **:** Gibbs Free Energy
ΔS_{conf} **:** Conformational entropy
Å **:** Angstrom

LIST OF TABLES

Page

LIST OF FIGURES

Page

MOLECULAR DOCKING STUDY ON THE DNA BINDING DOMAIN OF MASTER REGULATOR OF THE MAMMALIAN HSR PROTEIN

SUMMARY

Heat shock factor 1 (HSF1) is the primary regulator of heat shock response (HSR) in mammalian cancer cells and have been described as a family of transcription factors which are activated with stress conditions. Upon stress, HSF1 upregulates the heat shock proteins (HSP) that help misfolded proteins to refold for the cell functions. One of the heat shock proteins, HSP90, acts as chaperon in cancer cells as well and supports the proliferation and repair of cancer cells. For years, many efforts have been devoted to studies for finding molecules that inhibit these chaperon proteins but at the same time present drug properties, act compatibly with the physiological conditions with the lessened side effects. For this reason, in recent years, drug design studies have mainly concentrated on inhibiting proteins that trigger the process or production of HSP90. As a result of HSF1-HSE complex, formed by HSF1 and the promoter part of DNA, which is called Heat Shock Element, the unwanted heat-shock proteins are synthesized in the cancer cells. HSF1 protein has emerged as a potential target and in recent years shown as a target for specifically breast, pancreas and prostate cancers.

In the last decade, in silico drug design has been an integral part of drug development processes. Likewise, in this project, it was planned to design molecules that inhibit HSF1 protein via in silico methods. Herein, our target was to search for inhibitors via in silico methods and bring forward candidate lead molecules to literature. The starting point for the determination of the docking protocol was to calculate the docking scores of the ligands with the most effective IC_{50} value and to ensure qualitative agreement by comparing the binding energies calculated with the experimental IC_{50} values. This also enabled us to recognize the residues involved in the binding.

The candidate inhibitor structures of the library formed by virtual screening were studied by the receptor (selected flexible residues) - flexible ligand docking procedure. The candidate molecules with the best binding scores were refined by the MM-GBSA method. The molecular dynamic simulation was further performed for nine selected compounds. From the results, we propose candidate lead molecules that can effectively act in cancer cells, especially in breast, pancreas and prostate cancers.

BAŞ DÜZENLEYİCİ MEMELİ HSR PROTEİNİN DNA BAĞLANMA BÖLGESİ ÜZERİNE MOLEKÜLER KENETLENME ÇALIŞMASI

ÖZET

Isı şok faktörü 1 (HSF1), kanser hücrelerinde ısı şok tepkisinin (HSR) bir ana düzenleyicisidir ve stres koşulları ile aktif olan transkripsiyon faktörlerinin ailesinin bir üyesi olarak tanımlanmıştır. Stres üzerine HSF1, yanlış katlanmış proteinlerin hücre fonksiyonları için yeniden katlanmasına yardımcı olan ısı şoku proteinlerini (HSP) düzenler. Şaperon proteinlerinden olan HSP90 kanserli hücrelerde de şaperon görevi yapmakta ve üremesi istenilmeyen kanserli hücrelerin onarılmasını ve çoğalmasını desteklemektedir. Uzun yıllar bu şaperon görevi gören proteinleri inhibe eden, ilaç olabilme özelliklerini taşıyabilen, yan etkileri en aza indirgenmiş ve fizyolojik koşullara uyumlu yapılar elde edilememiştir. Bu nedenle, son yıllarda ilaç tasarımı çalışmaları HSP90'ı inhibe etmek yerine HSP90'ın üretimini ya da çalışmasını tetikleyen proteinler üzerine yoğunlaşmıştır. HSF1 proteininin DNA'nın HSE (ısı şok elementi) adlı promotör kısmı ile HSF1-HSE kompleksi oluşturması sonucu kanser hücresinde bulunmasını istemediğimiz ısı şok proteinleri sentezlenir. Bu nedenle, HSP90'ın çalışmasını sağlayan HSF1 proteini potansiyel bir hedef olarak tanımlanmış olup son yıllarda özellikle meme, pankreas ve prostat gibi kanser türlerinin tedavisinde hedef olarak gösterilmektedir.

Bilgisayar destekli ilaç tasarımı özellikle son 10 yılda ilaç geliştirme süreçlerinin ayrılmaz bir parçası olmuştur. Bu proje kapsamında HSF1 proteinini inhibe edebilecek olası moleküllerin çeşitli in silico yöntemler kullanılarak tasarlanması planlanmaktadır. Burada, HSF1 proteininin aktif bölge dinamiklerinin daha detaylı incelenmesi, in silico yöntemlerle inhibitörlerinin modellenmesi ve aday öncü yapıların literatüre kazandırılması hedeflenmektedir. Kenetlenme protokolünün belirlenmesi için başlangıç noktası ise en etkin IC50 değerine sahip moleküllerin kenetlenme skorlarının hesaplanması ve deneysel IC50 değerleri ile hesaplanan bağlanma enerjilerinin karşılaştırılarak kalitatif uyumun sağlanması olacaktır. Bu şekilde bağlanma sırasında rol oynayan rezidülerinde tanımlanması da gerçekleştirilecektir.

Devamında oluşturulan inhibitör aday kütüphanesindeki yapıların, reseptör (seçilmiş hareketli rezidüler) esnek-ligand prensibi ile kenetlenmesi çalışıldı. Son olarak en iyi skorlama değerine sahip ilaç adayları için MM-GBSA yöntemi ile bağlanma enerjileri rafine edildi. Moleküler dinamik simülasyon ayrıca seçilen dokuz bileşik için gerçekleştirildi. Böylelikle, özellikle meme, pankreas ve prostat gibi çeşitli kanser türlerinin tedavisinde kullanılabilecek ilaç öncüsü moleküller önerilmiştir.

1. INTRODUCTION

Based on World Health Organization research results, every one in six deaths is happening by the cause of cancer. The report with the number of 9.6 million deaths in 2018, cancer has been the second crucial reason of death (World Health Organization). Eighteen years ago, Darnell (2002) mentioned the activity of some of the transcription factors that increased during human cancer cell processes (Darnel, 2002). He suggested that transcription factors might promising targets for cancer therapy. However, transcription factors were described to be "undruggable" targets (Lambert et al, 2018). Recently, it is known that transcription factors can be inhibited by blocking their protein-protein interaction domains or their DNA binding sites to prevent transcription process directly. Lately, inhibitors which are found to block these domains are in clinical trials or being decided to use for cancer treatments. Some of the inhibitors are investigated to take the transcription factor to their degradation (Naidu, 2017). Today, exploring new inhibitors are simplified by computational chemistry which is rapidly developing with new technologies. In addition, due to enhanced knowledge in protein and a variety of small-molecule databases, it is possible to have more efficient ways of drug designing for transcription factors (Lambert et al, 2018).

One of the significant transcription factors in the human cancer cell is Heat Shock Factor 1 (HSF1), which is the master regulator of heat shock responses. HSF1 supports cancer cell survival, proliferation and transformations. High level of HSF1 is found in many types of cancer cell lines. These cancer cells were observed mostly in breast, colon and lung tumours taken from human patients (Mendillo et al, 2012). These studies showed that HSF1 is a promising cancer therapeutic target. Therefore, researchers were exploring new inhibitors to target HSF1 to block its active domains for transcription processes. There were many inhibitors investigated, but not all of them could pass the clinical trials. The inhibitor has to be water-soluble, and should not be poisonous if it accumulates in the body. Studies in the literature have shown

that the most efficient results are obtained from structures that inhibit HSF1 by directly interacting to DNA-binding domain (Sharma and Seo, 2018).

As stated above, the design of new inhibitors is still needed, and studies in this area are of great importance.

In this study, an in-silico analysis was performed to find inhibitors for the HSF1 protein, which has been stated as a promising candidate against various cancers. To this end, a virtual screening protocol was applied to select inhibitor candidates with tailored properties. The virtual screening was followed via different branches which involved a library formed from ZINC and PubChem databases. The libraries included small molecules, leading compounds, and FDA approved drugs. In one hand, small-molecule ligands were investigated; on the other hand, repurposing of drugs was examined. The structures that showed high binding to HSF1 via docking studies were further analyzed for their protein-ligand binding energies from MD simulations. The inhibitor candidates obtained from this study are expected to have potentials for preventing the binding of HSF1 to HSE, thereby preventing the growth and reproduction of cancer cells from the formation of chaperones produced by binding. The leading molecules that will be proposed will contribute to the studies in the field of cancer drugs.

2. LITERATURE REVIEW

2.1 History of Heat Shock Proteins

Temperature effect in cells was first revealed by Ritossa in 1962. He proved that change in the temperature was affecting the fruit flies or in another word Drosophila melanogaster. One of his colleagues increased the temperature by accident, and Ritossa found a new pattern of puffs in these fruit flies. The active site which is "puffs", on the salivary gland chromosomes of Drosophila melanogaster was found for transcription. In some cases, the loss of puffs was occurring, and on the other hand, puffs were existing in other places (Ritossa, 1996; Aufricht, 2005). Ten years later, this reaction against the stress in the cell was investigated at the protein level by Tissiéres et al, (1974) and heat shock proteins were explored. Proteins could be finally observed in details by radioactively labelled amino acids. In their research, it is observed that, under stress conditions, the normal protein level was decreasing while the number of a new type of proteins was increasing. In fact, these new proteins that appeared against stress were named after the heat, which was the first found stress, heat shock proteins (HSP). A stress condition for a cell can be defined as extreme of high temperature, high concentration of ions, radiation, cellular energy depletion and toxic substances. Most of these stress conditions are inducers of HSP. However, they can cause denaturation of a normal protein (Aufricht, 2005). In 1997, Gething reported that heat shock proteins can be classified according to their molecular weight, i.e. hsp110, hsp100, hsp90, hsp70, hsp60, hsp40, hsp10, and small HSP families. In Gething's study, seven extra families and undescribed 12 genes/proteins were observed.

Ellis and Van der Vies (1991) described stress proteins a family of 'molecular chaperones' as essential components for protein maturation. Therefore, these heat shock proteins are upregulated in the cell under stress conditions. Additionally, under normal conditions, heat shock proteins are interacting with other proteins which are not folded. Substantially, the viral thing about HSPs is that the interaction between

HSP and non-folded protein is temporary, and when the target protein is finally in its properly folded state, HSP is also released for the next round of interactions. However, when the cell goes under metabolic stress, the target protein and HSP can make a stable complex and remain as a complex (Beckmann et al, 1992).

Hightower (1980) suggested that under stress conditions when the protein denaturation started, a stress response also started. Following Hightower's works, denatured proteins were injected to cells to observe the activation of stress responses. (Ananthan et al, 1986).

Tumour cells produce more heat shock proteins than the normal cells (Jäättelä et al, 1999). This research showed that tumour cells need heat shock proteins to live longer (Westerheide et al, 2005). Some type of passive cancer vaccines was produced based on this reason with the logic of enhanced immune response. Cancer vaccines which are not used for killing tumour cells are specific active immunotherapy agents that use purified peptides and other antigens in the cell. HSP-based cancer vaccines have been used in many phases in cancer therapy, and still, many of them are being tested in clinical trials. These vaccines were made of purified peptides and then injected to patients with fibrosarcoma, hepatoma, colon cancer, melanoma, lung cancer, lymphoma and prostate cancer (Wang H. H. et al, 2006; ClinicalTrials.gov).

Today, researchers are investigating to find new inhibitors for HSP90, HSP70, HSP40 types of heat shock proteins. Heat shock proteins ATP-binding pockets are good targets for inhibitors which are used for cancer therapy. These proteins are upregulated by heat shock factors (HSF) in the cell (Aşkar et al, 2007; Öztürk et al, 2009).

HSP90, which is one of the heat shock proteins, is located in the cytoplasm and endoplasmic reticulum in the cell. HSP90 regulates the folding, activation and status of HSF1 proteins. Heat shock factors act as transcription factors in protein synthesis that takes place inside the cell. These factors bind to the genes necessary for the synthesis of chaperone-function proteins such as HSP90 and HSP70. These genes are in the promoter part of DNA called HSE (heat shock element). Transcription factors play an essential role in the realization of mRNA synthesis, followed by protein synthesis. HSF1, weighing 57 kDa, is inactive in the cytoplasm when it is bound to the HSP90 protein, it becomes free as a monomer as a result of stress, such as temperature or pressure, and turns into a trimer structure. In humans, when the temperature of the

cells rises to 41-42°C, heat shock proteins in the cell are stimulated. With the increasing temperature, released active HSF1 helps transcription in the cell nucleus to synthesize heat-shock proteins. HSF1 is a primary regulator of heat shock proteins (Naidu and Dinkova-Kostova, 2017).

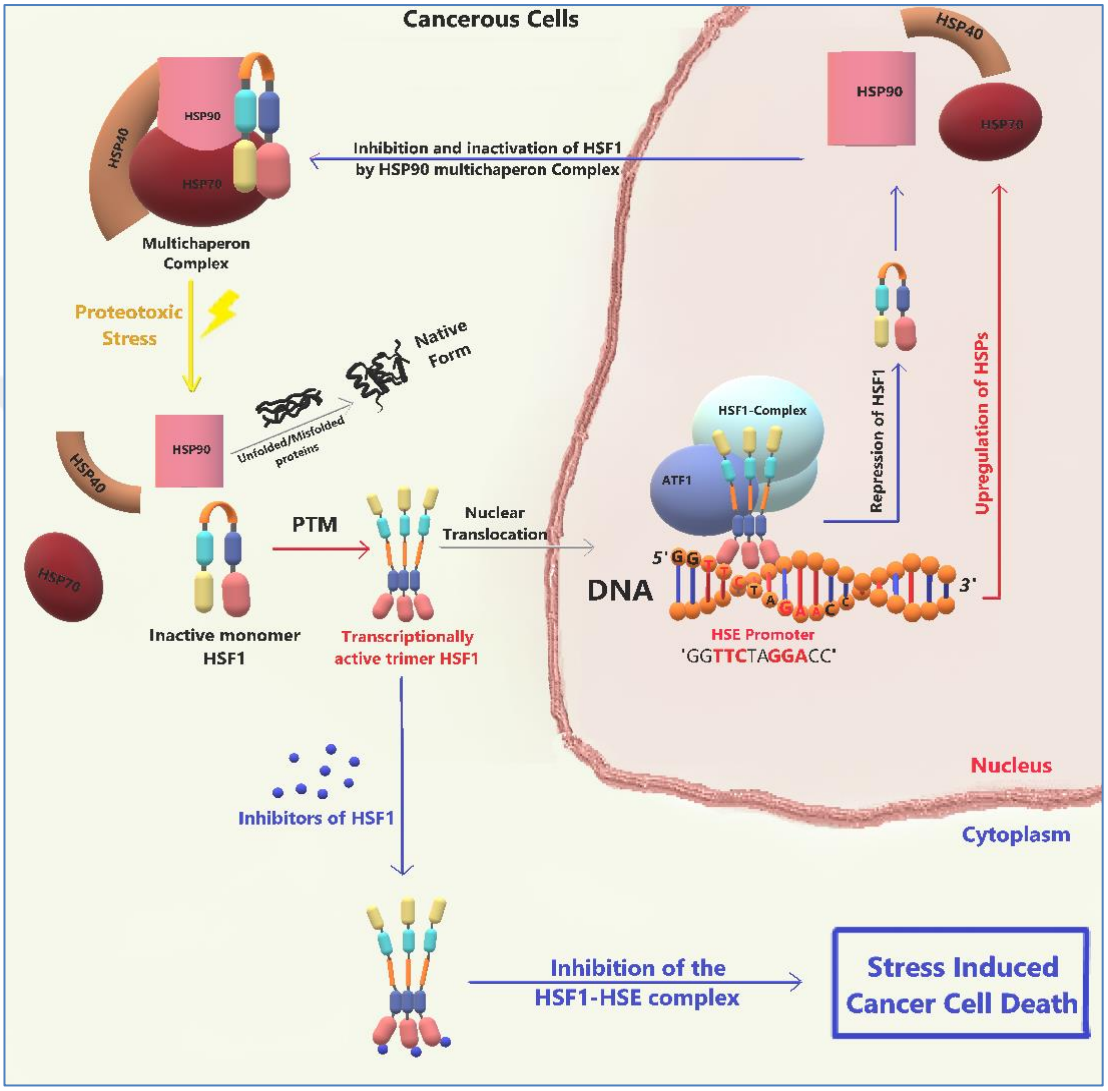


Figure 2.1: Cancer cell survival under stress. Upon stress, HSF1 is released from multichaperone complex of HSP90. Active HSP90 as chaperon machinery can fold the proteins which are unfolded or misfolded in the cells. Monomer and inactive HSF1 goes through post-translational modifications and becomes active homotrimer HSF1. After relocation, active HSF1 binds to HSE promoter on the DNA in the nucleus and upregulates heat shock proteins. In the end of upregulation, HSF1 repression and relocation to the cytoplasm, respectively goes on and terminates the cycle with binding to HSP90 multichaperone complex again.

As a result of HSF1 forming the HSF1-HSE complex with this part of DNA, mRNA synthesis is performed, and the synthesis of heat shock proteins is completed from there. These chaperones produced by the binding of HSF1 to HSE are proteins that are

not wanted in the cancer cell (Jäättelä, 1999). These proteins cause proliferation and repair of the proteins in cancer cells. If an inhibitor can inhibit the HSF1-HSE relationship, heat shock protein production in the cancer cell may also be stopped**.** (Figure 2.1) (Sharma & Seo, 2018).

2.2. HSF Family Members

Four types of heat shock proteins were identified as the family members that are HSF1, HSF2, HSF3, HSF4, HSF5, HSFX and HSFY (Schuetz et al, 1991; Nakai et al, 1997; Nakai et al, 1995; Xu et al, 2012). Helix-turn-helix DNA binding domain is the only common thing between all HSF family members (Xu et al, 2012). While HSF1, HSF2 and HSF4 have been analyzed in mammals, HSF3 was determined in the avian species. During unstressed conditions, HSF1 is a monomer while HSF2 is a dimer protein. Under stress, HSF2 helps the upregulation of the heat shock genes like tandem satellite III DNA repeats (Sandqvist et al, 2009; Ostling et al, 2007).

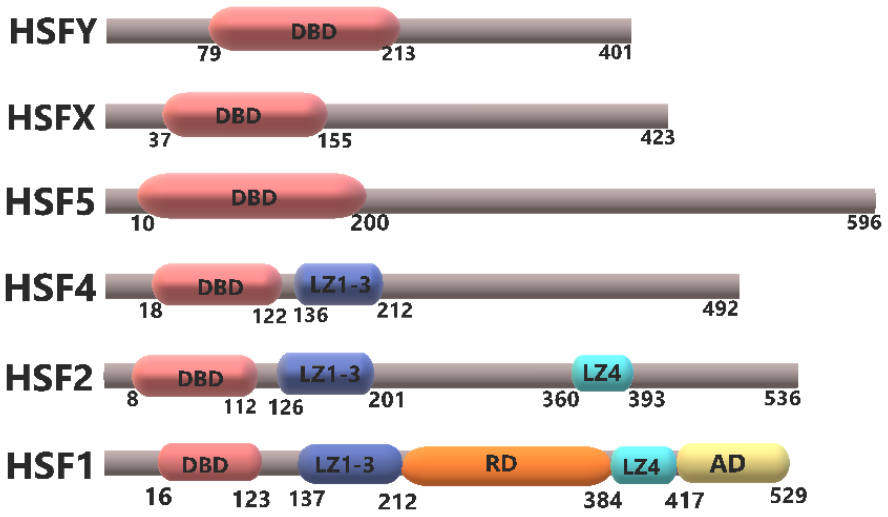


Figure 2.2: Schematic diagram of human heat shock factor family members. The domains which are the functional parts of the proteins are shown as follows: DNA-binding domain (DBD), the leucine zippers oligomerization domains (LZ1-3 and LZ4), the regulatory domain (RD) and the transactivation domain (AD).

HSF1 and HSF2 have connections from their coiled-coil domains close by DNA binding domains (Jaeger et al.,2016). HSF4 has a crucial role in the development of human cataracts (Fujimoto et al. 2004). Lastly, HSF3 is the fundamental transcription factor in the chickens (Kawazoe et al., 1999). From all heat shock factor members, HSF1 is the most studied one. While HSFX is named after finding it only on the X

chromosome, for the same reason, HSFY was found on the Y chromosome (Figure 2.2).

2.3. Structure and Function of HSF1

In general, it is known that under normal conditions, HSF1 is a monomeric phosphoprotein which is a complex structure with molecular chaperones such as Hsp70, Hsp90 etc. At the same time, HSF1 transcriptional activity and DNA binding capability are suppressed by these chaperones. Upon stress, HSF1 is released to the cytoplasm as an inactive monomer. Inactive HSF1 goes through many post-translational modifications (PTM), relocates to the nucleus and becomes active trimer structure (Naidu and Kostova, 2017).

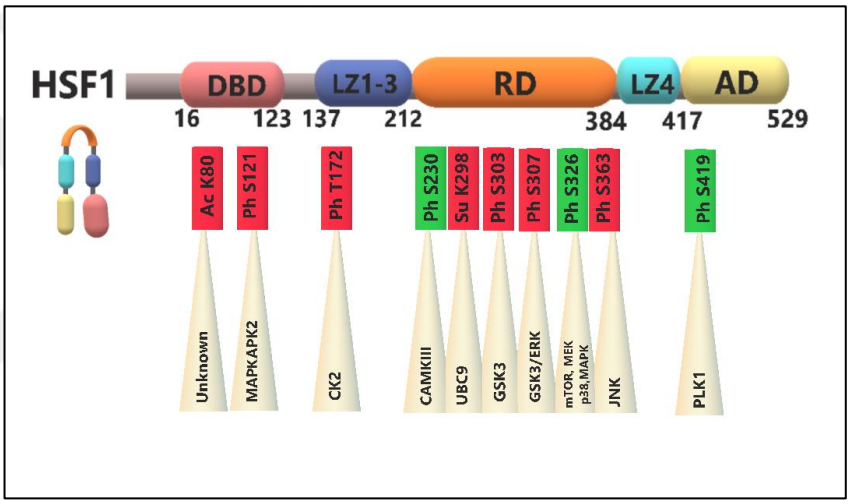


Figure 2.3: The repressive modifications on HSF1 can be seen on K80, S121, T172, K298, S303, S307 and S363 residues (red) and activating modifications shown above on S230, S326 and S419 residues (green). Also, the enzymes of these modifications are shown after residue names. The binding between the leucine zipper (LZ) domains LZ1–3 and LZ4 of HSF1 is disrupted, and the regulatory domain (RD) becomes hyperphosphorylated. C-terminal transactivation domain (AD) of HSF1 can inhibit its function (Figure is taken from Naidu and Kostova, 2017).

Acetylation, Sumoylation and Phosphorylations are some of the main PTMs of HSF1. Some of the protein kinases cause the activation or repression modifications on HSF1 specific residues. At standard temperature, negatively regulated HSF1 structure has intramolecular interactions between Leucine zippers which are the LZ1-3 domain and the LZ4 domain. Upon heat shock, homotrimer structure of HSF1 is formed by intermolecular interactions from LZ1-3 domain, while intramolecular interactions disappear (Rabindran, 1993; Farkas, 1998; Liu, 1999). The Sumoylation occurs at Lysine 298 residue on HSF1, which will be transcriptionally incapable after the

modification. A similar situation happens on the acetylation modification at Lysine 80 residue, and it interferes the transcription activity of HSF1 (Figure 2.3).

The Phosphorylation of HSF1 mostly appears on the Regulatory domain (RD) which is in the between LZ1-3 and LZ4 domains and known to be on serine residues. Even under unstressed conditions, lack of RD domain enables HSF1 to be active for the transcriptions. Most of the phosphorylation modifications on RD domain makes HSF1 transcriptionally inactive. Only phosphorylation modifications on Serine 230 and Serine 326 residues increase the activation of HSF1 by the specific enzymes such as DAXX and start the heat shock gene expression. (Naidu and Kostova, 2017).

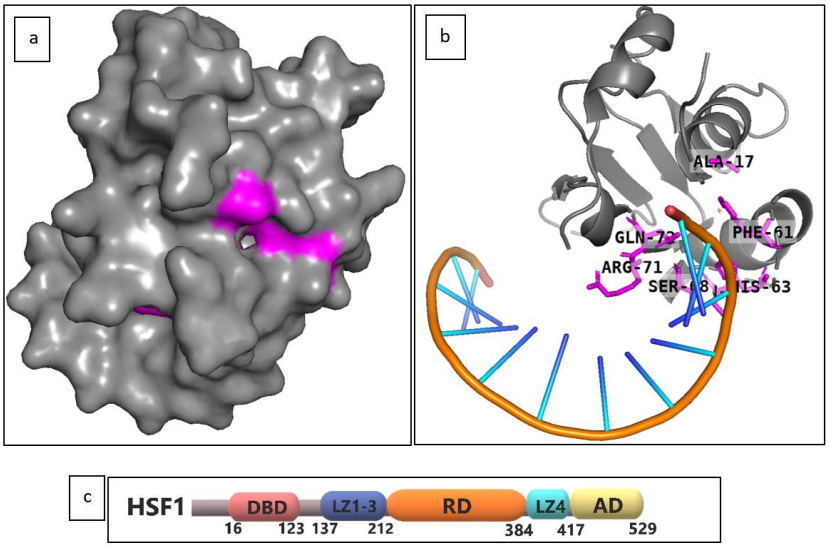


Figure 2.4: Crystal structure of HSF1-HSE complex. a) The binding site of HSF1 to HSE is shown as pink colour and by the surface application. b) The active site of the protein is shown as pink colours and labelled with residue names (PDB: 5D5U). c) HSF11 protein regions are shown as residue numbers (Neudegger et al, 2016).

C-terminal transactivation domain (AD) has less information than the other domains on HSF1 in the literature. It has two main domains which are TAD1 and TAD2. Mutation of TAD1 reduces the activation of HSF1. The domains, TAD1 and TAD2 are rich with hydrophobic amino acids and acidic residues, respectively. DBD domain of HSF1 is highly stable N-terminal DNA binding region. Also, this domain recognizes the most important sequences of DNA which is on the HSE promotor (Naidu and Kostova, 2017).

Neudegger et al. (2004) investigated the crystal structure of HSF1 in complex with DNA. They showed the interactions and the possible hydrogen bonds between HSF1 and DNA. The stabilization of DBD domain is under control by Phe104, Leu112 and

Ile115 residues with hydrophobic interactions and hydrogen bond between Met75 and Lys118 residues. Active HSF1 interacts with the pentameric sequence of tail-to-tail heat shock element (HSE) which is described as ggTTCtaGAAcc (Capital letters are HSF1 recognition bases). Residues which interact with HSE promoter site of DNA are Ser68, Arg71, Gln72, Asn74, Met75 and Tyr76. In the promoter, two thymine bases have van der Waals interactions with the side chains of Ser68, Arg71 and Gln72. Acetylation on lysine residues Lys80, Lys116 and Lys118 changes the transcriptional activity of HSF1 to interact with DNA backbone by removing the positive charges (Neudegger et al, 2016).

HSF1 has mainly regulator and transactive regions that bind to DNA to oligomerize. Only the region of HSF1 that binds to DNA takes part in transcription. This part has a crystal structure and is shown in Figure 2.4c. As can be seen in Figure 2.4c, the number of amino acids in the crystal structure between 16 and 123 are in the region of the DNA binding domain. Upon inhibition of this region, HSF1-HSE complex will not be formed, and the synthesis of the proteins in cancer cells will be stopped. For these reasons, HSF1 has been an effective target in cancer treatments such as breast, pancreas and prostate cancers. Literature data have shown that HSF1 is a better target and inhibition protein than HSP90 to fight against cancer (Sharma and Seo, 2018; Li et al, 2017).

Cantharidin (CLA) (Kim et al, 2013) and Rohinitib (RHT) (Santagata et al, 2013; Iwasaki et al, 2016) molecules were already known from traditional Chinese medicine. Although they are inhibitors of HSF1, both of them has low ligand efficiencies. In order to enhance the ligand efficiency, hybrid ligands of CLA and RHT were suggested by Agarwal et al. (2015) and those hybrid ligands were produced randomly and by changing the parts which do not have pharmacophoric elements of CLA and RHT. Hybrid ligands had better affinity than CLA and RHT. Homology modelling and structural validation of protein were studied before docking calculations. From all the suggested ligands, RC15 was found to have the best docking score and the ligand efficiency. Ligands specifically had hydrogen bonds with Ser68 and Gln72 residues. In addition to high ligand efficiencies, they were tested for ADME/Tox features as well. The molecular dynamic simulation was also carried out to find the active site which could interact with HSE promoter. Using different methods, Agarwal et al. (2015) have made theoretically active site estimation from cantharidin, rohinitib and

their hybrid ligands. The active site defined by Agarval et al. includes the amino acids of Ala17, Phe61, His63, Asn65, Ser68, Arg71 and Gln72, it is shown in Figure 2.4 (a,b) (Agarwal et al, 2015).

In stress conditions, HSF1 is known to have many different types of complexes with different proteins. Some of these proteins are CHIP (Dai et al, 2003), DAXX (Boellmann et al, 2004) and ATF1 (Takii et al, 2015). ATF1, which affects the DNA-binding domain of HSF1, makes the crucial transcription complex with HSF1 to regulate heat-induced genes. Therefore, inhibiting ATF1-HSF1 interaction domain can change the HSF1 interaction with DNA. ATF1-HSF1 interaction domain, which has been experimentally found by Vilaboa et al. (2017) and called cavity A, contains amino acids Val70, Lys80, Thr97 and Phe99. After eliminating many inhibitor candidates, they came up with $I_{HSF}115$ inhibitor, which is with the lowest IC50 value (0.7 +/- 0.1 μM) and the best docking pose. However, their docking result did not have any formal hydrogen bonds. Experimentally, they found out that $I_{HSF1}115$ molecule can strongly bind HSF1 DNA domain but with unknown affinity and it could kill the cancer cells. Based on their results, $I_{HSF}115$ prevents the formation of the ATF1-HSF1 complex, which is a fundamental structure for heat-induced genes transcription. The results showed that even though $I_{HSF}115$ inhibitor is not interfering with HSF1 oligomerization and does not reduce HSF1 binding to its chromatin, it is the only inhibitor found to bind directly to HSF1 with substantial shreds of evidence (Vilaboa et al, 2017; Takii et al, 2015).

2.4. Small Molecule Inhibitors of HSF1

To date, many inhibitors have been experimentally synthesized and recommended to inhibit HSF1 and prevent it from being active (Sharma and Seo, 2018). Quercetin is a class of natural inhibitors from the family of flavonoids. In the work of Nagai et al. (2004), quercetin obtained from natural flavonoids synthesized from plants, prevents the binding of HSF1 to the target DNA and stops the operation of heat shock gene transcription (Nagai et al, 1995). Although quercetin is an effective inhibitor, there are side effects and low water solubility (Miles et al, 2014). QC12 inhibitor has been tried, but no further progress has been made after 1st phase trials (Mulholland et al, 2001). Stresgenin B and CL-43 have been studied as an inhibitor for heat shock proteins. Although Stresgenin B is similar to quercetin in its biological activity, the inhibition of heat shock factors with CL-43 is more effective, and its water solubility is high.

Stresgenin B is from the chemical class of carboxamide (Akagawa et al, 1999). Another structure is from the diterpene epoxide family called Triptolide which inhibits molecular chaperones and causes the death of pancreatic cancer cells. Since triptolide, like quercetin, has low solubility in water, it had to be converted into a water-soluble molecule to prevent it from being toxic (Phillips et al, 2007). As a result of these studies, an inhibitor called Minnelide was found. (Chugh et al, 2012). This inhibitor has been shown to be highly effective and is currently being used after FDA approval in the treatment of pancreatic cancer. Fisetin, another inhibitor, prevents the binding of HSF1 to HSE and stops the upregulation of the chaperon proteins (Kim et al, 2015). Since Fisetin is a very reactive molecule, it was thought to have many side effects as a medicine, but the tests were continued as it has an anti-cancer effect. This molecule, a drug candidate, is in the 2nd stage of clinical tests. Another inhibitor was found to be 2,4-bis (4-hydroxybenzyl) phenol which could cause HSF1 protein to degrade by suppressing the heat shock proteins and stimulating phosphorylation in S326 residue. As a result, HSP27 and HSP72 protein amount decreased (Yoon et al, 2014).

Hybrid ligands of Cantharidin (CLA) and Rohinitib (RHT) were previously mentioned. Rohinitib was produced from the natural compound Rocaglamide A (RocA) which is a member of flavaglines. Even though it is known that RHT prevents the binding of HSF1 to its regulatory genes and the upregulation of HSPs does not happen, RHT inhibition of HSF1 has no clear information about direct binding to HSF1 (Santagata et al, 2013; Iwasaki et al, 2016). CLA is a terpenoid identified as an inhibitor of HSF1 and used as a traditional medicine in China (Figure 2.5).

In colon cancer cells, the level of Hsp27, Hsp70 and BAG3 were decreased by cantharidin inhibitor (Kim et al, 2013). BEZ235 has been characterized as an inhibitor, which is known as reducing tumour growth, and it is orally available. Similar to rohinitib, the mechanism between HSF1 and inhibitor BEZ235 is unknown (Maira et al, 2008). SNS-032 is found in the same analyses of BEZ235 inhibitor. SNS-032 is an effective inhibitor for tumour cells, and its Phase I is completed in the clinical trials.

Furthermore, 4,6-Disubstituted Pyrimidines and SNS-032 are also known inhibitors of cyclin-dependent kinases (CDK), and their mechanisms with HSF1 are still unknown as well (Rye et al, 2016; Chen et al, 2009). KNK437 is another HSF1 inhibitor which is a benzylidene lactam. Even though its mechanism with HSF1 is unknown, it

suppresses breast cancer cell lines (Oommen and Prise, 2012). The same unknown situation is valid for dorsomorphin, which is also an inhibitor of HSF1 (Li et al, 2019).

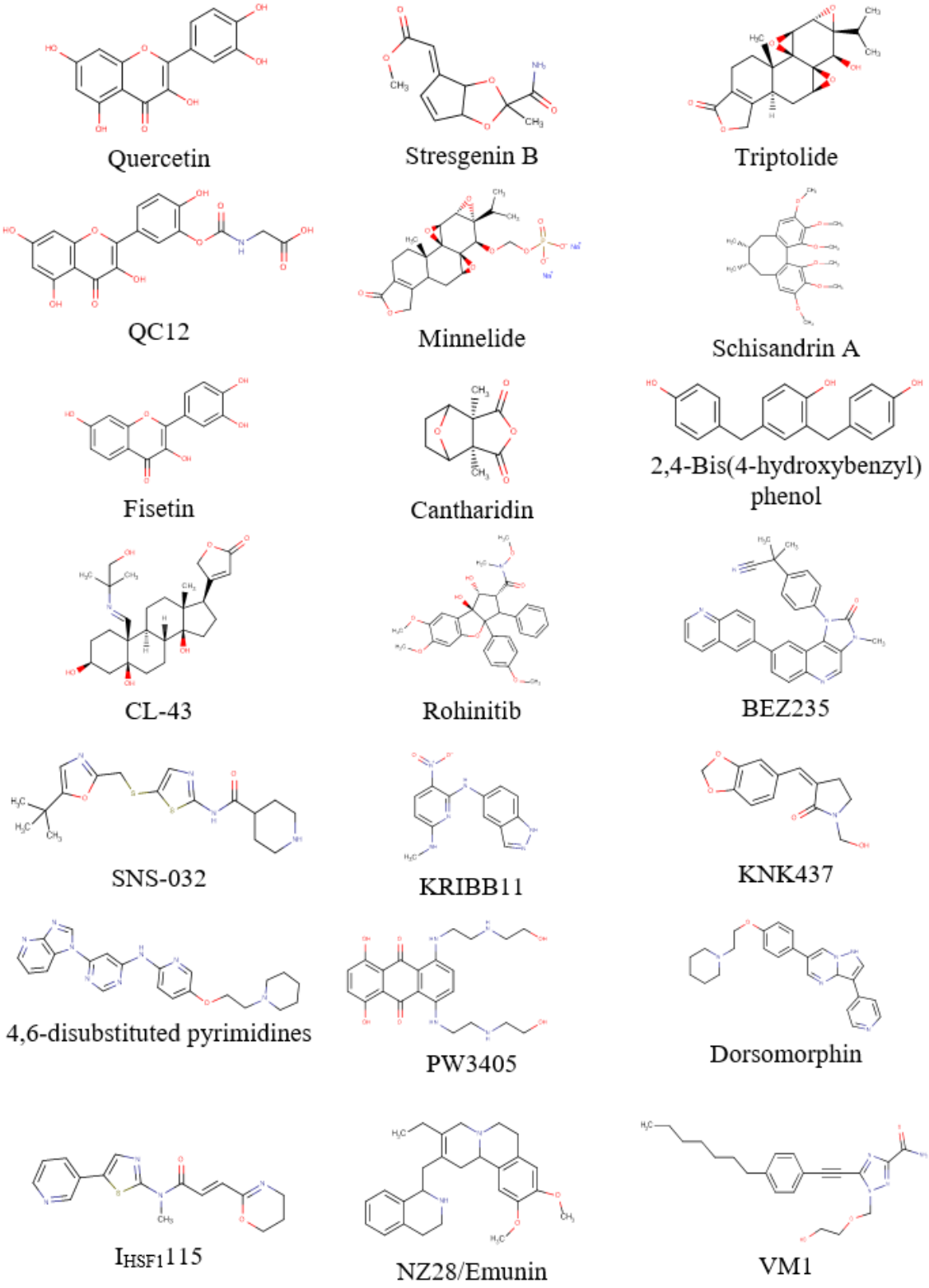


Figure 2.5: Some of the small molecule inhibitors of HSF1.

Other HSF1 inhibitors NZ28, emunin, PW3405, VM1 and KRIBB11 are small molecules which were found to inhibit the 'HSF1 pathway' from large databases (Zaarur et al, 2006; Yoon et al, 2011; Zhang and Zhang, 2016). The specificity of HSF1 was the biggest concern for these compounds. Nitropyridine compound named KRIBB11 has been observed to stop the cancer cells spreading. Unlike others, KRIBB11 was the first inhibitor that binds HSF1 directly as a target. However, the compound is in the class of reverse transcriptase inhibitors because of its lack of specificity (Yoon et al, 2011).

The last known inhibitor is Schisandrin A (Sch A) which is a significant bioactive dibenzocyclooctadiene ligand. It was used to cure any types of diseases in traditional Chinese medicine, mostly for chronic asthma. Lately, the antitumor molecular mechanism of SchA was investigated on human breast and ovarian cancer cells. After all, it has been reported that Sch A is inhibiting colorectal cancer cells by targeting the DBD domain of HSF1 directly. Even though Sch A has a low water solubility, researchers are working on its drug properties (Chen et al, 2020).

The clinical test results passed by the inhibitors were obtained from the ClinicalTrials.gov database, and the inhibitor binding information was obtained from the ChEMBL database.

2.5. Small Molecule Activators of HSF1

There are various types of HSF1 activators depending on the functions in the cell. Activators of HSF1 is essential for upregulation of HSPs and for the function of HSF1 in the cell (Figure 2.6). HSF1-mediated transcription can change with most of the activators, such as 4-Hydroxy-2-nonenal which was used on colon cancer cell lines and conclusion of the experiments indicated the accumulation of HSF1 in the nucleus (Jacobs et al, 2007). Also, other activators of HSF1, which is called 10-nitro-octadecenoic acid, have been shown to activate the HSRs in the cell (Kansanen et al, 2009). When the heart of male Wistar rats was exposed with 15d-PGJ2, upregulation of Hsp70 and DNA-binding activity of HSF1 was fully enhanced (Zingarelli et al. 2004). In addition, similar transcriptional activation of HSF1 happens in the presence of 17β-estradiol (Hamilton et al, 2004; Hou et al, 2010).

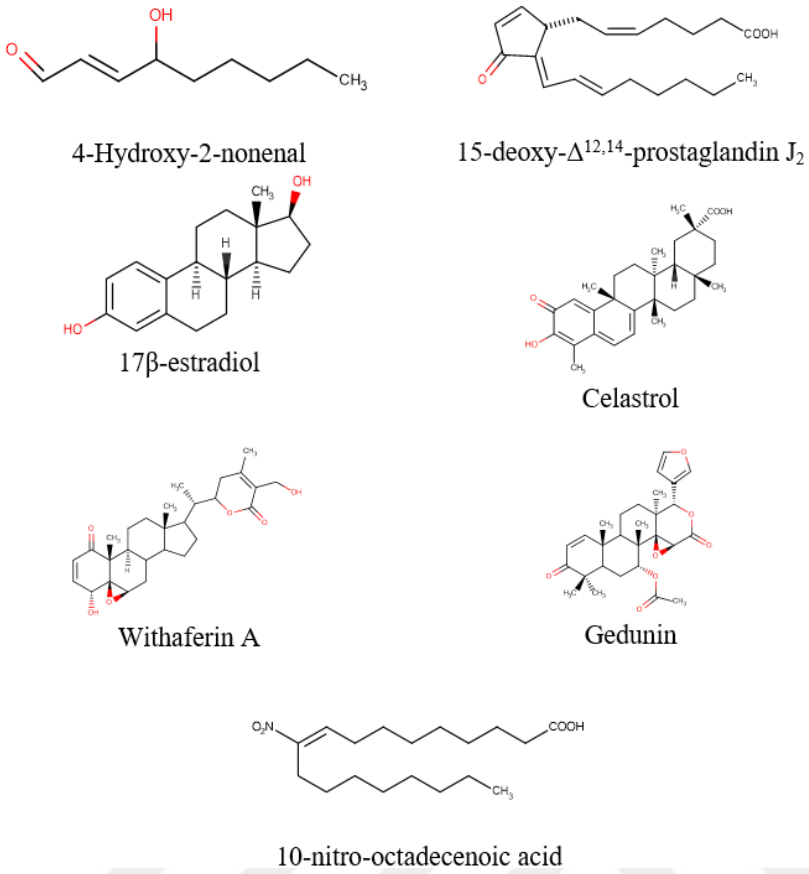


Figure 2.6: Some of the small molecule activators of HSF1.

HSF1 activation can happen during some type of physiological processes in the body. Reactive oxygens can induce the HSF1-mediated transcriptions by causing the hyperphosphorylation of HSF1, relocation to nucleus and upregulation of Hsp70 (Metzler et al, 2003).

Another way of activating HSF1 is going through inhibiting HSP90 chaperones. In this case, dissociated HSF1 can go under the process of activation. The other activators of HSF1, abundant with carbonyl functional groups; withaferin A, celastrol and gedunin have the common function which is inhibiting HSP90, then stimulating HSF1 to be released in the cytoplasm. These activators are also inducers of heat shock responses (HSR) (Santagata et al, 2012; Yu et al, 2010; Grover et al, 2011).

3. THEORY

3.1 Molecular Docking

Docking is the study of estimating the conformation of the ligand structure and their orientation during binding, mostly to the active site of the target protein, from time to time, to the allosteric binding sites. Three-dimensional structures of both the target structure and the ligands are needed for docking. If the target structure is a protein, its 3D structure can be obtained from the RCSB protein database (Research Collaboratory for Structural Bioinformatics Protein Databank), X-Ray, or illuminated by NMR method in PDB format (The Protein Data Bank, 2000). The reliability and image quality of the resulting three-dimensional structure depends on the resolution of the X-ray structure (Irving et al, 2001). Basically, molecular interlocking programs use repetitive search algorithms until the confirmation of the ligand converges to minimum energy. Subsequently, an affinity scoring function, in ΔG [kcal/mol], is used to rank candidate poses (Pagadala et al, 2017). The structure of the protein during the docking process can be defined as rigid or flexible. In standard virtual placement studies, flexible ligands are placed on a fixed (rigid) receptor. Fixed receptor-flexible ligand clamping produces a large number of clamped conformations with surface complementarity, thereby calculating affinity values and reordering structures (Jackson et al, 1998; Moon and Howe, 1991; Rotstein and Murcko, 1993a, b; Nishibata and Itai, 1993). Over the years, the side chain flexibility found to play a crucial role in ligand-receptor complexes. These elasticities allow the side chains of the residues in the binding site of the receptor to vary according to the orientation of the ligand. To find good docking poses, the search algorithms are used, and energy was estimated by different types of scoring functions. Efficiency and selectivity are the most important properties for these scoring functions (Teodoro et al, 2001). Efficiency is vital for the time of the docking process, and selectivity is crucial for the good docking poses. Nowadays, the efficiency of the docking results depends on the speed of computers and their costs.

3.2 Docking Methodology

The most popular docking methodology is rigid protein-flexible ligands. Recently, researchers are developing flexible docking protocols with flexible protein and flexible ligands. Flexible docking is a proper search only for few numbers of compounds for specific conditions like certain pH and solvation. Because of these limits, before starting to run a docking program, it is important to know that the program depends on the amount of the compounds from the database and the type of computer hardware (Hernández-Santoyo et al, 2013; Teodoro et al, 2001). It is essential to see the result of a docking procedure with atomic representation for all interactions (Figure 3.1).

3.2.1 Search algorithms

The search algorithms which are used to find suitable poses of the receptor-ligand complex are basically divided into three: Shape Matching, Systematic Search and Stochastic Algorithms.

3.2.1.1 Shape matching

This method matches the ligand to the binding site of the protein with a proper shape similarity. Here, protein and ligands are studied as pharmacophores, and the distance between each of these pharmacophores are measured (Meng et al, 2011).

Docking programs which use shape matching algorithm are DOCK, LigandFit, MDock, Surflex, etc.

3.2.1.2 Systematic search

The systematic search algorithm aims to scan all possible binding conformations of the ligand. The most used is the systematic rotation of all rotatable ties at specified intervals. However, due to the scanning of many conformations, it needs intensive use of computer time, e.g. Glide and FRED (Kitchen et al, 2004). Most of the time, systematic searches are tried to be solved by introducing certain restrictions. The second method, followed by systematic search algorithms, is fragmentation in which, that is, the ligand is pieced into the receptor region and recombined with covalent bonds after scanning all possible conformations with these fragments (Leach and Kuntz, 1992). DOCK program uses this algorithm for searching (Ewing et al, 2001; Allen et al, 2015). The last method used in the systematic search is conformational

ensemble methods. In this method, docking programs like OMEGA generates ligand conformations to find out ligand flexibility. Afterwards, binding modes of the ligands are arranged based on their binding energies (Roy et al, 2015).

3.2.1.3 Stochastic algorithms or random search

There are several types of docking programs which use stochastic algorithms: Monte Carlo (MC) methods, Evolutionary algorithms (EAs), Swarm optimization (SO) methods.

Metropolis Monte Carlo is a method looking for structures with global minimum energy. It generates random changes in ligand conformation, thermodynamically investigates different states, and selects energy-appropriate states to form protein-ligand complexes with better binding energy. Simulated Annealing (Kirkpatrick et al, 1983; Goodsell and Olson, 1990; Yue, 1990), on the other hand, is a technique that aims to achieve the global minimum by scanning the conformations at the minimum energy potential by increasing the temperature and then decreasing it while using the Monte Carlo algorithm.

Evolutionary Algorithms, on the other hand, are newer methods than Monte Carlo and Simulated Annealing methods (Morris et al, 2000; Thomsen, 2003). A candidate solution population represented by artificial "genes" has the ability to simultaneously explore the search area and combine existing ones to create new solutions. Thus, more confirmations are scanned at the same time compared to other methods. The AutoDock program uses the local search method called Lamarckian Genetic algorithm (Morris et al, 1998), which is an extension of the genetic algorithm but has improved performance compared to the traditional genetic algorithm. The binding energies of the search algorithms should be estimated and ranked by the scoring functions, and the best binding pose should be determined as a result of this evaluation. Scoring functions basically include electrostatic and van der Waals interactions to evaluate the binding mode of molecules (Sotriffer et al, 2002).

Particle Swarm Optimization method was an inspiration from the social behaviour of birds when they fly and found by Dr. Eberhart and Dr. Kennedy in 1995. There is a particle which flies in the search area and a swarm which is the behaviour of flying birds. Therefore, a ligand movement is based on this particle in a search space as a population of individuals.

3.2.2 Scoring functions

Scoring functions are mathematical functions which can estimate the interactions in the binding pose of the docked complex. Most known interactions that can be calculated in the scoring functions are hydrogen bonding, electrostatic interaction, van der Waals interaction, desolvation effect and loss of torsional entropy upon binding. The scoring functions can be basically grouped into four different types (Huang et al, 2010): Force field/Physics-based (GOLD, AutoDock, DOCK), Empirical (FlexX, ChemScore, SCORE, SCORE 1&2, GLIDE), Knowledge-based potential (PMF, DrugScore, SMoG, ASP, DSX), Descriptor-based (LigScore, SFCscore,RFScore, PHONEIX, ID-Score).

3.2.2.1 Force field/physics-based

Force field-based scoring functions consider non-bond terms of the molecular mechanical force field, i.e. estimate the energy of interaction between protein and ligand. Terms used for force field-based scoring represent electrostatic and van der Waals interactions. The electrostatic terms are predicted by Coulomb formulation, and the term van der Waals are predicted by the Lennard-Jones potential function.

DOCK and AutoDock programs basically calculate scoring functions using Amber force field parameters. (Kuntz et al, 1982; Wang et al, 2004). The scoring function consists of the terms electrostatic and VDW, as shown by equation 3.1.

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \quad (3.1)$$

While r_{ij} represents the distance between the protein atom i and the ligand atom j, A_{ij} and B_{ij} are VDW parameters and q_i, and q_j are atomic charges. Here, the effect of the solvent is indirectly by adding a simple distance-dependent dielectric constant $\varepsilon(r_{ij})$ to the term Coulomb.

Autodock 4.0 program has been using a new semiempirical free energy FF as its scoring function (Huey et al, 2007). An old version of this FF energy function was used by Autodock 3.0. However, new FF free energy has two critical advantages over AutoDock 3.0 FF, i.e., a) thermodynamic model of the process is developed by adding

new intramolecular terms, b) it consists all atom types which call favourable and unfavourable energies, this can be defined a full desolvation model as well.

In AutoDock 4.0, the free energy of binding is calculated with the equation below:

$$\Delta G = \left(V_{bound}^{L-L} - V_{unbound}^{L-L}\right) + \left(V_{bound}^{P-P} - V_{unbound}^{P-P}\right) + V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf} \quad (3.2)$$

The new FF scoring function consists of six pair-wise evaluations of V and missing part of the conformational entropy (ΔSconf). L and P represent 'ligand' and 'protein' respectively.

AutoDock Vina is an advanced docking program, developed after AutoDock software. Compare to AutoDock 4, Vina is running faster, and the results of Vina have higher accuracy. Vina is better for docking big databases. Vina scoring function is investigated to perform with the equation as follows:

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij}) \quad (3.3)$$

The summation is for all the atom pairs that might change relative to each other but generally eliminating atoms which are separated by three sequential covalent bonds. In this equation, for each atom i, t_i is referred and set of symmetric interaction functions, ftitj is described with interatomic distance r_{ij}. Here, the total amount of intermolecular and intramolecular additions is defined as 'c'.

$$c = c_{inter} + c_{intra} \quad (3.4)$$

An optimization algorithm is defined in the following section to investigate a global minimum of 'c' along with other low-scoring conformations.

From the lowest-scoring conformation, the predicted free energy of binding is then calculated and defined as follows.

Here, g function can be inconsistent fully enhancing nonlinear function. Also, the other low-scoring conformations are described as 's' values in the throughput. However, to keep the ranking under control, using 'c_{intra}' of the best binding score is given as:

$$s_1 = g(c_1 - c_{intra1}) = g(c_{inter1}) \quad (3.5)$$

Substantially, all these functions are being used as a parameter for the remaining of the code. The large scale of stochastic global optimization was studied, such as genetic algorithms, particle swarm optimization, and simulated annealing to accelerate the optimization. Ultimately, Iterated Local Search global optimizer, which includes the mutation and local optimization, was used. Broyden-Fletcher-Gordfarb-Shanno (BFGS) method was used for the local optimization by Trott and Olson in 2010. This method is a quasi-Newton method which uses the scoring function result and its gradient to speed up the optimization (Trott and Olson, 2010)

3.2.2.2 Empirical

The binding affinity of the receptor-ligand complex is obtained from the regression analysis of X-ray structural information using previously defined and known binding energies. Compared to power field scoring, empirical scoring functions are much faster due to simple energy terms (Böhm, 1994, 1998; Sotriffer et al, 2008).

Glide uses many steps to find out the best positions for docking poses. These docking poses occur with the position of a ligand relative to the receptor. In the field of the receptor, ligands are minimized for the best poses by exhaustive conformational search of Monte Carlo procedures.

Glide docking is using the empirical scoring function, which is based on ChemScore 2.5 function (Eldridge et al,18). The scoring function is shown as follows:

$$\Delta G_{bind} = C_0 + C_{lipo} \sum f(r_{1r}) + C_{hbond} \sum g(\Delta r)h(\Delta\alpha) + C_{metal} \sum f(r_{1m}) + C_{rotb} H_{rotb} \quad (3.6)$$

The second term is described as lipophilic term, and the third one is for hydrogen bond interactions between ligand and receptor. Functions of f and g are for distances and angles.

Glide 2.5 has two forms of calculations; GlideScore 2.5 SP which is for the calculation of big databases, can calculate the binding energy even if receptor-ligand glide pose has some problems, and GlideScore 2.5 XP which is mostly for lead optimization, has more significant violation in the physical chemistry principles to calculate even worse structures. After all, GlideScore 2.5 recreates the function of ChemScore as it is shown below:

$$\Delta G_{bind} = C_{lipo-lipo} \sum f(r_{1r}) + C_{hbond-neut-neut} \sum g(\Delta r)h(\Delta\alpha) + C_{hbond-neut-charged} \sum g(\Delta r)h(\Delta\alpha) + C_{hbond-charged-charged} \sum g(\Delta r)h(\Delta\alpha) + C_{max-metal-ion} \sum f(r_{1m}) + C_{rotb}H_{rotb} + C_{polar-phob}V_{polar-phob} + C_{coul}E_{coul} + C_{vdW}E_{vdW} + solvation\ terms \quad (3.7)$$

The terms from ChemScore, which are lipophilic-lipophilic and hydrogen-bonding terms were used, and the difference between donor and acceptors were added. The metal-ligand interactions took into account with the addition of coefficient to -2.0 kcal/mol. The other term was considered the polar site in the hydrophobic region. One of the major terms in this equation is the addition of Coulomb and vdW interaction energies which are reduced for specific atoms. The other main term is the solvation model which is important to be included in the calculations. GlideScore 2.5 is using explicit water for specific cases, and it has the 'water scoring' technology that is giving many advantages to the program (Friesner et al, 2004)

3.2.2.3 Knowledge-based potential

It is based on a statistical analysis of binary interactions observed for the receptor-ligand. Energy functions are improving with structural information which has been found from the experiments (Muegge and Martin, 1999). Sum of pair-wise statistical potentials between protein and ligand is giving as scoring function, and it is as follows:

$$Binding\ Score = \sum_i^{lig} \sum_j^{prot} w_{ij}(r) \quad (3.8)$$

Here, w_{ij} is pair-wise potential and calculated using Boltzmann formula:

$$w_{ij}(r) = -k_B T ln\left[g_{ij}(r)\right] = -k_B T ln\left[\frac{\rho_{ij}(r)}{\rho_{ij}^*}\right] \quad (3.9)$$

Here, k_B is the Boltzmann constant, while T is the temperature of the system. The number density of protein-ligand atom pair which is described as ρ(r) at distance r, and in the reference state pair density is ρ*. Also, (ρ(r))/(ρ*) is calculated for free energies using radial distribution theory.

3.2.2.4 Descriptor-based

Descriptor-based scoring functions are similar to the previously found empirical scoring function even though they have fundamental differences. For many biological

structure properties, Quantitative structure-activity relationship (QSAR) method, which has descriptors has been used. Descriptors in QSAR method comes from some symbols of chemical data which can be applied to work on interactions of protein-ligand.

3.3 MM-GBSA

Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method is used for the determination of the binding free energy (ΔG_{bind}) of the ligand and receptor complexes by the following sum (Lyne et al, 2006):

$$\Delta G = \Delta E_{MM} + \Delta G_{SOL} + \Delta G_{SA} \quad (3.10)$$

Here, ΔE_{MM} is the difference in the minimized energies between the Nmt–inhibitor complex and the sum of the energies of the unliganded Nmt and inhibitor. ΔG_{SOLV} is the difference in the GBSA solvation energy of the protein–inhibitor complex and the sum of the solvation energies for the unliganded Nmt and inhibitor. ΔG_{SA} is the difference in surface area energies for the complex and the sum of the surface area energies for the unliganded Nmt and inhibitor.

4. METHODOLOGY

Several crystal structures of the HSF1 are available at the Protein Data Bank; however, only one of them had a complex structure with HSE, which is the promoter part of DNA. The crystal structure of human HSF1 complexed with HSE was selected for the molecular docking studies (PDB: 5D5U). The X-ray diffraction structure of the protein has a resolution of 2.91Å (Neudegger et al, 2016).

First, the protein was prepared for the drug discovery approach. The protein preparation was performed using Protein Preparation Wizard of Maestro package of Schrödinger 2020-2. The crystal structure was corrected, and hydrogen atoms were automatically added to the protein, and water molecules which were only two and not located in the binding site of the protein were removed, using Protein Preparation Wizard. Missing sidechains and missing loops were filled by Prime module. The energy was minimized using the OPLS-2005 force field (Schrödinger Release 2020-2, Protein Preparation Wizard). The prediction of pKa was performed through Propka at pH 7.0 (Li et al, 2005; Bas et al, 2008)

AutoDock Vina 1.1.2 and Glide HTVS (High-Throughput Virtual Screening) were used for the docking studies. Four different libraries were prepared based on the classification of the candidate structures (Figure 4.1).

The first library of molecules was prepared from ZINC15 Tranches databases, molecules with the properties of molecular weight up to 300 Da, standard reactivity, purchasable from stocks, at pH 7.4, neutral charge, and LogP up to 3.5. The number of molecules with these criteria were 104.522.

The second library was consist of 1379 compounds which were all FDA approved. The compounds were downloaded from ZINC15 database.

The third library of ligands from the PubChem databases was made from 3 different classes. First, of them was leading compounds which were selected from literature: Ronhinitib (RHT), Cantharidin (CLA) and (E)-ethyl 4-oxo-4-(thiazol-2-yl amino)but-

2-enoat (I001). From leading compounds of RHT, 109 compounds; from CLA, 310 compounds; from I001, 444 compounds were downloaded from PubChem database (Kim et al, 2019).

The last library of ligands which prepared from ZINC15 (Sterling and Irwin, 2015) contains 114555 compounds that are tested in vivo and 1114461 compounds tested with cells (Library 4). The compounds in Library 4 were only docked with Glide Docking HTVS which includes Glide XP (extra precision) and SP (standard precision) scores.

Following the preparation of the first three libraries of ligands, the ligands were optimized with B3LYP/6-31G* level of methodology (Jain et al, 2004).

AutoDock Vina 1.1.2 was used for the docking studies of the first three libraries to find out binding poses and binding energies of ligands to the receptor (Sastry et al, 2013). The docking protocol was set based on the compounds with experimental IC50 values (Table 5.1). The rotatable bonds of ligand were assembled, non-polar hydrogens were merged, and Gasteiger charges were added as implemented in AutoDock Tools 1.5.6. Ligand and protein files were saved as PDBQT file format (Morris et al, 2009). The grid-box sized 20 × 20 × 20 points and centred by Gly72 residue with a spacing of 1.0 Å. At the beginning of calculations, all ligands were docked to the rigid receptor. However, after having binding scores with low affinities because of the structure of the binding pocket, we decided to take some residues flexible in the receptor. Side chains of Phe61, His63, Ser68, Arg71, Gln72, and Tyr76 residues were taken flexible, and the structure was saved as PDBQT file format with implementations in AutoDock Tools 1.5.6. Selecting side chains flexible in the binding pocket of the target protein creates a more realistic interaction environment, and that is why flexible docking improves AutoDock Vina docking score accuracy (Rui et al, 2011). AutoDock Vina gives the docking scores as free binding energy (ΔG). All interactions between ligand and receptor were visualized using PYMOL software (Schrodinger, 2010) and Discovery Studio Visualizer for 2D diagrams (Biovia, 2017).

Glide HTVS was used for the docking studies of the last library to find out binding poses and binding energies of ligands to the receptor (Friesner et al, 2004). After preparation of the protein, a grid box generation was made for the binding site definition using Receptor Grid Generation of Maestro. Residue Gln72 was chosen to

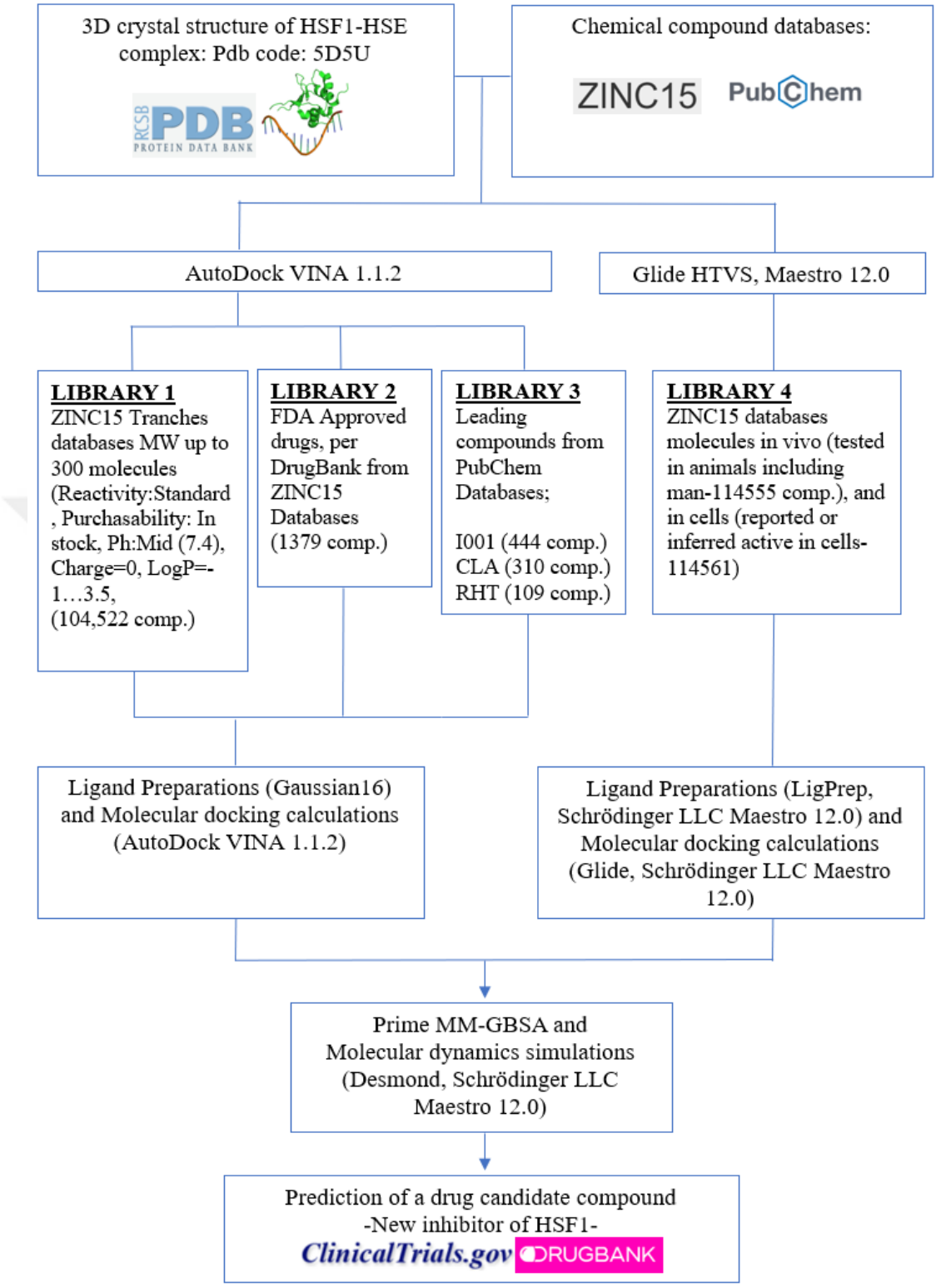


Figure 4.1: Flow-chart of the methodology.

be the centre of the grid box. The ligands only from library 4 were prepared by LigPrep module using the pH value of 7.0 +/- 2.0. Ligands from library 4 were docked with the protein using Glide HTVS (Schrödinger Release 2020-2, LigPrep)

Molecular Mechanics Generalized Born Surface Area (Prime/MM-GBSA) calculations were complete with receptor-ligand complex structures taken from the results AutoDock VINA and Glide docking to predict the binding mode and free energies. Receptor–ligand complex poses were optimized in Prime (Schrödinger Release 2020-2, PrimeX).

The molecular dynamic simulation was carried out for the apo form of the protein and the ligand-HSF1 complexes by Desmond simulation package of Schrödinger LLC (Schrödinger Release 2020-2, Desmond Molecular Dynamics System). Relax model system is one option to select before starting the simulation to prepare the system minimization. The NPT type of system was chosen with a temperature of 300 K and pressure 1 bar. The simulation lengths were 100 ns for the native protein and the other nine complexes of HSF1 with energy recording interval 100 ps. The long-range electrostatic interactions were found by the Particle Mesh Ewald method (Pierro et al, 2015). The cutoff radius in Columb interactions was chosen 9.0 Å, and SPC (single point charge) modes were used for water molecules characterization (Berendsen et al, 1981).

Characterization of ADME/Tox paramaters, prediction of druglike properties and pharmacokinetic properties of selected compounds (Table 5.6) were made using the SwissADME database (Daina et al, 2014).

5. RESULT AND DISCUSSION

Crystal structure of human HSF1 complexed with HSE DNA (Neudegger et al, 2016) was used for the docking studies (PDB:5D5U). The target area for potential inhibitors was chosen as the binding site HSE promoter to HSF1. As reported by Neudegger et al. (2016), there is a distinct region with probable interactions between the HSE and the HSF1 shown in Figure 2.4.(b). In a molecular docking study the residues Phe61, His63, Ser68, Arg71, Gln72 and Tyr76 were reported to be significant for HSF1-HSE interactions (Agarwal et al, 2015). For the validation and the determination of the docking protocol, the known inhibitors of HSF1 with IC_{50} values were taken from the research of Vilaboa et al, 2017. We decided on the docking protocol by comparing the rankings of the binding scores we found with the experimental IC_{50} values.

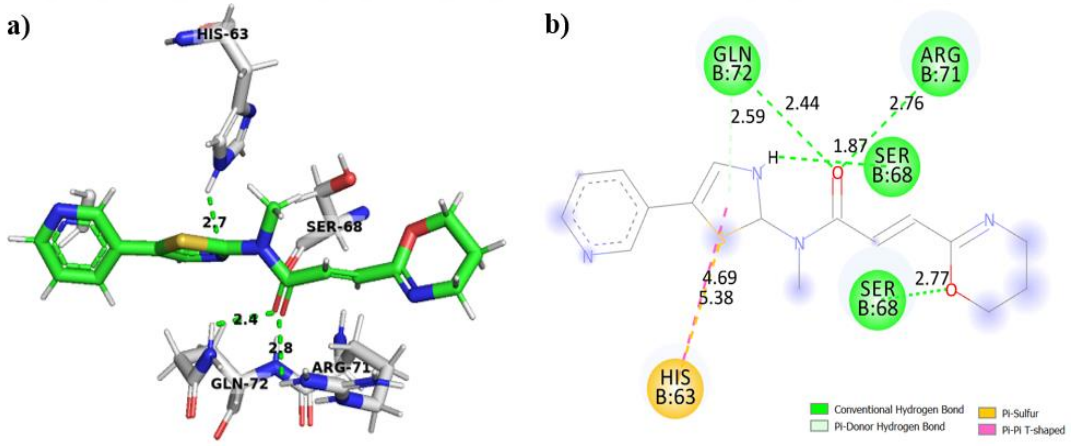
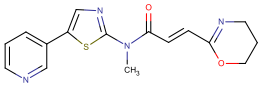
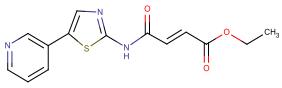
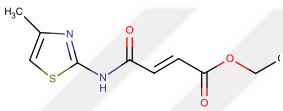
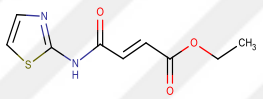
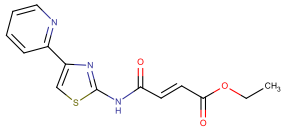
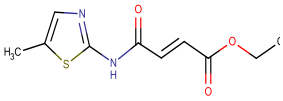


Figure 5.1: The binding site of HSF1 protein with I115. a) Hydrogen bond interaction between the HSF1 and side chains of Ser68, Arg71, and Gln72. b) All interactions can be seen between I115 and HSF1 with an interaction diagram.

In all Figures, the amino acid residues of HSF1 coloured as grey (carbon), blue (nitrogen), red (oxygen), and white (hydrogen). Moreover, intermolecular hydrogen bonds in HSF1 are shown by green dashed lines.

Table 5.1: Molecular docking and MM-GBSA protocol estimation of chosen molecules with IC50 values.

Compound	IC50 values (µM)	Structures	$\Delta G_{AutoDockVINA}$ (kcal/mol)	$\Delta G_{Prime\ MM\text{-}GBSA}$ (kcal/mol)
I115	0.7 +/- 0.1		-7.5	-57.50
I058	4.8 +/- 0.4		-6.6	-45.93
I011	7.0		-6.0	-34.03
I001	17.8 +/- 2.6		-5.6	-39.93
I054	20.2		-6.6	-44.13
I013	25.0		-5.6	-36.88

In Table 5.1. results of binding energies from AutoDock Vina and Prime MM-GBSA are given. The interactions between HSF1 and the compound I115, which has the best IC50 value and Vina score is shown in Figure 5.1. This compound had hydrogen bonds with the region of the HSF1 protein involving side chain of Gln72, Arg71 and His63 residues.

5.1 Molecular Docking

After determination of the Vina docking protocol, libraries were prepared from Zinc (Sterling and Irwin, 2015) and PubChem (Kim et al, 2019) as discussed in the methodology.

5.1.1 Library 1: ZINC15 tranches databases

A virtual screening on ZINC Tranches databases was performed by AutoDock Vina with flexible residues. Selected best 100 compounds in this library had the range of the binding energies from -9.9 to -8.7 kcal/mol. The docking free energies of the best compounds are given in Table 5.2.

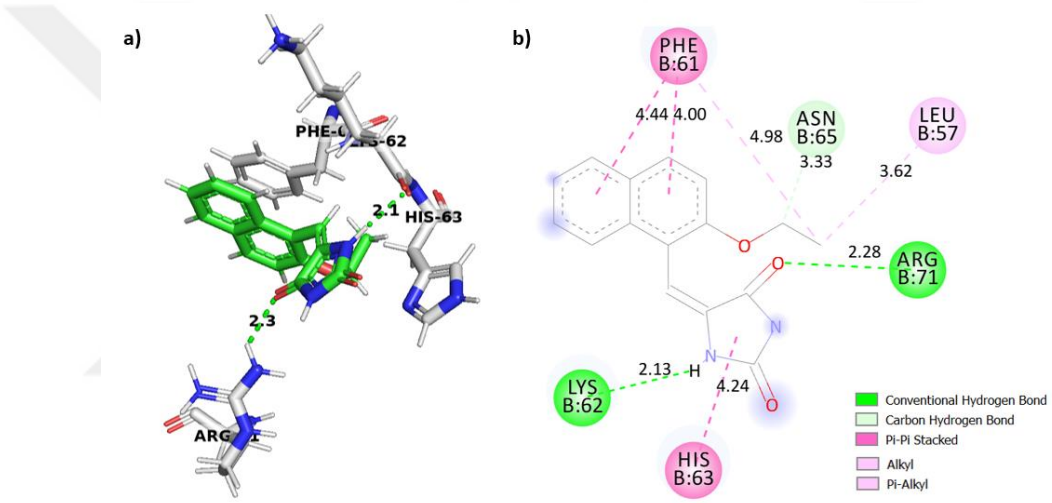
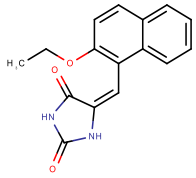
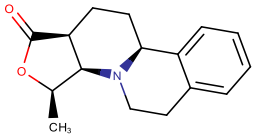
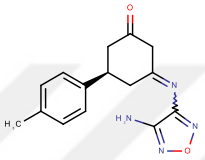
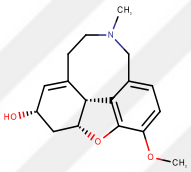
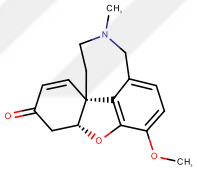
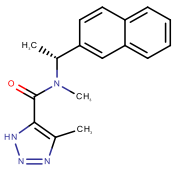
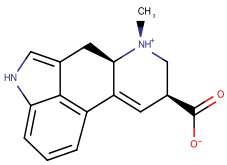
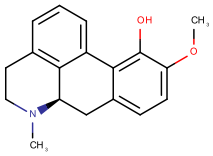
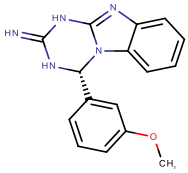


Figure 5.2: The binding site of HSF1 with compound 1(Zinc No: ZINC000004097653) a) Hydrogen bond interaction between the compound 1 and the Lys62, Arg71 residues. b) All interactions can be seen between the compound 1 and HSF1 with an interaction diagram.

The compound 1 **(**Zinc No: ZINC000409047067) from Table 5.2, has a strong hydrogen bond with Arg71 and Lys62 with a high docking score of -9.9 kcal/mol. It also makes a π–π stacked interaction with the aromatic side chain of Phe61 and His63. Also it has a π-alkyl interaction with Leu57 residue (Figure 5.2).

The compound 2 **(**Zinc No: ZINC000005530788) has a strong hydrogen bond with side chain of Lys62 with a docking score of -9.5 kcal/mol. It also makes a π–π stacked interaction with the aromatic side chain of Phe61 and a π-alkyl interaction with Lys62 residue. It also has π-cation interaction with Arg71 (Figure 5.3).

Table 5.2: The compounds with the highest docking scores from LIBRARY 1.

No.	ZINC code	Structures	$\Delta G_{AutoDockVINA}$ (kcal/mol)
1	ZINC000409047067		-9.9
2	ZINC000005530788		-9.5
3	ZINC000100903270		-9.5
4	ZINC000242473299		-9.5
5	ZINC000004097653		-9.4
6	ZINC001506396874		-9.4
7	ZINC000000897013		-9.3
8	ZINC000002033971		-9.3
9	ZINC000004622635		-9.3

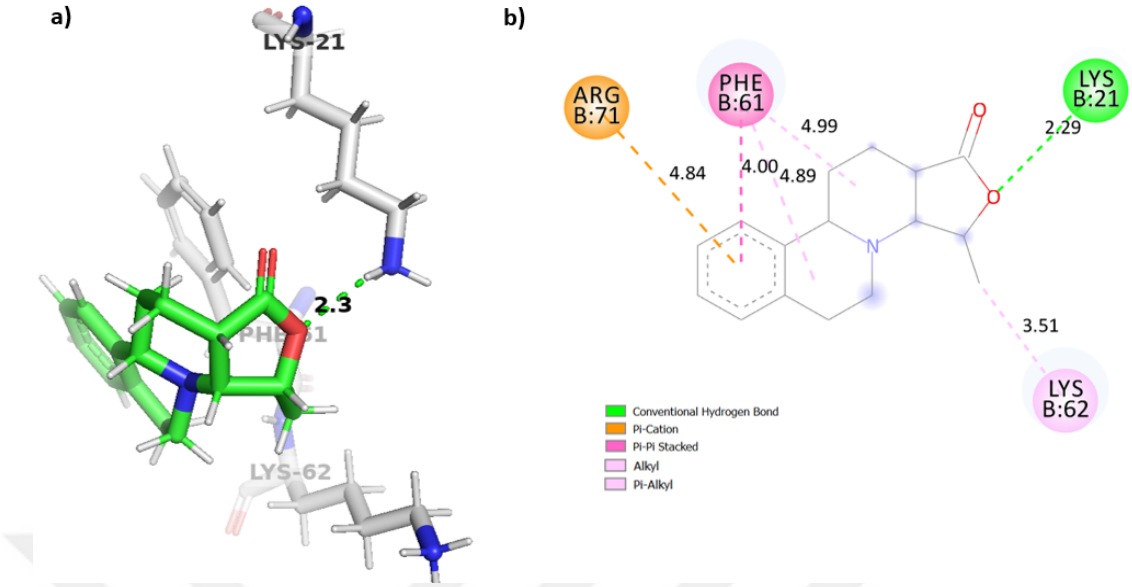


Figure 5.3: The binding site of HSF1 with compound 2 (Zinc No: ZINC000005530788) a) Hydrogen bond interaction between the compound 2 and the side chain of Lys21. b) All interactions can be seen between the compound 2 and HSF1 with an interaction diagram.

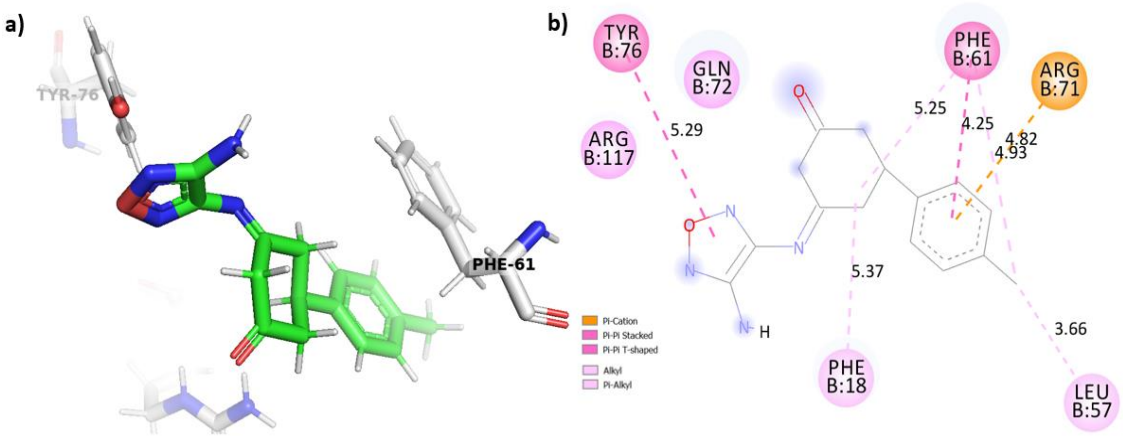


Figure 5.4: The binding site of HSF1 with compound 3 (Zinc No: ZINC000100903270) a) Hydrophobic interactions between the compound 3 and HSF1 b) All interactions can be seen between the compound 3 and HSF1 with an interaction diagram.

There is no hydrogen bond between the compound 3 **(**Zinc No: ZINC000100903270) and HSF1. However, there is strong hydropobic interactions with a docking score of -9.5 kcal/mol (Figure 5.4).

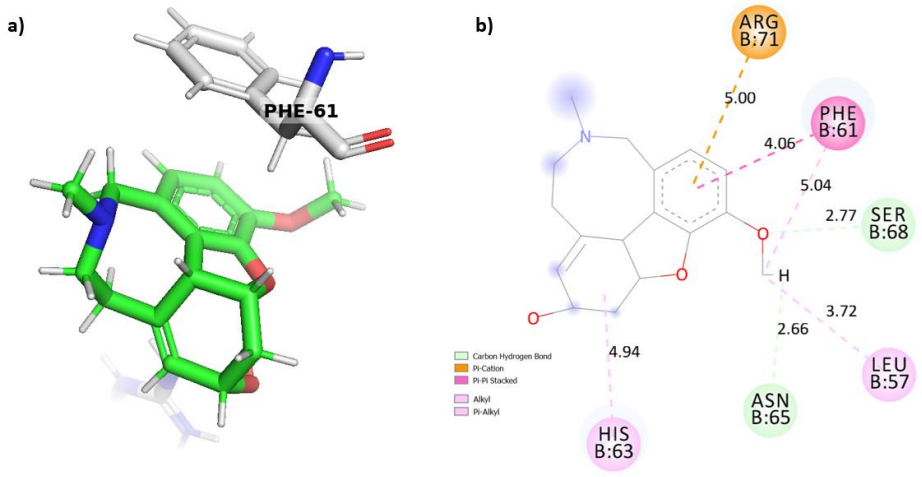


Figure 5.5: The binding site of HSF1 with compound 4 (Zinc No: ZINC000242473299) a) Hydrophobic interactions between the compound 4 and the HSF1. b) All interactions can be seen between the compound 4 and HSF1 with an interaction diagram.

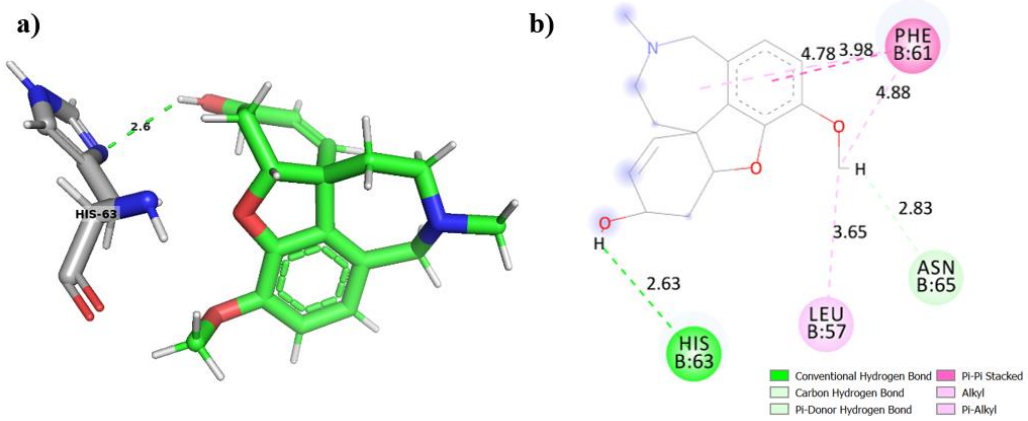


Figure 5.6: The binding site of HSF1 with compound 5 (Zinc No: ZINC000004097653) a) Hydrogen bond interaction between the compound 5 and the side chain of His63. b) All interactions can be seen between the compound 5 and HSF1 with an interaction diagram.

There is no hydrogen bond between the compound 4 **(**Zinc No: ZINC000242473299) and HSF1. However, there is strong hydropobic interactions with a docking score of -9.5 kcal/mol (Figure 5.5).

The compound 5 **(**Zinc No: ZINC000004097653) has a strong hydrogen bond with Nδ of His63 and a high docking score of -9.4 kcal/mol. It also makes a π–π stacked interaction with the aromatic side chain of Phe61 and a π-alkyl interaction with Leu57 residue (Figure 5.6).

In this library, even though compounds are with good binding scores, their interactions with HSF1 is mostly limited to hydrophobic interactions. These compounds are rather

small molecules and they do not have high binding affinities to HSF1-DNA domain . We further prepare a new Library based on FDA Approved Drugs.

5.1.2 Library 2: FDA approved drugs -repurposing of drugs

A virtual screening on ZINC FDA approved drug databases were studied for a possible repurposing of existing drugs. The docking study was carried out by AutoDock Vina with flexible residues. The binding energies of the eliminated best 196 compounds are from -10.7 to -8.6 kcal/mol in this library. The docking free energies of the best compounds are given in Table 5.3. All drug properties were checked from Drugs.com database.

Drug tetracycline (compound 10) has the best binding score with HSF1 protein in our results. Tetracycline has hydrogen bonds with His63, Arg71 and Arg117 in the binding pocket of HSF1. Also, it has π-alkyl interactions with the residues of Phe61, Lys21 and Ala17 (Figure 5.7). Tetracyclic is an antibiotic that is used for the treatment of many types of infections.

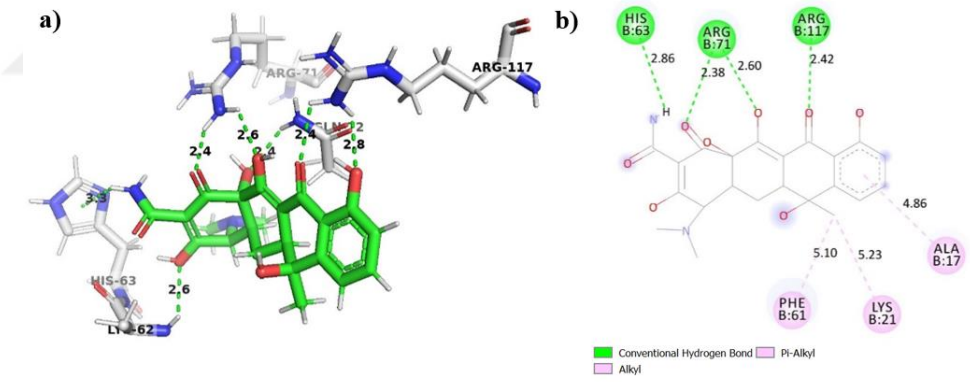


Figure 5.7: The binding site of HSF1 protein with Tetracycline (compound 10). a) Hydrogen bond interaction between compound 10 and the side chain of His63, Arg71, and Arg117. b) All interactions can be seen between the compound 10 and HSF1 with an interaction diagram.

The next compound is Midostaurin (compound 11) which is hydrogen-bonded by the guanidino group of Arg71, as well as stacking (π–π) interaction between its aromatic group and Phe61 residue to stabilize the binding (Figure 5.8). Midostaurin is a cancer therapy medicine used to prevent the proliferation of cancer cells in the body.

Aromasin is one of the important drugs in our FDA-approved drugs category. This compound is a known medicine for the treatment of breast cancer. For Aromasin

(compound 12), it is obtained that, it makes a hydrogen bond with the side chain of Ser68 residue. Moreover, the compound has alkyl interactions with Phe61, Lys62 and His63 (Figure 5.9).

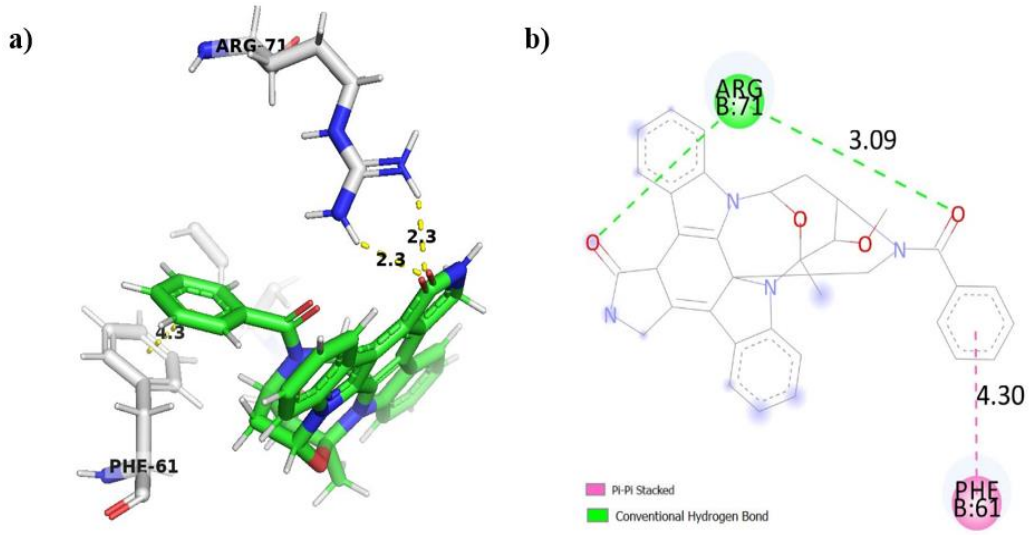


Figure 5.8: The binding site of HSF1 protein with Midostaurin (compound 11). a) Hydrogen bond interaction between compound 11 and the side chain of Phe61, and Arg71. b) All interactions can be seen between the compound 11 and HSF1 with an interaction diagram.

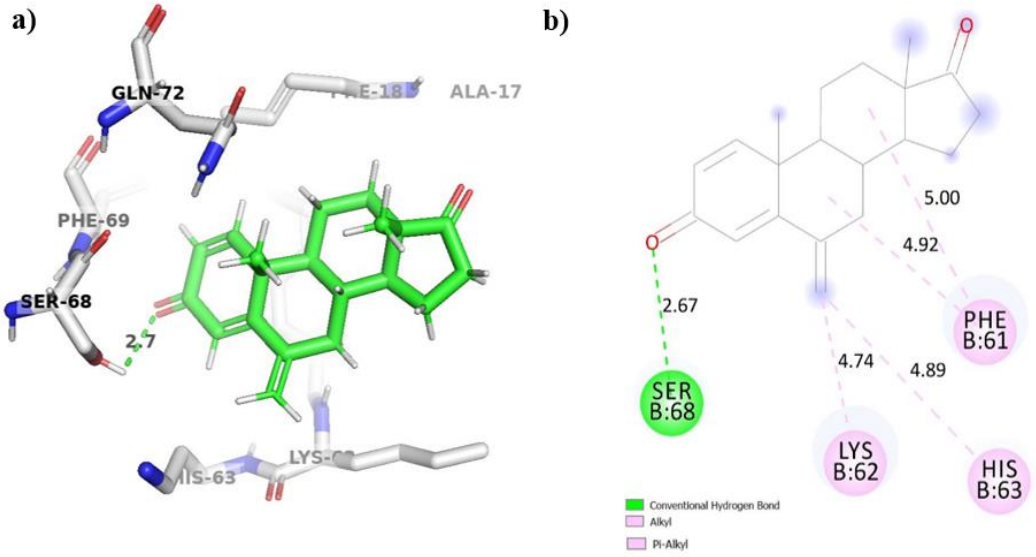


Figure 5.9: The binding site of HSF1 protein with Aromasin (compound 12). a) Hydrogen bond interaction between compound 12 and the side chain of Ser68. b) All interactions can be seen between the compound 12 and HSF1 with an interaction diagram.

Table 5.3: The compounds with the highest docking scores from LIBRARY 2.

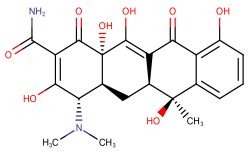
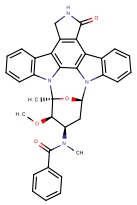
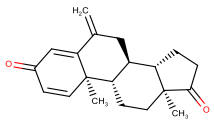
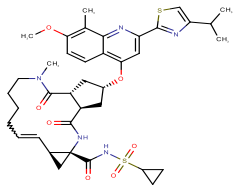
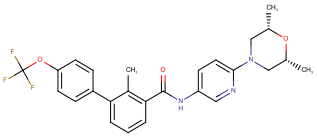
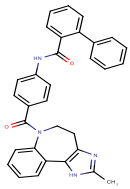
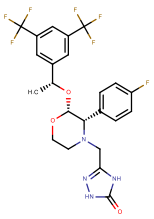
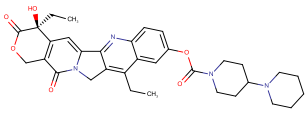
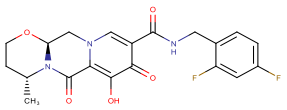
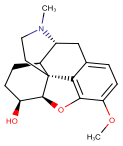
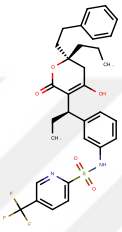
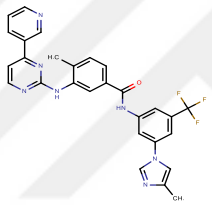
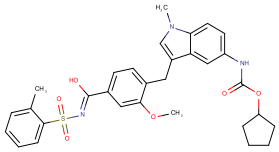
No.	Drug Name	Structures	$\Delta G_{AutoDockVINA}$ (kcal/mol)
10	Tetracycline Brand names: *Ala-Tet, Brodspec, Panmycin, Sumycin, Tetracap, Tetracon* (antibiotic)		-10.7
11	Midostaurin Brand name: *Rydapt* (medicine to prevent the proliferation of a cancer cell)		-10.7
12	Aromasin Brand name: *Aromasin* (medicine for the treatment of breast cancer)		-10.2
13	Simeprevir Brand name: *Olysio* (It is used for the treatment of hepatitis C infections)		-10.1
14	Erismodegib Brand name: *Odomzo* (It is used for the treatment of basal cell carcinoma which is a type of skin cancer)		-10.1
15	Conivaptan Brand name: *Vaprisol* (It is used for decreasing a level of hormone)		-9.8
16	Emend Brand name: *Emend* (medicine to prevent nausea and vomiting related to cancer chemotherapy)		-9.7
17	Irinotecan Brand name: *Onivyde* (medicine for the treatment of pancreatic cancer)		-9.6

Table 5.3 (continued): The compounds with the highest docking scores from LIBRARY 2.

No.	Drug Name	Structures	$\Delta G_{AutoDockVINA}$ (kcal/mol)
18	Dolutegravir Brand name: *Tivicay* (used with other medicines to treat HIV)		-9.6
19	Dihydroergotamine Brand name: *D.H.E. 45* (medicine to treat migraine headaches.)		-9.6
20	Tipranavir Brand name: *Aptivus* (Tipranavir is an antiviral medicine that is also used for the treatment of HIV)		-9.5
21	Nilotinib Brand name: *Tasigna* (It is used to treat Philadelphia chromosome-positive chronic myeloid leukaemia)		-9.5
22	Zafirlukast Brand name: *Accolate* (It is used to treat asthma)		-9.4

Simeprevir is another significant drug which is used for the treatment for hepatitis C virus. Simeprevir (compound 13) makes a hydrogen bond with Gln72, which is a crucial interaction for the binding mode of HSF1. This compound has additional π-sigma interactions with the aromatic group of Phe61 and alkyl interactions with the side chains of Pro16, Lys62, Lys116 and Arg117 residues (Figure 5.10).

Erismodegib (compound 14) which was already used in the treatment of advanced basal cell carcinoma (BCC) has hydrogen bonds with Phe18 and Tyr76 residues. Additionally, the compound binding mode is stabilized with surrounding residues. It has stacking (π–π) interaction with Phe61, alkyl interactions with Pro16, Ala17, Lys116, Val119, and interestingly halogen interaction between its fluorine group and Asn65 residue (Figure 5.11).

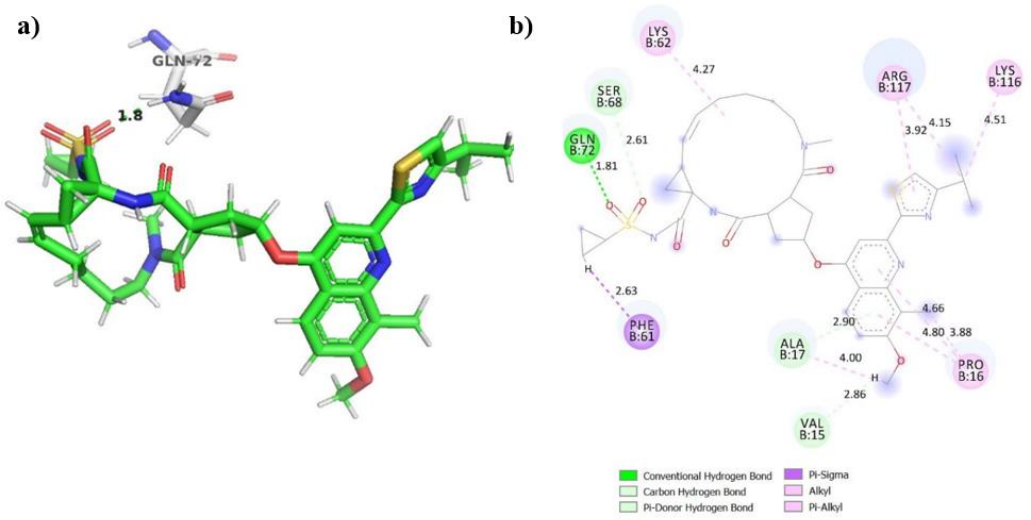


Figure 5.10: The binding site of HSF1 protein with Simeprevir (compound 13). a) Hydrogen bond interaction between compound 13 and the side chain of Gln72. b) All interactions can be seen between the compound 13 and HSF1 with an interaction diagram.

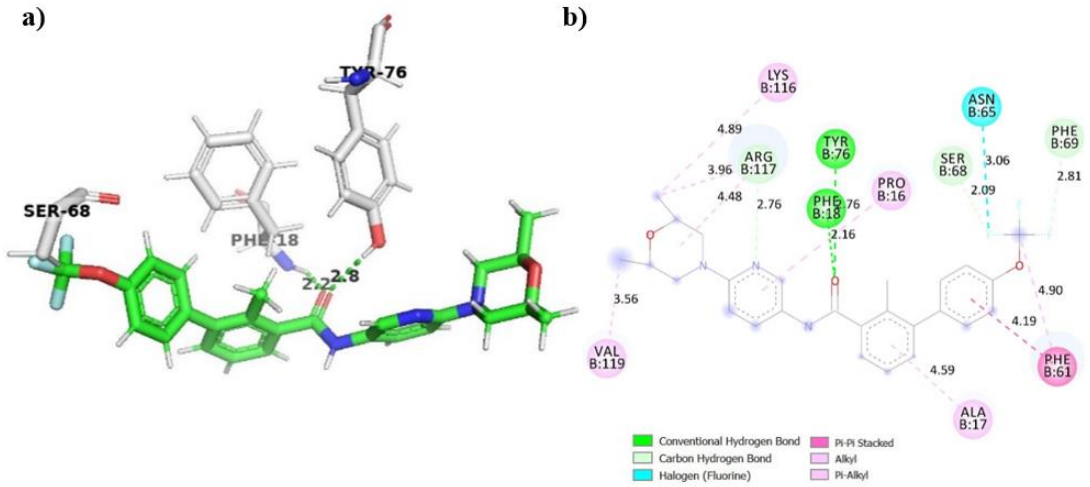


Figure 5.11: The binding site of HSF1 protein with Erismodegib (compound 14). a) Hydrogen bond interaction between compound 14 and Phe18, Tyr76 residues. b) All interactions can be seen between the compound 14 and HSF1 with an interaction diagram.

Tipranavir (compound 20) which was used for HIV positive patients has a variety of interactions with HSF1, including hydrogen bonds with the Ala17, Phe18, Phe69, Arg71, Gln72, Tyr76 residues. The compound has stacking (π−π) interaction with Phe61, alkyl interactions with Pro16, Leu57 and Ala117. Likewise, it has halogen interaction between its fluorine group and Asn65, His63 residues. Lastly, Arg71 is also making a π-cation interaction with the aromatic group of Tipranavir (Figure 5.12).

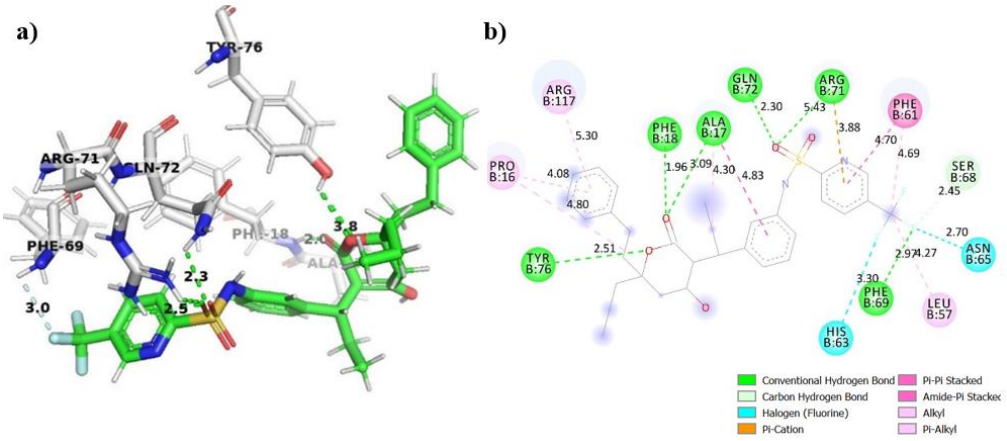


Figure 5.12: The binding site of HSF1 protein with Tipranavir (compound 20). a) Hydrogen bond interaction between the compound 20 and Ala17, Phe18, Phe69, Arg71, Gln72, Tyr76 residues. b) All interactions can be seen between the compound 20 and HSF1 with an interaction diagram.

Repurposing of drugs is the fastest way to find a potential cure for some diseases. In this library, we found some of the medicines can be potential inhibitors for HSF1 protein. Tetracycline (is an antibiotic), Midostaurin (is used to prevent the proliferation of a cancer cell), Aromasin (is used for the treatment of breast cancer), Simeprevir (is used for the treatment of hepatitis C infections), Erismodegib (is used for the treatment of basal cell carcinoma which is a type of skin cancer), and Tipranavir (is an antiviral medicine that is also used for the treatment of HIV) have great binding affinities with HSF1.

5.1.3 Library 3: Leading compounds library

Virtual screening of the compounds was completed by AutoDock Vina with flexible residues. This library was prepared with screening the three known inhibitors of HSF1, which are CLA, RHT and I001. However, pdb complexes of these inhibitors are not available. Therefore, before using the inhibitors for virtual screening, their binding modes were obtained individually (Figure 5.13). Selected best 73 compounds in this database had the range of the binding energies between -11.2 and -8 kcal/mol. After all, the docking free energies of the best compounds are given in Table 5.4.

Compound 23 (PubChem ID: 23281114), which was found by screening RHT ligand from PubChem database, has hydrogen bonds with side chains of Ser68, Arg71, Gln72 residues. It also has π-cation interaction between its aromatic group and Arg117 residue, π-π stacking interaction with Phe61 and weak alkyl interaction with Lys62 residue (Figure 5.14)

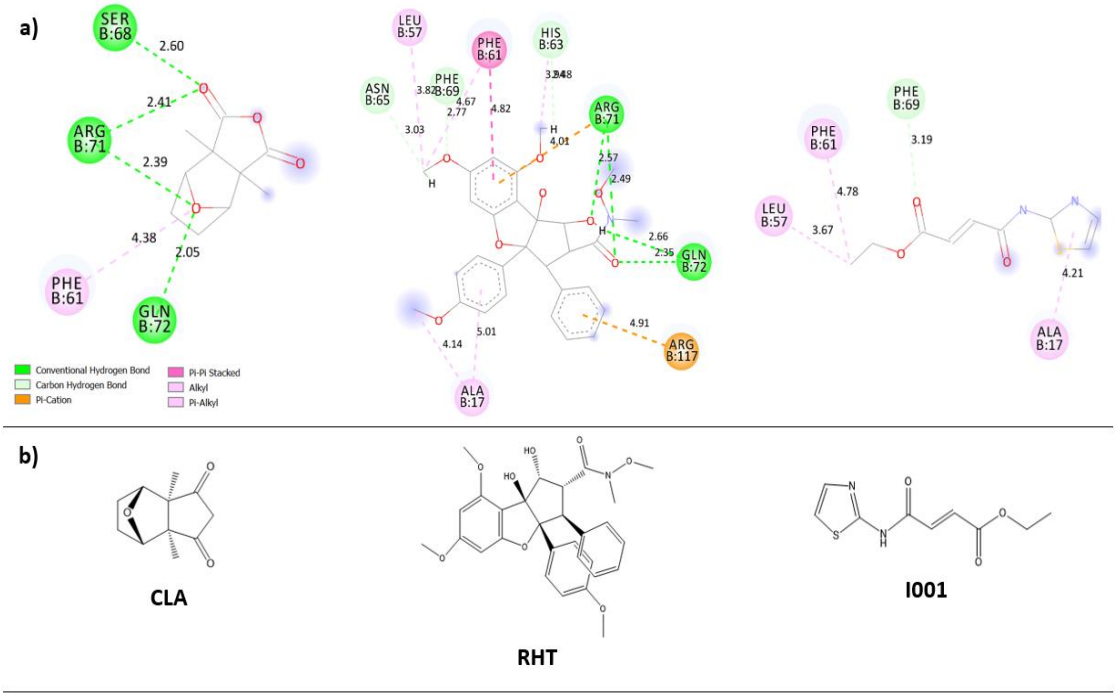


Figure 5.13: HSF1 Leading Compounds; Cantharidin, Rohinitib and (E)-ethyl 4-oxo-4-(thiazol-2-yl amino) but-2-enoat**,** a) All interactions with HSF1 with an interaction diagram, b) 2D structures, c) AutoDock Vina docking results and Prime MM-GBSA binding scores, respectively.

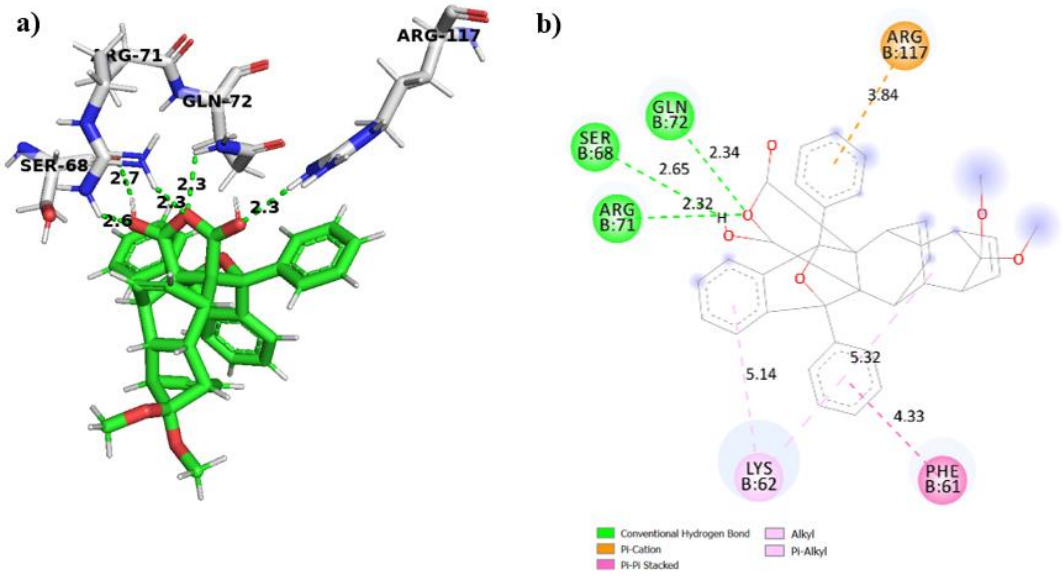


Figure 5.14: The binding site of HSF1 protein with compound 23 **(**PubChem ID: 23281114) a) Hydrogen bond interaction between the compound 23 and side chains of Ser68, Arg71, Gln72 residues. b) All interactions can be seen between the compound 23 and HSF1 with an interaction diagram.

Table 5.4: The compounds with the highest docking scores from LIBRARY 3.

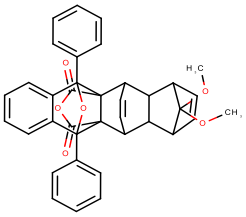
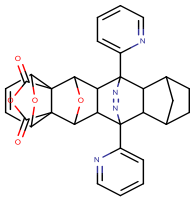
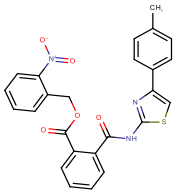
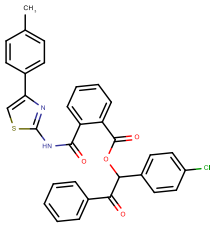
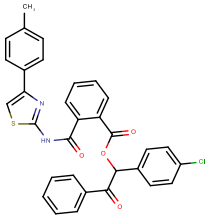
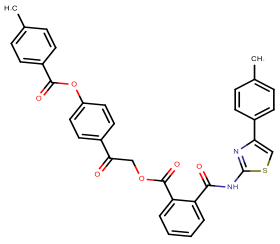
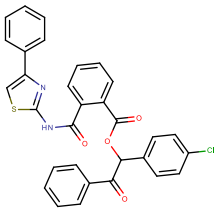
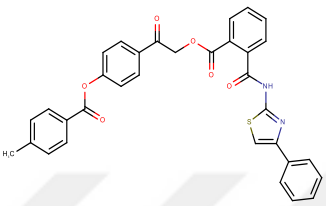
No.	PubChem ID / compound database	Structures	$\Delta G_{AutoDockVINA}$ (kcal/mol)
23	PubChem 23281114/ (RHT compound search)		-11.2
24	PubChem 122201797/ (CLA compound search)		-10.2
25	PubChem 126182188/ (I001 compound search)		-10.1
26	PubChem 126182038/ (I001 compound search)		-9.9
27	PubChem 126182039/ (I001 compound search)		-9.9
28	PubChem 126178230/ (I001 compound search)		-9.7

Table 5.4 (continued): The compounds with the highest docking scores from LIBRARY 3.

No	PubChem ID / compound database	Structures	$\Delta G_{AutoDockVINA}$ (kcal/mol)
29	PubChem 126189744 (I001 compound search)		-9.4
30	PubChem 126180578/ (I001 compound search)		-9.3

Compound 25 (PubChem ID: 126182188), which was found by screening I001 ligand from PubChem database, has hydrogen bonds with side chains of His63, Arg71, and Arg117 residues. Similarly, to compound 23, it also has π-cation interaction with Arg117 residue, π-π stacking interaction with Phe61 and alkyl interaction with Ala17, Lys116 residue (Figure 5.15).

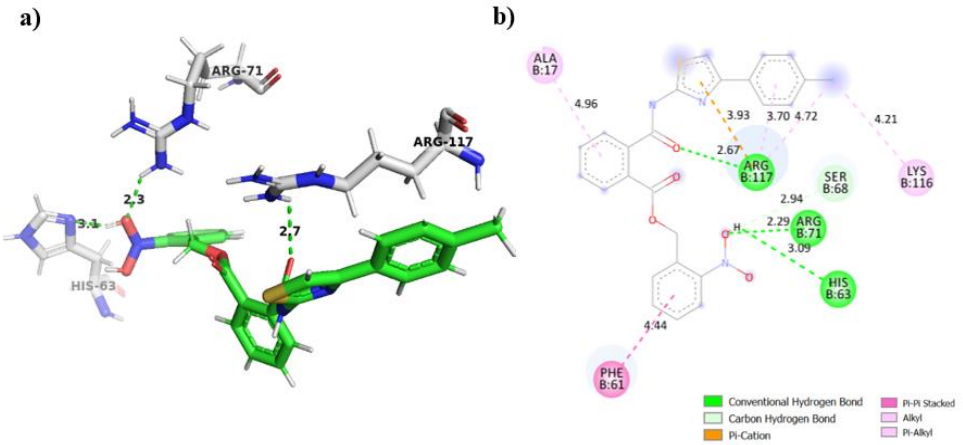


Figure 5.15: The binding site of HSF1 protein with compound 25 (PubChem ID: 126182188). a) Hydrogen bond interaction between the compound 25 and side chains of His63, Arg71, Arg117 residues. b) All interactions can be seen between the compound 25 and HSF1 with an interaction diagram.

Compound 26 (PubChem ID: 126182038), which was also found by the screening of I001 ligand from PubChem database, has hydrogen bonds with side chains of Gln72, and Arg117 residues. Furthermore, it has π-cation interaction with Arg117 residue, π-

π stacking interaction with Phe61 and alkyl interaction with Ala17, Leu57, Phe69, Arg71, and Lys116 residues (Figure 5.16).

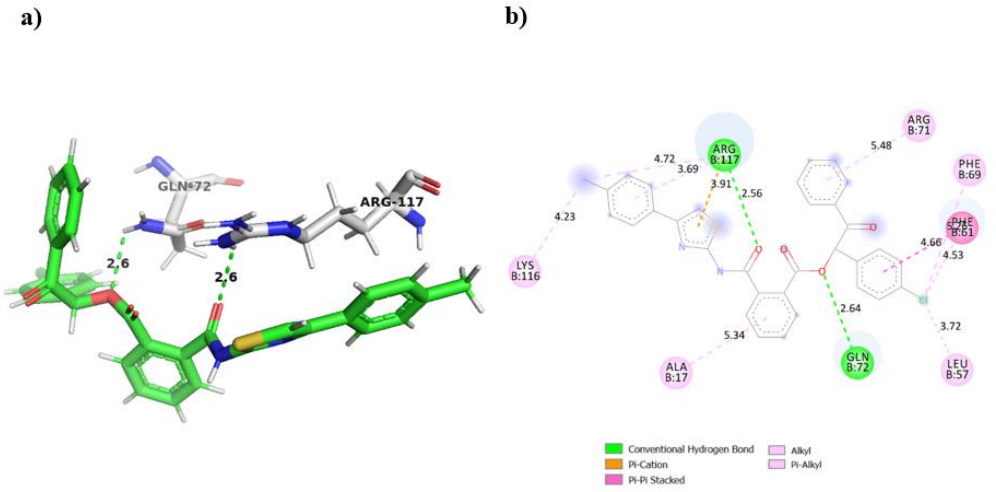


Figure 5.16: The binding site of HSF1 protein with compound 26 (PubChem ID: 126182038). a) Hydrogen bond interaction between the compound 26 and side chains of Gln72 and Arg117 residues. b) All interactions can be seen between the compound 26 and HSF1 with an interaction diagram.

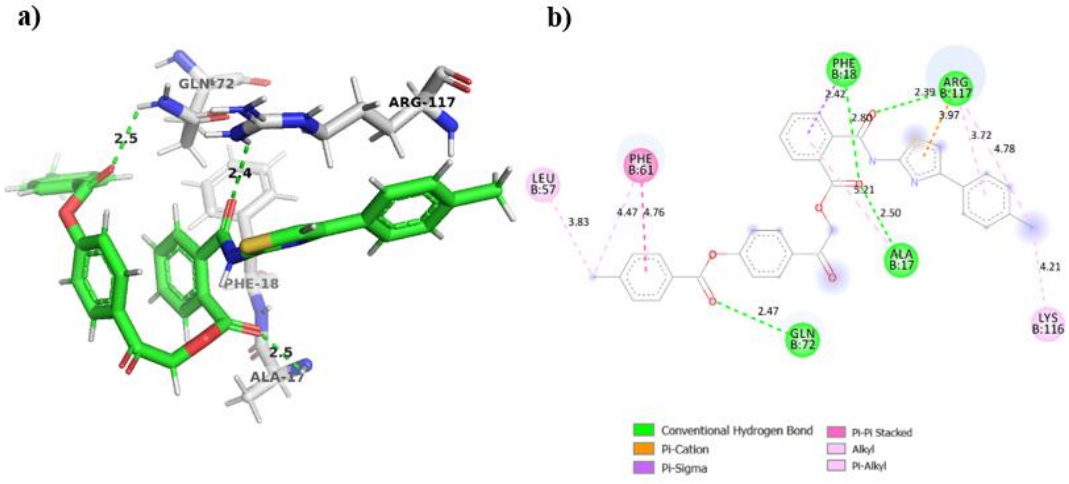


Figure 5.17: The binding site of HSF1 protein with compound 28 (PubChem ID: 126178230). a) Hydrogen bond interaction between the compound 28 and side chains of Ala17, Phe18, Gln72 and Arg117 residues. b) All interactions can be seen between the compound 28 and HSF1 with an interaction diagram.

Compound 28 (PubChem ID: 126178230) which was found similarly to compound 26, it has hydrogen bonds with side chains of Ala17, Phe18, Gln72 and Arg117 residues. In addition, it has π-cation interaction with Arg117 residue, π-π stacking interaction with Phe61 and alkyl interaction with Leu57, and Lys116 residues (Figure 5.17).

As a result, the compounds from this library have better binding affinity than CLA, RHT and I001 leading compounds. Compound 23 has better binding score and interactions than its own leading compound RHT. Similarly, compound 25 and 26 also have better binding scores which are -10.2 kcal/mol and -9.9 kcal/mol, than I001 compound with -5.6 kcal/mol.

5.1.4 Library 4: ZINC15 databases with molecules reported in vivo and cells

Virtual screening of ZINC15 databases for the molecules tested or reported in vivo or in cells were completed by Glide HTVS. The best 50 compounds in vivo database had the range of the binding energies between –7.3 and -6.1 kcal/mol. Also, 98 compounds in the cell database had the range of the binding energies between –8.1 and –6.0 kcal/mol.

Results are shown in Table 5.5. From the virtual screening results, compound 31 had the best docking score. Compound 31, which has a high molecular weight of 614.424 Da, is hydrogen-bonded with Phe18, Lys62, His63, Ser68, Gln72 and Tyr76 residues. Additionally, it has π-π stacking interaction with His63 residue. Lys21 has an attractive charge for the compound (Figure 5.18).

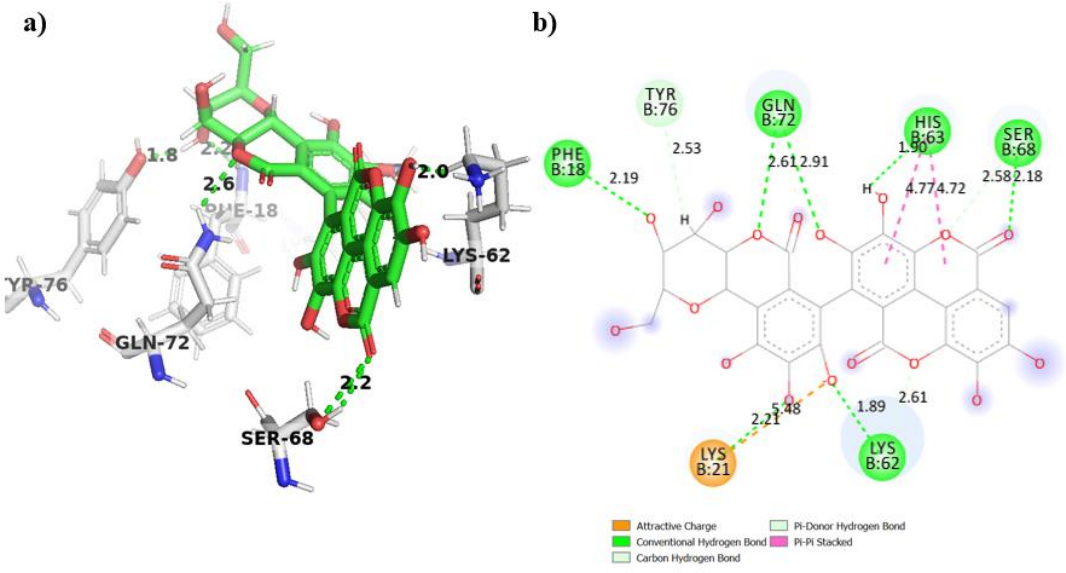
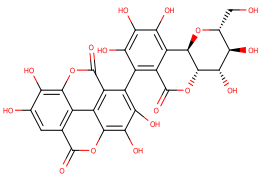
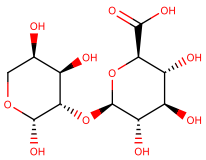
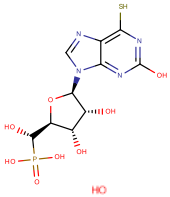
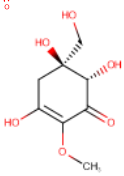
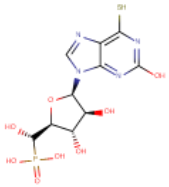
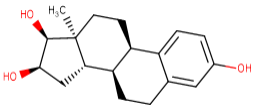
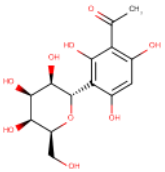


Figure 5.18: The binding site of HSF1 protein with compound 31 (Zinc ID: ZINC000100773913) a) Hydrogen bond interaction between the compound 31 and Phe18, Lys62, His63, Ser68, Gln72 and Tyr76 residues. b) All interactions can be seen between the compound 31 and HSF1 with an interaction diagram.

Table 5.5: The compounds with the highest docking scores from LIBRARY 4.

No.	ZINC code	Structures	$\Delta G_{GlideDocking}$ (kcal/mol)
31	ZINC000100773913		-9.5
32	ZINC000257466362		-8.0
33	ZINC000095618666		-7.7
34	ZINC000100829660		-7.5
35	ZINC000095618667		-7.5
36	ZINC000012493629		-7.3
37	ZINC000085807436		-7.3

Among the docked compounds of Library 4, Compound 31 has the highest binding affinity with strong hydrogen bond interactions with HSF1.

5.2 MM/GBSA Calculations

To compare the compounds that have good binding scores, we investigated MM-GBSA calculations for re-scoring the binding energies of ligand-protein complexes. The binding energies of these complexes are the sum of basic molecular mechanics energy terms (bond, angle, and dihedral) and electrostatic, van der Waals interactions are added to the terms. Also, polar and non-polar contributions to the solvation free energies are included. The binding energies in Table 5.6, are calculated for ligands (with best binding scores in top 10 from each library), with AutoDock Vina and Prime MM-GBSA. The negative values of binding energies represent the favourable interaction between the compound and receptor.

The results strongly show that the selected compounds may inhibit the HSF1 protein before the transcription. Binding scores are sorted for the best MM-GBSA scores since they already have similar values for Vina results (Table 5.6).

In Table 5.6, HBA, HBD, RB, and H-Bond represent the number of hydrogen bond acceptor, hydrogen bond donor, rotatable bonds and interactions in the binding mode, respectively.

5.3 Molecular Dynamic Simulations

The molecular dynamic simulation was carried out for the apo structure of the protein with 100 ns to observe the side chain movement of the active side residues. The thermodynamic properties against time (ps) were plotted to investigate the stability of the simulation. In these plots, the system equilibration is shown as E, E_P, T, P, and V which are the total energy (kcal/mol), barostat energy (kcal/mol), temperature (K), pressure (atm), and volume ($Å^3$), respectively (Figure 5.19).

Root mean square deviation (RMSD) of the HSF1 protein shows that if the simulation was equilibrated. During the simulation, conformational changes in the protein were observed to be around 2-3Å, which is perfectly acceptable for our small protein. Backbone RMSD data did not indicate any departure from the native structure (Figure 5.20.a).

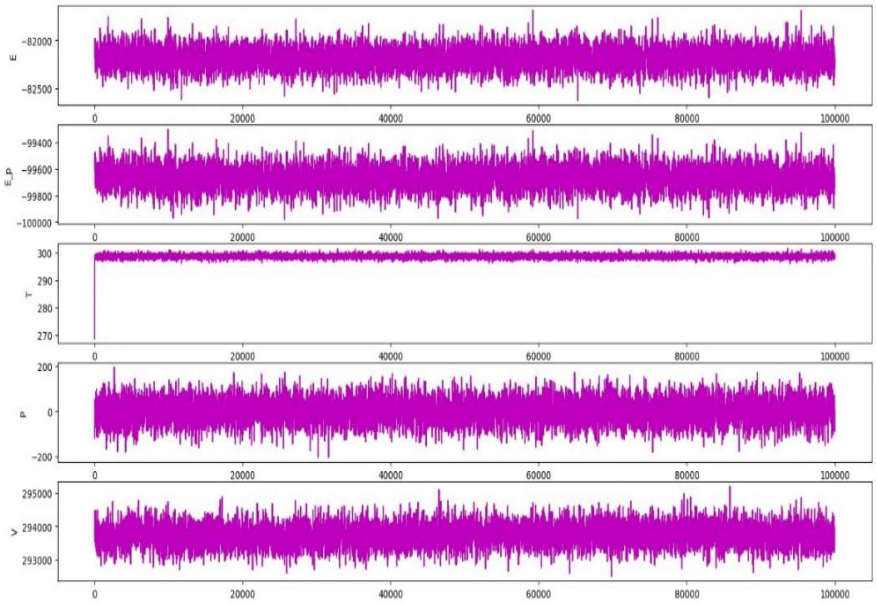


Figure 5.19: Stability of the thermodynamic properties of HSF1-apo form.

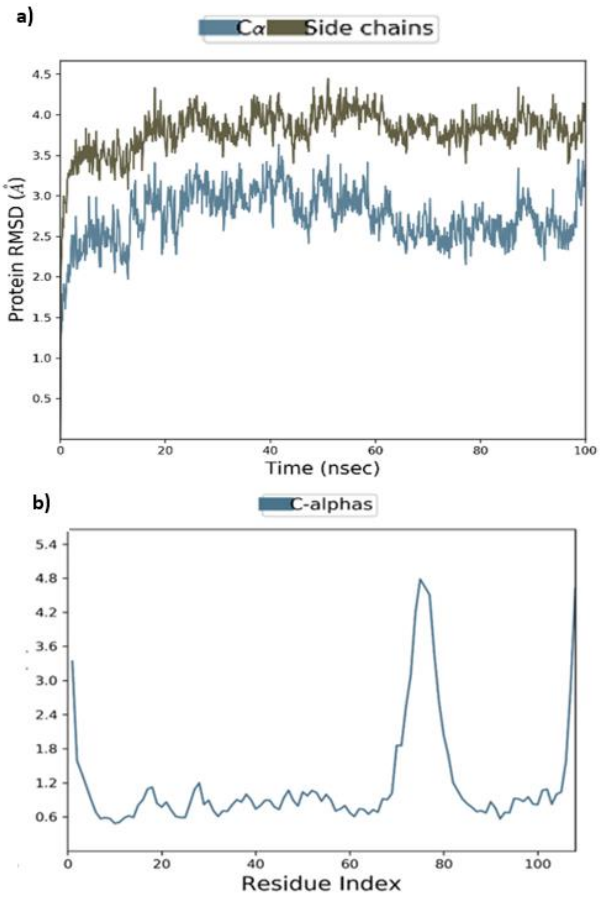


Figure 5.20: The HSF1 protein backbone and sidechain a) RMSD plots against time and b) RMSF plots against residue numbers.

Root mean square fluctuation (RMSF) of the protein indicates the local changes. In Figure 5.20.b, the most fluctuated areas in the protein are shown with the peaks, which has the amino acid residues between 84-91 possessed high flexibility. The Ser13 and Gly87 residues had the major peaks which 3.34 Å and 4.78Å fluctuations. HSF1 binding site to HSE had the least fluctuated peaks which were less than 1Å. After molecular dynamic simulation (MD) of the apo-protein, the first nine compounds from Table 5.6 (sorted according to $\Delta G_{MM\text{-}GBSA}$ values) were chosen for the MD calculations to understand the interaction between ligand and receptor in detail. RMSD of the ligand-protein complexes were demonstrated during 100ns simulations. On the left y-axis, protein RMSD is calculated, and on the right y-axis, ligand RMSD changes are calculated with respect to the protein and its binding pocket (Figure 5.21).

Compound 29 (PubChem ID: 126189744), has hydrogen bonds with Gln72 and hydrophobic interactions with Phe61, His63, Arg117 residues (Figure 5.22.a). Compound 26 (PubChem ID: 126182038) has hydrogen bonds with Lys62 and hydrophobic interactions with Phe61 residue (Figure 5.22.b). Compound 25 (PubChem ID: 126182188) has hydrogen bonds with Ser68, Gln72 and hydrophobic interactions with Phe61 residue (Figure 5.22.c). Compound 14 (Erismodegib) has hydrogen bonds with Gln72 and hydrophobic interactions with Phe18, His63 residues (Figure 5.22.d). Compound 20 (Tipranavir) has hydrogen bonds with Lys62 and hydrophobic interactions with Phe61, His63 residues (Figure 5.22.e). Compound 12 (Aromasin) has hydrogen bonds with Asn65 and hydrophobic interactions with Ala17, Phe18, Phe61 residues (Figure 5.22.f). Compound 28 (PubChem ID: 126178230) has hydrogen bonds with Gln72 and hydrophobic interactions with Phe61 residue (Figure 5.22.g). Compound 23 (PubChem ID: 23281114) has hydrogen bonds with Gln72 and hydrophobic interactions with Phe61 residue (Figure 5.22.h). Compound 13 (Simeprevir) has hydrogen bonds with Ala17 and Phe18 residues (Figure 5.22.i). These interactions were observed for half of the simulation time.

The simulation interactions diagrams in Figure 5.22 demonstrate that the most important hydrogen bonds are occurring between the ligands and Ala17, Phe18, Phe61, Lys62, His63, Asn65, Gln72, Ser68, Arg117 residues. Similarly, Phe61 residue also has hydrophobic interaction most of the simulation time (Figure 5.22)

Table 5.6: The Prime MM-GBSA and Vina flexible docking binding energies between the compounds and the HSF1 (kcal/mol).

Name of the molecules	MW (g/mol)	HBA	HBD	LogP	RB	LogS(ESOL)/Solubility	Druglikeness (ADME/Tox) Lipinski Violation/Bioavailability Score	H-Bonds (Residue of Amino Acid Group)	$\Delta G_{AutoDockVINA}$ (kcal/mol)	$\Delta G_{MM\text{-}GBSA}$ (kcal/mol)
PubChem 126189744	553.0	4	1	8.22	9	-9.77/ Low	2 violation/ 0.17	-	-9.4	-76.21
PubChem 126182038	567.1	4	1	8.73	9	-10.22/ Low	2 violation/ 0.15	Gln72, Arg117	-9.9	-73.80
PubChem 126182188	473.5	5	1	6.62	8	-8.28/ Low	0 violation/ 0.55	His63, Arg71, Arg117	-10.1	-72.50
PubChem 126178230	590.7	5	1	8.37	11	-10.26/ Low	1 violation/ 0.55	Ala17, Phe18, Gln72, Arg117	-9.7	-66.23
Erismodegib	485.5	5	1	6.76	6	-6.43/ Low	0 violation/ 0.55	Phe18, Tyr76	-10.1	-60.57
Tipranavir	602.7	5	2	7.14	11	-7.49/ Low	1 violation/ 0.56	Ala17, Phe18, Phe69, Arg71, Gln72, Tyr76	-9.5	-58.83
Aromasin	296.4	2	0	3.87	0	-3.61/ High	0 violation/ 0.55	Ser68	-10.2	-58.60
PubChem 23281114	570.6	5	0	5.26	4	-5.89/ Moderate	2 violation/ 0.17	Arg71, Ser68, Gln72	-11.2	-57.62
Simeprevir	749.9	9	2	4.56	7	-7.14/ Low	2 violation/ 0.17	Gln72	-10.1	-57.41
Nilotinib	529.5	6	2	5.36	7	-6.23/ Low	1 violation/ 0.55	Tyr76	-9.5	-56.83
PubChem 126180578	576.6	5	1	7.85	11	-9.82/ Low	1 violation/ 0.55	Ala17, Phe18, Gln72, Arg117	-9.3	-55.51
PubChem 122201797	534.6	8	0	2.29	2	-4.48/ Moderate	1 violation/ 0.55	Ser68, Arg71, Gln72	-10.2	-55.00
PubChem 126182039	567.1	4	1	8.73	9	-10.21/ Low	2 violation/ 0.17	Lys62, Gln72, Tyr76	-9.9	-54.54

Table 5.6 (continued): The Prime MM-GBSA and Vina flexible docking binding energies between the compounds and the HSF1 (kcal/mol).

Name of the molecules	MW (g/mol)	HBA	HBD	LogP	RB	LogS/Solubility	Druglikeness (ADME/Tox) Lipinski Violation/Bioavailability Score	H-Bonds (Residue of Amino Acid Group)	$\Delta G_{AutoDockVINA}$ (kcal/mol)	$\Delta G_{MM\text{-}GBSA}$ (kcal/mol)
Irinotecan	586.7	6	1	2.78	5	-5.71/ Moderate	1 violation/ 0.55	Lys62, Arg71	-9.6	-54.32
Zafirlukast	575.7	7	2	6.77	8	-6.93/ Low	1 violation/ 0.55	Phe18, Gln72	-9.4	-52.47
Dolutegravir	419.4	6	2	1.10	3	-4.01/ Moderate	0 violation/ 0.55	Ala17, Phe18, Ser68	-9.6	-51.95
Conivaptan	498.6	3	2	5.31	4	-6.67/ Low	1 violation/ 0.55	-	-9.8	-48.97
ZINC000242473299	287.4	4	1	1.03	1	-2.57/ High	0 violation/ 0.55	-	-9.5	-47.52
ZINC000002033971	281.4	3	1	3.24	1	-3.60/ High	0 violation/ 0.55	Phe61	-9.3	-45.80
ZINC000100903270	284.3	5	1	2.87	2	-2.73/ High	0 violation/ 0.55	-	-9.5	-45.25
ZINC000004622635	293.3	5	3	3.14	2	-3.51/ High	0 violation/ 0.55	Gln72, Phe61	-9.3	-44.50
ZINC000409047067	282.3	3	2	1.87	3	-3.43/ High	0 violation/ 0.55	Lys62, Arg71	-9.9	-42.91
Dihydroergotamine	583.7	6	3	2.71	4	-4.88/ Moderate	1 violation/ 0.55	Arg71, Gln72	-9.6	-40.79
Emend	534.4	5	2	5.22	8	-5.61/ Moderate	1 violation/ 0.55	Phe18, Lys21, Gln72, Tyr76	-9.7	-40.62
Midostaurin	570.7	4	1	5.43	3	-6.60/ Low	1 violation/ 0.55	Arg71, Gln72	-10.7	-36.09
ZINC000005530788	257.3	2	0	2.69	0	-3.21/ High	0 violation/ 0.55	Lys21	-9.5	-36.09
ZINC000000897013	268.3	2	2	-0.78	1	-3.21/ High	0 violation/ 0.55	Phe18	-9.3	-35.05
Tetracycline	444.4	9	6	-3.49	2	-1.78/ High	1 violation/ 0.11	His63, Arg71, Arg117	-10.7	-25.70

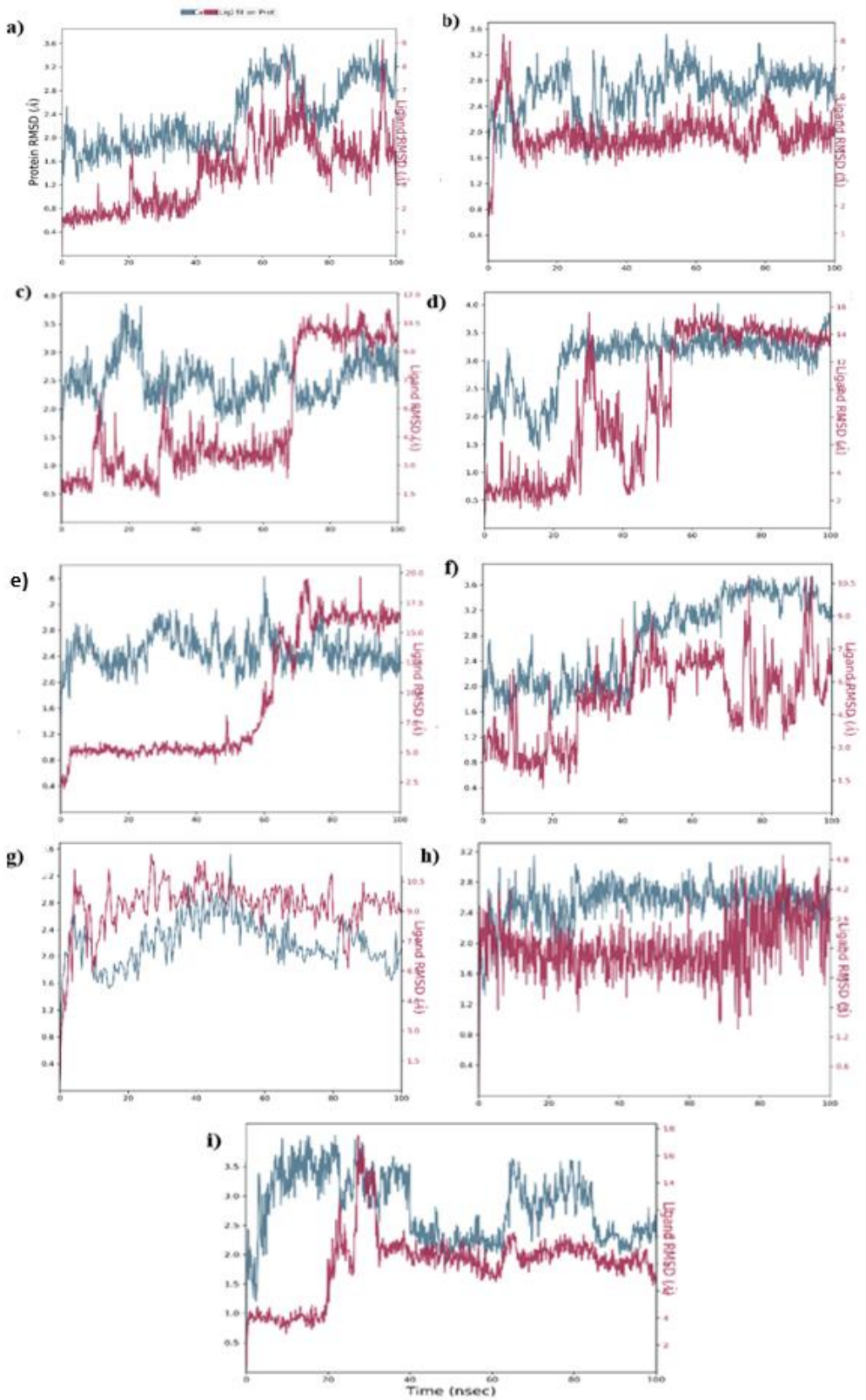


Figure 5.21: The protein and ligand a)PubChem126189744, b)PubChem126182038,c)PubChem126182188,d)Erismodegib, e)Tipranavir,f)Aromasin,g)PubChem126178230, h) PubChem23281114, i)Simeprevir, RMSD plot against time for 100 ns.

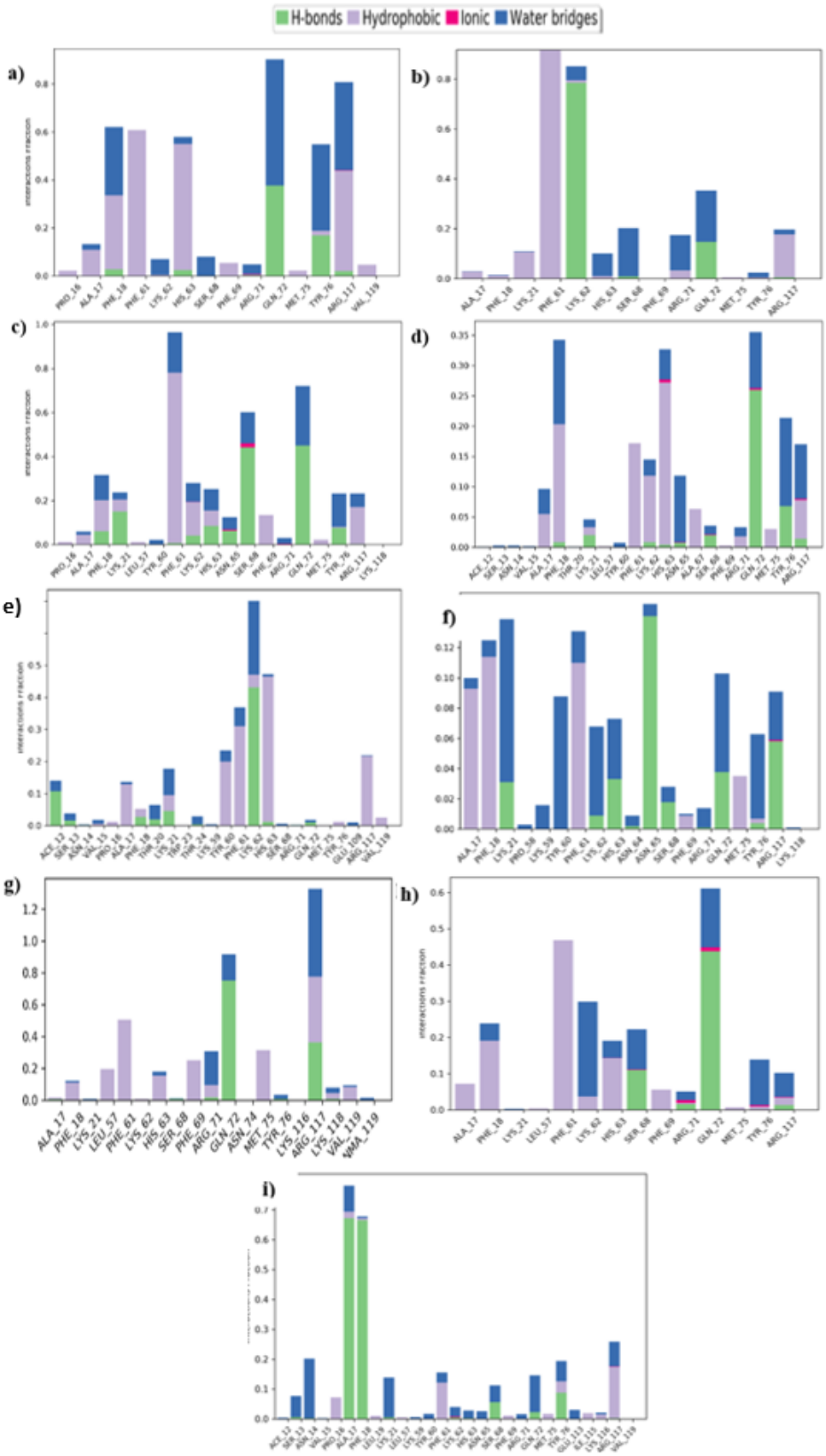


Figure 5.22: Interaction diagram between the protein and ligand a)PubChem126189744, b)PubChem126182038, c)PubChem126182188, d)Erismodegib, e) Tipranavir, f)Aromasin, g)PubChem126178230, h) PubChem23281114, i)Simeprevir, for 100 ns.

6. CONCLUSION

Recent studies have shown that HSF1 inhibition has an important role in cancer treatments such as colon cancer, breast cancer (Mendillo et al, 2012). In this study, inhibitors that targets the HSF1 were screened using computational chemistry methods. To fulfill this purpose, AutoDock 4.2 software and Glide docking methods were used. The ZINC and PubChem databases were scanned and four different libraries were prepared for the virtual screening of the ligands to predict the binding modes with HSF1.

The Library 1 was prepared from ZINC15 Tranches Database with compounds having molecular weight lowerthan 300 Da. However, these compounds are rather small molecular weight compounds and they do not have strong binding affinity to HSF1-DNA domain. The DNA binding domain of HSF1 is more suitable for compounds with higher molecular weight. In the Library 2, we studied for a possible repurposing of existing drugs and identified several potential candidates of HSF1 from FDA approved drugs database. Tetracycline, Midostaurin, Aromasin, Simeprevir, Erismodegib, and Tipranavir are some of these medicines that demonstrated good docking results. Experimental inhibition efficiencies must be further investigated to determine the possible repurposing of these candidate drugs. Commercially available compounds were also screened with the three known inhibitors of HSF1, which are CLA, RHT and I001 from ZINC and PubChem databases to prepare the Library 3. Based on binding affinities and interactions, four compounds (25, 26, 28 and 29) with the highest binding energies with Vina and Prime MM-GBSA, proposed to be potential inhibitors for HSF1. These compounds had better binding affinities than the leading compounds CLA, RHT and I001. Changing flexible functional group of RHT with more rigid cyclic groups, changed the interactions with HSF1. Lastly, compounds that are experimentally confirmed (in vivo/in cell) were further arranged for the Library 4. However, further research with flexible docking for these compounds is still needed.

The significant interactions of ligands with the amino acid residues Ala17, Phe18, Phe61, Lys62, His63, Ser68, Arg71, Gln72 and Tyr76 were found in the HSF1-HSE binding pocket as expected.

In order to investigate the interactions of the candidate inhibitors with HSF1, 100 ns MD runs were carried out. The molecular dynamics simulation results of HSF1-ligand complexes revealed that critical hydrogen bond and hydrophobic interactions with Ala17, Phe18, Lys21, Tyr60, Phe61, Lys62, His63, Asn65, Gln72, Tyr76, Ser68 and Arg117 residues of HSF1 might contribute to its potent HSF1 inhibition. Especially, Phe61 residue has shown hydrophobic interaction with aromatic fragments of the ligands. These residues were reported for the interaction of HSF1 with HSE. Agarwal et al, also pointed out similar binding interactions between HSF1 and Cantharidin-Rohinitib based hybrid ligands (Agarwal et al, 2015).

The obtained results enabled us to propose ligands with compound numbers 12 (Aromasin), 13 (Simeprevir), 14 (Erismodegib), 20 (Tipranavir), 23 (PubChem ID: 23281114), 25 (PubChem ID: 126182188), 26 (PubChem ID: 126182038), 28 (PubChem ID: 126178230), and 29 (PubChem ID: 126189744) as potential inhibitors of HSF1 for target-based cancer therapy

REFERENCES

Agarwal, T., Annamalai, N., Khursheed, A., Maiti, T. K., Arsad, H. B., & Siddiqui, M. H. (2015). Molecular docking and dynamic simulation evaluation of Rohinitib—Cantharidin based novel HSF1 inhibitors for cancer therapy, *Journal of Molecular Graphics and Modelling*, *61*, 141-149.

Akagawa, H., Takano, Y., Ishii, A., Mizuno, S., Izui, R., Sameshima, T., ..., Yoshioka, T. (1999). Stresgenin B, an inhibitor of heat-induced heat shock protein gene expression, produced by Streptomyces sp. AS-9, *The Journal of Antibiotics*, *52*, 960-970.

Allen, W. J., Balius, T. E., Mukherjee, S., Brozell, S. R., Moustakas, D. T., Lang, P. T., . Rizzo, R. C. (2015). DOCK 6: Impact of new features and current docking performance, *Journal of Computational Chemistry*, *36*, 1132-1156.

Aşkar, T. K., Ergün, N., Turunç, V. (2007). Isı şok proteinler ve fizyolojik rolleri, *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, *13*(1), 109-114.

Aufricht C. (2005). Heat–shock protein 70: molecular supertool?, *Pediatr Nephrol; 20*: 707–713.

Ananthan, J., Goldberg, A. L., & Voellmy, R. (1986). Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes, *Science, 232*(4749), 522-524.

Bas, D. C., Rogers, D. M., & Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein–ligand complexes, *Proteins: Structure, Function, and Bioinformatics*, *73*(3), 765-783.

Beckmann, R.P., Lovett, M. & Welch, W.J. (1992). Examining the function and regulation of hsp 70 in cells subjected to metabolic stress, *J. Cell Biol*, *117*, 1137-1150.

Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., & Hermans, A. J. (1981). Intermolecular forces.

Biovia, D. S. (2017). Discovery studio visualizer. San Diego, CA, USA, 936.

Boellmann, F., Guettouche, T., Guo, Y., Fenna, M., Mnayer, L. and Voellmy, R. (2004). DAXX interacts with heat shock factor 1during stress activation and enhances its transcriptional activity, *Proc. Natl. Acad. Sci. U.S.A.., 101*, 4100–4105.

Böhm, H. J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known

three-dimensional structure, *Journal of Computer-Aided Molecular Design, 8*(3), 243–256.

Böhm, H. J. (1998). Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs, *Journal of Computer-Aided Molecular Design, 12*(4), 309-323.

Chen, R., Wierda, W. G., Chubb, S., Hawtin, R. E., Fox, J. A., Keating, M. J., ... & Plunkett, W. (2009). Mechanism of action of SNS-032, a novel cyclin-dependent kinase inhibitor, in chronic lymphocytic leukemia, *Blood 113*(19), 4637–4645.

Chen, B. C., Tu, S. L., Zheng, B. A., Dong, Q. J., Wan, Z. A., & Dai, Q. Q. (2020). Schizandrin A exhibits potent anticancer activity in colorectal cancer cells by inhibiting heat shock factor 1, *Bioscience Reports, 40*(3).

Chugh, R., Sangwan, V., Patil, S. P., Dudeja, V., Dawra, R. K., Banerjee, S., ..., Saluja, A. K. (2012). A preclinical evaluation of Minnelide as a therapeutic agent against pancreatic cancer, *Science Translational Medicine, 4*(156), 156ra139.

Dai, Q., Zhang, C., Wu, Y., McDonough, H., Whaley, R. A., Godfrey, V., ... & Cyr, D. (2003). CHIP activates HSF1 and confers protection against apoptosis and cellular stress. *The EMBO journal, 22*(20), 5446-5458.

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

Darnell, J. E. (2002). Transcription factors as targets for cancer therapy, *Nature Reviews Cancer, 2*, 740-749.

Ellis, R.J. & van der Vies, S.M. (1991). Molecular chaperones, *A Rev. Biochem, 60*, 321-347.

Ewing, T. J., Makino, S., Skillman, A. G., Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, *Journal of Computer-Aided Molecular Design, 15*(5), 411-428.

Farkas, T., Kutskova, Y. A., Zimarino, V. (1998). Intramolecular repression of mouse heat shock factor 1, *Mol Cell Biol 18*, 906–918.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... & Shaw, D. E. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *Journal of medicinal chemistry, 47*(7), 1739-1749.

Fujimoto M, Izu H, Seki K, Fukuda K, Nishida T, Yamada S, Kato K, Yonemura S, Inouye S & Nakai A. (2004). HSF4 is required for normal cell growth and differentiation during mouse lens development, *EMBO J 23*, 4297–4306.

Goodsell, D. S., Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing, *Proteins: Structure, Function, and Bioinformatics, 8*(3), 195-202.

Grover, A., Shandilya, A., Agrawal, V., Pratik, P., Bhasme, D., Bisaria, V. S. Sundar, D. (2011). Hsp90/Cdc37 chaperone/co-chaperone complex, a novel junction anticancer target elucidated by the mode of action of herbal drug Withaferin A, *BMC Bioinformatics 12* (Suppl 1), S30.

Hamilton, K. L., Gupta, S., Knowlton, A. A. (2004). Estrogen and regulation of heat shock protein expression in female cardiomyocytes: cross-talk with NF kappa B signaling, *J Mol Cell Cardiol 36*, 577– 584.

Hightower, L.E. (1980). Cultured cells exposed to amino acid analogues or puromycin rapidly synthesize several polypeptides, *J. Cell Physiol. 102*, 407-424.

Hou, T., Wang, J., Li, Y., Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *Journal of Chemical İnformation and Modeling, 51*(1), 69-82.

Hou Y, Wei H, Luo Y & Liu G. (2010). Modulating expression of brain heat shock proteins by estrogen in ovariectomized mice model of aging, *Exp Gerontol 45*, 323–330.

Irving, J. A., Whisstock, J. C., Lesk, A. M. (2001). Protein structural alignments and functional genomics, *Proteins: Structure, Function, and Bioinformatics*, *42*(3), 378-382.

Iwasaki, S., Floor, S. N., & Ingolia, N. T. (2016). Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor, *Nature*, *534*(7608), 558-561.

Jackson, R. M., Gabb, H. A., Sternberg, M. J. (1998). Rapid refinement of protein interfaces incorporating solvation: application to the docking problem, *Journal of Molecular Biology*, *276*(1), 265-285.

Jacobs, A. T. Marnett, L. J. (2007). Heat shock factor 1 attenuates 4-hydroxynonenal-mediated apoptosis: critical role for heat shock protein 70 induction and stabilization of Bcl-XL, *J Biol Chem 282*, 33412–33420.

Jaeger, A. M., Pemble, C. W., Sistonen, L., Thiele, D. J. (2016). Structures of HSF2 reveal mechanisms for differential regulation of human heat-shock factors, *Nat Struct Mol Biol 23*, 147–154.

Jain, R., Ahuja, B. L., & Sharma, B. K. (2004). Density-Functional Thermochemistry. III. The Role of Exact Exchange, *Indian Journal of Pure & Applied Physics*, *42*, 43-48.

Jäättelä, M. (1999). Escaping cell death: survival proteins in cancer. *Experimental cell research*, *248*(1), 30-43.

Kansanen, E., Jyrkkänen, H. K., Volger, O. L., Leinonen, H., Kivelä, A. M., Häkkinen, S. K., ... & Freeman, B. A. (2009). Nrf2-dependent and-independent responses to nitro-fatty acids in human endothelial cells identification of heat shock response as the major pathway activated by nitro-oleic acid, *Journal of Biological Chemistry*, *284*(48), 33233-33241.

Kawazoe, Y., Tanabe, M., Sasai, N., Nagata, K., Nakai, A. (1999). HSF3 is a major heat shock responsive factor during chicken embryonic development, *Eur J Biochem 265*, 688–697.

Kim, J. A., Kim, Y., Kwon, B. M., & Han, D. C. (2013). The natural compound cantharidin induces cancer cell death through inhibition of heat shock protein 70 (HSP70) and Bcl-2-associated athanogene domain 3 (BAG3) expression by blocking heat shock factor 1 (HSF1) binding to promoters, *Journal of Biological Chemistry*, *288*(40), 28713-28726.

Kim, J. A., Lee, S., Kim, D. E., Kim, M., Kwon, B. M., Han, D. C. (2015). Fisetin, a dietary flavonoid, induces apoptosis of cancer cells by inhibiting HSF1 activity through blocking its binding to the hsp70 promoter, *Carcinogenesis*, *36*(6), 696-706.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2019). PubChem 2019 update: improved access to chemical data, *Nucleic acids research*, *47*(D1), D1102–D1109. https://doi.org/10.1093/nar/gky1033

Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P. (1983). Optimization by simulated annealin, *Science*, *220*(4598), 671-680.

Kuntz, ID., Blaney, JM., Oatley, SJ., Langridge, R., Ferrin, TE. (1982). A geometric approach to macromolecule-ligand interactions, *Journal of Molecular Biology*, *161*(2), 269–88.

Lambert, M., Jambon, S., Depauw, S., & David-Cordonnier, M. H. (2018). Targeting transcription factors for cancer treatment, Molecules, 23(6), 1479.

Leach, A. R., Kuntz, I. D. (1992). Conformational analysis of flexible ligands in macromolecular receptor sites, *Journal of Computational Chemistry*, *13*(6), 730-748.

Liu, P. C., Thiele, D. J. (1999). Modulation of human heat shock factor trimerization by the linker domain, *J Biol Chem 274*, 17219–17225.

Li, N., Wang, T., Li, Z., Ye, X., Deng, B., Zhuo, S., ... & Zhu, T. (2019). Dorsomorphin induces cancer cell apoptosis and sensitizes cancer cells to HSP90 and proteasome inhibitors by reducing nuclear heat shock factor 1 levels, *Cancer biology & medicine, 16(2), 220*.

Li, J., Labbadia, J., Morimoto, R. I. (2017). Rethinking HSF1 in stress, development, and organismal health, *Trends in Cell Biology*, *27*(12), 895-905.

Li, H., Robertson, A. D., Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values, *Proteins: Structure, Function, and Bioinformatics*, *61*(4), 704-721.

Lyne, P. D., Lamb, M. L., Saeh, J. C. (2006). Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring, *J Med Chem 49*, 4805-4808.

Mendillo, M. L., Santagata, S., Koeva, M., Bell, G. W., Hu, R., Tamimi, R. M., ... & Lindquist, S. (2012). HSF1 drives a transcriptional program distinct

from heat shock to support highly malignant human cancers, *Cell, 150*(3), 549-562.

Metzler, B., Abia, R., Ahmad, M., Wernig, F., Pachinger, O., Hu, Y., Xu, Q. (2003). Activation of heat shock transcription factor 1 in atherosclerosis, *Am J Pathol 162*, 1669–1676.

Miles, S. L., McFarland, M., & Niles, R. M. (2014). Molecular and physiological actions of quercetin: need for clinical trials to assess its benefits in human disease, *Nutrition Reviews, 72*(11), 720-734.

Moon, J. B., Howe, W. J. (1991). Computer design of bioactive molecules: A method for receptor-based de novo ligand design, *Proteins: Structure, Function, and Bioinformatics, 11*(4), 314-328.

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of Computational Chemistry, 19*(14), 1639-1662.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *Journal of Computational Chemistry, 30*(16), 2785-2791

Morris, G. M., Olson, A. J., Goodsell, D. S. (2000). Protein-ligand docking, *Evolutionary Algorithms in Molecular Design*, 31-48.

Mulholland, P. J., Ferry, D. R., Anderson, D., Hussain, S. A., Young, A. M., Cook, J. E., Hodgkin, E., Seymour, L.W., Kerr, D. J. (2001). Pre-clinical and clinical study of QC12, a water-soluble, pro-drug of quercetin, *Annals of Oncology, 12*(2), 245-248.

Nagai, N., Nakai, A., Nagata, K. (1995). Quercetin suppresses heat shock response by down-regulation of HSF1, *Biochemical and Biophysical Research Communications, 208*(3), 1099-1105.

Nakai, A., Tanabe, M., Kawazoe, Y., Inazawa, J., Morimoto, R. I., Nagata, K., (1997). HSF4, a new member of the human heat shock factor family which lacks properties of a transcriptional activator, *Mol Cell Biol 17*, 469–481.

Nakai, A., Kawazoe, Y., Tanabe, M., Nagata, K., Morimoto, R. I. (1995). The DNA-binding properties of two heat shock factors, HSF1 and HSF3, are induced in the avian erythroblast cell line HD6, *Mol Cell Biol 15*, 5268–5278.

Naidu, S. D., Dinkova-Kostova, A. T. (2017). Regulation of the mammalian heat shock factor 1, *The FEBS Journal, 284*(11), 1606-1627.

Neudegger, T., Verghese, J., Hayer-Hartl, M., Hartl, F. U., & Bracher, A. (2016). "Structure of human heat-shock transcription factor 1 in complex with DNA", *Nature Structural & Molecular Biology*, 23(2), 140.

Nishibata, Y., Itai, A. (1993). Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation, *Journal of Medicinal Chemistry, 36*(20), 2921-2928.

Oommen, D. and Prise, K.M. (2012). KNK437, abrogates hypoxia-induced radioresistance by dual targeting of the AKT and HIF-1alpha survival pathways, *Biochem. Biophys. Res. Commun. 421*, 538–543.

Ostling, P., Bjork, J. K., Roos-Mattjus, P., Mezger, V., Sistonen, L. (2007). Heat shock factor 2 (HSF2) contributes to inducible expression of hsp genes through interplay with HSF1, *J Biol Chem 282*, 7077–7086.

Öztürk, E., Kahveci, N., Özlük, K., Yılmazlar, T. (2009). Isı şok proteinleri, *Turkish Journal of Surgery/Ulusal Cerrahi Dergisi*, *25*(4), 131-136.

Pagadala, N. S., Syed, K., Tuszynski, J. (2017). Software for molecular docking: a review, *Biophysical Reviews*, *9*(2), 91-102.

Phillips, P. A., Dudeja, V., McCarroll, J. A., Borja-Cacho, D., Dawra, R. K., Grizzle, W. E., Saluja, A. K. (2007). Triptolide induces pancreatic cancer cell death via inhibition of heat shock protein 70, *Cancer Research*, *67*(19), 9407-9416.

Pierro, M., Elber, R., & Leimkuhler, B. (2015). A Stochastic Algorithm for the Isobaric–Isothermal Ensemble with Ewald Summations for All Long Range Forces, *Journal of chemical theory and computation*, *11*(12), 5624-5637.

Rabindran, S.K., Haroun, R.I., Clos, J., Wisniewski, J., & Wu, C. (1993). Regulation of heat shock factor trimer formation: role of a conserved leucine zipper, *Science 259*, 230–234.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in Drosophila, *Experientia*, *18*(12), 571-573.

Ritossa, F. (1996). Discovery of the heat shock response, *Cell Stress Chaperones 1*, 97–98.

Rye, C. S., Chessum, N. E., Lamont, S., Pike, K. G., Faulder, P., Demeritt, J., ... & Isaac, R. (2016). Discovery of 4, 6-disubstituted pyrimidines as potent inhibitors of the heat shock factor 1 (HSF1) stress pathway and CDK9. *Medchemcomm*, *7*(8), 1580-1586.

Sandqvist, A., Bjork, JK., Akerfelt, M., Chitikova, Z., Grichine, A., Vourc'h, C., Jolly, C., Salminen, TA., Nymalm, Y., & Sistonen L. (2009). Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli, *Mol Biol Cell 20*, 1340–1347.

Santagata, S., Xu, Y. M., Wijeratne, E. K., Kontnik, R., Rooney, C., Perley, C. C., ... & Lindquist, S. (2012). Using the heat-shock response to discover anticancer compounds that target protein homeostasis, *ACS chemical biology*, *7*(2), 340-349.

Santagata, S., Mendillo, M. L., Tang, Y. C., Subramanian, A., Perley, C. C., Roche, S. P., ... & Amon, A. (2013). Tight coordination of protein translation and HSF1 activation supports the anabolic malignant state, *Science*, *341*(6143), 1238303.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on

virtual screening enrichments, *Journal of computer-aided molecular design, 27*(3), 221-234.

Schrodinger, L. L. C. (2010). The PyMOL molecular graphics system, *Version, 1*(5), 0.

Schrödinger Release 2020-2: LigPrep, Schrödinger, LLC, New York, NY, 2020.

Schrödinger Release 2020-2: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2020.

Schrödinger Release 2020-2: PrimeX, Schrödinger, LLC, New York, NY, 2020.

Schrödinger Release 2020-2: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2020. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2020.

Schuetz, TJ., Gallo, GJ., Sheldon, L., Tempst, P., & Kingston, RE., (1991). Isolation of a cDNA for HSF2: evidence for two heat shock factor genes in humans, *Proc Natl Acad Sci USA 88*, 6911–6915.

Sharma, C., & Seo, Y. H. (2018). Small molecule inhibitors of HSF1-activated pathways as potential next-generation anticancer therapeutics, *Molecules, 23*(11), 2757.

Sotriffer, C. A., Gohlke, H., Klebe, G. (2002). Docking into knowledge-based potential fields: a comparative evaluation of DrugScore, *Journal of medicinal chemistry, 45*(10), 1967-1970.

Sotriffer, C. A., Sanschagrin, P., Matter, H., Klebe, G. (2008). SFCscore: scoring functions for affinity prediction of protein-ligand complexes, *Proteins: Structure, Function, and Bioinformatics, 73*(2), 395–419.

Sterling, T., Irwin, J. J. (2015). ZINC 15–ligand discovery for everyone, *Journal of Chemical İnformation and Modeling, 55*(11), 2324-2337.

Takii, R., Fujimoto, M., Tan, K., Takaki, E., Hayashida, N., Nakato,R., Shirahige, K., Nakai, A. (2015). ATF1 modulates the heat shock response by regulating the stress-inducible heat shock factor 1 transcription complex, *Mol. Cell. Biol., 35*, 11–25.

Thomsen, R. (2003). Flexible ligand docking using evolutionary algorithms: investigating the effects of variation operators and local search hybrids, *Biosystems, 72*(1-2), 57-73.

Tissiéres, A., Mitchell, H. K., Tracy, U. M. (1974). Protein synthesis in salivary glands of Drosophila melanogaster: relation to chromosome puffs, *Journal of Molecular Biology, 84*(3), 389-398.

Trott, O., Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *Journal of Computational Chemistry, 31*(2), 455-461.

Vilaboa, N., Boré, A., Martin-Saavedra, F., Bayford, M., Winfield, N., Firth-Clark, S., ..., Voellmy, R. (2017). New inhibitor targeting human transcription factor HSF1: effects on the heat shock response and tumor cell survival, *Nucleic Acids Research, 45*(10), 5797-5817.

Wang, H. H., Mao, C. Y., Teng, L. S., & Cao, J. (2006). Recent advances in heat shock protein-based cancer vaccines, *Hepatobiliary & pancreatic diseases international: HBPD INT*, *5*(1), 22-27.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., Case, D. A. (2004). Development and testing of a general AMBER force field, *Journal of Computational Chemistry*, *25*, 1157-1174.

Yoon, T., Kang, G. Y., Han, A. R., Seo, E. K., Lee, Y. S. (2014). 2, 4-Bis (4-hydroxybenzyl) phenol inhibits heat shock transcription factor 1 and sensitizes lung cancer cells to conventional anticancer modalities, *Journal of Natural Products*, *77*(5), 1123-1129.

Yoon, Y. J., Kim, J. A., Shin, K. D., Shin, D. S., Han, Y. M., Lee, Y. J., Lee J. S., Kwon B. M., Han, D. C. (2011). KRIBB11 inhibits HSP70 synthesis through inhibition of heat shock factor 1 function by impairing the recruitment of positive transcription elongation factor b to the hsp70 promoter, *Journal of Biological Chemistry*, *286*(3), 1737-1747.

Yue, S. Y. (1990). Distance-constrained molecular docking by simulated annealing, *Protein Engineering, Design and Selection*, *4*(2), 177-184.

Yu, Y., Hamza, A., Zhang, T., Gu, M., Zou, P., Newman, B., Li, Y., Gunatilaka, A. A., Zhan, C. G., Sun, D. (2010). Withaferin A targets heat shock protein 90 in pancreatic cancer cells, *Biochem Pharmacol 79*, 542–551.

Zaarur, N., Gabai, V. L., Porco, J. A., Calderwood, S., Sherman, M. Y. (2006). Targeting heat shock response to sensitize cancer cells to proteasome and Hsp90 inhibitors, *Cancer Research, 66(3)*, 1783-1791.

Zhang, D., Zhang, B. (2016). Selective killing of cancer cells by small molecules targeting heat shock stress response, *Biochemical and Biophysical Research Communications, 478(4)*, 1509-1514.

Zingarelli, B., Hake, P. W., O'Connor, M., Denenberg, A., Wong, H. R., Kong, S., Aronow, B. J. (2004). Differential regulation of activator protein-1 and heat shock factor-1 in myocardial ischemia and reperfusion injury: role of poly(ADP-ribose) polymerase-1, *Am J Physiol Heart Circ Physiol 286*, H1408–H1415.

Url-1 <https://www.drugs.com>, data retrieved 14.07.2020.

Url-2 <https://www.who.int/health-topics/cancer#tab=tab_1>, data retrieved 14.07.2020.

CURRICULUM VITAE

Name Surname **:** Gülbahar Bozan

Place and Date of Birth **:** Adıyaman/ 25.01.1993

E-Mail **:** bozang@itu.edu.tr

EDUCATION

- **B.Sc. :** 2011, Istanbul Technical University, Faculty of Science and Letters, Department of Chemistry