Identifying plant based small molecule inhibitors for biologically important targets in cancer

by

Derya AYDIN

A Thesis Submitted to the

Graduate School of Science and Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Computational Science and Engineering

Koç University

September, 2012

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Derya Aydın

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Dr. Burak Erman (Advisor)

_____

Assoc. Prof. Dr. Meltem Müftüoğlu (Co-advisor)

_____

Prof. Dr. Özlem Keskin Özkaya

_____

Assoc. Prof. Dr. İhsan Solaroğlu

_____

Assoc. Prof. Dr. I. Halil Kavaklı

Date: 11.09.2012_____

ABSTRACT

In this thesis, we investigated plant secondary metabolites as potential drugs for cancer and identified plant based small molecule inhibitors for biologically important targets by screening a plant based library that we constructed. We carried out two projects and developed new methods for this purpose.

In the first part, we aimed to inhibit base excision repair (BER) pathway that is the major system for correcting many types of oxidative DNA damages and mono functional base modifications caused by endogenous and exogenous agents including chemotherapeutics. Whereas DNA repair pathways mediate resistance to DNA damage and protect the cells from its killing effects, and maintain genomic stability, inhibition of DNA repair pathways is required for the toxicity of several anticancer drugs and ionizing radiation. Because the cytotoxic effects of most chemotherapeutic agents and radiation are related to their ability to induce DNA damage. The DNA repairing system of cancer cells is therefore disadvantageous to cancer treatment. We identified plant based small molecule inhibitors for two critical BER proteins, mitochondrial DNA polymerase γ and nuclear DNA polymerase β as potential therapeutic targets to sensitize tumors. We found an inhibitor for known active site of polymerase β with more significant interactions than found in previous studies. We also identified three promising inhibitors for polymerase γ targeting a specific interface of polymerase γ for the first time in literature.

In the second part, we investigated plant secondary metabolite specificities on three biologically important domains, PDZ, BROMO and SPRY, which have high drug target potency for a large variety of diseases including cancers. Interest in molecular target-class specificity information is getting increased due to the advantages into drug

design. However, studies about information on molecular targets are lacking systematic methods to identify the specificities. We developed a novel method that will enable to analyze and identify the domain-class specificity more quantitatively and for this purpose we used the Gibbs Distribution of binding energies in order to differentiate different ligands. We identified three specific plant secondary metabolite classes, Triterpenoids, Stilbenoids and Stereoidal alkaloids for PDZ, BROMO and SPRY domains, respectively.

ÖZET

Bu çalışmada ikincil bitki metabolitlerinin kanser türleri üzerinde potasiyel ilaç etkisi araştırılmış ve oluşturduğumuz bitki tabanlı ilaç kütüphanesi taranarak biyolojik olarak önemli hedeflere bitki kökenli küçük molekül inhibitörleri saptanmıştır. Bu çalışmada iki proje yürütülmüş ve bu amaçlara yönelik yeni metodlar geliştirilmiştir.

İlk bölümde kemoterapik ilaçlarıda içine alan iç ve dışetkenlerin neden olduğu bir çok oksidatif DNA hasarları ve fonKsiyonel tekil baz değişimlerini düzelten nükleotideksizyon tamir (NET) makenizması hedef alınmıştır. DNA tamir mekanizmları DNA hasarlarına karşı direnç sağlayarak hücrenin yaşamsallığını korur ve genetik materyalin stabil olarak korunmasını sağlar. DNA tamir mekanizmalarının inhibe edilmesi bir çok kanser ilacının hücre toksitesi ve iyonlaştırıcı radyasyonun etkisini artırmak için gereklidir. Çünkü genelde kemoterapik ilaçların ve radyasyonun hücre toksitesi yaptıkları DNA hasarından kaynaklanmaktadır. Biz bu çalışmada NET mekanizmasında bulunan iki önemli protein için bitkisel kökenli küçük molekül inhibitörleri bulduk. Bu proteinler DNA polimeraz β ve mitokondriyal DNA polimeraz γ tümörlerin direncinin düşürülmesi için olası hedefler olarak şeçilmiştir. DNApolimeraz β' nin bilinen aktif bölgesi için saptadığımızinhibitör daha önce literatürde bulunanlara göre protein ile dahakuvvetli etkileşim göstermektedir. Literatürde ilk kez mitokondriyal DNA polimeraz γ' nın etkileşim yüzeyi için 3 adet kuvvetle olası inhibitör bulunmuştur.

İkinci bölümde, biyolojik olarak önemli, kanseri de içeren çeşitli hastalıklar için hedef olma potansiyeli yüksek olan üç domaine özgü bitki sekonder metabolitlerini inceledik. İlaç tasarımına sağladığı avantajlardan dolayı, moleküler hedef-sınıf

özgülüğü bilgisine olan ilgi artmaktadır. Ancak, moleküler hedeflerin özelliklerini belirlemeye yönelik çalışmalar özgülükleri beliryebilecek sistematik metodlara haiz değildir. Biz domain-sınıf özgülüğünü analiz ve belirlemeyi sağlayacak yeni bir yöntem geliştirdik ve bu amaçla ligandları ayırt etmek için bağlanma enerjilerinin Gibbs dağılımını kullandık. PDZ, BROMO ve SPRY domainleri için sırasıyla Triterpenler, Stilbenler ve Steroid alkaloitleri belirledik.

ACKNOWLEDGEMENTS

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter1

1. INTRODUCTION

1.1. PART1: IDENTIFYING SMALL MOLECULE INHIBITORS TARGETTING BASE EXCISION REPAIR ENZYMES, DNA POLYMERASE γ AND β

Maintenance of the genome integrity is critical for the survival and health of an organism. The stability of our genome is constantly challenged by exogenous and endogenous damaging agents. It has been estimated that more than 100,000 endogenous DNA lesions occur in each mammalian cell per day from spontaneous decay, replication errors, and cellular metabolism alone [1]. A major endogenous source of DNA damage is the reactive oxygen species (ROS), mainly generated by the mitochondrial respiration process. Most of these DNA lesions are mutagenic and/or cytotoxic and have been associated with aging, neurodegeneration and carcinogenesis. Therefore, in order to maintain genomic integrity, it is essential that DNA damage is repaired [2].

Whereas DNA repair pathways mediate resistance to DNA damage and protect the cells from its killing effects, and maintain genomic stability, inhibition of DNA repair pathways is required for the toxicity of several anticancer drugs and ionizing radiation. Because the cytotoxic effects of most chemotherapeutic agents and radiation are related to their ability to induce DNA damage. The DNA-repairing system of cancer cells is therefore disadvantageous to cancer treatment. Base excision repair (BER) is the major system for handling many types of oxidative DNA damages and monofunctional base modifications such as alkylation of DNA caused by endogenous and exogenous agents including chemotherapeutics. For instance, ionizing radiation, bleomycin,

monofunctional alkylators, antimetabolites and platinum drugs like cisplatin have been utilized extensively in the treatment of several cancer types; and these drugs generate ROS causing oxidative DNA damage, and BER of these lesions inhibit the cytotoxicity of anticancer treatments [3-5].

BER is a multistep process and takes place both in nucleus and mitochondria with the similar mechanism, but the steps are carried out by a distinct protein or a unique form (a specific splice product) of a common gene. BER is initiated by lesion specific DNA glycosylases that recognize and cleave modified or inappropriate DNA bases [6]. 8-oxogunanine DNA glycosylase (OGG1) is the primary enzyme for the repair of 8-oxoguanine (8-oxoG) DNA lesion in both nuclear and mitochondrial DNA (mtDNA). OGG1 is a bifunctional enzyme, which also possess AP-lyase activity to cleave DNA at AP sites. After either removal of the substrate base or incision 3` to AP site by bifunctional DNA glycosylases, apurinic/apyrimidinic endonuclease I (APE1) incise the DNA backbone immediately adjacent to the lesion, creating a single strand break (i.e., the deoxyribose phosphate, dRP). The gap generated by a cleavage of abasic site is filled-in by DNA polymerase β (pol β) in nuclear DNA but by DNA polymerase γ (pol γ) in mtDNA. It is followed by strand ligation by DNA ligase [6, 7].

Mammalian mtDNA is found in large-protein DNA complexes known as nucleoids, which are associated with the inner side of the mitochondrial inner membrane [8]. This location exposes to mitochondrial genome to ROS, such as superoxide anion, hydrogen peroxide and the hydroxyl radical, produced by membrane-bound electron transport proteins. Thus, mtDNA is more vulnerable to being damaged by ROS in comparison to nDNA. Several groups have demonstrated that the levels of oxidized bases in mtDNA are higher than in nDNA. MtDNA is also lacks histone proteins [9, 10]. Thus, targeting selected BER repair enzymes to mitochondria may lead to an imbalance of the BER repair pathway and increased cell death at lower doses than with nuclear imbalancing BER. This approach could be useful in cancer therapy

approaches to sensitizing tumor cells to lower doses of chemotherapeutic agents with enhanced cell killing. Importantly, downregulation of mitochondrial BER proteins` expression including OGG1, pol γ or transcription factor A by RNA interference (siRNA) technique enhanced the susceptibility of oral squamous cell carcinoma cells to gamma-rays [11]. Other studies demonstrated that siRNA mediated pol β downregulation is sufficient to increase temozolomide efficacy that is effective against glioblastoma and melanoma [3]. However, these studies did not reach beyond the preclinical level. Therefore, we aimed to identify the plant based small molecule inhibitors of human nuclear or mitochondrial BER proteins, DNA polymerase β, or polymerase γ, respectively to improve the cancer therapy and to differentiate the biological outcome of mitochondrial and nuclear BER pathways in the therapy.

We identified plant based small molecules for DNA pol β and pol γ by screening a plant based molecule library which consists of plant secondary metabolites that we collected from various databases. The library is composed of about 8550 plant secondary metabolites that provide 3D structures that we obtained from various online available databases: KEGG Phytochemical Compounds [12], Analyticon-discovery MEGxp Pure Natural Plant Products [13], Seaweed Metabolites Database (SWMD) [14], Indofine Chemicals [15], and Indian Plant Anticancer Compounds Database (InPACdb) [16]. We determined the binding residues of the enzymes according to the previous studies that identified the active residues [17-20] [21, 22] depending on their contribution to DNA binding and processivitybu sozcugu iyi arastir lutfen of the enzyme using the techniques of deletion or mutation in residues. We performed dockings of plant based library for DNA pol β and pol γ by using the docking Software Autodock VINA [23] and identified hit molecules based on ranking average binding free energies. We identified Fucax23, Xntmf27 and Dxbd20 from the KEGG Phytochemicals Database as promising inhibitors of pol γ, with -10.7 kcal/mol, -9.8

kcal/mol and -9.8 kcal/mol binding free energies, respectively. Anxbd47 was identified as inhibitor for pol β with -10.7 kcal/mol binding free energy.

Additionally, we analyzed our docking results for pharmacophore modeling by Ligand Scout to model the key features in protein-ligand complexes. We observed the interactions of our models with important residues which are identified in previous studies [17, 19, 21, 22, 24-26].

1.2. PART 2: IDENTIFYING SPESIFIC SECONDARY METABOLITE CLASSES AS INHIBITORS FOR DOMAINS INVOLVED IN DISEASES INCLUDING MANY CANCERS

Plant secondary metabolites constitute an excellent reservoir of biologically active compounds. For centuries, extracts from plant products have been a main source of medicines. Plant secondary metabolites exhibit very large structural diversity and encompass many classes of metabolites. In this study we tried to identify the secondary metabolite specify of three biologically important domains, the

PDZ, BROMO and SPRY domains, which are promising, drug targets. We investigated class specify of these targets in order to provide insights into herbal medicine and drug design of many cancers and diseases.

Plants produce a vast range of organic compounds that are traditionally classified as primary and secondary metabolites. Primary metabolites are compounds that have critical roles associated with photosynthesis, respiration, and growth and development. These include phytosterols, acyl lipids, nucleotides, amino acids and organic acids. Other phytochemicals, Secondary metabolites are compounds produced in response to stress, such as the case when acting as a deterrent against herbivores [27]. They belong to extremely varied chemical groups such as organic acids, aromatic

compounds, steroids, terpenoids, alkaloids, flavonoids, carbonyles etc. Even though secondary metabolism generally accounts for less than 10% of the total plant metabolism, its products are the main plant constituents with pharmaceutical properties[28].

Plant products are also important sources for high-throughput screening which has become an important tool in drug discovery. However, this technique requires a large number of compounds to be effective. Secondary metabolites, previously regarded as waste products are now recognized for their resistant activity against pests and diseases and their chemodiversity in nature offers a valuable source for high-throughput screening. Chemodiversity in plants and the role of secondary metabolites should be investigated, which might be helpful in the optimization of the lead discovery process.

Plant compounds exhibit large structural diversity. However, only a small proportion of that diversity has been seriously explored for its pharmacological potential so far and there is a an important potential of new classes to be explored for their pharmacological properties [29].

Plant secondary metabolites have high drug potency and there are many metabolites known as drugs for various diseases including cancer [30]. However, comparing to the other areas of pharmaceutical research there is less appropriate data for the screening of natural products. On the other hand, pharmacological activity through molecular target specificity of Chinese Herbs is increasing [31, 32]. This information is used in the field of virtual screening by screening of these structures to various particular receptors in order to provide the knowledge for new therapeutic properties.

Information on molecular targets of plant compounds has also increased over the past few years However, only a few studies are published. In one of these studies, two

databases Chinese Herbal Constituents Database (CHCD) and Bioactive Plant Compounds Database (BPCD) were constructed in order to provide information of molecular structure data, botanic sources, specific molecular targets and therapeutic categories of bioactive plant compounds [29]. Although this study provides useful information about molecular target specifites of plant products, the study does not explain the methods used to identify the class specificities rationally. Since the numbers of molecules in secondary metabolite classes are different, a more systematic approach is required for determination and verification of the classes. The purpose of our study was to develop a rational method that will enable to analyze and identify the domain-class specifity more quantitatively and for this purpose we used the Gibbs Distribution of binding energies in order to differentiate different ligands for the first time in literature.

The other important point in our study is the selection of targets. While previous studies focused on various groups of proteins [29], we targeted 3 critical and well-known domains which are hot topics and promising drug targets. PDZ, BROMO and SPRY are mostly studied domains which have considerable drug target possibilities due to their roles in many proteins [33-36]. The domains are involved in various critical druggable protein-protein interactions which mediate many cancers and Familial Mediterranean Fever (FMF). The PDZ domain is known to be involved in many cancers such as colon, breast, liver, lung, pancreas, stomach, and prostate cancers [37-40]. The BROMO domain is also known as a possible drug target for different cancer types such as breast and lung cancers and leukemia [41, 42]. Several human diseases arise from mutations in genes coding for SPRY domain containing proteins, for example mostly mutations in the Spry domain of MEFV gene are associated with Familial Mediterranean Fever [43]. In literature there are many studies about structures, functions, drug possibilities and inhibitors of these domains [33-36]. However, there is not a study published yet about their secondary metabolite specifity. The knowledge of secondary class specifity of these domains will provide insights into drug design for

many cancers and diseases. Additionally it is important to know class specifities in order to decrease computational cost and focus on the specific groups instead of screening all molecule libraries of molecular targets.

We identified the secondary metabolite class specifity of domains as their bioactive inhibitors firstly based on high average binding free energies and then using Gibbs distribution method. The docking regions of target domains PDZ, BROMO, SPRY were selected according to previous studies in the literature [36, 43-50]. We performed virtual screening of KEGG Phytochemical compounds [12] which is exactly classified according to secondary metabolite classes for each domain by using docking software VINA and used ranked binding free energies. We analyzed the docking results based on the high average binding free energies and identified 5 promising classes for domains: Triterpenoids, Flavonoids, Stereoidal alkaloids, Stilbenoids, Anthroquinones. Since we do not have equal number of molecules in these classes, we derived new sets of compounds of these promising classes using the software BROOD. Then, we docked these sets for each of the three domains. We analyzed the docking results of derived sets using Gibbs distribution. Docking softwares calculate the Gibbs free energy of binding empirically. We use this to obtain the binding probabilities of each drug molecule. We sum the binding probabilities of members of the each class for different number of sample hit sets. Comparing binding probabilities of each class for each domain we identified Triterpenoids, Stereoidal alkaloids, Flavonoids as specific inhibitor candidates for PDZ, SPRY, BROMO domains respectively which also verify our previous results.

Additionally we analyzed our docking results for pharmacophore modeling by Ligand Scout to model the key features in protein-ligand complexes in order to understand binding pocket properties of domains. We observed the interactions of our models with important residues which are identified in previous studies [36, 43-50].

Chapter 2

2.   LITERATURE REVIEW

2.1.   DNA REPAIR MECHANISMS

DNA is damaged every day by a variety of factors. This damage must be repaired quickly and efficiently to sustain the integrity of the genome. Common errors include the loss of bases resulting in AP sites (abasic sites) [51]; base modifications, such as alkylations or deaminations which converts cytosine, adenine and guanine to uracil, hypoxanthine and xanthine, respectively. Photodamage by UV light can generate pyrimidine dimers, such as cyclobutane pyrimidine dimers (CPDs) and pyrimidine pyrimidone photoproducts. Chemical agents and ROS can also modify the bases. Replication errors and base conversions can generate mismatch nucleotide pairs. Failures in normal DNA metabolism by topoisomerases and nuclease or ionizing radiation can generate single-strand and double-strand breaks. Thus, cells have evolved a number of DNA repair mechanisms to respond to DNA damage.

Mammalian cells utilize five major DNA repair pathways: Homologous recombination (HRR); Nonhomologous End Joining (NHEJ); Nucleotide Excision Repair (NER); Base Excision Repair (BER) and Mismatch repair (MMR) [52].

BER is the predominant DNA repair system in mammalian cells for eliminating oxidative DNA base lesions [53]. BER is a multi-step process which starts with a DNA glycosylase that recognizes and removes incorrect or damaged bases. DNA glycosylase cleaves the N-glycosyl bond between the deoxyribose sugar moiety and the damaged DNA base, generating an AP site. Resultant AP sites are the substrate of APE1, which cleaves the sugar–phosphate backbone 5′ to the AP site, leaving a 3′-hydroxyl group

and a 5′-deoxyribose phosphate group next to the nucleotide gap [54]. In the next step, the DNA polymerases incorporates the undamaged nucleotides and then removes the abasic sugar–phosphate (dRP) group, producing a nicked duplex DNA molecule [55]. The resultant gap is subsequently sealed by DNA ligase [56] (Figure 1). The specificity of BER is provided by DNA glycosylases that act on specific substrates. Four major pathway components (DNA glycosylase, AP endonuclease, DNA polymerase and ligase) have to be precisely regulated and highly coordinated for error-free and efficient DNA repair [53].
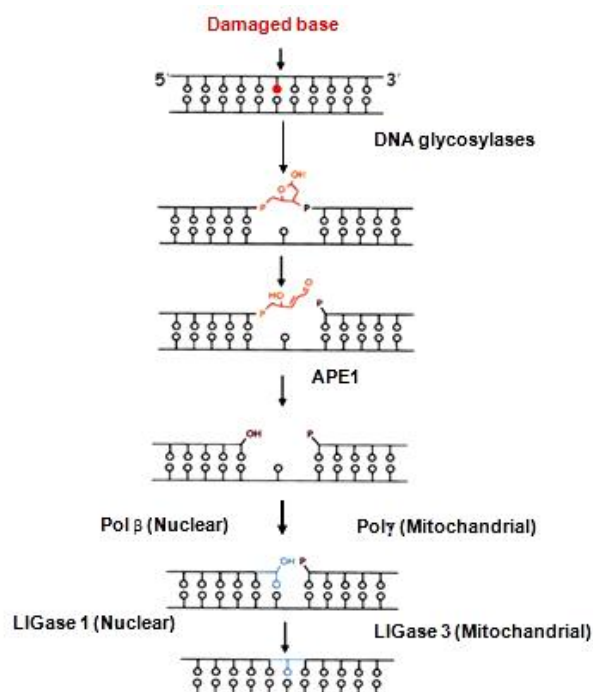
Figure 1 Mammalian Base Excision Repair Pathway

MMR pathway repairs mismatches and loops generated by insertions or deletions. The system is responsible for the post-replicative repair of mismatches and

small single stranded DNA loops, and it is critically involved in preventing recombination between homologous DNA sequences. MMR removes nucleotides mispaired by DNA polymerases and IDLs (Insertion/Deletion Loops) that result from slippage during replication of repetitive sequences or during recombination [57]. The eukaryotic MMR system involves two different heterodimeric complexes of MutS-related proteins, MSH2-MSH3 (known as MutSBeta) and MSH2-MSH6 (known as MutSAlpha), and each has different mispair recognition specificity. These proteins have important roles in genetic recombination beyond the repair of mispaired bases by regulating the resolution of holliday junctions and is involved in cancer especially Hereditary Non-Polyposis Colorectal Carcinoma. Hereditary non-polyposis colon cancer is linked to defects in genes of the MMR pathway. Therefore, in this cancer type MMR is inactive.

NER is perhaps the most flexible of the DNA repair pathways considering the diversity of DNA lesions it acts upon. The most significant of these lesions are pyrimidine dimmers (cyclobutane pyrimidine dimmers and 6-4 photoproducts) caused by the UV component of sunlight. The NER process involves damage recognition, local opening of the DNA duplex around the lesion, dual incision of the damaged DNA strand, gap repair synthesis, and strand ligation. Whereas BER processes are dependent upon specific N-glycosylases to recognize mismatches or damage, the NER pathway recognizes abnormal structures and chemistry via heterodimers composed of DDB1, XPE, XPC or HR23B proteins.

As a result it is obvious that repair mechanism includes consecutive steps of enzyme reactions and proteins have important roles in these steps. Therefore inhibiton of specific proteins of the repair mechanism can lead to deactivation of the mechanisms. But in case of inactivation of the certain repair system another mechanism can compensate and can carry on the repairing function. For example, in colon cancer due to

inactivation of MSH1-MSH2 MMR can not work and the BER mechanism carries on the DNA repair. Therefore, cancer treatments like γ-radiation or chemotheraphy are not effective and can not lead to apoptosis. For such cases, it is obvious that inhibiton of the BER system can increase the cancer treatment impression.

2.2.   TARGETING BER AS A THERAPEUTIC APPROACH

Targeting signaling pathways, cell cycle check points and antiangiogenesis are common therapeutic aproach [58]. A new and emerging approach is block to repair damaged DNA that chemotheraphy agents induce. Since the efficiacy of DNA repair in cancer cells results in drug resistance and decrease the effects of the therapy, balance between DNA damage and repair determines the final therapeutic consequences of these agents [59]. Therefore, interference with DNA repair has emerged as an important approach to combination therapy against such cancers.

The DNA damage induced by chemotherapeutic drugs is repaired primarily by the O6-methylguanine DNA methyltransferase (MGMT), MMR, and BER pathways. There are studies mainly target the MGMT and MMR pathways. The blockage of the BER pathway has been overlooked, although in the case of several DNA-alkylating drugs, including temozolomide BER is responsible for the repair of 70%, 5%, and 9% of N7-methylguanine (MeG), N3-MeG, and N3-methyladenine (MeA) lesions, respectively [60]. Any interruption of the BER pathway can cause an accumulation of these lesions which will be used as chemotherapeutic approach [61].

Many colon tumors become resistant to DNA-alkylating agents due to overexpression of MGMT or MMR deficiency [62]. In previous studies, the role of BER pathway has also been implicated in cellular resistance to TMZ, which depends on specific BER gene expression and activity [63]. Thus, downregulating the BER

pathway can reduce the resistance to DNA-alkylating agents and increase their efficacy to colon cancer cells.

In a few studies small molecule inhibitors have been identified by molecular docking to target the BER pathway. APE1,one of the main enzymes in the BER was targetted to prevent the elevation of APE1 expression which has been associated with poor outcome to chemoradiotherapy. The small molecule lucanthone has been shown to enhance oxidative damage and the killing ability of ionizing radiation in cancer cells [64]. In another study, NSC-666715 was identified as a pol β inhibitor that enhances the cytotoxicity of the DNA-alkylating drug TMZ by sensitizing the cancer cells [65].

2.3. DNA POLYMERASES

Key enzymes for the maintenance of the integrity of the genome are DNA polymerases involved in the DNA synthesis, repair of DNA lesions, and recombination.

2.3.1. DNA POLYMERASE BETA

DNA pol β is a polymerase enzyme which is involved in the DNA synthesis step of the single-nucleotide BER pathway, which processes small DNA lesions such as oxidized or alkylated bases.

Pol β is the lowest molecular mass (39 kDa) DNA polymerase and consists of an independently folded N-terminal 8-kDa domain and C-terminal a 31-kDa domain (Figure 2). Many structures of pol β in liganded states show that the enzyme has a modular domain organization. Pol β is composed of a polymerase (colored) and an amino-terminal lyase domain (grey). The polymerase domain (colored) is composed of

three subdomains: D (purple), C (gold), and N (green) subdomains (Figure 2). These subdomains are involved in duplex DNA binding, nucleotidyl transfer, and binding of the nascent base pair, respectively. These subdomains are referred to as C– (Catalytic), D– (Duplex DNA binding), and N– subdomains (Nascent base pair binding) to highlight their intrinsic function [66]. These would correspond to the palm, thumb, and fingers sub-domains, respectively, according to the nomenclature that utilizes the architectural analogy to a right hand. In addition to the polymerase domain, DNA polymerases often have an accessory domain that contributes a complementary enzymatic activity necessary for the polymerase to fulfill its biological task. For example, pol β contains an amino-terminal lyase domain that contributes a lyase activity required to remove a 50-dRP intermediate during BER. The panel on the right in (Figure 2) includes gapped DNA. The template strand (red) bends at 90º as it exits the nascent base pair binding pocket. The templating (coding) nucleotide is shown with the incoming nucleotide (blue surface). The remaining nucleotides are omitted for clarity [67].
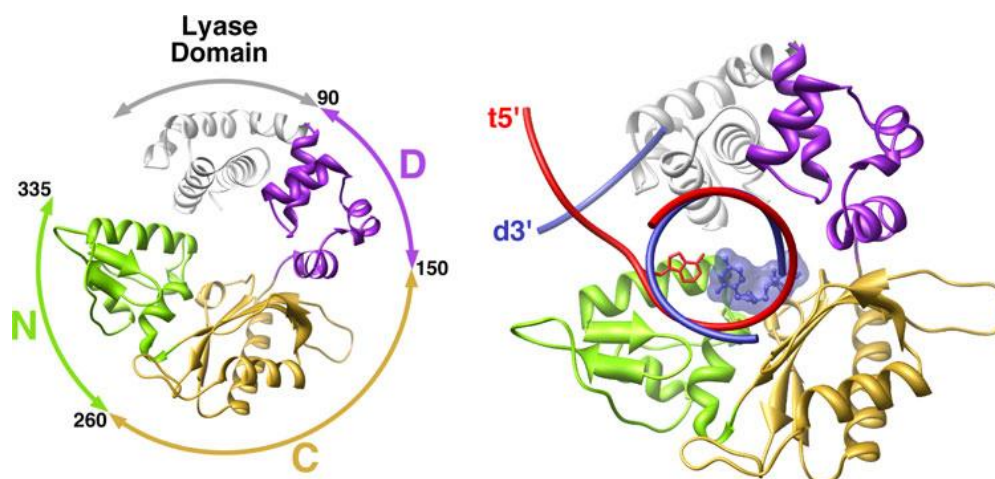
Figure 2 Domain and subdomain organization of DNA pol β [67]

The crystal and solution structures of the aminoterminal 8-kDa domain have been determined [20]. The 8-kDa domain (residues 1–87) is formed by four a-helices,

packed as two antiparallel pairs. The pairs of a-helices cross one another at 50° giving them a V-like shape. The 8-kDa domain of b-pol also contains a motif termed "Helixhairpin-Helix" (HhH) [17, 24, 25].

In several studies it is reported that N-terminal 8 kDa domain of pol β retains binding affinity for ssDNA, 5'-phosphate recognition in gapped DNA, and dRP lyase activity using an intact AP site containing DNA substrate [18, 19]. Residues of the HhH motif have been proposed to contribute to recognition and excision of damaged nucleotides in DNA, as well as AP lyase chemistry [17, 24, 25]. The Helix-3-hairpin-Helix-4 motif and residues in an adjacent V-type loop form the ssDNA interaction surface. The x-ray crystal structure of pol β bound to a template-primer substrate suggested that four lysine residues (residues 35, 68, 72, and 84) in this region of the protein coordinate the DNA 5'-phosphate that may exist in a gapped DNA [17, 24, 25]. However, in structures of pol β bound to a one-nucleotide DNA gap only Lys35 and Lys68 coordinate the 5'-phosphate in the short gap [68].

As a result of site directed mutagenesis of 9 residues in the 8-kDa domain of pol β reported by Prasad et al. [18], Lys35, Lys60 and Lys68 were impaired in the ssDNA binding activity, Lys35 was involved in 5'-phosphate recognition, and including Lys72 was significantly reduced in dRP lyase activity. The dRP lyase active site Lys72 was observed to be part of a lysine-rich pocket consisting of Lys35, 68, 72, 84 and 87 on the surface of the N-terminal 8 kDa domain that may also bind to ssDNA [18].

In one of the previous studies target the pol β, a pyhtotoxin, solanapyrone A has been found to reduce the ssDNA binding activity of pol β experimentally and by moleculer docking methods. In that study, solanapyrone A shows a high binding affinity with the residue Lys60 (-27.212 kcal/mol) which is known as one of important residue for ssDNA binding of pol β [18], whereas it does not have interaction with neither residues Lys35 and Lys68 which have a critical role in ssDNA binding activity nor also

the residues Lys72, 84 and 87 in lysine-rich binding pocket on the N-terminal domain binding affinity in the lysine rich pocket consisting of Lys35, 68, 72, 84 and 87 which may also bind to ssDNA.

DNA pol β maintains genome integrity by participating in nuclear BER. Studies showed that overexpression of pol β correlated with a number of cancer types, whereas deficiencies in pol β results in hypersensitivity to alkylating agents, induced apoptosis, and chromosomal breaking [69-72]. One of these studies showed that genetic instability increased in cells pol β overexpressed. Besides this, the cells showed decrease in sensitivity to chemothreapic agents [72]. Another study showed that increased activity and expression of this enzyme appear in several cancer cells including human ovarian tumor cells which are resistant to antitumor agents [71]. Therefore, it is essential that pol β expression is tightly regulated.
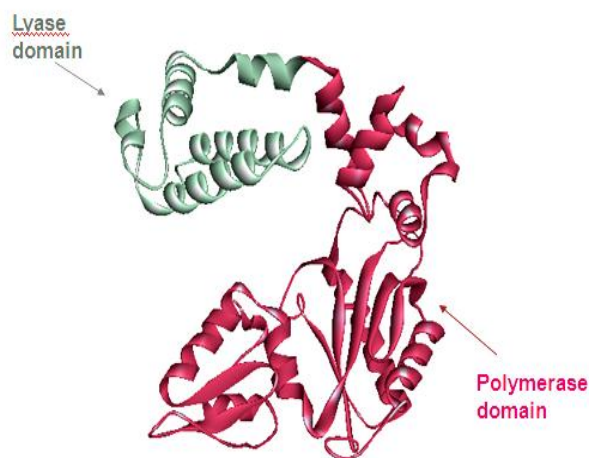
Figure 3 Domains of Pol β (The structure is visualized by Discovery Studio (PDB code: 9ICA)
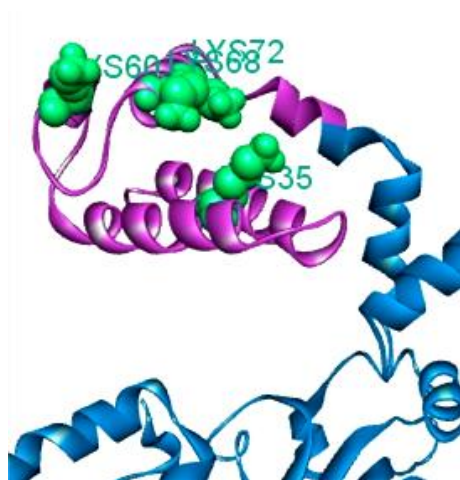
Figure 4 Lyase domain of Pol β (purple) with the active residues Lys35, Lys60, Lys68 and Lys72 (green). (The structure is visualized by Discovery Studio (PDB code: 9ICA)

In this study, inhibition of BER pathway is purposed by targeting pol β. We aimed to inhibit the binding of ssDNA in order to inhibit the polymerase function of the pol β enzyme during BER. We targeted lysine-rich pocket in the lyase domain (Figure 3 Figure 4) consisting of the residues Lys35, Lys60 and Lys68 which are critical in 5'-phosphate recognition in gapped DNA and the ssDNA binding activity and the other important residues for DNA binding in lysine-rich binding pocket including Lys35, 68, 72, 84 and 87.

2.3.2. DNA POLYMERASE γ

Pol γ is responsible for all aspects of mtDNA synthesis, including all replication, recombination of the mitochondrial genome, and repair of mtDNA damage during BER.

2.3.2.1.   STRUCTURE

Mitochondria contain a single DNA polymerase, pol γ, responsible for replication and repair of mtDNA. Human pol γ is isolated from mitochondria as a complex containing two subunits, a catalytic subunit, pol γA, of 139 kDa and an accessory subunit, pol γB, of 53 kDa (2–5). The catalytic subunit is a family A DNA polymerase with separate polymerase and 3'–5' exonuclease domains [26].

The catalytic subunit Pol γA contains domains for exonuclease (*exo*) and polymerase (*pol*) activities separated by a linker or spacer. Pol γA adopts the canonical polymerase 'righthand'configuration with subdomains of 'fingers', 'palm' and 'thumb' that bind template DNA and substrate nucleotide triphosphate, as well as catalyze phosphodiester bond formation (Figure 5A). The spacer domain is spatially far from the *exo* and *pol* domains and connects to them only through the long helices of the thumb subdomain. The spacer has two obvious subdomains, a globular IP subdomain (Intrinsic Processivity, residues 475-510 and 571-785) and an extended AID subdomain (Accessory Interacting Determinant, residues 511-570) that reaches more than 50 Å away from the main body of Pol γA (Figure 5) [73].
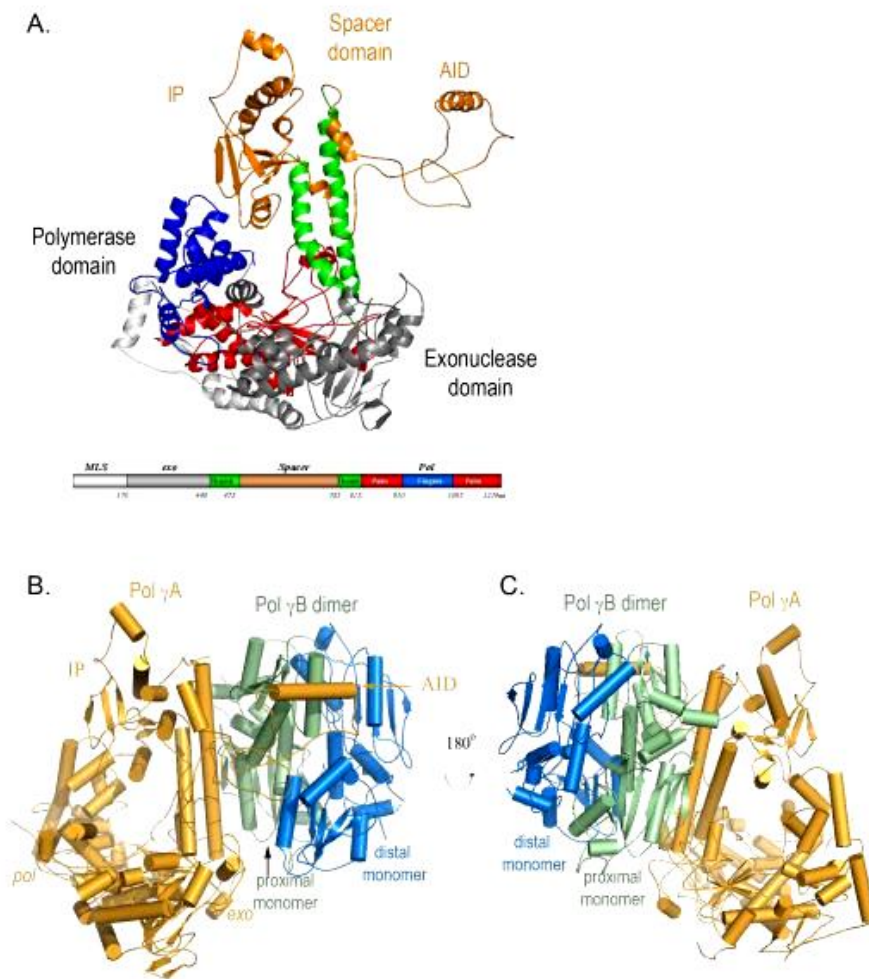
Figure 5 A)Structure of Pol γA   B) Structure of the heterotrimeric Pol γ holoenzyme containing one catalytic subunit Pol γA (orange) and the proximal (green) and distal (blue) monomers of Pol γB [73]

2.3.2.2. MECHANISM OF HALOENZYME

Pol γ, like other DNA replicases, has a catalytic subunit, pol γA, which possesses both polymerase and proofreading exonuclease activities, and an accessory subunit, pol γB. A novel mode of action was provided about the accessory subunit enhancement to mechanism of processivity [73]. Another feature of pol γB is that unlike other processivity factors that enhance processivity by increasing enzyme affinity for template DNA, pol γB enhances processivity by simultaneously accelerating polymerization rate and suppressing exonuclease activity, in addition to increasing affinity for DNA. It is predicted that the majority of the interaction between subunits occurs between one monomer of the dimeric accessory subunit and the catalytic subunit [26]. The accessory interacting determinant (AID) subdomain of the spacer region, approximately 50 Å from the polymerase domain, provides enhancement to DNA binding (Figure 6). This subdomain forms hydrophobic interactions between the L-helix of the catalytic subunit with the proximal accessory subunit [22].
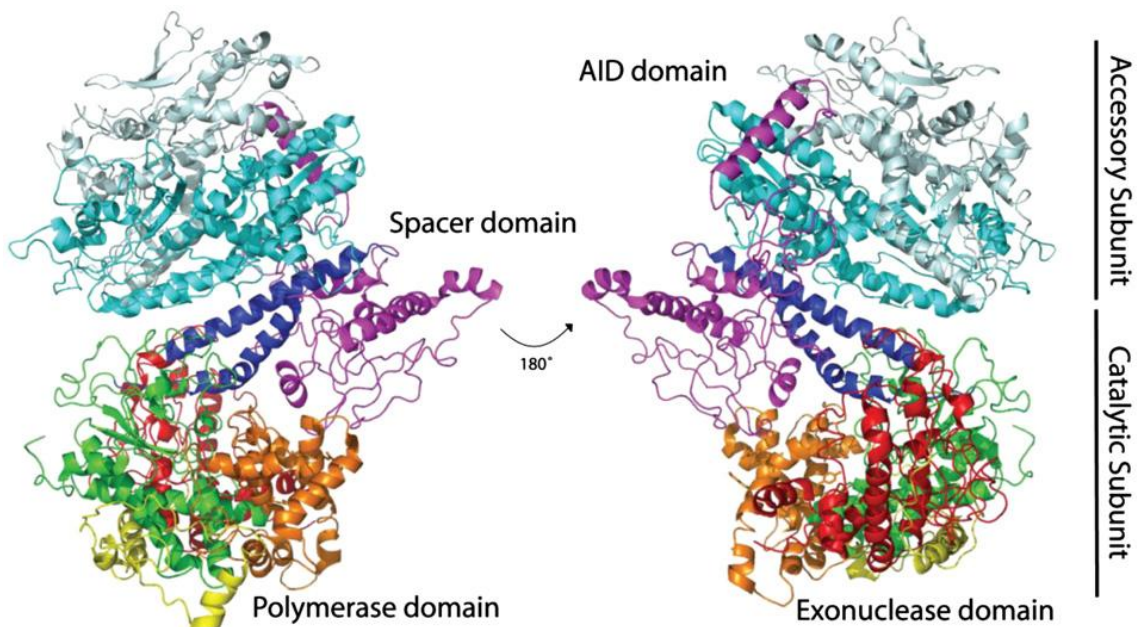
Figure 6 Crystal structure of Pol γ holoenzyme. Fingers (orange), palm (green) and thumb (blue). Domains highlighted include the mitochondrial localization sequence (yellow), exonuclease (red) and spacer subdomain (purple). The proximal (cyan) and distal (light cyan) monomers of the dimeric accessory subunit are shown [22]

Though pol γA has limited interaction with the distal monomer, it makes extensive interactions with the proximal Pol γB monomer (Figure 7B). Examination of the subunit interface shows two major areas of hydrophilic interactions: between pol γB R264, K373 and D459 and the pol γA thumb domain area (E454-D469, and R579) (Figure 7A), and between pol γB (D253 and D277) and pol γA (K1198, R1208 and R1209). In addition, hydrophobic interactions occur between a pol γB hydrophobic core (V398-L406, V441-L455) in the C-terminal region and pol γA AID subdomain L-helix (V543-L558) (Figure 7C). Studies show that pol γA AID causes the steric clash in the modeled A2B2 tetramer; in the absence of stabilizing hydrophobic forces for a second Pol γA monomer, the holoenzyme is therefore heterotrimeric. In turn, the modeling suggests that the hydrophobic interface is dominant in subunit interaction [73, 74]. A further study demonstrated was that the proximal monomer alone tightens the binding affinity between pol γ and the primer–template. However, rate enhancement is only achieved in the presence of the dimeric accessory subunit, presenting a role for the distal monomer of the accessory subunit complex. This effect can result from proper positioning of the primer–template which is only accomplished when the catalytic subunit is bound to the dimeric accessory subunit. With many tests it is showed that hydrophobic interactions between the pol γA L-helix and the C-terminal domain of pol γB are the dominant attractive forces that stabilize the AID subdomain so that it can support processive DNA synthesis by the holoenzyme [73].
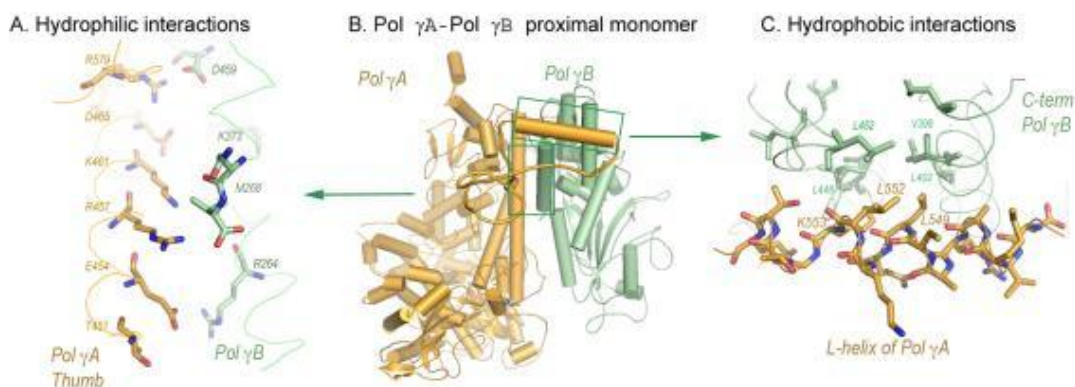
Figure 7 Interactions between proximal accessory subunit and catalytic subunit [73]

It is seen that the most important interactions that adopt the structure and enhance the processivity of the holoenzmye are the hydrophobic interactions that occur between the proximal accessory subunit and the L-helix of AID subdomain of the catalytic subunit (Figure 8). In this study we targeted this specific interaction site as a novel approach in order to inhibit the activity of the holoenzyme. We identified the active residues in the proximal accessory subunit that interact with the L-helix. We identified the binding pocket of the proximal accessory subunit containing the residues Leu402, Val398, Leu448, Leu395, Leu394 and Glu449 (green-brown) (Figure 9) which make important hydrophobic interactions with the catalytic subunit as shown in previous studies covered above. We aimed to identify a small molecule inhibitor for this pocket in order to prevent the interaction of proximal accessory subunit with L-helix of catalytic subunit for the first time in literature.
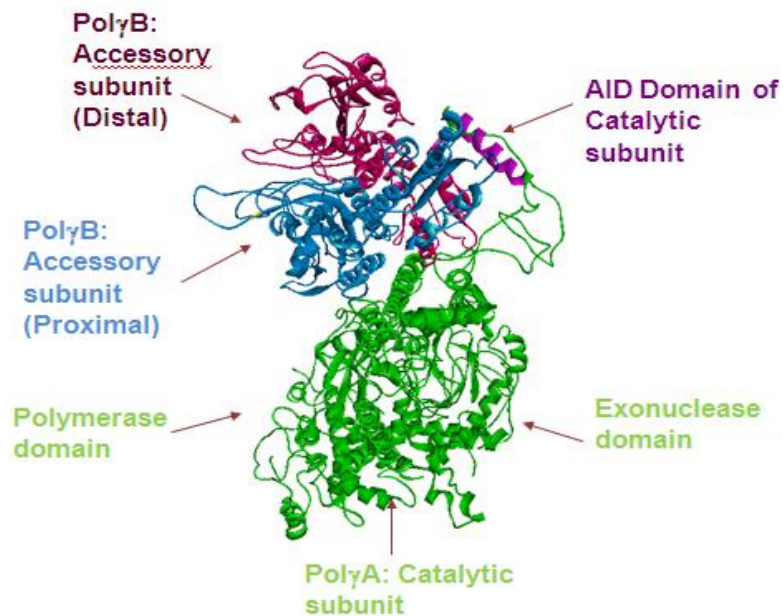
Figure 8 Domains of pol γ. (The structure is visualized by Discovery Studio (PDB code: 3IKM)
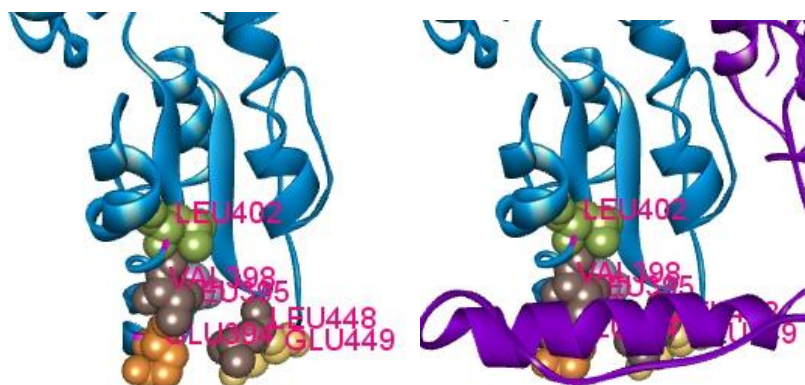
Figure 9 L-helix of AID domain in catalytic subunit (purple) interacts with critical residues of Proximal Accessory subunit (green-brown) in the binding pocket of Proximal Accesory subunit (blue). (The structures were visualized by Discovery Studio (PDB code: 3IKM)

2.4. PLANT SECONDARY METABOLITES

Plants produce primary and secondary metabolites which serve a wide array of functions [75]. Primary metabolites are necessary for cellular processes and include amino acids, simple sugars, nucleic acids, and lipids. Plant secondary metabolites are a diverse group of molecules that are involved in the adaptation of plants to their environment but are not part of the primary biochemical pathways of cell growth and reproduction.They include compounds produced in response to stress, such as the case when acting as a deterrent against herbivores [27]. They include aromatic compounds, steroids, terpenoids, alkaloids, flavonoids etc [28].

In old times, plant secondary metabolism had been neglected in science. Secondary metabolites were generally thought to be waste products of plants, without apparent function. Gradually, recognition of the important role of plant secondary metabolites has increased with the understanding of their functions in plants. They have a key role in protecting plants from herbivores and microbial infection, as attractants for pollinators and seed-dispersing animals, as allelopathic agents, UV protectants and signal molecules in the formation of nitrogen-fixing root nodules in legumes [75, 76].

2.5. PLANT SECONDARY METABOLITES AS DRUGS

Plant kingdom has been an important potential source of drugs since ancient times. This started with direct use of medicinal plants for the treatment of diseases and some drugs in Western medicine are based on these traditional drugs. Some of them are used as pure compounds such as atropine, morphine and quinine and some are modified, such as aspirin and local anaesthetics. In total, approximately half of the

world's 25 best-selling pharmaceutical agents are natural products [77, 78]. It shows that the hit rate of natural products is much higher than the libraries derived randomly.

Secondary metabolites previously known as waste products are now attracting attention as some appear to have a key role in protecting plants [28]. The bioactivities in plants are generally ascribed to the presence of plant secondary metabolites which could have beneficial or adverse effects [79]. The interest in plant secondary metabolites is getting increased due to their various roles in plants. The investigations in chemodiversity and roles of secondary metabolites in plants will contribute to drug discovery processes.

Plant compounds exhibit very large structural diversity. However, only a small proportion of that diversity has been seriously explored for its pharmacological potential so far, and there is a an important potential of new classes to be explored for their pharmacological properties [29].

One of the main problems in drug discovery from plant small molecules is the identification of their molecular targets: many compounds have been found to be more promiscuous than originally anticipated, which can potentially lead to side effects, but which may also open up additional medical uses. The drug poly-pharmacological activity can be understood only if its interactions with cellular components are comprehensively characterized. [80].

Thus the identification of target proteins and investigation of ligand-receptor interactions represents an essential step in the process of plant drug discovery and development [81, 82].

Information on moleculer targets is also useful for the elimination of libraries in high-throughput screening which has been an important tool in drug discovery. It

provides focusing on the certain groups instead of screening of all libraries. This is an important point for decreasing the computational cost and doing more investigations for a certain group.

Information on molecular targets of plant compounds has however also increased over the past few years. In one of these studies, two databases CHCD and BPCD were constructed in order to provide information of molecular structure data, botanic sources, specific molecular targets and therapeutic categories of bioactive plant compounds. The first covers details on chemical constituents of many of the major herbs used in Traditional Chinese Medicine (TCM), the other provides information on phytochemical compounds with known activity against a wide variety of targets, among them many of proven or suspected therapeutic significance. This is of particular significance for virtual screening in that the information can be used to identify other phytochemicals which may be expected to show similar behavior and affords the first opportunity to map the ligand-receptor space of plant compounds and their respective molecular targets [29].

2.6.   PLANT SECONDARY METABOLITES CLASSES

Based on their biosynthetic origins, plant secondary metabolites can be divided into four major groups: (i) Phenylpropanoids and related compounds (lignans, phenolic and polyphenolic compounds) (ii) polyketides (quinones, anthroquinones etc.) (iii) terpenoids (iv) alkaloids [12].

The general classification of plant secondary metabolites [12]:

1-) **Phenylpropanoids and related compounds**

    a) Monolignols

    b) Lignans

    c) Coumarins

    d) Flavonoids

        1) Chalcone

        2) Dihydrochalcone

        3) Aurone

        4) Flavone

        5) Isoflavone

        6) Flavonol

        7) Flavanone

        8) Flavan-3-ol

        9) Flavan-3,4-diol

        10) Dihydrofavanol

        11) Anthocyanidins

    e) Stilbenoids

        1) Stilbenes

        2) Bibenzyls

    f) Hydrolysable tannins

2-) Polyketides

    Quinones

        a) Anthroquinones

        b) Xanthones

3-) Terpenoids

    a) Hemiterpenoids

    b) Monoterpenoids

    c) Sesquiterpenoids

       d) Diterpenoids

       e) Triterpeoids

       f) Tetraterpenoids

       g) Polyterpenoids

4-) Alkaloids

       a) Tropane alkaloids

       b) Pyrrolizidine alkaloids

       c) Piperidine alkaloids

       d) Quinolizidine alkaloids

       e) Pyridine alkaloids

       f) Isoquinoline alkaloids

       g) Indole alkaloids

       h) Purin alkaloids

       i) Acridone alkaloids

       j) Quinoline Alkaloids

       k) Indolizine Alkaloids

       l) Stereoidal alkaloids

       m) Terpenoidal alkaloids

2.6.1. LIGNANS

Lignans are a class of secondary metabolites produced by oxidative dimerization of two phenylpropanoids through a C8-C8 linkage (Figure 10) [83] [84]. Lignans are widely distributed in the plant kingdom and have been found in various organs, including roots, rhizomes, stems, leaves, seeds and fruits in species belonging to more than 70 families [85]. Some lignans are used as medicines and dietary supplements because of their anticancer and antioxidative properties [86].
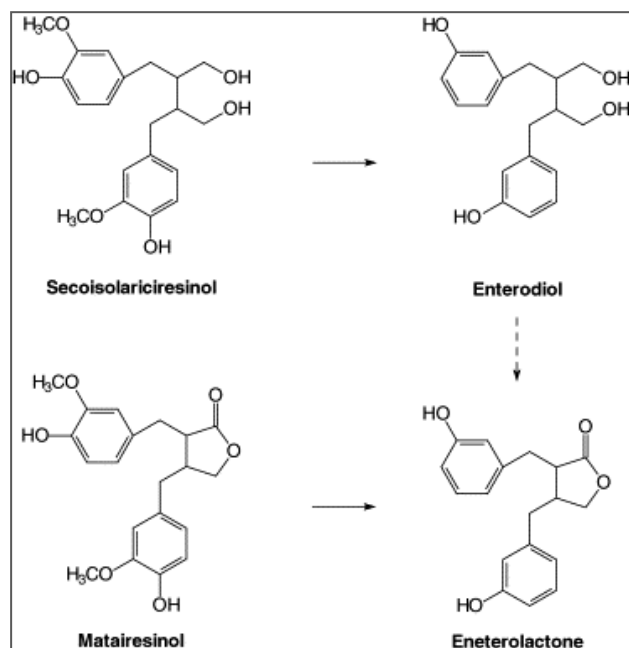
Figure 10 Structure of plant lignans—secoisolariciresinol and matairesinol—and mammalian lignans—enterolactone and enterodiol [84]

2.6.2. PHENOLIC COMPOUNDS

Phenolics are characterized by having at least one aromatic ring with one or more hydroxyl groups attached. In excess of 8000 phenolic structures have been reported and they are widely dispersed throughout the plant kingdom [28]. Phenolics range from simple, low molecular-weight, single aromatic-ringed compounds to large and complex tannins and derived polyphenols. They can be classified based on the number and arrangement of their carbon atoms and are commonly found conjugated to sugars and organic acids [28].

2.6.2.1.   FLAVONOIDS

Flavonoids are polyphenolic compounds which have fifteen carbons, with two aromatic rings connected by a three-carbon bridge in their basic structure. They are the most numerous of the phenolics and are found throughout the plant kingdom [87]. They are frequently found in leaves and the fruits of the plants. They take diverse and important roles as secondary metaboltes in plants such as UV protection, pigmentation, stimulation of nitrogen-fixing [88].

The main subclasses of flavonoids are the flavones, flavonols, flavan-3-ols, isoflavones, Flavanones and anthocyanidins (Figure 11). There are other flavonoids which found rather in less quantities are dihydroflavonols, flavan-3,4-diols, coumarins, chalcones, dihydrochalcones and aurones [28].
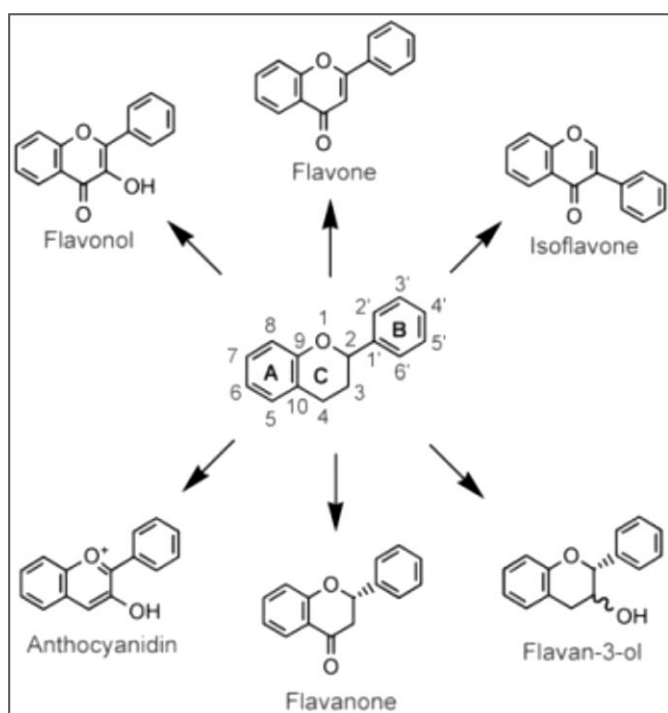
Figure 11 Generic structures of the major flavonoids [28]

The basic flavonoid skeleton can have many substituents such as hydroxyl groups which are usually present at the 4, 5 and 7 positions and sugars usually found as glycosides. Sugars and hydroxyl groups increase the water solubility of flavonoids, other substituents, such as methyl groups and isopentyl units, make flavonoids lipophilic [28].

2.6.2.2. STILBENOIDS

Members of the stilbene family which have the C6–C2–C6 structure, like flavonoids, are polyphenolic compounds. Stilbenes are phytoalexins, compounds produced by plants in response to attack by fungal, bacterial and viral pathogens. Resveratrol is the most common stilbene. It occurs as both the *cis* and the *trans* isomers and is present in plant tissues primarily as *trans*-resveratrol-3-*O*-glucoside which is known as piceid and polydatin (Figure 12). A family of resveratrol polymers, viniferins, also exists. The major dietary sources of stilbenes include grapes, wine, soya and peanut products [28].
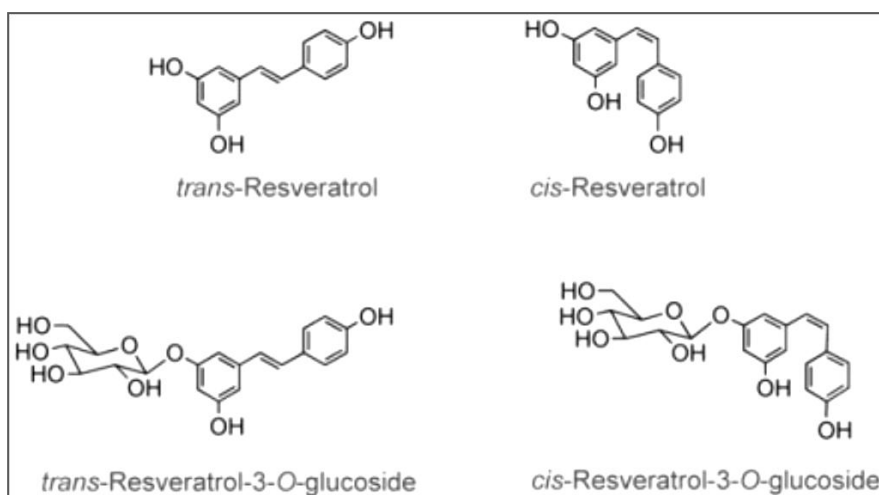
Figure 12 Structures of the stilbenoids trans- and cis-resveratrol and their glucosides [28]

2.6.3. TERPENOIDS

Terpenoids (isoprenoids) encompass more than 40 000 structures and form the largest class of all known plant metabolites. They are found as flavours and fragrances, antibiotics, plant hormones, membrane lipids, insect attractants and antifeedants, and mediators of the essential electron-transport [89].

Owing to their diverse biological activities and their diverse physical and chemical properties, terpenoid plant chemicals have been exploited by humans as traditional biomaterials since ancient times [90].

Basic unit of most secondary plant metabolites, including terpenes, consists of isoprene, a simple hydrocarbon molecule. The term terpene usually refers to a hydrocarbon molecule while terpenoid refers to a terpene that has been modified, such as by the addition of oxygen. The isoprene unit, is a five-carbon molecule. An isoprene unit bonded with a second isoprene is the defining characteristic of terpene, which is also a monoterpene (C10). While sesquiterpenes contain three isoprene units (C15), diterpenes (C20) and triterpenes (C30) contain two and three terpene units, respectively (Figure 13). Tetraterpenes consist of four terpene units and polyterpenes are those terpenes containing more than four terpene units [91].
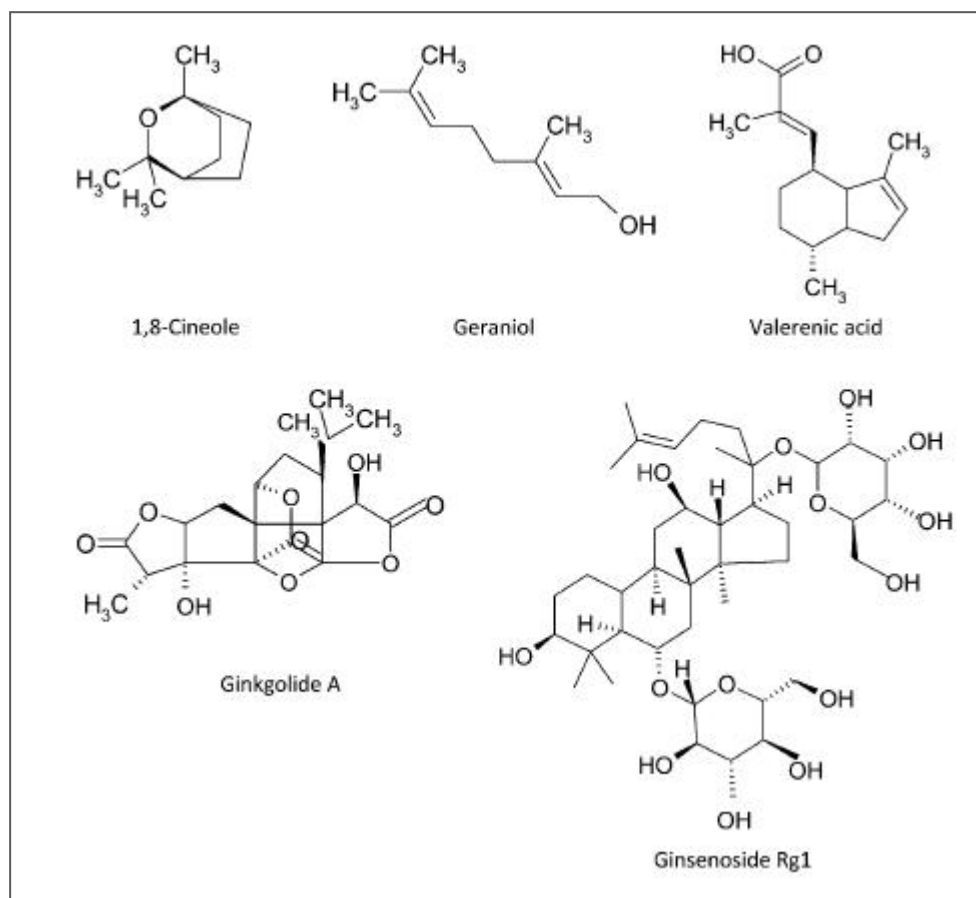
Figure 13 Structures of selected terpenes, including the monoterpenes 1,8-cineole and geraniol, the sesquiterpene, valerenic acid, the diterpene, ginkolide A, and the triterpene, ginsenoside [92]

Because of their many different structures, plant terpenoids as a group include compounds with many different physical and chemical properties. They may be lipophilic or hydrophilic, volatile or non-volatile, cyclic or acyclic, chiral or achiral. The chemical diversity of plant terpenoids originates from often complex terpenoid biosynthetic pathways [90].

2.6.4. ALKALOIDS

Alkaloids are a structurally diverse group of over 12,000 cyclic nitrogen-containing compounds that are found in over 20% of plant species [92]. The recorded use of alkaloids for medicinal purposes stretches back 5000 y and this chemical group has contributed the majority of the poisons, neurotoxins, and traditional psychedelics (e.g. atropine, scopolamine, and hyoscyamine, from the plant Atropa Belladonna) and social drugs [e.g. nicotine, caffeine, methamphetamine (ephedrine), cocaine, and opiates] consumed by humans (Figure 14) [93].

Alkaloids can be classified on the basis of biological activities, chemical structure and biosynthetic pathways. Following are few of major classes of alkaloids; (i) Pyrrolidine: is an organic compound with the molecular formula C_4H_9N; (ii) Quinoline: is a heterocyclic aromatic organic compound with formula C_9H_7N; (iii) Benzylisoquinoline: is a heterocyclic aromatic organic compound & structural isomer of quinoline; (iv) Indole: is an aromatic heterocyclic organic compound (v) Terpenoid: a diverse class of naturally occurring organic compounds.
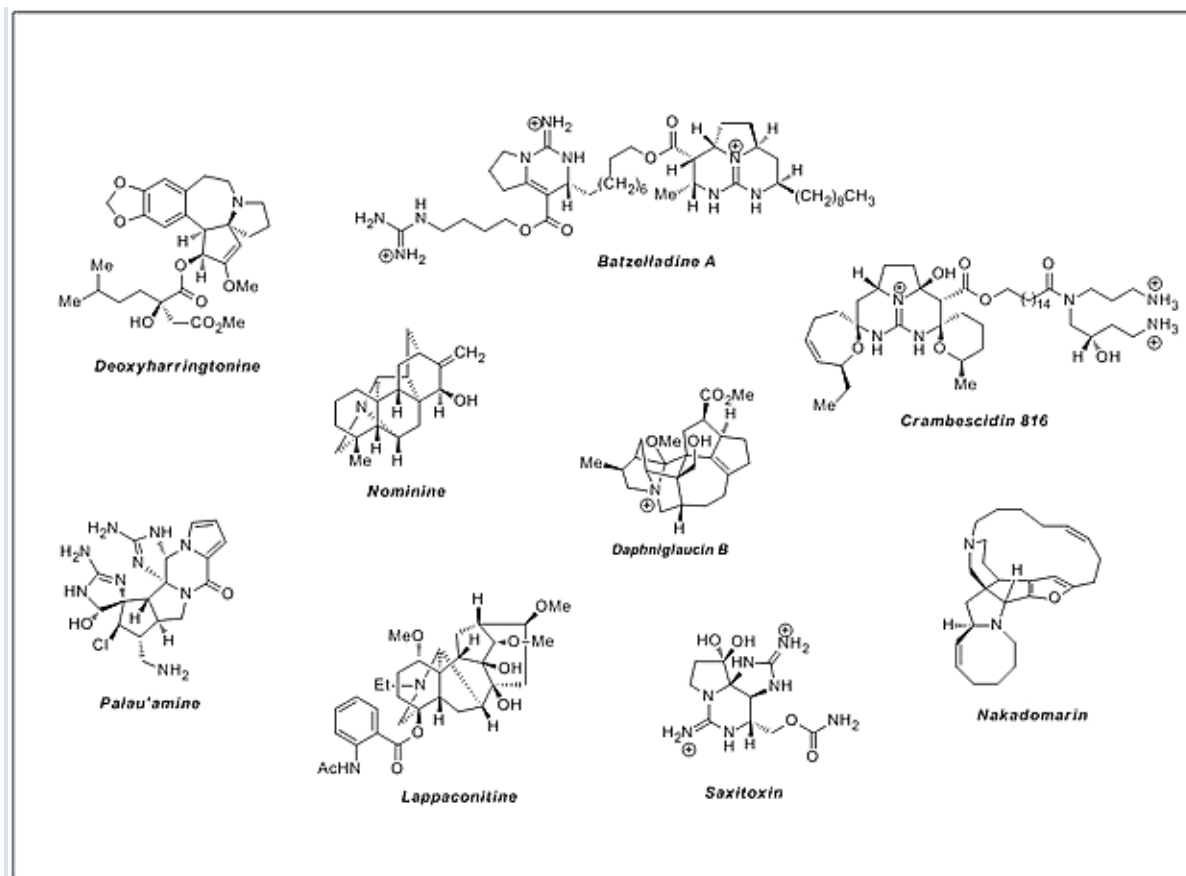
Figure 14 Structures of selected alkaloids [93]

2.7.   DOMAINS

One challenging therapeutic strategy that continues to attract attention is the potential to interfere with protein– protein interactions, which is of great importance for both understanding cell physiology and for developing novel treatments against disease. A number of reviews have recently described the structural and chemical intricacies involved in developing small molecules that modulate protein–protein interactions [94] [95]. Challenges to overcome include the detailed structural characterization of binding pockets involved in protein–protein interactions, and the acquisition of further chemical diversity within compound collections and improved accessibility to chemical building blocks, templates and early drug-like hits for this type of target [94, 95].

In this study we search for small molecule inhibitors for PDZ, BROMO and SPRY domains which have critical roles in various important protein-protein interactions.

2.7.1. PDZ DOMAIN

The PDZ domain is one of the most common protein−protein interaction domains in humans, and it is found in all kingdoms of life [44, 49, 96]. In the mouse genome, for example, 928 PDZ domains have been recognized in 328 proteins, which exist in single or multiple copies or in combination with other interaction modules (Figure 15). The types of PDZ proteins, their roles in cell function and the diversity of PDZ interactions have been extensively reviewed [34, 97]. From the abundance and diversity of PDZ domains in cells it is apparent that many cellular biological functions are affected by PDZ-mediated interactions [34, 39, 44, 48, 96, 98].
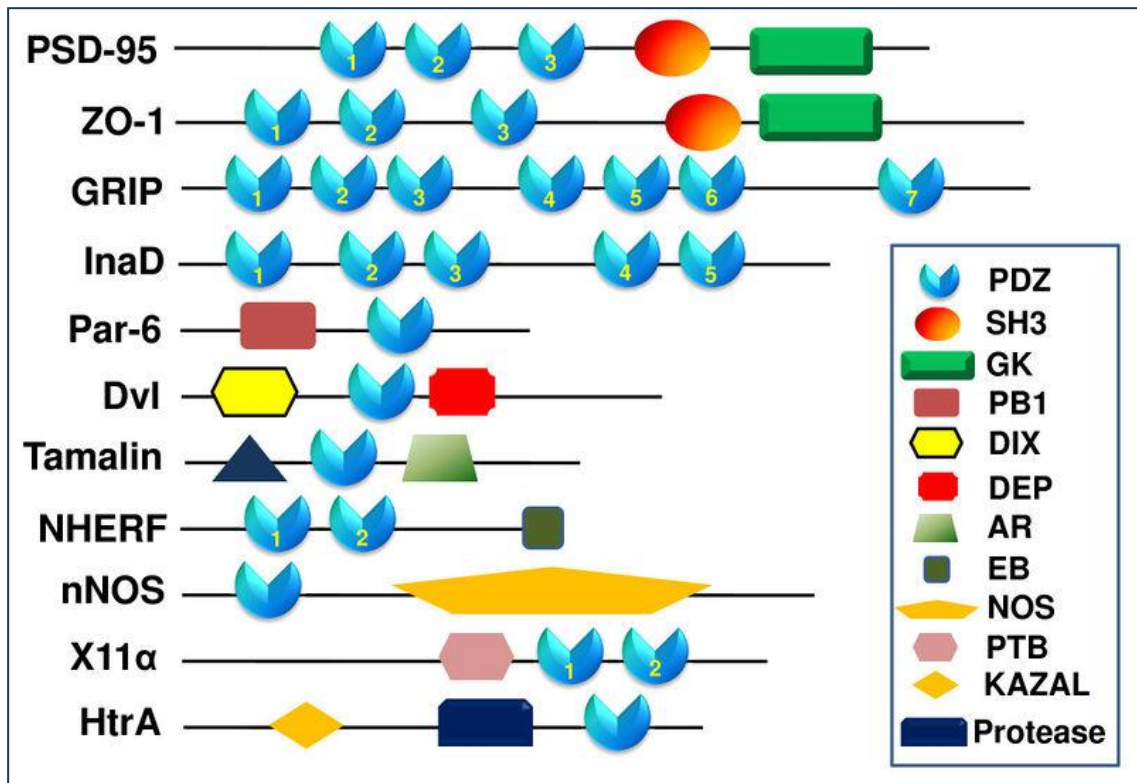
Figure 15 Examples of PDZ domain-containing proteins. Proteins are indicated by black lines scaled to the length of the primary sequence of the protein [44]

PDZ domains were first identified in the neuronexpressed postsynaptic density protein (PSD95), its Drosophila homologue discs large tumour suppressor (DlgA) gene product and zonula occludens-1 (ZO1), a tight-junction protein [34].

PDZ domains are usually 80-100 amino acid residues long and adopt a similar topology and specialized for binding of C-termini in partner proteins, generally transmembrane receptors and channel proteins and other PDZ domains [48]. Structural studies show that PDZ domains are usually composed of 6 β-strands (βA ~ βF), a short α-helix (αA) and a long α-helix (αB) (Figure 16). The N- and C-termini of canonical PDZ domains are in proximity to each other on the opposite side from the peptide-binding site in a groove between the αB-helix and βB-strand structures serving as an additional antiparallel strand within the PDZ domain (Figure 17) [44]
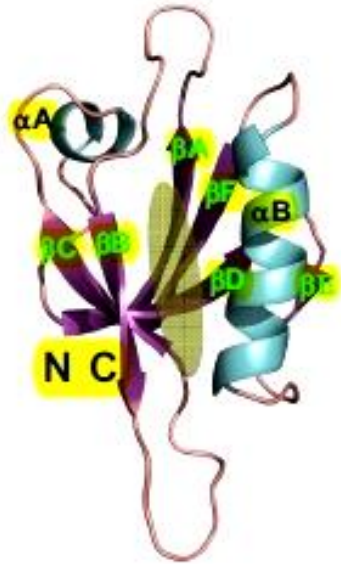
Figure 16 Structure of PDZ domain: Ribbon diagram of Dvl-1 PDZ (PDB code: 2KAW) [48]
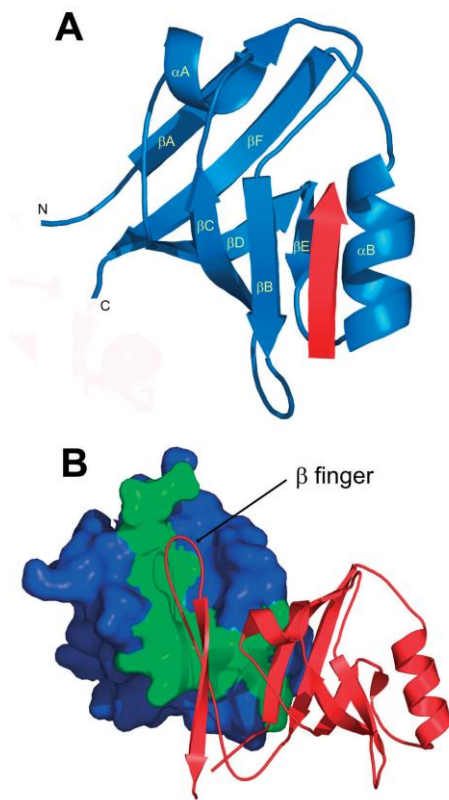
Figure 17 Structure of the PDZ domain bound to peptide and internal peptide motif. A. Ribbon representation of the third PDZ domain of PSD95 (blue) with KQTSV peptide (red arrow) (PDB code1be9 B. Complex of the syntrophin PDZ domain and nNOS PDZ domain (PDB code 1 qav) made by PyMOL [48]

There are studies about mutations on Val76 and Ile79 positions of PDZ domains showing the importance of the residues in protein binding [47]. In a study, reported by Tonikian et al, who have shown that binding preferences at these positions, these residues are determined: 23, 26, 28, 48, 49, 51, 79, 83 [99]. In further study positions are restricted to four ligand positions 0 (ligand C-terminal), -1 ,-2, and -3 [100]. In addition, the protein binding residues of first PDZ domain of Nherf are inidicated as Gly23, Tyr24, Gly25, Phe26, Leu28, Val75, Val76, Ile79 and Arg80 in BLAST database [50].

Several proteins that contain PDZ domains allow PDZ interactions to form complex protein-interaction. These interactions control targeting, clustering, cycling and membrane expression of receptors, transporters and ion channels. PDZ proteins are also essential in linking various components of intracellular signalling pathways and Networks. PDZ proteins have provisional roles in many disease states by regulating disease-associated proteins [34, 44, 97] [101].

There are some obvious advantages of discovering drugs against PDZ interactions. Blockade of PDZ interactions is a novel way to regulate signalling pathways in complex diseases. Targetting more than one disease-related protein simultaneously could be possible by targeting a single type of protein interaction. Thus, it might be more efficient than targetting a single type of protein. For example, by using a drug that inhibits the PDZ interaction between PTEN and MAG13 (two proteins that regulate the oncoprotein protein kinase Akt/PKB), it might be possible to regulate the role of Akt/PKB in cancer [102]. In other case, PSD95 which contains PDZ domain can enhance or depress synaptic strength depending on the frequency of neuronal firing [103]. The use-dependent modulation of particular protein pathways could therefore be another potential advantage of regulating PDZ interactions [34, 97].

Many PDZ containing proteins take roles in protein-protein interactions involved in some cancers. In a study it is suggested that AIPC which encodes a high molecular weight protein containing six PDZ domains is up-regulated in human prostate tumors [101].

One other study showed up-regulation of a PDZ domain-containing protein, PCD1, in many cancers, including malignant prostate tissues. PCD1, a novel gene containing PDZ and LIM domains, is overexpressed in human breast cancer and linked to lymph node metastasis [37].

In another study a series of cell-permeable, side chain-modified lipopeptides were modified that target the GIPC PDZ domain in vitro and in vivo. These peptides exhibit significant activity against pancreatic and breast cancers, both in cellular and animal models [104].

Similarly, PDZ domain mediated interactions have been implicated in several diseases, these interactions therefore represent a target for future drug discovery [34]. Because the potential benefits of making drugs that inhibit PDZ interactions remain hazy, this drug target still awaits the full attention of the drug discovery sector for the moment.

In this study we aimed to identify a specific secondary metabolite class as candidate inhibitors for PDZ domain. We target the PDZ1 domain of the Na(+)/H(+) exchanger regulatory factor (NHERF) which binds with nanomolar affinity to the carboxyl-terminal sequence QDTRL of the cystic fibrosis transmembrane conductance regulator (CFTR) and plays a central role in the cellular localization and physiological regulation of this chloride channel. We used the crystal structure of human NHERF PDZ1 bound to the carboxyl-terminal peptide QDTRL has been determined at 1.7-A resolution (PDB code:1I92) [105]. The structure reveals the specificity and affinity determinants of the PDZ1-CFTR interaction and provides insights into carboxyl-terminal leucine recognition by class I PDZ domains. We used the binding pocket of NHERF-PDZ1 domain that CFTR peptide inserts and identified the active residues in binding pocket with the insights from previous studies about binding specifities of the PDZ domains [34, 44, 97] [47, 99] [100].

2.7.2. BROMO DOMAIN

Bromdomains (BRDs) are the only known protein recognition module that selectively targets ε-N-acetylation of lysines. The human proteome encodes over 40 proteins that contain more than 60 diverse BRDs. Sixty-one unique bromodomains have been identified from the human genome [106] each containing a conserved tertiary structure. This tertiary structure is an "atypical left-handed four-helix bundle", with the hydrophobic KAc binding site at one end formed between the Z′ short helix, the ZA loop, and the BC loop (Figure 18A). This binding pocket is primarily hydrophobic, with the carbonyl oxygen of the acetyl-group forming two hydrogen bonds, one to a donor from either asparagine or threonine and the other to a conserved water molecule at the base of the pocket (Figure 18B,C) [33].
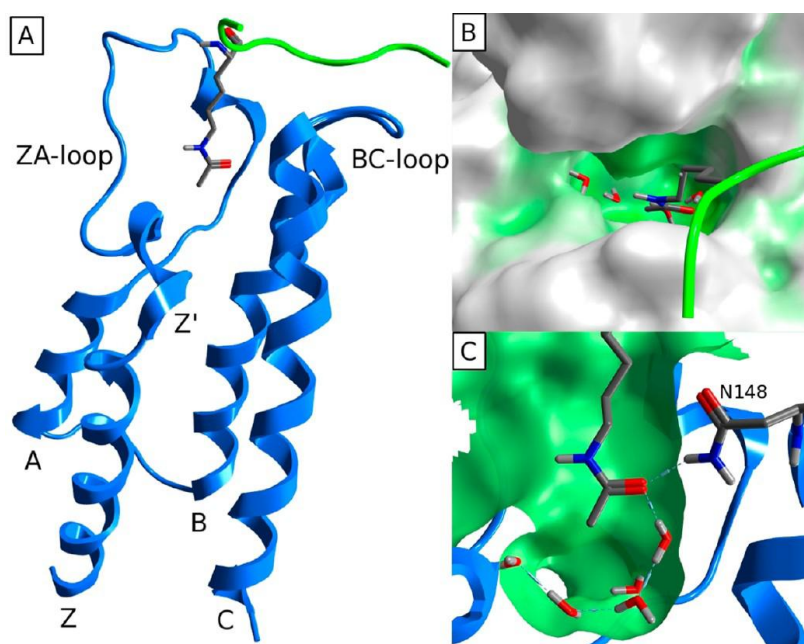
Figure 18 (A) Conserved protein fold of bromodomains comprising the four canonical helices αZ, αA, αB, and αC. (B) Surface representation of a typical KAc binding site. (C) Typical binding of KAc to bromodomain. All illustrated by FALZ (PDB 3QZS) [33]

Enzymes that 'write' (histone acetyltransferases, HATs) and 'erase' (histone deacetylases, HDACs) acetylation sites are an area of extensive research in current drug development. The principal readers of ε-N-acetyl lysine (K_{ac}) marks are bromodomains (BRDs), which are a diverse family of evolutionary conserved protein-interaction modules. The deep, largely hydrophobic acetyl lysine binding site of BRD domain is an attractive pocket for the development of small molecule inhibitors [107].

Bromodomain inhibitors are attracting interest as potential therapeutics in multiple disease areas [107]. RNA screen suggested inhibition of the BET family as a therapeutic strategy for AML [108]. Pan-BET family inhibitor GSK1210151A from the isoxazole class was discovered and inhibition of the BET family was suggested as therapeutic strategy for MLL-fusion leukemia. Pan-BET family inhibitor GSK525762A, from the benzodiazepine class, has been a potential drug in mouse models of inflammatory disease and sepsis [109].

Many proteins that contain bromodomains in their specific regulation complexes have been reported in the development of cancer. Though these diseases have not been directly related to the bromodomains, the correlation of overexpression of bromodomain proteins with the patient survival and dominant oncogenic rearrangements provide a strong claim for the role of bromodomains in cancer [107].

A recent study showed that ATAD2 is overexpressed in more than 70% of breast tumours and that higher protein levels correlate with tumour histological grades, poor overall survival and disease recurrence[110].

Aberrant expression has also been reported for TRIM24 in breast cancer, and high expression levels have been shown to negatively correlate with survival of breast cancer patients [111]. In liver, however, TRIM24 seems to function as a liver-specific tumour suppressor [112].

The testis-specific BET family member BRDT is frequently overexpressed in non-small-cell lung cancer and several other cancers [113].

Therapeutic potential of bromodomains is becoming increasingly recognized in both biology and drug design. Due to availability of many bromodomains structures published, it is possible to investigate the structure-based druggability across the protein family.

Targeting BRDs for the development of protein-interaction inhibitors has recently emerged as a strategy for the design of pharmacologically active reagents. The relatively weak interaction of BRDs with their substrates, the diversity and physicochemical properties of the acetyl lysine binding site, and the large number of available crystal structures will facilitate the rational design of such inhibitors.

In this study we targeted the bromodomain of histone acetyltransferase gcn5p protein. We identified the binding pocket and the critical binding residues with the insights of ligand-protein interactions of bromodomain- acetylated histone H4 complex [105] and the previous studies about bindng specifites of bromodomain. Studies show that bromodomain binds to acetyllysines through specific hydrophobic residues in alpha helices. In PCAF, Tyr-809 is shown to be the most critical residue for the interaction, although Tyr-802 also contributes to the binding. Tyr-802 in PCAF corresponds to Tyr406 in Gcn5p which includes our bromodomain, Tyr-139 and Tyr-433 in two bromo domains of Brd4, indicating conservation of Tyr in this position. Tyr-809 of PCAF corresponds to Tyr413 in Gcn5p, Ile146 and Val440 in Brd4, indicating substitutions in this position [114]. So Tyr406 and Tyr413 seem to critical and conserved residues in bromo domains which take important role in ligand binding. The protein binding residues of BROMO domain of Gcn5P are also indicated as Asp366, Pro371, Leu374, Tyr413, Ala417 and Phe423 in BLAST database [50].

2.7.3. B30.2/SPRY DOMAIN

B30.2 and SPRY domains share sequence similarity such that a SPRY domain can be identified in virtually all the known B30.2 domain-containing proteins in the sequence-based domain classification databases Pfam [115]and SMART [116]. Notably, the SPRY domains are shorter at the N-teminus than the B30.2 domains and appear as a subdomain of the latter. Out of 642 SPRY domain-containing proteins classified in the SMART database, 516 possess a PRY domain (50–60 residues) right next to the N-terminus of the SPRY domain. The functional relationship between the PRY and SPRY domains has been unknown. The B30.2/SPRY (for B30.2 and/or SPRY) domain is present in a large number of proteins with diverse individual functions in different biological processes [117]. The functional significance of the B30.2/SPRY domains is clear from numerous biochemical or genetic studies [36].

Pyrin/Marenostrin/MEFV is a B30.2 containing protein. Missense mutations in Pyrin/Marenostrin/MEFV are the cause of familial Mediterranean fever (FMF), which is characterized by recurrent episodes of fever and serosal inflammation [36, 118]. FMF is considered to be an autoinflammatory syndrome due to the lack of any obvious stimuli and the involvement of autoreactive T cells. Autoinflammatory syndromes are caused by a dysfunction of the innate immune system. Therefore, it is assumed that pyrin plays a role in the innate immune system[43]. Human Pyrin contains an N-terminal Pyrin domain, a B-box, a coiled-coil, and a C-terminal B30.2/SPRY domain [117]. Most FMF-causing mutations are located in the SPRY domain [36].

In a study, The 191-amino-acid polypeptide chain of human pyrin expressed with clear electron density folds into a compact β-barrel domain (Figure 19a). The β-barrel consists of two antiparallel β-sheets that are connected in a jelly-roll topology. These

six- and seven-stranded β-sheets possess pronounced left-handed twists and build up the hydrophobic core of the B30.2 domain. Besides some short helical loops, there is just one α-helix. This helix packs against the seven-stranded β-sheet and runs perpendicular to the directions of the β-strands (Figure 19a) [43].
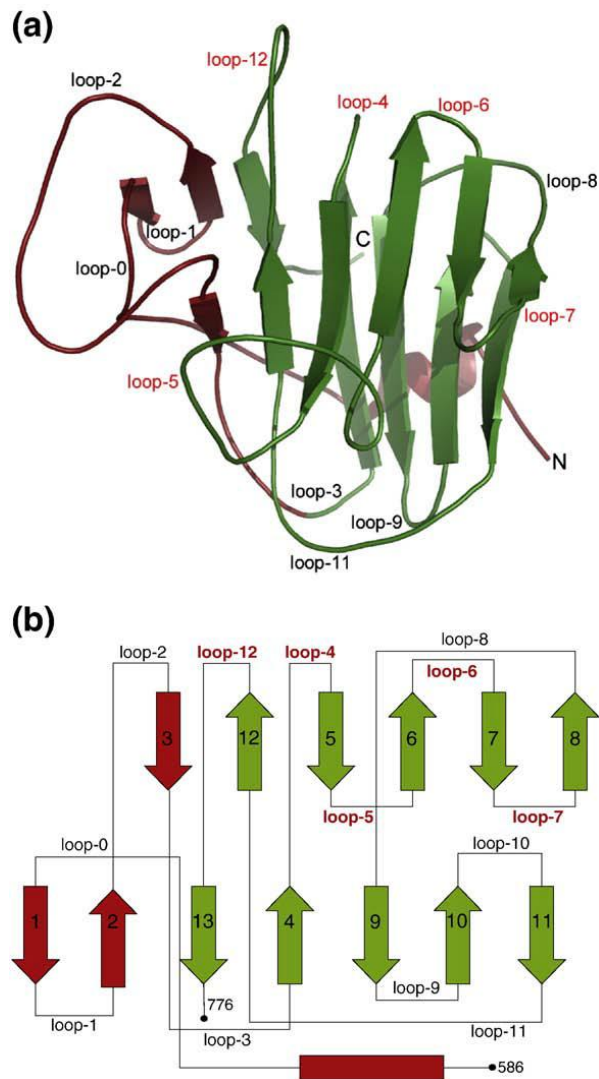
Figure 19 (a) Ribbon diagram of the pyrin B30.2 domain showing the Pry and Spry subdomains in red and green, respectively. (b)Topology diagram of the B30.2 fold [43]

Based on sequence profile methods, B30.2 domains were subdivided into N-terminal Pry and C-terminal Spry subdomains. The Spry subdomain contains two five-stranded β-sheets in a sandwich-like arrangement, and each of these sheetsis extended by additional β-strands from the Pry subdomain (Figure 19b). The Pry subdomain serves as a N-terminal lid to shield the hydrophobic core of the larger Spry subdomain. Residues involved in loops 0, 2, 4, 5, 6, 7, and 12 and β-strands 5, 6, 7, and 12 create a shallow cavity on the surface of the pyrin B30.2 domain. The bottom of this cavity is covered entirely by hydrophobic amino acids [43].

Most mutations that are associated with FMF cluster in the B30.2 domain of pyrin. All the known mutations in the B30.2/SPRY domain in human Pyrin were mapped on the structure of GUS (Figure 20B). They are the amino-acid changes at the positions of Arg653, Ser675, Met680, Thr681, Ile692, Met694, Lys695, Val726, Ala744, or Arg761. Of these, Met680, Met694, and Val726 are most commonly mutated in FMF patients [118].

In another study, mutations of 28 positions have been reported. Mutations that are associated with FMF are observed in three distinct areas of the pyrin B30.2 domain. Seventeen mutations (Arg761, Asp661, Lys695, Ile640, Met694, Ile692, Thr681, Val704, Met680, Ser702, Tyr688) cluster either directly inside or close to the central cavity, three mutations are located in the hydrophobic core, and eight mutations modify the molecular surface on the opposite side of the central cavity (Figure 21 a and b) [43].
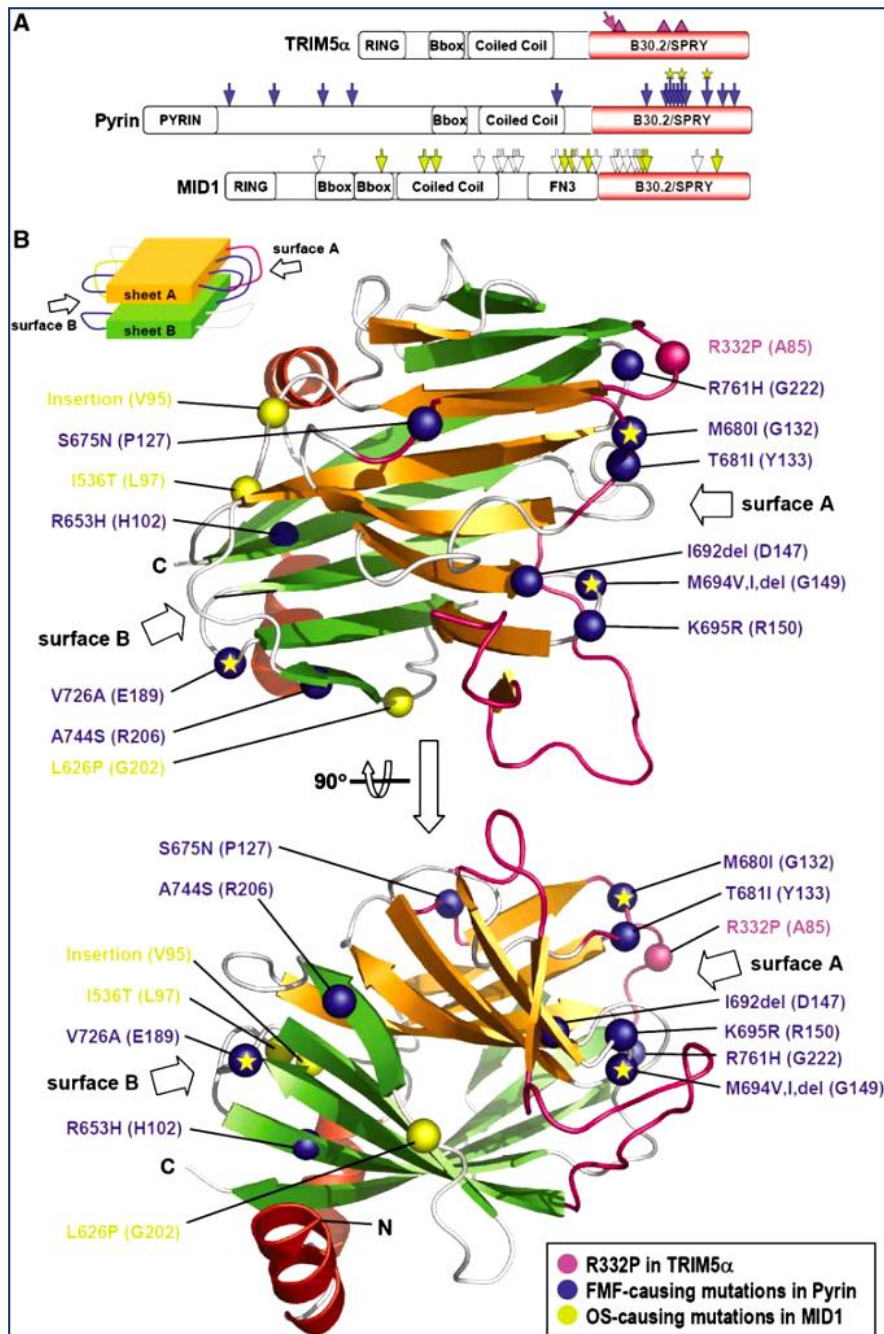
Figure 20 Sequence mutations in TRIM5a, Pyrin, and MID1. Mutations in FMF are indicated in purple [36]
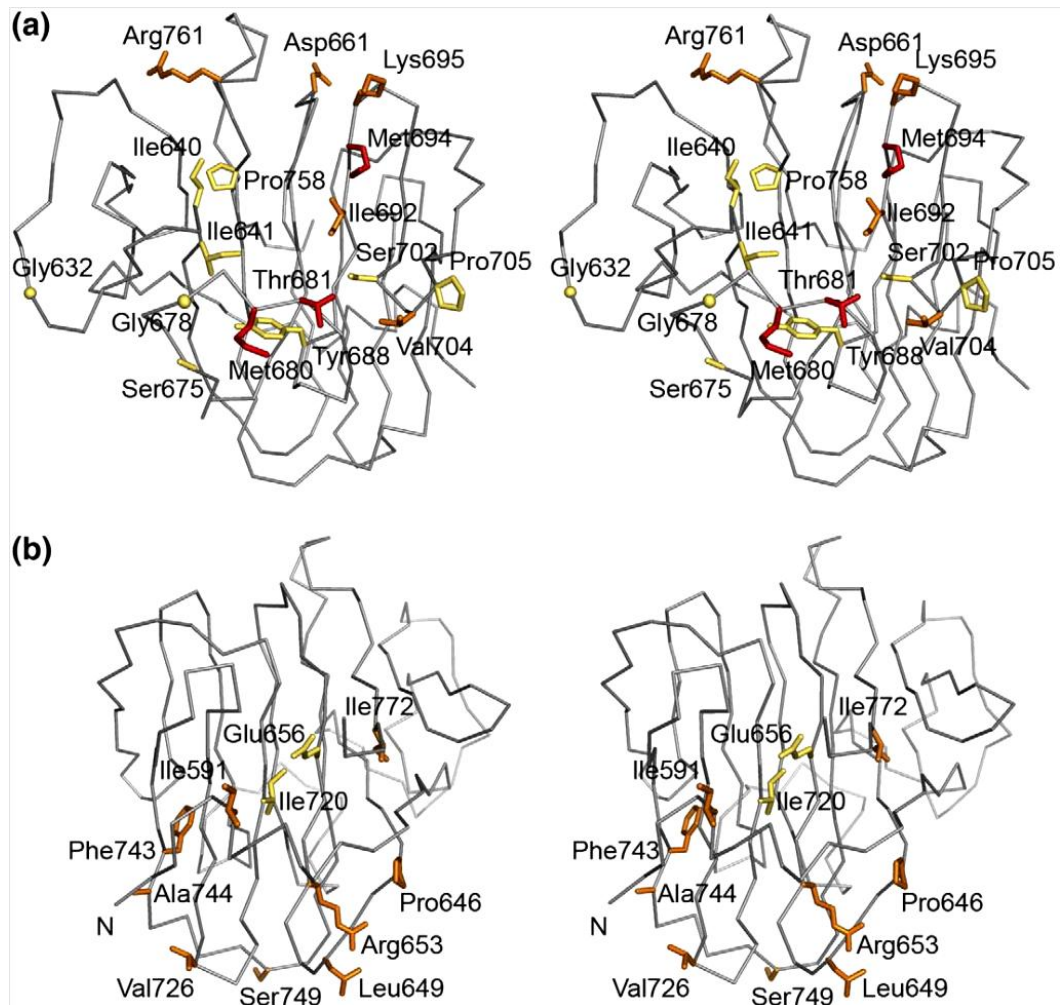
Figure 21 Mutations of pyrin B30.2 that are associated with FMF (a) Seventeen mutations map to the central cavity. (b) Eleven mutations map to the hydrophobic core and the N-terminal surface [43].

This analysis reveals that most mutations that are associated with FMF cluster around a putative peptide binding site. Since this binding site is blocked neither by ligands nor by crystal contacts, the described crystal structure will be of great value for probing the pyrin B30.2 peptide binding activity by fragment-based screening [43].

Some mutations on residues of SPRY domain that cause Mediterranean fever disease indicated as Val704 Ile, Ile640Met, Lys695Arg, Lys695Met, Met694Ile in DMDM (Domain mapping of disease mutations) database [45].

In this study we aimed to identify a specific secondary metabolite class of molecules targetting PyrinSpry domain. We used x-ray crystal structure of human pyrin protein (PDB code: 2WL1) for this study. According to previous studies that identify the hot spots of the domain covered above, we determined the residues in the ligand binding pocket. As a common result in these studies, the residues Arg653, Ser675, Met680, Thr681, Ile692, Met694, Lys695, Val726, Ala744, or Arg761 whose mutations lead to FMF are indicated as the ligand binding residues of PyrinSpry domain. The most common mutations are indicated as Met680, Met694, Val726 in the central cavity. We determined the central cavity as binding pocket including residues Met680 and Met694 for structure based screening studies.

Chapter 3

3.   METHODS

3.1.   MOLECULAR MODELLING AND DOCKING

3.1.1. MOLECULAR MODELLING AND VISUALIZATION [119, 120]

The structures of proteins, ligands and protein-ligand complexes were visualized and analyzed with a molecular modeling and visualization software Accelrys Discovery Studio [119] to determine the binding pockets and the active side residues of the proteins. The .mol2 and .pdb molecular structural file formats used in this study were also prepared by Discovery Studio.

We used Autodock MGL Tools [120] for determination of the size and the center coordinates of the grid-box which is required information to prepare conformation files for docking by VINA. The .pdbqt molecular structural file formats files which are required format for the docking software, VINA were also prepared by Autodock MGL Tools.

3.1.2. CONVERSION OF STRUCTURAL FILE FORMATS [121]

Open Babel 2.3.1 [121] was used to make conversion between various moleculer structural file formats. Preparation, combination and splitting of .smiles, .smi, and .mol file formats used in this study were done with Open Babel 2.3.1.

3.1.3. MOLECULER DOCKING [23]

We used Autodock VINA [23] docking software for moleculer docking studies. Autodock VINA calculates the free energy of binding by subtracting unbound ligands" and proteins" energy separately from the bound state of ligand-protein complex.

The energies were calculated by evaluating the intramolecular energetics of the transition from the unbound state to the bound conformation for both protein and ligand separately, and then by evaluating the intermolecular energetics of bringing the protein and ligand together into the bound complex. The formula is given below for the binding free energy calculation:

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

(1)

Here the first two terms indicate intramolecular energies of bound and unbound states of ligand while the following two terms indicate intramoleculer energies of protein for bound and unbound states. The change in intermolecular energy was calculated in the third parentheses where ΔSconf is the entropy loss upon binding. In unbound state, since two molecules are distant from each other, $V^{(P-L)}$ bound was accepted as zero. The pair-wise atomic terms (indicated as V) include evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation.

While docking, Gasteiger charges were added by Autodock MGL Tools. Polar hydrogens were added and unit atom approximation was applied. As a binding site, the most potential binding site was selected and taken as a grid centre. According to the maximum length and centre of the binding region measured by Discovery Studio, grid-

box size was determined by Autodock MGL Tools. Grid spacing between grid points was chosen as 0.375 Angstrom. The Lamarckian Genetic Algorithm was applied as the docking search parameter [122].

3.1.4. PHARMACOPHORE ANALYSIS [123]

The compounds were taken on the basis of higher scoring function for pharmacophore modeling. The protein-ligand complexes were analyzed and the key features in protein-ligand interaction were modeled by Ligand Scout [123]. Both docking results of KEGG Phytochemical compounds and derived compounds were analyzed and the important interactions were identified.

Ligand Scout 2.03 [123] is used for pharmacophore analysis. The software is used to analyze the protein-ligand interactions and model the pharmacophores. When .pdb file of a ligand-protein complex is given to Ligand Scout, the interactions such as hydrophobic interactions, negative ionizable areas, hydrogen bond donors and acceptors in the binding pockets can be observed.

3.2. PART 1

3.2.1. DECISION OF BINDING POCKETS

For structural based screening, important residues and active sites of domains were obtained from previous studies and analyzed by using the softwares Discovery Studio [119] and Autodock Tools [120]. The 3D coordinates of the crystal structure of pol γ (PDB ID: 3IKM) and pol β (PDB ID: 9ICA) were selected as the receptor models for pol γ and pol β, respectively for moleculer docking.

We identified the binding pocket in the proximal accessory subunit of pol γ including the residues Leu402, Val398, Leu448, Leu395, Leu394 and Glu449 (Figure 22). We identified these residues from previous studies as covered in detail in Literature Review part [21, 22, 26]. The accessory subunit and its binding pocket with the binding residues were visualized by Discovery Studio as shown in (Figure 22).

.
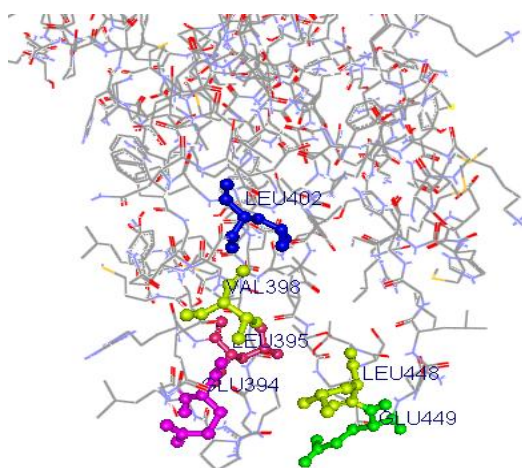
Figure 22 Active residues (colored) of proximal accessory subunit of pol γ interact with catalytic subunit. (The structure is visualized by Discovery Studio (PDB code: 3IKM)

We identified the residues Lys35, Lys60, Lys68, Lys72 (Figure 23) from previous studies as covered in detail in Literature Review part [19, 20, 25, 67] [18]. We identified the binding pocket of the lyase domain of DNA Polymerase β including the residues Lys35, Lys68, Lys72, Gly62, Gly66, Tyr39, Ile69, Met18 and Ala38. Lyase domain and its binding pocket with the binding residues were visualized by Discovery Studio as shown in (Figure 23).
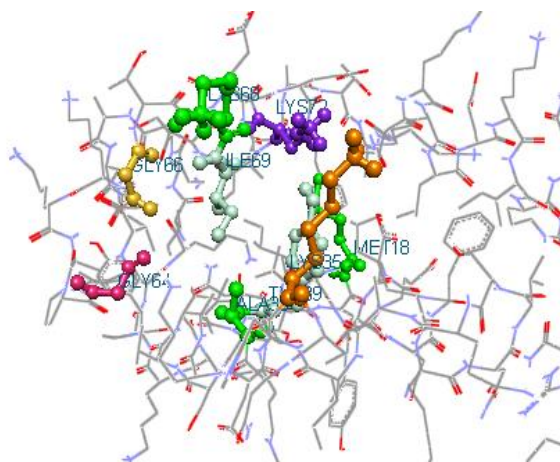
Figure 23 Binding pocket of lyase domain with active residues (coloured) of pol β. (The structure is visualized by Discovery Studio (PDB code: 9ICA)

3.2.2. MOLECULER DOCKINGS FOR DNA POLYMERASE γ AND β

We have targeted DNA pol β that involves in nuclear BER pathway, and DNA pol γ that involves in mitochondrial BER pathway using structure based drug design techniques. The hit molecules were searched by virtual screening of a Plant Based Compounds Library that we have constructed. The library is composed of about 8550 plant secondary metabolites that provide 3D structures that we obtained from various online available databases: KEGG Phytochemical Compounds [12], Analyticon-discovery MEGxp Pure Natural Plant Products [13], Seaweed Metabolites Database (SWMD) [14], Indofine Chemicals [15], Indian Plant Anticancer Compounds Database (InPACdb) [16]. These compounds are plant secondary metabolites which are extracted and derived from various plant species. Dockings of molecules were performed by using the docking soft ware Autodock Vina [23]. Autodock Vina calculated the Gibbs free energy of binding of molecules to the targets and the hit molecules were evaluated according to the these energies.

3.3. PART2

We performed the virtual screening of a part of plant based library which consists of plant secondary metabolites that we obtained from KEGG Phytochemical Compounds [12]. We performed docking of KEGG Phyochemicals for the PDZ, BROMO, and SPRY domains by using the docking software VINA. We analyzed the docking results of this set for the domains PDZ, BROMO, SPRY on the basis of higher binding scores and determined 5 promising classes for domains. We derived sets of compounds for each class adding fragments to the basic structures of the classes and performed docking of these sets for verification. We used a ligand modifying software BROOD for this purpose.

We analyzed the docking results of the derived compounds for domains PDZ, BROMO, and SPRY to identify the specific classes that have the best binding affinity to each domain. For this purpose, we compared the binding probabilities of each class to each domain obtained according to Gibbs distribution of the binding free energies of the compounds.

3.3.1. KEGG PHYTOCHEMICALS

We performed virtual screening of the part of plant based library which consists of plant secondary metabolites that we obtained from KEGG Phytochemical Compounds [12]. This set of molecules is classified according to their type of secondary metabolite such as phenylpropanoids (lignans, flavonoids, stilbenoids etc.), terpenoids, alkaloids and their subclasses. The set is composed of about 2650 plant secondary metabolites.

KEGG Phytochemical Compounds provides the electronic data of chemical structures in .mol structural file formats. We converted structure files into the form of .mol2 files using the visualization program Accelrys Discovery Studio [119]. Mol2 is generally accepted format for most of the docking softwares. KEGG Phytochemical Compounds are stored under the classes of secondary metabolite types with C codes as shown in the Figure 24. For the systematic calculations of classes we gave cluster codes to the compounds to represent the class that the compounds belong to (Table 1). We used these cluster codes instead of name to do calculations and analysis of the results based on the classes. We considered the classes shown in Table 2 which contain more than 5 compounds in our calculations.

```
D       C10788   Ellagic acid
D       C10836   3,4,3'-Tri-O-methylellagic acid
C    Miscellaneous galloyl derivatives
D       C10242   Terminalin
B   Styrylpyrones
C      C09925   5,6-Dehydrokawain
C      C09926   Dihydromethysticin
C      C09947   Kawain
C      C09952   Methysticin
C      C09980   Yangonin
A<b>Polyketides</b>
B   Quinones
C     Anthraquinones
D       Anthrone type
E          C10305   Barbaloin
E          C10307   Cascaroside A
E          C10314   Chrysophanic acid 9-anthrone
E          C10404   Sennoside A
D       Anthraquinone type
E          C01448   Questin
E          C01474   Alizarin
E          C10291   Alizarin 2-methyl ether
E          C10294   Aloe-emodin
```

Figure 24 A sample for format of KEGG Phytochemicals

Table 1 A sample for modified format of KEGG Phytochemicals with given cluster codes

Kegg Code	Name of Compound	Cluster Code
C10788	Ellagic_acid	1_6_2_20
C10836	3,4,3'-Tri-O-methylellagic_acid	1_6_2_21
C10242	Terminalin	1_6_3_1
	Styrylpyrones	1_7
C09925	5,6-Dehydrokawain	1_7_1
C09926	Dihydromethysticin	1_7_2
C09947	Kawain	1_7_3
C09952	Methysticin	1_7_4
C09980	Yangonin	1_7_5
	Polyketides	2
	Quinones	2_1
	Anthraquinones	2_1_1
	Anthrone type	2_1_1_1
C10305	Barbaloin	2_1_1_1_1
C10307	Cascaroside_A	2_1_1_1_2
C10314	Chrysophanic_acid_9-anthrone	2_1_1_1_3
C10404	Sennoside_A	2_1_1_1_4
	Anthroquinone type	2_1_1_2
C01448	Questin	2_1_1_2_1
C01474	Alizarin	2_1_1_2_2
C10291	Alizarin_2-methyl_ether	2_1_1_2_3
C10294	Aloe-emodin	2_1_1_2_4

Table 2 Number of Compounds in Secondary Metabolite Classes which consist of larger than 5 compounds

Secondary Metabolite Class	Sub-classes	Cluster Codes	Number of Compounds
Other Phenolics	Monolignols	1_1	37
	Lignans	1_2_1	57
	Neolignans	1_2_2	14
	Coumarins	1_3	61

Flavonoids	Flavonoids	1_4_1	348
	Chalcones	1_4_1_1	38
	Dihydrachalcones	1_4_1_2	40
	Flavones	1_4_1_3	86
	Dihydrafalvones	1_4_1_4	17
	Flavonols	1_4_1_5	72
	Flavans	1_4_1_6	27
	Anthocyanidins	1_4_1_7	71
	Isoflavonoids	1_4_2	124
	Biflavonoids and Polyflav.	1_4_4	11
	Condensed tannins	1_4_5	20
Non-Flavonoid Phenolics	Stilbenoids	1_5	39
	Hydrolysable Tannins	1_6	14
Polyketides	Anthroquinones	2_1_1	41
	Gamma-pyrones	2_2	56
	Chromones	2_2_2	17
	Xanthenes	2_2_4	34
Terpenoids	Monoterpenoids	3_2	164
	Sesquiterpenoids	3_3	298
	Diterpenoids	3_4	170
	Triterpenoids	3_6	275
	Tetraterpenoids	3_7	36
	Polyterpenoids	3_8	8
	Tropane Alkaloids	4_1_1	46
Alkaloids	Pyrrolizidine Alkaloids	4_1_2	57
	Piperidine Alkaloids	4_2_1	37
	Quinolizidine Alkaloids	4_2_2	33
	Indolizine Alkaloids	4_2_3	6

	Pyridine Alkaloids	4_3_1	18
	Isoquinoline Alkaloids	4_4_1	200
	Indole Alkaloids	4_5_1	111
	Quinoline Alkaloids	4_5_2	54
	Pyrroloindole Alkaloids	4_5_3	5
	Quinazoline Alkaloids	4_5_4	16
	Acridone Alkaloids	4_5_5	13
	Stereoidal alkaloids	4_7_4	95
	Purine Alkaloids	4_8_1	17

3.3.2. FINDING DOMAIN BINDING REGIONS

For structure based screening, important residues and active sites of domains were found from previous studies and analyzed by using the softwares Discovery Studio [119] and Autodock Tools [120]. We used PDZ1 domain of NHERF (Pdb ID: 1I92), BROMO domain of GCN5P (Pdb ID: 1E6I), Pyrin Pryspry domain (Pdb ID: 2WL1) as targets in dockings [124].

We identified the binding pocket including the residues ASN22, HIS72, PHE26, LEU28, THR71, VAL76, ILE79 and ARG80 for PDZ domain according to the previous studies covered in detail in Literature Review part [44, 45, 47-50]. PDZ domain and its binding pocket with the binding residues are visualized by Discovery Studio as shown in Figure 25.
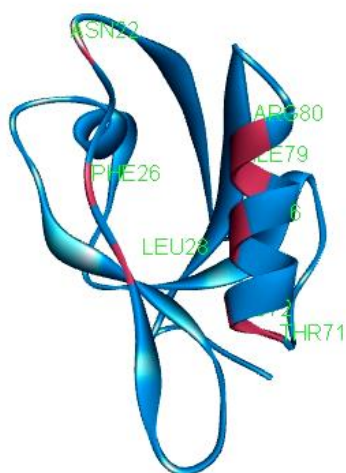
Figure 25 Visualization of PDZ domain (blue) and the binding residues (green) by Discovery Studio (Pdb ID: 1I92)

The binding pocket for BROMO domain was identified from the structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p [46]. We visualized the ligand-protein interactions of bromodomain- acetylated histone H4 complex using Discovery Studio and identified the binding pocket with the residues TYR364, ASP366, TYR406, ASN407 and TYR413. BROMO domain (blue) and its binding pocket with the binding residues (green) are visualized by Discovery Studio as shown in Figure 26.
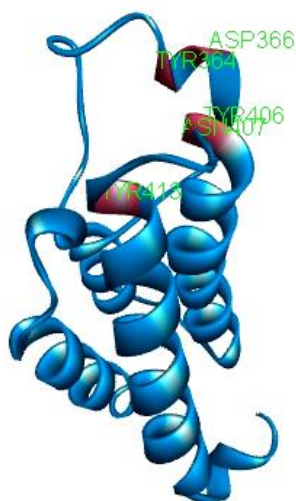
Figure 26 Visualization of BROMO domain (blue) and the binding residues (green) by Discovery Studio (Pdb ID: 1E6I)

We identified the binding pocket including the residues MET680, MET694, TYR608 and ILE666 for SPRY domain according to the previous studies covered in detail in Literature Review part [36, 43, 45] . SPRY domain and its binding pocket with the binding residues are visualized by Discovery Studio as shown in Figure 27.
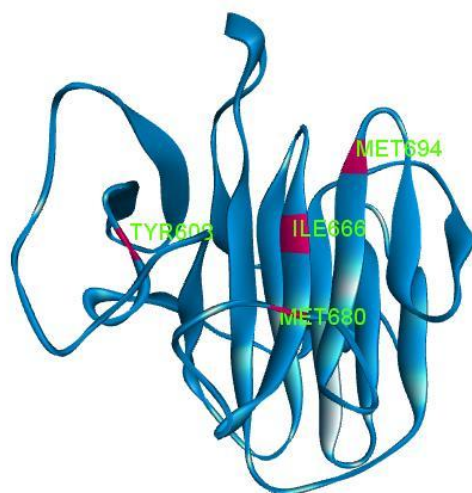
Figure 27 Visualization of SPRY domain (blue) and the binding residues (green) by Discovery Studio (Pdb ID: 2WL1)

3.3.3. DOCKING OF KEGG COMPOUNDS

We performed docking of KEGG Phyochemicals for PDZ, BROMO, SPRY domains by using the docking software VINA [23]. We calculated the binding free energies of 2386 compounds which were docked successfully. We analyzed the docking results of this set for the domains PDZ, BROMO, SPRY on the basis of higher binding scores and determined the promising classes for domains. We calculated average binding free energies of each classes and subclasses and identified 5 classes with the lowest average binding energies that indicate the best binding affinities.

3.3.4. DERIVATION OF CLASSES

Since the number of compounds in the promising 5 classes are different and not enough for appropriate conclusions, we derived 900 compounds for each class by adding fragment to the basic structures of the classes. We used a ligand modifying and fragment replacement software BROOD [125] for this purpose. BROOD generates analogs of the molecule by replacing selected fragments in the molecule with fragments that have similar shape and electrostatics, yet with selectively modified molecular properties.

We derived compounds by replacing one selected fragment of the compounds of the classes with the fragments of the BROOD 2.0 Fragment database with desired molecular properties. Since we search for molecules for PDZ, BROMO, SPRY domains that have hydrophobic binding pocket properties, we replaced the fragments with hydrophobic donors as selective molecular properties.

3.3.5. DOCKING OF DERIVED COMPOUNDS

We performed the docking of the derived compounds of 5 classes for domains PDZ, BROMO, SPRY and analyzed the docking results to identify spesific classes for each domain. For this purpose, we compared the binding probabilities of the classes obtained according to Gibbs Distribution.

3.3.6. GIBBS DISTRIBUTION AND ANALYSIS OF DERIVED COMPOUNDS

Gibbs distribution is a certain distribution function or probability measure for the distribution of the states of a system.

When a ligand binds to the protein with a binding energy, E, it is not stationary and may assume various other configurations around the calculated one. It may also replace water molecules located at the cavity of binding. The Boltzmann factor,

$$e^{-E/k_B T} \quad (1)$$

is proportional to the relative probability of a ligand bound on the protein with a given energy. If the system is at constant pressure, enthalpy can be used as the relevant energy. Then the probabilities are proportional to

$$e^{-H/k_B T} \quad (2)$$

The bound state actually corresponds to several different arrangements of the ligand (with approximately the same energy) at the surface and to several arrangements of the water molecules. So we have to include the entropy as well.

If S is the entropy and W is the number of arrangements of the system at the stated energy, then

$$S = k_B \ln W \quad (3)$$

Solving for W one obtains

$$W = e^{S/k_B} \quad (4)$$

So to find the overall probability that a system is in a given state, the two contributions must be multiplied together:

$$W * e^{-H/k_B T} = e^{S/k_B} * e^{-H/k_B T} = e^{TS/k_B T} * e^{-H/k_B T} = e^{-(H-TS)/k_B T} \quad (5)$$

The term in parenthesis in the last exponent is the Gibbs free energy.

$$G = H\text{-}TS \quad (6)$$

So the overall probability is actually proportional to

$$e^{-G/k_B T} \quad (7)$$

The Gibbs free energy combines the effects of energy (here, that's the original Boltzmann factor) and of entropy (the number of possible arrangements).

Therefore, for systems at constant pressure and temperature, the Gibbs free energy tells us which states are more probable, and therefore what a system is actually going to do.

In our system, we assumed our protein and ligands of all classes in a closed system and aimed at finding the binding probability of each promising secondary metabolite class to this protein. First of all, we performed the dockings of all classes which consist of the same number of molecules in each and calculated the binding free energies. We used VINA for dockings which calculate the Gibbs free energy of binding empirically. We used this Gibbs free energy to obtain the binding probabilities of each drug molecule.

The probability of a state's occurrence, is proportional to

$$e^{-G/k_B T} \quad (8)$$

where G is Gibbs free energy of binding, T is the system's absolute temperature and where k is the Boltzmann constant. For such a system in equilibrium at the temperature $T,$ the probability Pi of finding it in the particular state i is

$$P_i = \frac{e^{\frac{-G_i}{K_B T}}}{\sum_i e^{\frac{-G_i}{K_B T}}} \quad (9)$$

The denominator is the sum of exp(−Gi /kT) over all states i (and so does not depend on i, which is there just a dummy summation index), and is what guarantees that Pi is properly normalized:

$$\sum_{i=1} P_i = 1 \quad (10)$$

The normalization denominator is called the partition function of the system [126].

We obtained the binding probability of the each molecule using Gibbs free energies;

$$\text{Pi} = \frac{e^{\frac{-G_i}{K_B T}}}{\sum_{i=1}^{N} e^{\frac{-G_i}{K_B T}}} \qquad (11)$$

Where :

G_i=Gibbs free energy of binding of the ith compound (kcal/mol)

N_0=number of all ligands

P_i= binding probability of ith compound

T=temperature

k= Boltzmann constant

We used T=300K and k=1.987x 10^{-3} kcal/mol.K

Then, we sum the probabilities of the compounds that belong to each class in the set of best binding N drugs to determine the binding probability of each class (S_i).

$$N = \sum_i n_i \qquad (12)$$

$$S_i = \sum_{j=1}^{n_i} p_{ij} \qquad (13)$$

Where :

N=number of ligands in the set best binding drugs

n_i= number of ligands from i[th] class that appear in the set of best binding N drugs

S_i=binding probability of ith class

P_{ij}=binding probability of jth compound of ith class

Since the denominator is the sum of $\exp(-G_i/kT)$ over all states i in (4) and (6), and is what guarantees that P_i and S_i is properly normalized:

$$\sum_{i=1}^{N} p_i = 1 \quad (14)$$

$$\sum_i S_i = 1 \quad (15)$$

Where :

N=number of ligands in the set best binding drugs

S_i=probability of the ith class

We compared the probability of the classes (S_i) for each domain in order to identify the class which has the best binding affinity for each domain.

Chapter 4

4. RESULTS AND DISCUSSION

4.1. RESULTS OF PART1

4.1.1. RESULTS FOR POL γ

Many studies show that proximal accessory subunit of pol γ contacts the catalytic subunit through interaction with the AID domain of polymerase domain, whereas the other subunit provides little surface interaction [22]. The majority of the interactions occur between one monomer of the dimeric accessory subunit and the catalytic subunit [21]. Studies indicated that AID subdomain predominantly interacts with the proximal monomer and these interactions result in enhanced DNA binding affinity which determines the processivity of holoenzyme [21, 26]. Since these interactions are so important in the activity of holoenzyme, we identified small molecule inhibitors for accessory subunit which will prevent the binding of L-helix of the AID domain to the accessory subunit to inhibit the haloenzyme.

We have identified Fucax23, Xntmf27 and Dxbd20 from KEGG Phytochemicals Database as promising inhibitors of pol γ, with -10.7 kcal/mol, -9.8 kcal/mol and -9.8 kcal/mol binding free energies, respectively.

Protein-Ligand complexes are visualized in Discovery Studio and analyzed as 2D-pharmacophore models in Ligand Scout (Figure 28 Figure 29 Figure 30). All three promising molecules have many hydrophobic interactions with the active residues in the binding pocket of accessory subunit of pol γ. It is seen that ligands have hydrophobic

interactions (yellow sphere) with the active residues Leu402 and Ile468 as shared common features. Xntmf27 and Dxbd20 have additional hydrophobic common interactions with the other active residues Leu448 and Val398.
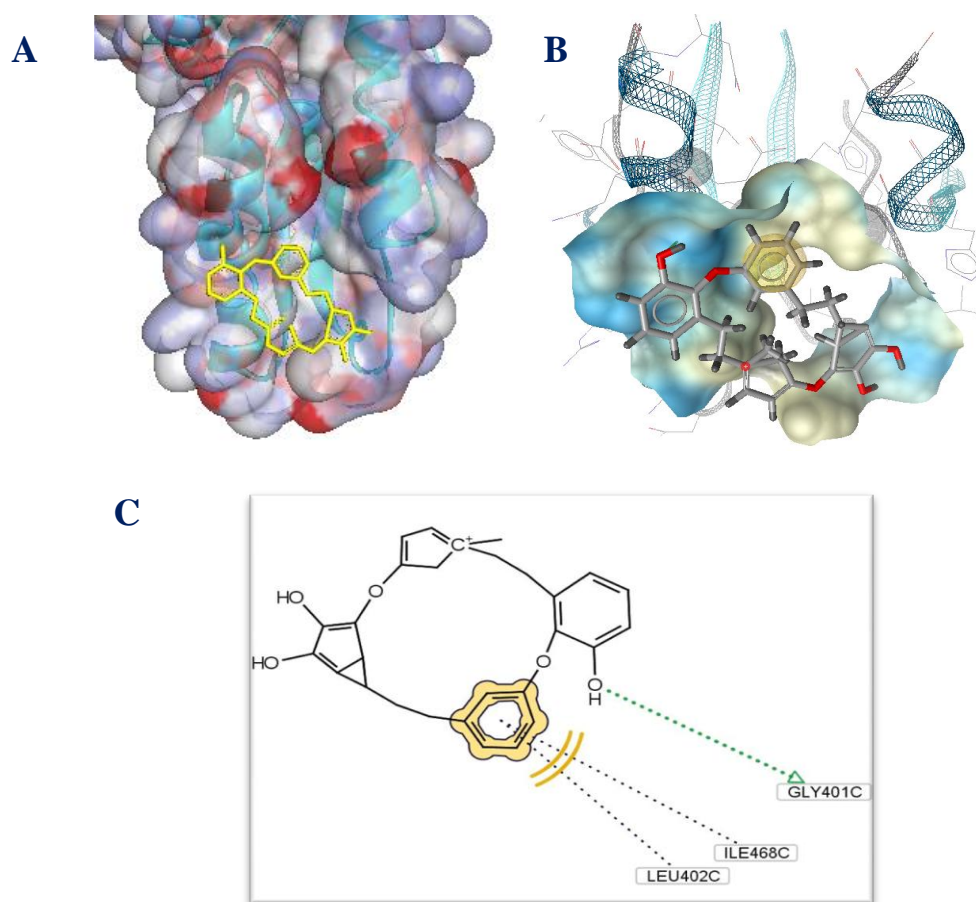
Figure 28 A) Visualization of Fucax23 in binding pocket of proximal accessory subunit of pol γ using Discovery Studio. B) 3-D and C) 2-D Pharmacophore models generated by Ligand Scout
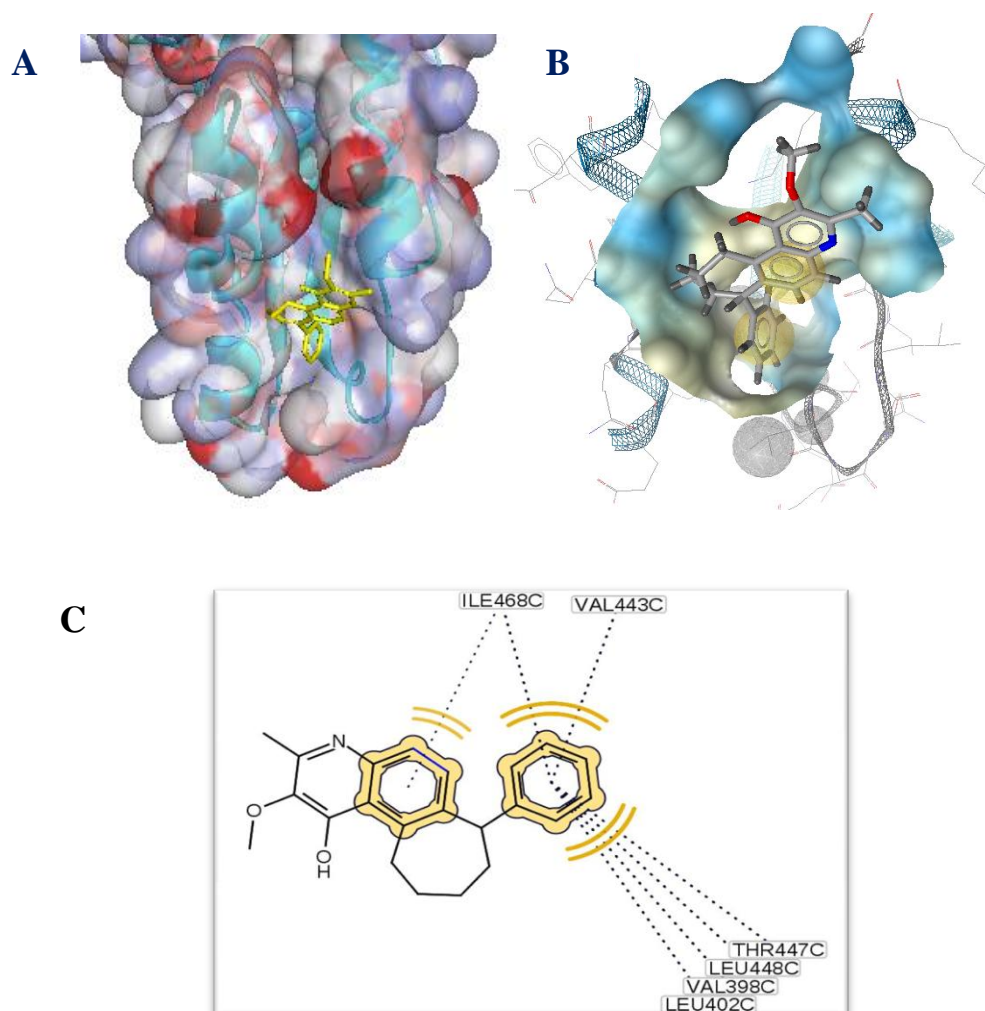
Figure 29 A) The visualization of Xntmf27 in the binding pocket of proximal accessory subunit of pol γ using Discovery Studio. B) 3-D and C) 2-D Pharmacophore models generated by Ligand Scout
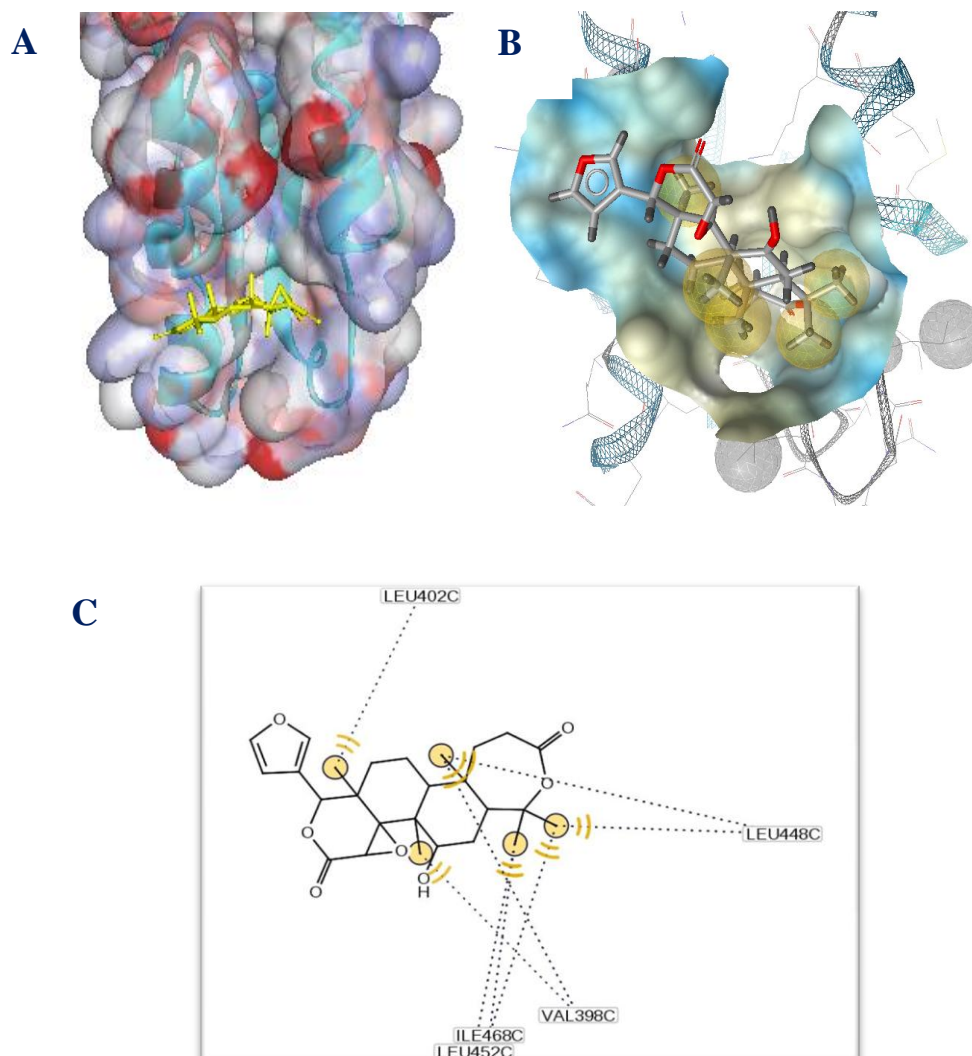
Figure 30 A) The visualization of Dxbd20 in binding pocket of proximal accessory subunit of pol γ using Discovery Studio. B) 3-D and C) 2-D Pharmacophore models generated by Ligand Scout

4.1.1. RESULTS FOR POL β

In previous studies, it is reported that the active pocket including residues Lys35, Lys60, Lys68, Lys72 in N-terminal 8 kDa domain of pol β retains binding affinity for ssDNA, 5'-phosphate recognition in gapped DNA, and dRP lyase [18]. We targeted this pocket and identified a molecule that interacts with these active residues. Anxbd47 was identified as the small molecule inhibitor for pol β with -10.7 kcal/mol binding free energy.

Anxbd47, the hit molecule for pol β, was visualized in the binding pocket of Lyase domain of pol β and analyzed in Ligand Scout. We showed that it interacts with the active residues in the binding pocket of Lyase domain (Figure 31). It is seen that the Ligand has interactions with most of the active residues discussed. It has negative ionizable interactions (red) with the active residues of lyase domain Lys35, Lys68 and Lys72 and hydrogen bond acceptors (red) with Lys35 and Lys72 in the binding pocket.
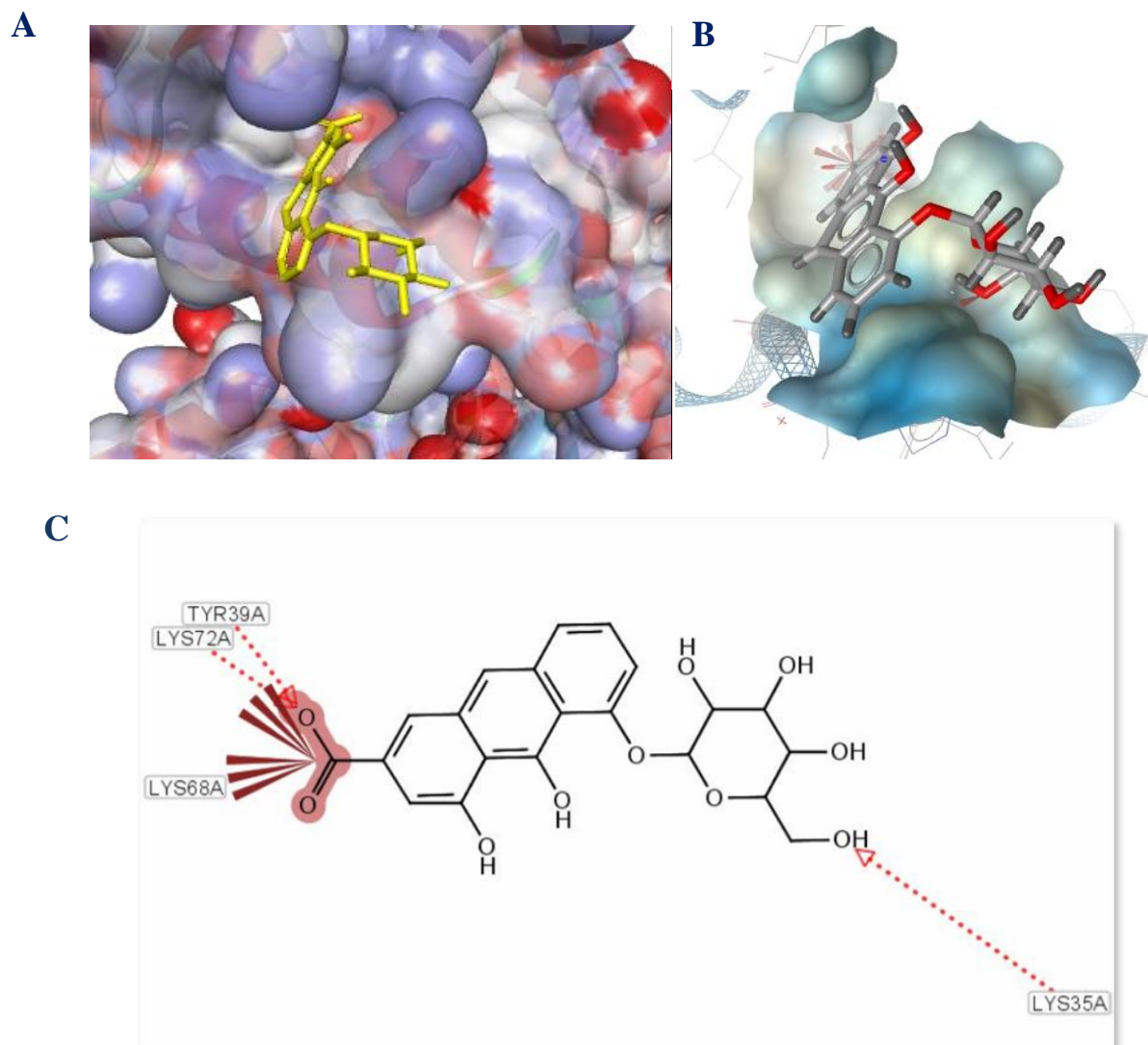
Figure 31 A) The visualization of Anxbd47 in binding pocket in lyase domain of pol β using Discovery Studio. B) 3-D and C) 2-D Pharmacophore models generated by Ligand Scout

4.2. RESULTS OF PART2

4.2.1. RESULTS OF KEGG LIBRARY DOCKINGS AND ANALYSIS

200 drugs of 2386 docked plant based compounds with the highest binding affinity were selected as a set of best binding drugs with -8.4 kcal/mol, -8.9 kcal/mol and -8,9 kcal/mol cut off binding free energies for PDZ, BROMO and SPRY domains, respectively. The average binding free energies of classes to each domain are showed in Table 3-5.

Table 3 Average binding free energies of secondary metabolite classes for pdz domain

Class of Secondary metabolite	Average binding free energy of the class (kcal/mol)
Triterpenoids	-9,1
Stereoidal alkaloids	-9,04
Flavonoids	-8,9
Stilbenoids	-8,5
Anthroquinone	-9
Indole alkolodis	-8,8
Quinoline alkaloids	-8,6
Pyrrolizidine alkolodis	-8,6
Pyrroloindole alkolodis	-8,7
Acridone	-9
Dhydraflavonols	-9,03
Flavones	-8,91
Dhydrachalcones	-8,9
Anthocyanidins	-9,02

Table 4 Average binding free energies of secondary metabolite classes of for bromo domain

Class of Secondary metabolite	Average binding free energy of the class (kcal/mol)
Triterpenoids	-9,3
Stereoidal alkaloids	-9,7
Flavonoids	-9,4
Stilbenoids	-10
Anthroquinone	-9,4
Isoflavonoids	-9,05
Xanthenes	-9,3
Monoterpenoids	-9,2
Sesquiterpenoids	-9,15
Diterpenoids	-8,95
Quinolizidine alkaloids	-9,1
Isoquinoline alkaloids	-9,2
Flavans	-9,13
Dhydraflavonols	-9
Flavones	-9,33
Dhydrachalcones	-9,57
Anthocyanidins	-9,6

Table 5 Average binding free energies of secondary metabolite classes for spry domain

Class of Secondary metabolite	Average of binding free energy of the class (kcal/mol)
Triterpenoids	-9,3
Stereoidal alkaloids	-9,8
Flavonoids	-9,3
Stilbenoids	-9,15
Anthroquinone	-9,6
Isoflavonoids	-9,1
Diterpenoids	-9,1
Quinoline alkaloids	-9,2
Isoquinoline alkaloids	-9,07
Dhydraflavonols	-9,2
Flavones	-9,44
Dhydrachalcones	-9,45
Anthocyanidins	-9,4
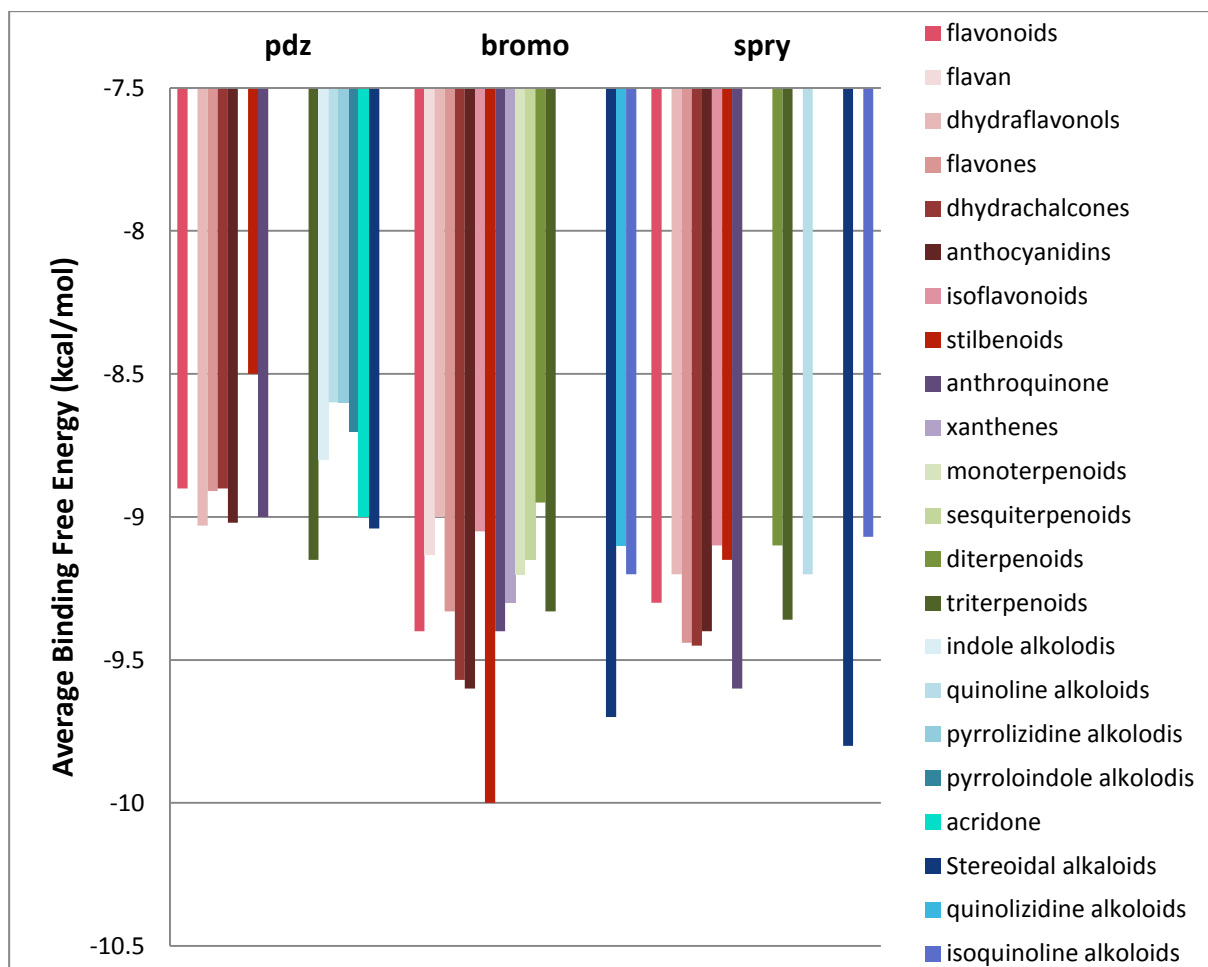
Figure 32 Average binding free energies of the classes for PDZ, BROMO and SPRY domains.

As seen in Figure 32, the classes that bind to the domains with lowest binding free energies are different. Since some classes are above the cut off binding free energies, they are not shown in the graph. The colors of the columns are same for the same classes with changing light to dark according to the subclasses. It is seen that out of terpenoids, Triterpenoids is the most promising class with -9.1 kcal/mol binding energy for PDZ domain. However, there are many subclasses from flavonoids and alkaloids which bind to domain with rather lower binding affinities. Stilbenoids is the

most promising class for BROMO domain with -10 kcal/mol binding free energy. Terpenoid alkaloids is the second group that have the most binding affinity to BROMO domain and some flavonoids subclasses which are also phenlyproponaid related classes like stilbenoids follow them. For SPRY domain, Stereoidal alkaloid is the most promising class with -9.8 kcal/mol binding free energy followed by Anthroquinones which bind to other domains with rather higher binding free energies.

When we look at the overall picture we see that many Flavonoids subclasses bind to all domains with very close but acceptable binding free energies. For this reason, the average of these subclasses is shown under flavonoids. On the other hand, triterpenoids are the most potential candidates out of terpenoids and same condition is valid for Stereoidal alkaloids out of alkaloids. Anthroquinones seem to be also potential candidates for SPRY domain. Finally, it is clearly seen that Stilbenoids is the most promising one for BROMO domain.

As a result we identified Flavonoids and Stilbenoids from Phenylproponaids, Stereoidal alkaloids, Anthroquinones and Triterpenoids are potential classes for these 3 domains and decided to investigate further into these groups. Since the numbers of molecules that are included in these groups are such different such as shown in Table 6, we derived new sets of molecules using the basic structure of the classes.

Table 6 Number of compounds in the promising classes identified
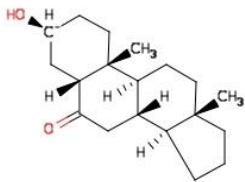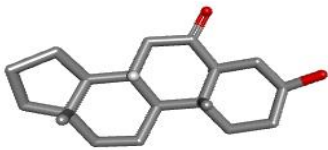
Class of Secondary metabolite	n (number of compounds of the class)
Triterpenoids	275
Stereoidal alkaloids	95

Flavonoids	348
Stilbenoids	39
Anthroquinone	41

4.2.2. DERIVATION OF SECONDARY METABOLITES

The basic structures of the classes identified in previous part are shown in Table 7.

Table 7 Basic Structures of secondary metabolite classes and their visualizations by Discovery Studio

Class of Secondary metabolite	Basic Structure of Secondary Metabolite Class	3-D Visualization of basic structures used in derivations
Triterpenoids		

Stereoidal
alkaloids

Flavonoids

Stilbendoids
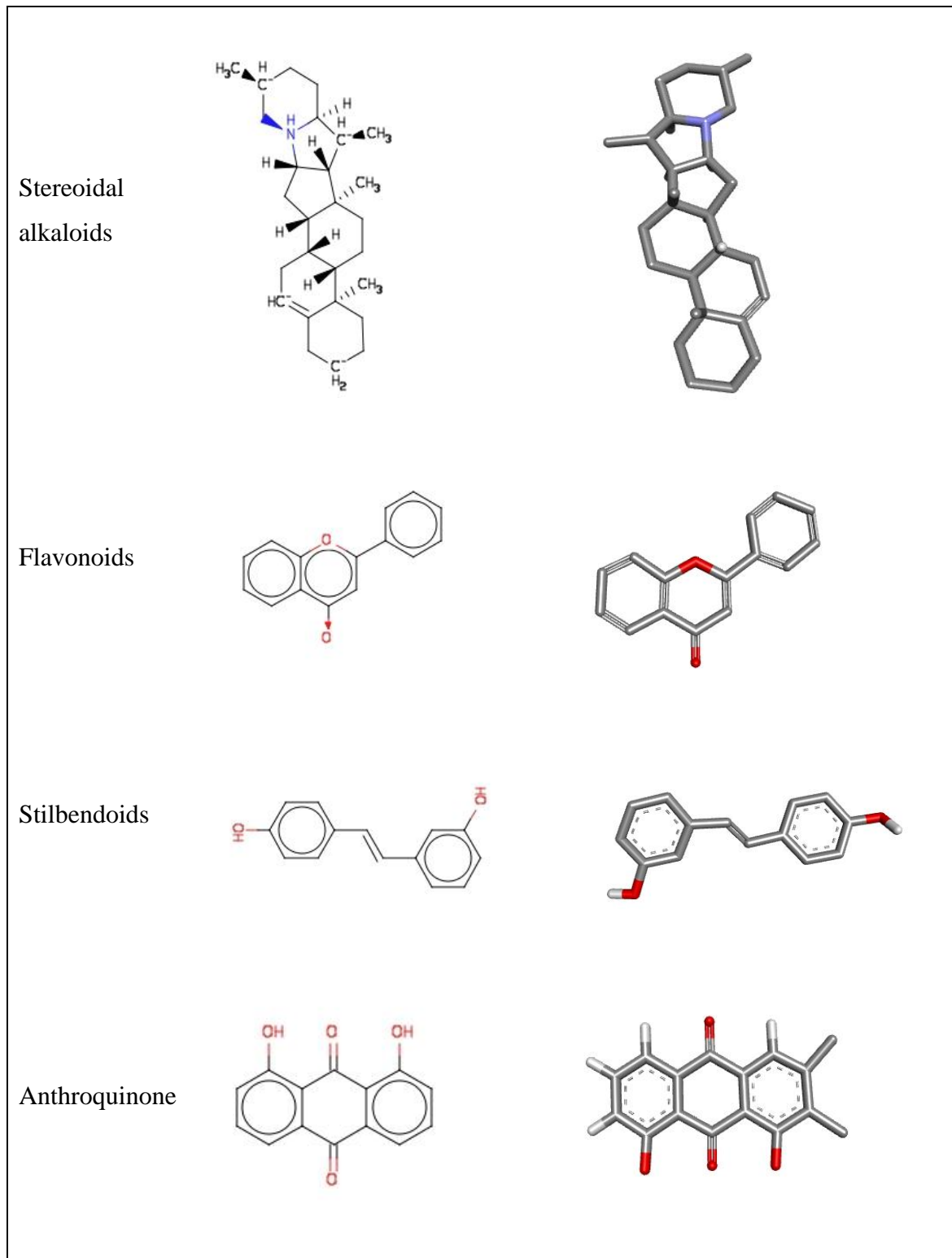
Anthroquinone

We derived 900 compounds for classes Flavonoids, Stilbenoids, Triterpenoids, Stereoidal alkaloids and Anthroquinones adding one fragment to the basic structures of the classes. The derived molecules were generated by using BROOD [125]. The derived molecules were visualized (Figure 33 Figure 34 Figure 35 Figure 36 Figure 37) using a Visualization and Modelling Software VIDA which is a Openeye Software Tool [127].
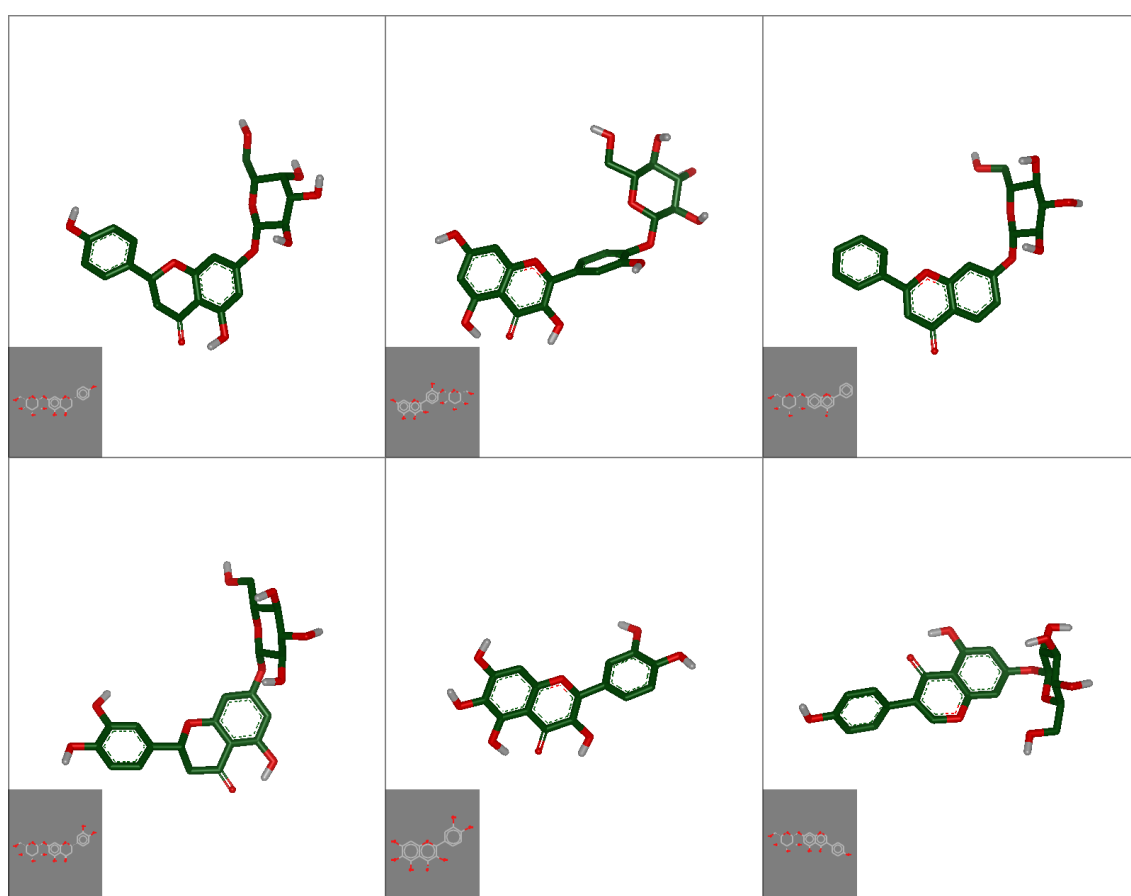
Figure 33 Visualization of derived Flavonoids by VIDA[127]
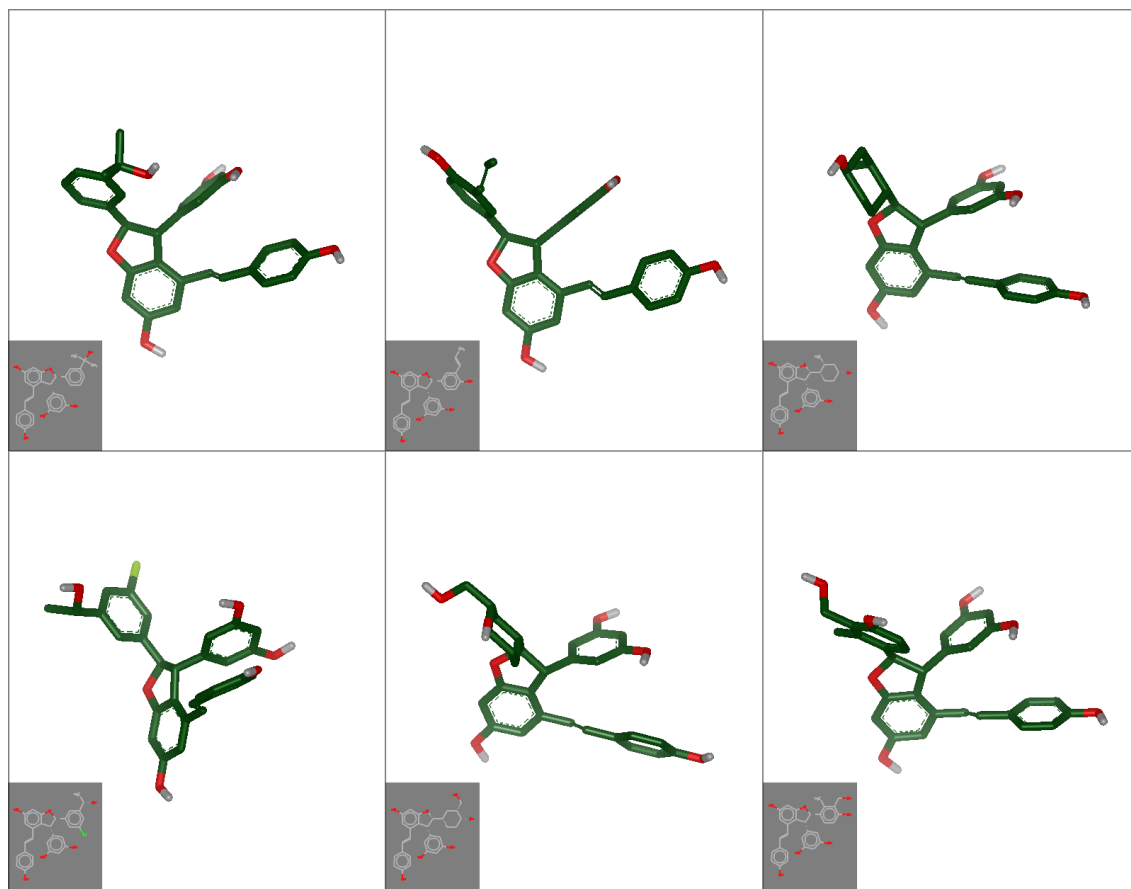
Figure 34 Visualization of derived Stilbenoids by VIDA[127]
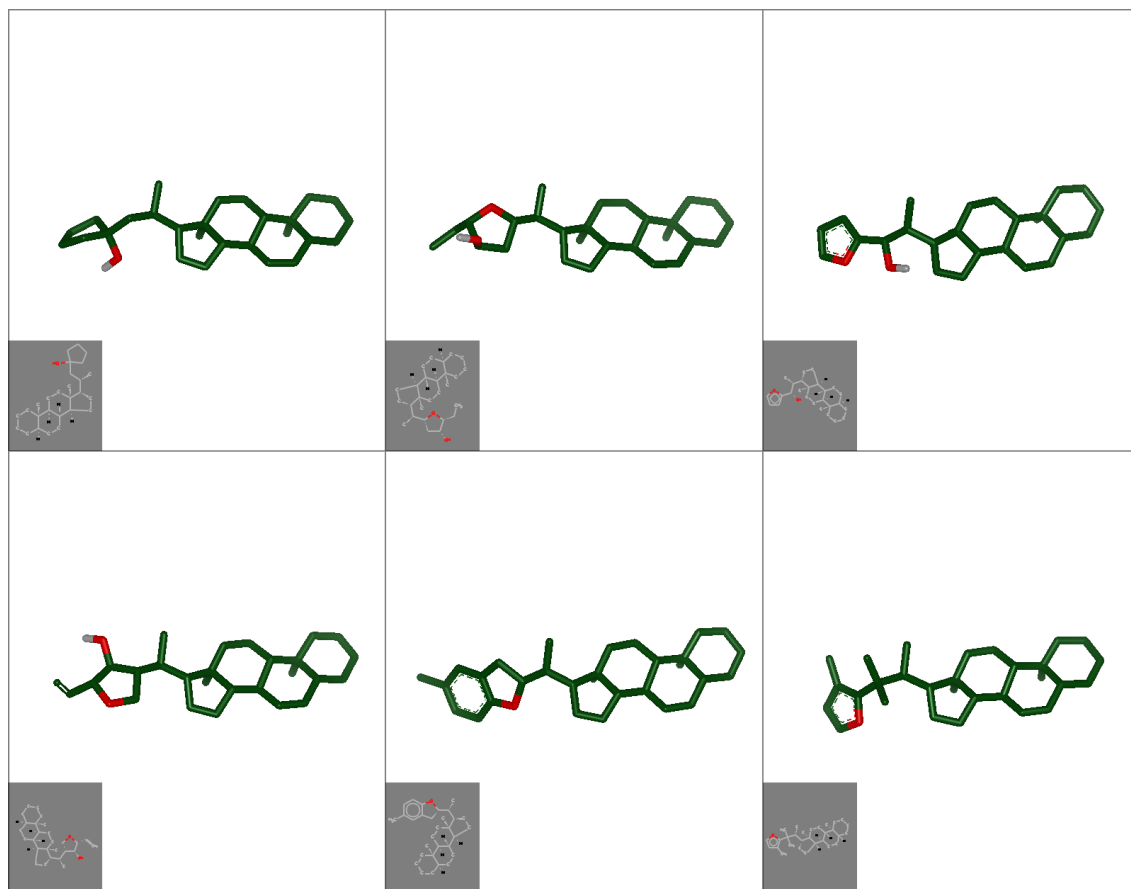
Figure 35 Visualization of derived Triterpenoids by VIDA [127]
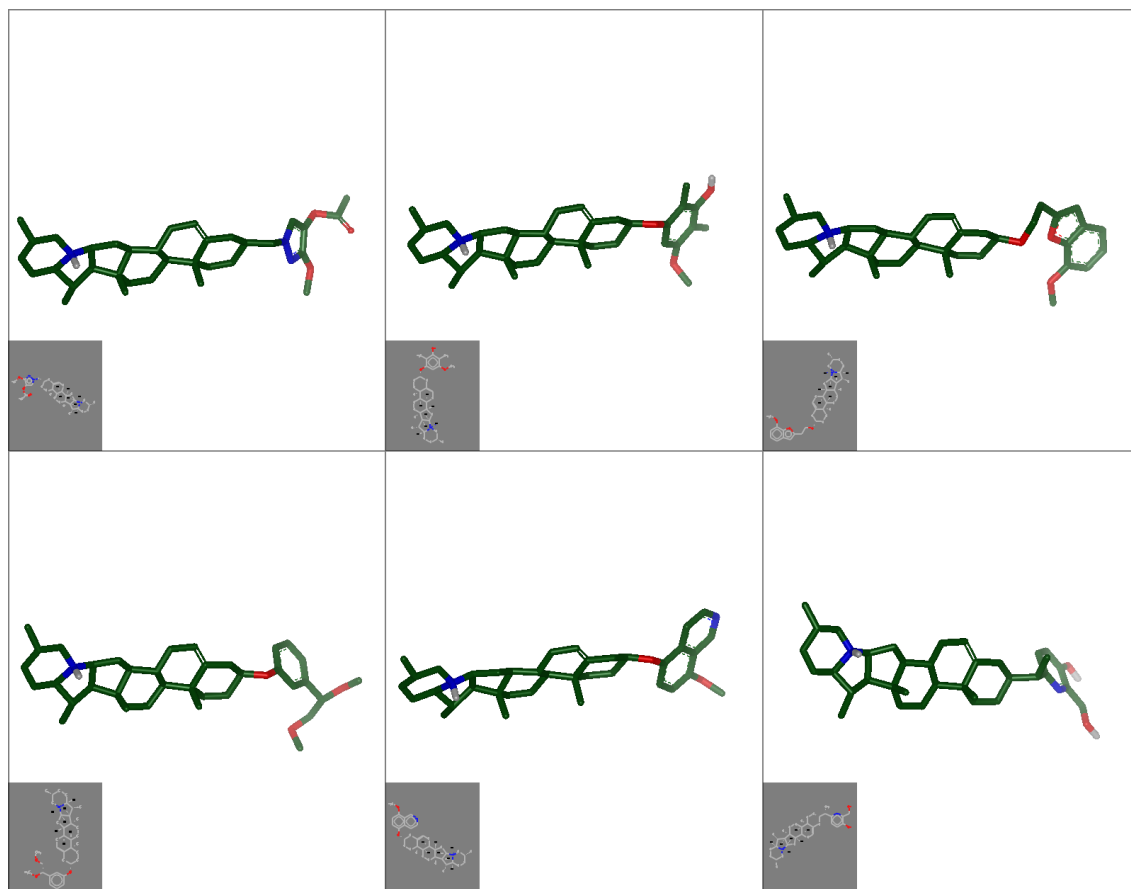
Figure 36 Visualization of derived Stereoidal alkaloids by VIDA [127]
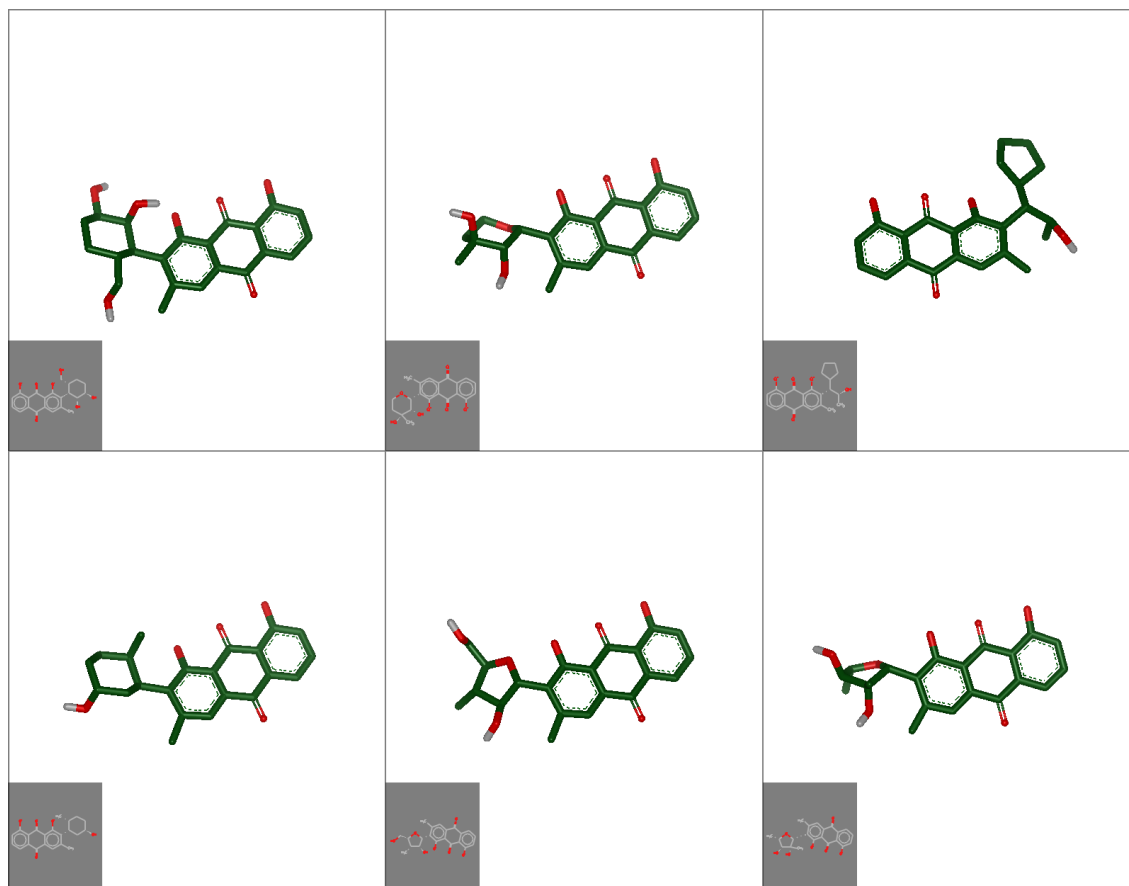
Figure 37 Visualization of derived Anthroquinones by VIDA [127]

4.2.3. RESULTS OF DERIVED COMPOUNDS

We performed virtual screening of derived compounds of Triterpenoids, Flavonoids, Stereoidal alkaloids, Stilbenoids and Anthroquinones for PDZ, BROMO, SPRY domains.

Table 8 Binding probabilities of the plant secondary metabolite classes to the domains obtained for all ligands according to the Gibbs Distribution

	PDZ	BROMO	SPRY
Stilbenoids	0,15	0,26	0,2
Triterpenoids	0,32	0,23	0,25
Stereoidal alkaloids	0,26	0,18	0,33
Flavonoids	0,05	0,1	0,03
Anthroquinones	0,22	0,23	0,19
	1	1	1

Table 9 Binding probabilities of the plant secondary metabolite classes to the domains obtained among the set of best binding 200 drugs according to the Gibbs Distribution.

	PDZ	BROMO	SPRY
Stilbenoids	0,004	0,34	0,004
Triterpenoids	0,55	0,25	0,18
Stereoidal alkaloids	0,21	0,18	0,77
Flavonoids	0,14	0,19	0,046
Anthroquinones	0,096	0,04	0
	1	1	1

Table 10 Binding probabilities of the plant secondary metabolite classes to the domains obtained among the set of best binding 100 drugs according to the Gibbs Distribution

	PDZ	BROMO	SPRY
Stilbenoids	0	0,31	0
Triterpenoids	0,44	0,25	0,11
Stereoidal alkaloids	0,18	0,16	0,84
Flavonoids	0,25	0,24	0,05
Anthroquinones	0,13	0,04	0
	1	1	1

When we analyze the binding probabilities obtained according to Gibbs Distribution for all ligands, we see that Triterpenoids have the highest binding probability to PDZ domain, while Stereoidal alkaloids are more specific class for SPRY domain. In the case of BROMO domain, however Stilbenoids have the highest binding probability to BROMO domian, Bromo domain does not seem to show a significant differentiaton of ligands (Table 8).

When we look at the picture of results for the set of 200 best binding drugs which have rather considerable binding energies for domains, we see an obvious increase in the differentiation of PDZ and SPRY domain, whereas BROMO domain still have close binding probabilities of classes (Table 9).

When we examine the set of 100 best binding drugs with rather lower binding energies, a significant speficity appear for Stereoidal alkaloids in the case of SPRY domain and PDZ domain still have an important differentiation of Triterpenoids. Stilbenoids keep being most specifc class for BROMO domain (Table 10).

We obtained the binding probabilities of each class for each domain according to Gibbs distribution for the sets of best binding 10, 20, 30, 40…3790, 3800 drugs according to -5 kcal/mol cut off energy to see the overall picture. We plotted the probability of the each class (S_i) vs. number of best binding drugs (N) for each domain.
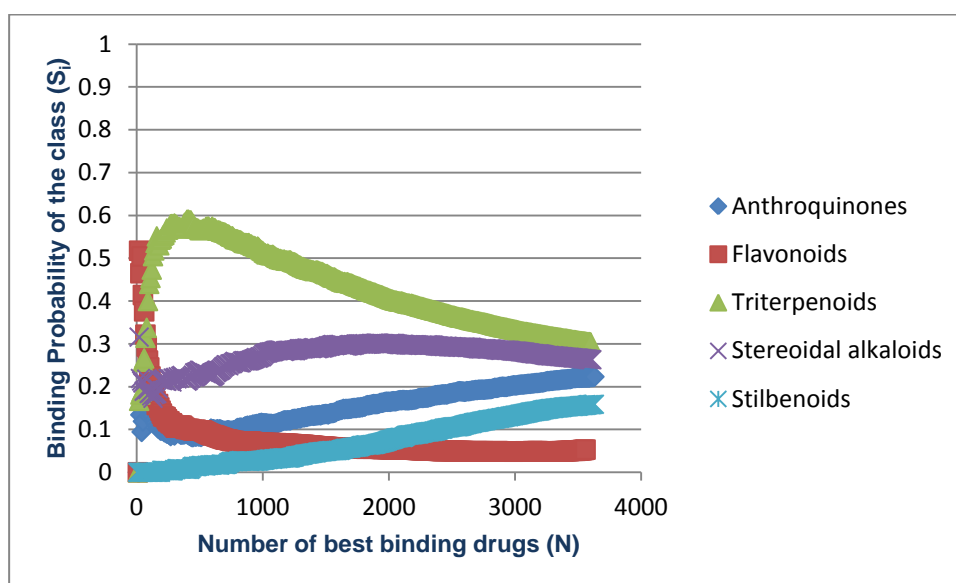
Figure 38 Binding probabilities of the classes for PDZ domain

Figure 38 shows the binding probability changes of classes for PDZ domain depend on the number of best binding drugs. However there is a continous decrease, it is seen that the binding probabilitiy of Triterpenoids class is the highest on whole with the range of 0.3-0.6 binding probability. Stereoidal alkaloids rank the second with the binding probability changes in a range of 0.2-0.3 which is very low compared to Triterpenoids. The results verify the results of previous analysis done for KEGG compounds and showed that triterpenoids are the most promising secondary metabolite class for PDZ domain.
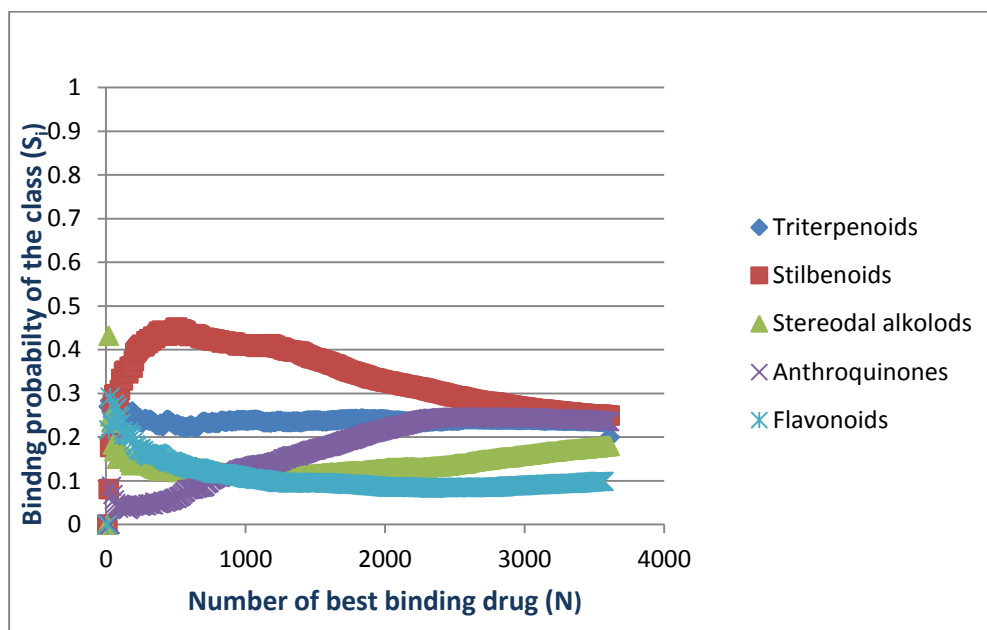
Figure 39 Binding probabilities of the classes for BROMO domain

In Figure 39 we see the (Si) vs. number of best binding drugs (N) graph for BROMO domain. Though the binding probabilities of the classes are close to each other, Stilbenoids show rather higher binding affinity for BROMO domain with the binding probabilities obtained between 300-2000 best binding drugs. The result is again consistent with the previous calculations. Because the results from KEGG library showed that there is more distributed profile of classes and subclasses for BROMO domain.
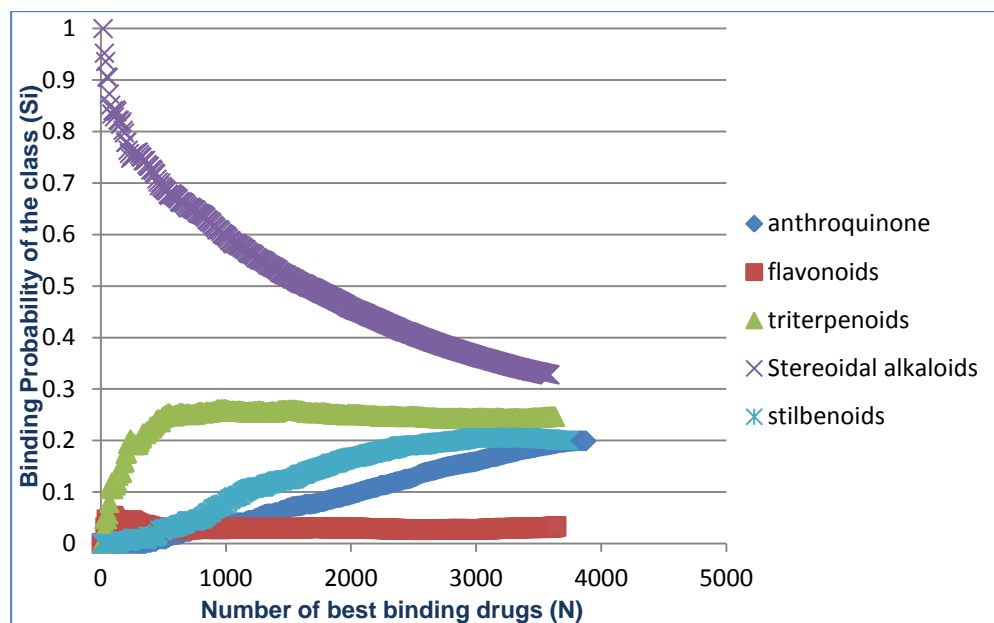
Figure 40 Binding probabilities of the classes for SPRY domain

Figure **40** shows that Stereoidal alkaloids have the highest binding affinity with the overall binding probability to the SPRY domain. However, this probability is decreasing continuously until it is like to converge.

4.3.   PHARMACOPHORE ANALYSIS

Compounds for PDZ, BROMO and SPRY domians were taken on the basis of higher scoring function for pharmacophore modeling in order to model the key features in protein-ligand complexes. The phamacophore models were generated by Ligand Scout. In models, hydrophobic interactions (yellow sphere), aromatic rings (blue), negative ionizable area (red), hydrogen bond acceptors (red arrow), hydrogen bond donors (green arrow) are observed.

Stilbenoids that are bound to BROMO domain with the highest binding scorings were analyzed. Protein-ligand interactions were shown in both 3-D and 2-D pharmacophore models. Epsilon-viniferin, a stilbenoid from plant secondary metabolites library binds with -10.3 kcal/mol binding energy which is the lowest one. Epsilon-viniferin is a resvaratrol dimer extracted from Vitis vinifera. It is found in grapevines and known for its antioxidant activity [128]. Pharmacophore model for Epsilon-viniferin is shown in Figure 41 and a derived stilbenoid with -10.2 kcal/mol binding energy is shown in Figure 42. It is seen that both ligands have hydrophobic interactions (yellow sphere) with the residues Tyr364, Val356 and Phe352 as common features. Derived stilbenoid has an aromatic ring (blue) interacts with TYR 413 and hydrogen bond donor (green arrow) with Asn407 (Figure 41) which are active binding residues that have been focused on in previous studies [114, 129].

Figure 41 3-D and 2-D Pharmacophore model of a stilbenoid from secondary metabolites library in the binding pocket of BROMO domain generated by Ligand Scout

Figure 42 3-D and 2-D Pharmacophore model of a derived stilbenoid in the binding   pocket of BROMO domain generated by Ligand Scout.

Triterpenoids that are bound to PDZ domain with the highest binding scorings were analyzed. Rotundioside, a tirterpenoide from plant secondary metabolites library binds with -10.3 kcal/mol binding energy which is the lowest one. Rotundioside is

shown in 43 and a derived triterpenoid with -9,1 kcal/mol binding energy is shown in Figure 17. Both ligands have hydrophobic interactions (yellow spheres) with the residues Val76, Leu28, Phe 26 and Ile79 which are identified as active residues and important in the protein-protein interactions in the literature [44, 47, 50] [49]. Another common residue His72 interacts with a negative ionizable area (red) in Rotundioside (Figure 43) and interacts with two aromatic rings (blue) in derived triterpenoid (Figure 44).

Figure 43 3-D and 2-D Pharmacophore model of Rotundioside from secondary metabolites library in the binding pocket of PDZ domain generated by Ligand Scout
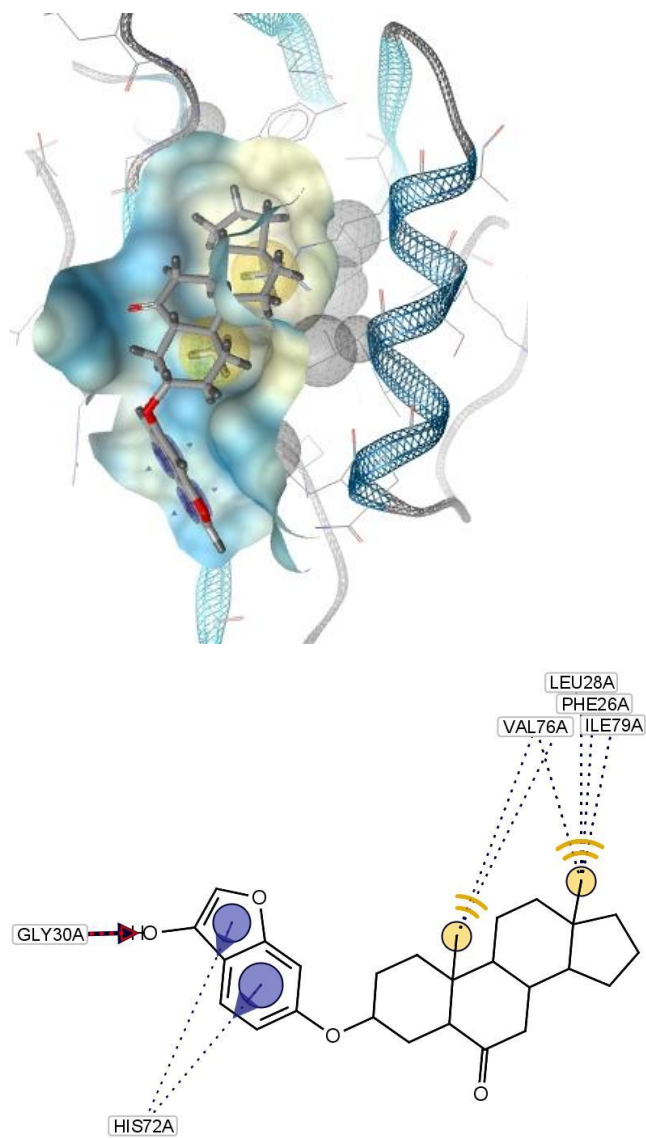
Figure 44 3-D and 2-D Pharmacophore model of a derived triterpenoid in the binding pocket of PDZ domain generated by Ligand Scout

Stereoidal alkaloids that are bound to SPRY domain with the highest binding scorings were analyzed. Alpha chaconine, a Stereoidal alkaloid from plant secondary metabolites library binds with -10.2 kcal/mol binding energy which is the lowest

energy. Alpha chaconine is potato glycoalkaloid which have shown cytotoxic effects on human cancer cells [130]. Alpha chaconine (Figure 45) and a derived Stereoidal alkaloid with -10.2 kcal/mol binding energy (Figure 46) were analyzed. Both ligands have hydrophobic interactions (yellow sphere) with the residues Ile692, Met 694, Ile666, Val 704 and Ala 664 which are binding residues of SPRY domain and critical in FMF [36, 43, 45]. The other common interaction is Thr760 as hydrogen bond donor and acceptor.
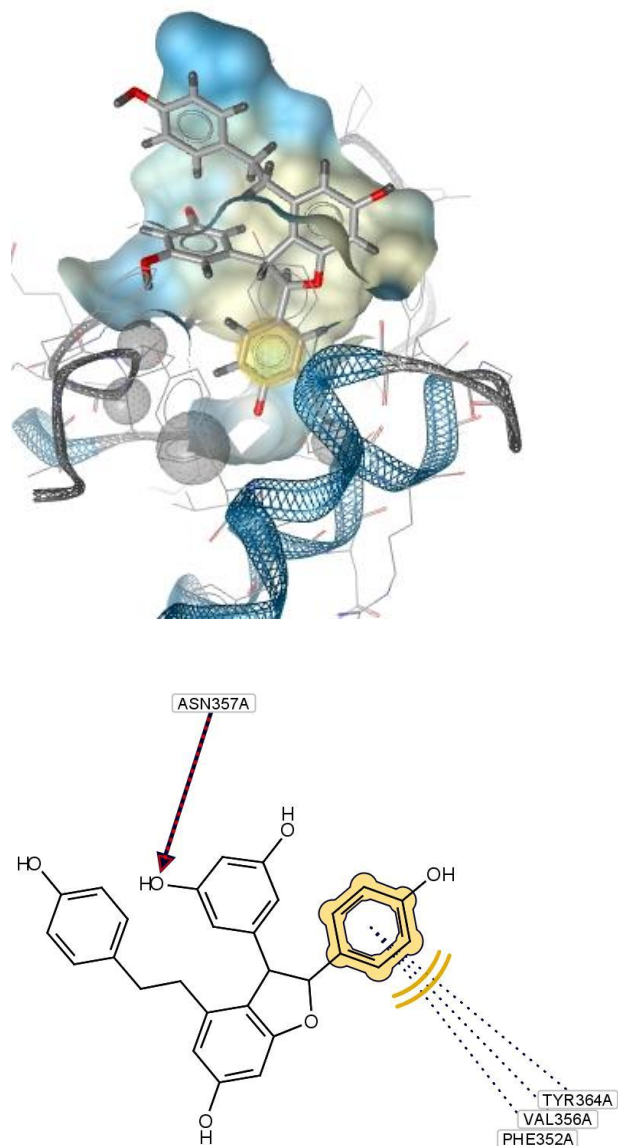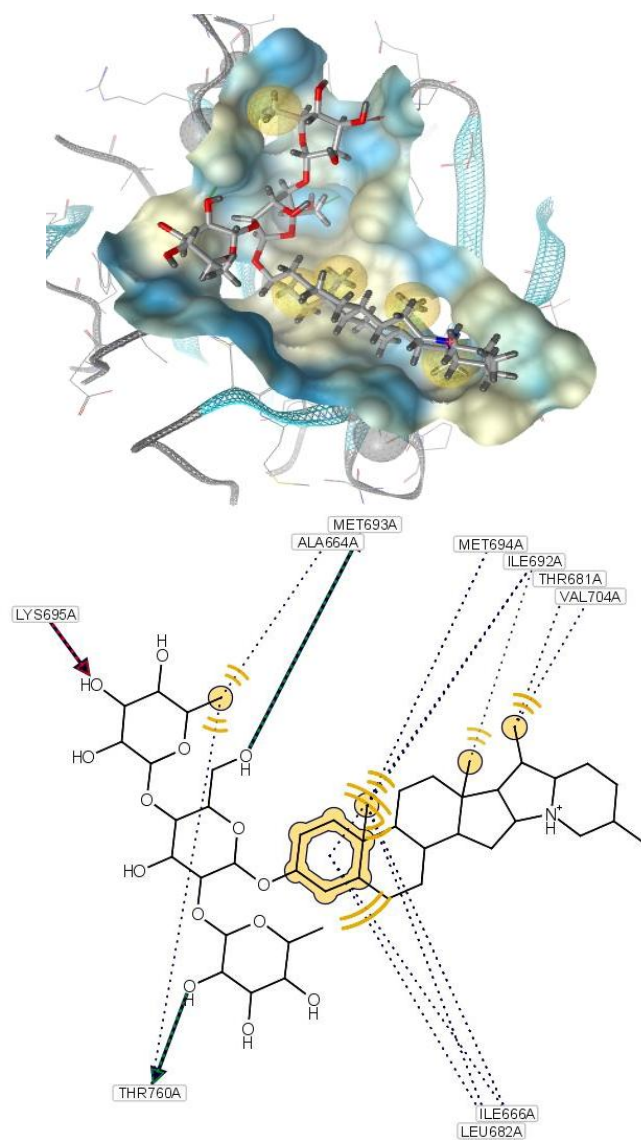
Figure 45 3-D and 2-D Pharmacophore models of Alpha chaconine from plant secondary metabolites library in the binding pocket of SPRY domain generated by Ligand Scout
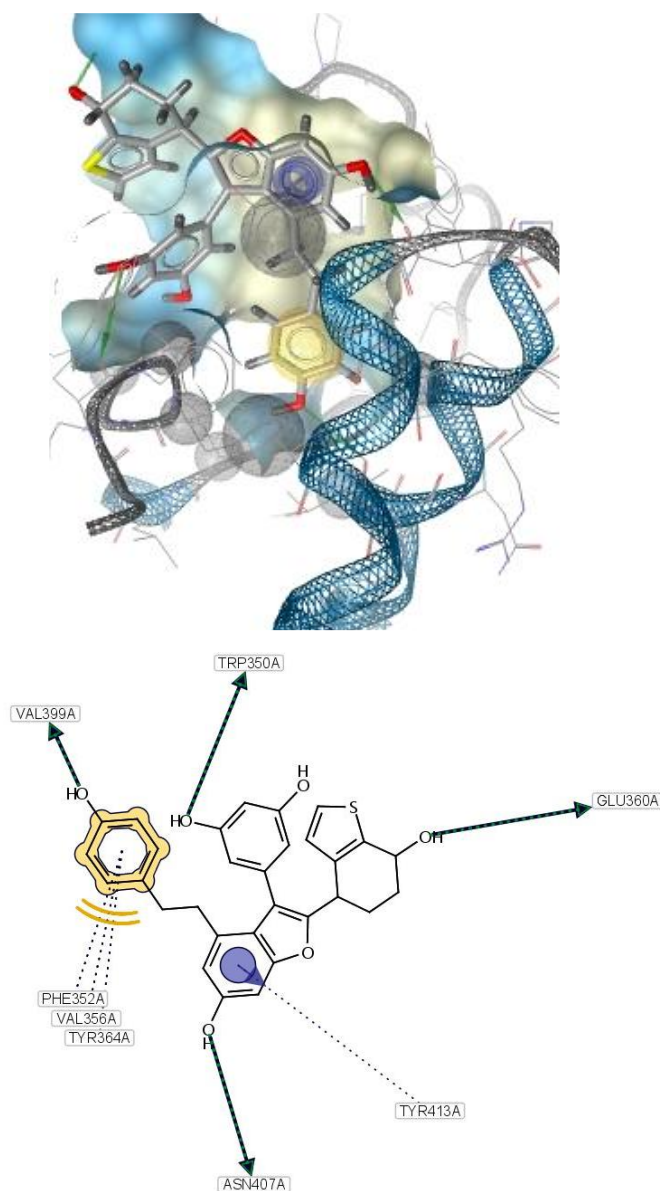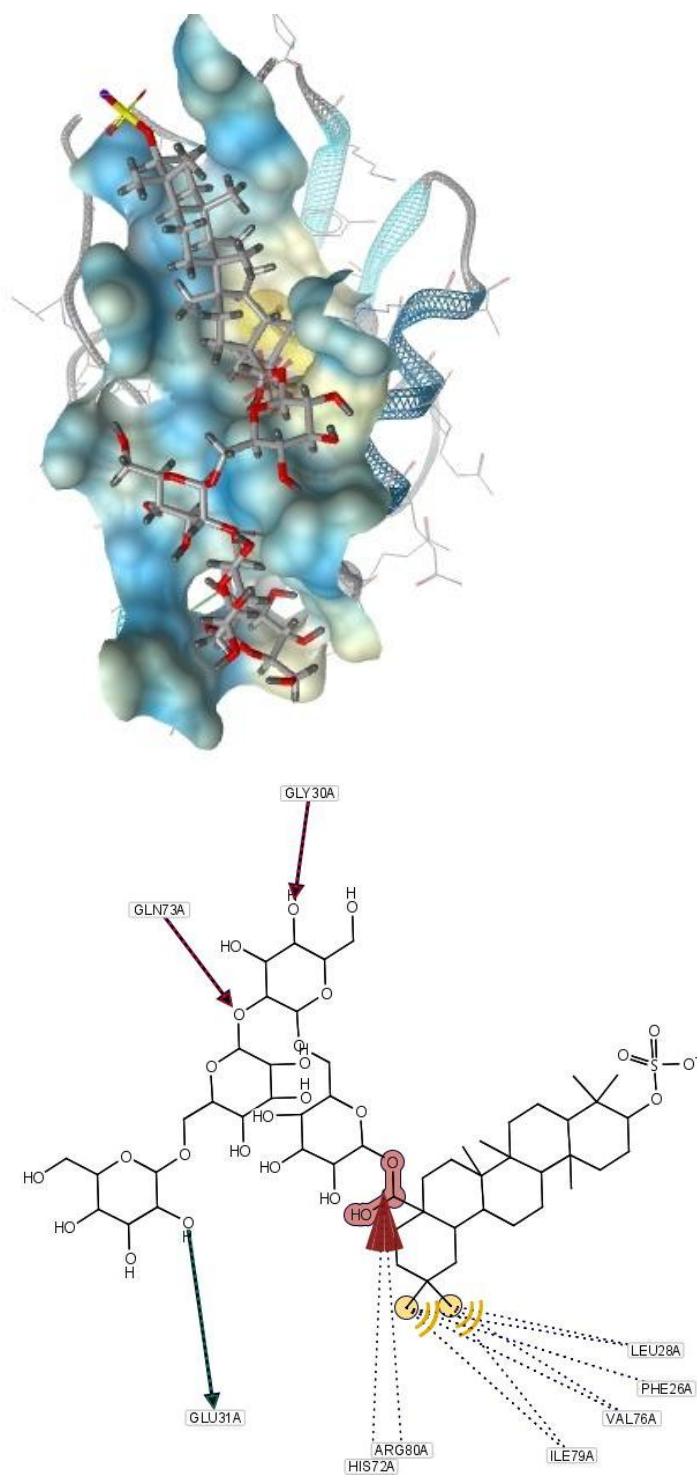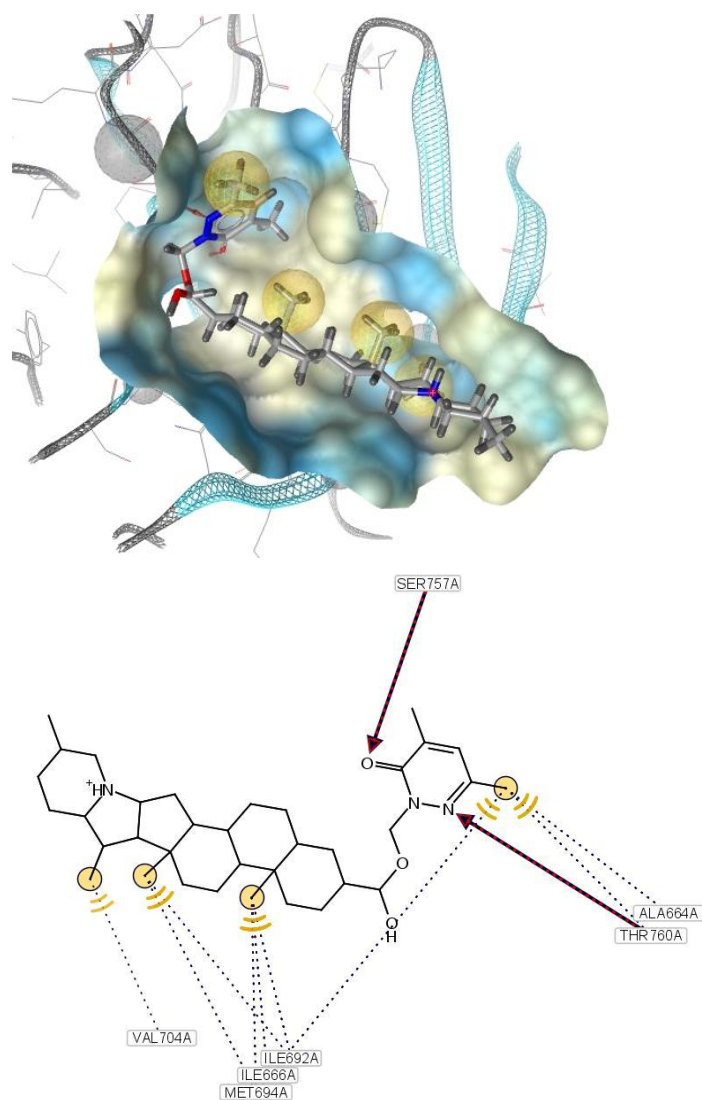
Figure 46 3-D and 2-D Pharmacophore models of a derived Stereoidal alkaloid in the binding pocket of SPRY domain generated by Ligand Scout

Chapter 5

5. CONCLUSION

5.1. PART 1: IDENTIFYING SMALL MOLECULE INHIBITORS TARGETTING BASE EXCISION REPAIR ENZYMES, DNA POLYMERASE γ AND β

Most cancer treatments, chemoradiation and most drugs used for cancer therapy, induce DNA damage, which is primarily repaired by the base excision repair (BER) pathway which is an important DNA repair mechanism. Thus, creating a blockade of BER pathway is an attractive option to increase the effects of cancer treatments. In this study plant based small molecule inhibitors were identified for BER proteins, mitochondrial DNA polymerase γ and nuclear DNA polymerase β using structural based drug design techniques.

First of all, a plant based molecules library that consists of plant secondary metabolites were constructed to be used in structure based screening. The molecules were obtained from various online available databases. 3-D coordinates of molecules were prepared in different file formats to be enabled to use in different molecular docking softwares.

Secondly, important residues and active sites of proteins were found from previous studies for decision of binding pockets and molecular docking. The hit molecules were searched by virtual screening of the Plant Based Molecules Library using the docking software Autodock Vina. Hit molecules were selected based on the binding energies obtained from VINA. Three molecules from KEGG Phytochemicals Database, Fucax23, Xntmf27 and Dxbd20, were identified as promising inhibitors of pol γ, with -10.7 kcal/mol, -9.8 kcal/mol and -9.8 kcal/mol binding free energies,

respectively. Anxbd47 was identified as small molecule inhibitor for pol β with -10.7 kcal/mol binding free energy.

Additionally, hit molecules were analyzed to model pharmacophores. All three drug candidates for pol γ were observed to have many hydrophobic interactions with the active residues in the binding pocket of accessory subunit of pol γ which indicate their potential to inhibit the enzyme. The hit molecule for pol β was observed to have interactions with most of the active residues in the binding pocket of lyase domain that mediates the lyase activity, 5'phosphate recognition in gapped DNA and ssDNA binding which determines the activity of the enzyme.

As a result, four plant secondary metabolites were identified as potential drugs for mitochondrial DNA polymerase γ and nuclear DNA polymerase β for inhibition of the BER which is a new therapautical approach. Whereas the active site of the enzyme was targeted in the case of pol β, inhibition of pol γ by targeting the interaction site between the subunits was a novel approach which might provide insights in drug design. As future work, in vivo and in vitro experiments should be conducted to verify the activities of these molecules and to understand the enzyme kinetics.

5.2.   PART 2: IDENTIFYING SPESIFIC SECONDARY METABOLITE CLASSES AS INHIBITORS FOR DOMAINS INVOLVED IN DISEASES INCLUDING MANY CANCERS

Interest in knowledge about secondary metabolite spesificity of molecular targets is increasing due to its importance and possible many advantages for rational drug design. In this study, secondary metabolite class spesificity of PDZ, BROMO and SPRY domains were investigated and 3 spesific secondary metabolite classes were identified for these domains. The three systems, PDZ, BROMO and SPRY domains were investigated due to their important drug target potential. They are found in many

proteins and take role in various critical protein-protein interactions which are involved in several cancer types.

Firstly, a well classified plant secondary metabolites library, KEGG Phytochemicals were screened for PDZ, BROMO and SPRY domains by using docking software VINA for identification of the promising classes out of four main classes Phenylpropanoids, Polyketides, Terpenoids, Alkaloids and their many subclasses. The results were analyzed on the basis of high binding scorings. Stilbenoids and Flavonoids from Phenylpropanoids, Triterpenoids from terpenoids, Anthroquinones from polyketides and Stereoidal alkaloids from alkaloids were identified as promising classes with the lowest average binding free energies.

Secondly, a more systematic method was developed in order to investigate the spesificities further and to obtain more exact results. For this purpose, these classes were derived to obtain equal and sufficient number of molecules for each. After the screening of derived compounds of classes, Gibbs Distribution was used to analyze the results. Gibbs Distribution is a function for probability measure of states depending on the energies of the states. This probability measure was applied to our system which is composed of a domain and ligands from different classes. Binding probabilities of molecules were obtained according to Gibbs Distribution using binding free energies. Finally, three classes Triterpenoids, Stilbenoids and Stereoidal alkaloids were identified for PDZ, BROMO and SPRY domain, respectively.

Additionaly, the molecules with the highest binding affinities were analyzed to model pharmacophores for each domain. The ligands were observed to have interactions with the important binding residues which were discussed in the literature.

As a result, secondary metabolites-domains specfitiy was investigated and a novel calculation approach was developed to identify this relation. Potent inhibitor

classes, Triterpenoids, Stilbenoids and Terpenoid alkaloids were found for three important domains PDZ, BROMO and SPRY.

This method and information can provide valuable source to drug design with the knowledge of such class-domain specifity and pharmacological activities of these classes. Since PDZ and BROMO domains are critical and are major players in many cancers such as breast, colon, pancreatic, colon, prostate and leukemia, this study has shown the pharmacological activities of Triterpenoids and Stilbenoids as anti-cancer and anti-tumor agents. Similarly, finding Stereoidal alkaloids as candidate inhibitors for the SPRY domain will provide insights for understanding the mechanism of the SPRY domain and FMF which is directly related with the mutations of this domain. It provides the advantage of focusing on specific classes and decreases the computational cost for screening all the molecule libraries. As future work, various further target-domain relations can be found which will serve the field of virtual screening. With the knowledge of these relations pharmacological activity of the classes can be identified through specific targets of many diseases.

6. BIBLIOGRAPHY

1. Wallace, S.S., *Biological consequences of free radical-damaged DNA bases.* Free Radic Biol Med, 2002. **33**(1): p. 1-14.
2. Wilson, D.M., 3rd and V.A. Bohr, *The mechanics of base excision repair, and its relationship to aging and disease.* DNA Repair (Amst), 2007. **6**(4): p. 544-59.
3. Trivedi, R.N., et al., *The role of base excision repair in the sensitivity and resistance to temozolomide-mediated cell death.* Cancer Res, 2005. **65**(14): p. 6394-400.
4. Wang, D. and S.J. Lippard, *Cellular processing of platinum anticancer drugs.* Nat Rev Drug Discov, 2005. **4**(4): p. 307-20.
5. Preston, T.J., et al., *Base excision repair of reactive oxygen species-initiated 7,8-dihydro-8-oxo-2'-deoxyguanosine inhibits the cytotoxicity of platinum anticancer drugs.* Mol Cancer Ther, 2009. **8**(7): p. 2015-26.
6. Hegde, M.L., T.K. Hazra, and S. Mitra, *Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells.* Cell Res, 2008. **18**(1): p. 27-47.
7. de Souza-Pinto, N.C., et al., *Mitochondrial DNA, base excision repair and neurodegeneration.* DNA Repair (Amst), 2008. **7**(7): p. 1098-109.
8. Kucej, M. and R.A. Butow, *Evolutionary tinkering with mitochondrial nucleoids.* Trends Cell Biol, 2007. **17**(12): p. 586-92.
9. Richter, C., J.W. Park, and B.N. Ames, *Normal oxidative damage to mitochondrial and nuclear DNA is extensive.* Proc Natl Acad Sci U S A, 1988. **85**(17): p. 6465-7.
10. Hudson, E.K., et al., *Age-associated change in mitochondrial DNA damage.* Free Radic Res, 1998. **29**(6): p. 573-9.
11. Ueta, E., et al., *Enhancement of apoptotic damage of squamous cell carcinoma cells by inhibition of the mitochondrial DNA repairing system.* Cancer Sci, 2008. **99**(11): p. 2230-7.
12. KEGG, K.E.o.G.a.G., *KEGG Phythochemicals.*
13. Analytcon, D., *Analyticon-discovery MEGxp Pure Natural Plant Products available at http://www.ac-discovery.com/content/Products&Technologies/MEGAbolite.php.*
14. Davis, G.D. and A.H. Vasanthi, *Seaweed metabolite database (SWMD): A database of natural compounds from marine algae.* Bioinformation, 2011. **5**(8): p. 361-4.
15. INDOFINE, C.C., *Flavonoids and Coumarins available at http://www.indofinechemical.com/products/flavonoids-coumarins.aspx.*
16. Vetrivel, U., N. Subramanian, and K. Pilla, *InPACdb--Indian plant anticancer compounds database.* Bioinformation, 2009. **4**(2): p. 71-4.
17. Pelletier, H., et al., *Crystal structures of human DNA polymerase beta complexed with DNA: implications for catalytic mechanism, processivity, and fidelity.* Biochemistry, 1996. **35**(39): p. 12742-61.
18. Prasad, R., et al., *Functional analysis of the amino-terminal 8-kDa domain of DNA polymerase beta as revealed by site-directed mutagenesis. DNA binding and 5'-deoxyribose phosphate lyase activities.* J Biol Chem, 1998. **273**(18): p. 11121-6.
19. Mizushina, Y., et al., *A plant phytotoxin, solanapyrone A, is an inhibitor of DNA polymerase beta and lambda.* J Biol Chem, 2002. **277**(1): p. 630-8.
20. Pelletier, H., et al., *Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP.* Science, 1994. **264**(5167): p. 1891-903.
21. Young-Sam Lee, W.D.K., and Y. Whitney Yin, *Structural insights into human mitochondrial DNA replication and disease-related polymerase mutations.* Cell. 2009 October 16; 139(2): 312–324.

22.   Bailey, C.M. and K.S. Anderson, *A mechanistic view of human mitochondrial DNA polymerase gamma: providing insight into drug toxicity and mitochondrial disease.* Biochim Biophys Acta, 2010. **1804**(5): p. 1213-22.

23.   Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-61.

24.   Pelletier, H. and M.R. Sawaya, *Characterization of the metal ion binding helix-hairpin-helix motifs in human DNA polymerase beta by X-ray structural analysis.* Biochemistry, 1996. **35**(39): p. 12778-87.

25.   Pelletier, H., et al., *A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase beta.* Biochemistry, 1996. **35**(39): p. 12762-77.

26.   Yakubovskaya, E., et al., *Functional human mitochondrial DNA polymerase gamma forms a heterotrimer.* J Biol Chem, 2006. **281**(1): p. 374-82.

27.   Keeling, C.I. and J. Bohlmann, *Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens.* New Phytol, 2006. **170**(4): p. 657-75.

28.   Alan Crozier, M.N.C., Hiroshi Ashihara, *Plant Secondary Metabolites Occurrence, Structure and Role in the Human Diet 2006*

29.   Thomas M. Ehrman, D.J.B., and Peter J. Hylands, *Phytochemical Databases of Chinese Herbal Constituents and Bioactive Plant Compounds with Known Target Specificities.* J. Chem. Inf. Model. 2007, 2006. **47**: p. 254-263.

30.   Joy, P.P., Thomas, J., Mathew, S., and Skaria, B.P. (eds. Bose, T.K., Kabir, J., and P.a.J. Das, P.P.). Naya Prokash, Calcutta, *Medicinal Plants* Tropical Horticulture 2001. **2** p. 449-632.

31.   *Traditional Chinese Medicines 2.1; CambridgeSoft Inc. Cambridge, MA*.

32.   Leung S.; Chan, K.S., *China Natural Products Database; NeoTrident Technology Ltd. Beijing, P. R. China.*

.

33.   Vidler, L.R., et al., *Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites.* J Med Chem, 2012.

34.   K.Dev, K., *MAKING PROTEIN INTERACTIONS DRUGGABLE:TARGETING PDZ DOMAINS.* Nature Reviews, 2004. **3**: p. 1047-1056.

35.   Shan, J., et al., *Identification of a specific inhibitor of the dishevelled PDZ domain.* Biochemistry, 2005. **44**(47): p. 15495-503.

36.   Jae-Sung Woo1, J.-H.I., Chang-Ki Min1, Kyung-Jin Kim2, Sun-Shin Cha2 and Byung-Ha Oh1,, *Structural and functional insights into the B30.2/SPRY domain.* The EMBO Journal, 2006. **25**: p. 1353-1363.

37.   Sasaki, M., et al., *PCD1, a novel gene containing PDZ and LIM domains, is overexpressed in human breast cancer and linked to lymph node metastasis.* Anticancer Res, 2003. **23**(3B): p. 2717-21.

38.   Furuya, M., et al., *A novel gene containing PDZ and LIM domains, PCD1, is overexpressed in human colorectal cancer.* Anticancer Res, 2002. **22**(6C): p. 4183-6.

39.   Kang, S., et al., *PCD1, a novel gene containing PDZ and LIM domains, is overexpressed in several human cancers.* Cancer Res, 2000. **60**(18): p. 5296-302.

40.   Tam, C.W., et al., *Inhibition of prostate cancer cell growth by human secreted PDZ domain-containing protein 2, a potential autocrine prostate tumor suppressor.* Endocrinology, 2006. **147**(11): p. 5023-33.

41.   Denis, G.V., *Bromodomain coactivators in cancer, obesity, type 2 diabetes, and inflammation.* Discov Med, 2010. **10**(55): p. 489-99.

42.   Santillan, D.A., et al., *Bromodomain and histone acetyltransferase domain specificities control mixed lineage leukemia phenotype.* Cancer Res, 2006. **66**(20): p. 10032-9.

43. Christopher Weinert, C.G., Heidi Roschitzki-Voser, Peer R. E. Mittl and Markus G. Grütter, *The Crystal Structure of Human Pyrin B30.2 Domain: Implications for Mutations Associated with Familial Mediterranean Fever.* Journal of Moleculer Biology, 2009. **394**: p. 226-236.

44. Zheng, H.-J.L.a.J.J., *PDZ domains and their binding partners: structure, specificity, and modification.* Cell Communication and Signaling, 2010. **8**.

45. Peterson, T.A., et al., *DMDM: domain mapping of disease mutations.* Bioinformatics, 2010. **26**(19): p. 2458-9.

46. Owen, D.J., et al., *The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p.* EMBO J, 2000. **19**(22): p. 6141-9.

47. Lockless, S.W. and R. Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families.* Science, 1999. **286**(5438): p. 295-9.

48. Jelen, F., et al., *PDZ domains - common players in the cell signaling.* Acta Biochim Pol, 2003. **50**(4): p. 985-1017.

49. Doyle, D.A., et al., *Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ.* Cell, 1996. **85**(7): p. 1067-76.

50. BLAST: Basic Local Alignment Search Tool, B.

51. Kokoska, R.J., S.D. McCulloch, and T.A. Kunkel, *The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase eta and Sulfolobus solfataricus Dpo4.* J Biol Chem, 2003. **278**(50): p. 50537-45.

52. Ataian, Y. and J.E. Krebs, *Five repair pathways in one context: chromatin modification during DNA repair.* Biochem Cell Biol, 2006. **84**(4): p. 490-504.

53. Zhang, H., et al., *Targeting human 8-oxoguanine DNA glycosylase (hOGG1) to mitochondria enhances cisplatin cytotoxicity in hepatoma cells.* Carcinogenesis, 2007. **28**(8): p. 1629-37.

54. Mitra, S., et al., *Complexities of DNA base excision repair in mammalian cells.* Mol Cells, 1997. **7**(3): p. 305-12.

55. Matsumoto, Y. and K. Kim, *Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair.* Science, 1995. **269**(5224): p. 699-702.

56. Srivastava, D.K., et al., *Mammalian abasic site base excision repair. Identification of the reaction sequence and rate-determining steps.* J Biol Chem, 1998. **273**(33): p. 21203-9.

57. Kunkel, T.A. and D.A. Erie, *DNA mismatch repair.* Annu Rev Biochem, 2005. **74**: p. 681-710.

58. Overdevest, J.B., D. Theodorescu, and J.K. Lee, *Utilizing the molecular gateway: the path to personalized cancer management.* Clin Chem, 2009. **55**(4): p. 684-97.

59. Helleday T, P.E., Lundin C, Hodgson B, Sharma RA., *DNA repair pathways as targets for cancer therapy.* Nat Rev Cancer, 2008. **Mar;8(3):193-204**.

60. Liu, L. and S.L. Gerson, *Therapeutic impact of methoxyamine: blocking repair of abasic sites in the base excision repair pathway.* Curr Opin Investig Drugs, 2004. **5**(6): p. 623-7.

61. Adhikari, S., et al., *Targeting base excision repair for chemosensitization.* Anticancer Agents Med Chem, 2008. **8**(4): p. 351-7.

62. Kaina, B., et al., *BER, MGMT, and MMR in defense against alkylation-induced genotoxicity and apoptosis.* Prog Nucleic Acid Res Mol Biol, 2001. **68**: p. 41-54.

63. Trivedi, R.N., et al., *Human methyl purine DNA glycosylase and DNA polymerase beta expression collectively predict sensitivity to temozolomide.* Mol Pharmacol, 2008. **74**(2): p. 505-16.

64. Luo, M. and M.R. Kelley, *Inhibition of the human apurinic/apyrimidinic endonuclease (APE1) repair activity and sensitization of breast cancer cells to DNA alkylating agents with lucanthone.* Anticancer Res, 2004. **24**(4): p. 2127-34.

65. Jaiswal, A.S., et al., *A novel inhibitor of DNA polymerase beta enhances the ability of temozolomide to impair the growth of colon cancer cells.* Mol Cancer Res, 2009. **7**(12): p. 1973-83.

66. Beard, W.A., et al., *Efficiency of correct nucleotide insertion governs DNA polymerase fidelity.* J Biol Chem, 2002. **277**(49): p. 47393-8.

67.   Wilson, S.H., et al., *Base excision repair and design of small molecule inhibitors of human DNA polymerase beta.* Cell Mol Life Sci, 2010. **67**(21): p. 3633-47.

68.   Sawaya, M.R., et al., *Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism.* Biochemistry, 1997. **36**(37): p. 11205-15.

69.   Srivastava, D.K., et al., *DNA polymerase beta expression differences in selected human tumors and cell lines.* Carcinogenesis, 1999. **20**(6): p. 1049-54.

70.   Bergoglio, V., et al., *Deregulated DNA polymerase beta induces chromosome instability and tumorigenesis.* Cancer Res, 2002. **62**(12): p. 3511-4.

71.   Bergoglio, V., et al., *Enhanced expression and activity of DNA polymerase beta in human ovarian tumor cells: impact on sensitivity towards antitumor agents.* Oncogene, 2001. **20**(43): p. 6181-7.

72.   Canitrot, Y., et al., *Overexpression of DNA polymerase beta in cell results in a mutator phenotype and a decreased sensitivity to anticancer drugs.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12586-90.

73.   Lee, Y.S., W.D. Kennedy, and Y.W. Yin, *Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations.* Cell, 2009. **139**(2): p. 312-24.

74.   Lee, Y.S., et al., *Each monomer of the dimeric accessory protein for human mitochondrial DNA polymerase has a distinct role in conferring processivity.* J Biol Chem, 2010. **285**(2): p. 1490-9.

75.   Croteau, R., Kutchan, T.M. and Lewis, N.G., *Natural products (secondary metabolites. In B.B. Buchannan, W. Gruissem and R.L. Jones (eds), .* Biochemistry and Molecular Biology of Plant Physiol, 2000: p. 1250–1318.

76.   Dewick, P.M., *Medicinal Natural Products: A Biosynthetic Approach, 2nd edn., JohnWiley and Sons, Chichester.* 2002.

77.   *Farnsworth, N.R. et al. (1985) WHO Bull. 63, 965–981.*

78.   *Baker, J.T. et al. (1995) J. Nat. Prod. 58, 1325–1357.*

79.   Rochfort, S. and J. Panozzo, *Phytochemicals for health, the role of pulses.* J Agric Food Chem, 2007. **55**(20): p. 7981-94.

80.   Rix, U. and G. Superti-Furga, *Target profiling of small molecules by chemical proteomics.* Nat Chem Biol, 2009. **5**(9): p. 616-24.

81.   Rix, U., et al., *A comprehensive target selectivity survey of the BCR-ABL kinase inhibitor INNO-406 by kinase profiling and chemical proteomics in chronic myeloid leukemia cells.* Leukemia, 2010. **24**(1): p. 44-50.

82.   Bantscheff, M., A. Scholten, and A.J. Heck, *Revealing promiscuous drug-target interactions by chemical proteomics.* Drug Discov Today, 2009. **14**(21-22): p. 1021-9.

83.   Slanina, J. and Z. Glatz, *Separation procedures applicable to lignan analysis.* J Chromatogr B Analyt Technol Biomed Life Sci, 2004. **812**(1-2): p. 215-29.

84.   Pan, J.Y., et al., *An update on lignans: natural products and synthesis.* Nat Prod Rep, 2009. **26**(10): p. 1251-92.

85.   Saleem, M., et al., *An update on bioactive plant lignans.* Nat Prod Rep, 2005. **22**(6): p. 696-716.

86.   Adlercreutz, H., *Lignans and human health.* Crit Rev Clin Lab Sci, 2007. **44**(5-6): p. 483-525.

87.   Harborne, D.J. and J. Worrell, *Accident and emergency in London. Good primary care reduces workload.* BMJ, 1993. **306**(6894): p. 1752.

88.   *Koes, R.E., Quattrocchio, F. and Mol, J.N.M. (1994) The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays, 16, 123–132.*

89.   Buckingham, J., *Dictionary of Natural Products, Web Version 2004. London: Chapman and Hall. Available at: http:// www.chemnetbase.com; last accessed 9 March 2008.*

90.   Bohlmann, J. and C.I. Keeling, *Terpenoid biomaterials.* Plant J, 2008. **54**(4): p. 656-69.

91.   Kennedy, D.O. and E.L. Wightman, *Herbal extracts and phytochemicals: plant secondary metabolites and the enhancement of human brain function.* Adv Nutr, 2011. **2**(1): p. 32-50.

92. *Zulak K, Liscombe D, Ashihara H, Facchini P. Alkaloids. Plant secondary metabolism in diet and human health. Oxford: Blackwell Publishing; 2006. p. 102–36.*

93. Zenk, M.H. and M. Juenger, *Evolution and current status of the phytochemistry of nitrogenous compounds.* Phytochemistry, 2007. **68**(22-24): p. 2757-72.

94. Arkin, M.R. and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing towards the dream.* Nat Rev Drug Discov, 2004. **3**(4): p. 301-17.

95. Berg, T., *Modulation of protein-protein interactions with small organic molecules.* Angew Chem Int Ed Engl, 2003. **42**(22): p. 2462-81.

96. Morais Cabral, J.H., et al., *Crystal structure of a PDZ domain.* Nature, 1996. **382**(6592): p. 649-52.

97. Nourry, C., S.G. Grant, and J.P. Borg, *PDZ domain proteins: plug and play!* Sci STKE, 2003. **2003**(179): p. RE7.

98. Levi, M., *Role of PDZ domain-containing proteins and ERM proteins in regulation of renal function and dysfunction.* J Am Soc Nephrol, 2003. **14**(7): p. 1949-51.

99. Tonikian, R., et al., *A specificity map for the PDZ domain family.* PLoS Biol, 2008. **6**(9): p. e239.

100. Zaslavsky, E., P. Bradley, and C. Yanover, *Inferring PDZ domain multi-mutant binding preferences from single-mutant data.* PLoS One, 2010. **5**(9): p. e12787.

101. Chaib, H., et al., *Activated in prostate cancer: a PDZ domain-containing protein highly expressed in human primary prostate tumors.* Cancer Res, 2001. **61**(6): p. 2390-4.

102. Fujii, N., et al., *A selective irreversible inhibitor targeting a PDZ protein interaction domain.* J Am Chem Soc, 2003. **125**(40): p. 12074-5.

103. Migaud, M., et al., *Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein.* Nature, 1998. **396**(6710): p. 433-9.

104. Patra, C.R., et al., *Chemically modified peptides targeting the PDZ domain of GIPC as a therapeutic approach for cancer.* ACS Chem Biol, 2012. **7**(4): p. 770-9.

105. Karthikeyan, S., T. Leung, and J.A. Ladias, *Structural basis of the Na+/H+ exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator.* J Biol Chem, 2001. **276**(23): p. 19683-6.

106. Filippakopoulos, P., et al., *Histone recognition and large-scale structural analysis of the human bromodomain family.* Cell, 2012. **149**(1): p. 214-31.

107. Muller, S., P. Filippakopoulos, and S. Knapp, *Bromodomains as therapeutic targets.* Expert Rev Mol Med, 2011. **13**: p. e29.

108. Zuber, J., et al., *RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia.* Nature, 2011. **478**(7370): p. 524-8.

109. Nicodeme, E., et al., *Suppression of inflammation by a synthetic histone mimic.* Nature, 2010. **468**(7327): p. 1119-23.

110. Ciro, M., et al., *ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors.* Cancer Res, 2009. **69**(21): p. 8491-8.

111. Tsai, W.W., et al., *TRIM24 links a non-canonical histone signature to breast cancer.* Nature, 2010. **468**(7326): p. 927-32.

112. Khetchoumian, K., et al., *Loss of Trim24 (Tif1alpha) gene function confers oncogenic activity to retinoic acid receptor alpha.* Nat Genet, 2007. **39**(12): p. 1500-6.

113. Grunwald, C., et al., *Expression of multiple epigenetically regulated cancer/germline genes in nonsmall cell lung cancer.* Int J Cancer, 2006. **118**(10): p. 2522-8.

114. Dey, A., et al., *The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8758-63.

115. Bateman, A., et al., *The Pfam protein families database.* Nucleic Acids Res, 2004. **32**(Database issue): p. D138-41.

116. Schultz, J., et al., *SMART, a simple modular architecture research tool: identification of signaling domains.* Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5857-64.

117. Henry, J., et al., *B30.2-like domain proteins: update and new insights into a rapidly expanding family of proteins.* Mol Biol Evol, 1998. **15**(12): p. 1696-705.
118. Bakkaloglu, A., *Familial Mediterranean fever.* Pediatr Nephrol, 2003. **18**(9): p. 853-9.
119. DS, *Accelrys Discovery Studio Visualizer 3.5 available from http://accelrys.com/products/discovery-studio/.*
120. MGL, M.G.L., *MGL Tools available at http://mgltools.scripps.edu/.*
121. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* J Cheminform, 2011. **3**: p. 33.
122. *Huey, R., et al., A semiempirical free energy force field with charge-based desolvation. J Comput Chem, 2007. 28(6): p. 1145-52.*
123. Wolber, G.a.T.L., *LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model, 2005. 45(1): p. 160-9.*
124. RSCB, *Protein Data Bank available from http://www.rcsb.org/pdb/home/home.do.*
125. BROOD, *OEChem, version 1.7.4, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, 2010.*
126. Widom, B., *Statistical Mechanics: A Concise Introduction for Chemists*: Cambridge University Press.
127. VIDA, *OEChem, version 1.7.4, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, 2010.*
128. Privat, C., et al., *Antioxidant properties of trans-epsilon-viniferin as compared to stilbene derivatives in aqueous and nonaqueous media.* J Agric Food Chem, 2002. **50**(5): p. 1213-7.
129. Li, S. and M.A. Shogren-Knaak, *The Gcn5 bromodomain of the SAGA complex facilitates cooperative and cross-tail acetylation of nucleosomes.* J Biol Chem, 2009. **284**(14): p. 9411-7.
130. Yang, S.A., et al., *Alpha-chaconine, a potato glycoalkaloid, induces apoptosis of HT-29 human colon cancer cells through caspase-3 activation and inhibition of ERK 1/2 phosphorylation.* Food Chem Toxicol, 2006. **44**(6): p. 839-46.

VITA

Derya Aydın was born in Istanbul (Turkey) in 1985. She had been to İstanbul Atatürk Science High School for her high school education She received her Bachelor of Science Degree in Food Engineering from Hacettepe University. In September 2010 she attended Copmutational Sciences and Engineering M.S. program at Koç University. From 2010 to 2012, she worked as a research and teaching assistant at the same institution.