Combined Ligand and Structure Guided Protocol for the Virtual Screening of Activators of NRF2- Associated Cellular process in Cardiovascular Diseases

By

Nimra Nisar

(NUST00000274523-MSBI-Fall18)

(MS Bioinformatics)

Supervised by:

Dr. Ishrat Jabeen

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Sciences and Technology

Islamabad, Pakistan

Combined Ligand and Structure Guided Protocol for the Virtual Screening of Activators of NRF2- Associated Cellular process in Cardiovascular Diseases

A thesis submitted in partial fulfilment of the requirement for the degree of Master's in Bioinformatics.

By

Nimra Nisar

(NUST00000274523-MSBI-Fall18)

(MS Bioinformatics)

Supervised by:

Dr. Ishrat Jabeen

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Sciences and Technology

Islamabad, Pakistan.

Dedication

*I dedicate this thesis to **my parents** and **friends** who have been a great source of inspiration and support.*

Certificate of Originality

I hereby declare that this thesis covers my very own struggle and research work. Moreover, none of its contents are plagiarized or submitted for any other higher degree. The assistance of other people in this work is acknowledged and referenced.

_____

Nimra Nisar

(NUST00000274523-MSBI-Fall18)

Acknowledgment

In the name of Allah, the Most Gracious and the Most Merciful, all praises to Allah for the strengths and His blessing in completing this research project.

First and foremost, I would like to express my greatest gratitude to my supervisor Dr. Ishrat Jabeen, who has motivated, guided, and helped me a lot throughout the project and believed in me. Her invaluable help of constructive comments and suggestions throughout the research and thesis work have contributed to the success of this study. I am very obliged that she has given me the opportunity to work and devoted her time from the busy schedule. It was her firm direction and encouragement in the hard period of this project that makes me capable to complete my project in the due time. Her positive attitude and dedication to the research always keeps me on the track and raise me from the scratch to the internationally capable researcher in the field of "Computational Drug Design" throughout my master's journey that has led to this thesis.

I would like to extend my thanks to the NUST and I treasure the research environment provided by SINES. Moreover, I am thankful to my Guidance and Examination Committee (GEC) members; Dr. Muhammad Tariq Saeed (SINES) and Dr. Zamir Hussain (SINES) who provide me with constant assistance throughout my research phase.

I'd like to express my appreciation to Principal RCMS, all my teachers, all the technicians, attendants, and exam branch staff for their co-operations. I owe my sincere gratitude to Pharmacoinformatics research group specially, Dr. Yusra Sajid Kianai, Maharij Jadoon, Maria Ehsan, Fatima, and Humaira Ismatullah for their encouragement, countless support and healthy discussions. Very special appreciations must be given to my friends; Umar Ali, Ali Raza Khosa, Dur-e-Shahwar and Monaza Irshad for the endless support and encouragement.

Last but not the least I want to say thanks to my beloved parents, Mr. Nisar Ahmed and Mrs. Nisar for their prayers, understandings, and never-ending support. Finally, I want to thank all those wished this memoir to reach its successful completion.

Contents

Figures

Tables

Lists of Abbreviations

WHO	**World Health Organization**
FDA	Food and Drug Administration
clogP	calculated logP
LipE	Lipophilic Efficiency
MIF	Molecular Interaction Field
MOE	Molecular Operating Environment
2D	2 Dimensional
3D	3 Dimensional
MSA	Multiple Sequence Alignment
GO	Genetic Optimization for Ligand Docking
LDL	Low- Density Protein
DOPE	Discrete Optimized Protein Energy
RMSD	Root Mean Square Deviations
SAR	Structure-Activity Relationship
HB	Hydrogen Bond
MOE	Molecular Operating Environment
QSAR	Quantitative Structure-Activity Relationship
GRIND	GRID Independent Molecular Descriptors
PLIF	Protein Ligand Interaction Field
CLACC	Consistently Large Auto and Cross Correlograms
KEAP1	Kelch-like ECH-associated protein 1
NRF2	NF-E2 p45-related factor 2
VRS	Virtual Receptor Site
PLS	Partial Least Square
FFD	Fractional Factorial Design
ROS	Reactive Oxygen Species
ARE	Antioxidant Response Element
LDL	Low Density Lipoprotein
CVDs	Cardiovascular Diseases

NO	Nitric Oxide
DMF	Dimethyl Fumarate
ECGE	Epigallocatechin Gallate
SGLT2	Sodium glucose contransporter-2 inhibitors
miRNA	Single Standard non coding RNA
SiRNA	Small inhibitory RNA
SMC	Smooth Muscle Cells
EC	Endothelial Cell
IHD	Ischemic Heart Disease
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
PKC	Protein Kinase C
CS	Cigarette Smoke
NOX	Nitrogen Oxide
EC	Endothelial Cells
UPC	Uncoupling Protein
AKT2	Protein Kinase
IHD	Ischemic Heart Disease
ONOO	Peroxynitrite
HO-1	Hemeoxygenase-1
GA	Genetic Algorithm
LOGP	Octanol–water Partition
LOGD	Octanol–water Distribution
PDB	Protein Data Bank
IC50	Half-maximal inhibitory concentration
HBD	Hydrogen Bond Donor
HBA	Hydrogen Bond Acceptor

Abstract

NRF2 (NF-E2 p45-related factor 2) is a transcriptional factor that controls the production of antioxidant and cytoprotective enzymes that are produced in response to oxidative stress. NRF2 and its primary inhibitor, the E3 ligase adaptor Kelch-like ECH-associated protein 1 (KEAP-1), are essential for redox and metabolic processes to occur. The binding of NRF2 to KEAP-1 in the cytoplasm keeps NRF2 at a low level. Stressors, such as free radicals, however, promote NRF2 translocation to the cell nucleus. Nuclear NRF2 accumulates in the nucleus, allowing it to bind to the antioxidant response element (ARE) of genes that code for antioxidant proteins. Several studies have shown that excessive synthesis of free fatty acids produces significant reactive oxygen species (ROS) production, which reduces NO bioavailability and results in decreased expression of NRF2, resulting in a poor anti-oxidative response. Furthermore, it results in endothelial dysfunction, atherosclerosis, myocardial ischemia-reperfusion injury, hypertension, diabetic vascular disease, and other NO-mediated cardiovascular diseases emerge. As a result, inhibiting KEAP-1 or activating NRF2 is a viable strategy for avoiding endothelial dysfunction caused by the NRF2 pathway. Therefore, in present project, various molecular models were built to probe the 3D structural features of NRF2- KEAP-1 modulator that shields NRF2 from degradation and allows it to move into the nucleus where it controls antioxidant genes. Briefly, the GRIND model for the KEAP1 was developed against the dataset of 91 inhibitors to extract the features that had a positive and negative impact on the activity of the inhibitors. One feature NI- N1 at a distance of **2.80 – 3.20** Å positively impacted the activity and other O-O feature at a distance of **2 – 2.40** Å showing the negative impact on the activity. This model is important since it can help in the development of new drugs/compounds based on these features.

Chapter 1

Introduction

Chapter 1
Introduction

1.1 Cardiovascular diseases

Cardiovascular disease (CVD) is a collective term for the set of heart disorders consisting of ischemic heart disease, heart failure, peripheral arterial disease, injured arteries, thrombosis, stroke, and atherosclerosis [1]. An inflammatory disease of the blood vessel wall called atherosclerosis is considered a dominant cause of CVD. Diagnosing a cardiovascular disease and finding their treatment is the main challenge due to their complex pathophysiology. CVDs are the major cause of death and disability globally [1].

1.2 Symptoms

The majority of the time, there are no symptoms shown by blood vessel disease. A heart attack or stroke is an alarming condition towards the first sign of CVD disease. Symptoms include pain or uneasiness in the center of the chest, in the arms, the left shoulder, elbows, followed by dyspnea; nausea or vomiting; diaphoresis, and turning pale [2]. Fatty deposits escalate the formation inside the arteries (atherosclerosis) and increase the risk of blood clotting that becomes the reason for different chronic diseases and infections. Cardiovascular Disease (CVDs) is a biological term that describes the different groups of diseases or conditions that affect the blood vessels and heart. The major causes of CVDs and diseases related to them are substantial factors, heritable vulnerability, unfortunate choices for lifestyle, and certain momentous circumstances, so on and so forth [8]. People with these symptoms should seek immediate medical attention.

1.3 Risk Factors

The major risk factors that remained consistent in all populations for CVD were explained by inter heart study and include diabetes, dyslipidemia, hypertension, smoking, abdominal obesity [3]. The basic and foremost cause of CVD disease is atherosclerosis. This is caused due to certain reasons. It narrows down the arteries and makes it hard, which is the main reason for the reduction of the flow and delivery of blood and oxygen in the whole body. The cellular waste, bulks of LDL cholesterol, and surrounding materials are involved with the inner coronary artery walls characterized by commemorative formation accordingly. Mainly, when different kinds of interactions occur between endothelial cells, macrophages, and LDLs it causes Atherosclerotic plaques by molecules that bring changes persuaded by hormones, oxidative species, cytokines, and growth factors, etc. Hence, the leading

accumulation to the intima results for the development of atherosclerosis is due to the endocytosis oxidized LDLs that ensue inside the macrophages and makes the whole process very slow [9]. These risk factors help in finding the possible approaches for the prevention of CVD worldwide.

1.4 Prevention of CVD

Main areas that are targeted for basic prevention of CVD include lifestyle modifications focusing on exercise, diet, weight, smoking. Exercise is having a positive impact on CVD as an increase in physical activity decreases the risk of CVD [3]. A person's diet also plays a major role in CVD risk. A diet that is rich in fiber, vegetable and fruits intake and less use of simple sugars and salt is considered as cardio protective. Smoking is considered a crucial risk factor for CVD as it doubles the 10 years CVD mortality rate, so quitting smoking can reduce the risk of CVD [3]. The weight of an individual is also a risk factor for CVD so for reduction of CVD maintenance of a healthy weight is important.

1.5 Global Burden

CVDs are the prominent reason for worldwide mortality, hence decreasing the quality of life. About 17.9 million people die each year from CVDs reported by WHO, and an approximate 32% of all fatalities globally [4]. According to recent studies, approximately 17.8 million deaths occur worldwide annually, along with another 35.6 million years lived with disability. Moreover, in developing and underdeveloped countries nearly 80% of global CVD deaths[4]. Cardiovascular disease (CVD) is considered one of the vital causes of death around the globe. According to the Global Heath Projection Organization's report, CVD will be marked as the most prominent cause of death all around the world in 2030. In 2016 almost 31 percent of world fatalities were caused to cardiovascular disease (CVD) [8].

In September 2018, a renowned cardiologist, and a head of the pre-cautionary cardiology Department, at the National Institute of Cardiovascular Diseases (NICVD) Prof. Khawar Kazmi, have written an article in the international 'THE NEWS'. The article is all about the ratio of people who died due to heart diseases. He said and I quote that "In Pakistan, three years back, twelve (12) people die due to heart diseases hourly, but just in three years the figures have changed and increased very surprisingly and now forty-six (46) dying because of hearts attacks per hour" [9]. Similarly, the World Health Organization (WHO) accounts that, mortalities involve 29% of the death ratio due to heart attacks. Generally,

they claim that around 406,870 individuals die each year in Pakistan due to heart diseases [10].

1.6 Reason of Cardiovascular Diseases

Cardiovascular disease (CVD) is a serious problem caused approximately 15.2 M death globally in 2016 [6]. The endothelium's loss of function is a major reason for risk of (CVD) which leads to the beginning of athero-sclerosis [7]. Deterioration of endothelial is linked to efficient alterations that reduce bioavailability of nitric oxide (NO) and, as a result, increase the risk of cardiovascular disease (CVD) [8]. Cellular damage and dysregulation are caused by a sustained inability to prevent excessive generation of reactive-oxygen-species

(ROS), and deregulation of the anti-oxidant defense mechanism of endothelium [7]. However, in recent years, increased awareness of the crucial signaling role free radicals play within and between cells has supplanted the basic assumption that mostly unrestricted radicals are harmful and causes disease, which may cause the cardiovascular disease. There are few symptoms of cardiovascular disease which includes tightness of chest or pressure, difficulty in breathing, fainting and dizziness, severe fatigue, building up fluid in blood, fasten the heartbeat, pain in the body especially legs and arms, abdomen pain and vomiting. However, there are many defenses mechanism which prevent our body from the attack of cardiovascular diseases.

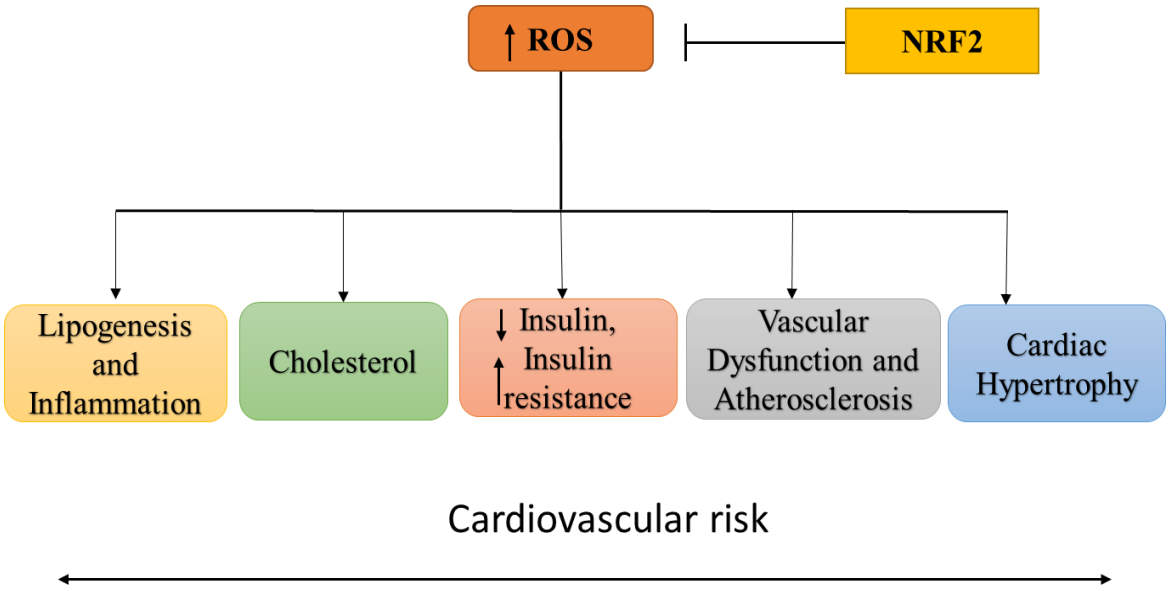


Figure 1. 1: Nrf2 role in cardiovascular diseases

1.7 Importance of NRF2

The transcription factor Nrf2 (nuclear factor erythroid 2–related factor 2), a major regulator of

antioxidant and cellular protective genes, is predominantly active in response to oxidative stress. Nrf2 is generally linked with Keap-1, a repressor protein, and is produced in the cytoplasm of healthy cells on a regular basis. As a result of this relationship, low levels of free Nrf2 are maintained. The translocation of Nrf2 to the cell nucleus is facilitated by free radicals, for example. Nuclear Nrf2 binds to the antioxidant response element of genes that code for antioxidant proteins, causing it to accumulate in the nucleus. Although little is known about the role of Nrf2 in the cardiovascular system, statistics suggest that decreased Nrf2 activity contributes to oxidative stress, which facilitates in the pathogenesis of cardiovascular diseases such obesity, diabetes, and atherosclerosis.

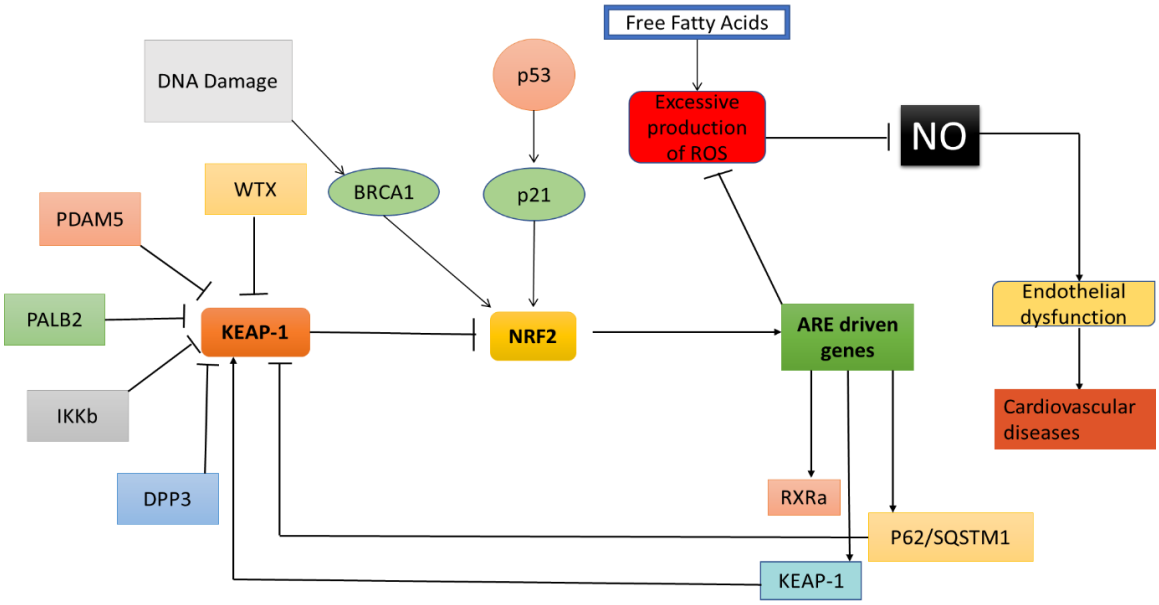


Figure 1. 2: Biological Regulatory Network (BRN) of cardiovascular diseases

Exogenous free radical sources combined with high abundance in endogenous production result in an imbalance that leads to a variety of degenerative diseases, including cardiovascular disease and ageing in general. By binding to and activating promoters encoding the antioxidant response element (ARE), NFE2L2 (also known as Nrf2) regulates the production of antioxidant and cytoprotective genes. Many pathways govern Nrf2 activity, demonstrating the need of precise regulation for healthy cell function. Both hypo- and hyperactivation of Nrf2 have been linked to various aspects of cardiovascular disease. In the development of drugs to lessen the burden of cardiovascular disease, regulated activation of Nrf2 or downstream genes might be beneficial.

In reaction to oxidative stress, NRF2 (NF-E2 p45-related factor 2) is a transcription factor that regulates the production of antioxidant and cytoprotective enzymes. NRF2 and its major inhibitor, the E3 ligase adaptor Kelch-like ECH-associated protein 1 (KEAP-1), are

also necessary for redox and metabolic balance. NRF2 is maintained at a low level in the cytoplasm via binding to KEAP-1. NRF2 translocation to the nucleus, on the other hand, is promoted by free radicals. Nuclear NRF2 binds to the antioxidant response element (ARE) of antioxidant protein-coding genes as it accumulates in the nucleus.

1.8 KEAP-1 Dependent Regulation of NRF2 Activity

NRF2 is suppressed in the cytoplasm during physiological conditions, when oxidant species are in balance with antioxidant protein activity. The E3 ubiquitin ligase KEAP-1 (Kelch ECH-associating protein 1) retains NRF2 protein at low levels and targets it for breakdown by the 26S proteasome. NRF2 has a half-life of around 20 minutes under normal homeostatic conditions. [33]. When ROS or NRF2 inducers such as EGCG, sulforaphane, dimethyl fumarate (DMF), and tert-butylhydroquinone (tBHQ) are present, the conformation of KEAP-1-cysteine residues changes, and the DLG motif no longer interacts with the Kelch domain, although the ETGE motif stays attached to NRF2. Because KErdeAP-1 fails to link with the E2 ubiquitin-conjugating enzyme, NRF2 is no longer targeted for ubiquitination and degradation. (See Figure 2) NRF2 is translocated from the cytosol to the nucleus, where it forms a dimer with sMaf protein and binds to ARE sequences, allowing downstream target genes to be produced. [34][35].

1.9 Problem Statement and Proposed Solution

In this research work the following problem statement was identified and give proposed solution

- Hypoactivation of NRF2 due to excessive production of ROS may cause to many cardiovascular diseases.
- Modulation of normal physiological function of NRF2 by using combined ligand and structure guided protocol could inhibit the KEAP-1 or activating physiological function of NRF2. It is a viable strategy for avoiding endothelial dysfunction caused by the KEAP1-NRF2 pathway.

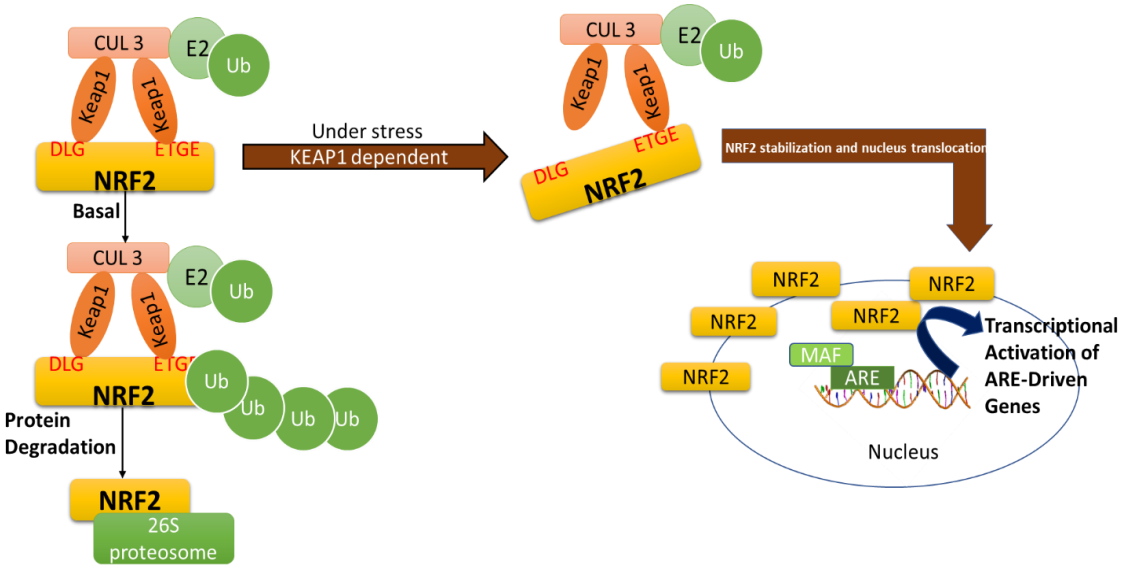


Figure 1. 3: Activation of NRF2

1.10 Aims and Objective of the study:

- Activation of NRF2 and inhibition of KEAP-1 rescue the normal physiological function of NRF2 in NO-mediated cardiovascular disorders.
- To develop both NRF2 and KEAP-1 binding hypotheses.
- To build prediction models for NRF2 activation and KEAP-1 inhibition.

A new library of NRF2-KEAP-1 modulators was designed using structure and ligand guidance.

Chapter 2
Literature Review

Chapter 2
Literature Review

2.1 Oxidative Stress Pathway

Nowadays, oxidative stress is known to have at least two roles within the cell: the creation of cellular damage and the engagement in various signaling cascades. The hunt for a clear link among cardiovascular disease (CVD) as well as the consequences of oxidative stress has consumed a significant amount of effort thus far. Researchers offered an overview of the various origins and forms of reactive oxygen species in CVD, highlighted the link between CVD with oxidative stress. They described the most notable molecules involved in CVD pathogenesis in this study [12]. The details of typical therapeutic interventions for cardiovascular discomfort, as well as how a few of them work on ROS-related pathways including molecules, have been revealed by researchers. Novel therapeutics are discussed, as well as newly proposed ROS diagnostics and future hurdles in the field. The quest for a greater understanding of how ROS contributes to the pathogenesis of CVD is far from over. New techniques and more appropriate biomarkers, according to the authors, are required to achieve the latter.

Authors of another study stated that heart failure is indeed a global pandemic that affects 26 million people worldwide and is now on the rise [13]. In patients with cardiovascular disease, unbalanced redox homeostasis changes the structure and activity of cardiac cells, leading to contractile inefficiency, myocardial enlargement, and fibrosis. Several targets and medicines acting on individuals, such as SGLT2 inhibitors, siRNA, opioids, miRNA, interleukin-1 and vasodilators are being studied for heart failure. According to the authors, and Nuclear factor erythroid-2-related factor-2 NRF2 is one of them. NRF2 is a dominant transcription factor that is found throughout most tissues and plays a key role in the activation of antioxidants pathways involving enzymes found in the myocardium. As per the researchers, overexpression of NRF2 will reduce the increase in circulatory stress and play a favorable function in cardiovascular illnesses. Researchers also highlighted the present evidence that NRF2 activators have a positive role in heart failure.

Cardiovascular Diseases (CVD) are highly complex with heterogeneous pathophysiologic pathways, according to the experts, and elevated oxidative stress has already been identified as one of the peers with similar etiologies. The presence of antioxidants and Reactive Oxygen Species (ROS) must be maintained in a delicate balance for such cells to operate

normally. Excessive quantities of ROS destroy cellular macromolecules including DNA, proteins, and lipids resulting in necrosis and induced apoptosis. Researchers went on to say that CVD is the leading cause of mortality globally, and that oxidative stress affects a variety of illnesses [14]. Increased ROS causes nitric oxide depletion and vasoconstriction, resulting in arterial hypertension. ROS has been proven to increase the formation of atherosclerotic plaques. The pictural vision of the Atherosclerotic plaque development process is shown in figure 2.1.

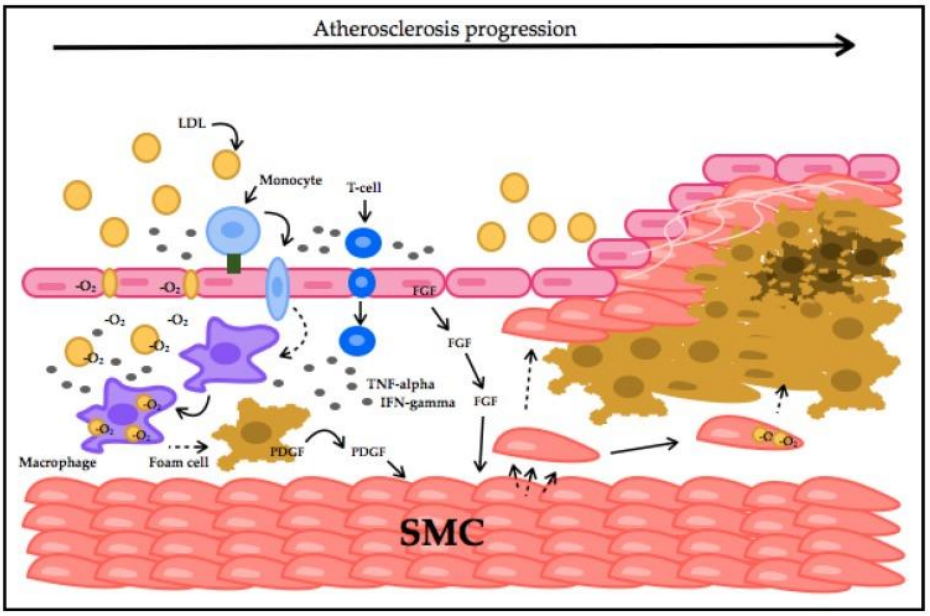


Figure 2. 1: Atherosclerotic plaque development process.

In a nutshell, the goal of this study was to offer an overview of oxidative stress, with a special emphasis on endothelial function, before delving deeper into the function of oxidative stress in its most common of these disorders. Finally, prospective nutraceuticals and diets that may help reduce the impact of oxidative stress in CVD were discussed.

Natural products have long been used to improve health and have proven to be a great source for novel medicine development. The significant potential of the natural chemicals and medicinal herbs for the cure of metabolic and cardiovascular illnesses, which are worldwide health issues with rising frequency, was examined in this research [15]. There was also a look back at the history of cholesterol and biguanides, two major natural compounds in the combat in contradiction of metabolic disorders. The goal of the provided study was to assist like an ‘opening’ to this recent issue of Molecules by highlighting major historical discoveries, new progress, and future directions on the properties of natural compounds for cardiovascular as well as metabolic illness prevention and treatment.

According to the findings, free radicals as well as other reactive nitrogen and oxygen species play a dual role in biological systems. As they can induce oxidative damage and tissue malfunction as well as operate as molecular signals that activate stress reactions that are advantageous to the body. Mitochondria have been considered to play a significant role in the tissue oxidative injury and malfunction and to provide safety against high tissue dysfunction over multiple mechanisms, involving stimulation of permeability transition pores opening, according to the researchers. However, researchers showed that, in addition to establishing mitochondria's involvement in the tissue of oxidative stress and protection, there is a synergistic relationship between mitochondria and cellular sources of ROS [16]. As a result, it is now widely accepted that, under varying conditions, all ROS cellular sources contributed significantly to procedures that oxidatively damaged tissues while also ensuring their survival, like autophagy and apoptosis.

2.2 Importance of NRF2

According to recent research, Vascular cells can generate Reactive Oxygen Species (ROS) via NAD(P)H oxidase, which could be relevant in vascular injury. Pathogenic involvement of vascular NAD(P)H oxidase in diabetics, on the other hand, is uncertain. Researchers used nuclear magnetic resonance spectroscopy to investigate the effects of excess glucose plus free fatty acid on ROS manufacture in the Smooth Muscle Cells (SMC) and the Endothelial Cell (EC) [17]. D-iphenylene iodonium and GF109203X were able to reverse the increase of free radical generation due to high glucose levels. Palmitate exposure enhanced free radical generation as well as diacylglycerol levels and PKC activity. Both were effective in restoring this rise to the values obtained. According to their results, high glucose levels and phosphonates may both induce ROS generation in vascular SMC and EC via PKC activation of the NAD(P)H oxidase. So, their findings could play a key role in the excessive development of atherosclerosis in insulin resistance syndrome and diabetes patients in the future.

Although pathophysiological processes that cause diabetic retinopathy (DR) are well understood, the participation of preventive pathways has attracted less attention. NRF2 is a transcription feature of oxidative stress and has anti-inflammatory properties. The goal of this research was to look at the possibility of NRF2 acting as a protective mechanism in DR [18]. Immunoreactivity was used to look at NRF2 activity in the retinas of human donors and mice. In the humanoid Muller cell line MIO-M1, the effect of NRF2 regulation

on oxidative stress was investigated. Finally, findings showed that NRF2 is a key protective component in the advancement of DR and that improving the NRF2 pathway could be a possible therapeutic strategy.

Based on accumulating evidence from laboratory oxidative disease reproductions and human investigations, researchers addressed the importance of Nrf2 in the airways in this study [19]. Nrf2 is a ubiquitous key transcription factor which controls the expression of antioxidant enzymes and cytoprotective protein via antioxidant response elements (AREs). Kelch-like ECH-associated protein 1 (Keap1) reduces cellular Nrf2 in cytoplasm and indorses its dephosphorylation in the absence of stress. Various stimuli, such as oxidants, antioxidants, and chemo preventive drugs, can stimulate Nrf2. Figure 2.2 shows balance among oxidative stress and antioxidant defense system.

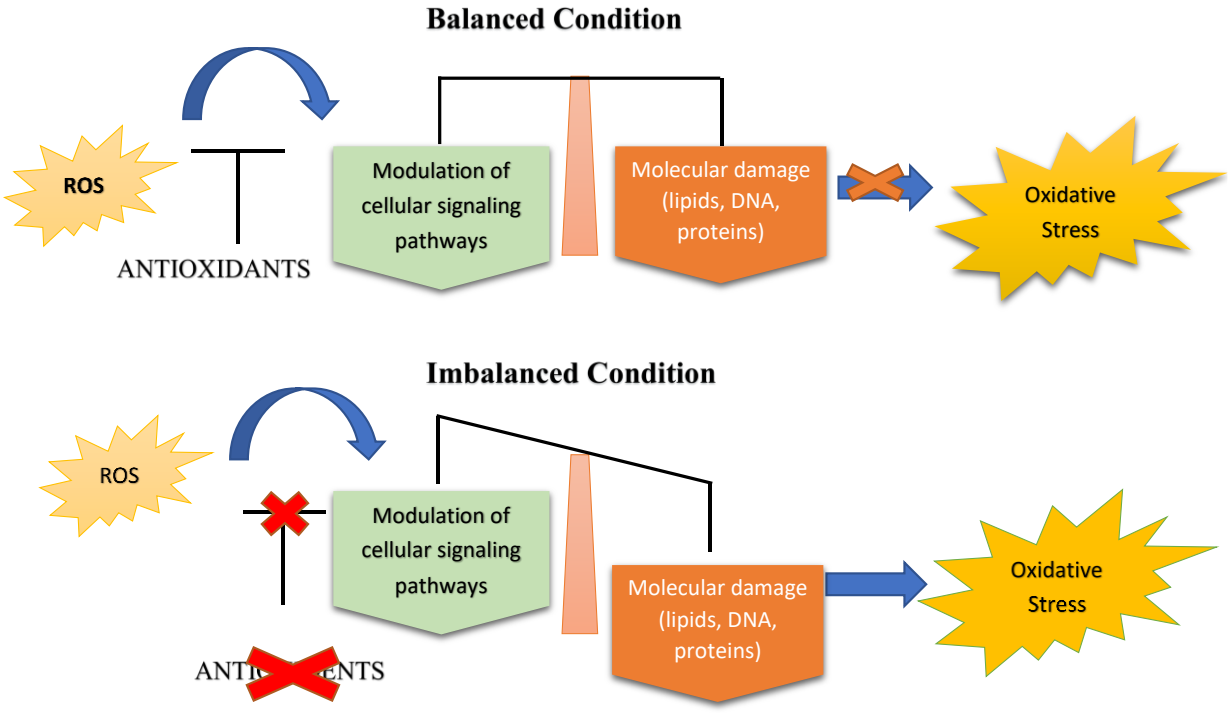


Figure 2. 2: Balance among antioxidant defense system and oxidative stress.

Against oxidative stress, aberrant inflammatory and immunological responses, apoptotic, and carcinogenesis, Nrf2 activates cellular rescuing pathways. The use of Nrf2 germ-line transgenic mice in laboratory examples of human illnesses in liver, intestinal system, lung, brain, kidney, circulation, and immunological or nervous system has revealed a wide range of interest in the role for Nrf2. Absence of Nrf2 in the lungs exacerbated toxic effects caused by a variety of oxidative slurs, including additional respiratory therapy, cigarette smoke,

allergens, viruses, bacterial endotoxin, and environmental pollution. In associates of individuals with acute lung respiratory failure syndrome as well as lung disease, loss of functionality has been linked to NRF2 somatic and the epigenetic changes in KEAP1 and NRF2, indicating that NRF2 plays a key role in these disorders.

In this research, authors proposed that Nrf2 is a promising therapeutic approach for the cardiovascular illness, but they included experiments to demonstrate the processes of Nrf2 cardio protection [20]. Nrf2 is a key regulator the appearance of a variety of antioxidant genes as well as other cytoprotective Phase-2 detoxifying enzymes at both the basal and inactivating levels. In the cardiovascular health system, Nrf2 is widely expressed. While some Nrf2 downstream proteins have been linked to protection versus cardiovascular pathogenicity, the exact role of the Nrf2 in cardiovascular system has yet to be determined. Conversely, accumulating data indicates that Nrf2 is a crucial regulator of the cardiovascular balance by suppressing oxidative stress, a significant cause of cardiovascular disease initiation and progression. As a result, Nrf2 seems to be a promising therapeutic target in cardiovascular treatment diseases.

2.3 NRF2 and Mitochondrial Dynamics in the Cardiovascular Dynamics

Cardiovascular diseases are leading cause of mortality in the world, and it encompasses a wide range of illnesses. Stroke, congestive heart failure, and Ischemic Heart Disease (IHD) are the most likely reasons of CVD morbidity and mortality. ROS is a fundamental mediator and a common element in CVD, and it is elevated by various illnesses such as diabetes, stress, and smoking. Smoking is among the leading risk factors of CVD, yet it is still one of the most preventable [21]. Oxidative stress, a prothrombotic condition, inflammation, abnormal lipid metabolism, plus hypoxia are all essential aspects of the pathophysiology linked with smoking. Tobacco CVD involves many molecular processes. Dysregulations involving reactive oxygen species (ROS) formation and metabolism, on either hand, are primarily responsible for development of various CVDs, and NADPH oxidase has been identified as a source of ROS involved in CVD pathophysiology [22]. Cigarette smoke (CS) treatments have been shown to activate NOX and produce ROS in isolated blood arteries and cultured vascular endothelial cells, including endothelium and smooth muscle cells. Oxidative stress caused by NOX has also been proven in animal experiments. To assess the danger of smoking to human health, designate high-risk

groupings, and create ways to prevent or cure cigarette CVD, a deep understanding of the NOX's role would be beneficial.

According to this study, ROS can promote EC apoptosis and the activate nuclear factor kappa-B (NF-B), boosting adhesion molecules including cytokines which improve monocyte adherence [23]. ROS appear to have a substantial role in the induction and progression for cardiovascular failure in diseases like hyperlipidemia, ischemic heart disease, hypertension, diabetes mellitus, and sudden cardiac death, according to mounting data. The functional effects of ROS produced by migratory inflammatory cells and vascular cells on each cell type are different. Cell proliferation, death, migration, inflammatory epigenetics, and matrix modulation are all instances of these processes. ROS can play a key role in the vascular wall physiology and assist immensely in the improvement of vascular disease through influencing vascular cell function.

The authors of this study concluded that oxidative stress contributes to mitochondrial malfunction, which is linked to bioenergetic abnormalities and changes in mitochondria [24]. This leads to a reduction in transcription resulting in cell damage. When the electron transport in complex-III of the mitochondria is blocked, electrons are released, reducing molecular oxygen into superoxide, and increasing intracellular ROS generation. Considering the current weight of experimental data, it is now commonly accepted that ROS plays a role in the cell and the tissue malfunction and damage induced by liposomes in the diabetes. Researchers added to the growing body of data indicating NADPH oxidase-dependent ROS production in both pancreatic as well as insulin-sensitive tissues. While mitochondria ROS may be significant for regulating Uncoupling Protein (UCP) movement and thereby disrupting cellular energy consumption, NADPH hydroxylase ROS may change signal transmission, insulin production, insulin action, and the cell growth or cell death characteristics. As a result, NADPH oxidase could be a good target for diabetes therapeutic strategies aimed at alleviating the unfavorable effects of glucolipotoxicity.

Furthermore, the ROS can stimulate membrane oxidases, resulting in higher dimethylarginine levels, which compete against L-arginine transporter and active sites on eNOS. The function of respiration chain is modulated by Nrf2, and pharmacological stimulation of Nrf2 protects against toxicity and maintains mitochondrial

homeostasis, presumably via inhibiting Akt2 signaling. Paraquat is a cationic nitrogen herbicide that causes oxidative stress, mitochondrial dysfunction, and injury to multiple organs, including the heart. To date, there are no effective measures in place to combat parathion toxicity. Akt appears to play a role in heart homeostasis, according to latest research. Researchers used a unique Akt2 deletion mouse model to investigate the involvement of Akt2 in acute paraquat exposed heart contractile and mitochondrial damage [25]. Their outcomes from an in vitro investigation demonstrated that stimulating Nrf2 with sulforaphane neutralized the good effect of Akt2 ablation versus paraquat but inhibiting Nrf2 with luteolin replicated the positive effect of Akt2 ablation versus paraquat challenge. Akt2 ablation might protect towards paraquat toxicity-induced cardiomyocyte dysfunction and apoptosis, according to the findings, perhaps through modulation of Nrf2 activation with mitochondrial homeostasis.

Acrolein, a residue of cigarette burning, was discovered to eliminate the KEAP1/Nrf2 pathway and reduce mitochondrial function in a recent study. Using hydroxytyrosol to stimulate Nrf2, researchers demonstrated the ability of Nrf2 to avoid mitochondrial damage in this study [26]. Zeaxanthin (Zea) is indeed a primary carotenoid pigment found in the retina of human and has been linked to a lower risk of age-related macular degeneration when taken regularly. Although its antioxidant properties, the underlying chemical modes of action of Zea are currently unclear. The goal was to figure out how to control Zea's effect on phase-II detoxifying enzymes. According to the findings, the Zea demonstrated effects of stimulating phase-II enzymes and increasing GSH content and related to abridged nutritional quality peroxidation inside the retina, liver, heart, and serum of rats. So, it was proved that Zea described as an activator of phase-II enzymes rather than an antioxidant. Zea could improve anti-oxidative capability and reduce cell damage in vivo and in vitro through activating Nrf2-mediated phase-II enzymes.

2.4 NRF2 in Endothelial Dysfunction

The vascular endothelium regulates vascular homeostasis by producing and releasing a variety of active substances. Endothelium loss is a major risk factor for CVD, since it causes atherosclerosis and is linked to functional changes that reduce NO bioavailability and, as a result, contribute to CVD. Endothelial dysfunction is caused by hypoxia, flow disturbances,

and oxidative stress. Cellular damage and dysfunction result from a failure to counteract excessive ROS generation and modulate the antioxidant defense mechanism in the endothelium. NO is produced by combining the eNOS heme group with L-arginine using tetrahydrobiopterin (BH4) as a cofactor, which is required for normal vascular endothelial function. Excess ROS cause BH4 to be converted to 7,8-dihydrobiopterin (BH2), resulting in eNOS uncoupling and the production of O2• instead of NO. (Fig 6). When O2• reacts with NO, the oxidant peroxynitrite (ONOO) is formed. Under hyperglycemic conditions, iNOS overexpression and eNOS uncoupling are well established. Arginase, which is increased in the endothelium of coronary arterioles in hypertension and leads to reduce NO-mediated dilation, is a substrate for L-arginine. Furthermore, ONOO and hydrogen peroxide (H2O2) have been shown to enhance arginase activity or expression in endothelial cells, causing myogenic tone abnormalities. As a result of the depletion of the substrate L-arginine, ROS can cause eNOS uncoupling. The findings of Romero et al., who found that elevated arginase activity caused L-arginine depletion and led to diabetic endothelium dysfunction, back up this theory. ONOO can also impact the formation of ROS by stimulating NADPH oxidases. Furthermore, when the mitochondrial electron transfers in complex III is blocked in diabetes, electrons are released, reducing molecular oxygen to O2• and increasing intracellular ROS generation. Additionally, ROS can activate membrane oxidases, resulting in a increase in asymmetric dimethylarginine levels, which competes for L-arginine transporters and active sites on eNOS. [7].

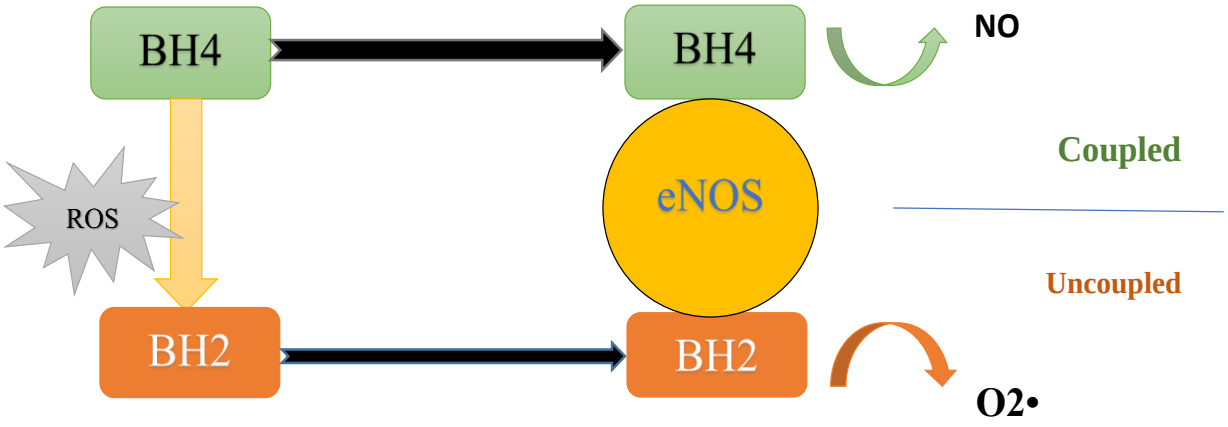


Figure 2. 3: ROS-induced uncoupling of eNOS and the generation of O2•

Recent research has shown that high ROS production caused by free fatty acids (FFAs) in

endothelial cells reduced gene and protein expression of NRF2, NQO1, and HO-1 (Fig 2.3). Likewise, increased NRF2/ARE/HMOX1 signals protected human endothelial cells against TNF-activation. It's possible that mitochondrial ROS activates NRF2 in endothelial cells, triggering a protective response. According to Lo and Hannink's research, the NRF2–KEAP-1 complex attaches to mitochondria via contact with PGAM5, a mitochondrial outer membrane protein, and senses mitochondrial ROS production directly [35].

2.5 Structure of NRF2

Nrf2 is a 605-amino-acid cofactor that was cloned from K562 human leukemia cell line and corresponds to cap-n-collar (CNC) class of fundamental leucine zippers [27]. It is made up of 7 evolutionarily conserved domains designated Neh1 through Neh7, each of which has a distinct function shown in figure 2.4. The Neh1 domain has a CNC-bZIP zipper configuration that produces a heterodimer involving Small Musculoaponeurotic Fibrosarcoma (sMaf), which is essential for Nrf2 engagement to ARE [28]. The regulatory component of Nrf2 for its connection with Keap1 is N-terminal Neh2 region. ETGE motif, that has a strong affinity towards Keap1, and DLGex motif, that has a low selectivity for Keap1, constitute the fundamental structure of "latch and hinge" altogether [29]. The Nrf2-Keap1 connection is disrupted when the ETGE as well as DLG motif is deleted or mutated, resulting in the activation of Nrf2 pathway. The chromo-ATPase DNA-binding (CHD6), that is required for gene transcription of ARE-dependent gene during chromatin remodelling, interacts with C-terminal Neh3 domain. The Coactivators Cyclic Adenosine Monophosphate (cAMP), steroid receptor coactivator 3 and element-binding Proteins (CREB) are essential for the translational activation of Nrf2 [30]. The Neh6 domain plays an essential role in phosphorylation-based and ubiquitination-based regulation of the Nrf2 activity. The DSAPGS and DSGIS modules of Neh6 engage with the kinase-3β and β-transducin string protein to engage the kinase-associated proteins or Roc1 core E3 ubiquitin binder to promote Nrf2 degradation [31]. By associating with the retinoid X receptor, the Neh7 region can decrease Nrf2. Figure 2.4 shows the Structure of Nrf2.

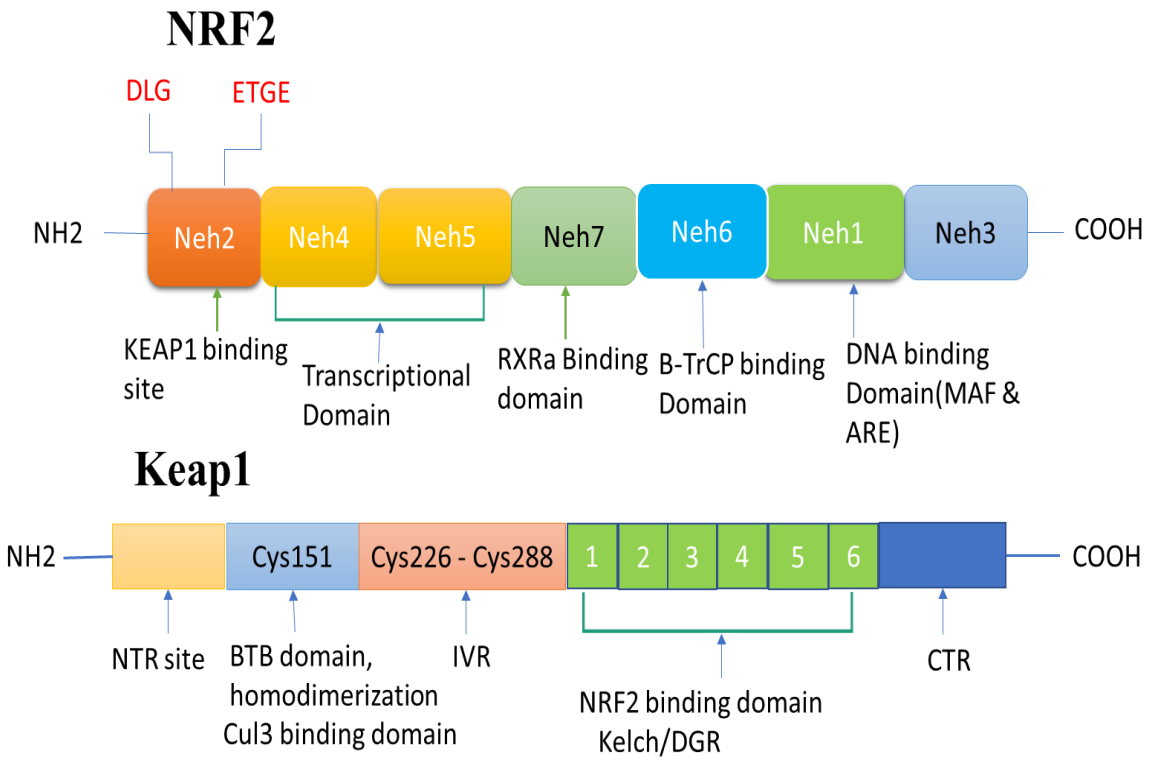


Figure 2. 4: Demonstration of Neh2 and Kelch active domains of NRF2 and KEAP1.

There are seven domains in Nrf2 protein. Neh2 binds to Keap1 via ETGE motifs and DLG in Neh1-Neh7. The ETGE motif has a 200-fold stronger affinity for the Keap1 than DLG motif. CREB/SRC3 binds to Neh4 and Neh5, promoting Nrf2 transcription. Neh7 inhibits Nrf2 through interacting with the retinoid X receptor. Neh6 interacts with TrCP, causing Nrf2 to be ubiquitinated and degraded by proteasomes. The transcription coactivator CHD6 links with Neh3.

Chapter 3
Methodology

Chapter 3
Methodology

3.1 Collection of Datasets:

A diverse dataset of already reported 124 modulators of human NRF2/KEAP1 with known inhibitory potency (IC_{50}) were retrieved from ChEMBL [36] and PubChem databases [37]. Data preprocessing was performed to refine and remove inconsistencies of ligands data set. This preprocessing includes removal of ligands redundancy, removing ligands related to other than humans NRF2/KEAP1. A total of 124 ligands data set from CHEMBL and PubChem, was initially retrieved but after data preprocessing 93 ligands leftover for further analysis. Therefore, we used these 93 ligands for *in-silico* based drug designing [shown in Table 1.1]. The 3D structure of selected compounds were constructed and then energy minimized through MOE 2019.

3.2 Target (Human NRF2/KEAP1) crystal structure retrieval

Crystal structure of human NRF2/KEAP1 in wild type (PDB ID: 6TYP) [38] was taken from PDB. We selected these structured based on these

1. 6typ has 2.50 Å resolution value. This resolution value can be considered as high-resolution structure (the lower the resolution in Å greater will be structural resolution).
2. 6typ has 0.21 R-values, which is measure of the quality of the atomic model obtained from the crystallographic data.
3. The most important reason for select this structure is that mostly most of the research analyst used 6typ structures for their analysis.

These structures were prepared using MOE following protonation, charge fixation and energy minimization using AMBER10 force field [39]. We have 3D structure of our target receptor now we should have knowledge about its binding pocket or residues involved in this pocket. So, we observed binding residues of NRF2/KEAP1 from literature and computational analysis.

3.3 Molecular Docking

Molecular docking against human NRF2/KEAP1 was performed by using Gold suite 53.0 to probe the binding confirmations of 93 inhibitors data within the binding cavity. For the selection of binding cavity using pose selection method where the x, y, z coordinates of an estimated centroid point were specified i.e., x = -17.1700, y = 5.5030, z = 21.7770 respectively. We select all the important binding residue which are confirmed from literature including ARG483, ARG415, TYR334, SER602, SER555, GLN530, and TYR525 [39]. Some binding site residues kept flexible during docking by defining torsion around one or more of their bonding. For each ligand, a maximum of 100 GA (genetic algorithm) runs were configured. Because of its high precision, the slow protocol was chosen for docking. The fitness function in GA was used to rate the generated poses using gold score. External protein-ligand interaction energies (external h-bond and Vander Waal energies) and ligand internal energies (ligand internal Vander Waal energy, torsional strain energy, and ligand intra molecule hydrogen bond energy) make up the gold score fitness function. Therefore, the gold score is calculated through expression as stated below.

GOLD Fitness = Score (hb_ext) + Score (vdw_ext) + Score (hb_int) + Score (vdw_int)

The compounds with high activity generally have a high gold score. However, few compounds show outlier behavior in our dataset. These compounds had high gold score with lower activity or vice versa.

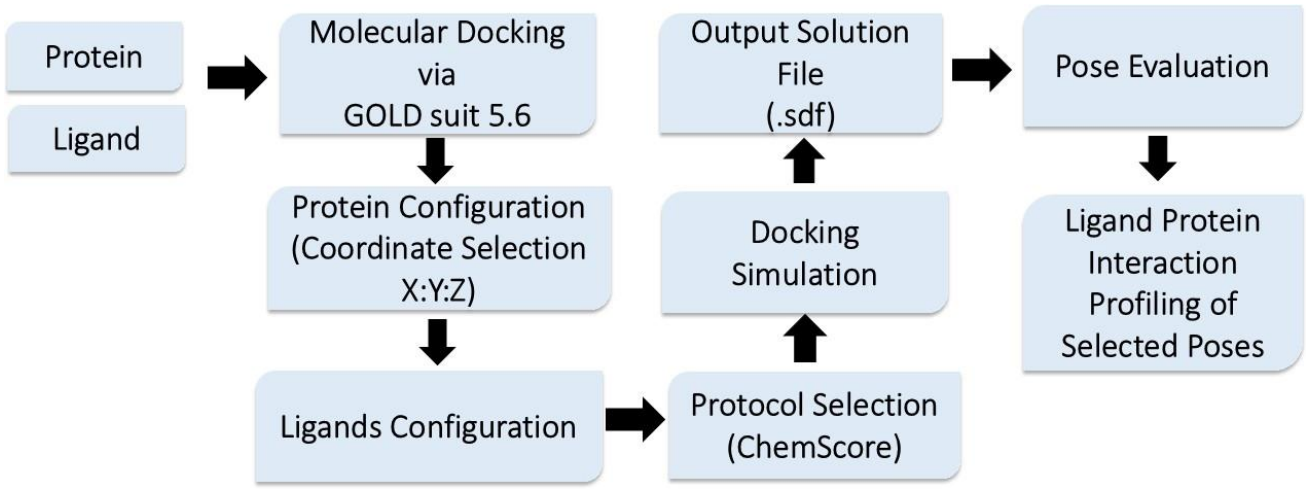


Figure 3. 1: Overall workflow of molecular docking

3.4 Pose Selection and Analysis:

After completing docking, the best pose of each ligand was then studied for the ligand interaction via MOE showed the behavior of ligand binding in a protein to form complex. IC_{50} value converts into the pIC_{50} value by using formula pIC_{50} = -log (IC_{50} in M). Poses were selected from the correlation plot between biological activity on y-axis and gold_score on x-axis. The best pose of each ligand was then studied for the ligand interaction analysis. Furthermore, the interactions generated by ligand interaction via MOE showed the behavior of ligand binding in a protein to form a complex. The highly active compound shows high gold_score and vice versa.

Protein-ligands Interaction Fingerprints technique was used to analyse the protein-ligand interactions through the fingerprints scheme. In this method, interaction of surface contact, hydrogen bonding and hydrophobic are categorized according to the binding cavity through fingerprint method that is representative of the training set protein-ligands complexes. Protein ligand interactions in the form of bits, and the computed output was analyzed in the form of frequency histogram. Furthermore, the final poses were subjected for the identification of binding patternusing GRIND

3.5 GRIND (GRID independent descriptors):

Three-dimensional Quantitative Search Activity Relationship (3D-QSAR) models were built

using an alignment independent technique for non-congeneric series, GRIND. GRIND is suitable for structurally diverse dataset, as it auto-computes descriptors that are independent of alignment. 3D confirmations are the pre-requisite for GRIND, therefore, 3D structures along with their experimentally calculated activity (pIC_{50}) values were imported into the Software Pentacle version 1.06 [45].

3.5.1 Molecular Interaction Field (MIFs)

Interaction between the molecule of ligands and the probes were calculated using field of interactions. These interaction fields represent the virtual receptor site for the actual protein molecule. The 3D lattice mapped the features between the ligand and the probes that contributes towards the activity of the inhibitor data. The probes used for MIFs computation are: N1: Amide Nitrogen – Hydrogen Bond Donor, O: Carboxyl Oxygen -- Hydrogen Bond Acceptor, Dry: Hydrophobic or Lipophilic, Tip: Defines the shape and boundaries in the 3D lattice

3.5.2 Discretization

The algorithm in Amanda [83] was scripted to discretize the computed MIFs. Amanda shortlisted the molecular fields by applying suitable energy cutoffs depending on the size of the molecule contrary to Almond algorithm. The energy cutoffs for probe Hydrogen Bond donor was -4.2 while for Hydrogen Bond Acceptor was -2.6. Furthermore, -0.5 was the energy cutoff for Dry and -0.7 for the Tip probe, respectively.

3.5.3 Encoding

GRIND uses Consistently Large Auto and Cross Correlation (CLACC) encoding analysis for the consistent selection of probes along with their computed distance [83]. Unlike, previous MACC (Maximum Auto and Cross Correlation) algorithm that arbitrarily select the probes leading to inconsistent results. CLACC also aligns the molecules on their movement of inertia, the bulky part of the molecule moves downwards lifting the light part in the upward direction. The interaction energies were computed as per scripted by Amanda by summation of the interacting energies i.e. Lennard-Jones potential (ELJ), hydrogen bond energies (EHB) and electrostatic energy (EEL).

$$E_{ijk} = \sum E_{LJ} + \sum E_{EL} + \sum E_{HB}$$

The calculated distance with respect to their interacting energies were represented in the form of a graph called correllogram. Correllogram comprises of peaks that represents the distance with their respected energies computed by the four Auto (N1-N1, O-O, Dry- Dry and Tip-Tip), and six cross (N1-O, N1- Dry, N1-Tip, O-Dry, O-Tip, Dry-Tip) probes.

3D-QSAR models were built using a powerful regression analysis method, Partial Least Square (PLS). The features of the 3D structures are correlated with respective activity (pIC50) by PLS. The predictive ability of q^2 was validated by Leave One Out (LOO) cross validation. The QSAR models were evaluated based on statistical parameters; the correlation coefficient (r^2), Standard Deviation of Error Prediction (SDEP) and predictive ability of model (q^2). Statistically significant models with $q^2 > 0.5$, SDEP < 1 were analyzed for further interaction studies. Moreover, the structural variance within the imported confirmations was elucidated by Principal Component Analysis (PCA). As we knew that GRIND computed independent set of variables according to the data. Therefore, Fractional Factorial Design (FFD) was performed for the reductions of dimensions for calculated descriptors.

Multiple GRIND models with training data were generated using 3D confirmations as input. The data was divided into 80 to 20 ratio, the 80% of training data was classified into classes according to the docking results and was tested against the 20% of test set data to test the predictive ability of the models with precision and accuracy. Final templates were selected from the final models based on their statistical significance, consistent correlogram and the features that differentiate the actives from in actives.

The complexes of the final templates were simulated at 50ns using Schrodinger. Root Mean Square Deviation (RMSD) plot were analyzed to check the stability of the complexes. Overall workflow of the GRIND is shown in figure 3.2

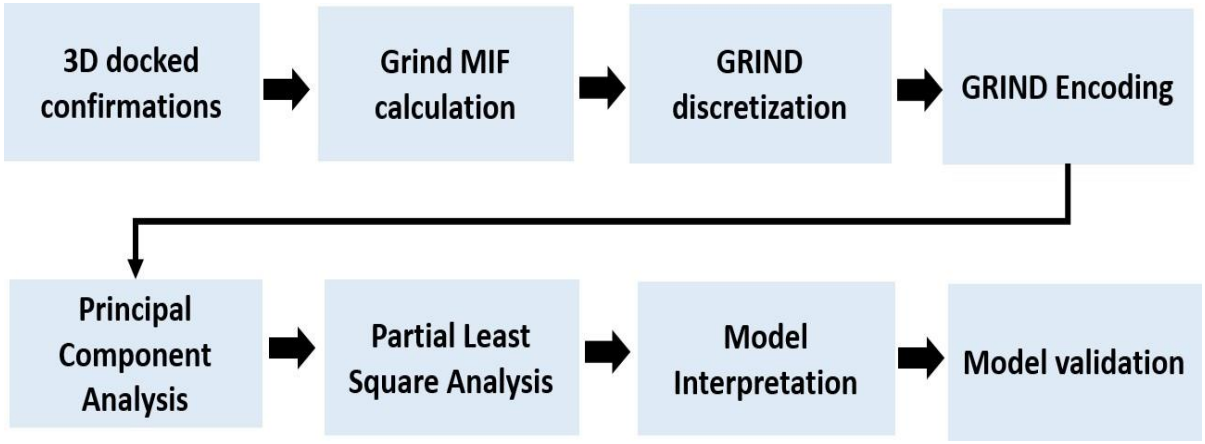
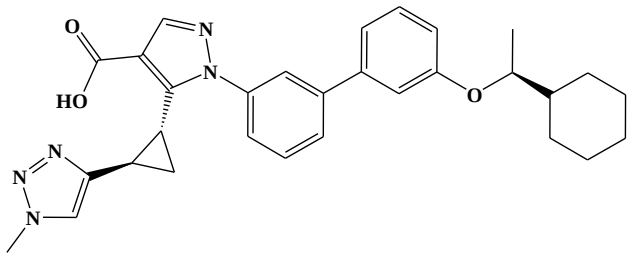
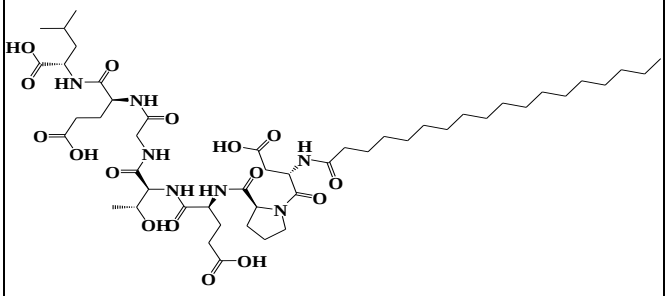
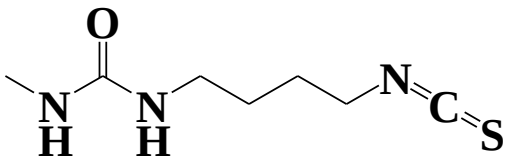
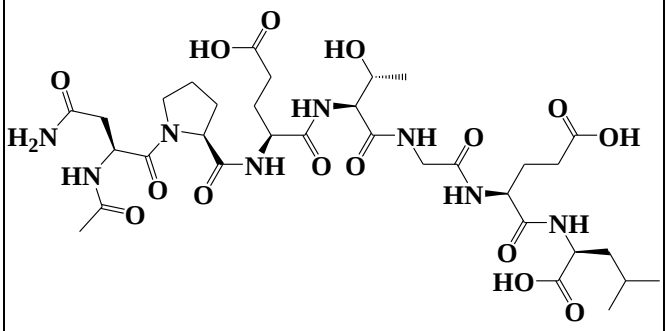
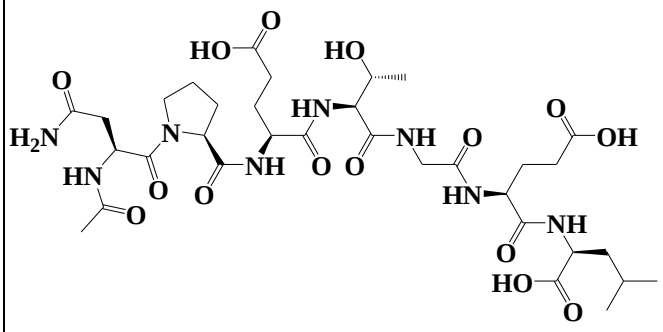
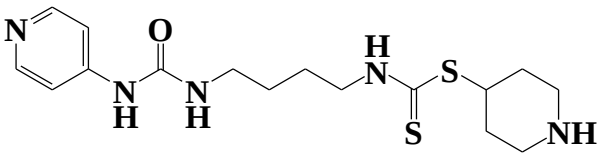
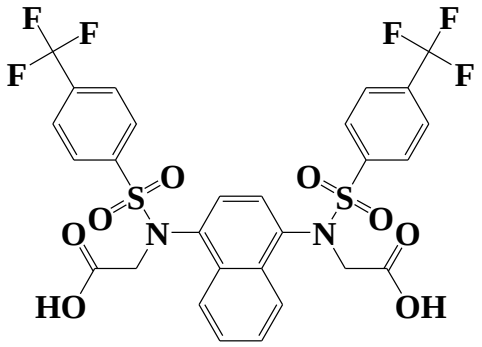
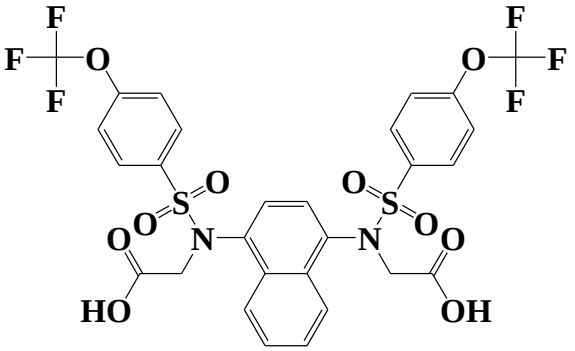
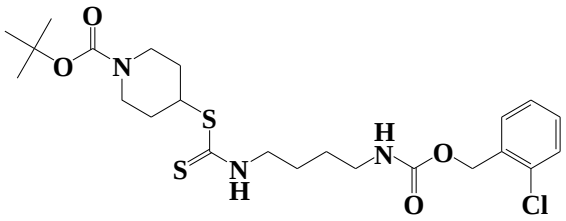
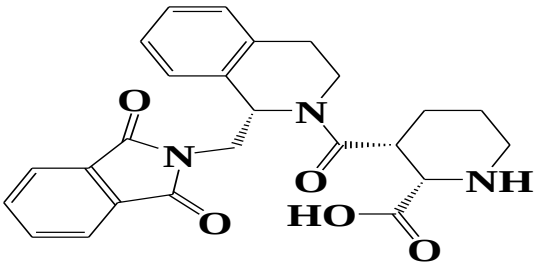
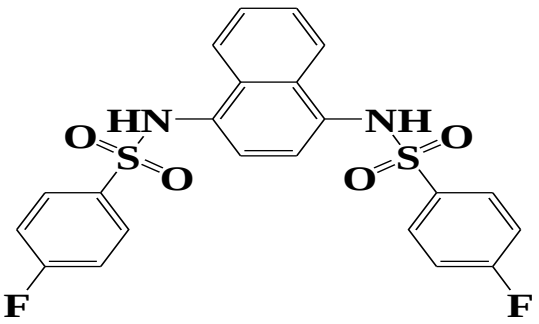
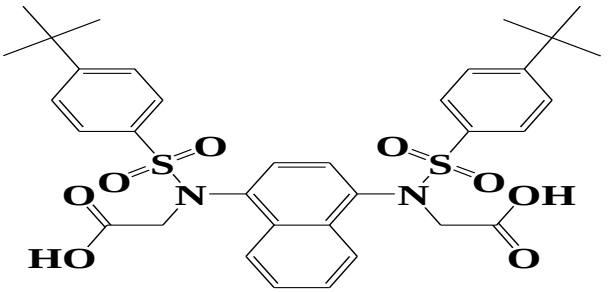
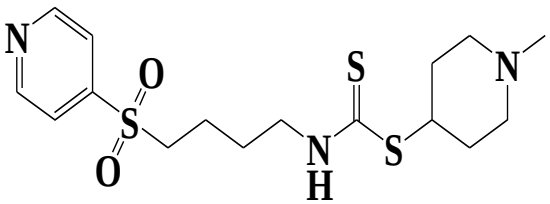
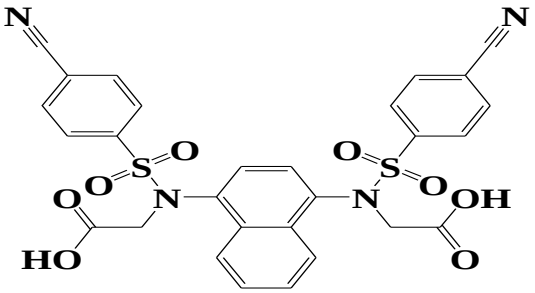
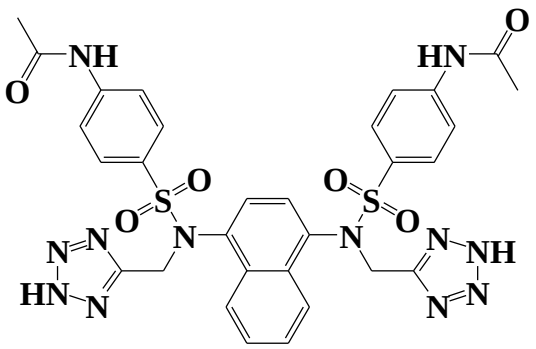
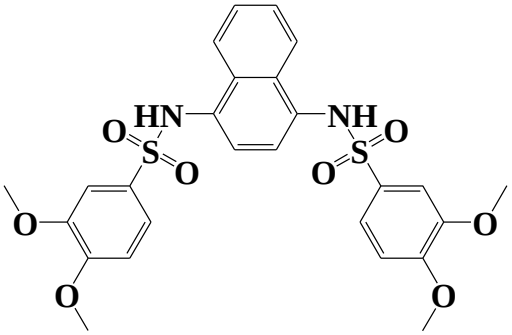
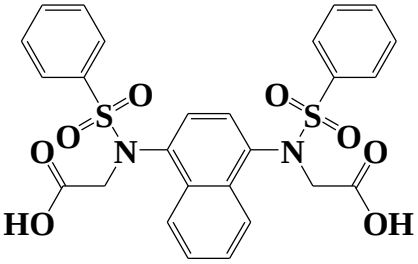
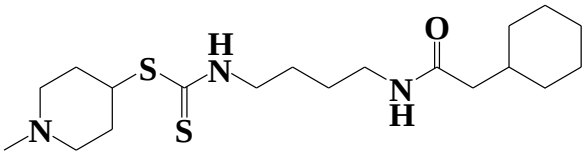
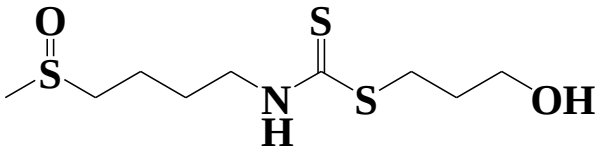
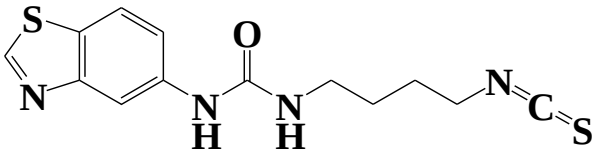
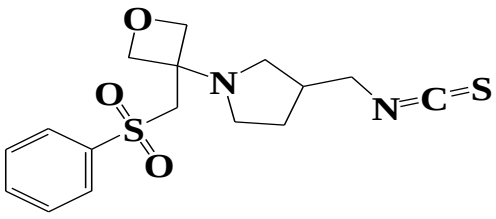
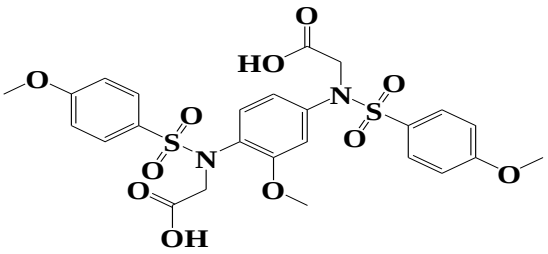


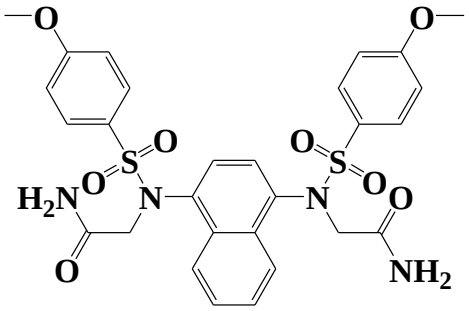
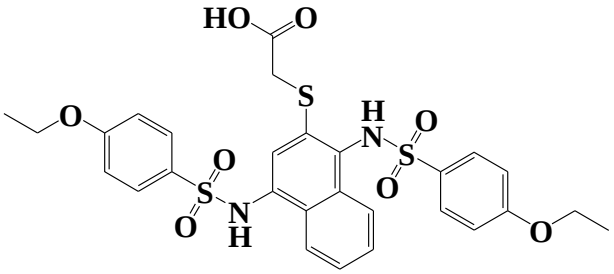
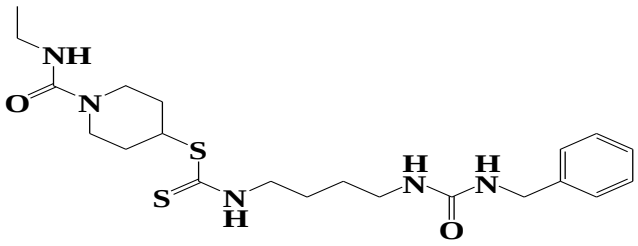
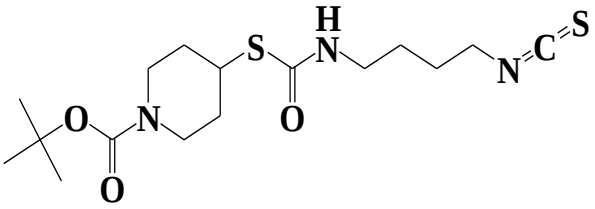
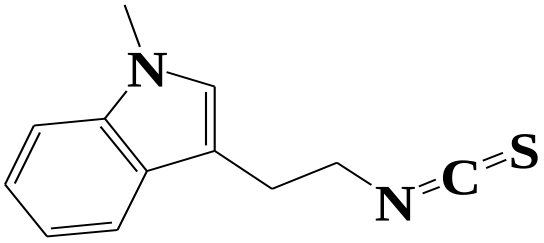
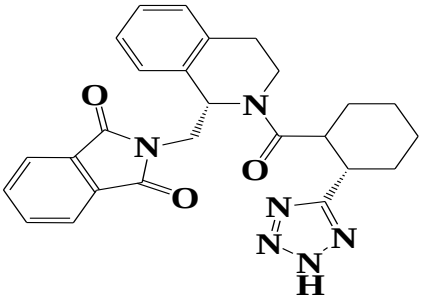
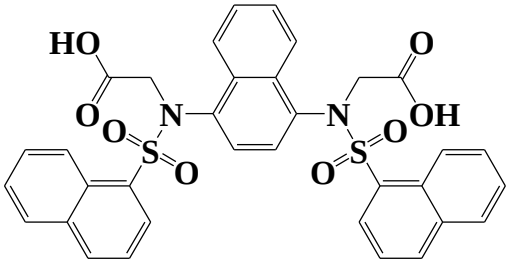
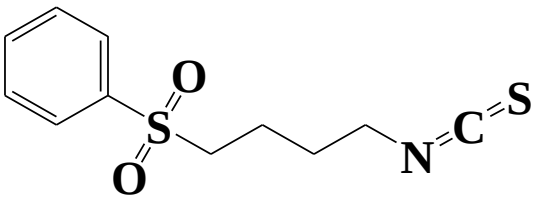
Figure 3. 2: Overall workflow of GRIND methodology. MIF computation, discretization, and encoding are the three major steps of GRIND.

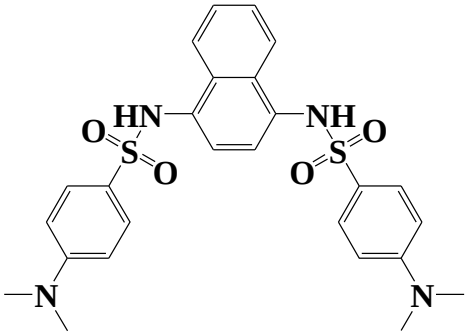
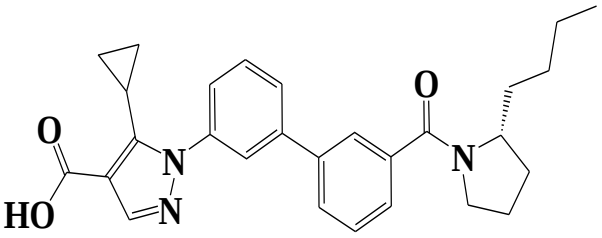
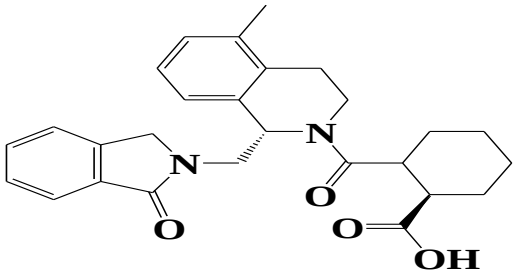
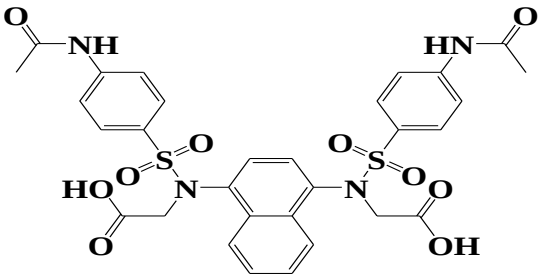
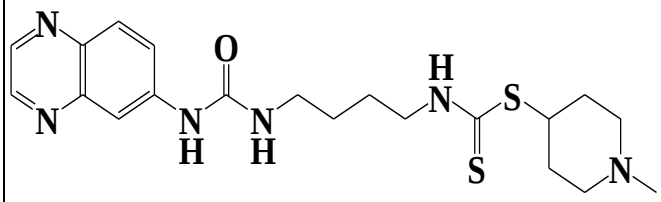
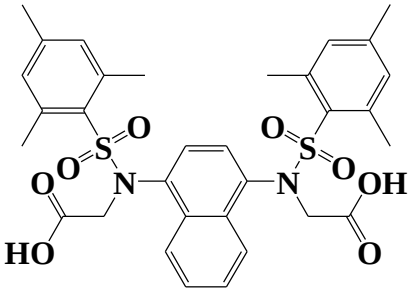
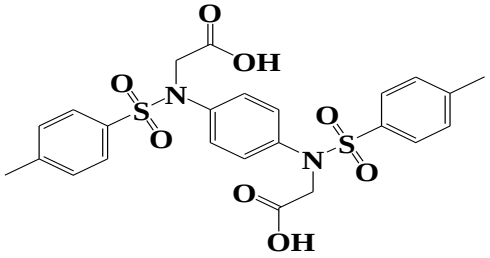
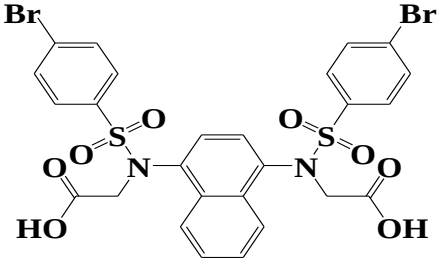
Table 1. 1 : The 2D chemical structures of 93 compounds

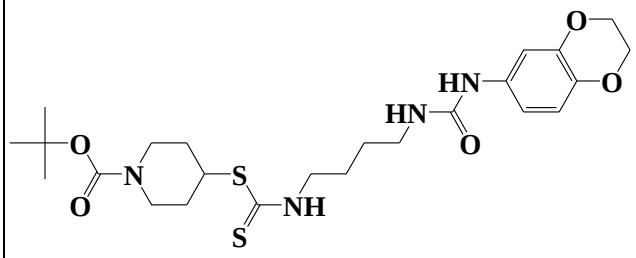
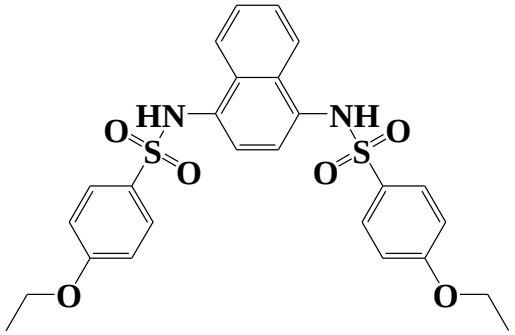
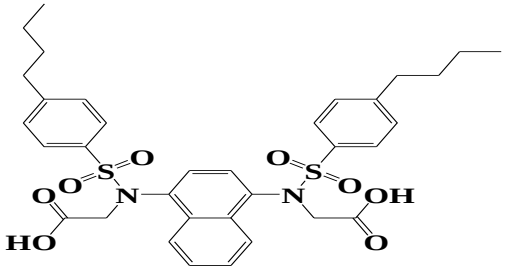
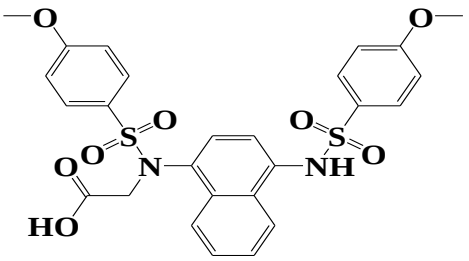
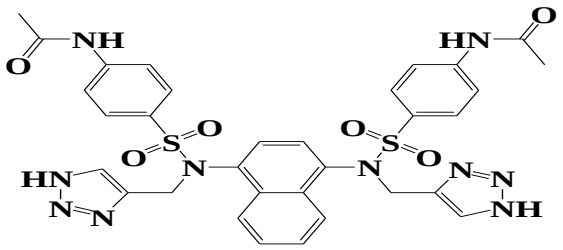
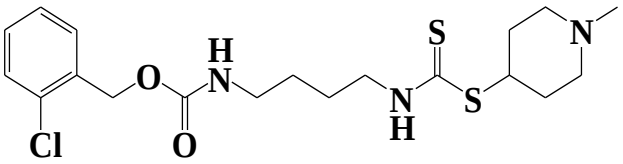
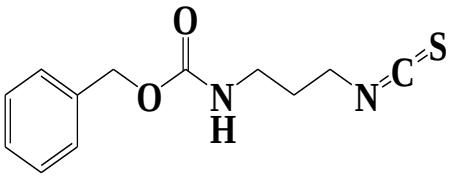
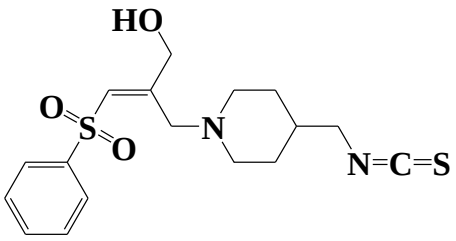
NRF2_1  **pIC50 = 2.9087 Gold_Score = 37.5629**	**NRF2_6**  **pIC50 = 2.5662 Gold_Score =94.2669**
NRF2_87  **pIC50 = -0.7913 Gold_Score = 45.6498**	**NRF2_36**  **pIC50 = 0.9666 Gold_Score = 81.1759**
NRF2_48  **pIC50 = 0.4937 Gold_Score = 77.4697**	**NRF2_78**  **pIC50 = -0.7913 Gold_Score = 70.2106**

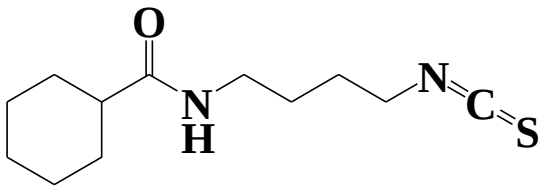
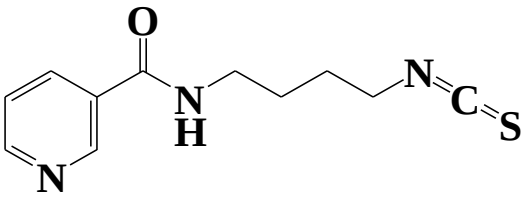
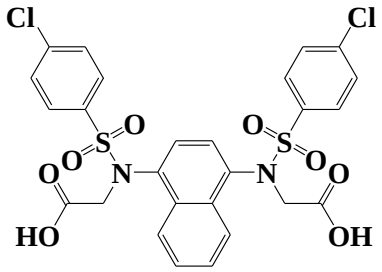
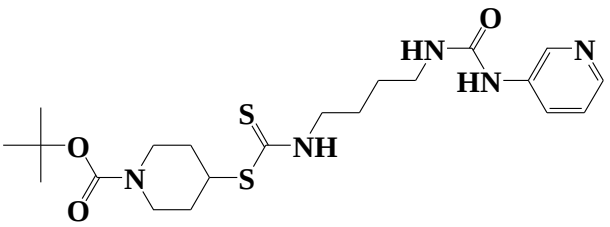
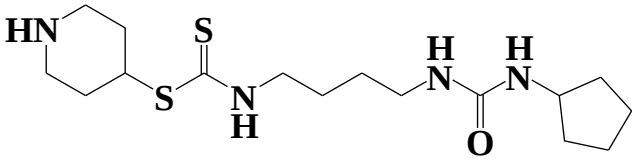
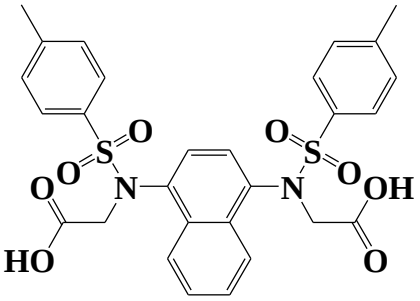
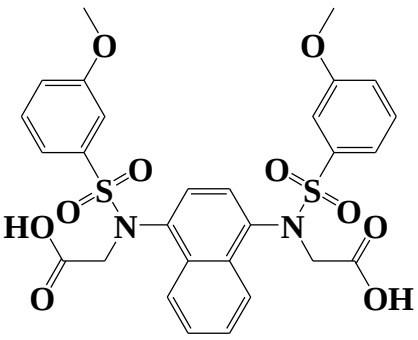
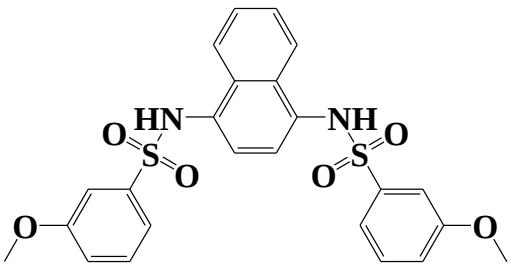
NRF2_21  **pIC50 = 1.8223 Gold_Score = 62.1183**	**NRF2_23**  **pIC50 = 1.6558 Gold_Score = 58.1538**
NRF2_47  **pIC50 = 0.5086 Gold_Score = 80.3281**	**NRF2_92**  **pIC50 = -0.9346 Gold_Score = 56.7111**
NRF2_33  **pIC50 = 1.0891 Gold_Score = 69.8134**	**NRF2_13**  **pIC50 = 2.2079 Gold_Score = 58.0764**
NRF2_93  **pIC50 = -1.0913 Gold_Score =74.1478**	**NRF2_20**  **pIC50 = 1.8368 Gold_Score =58.7179**

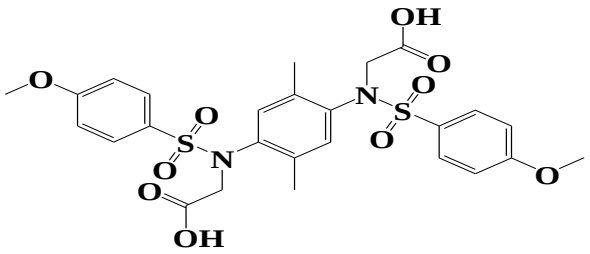
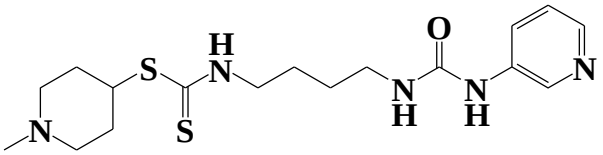
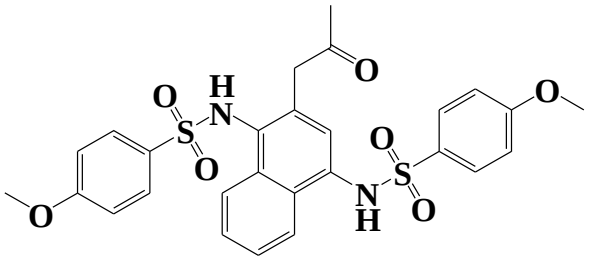
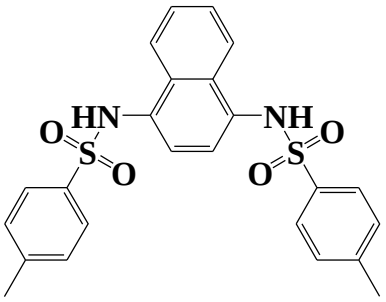
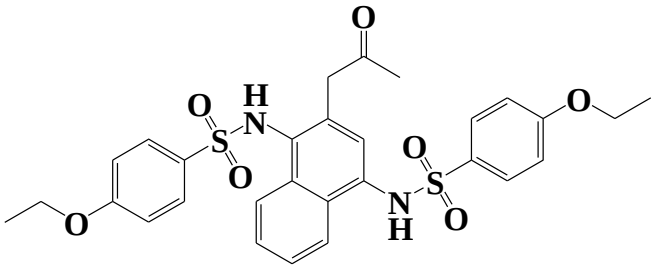
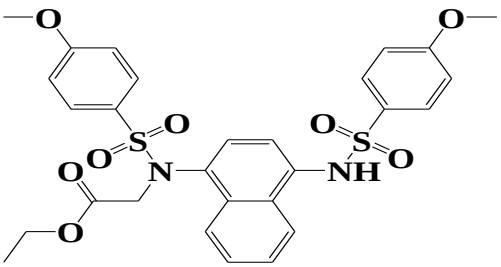
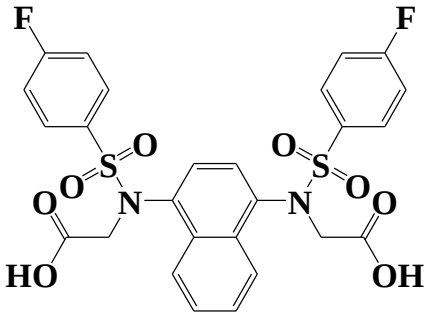
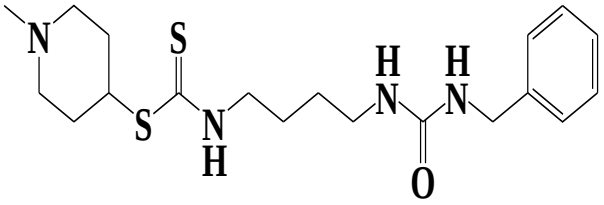
NRF2_4  **pIC50 = 2.7099 Gold_Score =65.8401**	**NRF2_40**  **pIC50 = 0.8672 Gold_Score =58.4317**
NRF2_8  **pIC50 = 2.3733 Gold_Score =49.7719**	**NRF2_66**  **pIC50 = -0.2913 Gold_Score =73.4223**
NRF2_71  **pIC50 = -0.4913 Gold_Score = 58.3767**	**NRF2_55**  **pIC50 = 0.2087 Gold_Score = 63.5079**
NRF2_85  **pIC50 = -0.7913 Gold_Score =54.4401**	**NRF2_29**  **pIC50 = 1.1290 Gold_Score =76.8331**

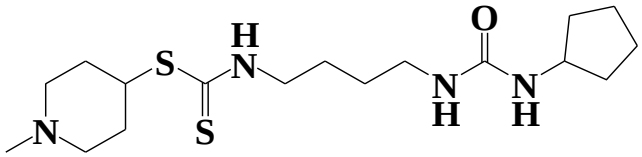
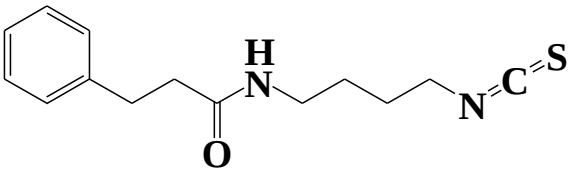
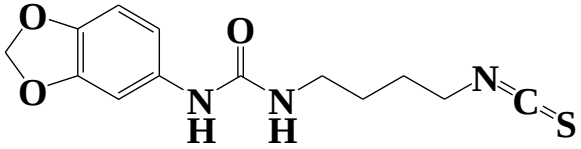
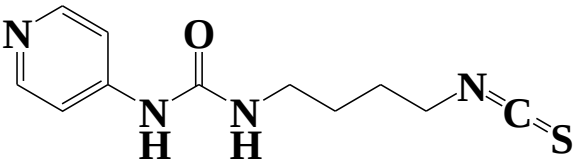
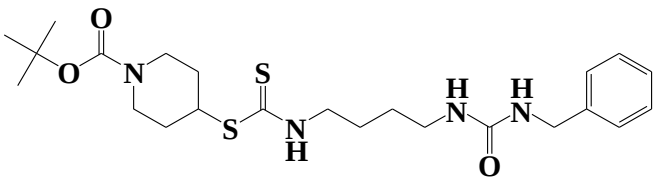
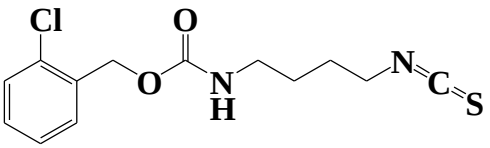
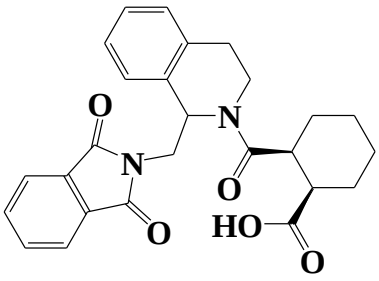
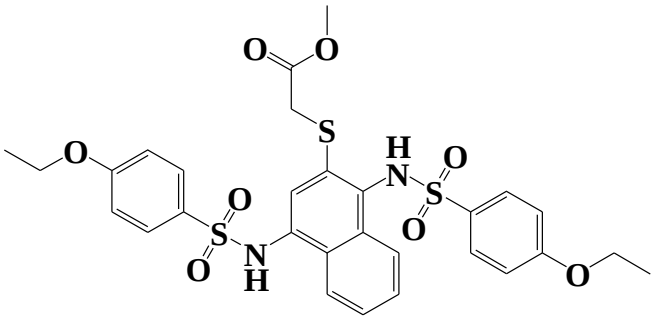
NRF2_16  **pIC50 = 2.1093 Gold_Score =74.6227**	**NRF2_25**  **pIC50 = 1.6076 Gold_Score =64.7918**
NRF2_90  **pIC50 = -0.891 Gold_Score =76.062**	**NRF2_64**  **pIC50 = -0.191 Gold_Score =66.337**
NRF2_61  **pIC50 = -0.091 Gold_Score =50.173**	**NRF2_58**  **pIC50 = 0.039 Gold_Score =60.287**
NRF2_12  **pIC50 = 2.232 Gold_Score =66.707**	**NRF2_88**  **pIC50 = -0.791 Gold_Score =53.751**

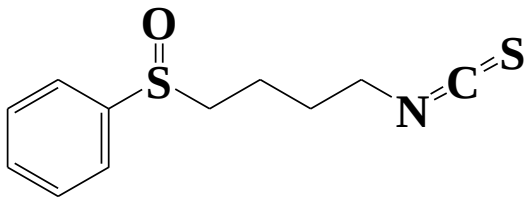
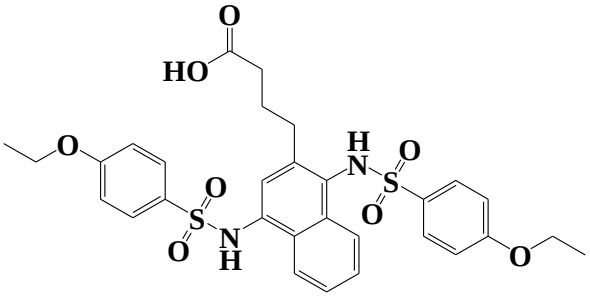
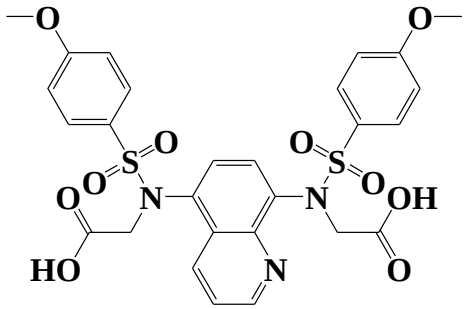
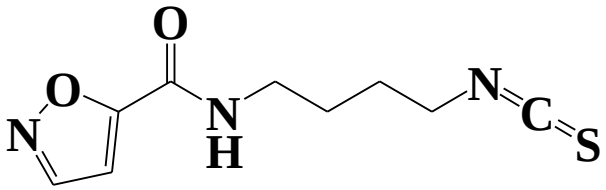
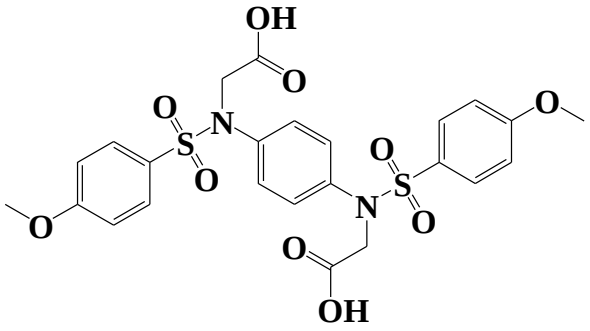
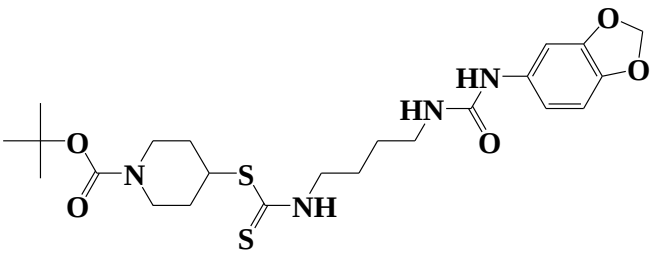
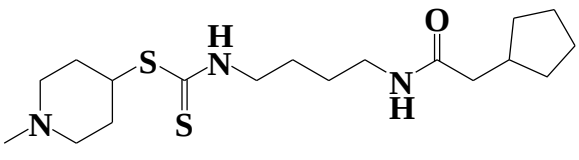
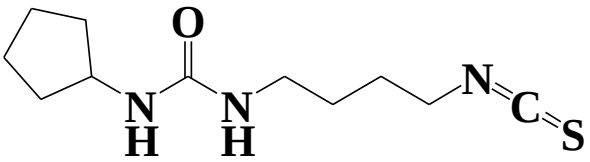
NRF2_24  **pIC50 = 1.629 Gold_Score =55.8189**	**NRF2_2**  **pIC50 = 2.908 Gold_Score =69.126**
NRF2_35  **pIC50 = 1.0336 Gold_Score =53.1461**	**NRF2_3**  **pIC50 = 2.7815 Gold_Score =63.2865**
NRF2_62  **pIC50 = -0.0913 Gold_Score =80.9053**	**NRF2_39**  **pIC50 = 0.9179 Gold_Score =53.6511**
NRF2_38  **pIC50 = 0.9237 Gold_Score =72.4054**	**NRF2_9**  **pIC50 = 2.3463 Gold_Score =67.1228**

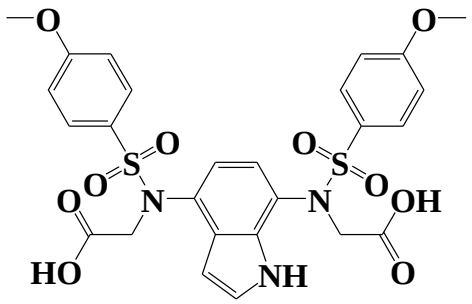
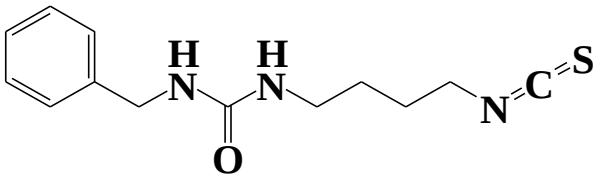
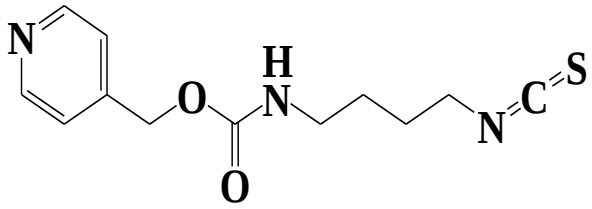
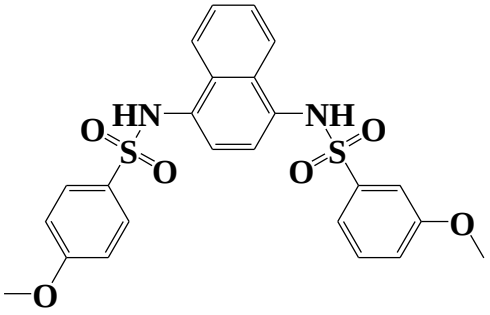
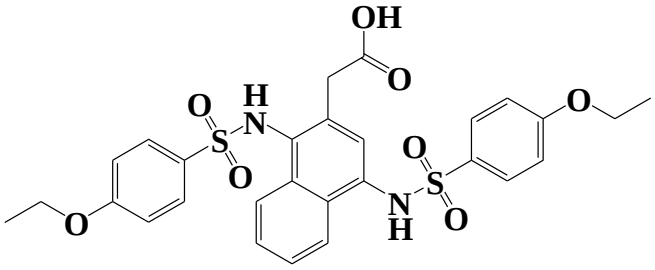
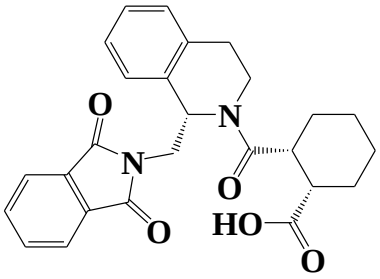
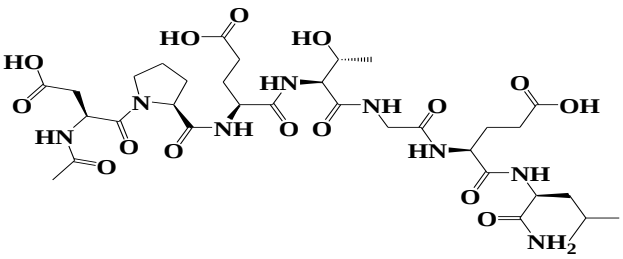
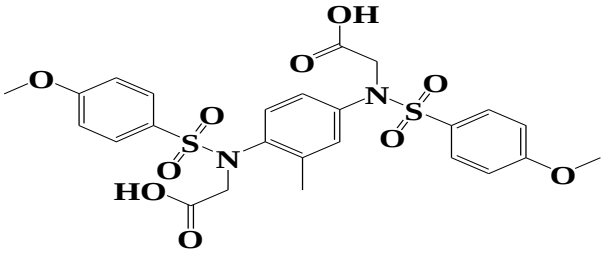
NRF2_74  **pIC50 = -0.5913 Gold_Score =81.1652**	**NRF2_49**  **pIC50 = 0.4615 Gold_Score =71.1391**
NRF2_10  **pIC50 = 2.2833 Gold_Score =63.4189**	**NRF2_15**  **pIC50 = 2.1233 Gold_Score =53.6518**
NRF2_30  **pIC50 = 1.1240 Gold_Score =62.8616**	**NRF2_52**  **pIC50 = 0.3087 Gold_Score =77.6362**
NRF2_53  **pIC50 = 0.20865 Gold_Score =57.7905**	**NRF2_70**  **pIC50 = -0.4913 Gold_Score =69.4547**

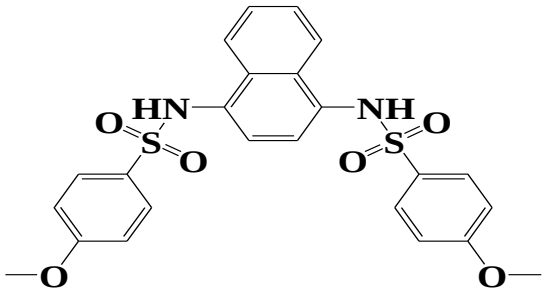
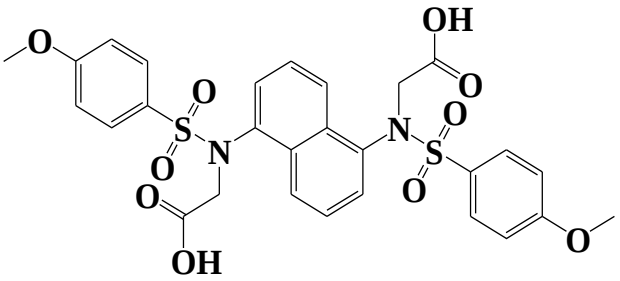
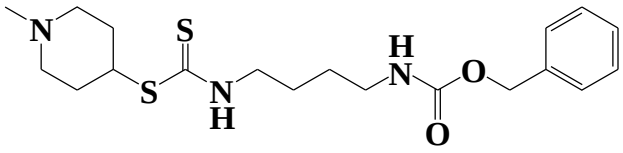
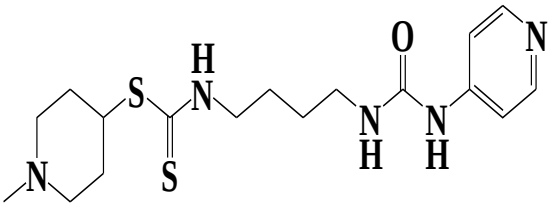
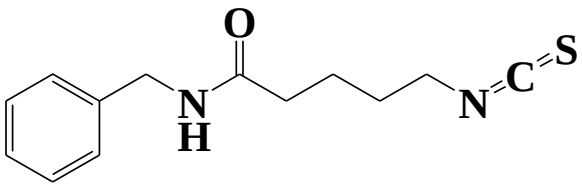
NRF2_82  **pIC50 = -0.7913 Gold_Score =55.2817**	**NRF2_79**  **pIC50 = -0.7913 Gold_Score =54.4507**
NRF2_11  **pIC50 = 2.2823 Gold_Score =47.7882**	**NRF2_80**  **pIC50 = -0.7913 Gold_Score =76.1444**
NRF2_86  **pIC50 = -0.7913 Gold_Score =65.6948**	**NRF2_5**  **pIC50 = 2.6510 Gold_Score =50.1343**
NRF2_17  **pIC50 = 2.0884 Gold_Score =51.6496**	**NRF2_31**  **pIC50 = 1.1093 Gold_Score =65.2389**

NRF2_63  **pIC50 = -0.0987 Gold_Score =47.498**	**NRF2_75**  **pIC50 = -0.6913 Gold_Score =67.8335**
NRF2_51  **pIC50 = 0.4035 Gold_Score =68.6768**	**NRF2_37**  **pIC50 = 0.9593 Gold_Score =65.6175**
NRF2_56  **pIC50 = 0.1163 Gold_Score =67.611**	**NRF2_18**  **pIC50 = 1.9792 Gold_Score =68.7372**
NRF2_14  **pIC50 =2.1738 Gold_Score =49.843**	**NRF2_68**  **pIC50 =-0.3913 Gold_Score =70.1821**

NRF2_76  **pIC50 = -0.6913 Gold_Score =66.042**	**NRF2_72**  **pIC50 = -0.4913 Gold_Score =59.1319**
NRF2_57  **pIC50 = 0.1086 Gold_Score =62.3908**	**NRF2_73**  **pIC50 = -0.5913 Gold_Score =57.8101**
NRF2_59  **pIC50 = 0.0086 Gold_Score =81.08910**	**NRF2_54**  **pIC50 = 0.2086 Gold_Score =59.3038**
NRF2_50  **pIC50 = 0.4315 Gold_Score =55.6857**	**NRF2_34**  **pIC50 = 1.0513 Gold_Score =75.2179**

NRF2_89  **pIC50 = -0.7913 Gold_Score =55.1653**	**NRF2_22**  **pIC50 = 1.7325 Gold_Score =64.6634**
NRF2_19  **pIC50 =1.9043 Gold_Score =62.5101**	**NRF2_91**  **pIC50 = -0.8913 Gold_Score =50.2337**
NRF2_42  **pIC50 = 0.7479 Gold_Score =71.3239**	**NRF2_83**  **pIC50 =-0.7913 Gold_Score =83.1111**
NRF2_84  **pIC50 = -0.7913 Gold_Score =71.7175**	**NRF2_81**  **pIC50 = -0.7913 Gold_Score =57.7145**

NRF2_41  **pIC50 = 0.7947 Gold_Score =72.0884**	**NRF2_65**  **pIC50 = -0.2913 Gold_Score =56.2922**
NRF2_77  **pIC50 = -0.6913 Gold_Score =58.2991**	**NRF2_28**  **pIC50 = 1.2959 Gold_Score =64.1942**
NRF2_27  **pIC50 =1.4035 Gold_Score =65.4175**	**NRF2_46**  **pIC50 = 0.5469 Gold_Score =52.6356**
NRF2_32  **pIC50 = 1.1065611 Gold_Score =81.513298**	**NRF2_43**  **pIC50 = 0.68697536 Gold_Score =65.615997**

NRF2_44  **pIC50 =0.6534 Gold_Score =66.6813**	**NRF2_45**  **pIC50 =0.6391 Gold_Score =57.4852**
NRF2_60  **pIC50 = -0.0913 Gold_Score =70.7623**	**NRF2_69**  **pIC50 = -0.3913 Gold_Score =65.5001**
NRF2_67  **pIC50 = -0.3913 Gold_Score =56.3595**	

Chapter 4
Results & Discussion

Chapter 4

Results and Discussion

4.1 Molecular Docking

4.1.1 Protein Structure Selection

The dataset (93 compounds shown in table 1.1) and crystal structure of NRF2/KEAP1 (PDB_ID 6TYP) with resolution 2.50 Å were initially energy with MM94x and Amber 10 force field protocol in MOE. Then this energy minimized structures of both (protein and ligands) were used in Molecular Docking to assess their inhibitory process. Although, the reported x-ray structure of protein 6TYP is of 293 amino acids length. Therefore, 93 ligand molecules were docked with protein to identify the binding pattern of KEAP1 protein.

4.1.2 Selection of Binding Cavity

The binding site of protein was chosen using Gold suite 5.3.0 and select a centroid point where the coordinates were x = -17.1700, y = 5.5030, z = 21.7770 respectively with a radius of 25Å. The selected bind cavity contained all the amino acid residues which are reported in literature till date ARG483, ARG415, TYR334, SER602, SER555, GLN530, and TYR525 [39] and for each ligand 100 genetic runs were executed, and their flexibility was maintained. The actual binding cavity in the receptor site was observed and here shown in Figure:4.1

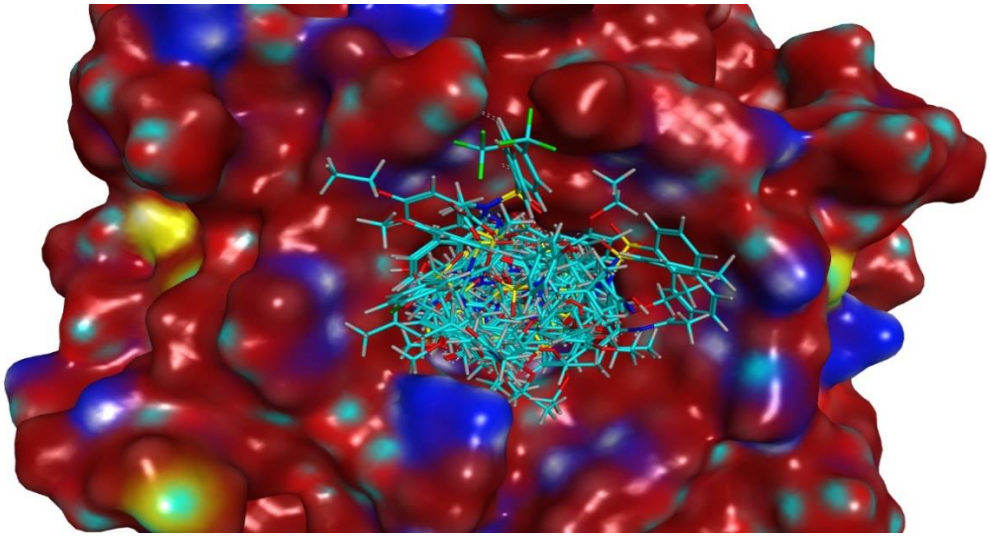
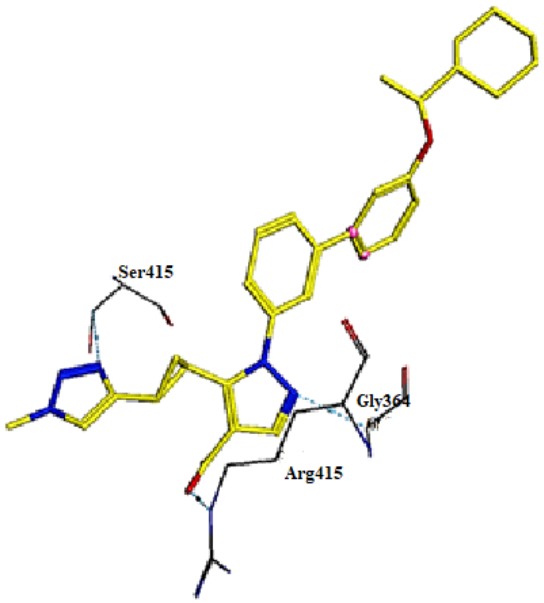
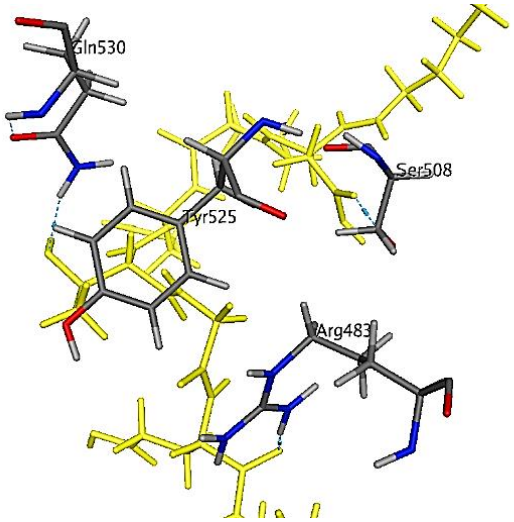


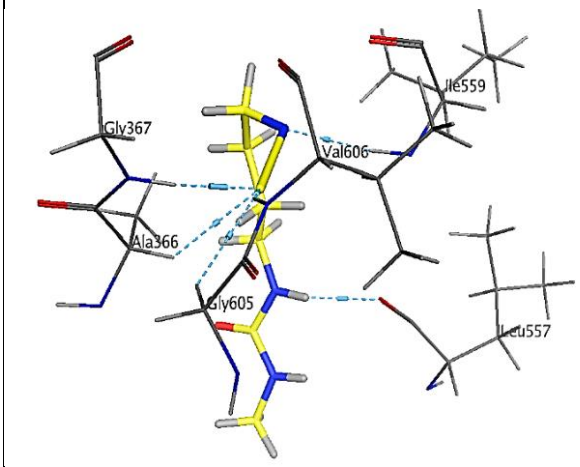
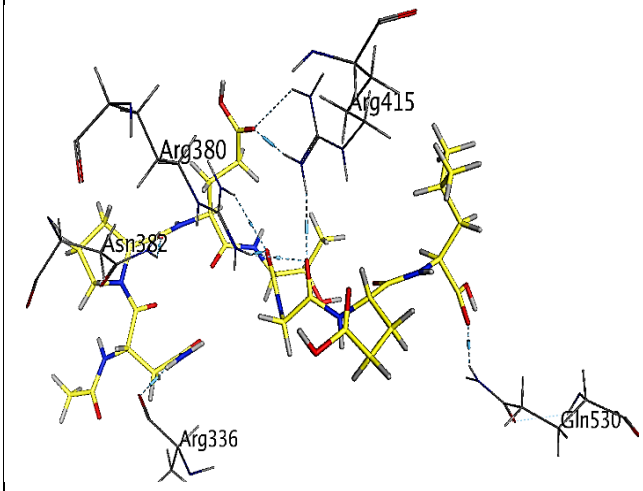
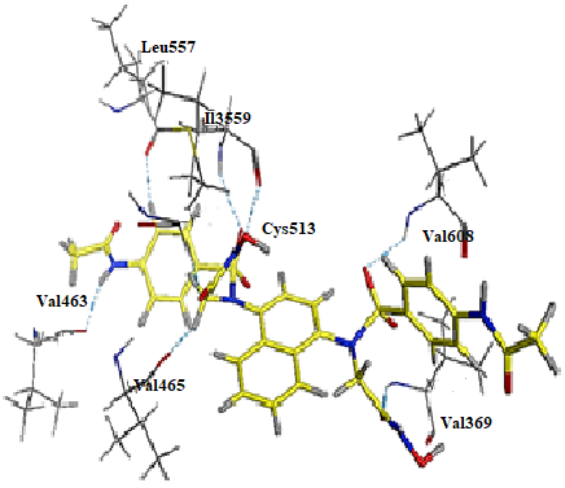
Figure 4. 1: All docked ligands were representing at one region which depicts the actual binding pocket in receptor site

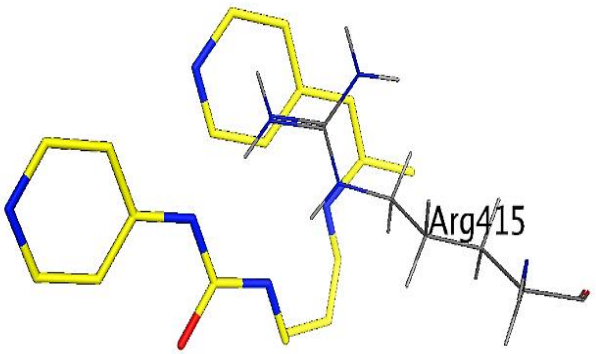
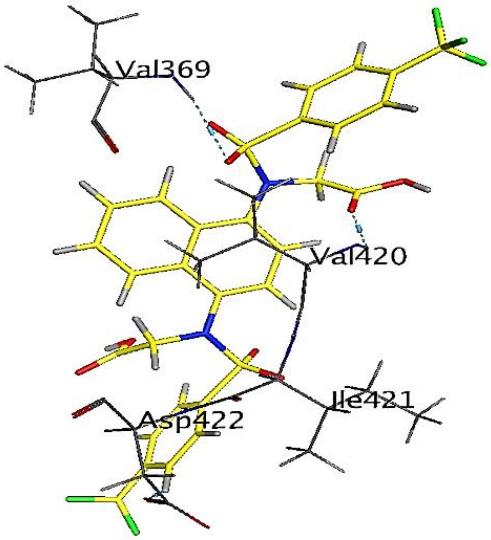
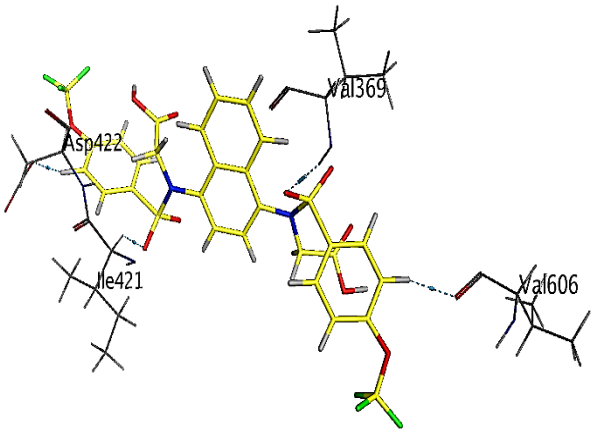
4.2 Protein-Ligand interactions

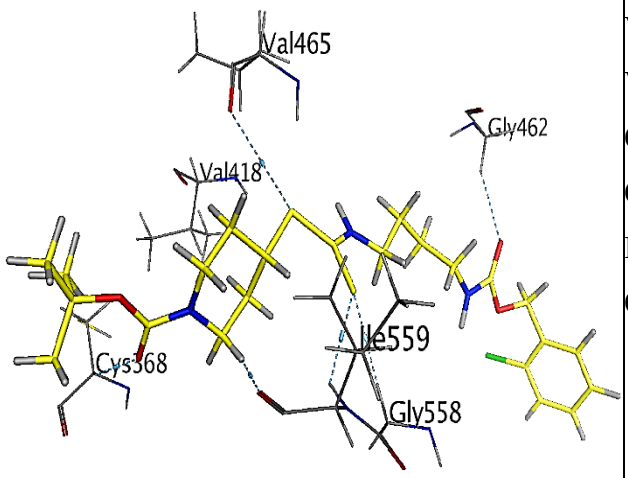
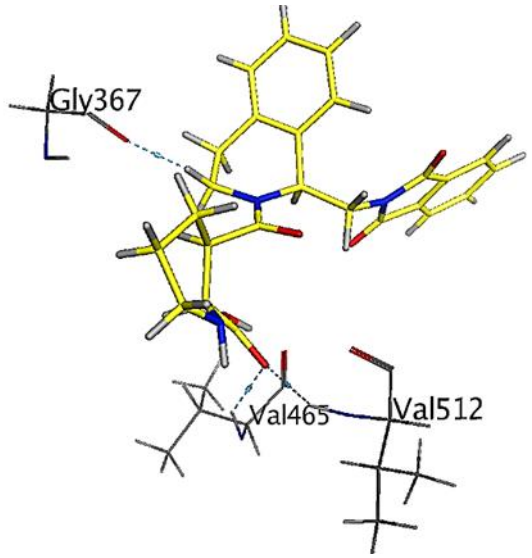
The best-scored pose of each ligand was selected for analysis of the interaction between protein and ligands. The selected pose of each highly active and least active ligand was chosen further for interactions study between protein and ligand. Each interaction of the protein with ligands was studied, and interactions that were perceived include π- H and hydrogen bond interactions, involving amino acids residues ARG 336, ASN 382, ARG 415, ARG 380, GLN 530, GLY 364, SER 555, SER 508, ARG 483, TYR 525 ARG 336 GLY 367, ILE 421, VAL 608, ILE 421 shown in Table 4.1.

Table 1. 2: Protein Ligand interaction of KEAP1 with their Residues and interaction.

Compound ID	Interaction pattern	Residues No	Interaction
NRF2_1		GLY 364 ARG 415 SER 555	H- Bond H- Bond H- Bond
NRF2_6		SER 508 ARG 483 GLN 530 TYR 525	H-Bond H-Bond H-Bond Π - H

NRF2_87		LEU 557 ILE 559 ALA 366 GLY 367 GLY 605 VAL 606	H-Donor H-Acceptor H-Acceptor H-Acceptor H-Acceptor H-Acceptor
NRF2_36		ARG 336 ASN 382 ASN 382 ARG 415 ARG 380 ARG 415 GLN 530	H- Donor H- Donor H-Acceptor H-Acceptor H-Acceptor H-Acceptor H-Acceptor
NRF2_48		VAL 463 VAL 465 ILE 559 LEU 557 ILE 559 CYS 513 VAL 369 VAL 608	H-Donor H-Donor H-Donor H-Donor H-Acceptor H-Acceptor H-Acceptor H-Acceptor

NRF2_78		ARG 415	H-Acceptor
NRF2_21		ASP 422 VAL 420 ILE 421 VAL 369	H-Donor H-Acceptor H-Acceptor H-Acceptor
NRF2_23		VAL 606 ASP 422 VAL 369 ILE 421	H-Donor H-Donor H-Acceptor H-Acceptor

NRF2_47		ILE 559 VAL 418 VAL 465 CYS 368 GLY 558 ILE 559 GLY 462	H-Donor H-Donor H-Donor H-Acceptor H-Acceptor H-Acceptor H-Acceptor
NRF2_92		VAL 465 GLY 367 VAL 465 VAL 512	H-Donor H-Donor H-Acceptor H-Acceptor

4.3 ChemScore and pIC_{50} Values Correlation:

Molecular docking of the KEAP1 inhibitors was performed through generating a maximum of 100 poses per ligands where each posed was assigned ChemScore through the GOLD suite. The correlation R^2 between biological activity (pIC_{50} value) and ChemScore of the active compound were estimated at -0.09 which shows not any correlation between biological activity and ChemScore of KEAP1 as presented in figure 4.2. As the ChemScore scoring function relies mainly on the binding affinity so ChemScore should be correlated with biological activity against targets. So, in these datasets, no correlation found which indicated that there were other reasons involved in the violation of the behavior of highly active compounds. Therefore, we explored

different physiological factors and preformed Grind because it is suitable for diverse dataset.

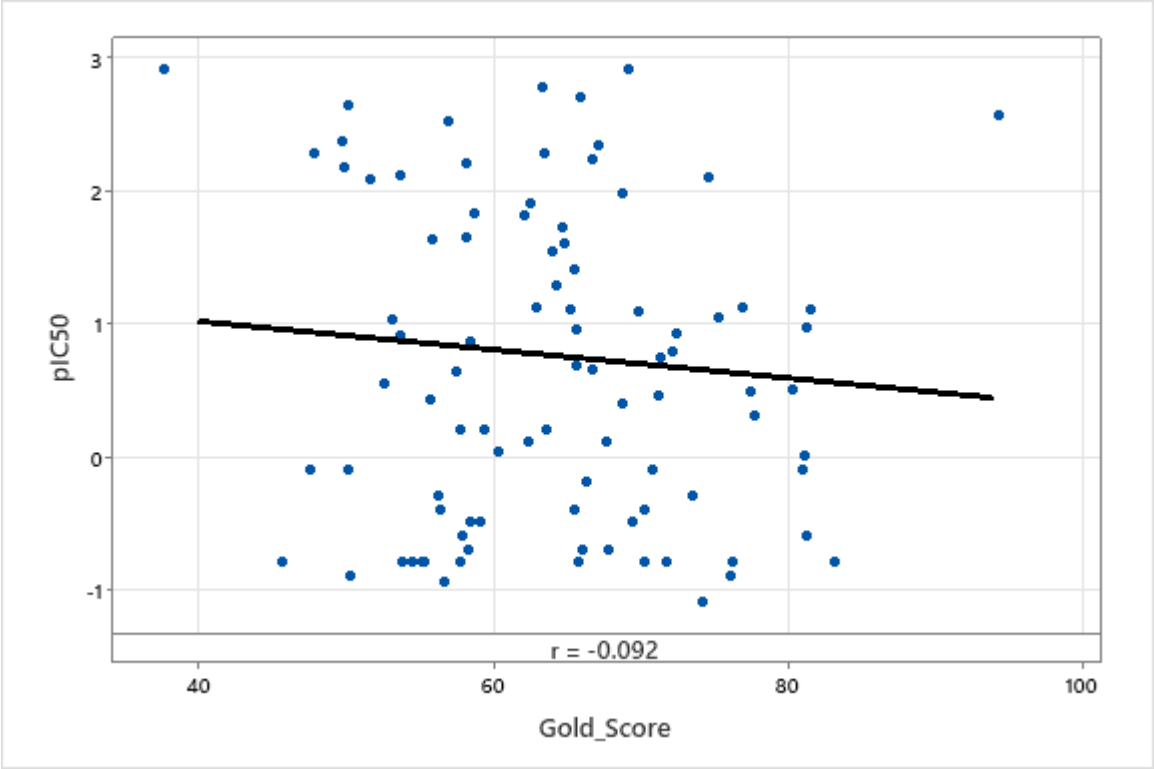


Figure 4. 2 : The biological activity (PIC_{50}) plotted on y-axis and binding affinity (gold score) on x-axis showed the correlation of R2 -0.09 for overall data

4.4 PLIF

Protein-ligands interaction was analyzed after the final selected pose using a protein ligand interaction finger printing scheme [83]. PLIF explained the interaction type including the SAC (Surface contact), HBD (hydrogen bond donor Sidechain), HBA (Hydrogen Bond

Acceptor Sidechain), HBD (Hydrogen Bond donor Backbone) and HBA (Hydrogen Bond acceptor Backbone). MOE used to generate PLIF of the selected poses. Protein-ligands interaction for the complex dataset was explained efficiently by using PLIF. The PLIF was computed to see the most common amino acid residues interactions of the docked inhibitor data with receptor protein.

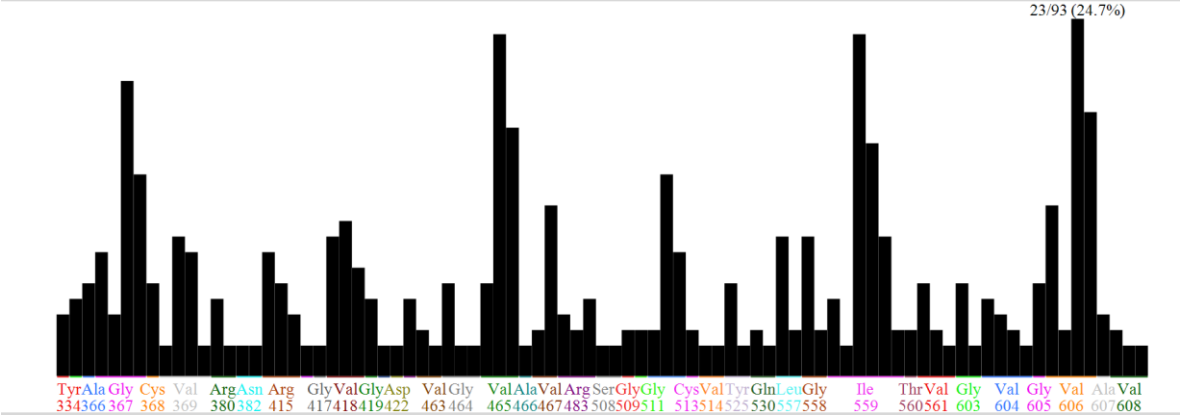


Figure 4. 3: PLIF analysis that are elaborate in the interaction presented in the bottom and percentage of the interaction of residues presented in the top.

According to PLIF, the residue Val606 forms most of the interactions with the ligands while Val465 and Gly367 and Ile 559 forms 40% of interactions in comparison, as presented in the Figure 4.2. Furthermore, Val606, Ile 559, Val465 and Gly367 forms interaction like backbone donor and acceptor interactions. These four are the most common residues are involved in the interaction with the ligands Figure: 4.2.

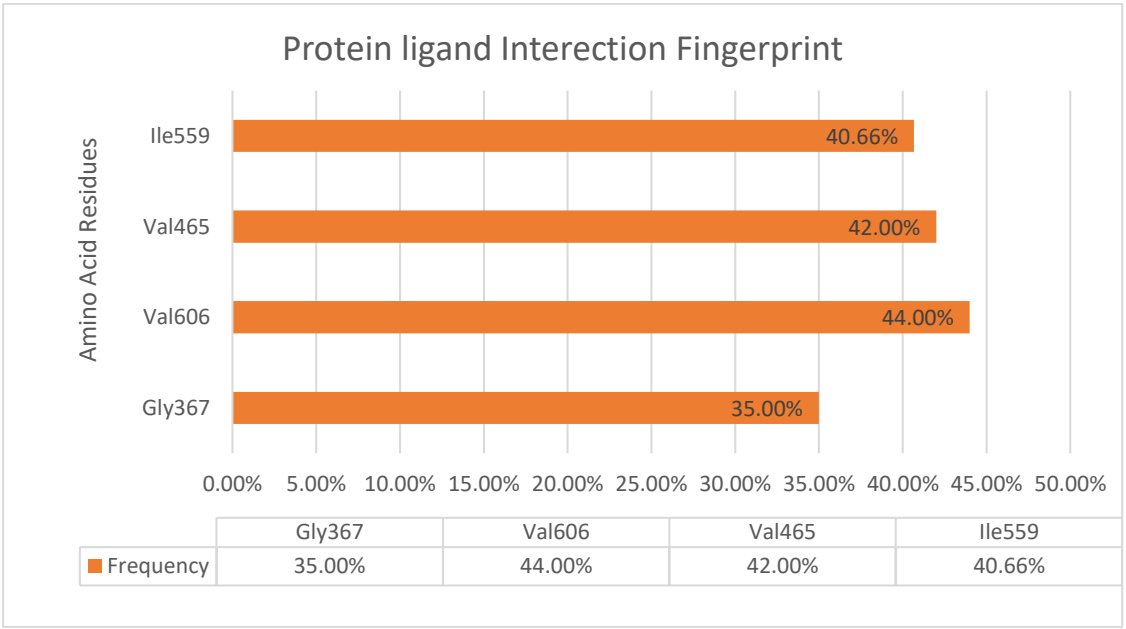


Figure 4. 4: PLIF analysis of selective datasets of protein residues that are involved in the interaction

4.5 GRIND

Based on the common scaffold analyzed from the docked poses, confirmations were obtained, accordingly. Analysis of GRID independent molecular descriptor of the final binding conformations of all chemical data set has been carried out by using software Pentacle v 1.06 [84]. The molecules and their biological activities (pIC_{50}) were loaded in

the software. As GRIND is largely dependent on conformations, Molecular descriptors were computed and discretized using the AMANDA algorithm [85] implemented in Pentacle software. Finally, encoding was done using the CLACC algorithm and the graph plotted between the Node – Node distances among the probes and their respective energy product is referred to as correlogram. Therefore, the correlograms were analyzed consistent peaks defining the feature represented by probes in the virtual receptor site. Furthermore, the predictive ability of PLS models were tested from q^2. The statistically significant models with q2 > 0.50 and SDEP <1.

PLS (partial least square) model was built using 5 latent variables. Initially, a complete set of variables of KEAP1/NRF2 was used to build the model and the result was not satisfying as Q2 value 0.46 and R2 value 0.59. Thus, Fractional Factorial Design (FFD) was applied to reduce the number of inconsistent variables in each model. After the first cycle of FFD, the statistical parameters of each model were slightly improved.

However final model with good statistical parameters at the second latent variable (LV2) showing $Q^2 = 0.58$, $R^2 = 0.68$ and standard error of prediction SDEP = 0.78 (Table 1.2) was obtained after three cycles of FFD.

Table 1. 3: Statistics of a preliminary model of GRIND using a complete set of active variables and after applying the FFD variable selection algorithm.

	R2	Q2	FFD variables
Complete set of variable	0.59	0.46	426
FFD1	0.62	0.52	696
FFD2	0.66	0.55	637
FFD3	0.68	0.58	599

Partial least square (PLS) co-efficient plot of GRIND model was used to analyses and identify features that show a significant role in the interaction of inhibitors with KEAP1.

This correlogram (Figure 4.3) includes auto (Dry-Dry, Tip-Tip, O-O, and N1-N1) and cross variables (Dry-O, Dry-Tip, Dry-N1, Tip-O, Tip-N1, O-N1) variables linked positively and negatively with NRF2/KEAP1 inhibitory potency (pIC50). The bar plot demonstrate that N1-N1 is an important feature that show positive contribution in the inhibitory activity of highly active ligands and O-O features also explains the negative contribution in the inhibitory activity of the least active ligand. For explaining the principal component

analysis (PCA) and partial least square analysis (PLS), the second latent variable (LV) was nominated to elucidate several chemical features correlation with biological activity and described structural variance among datasets. Here, N1-N1 auto and cross correlograms was made possible to analyses a good PLS.

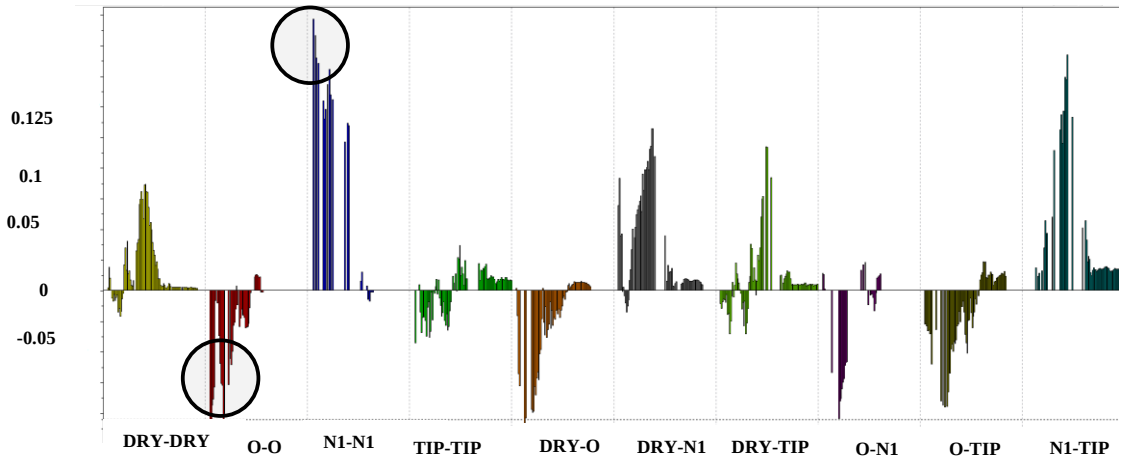


Figure 4. 5: PLS co-efficient correlogram profile of positively and negatively correlated variables of KEAP1

N1-N1 was one of the first prominent and consistent peak, demonstrating that two amide nitrogen features with a mutual distance of 2.80 – 3.20Å were present in the most promising KEAP1 inhibitors. So, Val 209 residue is complementary in the actual receptor site. The complementarity of virtual-actual receptors sites was estimated for NRF2_1 from the dataset (from table 1.2) collected which was one of the most prominent inhibitors of KEAP1 shown in fig 4.5

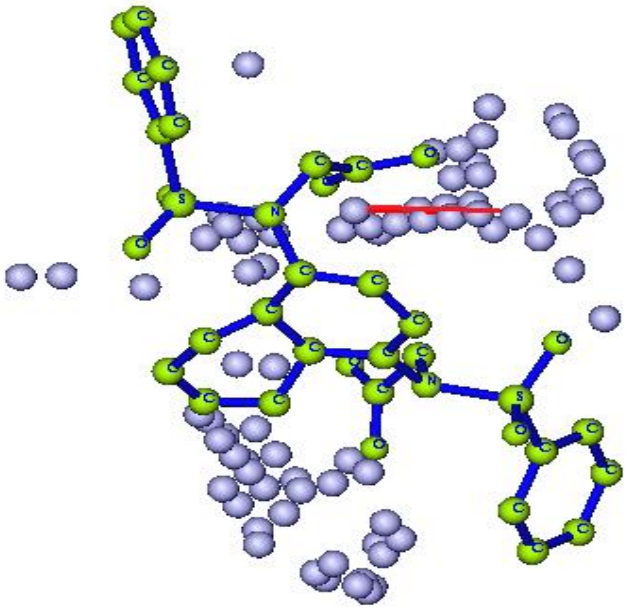


Figure 4. 6: Interaction shows the most prominent peak in the pentacle, N1-N1. In sphere, shapes coloured blue shows the predicted virtual space of interaction

O-O was the second important peak demonstrating the features of two hydrogen bond donor groups present in the ligand dataset at a shorter distance of 2 – 2.80Å presents at the least active compounds and shows a negative effect on the inhibitory potency of a compound against NRF2/KEAP1 Figure 4.6.

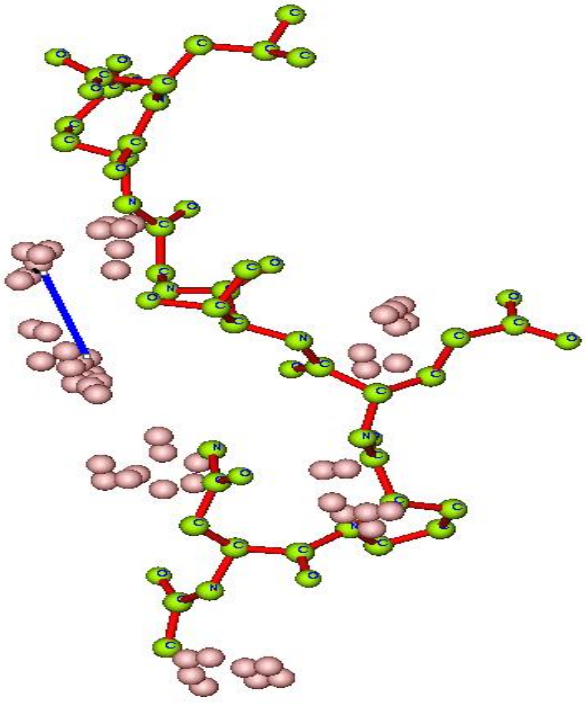


Figure 4. 7: Interaction shows peak in the pentacle, O-O. In sphere, shapes coloured blue shows the predicted virtual space of interaction

The predicted feature design can help in improving the properties of the drug. Also, these predictions can aid in the design of new compounds/drug-like entities. It can also enhance the properties of the drug. The recent research added value by predicting the activity of drugs based on several interactions. Same interactions were found around all the ligands and the GRIND analysis too provided features that were consistent with the docking output. The research can likely aid in the development of potential drug targets that can be elective against KEAP1 target protein which degrades the NRF2 and causes cardiovascular diseases.

Chapter 5
Conclusion

Chapter 5

Conclusion

Cardiovascular diseases are globally very common and there is not an effective treatment that can completely cure the disease, although a better lifestyle decreases the risk of CVD diseases. NRF2 plays a very important role in protecting the heart. The current study was designed to modulate the normal physiological function and feature extraction of target protein KEAP1/NRF2 using molecular docking and molecular dynamics studies. In the structure-based study, molecular docking simulations were performed for the prediction of NRF2/KEAP1 binding hypothesis. Most of the ligands bind in the domain have following residues that were identified ARG 336, ASN 382, ARG 415, ARG 380, GLN 530, GLY 364, SER 555, SER 508, ARG 483, TYR 525 ARG 336. CHEMBL4160447 named as NRF2_1 shows maximum interaction with binding pocket residues after docking. The molecular docking of the dataset of inhibitors against the KEAP1 helped in attaining the best conformations of the KEAP1 that were able to differentiate the highly active compounds/inhibitors with the least active compounds/inhibitors. Also, it predicted the best interaction between the ligands dataset and the protein structure is mediated by **Val 606** residue in the protein that facilitates maximum interactions with the ligand dataset.

The GRIND model for the KEAP1 was constructed against the dataset of 91 inhibitors to extract the features that had a positive and negative impact on the activity of the inhibitors. One feature represented by correlogram NI- N1 positively impacted the activity. This feature is a two amide nitrogens feature at a distance of **2.80 – 3.20** Å in the virtual receptor site. Val 209 residue is complementary in the actual receptor site. O-O feature represented by correlogram is two hydrogen bond acceptors at a distance **2 – 2.40** Å showing the negative impact on the activity.

The GRIND model, presented in this research, can be important since it can help in the development of new drugs/compounds based on these features. Therefore, present research could assist in the in-silico analysis of the NRF2/KEAP1 pathway and the design of new drugs that can treat the disease associated with NRF2.

Chapter 6
References

References:

[1] Wu, Y., et al., Recent Advances in the Development of Theranostic Nanoparticles for Cardiovascular Diseases. Nanotheranostics, 2021. 5(4): p. 49.

[2] Cramm, J.M. and A.P. Nieboer, Self-management abilities, physical health and depressive symptoms among patients with cardiovascular diseases, chronic obstructive pulmonary disease, and diabetes. Patient education and counseling, 2012. 87(3): p. 411-415.

[3] Stewart, J., et al., Primary prevention of cardiovascular disease: Updated review of contemporary guidance and literature. JRSM Cardiovascular Disease, 2020. 9: p. 2048004020949326.

[4] Mensah, G.A., G.A. Roth, and V. Fuster, The global burden of cardiovascular diseases and risk factors: 2020 and beyond. 2019, American College of Cardiology Foundation Washington, DC.

[5] Moore, Kathryn J., Frederick J. Sheedy, and Edward A. Fisher. "Macrophages in atherosclerosis: a dynamic balance." Nature Reviews Immunology 13.10 (2013): 709-721.

[6] S. Satta, A. M. Mahmoud, F. L. Wilkinson, M. Yvonne Alexander, and S. J. White, “The Role of Nrf2 in Cardiovascular Function and Disease,” Oxidative Medicine and Cellular Longevity, vol. 2017, p. e9237263, Sep. 2017, doi: 10.1155/2017/9237263.

[7] B. Chen, Y. Lu, Y. Chen, and J. Cheng, “The role of Nrf2 in oxidative stress-induced endothelial injuries,” J Endocrinol, vol. 225, no. 3, pp. R83-99, Jun. 2015, doi: 10.1530/joe-14-0662.

[8] D. Tousoulis et al., “Diabetes Mellitus-Associated Vascular Impairment,” Journal of the American College of Cardiology, vol. 62, no. 8, pp. 667–676, Aug. 2013, doi: 10.1016/j.jacc.2013.03.089.

[9] Hernandez, L. M., & Blazer, D. G. (2006). Genetics and Health. In Genes, behavior, and the social environment: Moving beyond the nature/nurture debate. National Academies Press (US).

[10] https://www.thenews.com.pk/print/374804-46-people-dying-of-heart-diseases-in-pakistan-every-hour

[11] https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[12] K. Cervantes Gracia, D. Llanas-Cornejo, and H. Husi, “CVD and Oxidative Stress,” J. Clin. Med., vol. 6, no. 2, Feb. 2017, doi: 10.3390/jcm6020022.

[13] “NRF2 in Cardiovascular Diseases: a Ray of Hope! | SpringerLink.”

[14] https://link.springer.com/article/10.1007%2Fs12265-020-10083-8

(accessed Sep. 18, 2021).

[15] T. Senoner and W. Dichtl, “Oxidative Stress in Cardiovascular Diseases: Still a Therapeutic Target?,” Nutrients, vol. 11, no. 9, p. 2090, Sep. 2019, doi: 10.3390/nu11092090.

[16] “Molecules | Free Full-Text | Natural Products to Counteract the Epidemic of Cardiovascular and Metabolic Disorders. |HTML.” https://www.mdpi.com/14203049/21/6/807/htm?utm_content=buffer404a2&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer (accessed Sep. 18, 2021).

[17] S. Di Meo, T. T. Reed, P. Venditti, and V. M. Victor, “Role of ROS and RNS Sources in Physiological and Pathological Conditions,” Oxid. Med. Cell. Longev., vol. 2016, p. 1245049, 2016, doi: 10.1155/2016/1245049.

[18] “High glucose level and free fatty acid stimulate reactive oxygen species production through protein kinase C--dependent activation of NAD(P)H oxidase in cultured vascular cells. | Diabetes.” https://diabetes.diabetesjournals.org/content/49/11/1939.short (accessed Sep. 18, 2021).

[19] “NRF2 plays a protective role in diabetic retinopathy in mice | SpringerLink.” https://link.springer.com/article/10.1007/s00125-013-3093-8 (accessed Sep. 18, 2021).

[20] H.-Y. Cho and S. R. Kleeberger, “Nrf2 protects against airway disorders,” Toxicol. Appl. Pharmacol., vol. 244, no. 1, pp. 43–56, Apr. 2010, doi: 10.1016/j.taap.2009.07.024.

[21] “Targeting the Nrf2 pathway against cardiovascular disease: Expert Opinion on Therapeutic Targets: Vol 13, No 7.” https://www.tandfonline.com/doi/abs/10.1517/14728220903025762 (accessed Sep. 18, 2021).

[22] “Cardiovascular diseases (CVDs).” https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed Sep. 18, 2021).

[23] “NADPH Oxidase and the Cardiovascular Toxicity Associated with Smoking | SpringerLink.” https://link.springer.com/article/10.5487/TR.2014.30.3.149 (accessed Sep. 18, 2021).

[24] Y. Taniyama and K. K. Griendling, “Reactive Oxygen Species in the Vasculature,” Hypertension, vol. 42, no. 6, pp. 1075–1081, Dec. 2003, doi: 10.1161/01.HYP.0000100443.09293.4F.

[25] P. Newsholme et al., “Diabetes associated cell stress and dysfunction: role of mitochondrial and non-mitochondrial ROS production and activity,” J. Physiol., vol. 583, no. 1, pp. 9–24, 2007, doi: 10.1113/jphysiol.2007.135871.

[26] S. Wang, X. Zhu, L. Xiong, and J. Ren, “Ablation of Akt2 prevents paraquat-induced myocardial mitochondrial injury and contractile dysfunction:

Role of Nrf2," Toxicol. Lett., vol. 269, pp. 1–14, Mar. 2017, doi: 10.1016/j.toxlet.2017.01.009.

[27] "Zeaxanthin induces Nrf2-mediated phase II enzymes in protection of cell death | Cell Death & Disease." https://www.nature.com/articles/cddis2014190 (accessed Sep. 18, 2021).

[28] "Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region - PubMed." https://pubmed.ncbi.nlm.nih.gov/7937919/ (accessed Sep. 18, 2021).

[29] "The small MAF transcription factors MAFF, MAFG and MAFK: Current knowledge and perspectives - ScienceDirect." https://www.sciencedirect.com/science/article/pii/S0167488912001693 (accessed Sep. 18, 2021).

[30] T. Fukutomi, K. Takagi, T. Mizushima, N. Ohuchi, and M. Yamamoto, "Kinetic, Thermodynamic, and Structural Characterizations of the Association between Nrf2-DLGex Degron and Keap1," Mol. Cell. Biol., vol. 34, no. 5, pp. 832–846, Mar. 2014, doi: 10.1128/MCB.01191-13.

[31] "The Carboxy-Terminal Neh3 Domain of Nrf2 Is Required for Transcriptional Activation." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1316965/ (accessed Sep. 18, 2021).

[32] S. Chowdhry, Y. Zhang, M. McMahon, C. Sutherland, A. Cuadrado, and J. D. Hayes, "Nrf2 is controlled by two distinct β-TrCP recognition motifs in its Neh6 domain, one of which can be modulated by GSK-3 activity," Oncogene, vol. 32, no. 32, pp. 3765–3781, Aug. 2013, doi: 10.1038/onc.2012.388.

[33] Ooi, B. K., Chan, K. G., Goh, B. H., & Yap, W. H. (2018). The Role of Natural Products in Targeting Cardiovascular Diseases via NRF2 Pathway: Novel Molecular Mechanisms and Therapeutic Approaches. *Frontiers in pharmacology*, *9*, 1308.

[34] da Costa, R. M., Rodrigues, D., Pereira, C. A., Silva, J. F., Alves, J. V., Lobato, N. S., & Tostes, R. C. (2019). NRF2 as a potential mediator of cardiovascular risk in metabolic diseases. *Frontiers in Pharmacology*, *10*, 382.

[35] Canning, P., Sorrell, F. J., & Bullock, A. N. (2015). Structural basis of KEAP-1 interactions with NRF2. *Free Radical Biology and Medicine*, *88*, 101-107.

[36] Satta, S., Mahmoud, A. M., Wilkinson, F. L., Yvonne Alexander, M., & White, S. J. (2017). The role of NRF2 in cardiovascular function and disease. *Oxidative medicine and cellular longevity*, *2017*.

[37] Gaulton, A., et al., *The ChEMBL bioactivity database: an update.* Scientific Data, 2013. 2: p. 150032.

[38] Kim, S., et al., *PubChem substance and compound databases.* Nucleic acids research, 2016. 44(D1): p. D1202-D1213.

[39] Ma, B., Lucas, B., Capacci, A., Lin, E. Y. S., Jones, J. H., Dechantsreiter, M., ... & Richter, K. (2020). Design, synthesis and identification of novel, orally bioavailable non-covalent Nrf2 activators. *Bioorganic & medicinal chemistry letters*, *30*(4), 126852.

[40] Jiménez, J., et al., *DeepSite: protein-binding site predictor using 3D-convolutional neural networks.* Bioinformatics, 2017. 33(19): p. 3036-3042.

[41] H. Chen, O. Engkvist, and T. Kogej, "Compound Properties and Their Influence on Drug Quality," in *The Practice of Medicinal Chemistry: Fourth Edition*, 2015.

[42] M. D. Shultz, "Improving the plausibility of success with inefficient metrics," *ACS Medicinal Chemistry Letters*. 2014.

[43] Y. Gao, C. Gesenberg, and W. Zheng, "Oral Formulations for preclinical studies: Principle, design, and development considerations," in *Developing Solid Oral Dosage Forms: Pharmaceutical Theory and Practice: Second Edition*, 2016.

[44] D. Kuntz, K. Chen, K. A. Sharp, and P. A. Kollman, "The maximal affinity of ligands,"
[45] *Proc. Natl. Acad. Sci. U. S. A.*, 1999.

[46] Cook, J.P., et al., *Identification of contact residues in the IgE binding site of human FcεRIα.* Biochemistry, 1997. 36(50): p. 15579-15588.

[47] M. Clementi, S. Clementi, S. Clementi, G. Cruciani, M. Pastor, and J. E. Nilsson, "Chemometric Detection of Binding Sites of 7TM Receptors," in *Molecular Modeling and Prediction of Bioactivity*, 2000.

[48] Á. Durán, G. C. Martínez, and M. Pastor, "Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in molecular interaction fields," *J. Chem. Inf. Model.*, 2008.

[49] S. Zafar and I. Jabeen, "GRID-independent molecular descriptor analysis and molecular docking studies to mimic the binding hypothesis of γ-aminobutyric acid transporter 1 (GAT1) inhibitors.," *PeerJ*, vol. 7, p. e6283, 2019.

[50] M. Pastor, G. Cruciani, I. McLay, S. Pickett, and S. Clementi, "GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors," *J. Med. Chem.*, vol. 43, no. 17, pp. 3233–3243, Aug. 2000.

[51] M. Clark and P. Labute, "2D depiction of protein-ligand complexes," *J. Chem. Inf. Model.*, 2007.

[52] M. Pastor, G. Cruciani, Iain McLay, A. Stephen Pickett, and S. Clementi, "GRid- INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three- Dimensional Molecular Descriptors," 2000.

[53] Á. Durán, G. C. Martínez, and M. Pastor, “Development and Validation of AMANDA, a New Algorithm for Selecting Highly Relevant Regions in Molecular Interaction Fields,” *J. Chem. Inf. Model.*, vol. 48, no. 9, pp. 1813–1823, Sep. 2008