1-1-2019

Cheminformatic Approach for Deconvolution of Active Compounds in a Complex Mixture - phytoserms in Licorice

Manal Mohammad Alhusban

CHEMINFORMATIC APPROACH FOR DECONVOLUTION OF ACTIVE COMPOUNDS IN A COMPLEX MIXTURE - PHYTOSERMS IN LICORICE

A Dissertation
Presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in Pharmaceutical Sciences, Emphasis in Pharmacognosy
The University of Mississippi

Manal M. Alhusban

August 2019

ABSTRACT

The genus *Glycyrrhiza,* encompasses several species exhibiting complex structural diversity of secondary metabolites and hence biological activities. The intricate nature of botanical remedies, such as licorice, rendered them obsolete for scientific research or medical industry. Understanding and finding the mechanisms of efficacy or safety for a plant-based therapy is very challenging, yet it remains crucial and warranted.

The licorice plant is known to have Selective Estrogen Receptor Modulatory effects (SERMs), with a spectrum of estrogenic and anti-estrogenic activities attributed to women's health. On the contrary, licorice extract was shown to induce pregnane xenobiotic receptor (PXR), which may manifest as a potential route for deleterious effects such as herb-drug interaction (HDI). While many studies attributed these divergent activities to a few classes of compounds such as liquiritigenin (a weak estrogenic SERM) or glycyrrhizin (weak PXR agonist), no attempt was made to characterize the complete set of compounds responsible for these divergent activities. A plethora of licorice components is undermined, which might have the potential to be developed into novel phytoSERMS or to trigger undesirable adverse effects by altering drug metabolizing enzymes and thus pharmacokinetics.

In this work, we explored the mechanism associated with the efficacy and safety of components reported in the licorice plant. We utilized smart screening techniques such as cheminformatics tools to reveal the high number of secondary metabolites produced by licorice, which are capable of interfering with the human Estrogen Receptors (*h*ERs) and/or PXR or other

vital cytochrome P450 enzymes.

After the validation of our *in silico* models by using the previous knowledge in this area, the alerting phytochemicals from two *Glycyrrhiza* species (*G. glabra and G. uralensis*) were clustered. Exhaustive computational mining of licorice metabolome against selected endocrinal and metabolic targets led to the discovery of a unique class of compounds, which belong to the dihydrostilbenoids (DHS) class appended with prenyl groups at various positions. To the best of our knowledge, this interesting group of compounds has not been studied for their estrogenic activities or PXR activation. In addition, some of the bis-prenylated DHS have been reported to be present only in *G. uralensis*.

Thus, we have ventured to synthesize a set of constitutional isomers of stilbenoids and DHS (archetypal of those found in licorice) with different prenylation patterns. Sixteen constitutional isomers of stilbenoids (**M2-M10**) and DHS (**M12-M18**) were successfully synthesized, of which six of them (**M8**, **M9**, **M14**, **M15**, **M17**, **and M18**) are synthesized for the first time to be further tested and validated with cell-based methods for their estrogenic activities.

We have unveiled a novel class of compounds, which possess a strong PXR activation. These results, which were in accord with the *in silico* prediction, were observed for multiple synthesized prenylated stilbenoid and DHS by the luciferase reporter gene assay at μM concentrations. Moreover, this activation was further validated by the six-fold increase in mRNA expression of Cytochrome P450 3A4 (CYP3A4), where three representative compounds (**M7, M10, and M15**) exceeded the activation fold of the positive control.

Another aspect of the current project was to predict the phase I primary metabolites of

compounds found in both species of *Glycyrrhiza* and assess them with computational tools to predict their binding potential against both isoforms of *h*ERs or drug metabolizing enzymes such as (CYP) inhibition models. Our investigations revealed estrogenic character for most of the predicted metabolites and have confirmed earlier reports of potential CYP3A4 and CYP1A2 inhibition.

Compilation of such data is essential to gain a better understanding of the efficacy/safety of licorice extracts used in various botanical formularies. This approach with the involved cheminformatic tools has proven effective to yield rich information to support our understanding of traditional practices. It also can expand the role of botanical drugs for introducing new chemical entities (NCEs) and/or uncovering their liabilities at early stages.

DEDICATION

My Mother Fawzeya

Who taught me how to dream, imagine and work hard

My Father Mohammad

Who taught me how to concentrate on the goals, not obstacles

My Husband Ali

Who taught me that there are no limits except the limits in our minds

My Kids Ghaith, Ra'ad and Salma

Who suffered with me the most

LIST OF ABBREVIATIONS AND SYMBOLS

μL	Microliter
μM	Micromolar
AF-2	Active Functional Region 2
AKT	Protein Kinase B
Al_2O_3	Alumina
CAR	Constitutive Androstane Receptor
CART	Classification and Regression Tree Algorithm
CERAPP	Collaborative Estrogen Receptor Activity Prediction database
CYP3A4	Cytochrome P450 3A4
DBD	DNA-Binding Domain
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
DHS	Dihydrostilbenoids
DMF	Dimethylformamide
EC_{50}	The concentration at the half maximum efficacy
EtOAc	Ethyl Acetate
EtOH	Ethanol
US FDA	US Food and Drug Administration
GST	Glutathione S-Transferase

HDI	Herb-Drug Interaction
*h*ERs	Human Estrogen Receptors
Hex	Hexanes
HMBC	Heteronuclear Multiple-Bond Correlation Spectroscopy
hrs.	Hours
LBD	Ligand-Binding Domain
m/z	Mass-to-Charge Ratio
MDR1	Phase III P-Glycoprotein or Multidrug Resistance-Associated Protein 1
MEMCl	2-Methoxyethoxymethyl Chloride
MeOH	Methanol
mL	Milliliter
MOMCl	Methoxymethyl Chloride
NaH	Sodium Hydride
*n*BuLi	*n*-Butyl Lithium
NCEs	New Chemical Entities
NFkβ	Nuclear Factor kappa-Light-Chain-Enhancer of Activated B Cells
NOESY	Nuclear Overhauser Effect Spectroscopy
NMR	Nuclear Magnetic Resonance
NR1I2	Nuclear Receptor Subfamily 1 Group I
OATP2	Organic Anion Transporting Polypeptide 2

OCHEM	Online Chemical Environment
PDB	Protein Data Bank
PXR	Pregnane Xenobiotic Receptor
phytoSERMs	Phytochemicals that act as Selective Estrogen Receptor Modulators
QSAR	Quantitative Structure-Activity Relationships
RBA	Relative Binding Affinity
RMSD	Root-Mean-Square Deviation
r.t.	Room Temperature
SDF	Spatial Data Files
SERMs	Selective Estrogen Receptor Modulators
Sult	Sulfotransferases
TEA	Triethyl Amine
THF	Tetrahydrofuran
TLC	Thin Layer Chromatography
UGTs	Uridine Diphosphate (UDP)-Glucuronosyltransferases
VDR	Vitamin D Receptor
VEGA	Virtual Models for Property Evaluation of Chemicals Within a Global Architecture

ACKNOWLEDGMENTS

I would like to express my deep gratitude for all the people who helped and supported me throughout my Ph.D. at Ole Miss. Special thanks for my advisor Dr. Ikhlas Khan, who accepted me as a member of his group and supported me and my family. He was always cautious to offer us the best environment to learn and succeed. The guidance, supervision, honesty and critique of Dr. Amar Chittiboyina, my co-advisor, and mentor made me a better scientist. His passion for exploration and discovery is contagious. Although he had a tight schedule, he never skipped a chance for discussion and explanation. When things become impossible in the lab, he would always come and teach us how to troubleshoot. I will always be grateful for him, for his concern to give me the opportunity to engage in diverse projects and for his constant encouragement, kindness, and inspiration.

I would also express my sincere thanks to Dr. Robert Doerksen, for his candid lectures and training, help, and guidance throughout the years. I am greatly indebted to him for his relentless support, mentoring and critiques, which greatly affected my academic development. I would like to thank Dr. Samir Ross for his support, advice, and kindness. To Dr. Cole Stevens for his support, effort and time for revising my ORP and dissertation. My special thanks for to Dr. John Rimoldi and Dr. Daneel Ferreira for their great lectures.

I am also thankful to all the people in NCNPR who supported me, especially, Dr. Junaid Rehman, I will always be grateful for him and his family for their continuous support and help for my family. For Dr. Ali Zulfiqar, for making NMR easily understandable. I would also thank

Dr. Shabana Khan, Dr. Jon Parcher, Dr. Bharathi Avula, and Dr. Yan-Hong Wang for their collaboration and advice, and for the post docs who trained me well in the labs, Dr. Saqlain Haider, Dr. Mohammad Albadry and Dr. Pankaj Pandey. Also I am thankful to my friends, Abidah Parveen, and Nasma for sharing all the difficulties and great moments in my Ph.D. journey.

Many thanks to Dr. Jaber and Philadelphia University in Jordan for the scholarship they have provided for me. I would also thank the financial support for this work provided by the grants from the food and drug administration (FDA-BAA, No. HHSF223201810175C): "Hollistic approach for potential drug interactions with botanical drugs" and (FDA) grant number 2U01FD004246-08: "Science Based Authentication of Botanical Ingredients."

I have special appreciation for the staff at NCNPR, particularly, Ms. Jennifer Taylor, and Steven Hopper, and in the Department of BioMolecular Sciences, Ms. Sherry Gussow, for their remarkable support, kindness and constant help. I also appreciate the Oxford School District, for offering a safe and incredible environment for my kids to learn and engage in several programs and activities during these years.

Lastly, I would thank my family members, my brother Anas and my sisters Muna, Ibtesam, and Eman, for their unconditional love and support. Also, my cousins Raeda, Bothanya, Wafa and Layla Alhusban for their love and support for my kids and me. Without your support, I would have never been able to continue in this program. Many thanks for you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF APPENDICES
SUPPLEMENTARY INFORMATION TABLES

CHAPTER I
INTRODUCTION

I.1 Natural product research remodeling: Medical research and clinical practice perspectives

The inherently complicated nature of an herbal remedy usually precludes the full understanding of its activity or safety. The polypharmacology represented by modulation of a network of targets and the infeasibility of the experimental testing of every reported secondary metabolite denotes the major hurdles in botanical research. This is coupled with the natural products chemical complexity, which imposes vast difficulties regarding their supplementation and progression in medical research. Therefore, in the past few decades, we have witnessed a biased investment into focused small molecule libraries generated by combinatorial chemistry as the mainstream for drug development. Unfortunately, the results were seemingly contradictory.[1-5]

The clinical attrition of the vast majority of the new chemical entities was devastating and described as "productivity crises."[6-7] In part, this situation has been correlated to the phasing out of NPs from the drug discovery pipeline.[8] On the other hand, one analysis revealed that 60% of the registered chemical entities in 1981 to 2010 are either natural products or natural products related molecules including NP derivatives or NP synthetic mimics.[9] Undeniably, natural products are the ideal primary resource of new chemical entities by natural selection. The interactive environment in which these natural products have evolved should have indulged them with exceedingly incomparable therapeutic potentials. Therefore, they are predicted to harbor

chemical pharmacophores with preferred interactions toward surrounding biological targets.[10-11] One interesting example is the blockbuster drug rosuvastatin, which is used for hypercholesteremia. This drug is considered as a synthetic mimic of mevastatin, a natural product from *Penicillium citrinum* fungus but with a less intricate scaffold for chemical synthesis **(Figure 1**).[12]

In addition to the above, NPs are still used as a basic therapeutic option side by side as traditional medicine worldwide. In general, there was a global drift towards the use of natural products supplements, which was driven by the perception that they are safe and efficacious.[13] Correspondingly, the sales of these supplements peaked to $7.45 billion in 2016.[14] The current cliffs where both the medical research and clinical practices stand has created a chain reaction in the natural product field. At this moment, it became a requisite to understand the drawbacks of the existing strategies and practices in this field, and to improve their competences. For instance, although coupling the biological assays to the chemical investigation of the NP was meant to rationalize the process of isolation in this field, but soon it was clear that it failed to prove their efficacy in the clinical trials in many occasions.[15-16] In fact, these practices were incomplete to draw the complete picture of how these systems operate. Some of the reasons behind these failures are summarized in **Figure 2**.[17-18]

These concomitant conclusions in both fields pointed out to the necessity of understanding the mechanism of action of NP or herbal remedy as to a priority to prove their efficacy and safety in the clinical practice and to eliminate the difficulties that undermined their engagement into the pipelines of the medical research. *In silico* investigation of natural products provides a versatile toolbox to deal with its complexity.[4] In the past few decades, virtual screening (both structure-based and ligand-based approaches) has been utilized as a basic tool to

capture novel scaffolds out of diverse chemical libraries against known pharmacological targets.[19-20]

Additionally, the growing field of genome mining, chemogenomics and molecular networking constitute a promising avenue to approach novel and effective NPs leads by application of multiple sophisticated computational tools.[21] Nevertheless, in this context, natural product libraries are only partially and passively integrated into this procedure. However, to enrich their engagement in the process they should be treated separately before their direct employment. Creating this shortcut will inspire drug design and development effectively by capturing NP hits or generating natural product fragment libraries as fingerprints of active motifs, specifically, with the accumulating knowledge and computational tools to predict the pharmacodynamics and pharmacokinetic properties of molecules along with the growing numbers of NP entries in the databases. Moreover, it would also help us to explore novel candidates as natural-product-lead-structures in a time and cost-effective manner (**Table 1.**).[4, 12, 22]

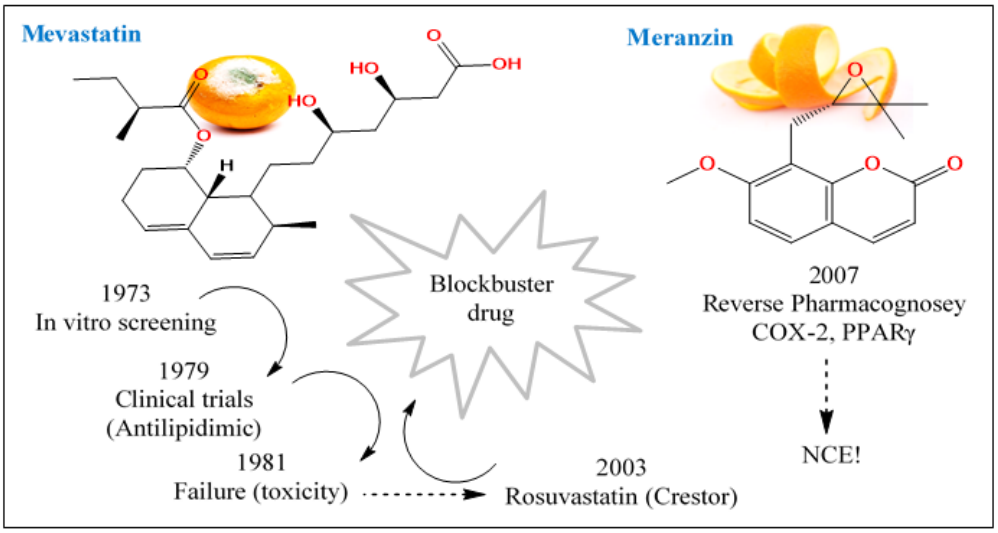


Figure 1. Natural products as a source of NCEs: Application of novel methodologies to upgrade sustainable resources

However, searching for lead compounds from NP origin is not the sole application of *in silico* procedures and virtual screening. They should be utilized for understanding the multi-targeted (polypharmacology) and off-targeted effects (toxicity) of the known plant therapy, defining the best components, which are capable of interacting with specific targets including enzymes involved in drug metabolism. The prediction of macromolecule targets for molecules of natural origin (reverse pharmacognosy), might solve the long lasted debate about the efficacy of herbal remedies. These concepts came along with the growing paradigm of reverse virtual screening or inverse docking, which is a direct consequence of thousands of accumulating solved protein crystal structures into data banks. Other approaches explored feature and similarity searches for target fishing.[23-24]

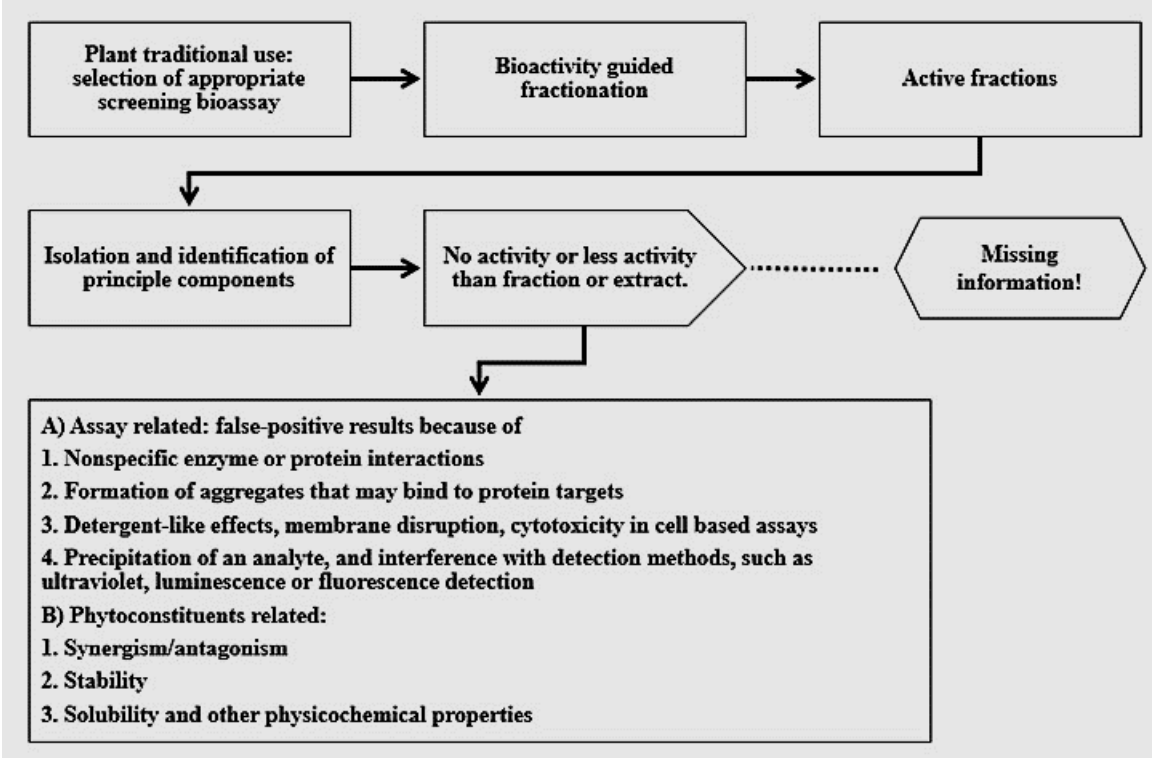


Figure 2. An overview of BGF drawbacks that might lead to sub-clinical significance of intended NP therapeutics

In the context of plant therapy, this would achieve the ultimate goal of defining a spectrum of potential activities and spotting the possible natural products candidates behind them. Conversely, it might predict or explain in early stages why such therapies end up with controversial conclusions when extrapolating to *in vivo* settings. For instance, adapting such methodologies have unveiled both peroxisome proliferator-activated receptor gamma and cyclooxygenase-2 as the main targets of Meranzin. Interestingly, the latter molecule was shown to have a potency comparable to their conventional drug therapies, rosiglitazone, and indomethacin.[20] By the application and benchmarking of such methods, we should be able to define and rationally prioritize the targets, which should be screened.

The diverse nature of plant secondary metabolites and their natural resources inconsistency has also dampened the study of their pharmacokinetic interaction with conventional therapeutics. They are the major contributors to why these effects are left underestimated and thus understudied. Although several HDIs are well known in the literature, the study of these events is still retrospective. This is quite concerning as one study showed that 4 out of 10 adults in the US engage alternative and complementary medicines in their life. Around 20% of these are natural products.[25-28] Moreover, concerning the HDI, it became clear that the NP entities play a complex role at the molecular level. Besides the direct interaction with the metabolizing enzymes and transporters, recent studies uncovered the significant role played by PXR mediating or exaggerating these effects.[13]

The recent development of a systematic approach by the FDA to prioritize the natural supplements with the high-risk potential of precipitating an HDI is still dependent on the ranking and scoring of gathered information. However, there is a clear basic need for adapting prospective methodologies.[25] In this context, the *in silico* oriented approaches, which try to

tackle the human-xenobiotic disposition system should take the lead. This trend has been motivated by the low-cost and the efficiency of such approaches, and it is aggravated by the clinical attrition of drug candidates in the clinical trials. In fact, the latter events along with several cases of FDA "approved drugs" withdrawals (because of CYP enzymes related drug-drug interaction) have escalated the need to screen for drug safety at an early stage of any drug development. This situation has created a surge of *in vitro* data measuring different metabolic endpoints such as metabolic stability or inhibition for thousands of chemical entities. In addition, the last few years have witnessed crucial advancements in the knowledge of metabolic pathways and their regulation as well as the structural and functional understanding of their operating machinery.[29-30]

In all, these events have created a momentum to develop multiple computational models to screen for different metabolic endpoints. Similar to the efficacy predictions, these computational approaches are also classified as ligand-based and structure-based models. On the grounds of reality, metabolism is among the most complex properties to predict. Unlike conventional targets, metabolic processes recruit a variety of enzymes with multiple isoforms and transporters, which are created to process a diverse group of chemical structures known as "xenobiotics".

Table 1: Selected natural product databases and number of entries

Database	**NPs entries**	**Content**
Dictionary of Natural products	40,000	NPs found in the literature
SymMAP	19,595	NPs from listed in Chinese medical texts with target prediction
SuperNatural II	326,000	NPs with the description of toxicity and target prediction
Zinc 15	80,617	NPs found in the literature with a purchasable list
CheMBL	>75,000	NPs structures from the literature
MarinLit	26,490	Marine NPs found in the literature

Hence, making them usually larger, more flexible and promiscuous.[31] Additionally, their expression is further regulated by several chemical, genetic and environmental factors. Nevertheless, assessing these properties is vital for developing NCEs as well as for herbs in order to avoid drug-drug or HDI that may lead to suboptimal dosing of accompanying drugs or even adverse effects.

Thus, the best rationale in such a scenario is "to divide and conquer." Incorporation of multiple computational models including the CYP enzymes and their regulators such as PXR or CAR, which are most involved in drug disposition, would be a better practice. Particularly the CYP enzymes, where multiple models can measure different aspects of the process such as substrate/inhibition properties, substrate selectivity, regioselectivity (the site of metabolism the expected metabolites), and the rate or extent of metabolite generation or inhibition. The vast majority of metabolism occurs via CYP enzymes, particularly, CYP3A4, which is solely responsible for the metabolism of nearly half of the prescription drugs.[31-33]

As a direct result of this discussion and to approach the pharmacokinetic and metabolism effects of herbal drug therapy with a known multicomponent system, the *in silico* models offer a good starting point to untangle the complexity of the situation.

I.2 Cheminformatic approaches to unravel the biological functions and potential safety issues associated with herbal drugs.

Application of cheminformatics to solve the complexity of an herbal remedy is enticing. Cheminformatics offers a diverse set of tools, which includes physics-based models comprising both quantum chemistry and molecular dynamics simulations. In addition, it encompasses the machine learning approaches that apply certain algorithms to recognize similar patterns and to find mathematical relationships between empirical observations. This could be applied to small

molecules in order to predict their properties (chemical, physical or biological). Furthermore, these techniques can be classified into structure-based and ligand-based approaches depending on the experimental design. Some of the commonly applied approaches include docking, pharmacophores, quantitative structure-activity relationships, similarity searches, and machine learning.

In fact, such approaches have proven to be effective in multiple research areas in medicine and drug design, which emphasizes the expansion of these terms to the phytochemical field in various applications such as activity profiling, mechanism of action and ADMET prediction including safety issues and HDI. The exponential increase in computational power represented by speed, accuracy, and cost, reinforces the importance of such techniques in every project agenda.

The following will be a brief overview of the implemented approaches in our study.

I.2.1 Structure-based approaches: Molecular Docking

The three dimensional and structural information extracted from biological targets forms a fundamental aspect of the structure-based approaches, which mainly includes molecular docking, virtual screening, and molecular dynamics. Exploring the transient interactions between small molecules and their targets became more and more feasible with the increasingly accumulating solved protein crystal structures in the databases. Moreover, the prediction of these interactions became sufficiently accurate with the advent of more sophisticated algorithms and scoring functions.[34]

The process of docking is a multistep procedure that will predict the posing of the ligand by defining specific conformation and orientation within the binding site of a certain target. The aim of this process is to predict a correct pose and to estimate the binding energy of the small

molecules toward their targets. Docking procedures involve three premises. The first one is the flexibility of small molecules. The second premise is the flexibility of the protein followed by the scoring method.[35] Small molecules usually have multiple degrees of freedom.

The conformational sampling algorithm will provide a multi-conformer database to find the best conformation, which matches the binding site. The sampling procedure could be systematic, but most often it is stochastic to comply with the high throughput nature of the docking virtual screening. This ligand sampling procedure could be a separate step, or it can be embedded within the docking protocol.

However, protein flexibility is a bit more challenging issue due to its huge nature and complexity. Docking procedures differ in the degree of freedom allowed for the binding site flexibility which could be advantageous in multiple occasions, especially where the key and lock principle fails to explain the protein-ligand interaction. Accordingly, the docking procedure could be rigid, or semi-flexible. In the first case, the protein is restricted to one conformation while in the second case the protein is treated as a soft body by relaxing the potentials of the van der Waals radii, or by considering a rotamer library for the protein side chains. In fact, these protocols may or may not adequately select a true or active conformation of the protein.

An alternative approach that we have adopted in this research is to consider multiple rigid and known active conformations of the protein-ligand complexes, which is known as ensemble docking.[23, 35-36]

For the third premise of the docking procedure, the direct evaluation and comparison between various ligand conformers inside the protein pockets are required. Most docking programs generate potential energy grids. In this context, grid points are set as reference points to store information about the binding site such as steric, electrostatic and van der Waals

potentials that will be used for scoring the ligand pose using atomic pair-wise interactions. However, the overall success of the docking procedure does not end by identifying the correct binding conformation. Rather, it will be judged for the correct pose prediction, differentiating between actives and inactives and the correct estimation of the binding energy. In fact, the latter aspect is not easily achieved in the regular setting, which relies on the scoring functions solely. Most often, the scoring function is inherently deficient for oversimplification of the energetic terms.[35, 37-39] Accordingly, to achieve true binding energy calculations both enthalpy and entropy terms such as desolvation, translational and rotational entropy of the binding process should be considered. Alternatively, quantum mechanics or mixed approaches might be adapted. Nonetheless, the scoring methods are usually sufficient to rank the small molecules according to their simplified predicted energies.[35]

I.2.2 Ligand-based approaches: Quantitative structure-activity relationships (QSAR)

Although it is reasonable to think of the protein's 3D structure as a straightforward way to get information about the ligand-binding process, this technique comes to the sheer complexity of multi-factorial drawbacks. These are related to the X-ray crystal structure itself, the nature of the target protein or to the interactive binding process. For instance, some of the major problems of crystallographic structures are their explicit waters, their resolution and other artifacts related to the crystallization process. Hence, the computational chemist should decide which waters to keep and which ones to discard.

Moreover, the inherent limitation of their resolution makes them blind to the hydrogen atoms. Then, it becomes problematic to decide if the ionizable groups inside the pocket are in their ionized form or not. Nevertheless, careful treatment and examination of these drawbacks in the structure-based approach will increase their rate of success. However, the docking process

could become more and more challenging with proteins, having large binding sites or greater flexibility. Experimentally, proteins could be heterogeneously expressed with different ligands or they could bind to multiple ligands at the same time. In addition, some of them have multiple isoforms, which mostly will create selectivity challenges.[40]

In such cases, considering single conformation of the protein, especially in a highly flexible one, would certainly undermine the dynamic process of a ligand interacting with a protein. In fact, there are multiple proposed solutions for such problems such as considering multiple protein structures, the soft docking, or by considering rotamer libraries of the protein side chains. On the other hand and as we have envisaged in this research with regards to the critical targets, combining an orthogonal approach along with the above mentioned would increase the true positives and decrease the false positives rates.[36, 41]

The ligand-based approaches provide an alternative solution to reach out for the best understanding of ligand-target interactions in the absence of the experimental structure of a protein or in the case of a challenging protein. Even though both approaches are divergent, they serve the same role of finding a true binder. The building blocks of ligand-based approaches are the QSAR and pharmacophore models. Emanating from the concept that actives should share common features or physicochemical properties that would yield similar bioactivity. Furthermore, better solutions (potent ligands) could be found in the neighborhood of good solutions (actives). Thus, a QSAR model is simply the mathematical representation that discerns and correlates the best physicochemical properties or "the descriptors" that can explain the affinity of ligands toward a biological target.

The first step in this method is data mining. The availability of diverse and reliable experimental data is a fundamental step in this method. After which, energy minimization and

calculation of diverse structural and physicochemical descriptors should take place. Consecutively, the researcher should find the best combination of descriptors that can explain the variation in the biological activity followed by the statistical examination of the model power in predicting the activity of an external testing set. This approach uses 2D and 3D information embedded in the ligands. For instance, it could be coupled to a pharmacophore model which can enrich the 3D conceptual match of how active molecules could bind to their targets. A pharmacophore is defined as "the 3D spatial orientation of various features, such as hydrogen bond donors or acceptors, which are essential for the desired biological activity".[40, 42]

There are two general classes for the derivatization of structure-activity relationships, the linear and the nonlinear methods. In fact, there is no method superior upon the other; it is more related to the researcher choice and preference. Linear methods include multiple linear regression and partial least squares while the non-linear includes mainly the support vector machines and artificial neural networks.

I.3 Cheminformatic approaches to understanding the efficacy and safety of herbal remedies - Licorice plant as a proof of concept

The licorice plant is among the most popular medicinal plants that are marketed in the U.S. to alleviate multiple ailments, including cough, asthma and menopausal complaints, etc. In addition, it is recognized as one of the most studied herbs in the contemporary alternative medicine. Nevertheless, significant research gaps are still found. For instance, there is no definite explanation or even an appropriate recommendation of either its efficacy or safety. This includes its use as an alternative medicine for hormone replacement therapy, and as a chemopreventive agent, or adjuvant therapy in cancer treatments.[17, 25]In addition, consumed licorice supplements are prepared from different mixed species. Every species has its distinctive chemical profile,

which will yield a diverse spectrum of biological effects. This situation warrants the use of more sophisticated methods to empower and delineate activity endpoint detection in clinical trials or to detect pharmacokinetic liabilities as well.

I.3.1 Licorice plant origin and medicinal use

Licorice plant was recognized as a medicinal plant since the genesis of early civilization. The use of licorice is predated to the ancient Assyrians, Chinese and Egyptians and documented as a folk remedy in both Greek and Roman empires. The name *Glycyrrhiza* is a combination of two Greek words, which means the sweet (glycos) root (rhiza). In addition to its healing properties and as the name indicates this plant has been used as a sweetener, which was usually mixed with other bitter therapeutic herbs. This herbaceous, perennial herb is native to the Mediterranean region but it is also found in other parts of the world such as China, Russia, and India. The *Glycyrrhiza* genus belong to the family Fabaceae and consist of more than 28 species. However, three clinically relevant species are employed in the pharmacopeias namely *Glycyrrhiza glabra* L., *Glycyrrhiza uralensis* Fisch. and *Glycyrrhiza inflata* Batal.

The roots and rhizomes are the most widely used parts in both industry and therapeutic settings.[43] Traditionally, the roots are used to treat cough, diabetes, stomachache, ulcers, and tuberculosis. In addition to the fresh leaves were used to treat wounds. In the past few decades, a variety of pharmacological activities have been described for licorice or its secondary metabolites such as antioxidant, anti-inflammatory, antitussive and expectorant, antimicrobial, antiviral, anti-carcinogenic, neuroprotective, estrogenic and antidepressant. The astonishing diverse spectrum of activities is not surprising since more than 400 compounds have been isolated from different species of *Glycyrrhiza*.[44-45]

I.3.2 The licorice secondary metabolome: a vast interspecies diversity

The secondary metabolites of medicinal herbs comprise a large reservoir of bioactive or toxic compounds.[11] Generally, a substantial group of compounds might be shared between closely related species. However, their abundances would not be the same. In addition, different species usually express species-specific or marker compounds.[46] As a result, a variation in therapeutic effect is expected when different species are used. In fact, it is of great importance to address the dissimilarities between plant species to assure safe and efficacious delivery of their therapeutic effects.

Licorice species are among the richest and diverse plants in terms of secondary metabolites, producing triterpene saponins and a variety of phenolic compounds such as flavons, flavonols, isoflavones, isoflavenes, coumarines, and chalcones, and others.[47-48] Multiple species are, usually, mixed and used without discrimination in the preparation of licorice based herbal supplements. For instance, *G. uralensis, G. glabra, and G. inflata* are used in Europe and China interchangeably as licorice. Whereas, in the United States and Japan only *G. uralensis and G. glabra* are used. The identity of the constituents is directly correlated to the efficacy or safety of the supplement, which emphasizes the significance of understanding the secondary metabolite in each species, the correct labeling of the herbal supplements and the standardization of the botanical to their active components rather than to their marker compounds. Fortunately, the licorice metabolome has been vigorously investigated by both targeted and untargeted analysis.[46-47, 49-53] One recent study, analyzed 95 plant samples of the three species of licorice (*G. uralensis, G. glabra, and G. inflata*) and targeted 151 pure compounds known for licorice, has identified 12 and 13 species-specific marker compounds in both *G. uralensis* and *G. glabra*.[54] In addition, they have deduced characteristic biosynthetic pathways in each kind. Based on their

findings, the 3-aryl-5-methoxyl coumarins, the isoprenyl isoflavanes and the 2'-H chalcones were expressed discriminatively in *G. uralensis, G. glabra, and G. inflata* respectively. This interspecies diversity can remarkably affect the intended medicinal use (**Figures 3 and 4**).[54]

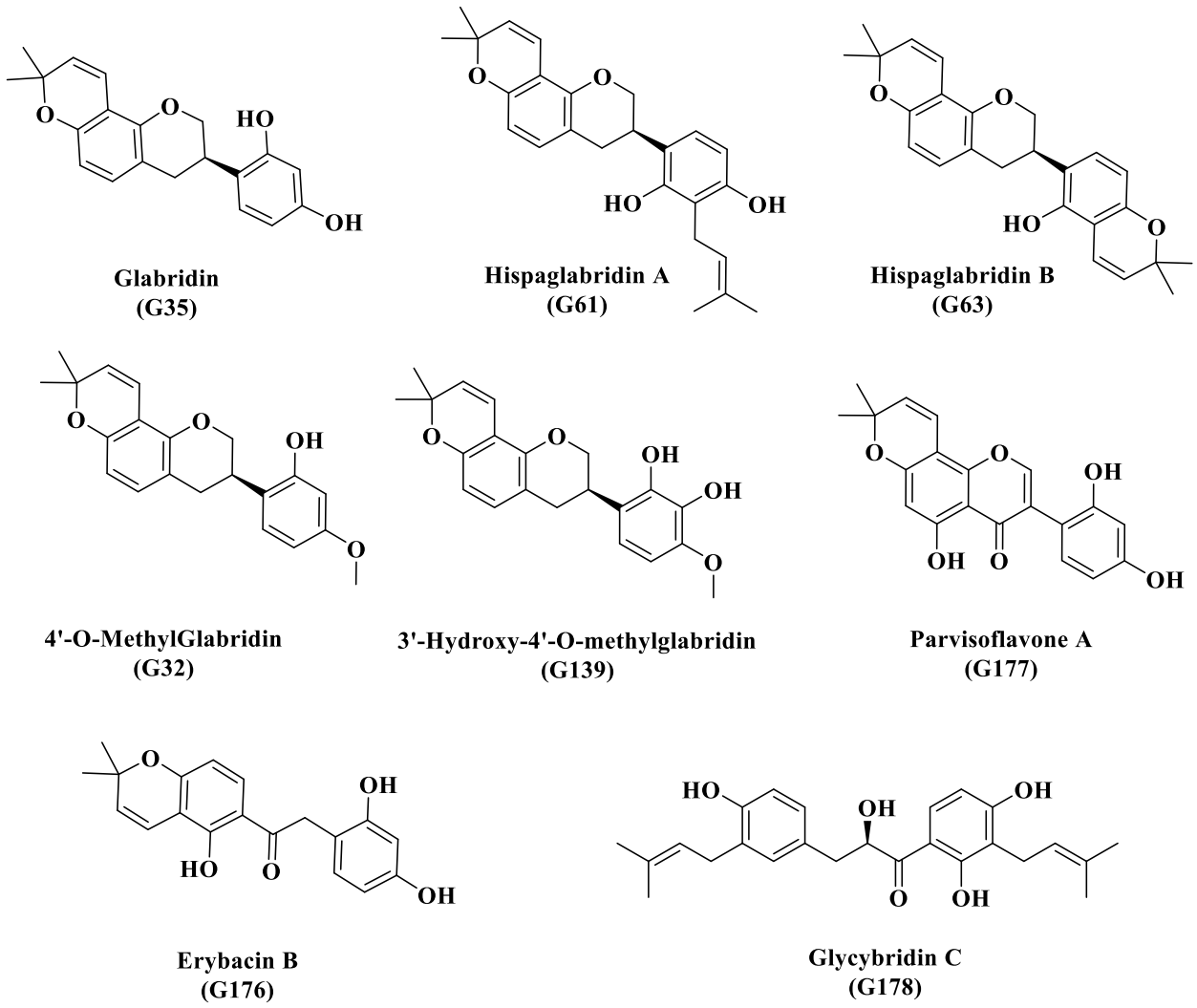


Figure 3. Species-specific chemical markers of *G. glabra*

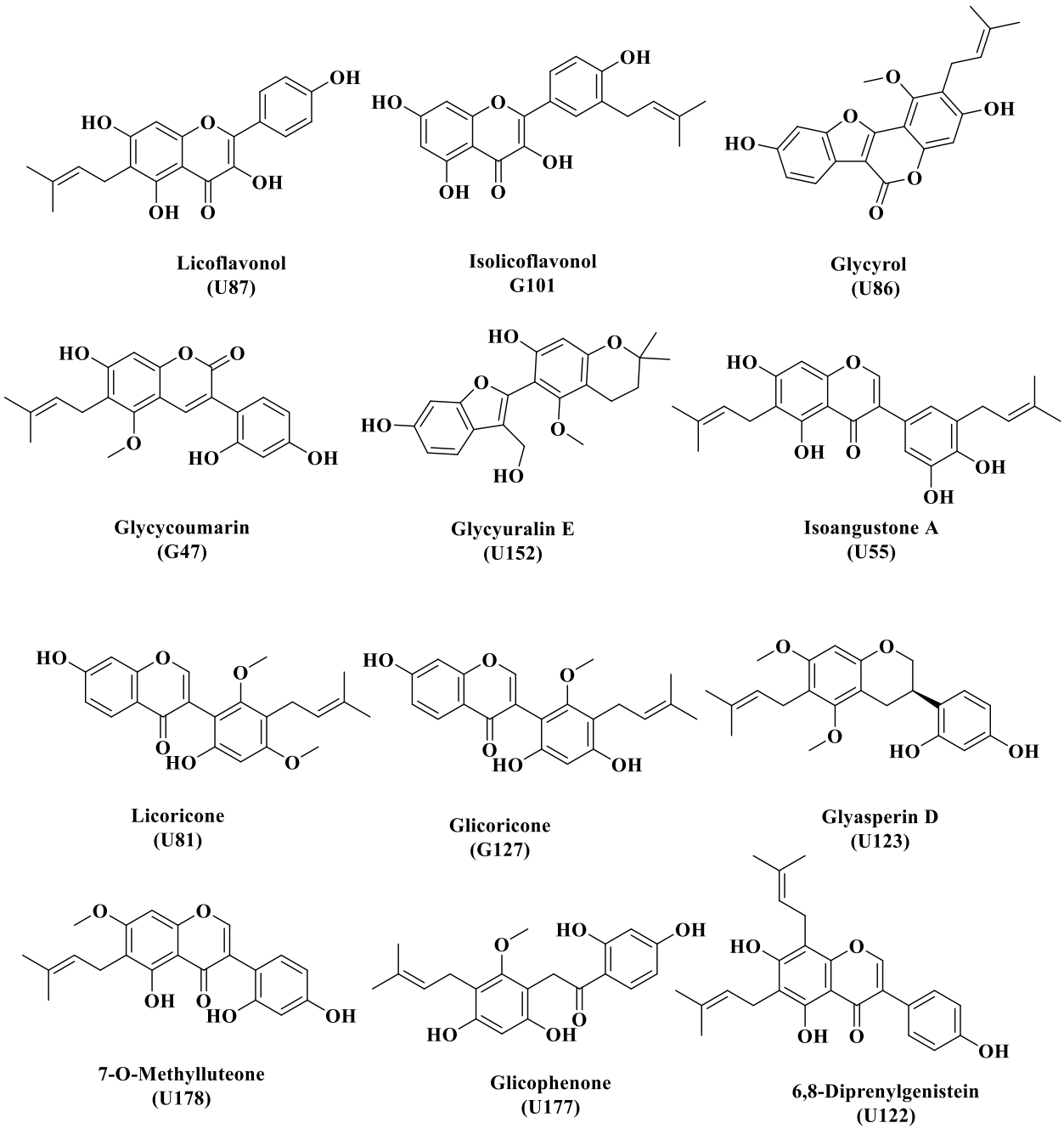


Figure 4. Species-specific chemical markers of *G. uralensis*

I.3.3 Licorice plant for woman health

There is a growing tendency among women to consume botanical supplements for a variety of reasons.[17] This trend is predominant in older women for the chemoprevention or the relief of postmenopausal symptoms such as vaginal atrophy, hot flashes, bone loss, and cardiovascular and metabolic functional changes. Popular opinion drifted to natural sources after the findings that pharmaceutical hormone replacement therapy (HRT), represented by estrogen pills or the available selective estrogen receptor modulators, may induce or exacerbate already existing uterine or breast cancer.[55] Moreover, several other studies encouraged the use of botanicals for their chemopreventive properties or protective effects against osteoporosis. However, none of these herbal treatments was carefully studied for their safety.[17, 56-58]

Glycyrrhiza is commonly encountered in herbal supplements intended for postmenopausal symptoms relief in the United States. Many studies detected the estrogenic behavior of the clinically relevant licorice plant as well as for other defined phenolics isolated from them.[17, 55, 59-63] In one recent study, *G. inflata* was shown to have the maximum efficacy for *h*ERα activity in Ishikawa cells followed by *G. uralensis* and *G. glabra* respectively. On the other hand, *G. uralensis* was ranked first for the maximum efficacy toward *h*ERβ followed by *G. inflata* and *G. glabra*.[62]

Liquiritigenin, a ubiquitous and major component in different licorice species was shown to produce weak estrogenic activity. However, several other compounds were also characterized by mixed estrogenic and anti-estrogenic behavior. Apparently, the estrogenic behavior of these species is a blend of multiple factors and can be traced back to many compounds that show a diverse estrogenic selectivity and functionality.[55]

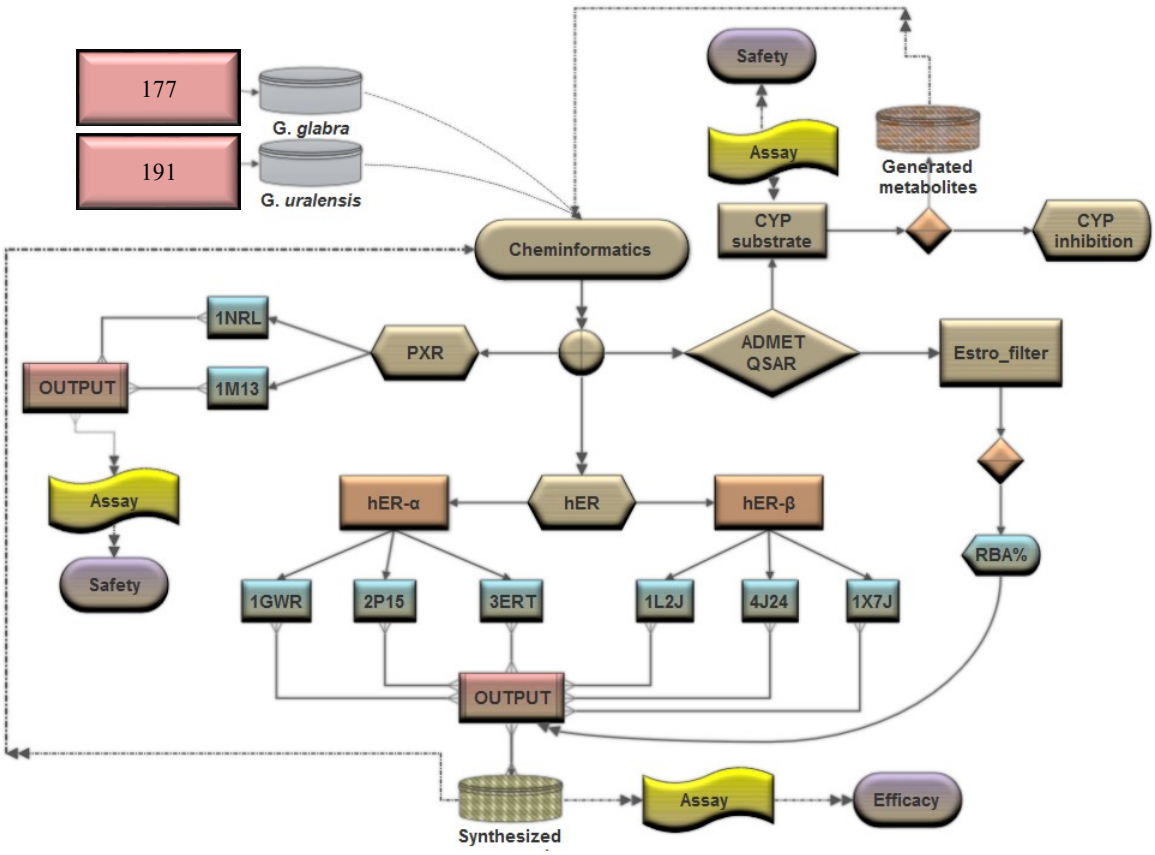
A direct comparison between these species is intricate. Nevertheless, rigorous identification of these compounds is critical and further evaluation of their long term interaction with the estrogenic receptors is required in order to guarantee the safety of their use.

I.4 Overall aims

The grand challenges in botanical research starting from efficacy issues (how effective are extracts, marker compounds, and potent active constituents?) or safety issues (off-target effects, toxicity, HDIs, and other adverse effects) warrant the implementation of new strategies such as physiologically-based models and informatics. The analysis of multicomponent herbs using cheminformatics logic may be one of the effective, innovative tools that could be used. An approach that includes identification of active component(s) with computational tools, validation with cell-based methods will facilitate the identification of component(s) that could explain the activity of the botanical drugs. Such an approach could lead to a method to provide a mass comparison between multiple plant species.

In view of the above discussion, we have envisaged that licorice plant would serve as the perfect demonstration in how to deal with a botanical issue of different levels or multiplicities. In this research, we have investigated the utility of cheminformatics methodologies as new strategies to identify the pool of the possible chemicals that could be attributed for certain activities or safety issues related to a plant species.

1. Deconvolution of Estrogenic Potential of PhytoSERMs in Licorice (Chapter II)
2. Synthesis of Unique Chemical Entities and their Estrogenic Activities (Chapter III)
3. Activation of PXR by Licorice Compounds (Chapter IV)



Scheme 1. General representation of the proposed research methodology

CHAPTER II
DECONVOLUTION OF ESTROGENIC POTENTIAL OF PhytoSERMs IN LICORICE

II.1 Introduction

II.1.1 PhytoSERMS: Chemoprevention vs endocrine disruption

The recent term of "PhytoSERMs" came after two consecutive serendipities. The first one was the red clover disease, which was reported in the 1940s. Remarkably, a "*botanical*" estrogenic behavior was inferred from the observed infertility of the sheep grazed on grasslands that had the red clover.[57] The second one was the discovery of tamoxifen in the 1960s, upon which the concept of "SERM behavior" was born. Tamoxifen had transformed from a failed contraceptive to the gold standard of the targeted chemotherapy of breast cancer in that era.[56] The estrogenic behavior of tamoxifen was puzzling; it showed different agonistic/antagonistic behaviors depending on the analyzed specific tissue. It has triggered antagonism and agonism in the breast and bone tissue respectively.[64] The main underlying mechanisms behind this non-classical activity involved the presence or absence of certain coactivators/corepressors proteins as well as the abundance of specific estrogen receptor isoform in the targeted tissue.[56, 65]

In fact, SERM tissue selectivity serves a multi-targeted therapy by antagonizing the estrogen-driven carcinogenesis in the breast tissue, while at the same time mimicking and maintaining the beneficial bone health. However, large-scale studies showed later that tamoxifen and other synthetic SERMs have detrimental effects on the uterine and endometrial tissues, which might lead to carcinogenesis. Consequently, to solve this estrogen paradox, the search for

alternatives SERMs with improved profiles was commenced. Finding novel SERMS with unique downstream expression events is extremely important, especially for the genes distinctly regulated by unique estrogen isoform.[66-67]

Over time, phytoestrogens were recognized as a vital source of natural estrogens and multiple studies have verified their SERM behavior. Many of the first generation phytoestrogens belong to the isoflavonoids and flavonoids classes, but other structural types such as lignans, coumestans, and stilbenes or their prenylated congeners share the same properties.[56] The significance of these phytoSERMS has grown immediately after the epidemiological studies among the Asian women. These studies found an interesting correlation between the phytoestrogen-rich diets and the diminished incidence of breast cancer as compared to the typical western women (These events have encouraged the western community to consume isoflavone-rich diets in addition to other nutraceuticals prepared for this purpose (**Figure 5**).[68] Presently, the debate about the efficacy and safety of these trends is still going on. The clinical studies conducted were controversial and most likely they were designed only for short spans.[69-70]

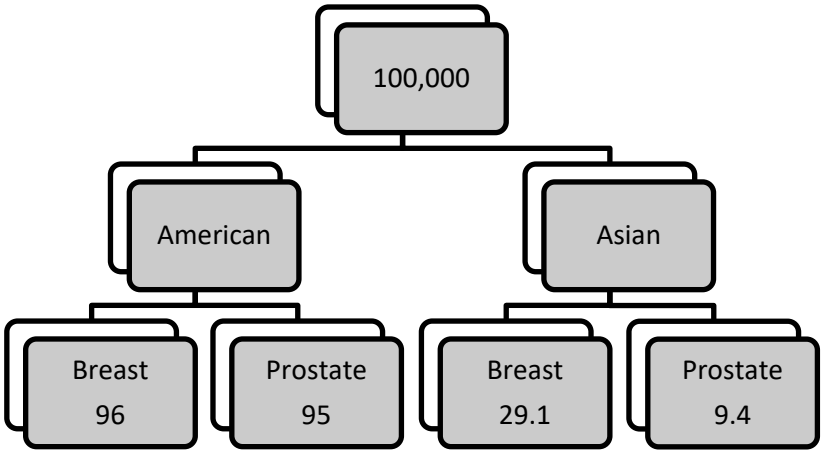


Figure 5. Comparison of the incidence of breast and prostate cancer among the Asian and American people.

A recent study has shown that isoflavones -like genistein- are capable of inducing proliferation in cancer cells by altering the $NF_k\beta$ (nuclear factor kappa-light-chain-enhancer of activated B cells) and AKT (Protein Kinase B) pathways in addition to *h*ERs. This study revealed that they are capable of upregulating the gene expression associated with tumor proliferation via *h*ERs.[71] Conversely, many other studies concluded appealing results, where they recommended their use in both pre and post-menopausal women.[68, 72] Taken together, there has been a steady conceptual buildup that phytoSERMs have the potential to act as chemopreventive agents, or as an estrogen surrogate to relief the vasomotor complications and to avoid osteoporosis.[73]

Since then, the targeted population has free access to these phytoSERMs nutraceuticals and herbal preparations without any rigorous attention. This situation has undermined the safety of these products, where they might behave as endocrine disruptors. In addition, it underscores the potential harms of transforming these products from regular diets to every day pills. It ignores the fact that these products are mixtures of chemicals with high capability of manipulating the estrogenic receptors and their various gene expression responses. Actually, the short-term and small-sized studies are incompetent to rule out any possibilities. Moreover, although the FDA divines the potentials of these therapies, it recommends people and special population to take some precautions when consuming such products.[74] On the other hand, these botanicals are considered as reservoirs of estrogens or estrogenic scaffolds each with a unique matrix of tissue, coregulator, gene, and isoform selectivity. A demanding requisite here is the detailed investigation of these components to unravel their potentials or harm spectrums.

II.1.2 The Human Estrogen Receptors

The first step to capture novel phytoSERMs from botanicals, which will aid in designing better SERMs, is to understand their mechanism of action. The estrogenic receptors belong to the nuclear receptor family, which are described as gene transcription regulators. Estradiol is the native ligand, which is responsible for triggering a cascade of gene transcriptional events that control the growth, proliferation, differentiation and the function of different tissues.[75] There are two isoforms of estrogen receptors (*h*ERα and *h*ERβ); both are widely distributed in our bodies and both can be stimulated by estradiol. Higher *h*ERα ratios found in the breast, ovarian, endometrial and hypothalamus tissues. On the other hand, higher *h*ERβ ratios are found in the heart, kidney, lung, bone, brain, prostate, intestine and endothelial tissues.[56, 75-76]

Additionally, unique tissues selectively express the specific type or concentration of coregulator proteins, which could act as coactivators or corepressors. *h*ERβ has been shown to exert regulatory functions over *h*ERα, while at the same time, they are both triggered by estradiol and share many target genes, indicating high similarity.[76-77] Similar to other nuclear receptors, it consists of five main domains but the most important ones in terms of transcriptional activity are the N-terminal (DNA-binding Domain, DBD) and the C-terminal domain (ligand-binding domain, LBD).

The LBD consists of 12 helices (H) and 4 β sheets, where the core layer that is in direct contact with the ligand is formed by H5, H6, H9, and H10. Three polar residues are established for receptor activation, in one side E353 (305 in *h*ER β) and R394 (R346) and on the other side H524 (475). Generally, to fulfill the best H-bond network required for activation, the ligand should have a hydrophobic scaffold flanked with two polar sides separated by 11 Å. The amino acids lining the LBD of both *h*ERs isoforms are almost identical except for two residues.[77-81]

Only two residues among those that line the binding pocket are found to be different: L384 (*h*ERα) vs. M336 (*h*ERβ) and M421 (*h*ERα) vs. I373 (*h*ERβ).[82-83]

Interestingly, ligands, which interacts with *h*ERs, can direct the function of these receptors toward their active or inactive states. Previous experimental data correlated the conformation of the LBD of *h*ER crystal structures to the ligand-induced activity. They can modulate the flexible helix H12 (one part of the active functional region 2 (AF-2) in the LBD), which allows variant conformations to take place as shown in **Figure 6.** Of note, the agonist, partial agonist and the antagonist conformations which will determine whether or not and which type of coregulator (coactivator or corepressor) might be hosted in the hydrophobic groove found in the AF-2.[75]

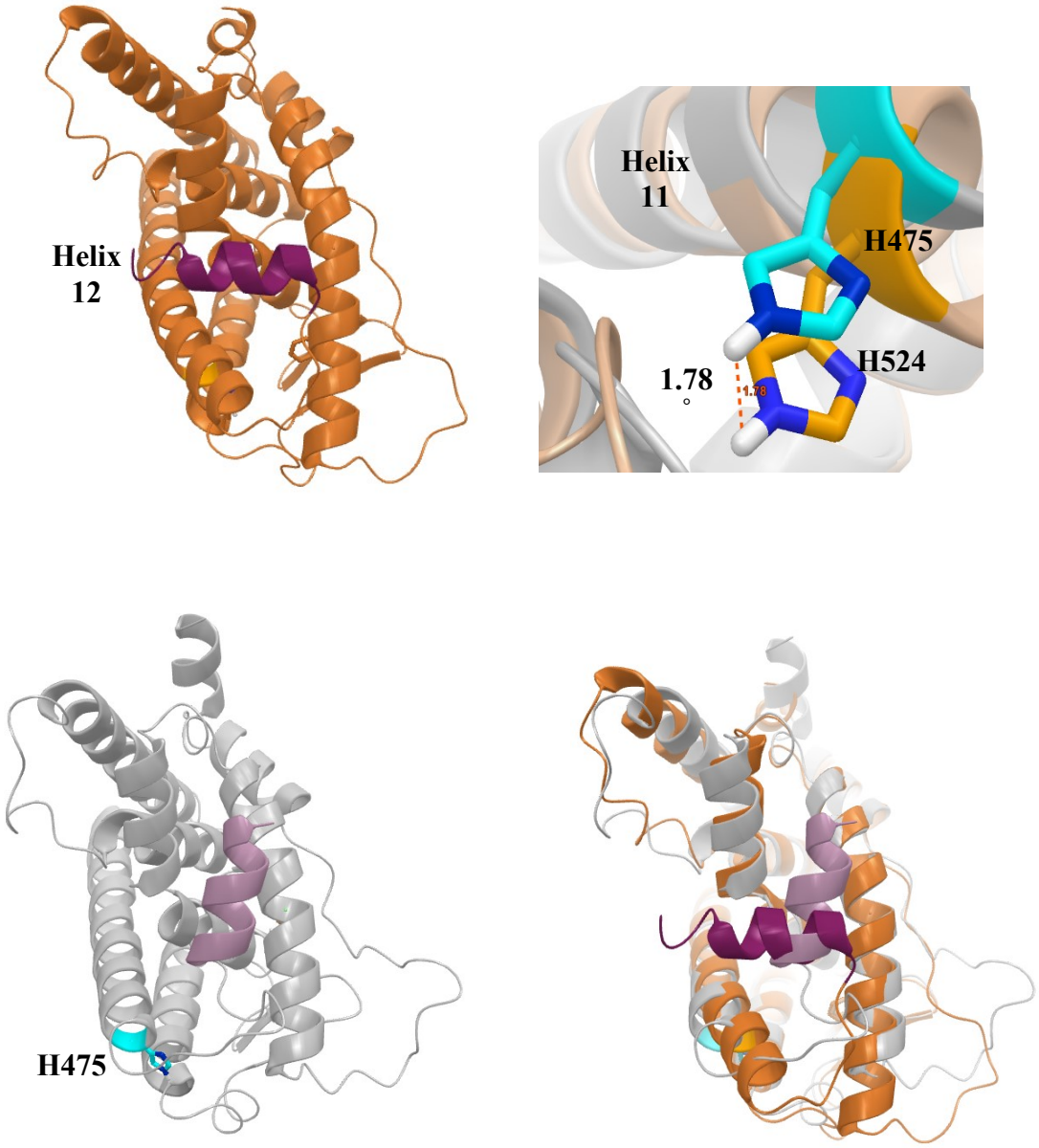


Figure 6. (Upper right) Open conformation of *h*ER in brown cartoon (PDB: 2P15) with helix 12 in maroon. **(Bottom left)** Open conformation of *h*ER in grey cartoon (PDB: 1L2J) with helix 12 in light pink. **(Upper right)** Helix 11 flexibility as shown by different orientations of the key residue histidine H524 (475). **(Bottom right)** Open and closed conformations overlaid.

II.1.3 Available cheminformatic information for the prediction of estrogenicity.

Since the discovery of *h*ERs in the 1960s up to date, the estrogenic behavior of synthetic and natural compounds became a hot topic in the scientific arena. This has been further aggravated by the emergence of the endocrine disruptors.[84] Essentially, the world has witnessed an increasing number of exogenous chemicals that can interfere with the human endocrine system and induce adverse effects.

This group of heterogeneous compounds includes many pesticides, herbicides, and fuel combustion by-products. This high burden of thousands of chemicals emerging has created a gap that exceeds the capacities of the regulatory agencies and the industrial sectors to follow and to evaluate their endpoint toxicities. Hence, to address this issue, the incorporation of multiple *in silico* and cheminformatic-based techniques have been created for the prediction of estrogenicity as well as other toxicity endpoints.[85-87]

Among the list, there are two commercially available models viz., ADMET Predictor™ and MetaDrug™, and other open access sources viz., the free docking software the Endocrine Disruptome.[88-90] Other free QSAR-based models are VEGA (Virtual models for property Evaluation of chemicals within a Global Architecture)[91] and OCHEM (Online Chemical Environment).[92] The first two models were generated from the same data set; the DSSTox database (Distributed Structure-Searchable Toxicity).[93] In addition, they both offer a qualitative and a quantitative assessment for the estrogenic activity built by QSAR models of nonlinear analysis such as ensemble neural networks or recursive partitioning algorithms, respectively. The OCHEM model was generated using a different dataset; the CERAPP database (Collaborative Estrogen Receptor Activity Prediction) which involved a diverse set of estrogenic activity endpoints that could not be modeled using regular linear methods.[94] Instead, they have applied a

network modeling with two-dimensional hierarchical clustering to get a consensus decision for their assessment, which provides both qualitative and quantitative measures. The VEGA model is only a classification model, which has applied the CART algorithm (Classification and regression tree) on a data set published from the Japanese Ministry of Economy, Trade, and Industry database.

The last free tool is the Endocrine Disruptome, which offers a structure-based solution rather than a QSAR based analysis. This tool provides a free platform for the assessment of potential activity toward fourteen nuclear receptors including the estrogenic receptors. It is a classification model that provides the illustration of the docking protocols and the data sets used for building the model, in addition to the sensitivity thresholds that would be applied for identification.[85]

One recent study has compared the performance of these five tools against the CERAPP database. As expected, the performance of OCHEM was the best since it has been trained with the same data set. However, the predictability of all models was described as convenient for selecting high priority chemicals that could alter the ERs. Even though they have shown low sensitivity (ratio of the true positive rate ~ 0.50), a higher specificity (ratio of the true negative rate ~ 0.80) was accomplished. The reason being that they were trained using imbalanced training sets that have incorporated more inactives. However, the addition of applicability domains would enhance the reliability and appropriateness of use according to the chemical structure similarity to the training set. The study concluded that these models would have more power to identify inactives and the combination of multiple models will be more advantageous to enhance the true positives rates.[85]

II.1.4 Ensemble docking technique

Docking and virtual screening are increasingly engaged in multiple processes of drug discovery, design and development. To speed up the docking processes, the flexibility of the protein is overlooked or partially considered.[36] However, the proteins are intrinsically dynamic structures that experience perturbations between high and low energy levels expressed as multiple conformations.

Apparently, in this environment, the ligand-binding event itself could not be explained by the key and lock model anymore; instead, a conformational selection or induced fit model will take place. In fact, overlooking a profound characteristic as protein flexibility will definitely compromise the predictability of the 3D model under consideration. On the other hand, implementing these properties will sensitize the performance of the model and permit the exploration of novel chemospheres for potential target-ligand pairs. The applicability of these models exceed the classic task of hit identification; it has open the doors to explore toxicity, selectivity profiling, target fishing, polypharmacology and drug repurposing.[95]

To overcome the limitations of rigid docking, multiple receptor docking or "ensemble docking" has been proposed. An experimental sampling of protein-ligand complexes will track back true active or inactive snapshots among the infinite number of possible protein conformations.[96] Furthermore, molecular dynamics offer another source of probable relevant conformations.[97]

One addition in this area is the "Pocketome" search engine. This engine is a collection (3313 entries in the last update in 2018) of conformational ensembles of druggable protein pockets for major biologicals targets pathways such as kinases, nuclear receptors family, CYP340 family, G-protein coupled receptor 1 family, etc. Treatment of protein flexibility within

the context of conformational ensembles has proven effective.[23] In one benchmarking study, the applicability of the conformational ensembles libraries for virtual screening was investigated by utilizing thirteen nuclear receptors (66 structures) and a focused library of known nuclear receptors modulators (157 structures).[98] The selectivity of active ligands was found to be high, where the area under the receiver operating characteristic curve (ROC curves) averaged to 85% and peaked to 99%. Compared to single receptor mode, there were major improvements in specific cases such as the estrogen receptors or at least keeping a high-level of detection. Specifically, the ensemble docking mode has increased the docking accuracy from 78% to 89% as measured by the selection of the near-native pose as one of the top three (out of a list of scored and ranked poses). This study concluded that the application of the ensemble strategy would enhance the performance of docking and virtual screening by careful selection of a limited number of distinctive conformations.[98]

For the assessment of licorice compounds for their estrogenic potential, we have incorporated the docking technique, which was elaborated with the ensemble of available receptors of both ER isoform. Additionally, we have investigated the commercially available QSAR model "estro-model" that is found in (ADMET Predictor™) to empower the performance of our predictions.

II.1.5 PhytoSERMs in licorice: A case study

Licorice (Radix *Glycyrrhizae*) is one of the herbs endowed with estrogenic character. Indeed, it is increasingly used both as a natural substitute for endogenous estrogens in case of menopause or as a chemopreventive agent.[17] However, licorice extract is derived from various species of *Glycyrrhiza*. Yet, most often, the identity of the botanical material in the licorice herbal product is not well characterized. The most common variety of licorice is *Glycyrrhiza*

glabra, commercially cultivated throughout the US and Europe, but other varieties are often used in various regions, such as *G. glabra var. glandulifera* (Russian and southwestern Asian), *G. echinata* (Russian licorice) and *G. uralensis* Fisch. Ex DC. (Chinese licorice/Manchurian licorice).[49] In addition to the presence of species-specific secondary metabolites, different licorice species demonstrate a wide spectrum of chemovariation. Metabolic profiling of multiple licorice species has shown widely heterogeneous chemical profiles. These distinct chemical patterns are expected to elicit variable biological responses. To date, more than 20 triterpenoids and 300 flavonoids have been isolated from different Glycyrrhiza species.[46, 52-53, 99]

A common phytoestrogen to all licorice species is liquiritigenin, which displays weak estrogenic activity with moderate selectivity for *h*ERβ.[55] Previous studies have shown significantly higher concentrations of liquiritigenin and its analogs; liquiritin or isoliquiritin in *G. uralensis* as compared to those found in *G. glabra* or *G. inflata*.[63] Species-specific compounds are also identified, such as glycycoumarin in *G. uralensis*, glabridin in *G. glabra*, and licochalcone A in *G. inflata*.[100] To elucidate the estrogenic behavior of these plant species, many researchers compared their efficacy for stimulating an estrogenic response in cell-based assays. One recent study showed that the estrogenic activity decreased in the order *G. uralensis* > *G. inflata* > *G. glabra* with a higher preference toward *h*ERβ.[62, 101]

This study and several previous ones encouraged researchers to carry out the bioactivity-guided fractionation to identify the compounds responsible for this estrogenic activity. Besides liquiritigenin, researchers identified other phytoestrogens in licorice such as (glabridin, glabrene, vestitol, calycosin, methoxychalcone, vestitol, glyasperin C, glycycoumarin, and glicoricone) demonstrating the estrogenic activity with moderate ERβ selectivity (**Figure 7**). Screening the estrogenic activity of the latter compounds via transcriptional assays in breast cell lines (MCF-7)

and liver cancer cell lines showed a unique mixed agonist/antagonist character for each compound in different tissue/target gene subsets, indicating a selective estrogen receptor modulating (SERM) behavior toward *h*ERs.[55, 59, 102]

However, there is enough evidence in the literature that crude plant extracts have greater estrogenic activity compared to the isolated constituents at an equivalent dose. In many cases, there is evidence of synergy including pharmacodynamics interactions, but the exact mechanisms have not been elucidated. Several mechanisms may also be operating in parallel.[55] Generally, phytoestrogens are estrogen mimics capable of interacting with the estrogenic receptors (*h*ERα and *h*ERβ) but in a discriminative fashion. They were shown to be *h*ERβ selective in multiple levels of action including binding affinity, co-regulators recruitment, and gene transcription. The licorice SERM activity can be explained by the distinct ability of its components to stabilize the different conformation of *h*ERs, which will preferentially recruit different types of coregulators.[101]

Nevertheless, none of the individual identified components matches the activity of the extract. From the above discussion, it is clear that only a few licorice components were tested for their estrogenic behavior mostly due to their minor presence in the fractions. Likewise, they might have different unique phytoSERMS profiles and they could be involved in the observed activity. The aim of this study was to dissect the estrogenic behavior of licorice components by the application of comparative ensemble docking study. Hence, validating an exciting methodology to identify active compounds in complex mixtures and to understand the mechanism of action behind an herbal remedy.

By the application of such a method, we anticipate circumventing some of the shortcomings accompanied by the classic bioactivity-guided fractionation. Particularly, when the

tested fraction includes multiple compounds with potential paradoxical estrogenic effects or in the opposite scenario where you have the additive effect of multiple weakly active components. Furthermore, selecting the right measures of certain activity for screening would certainly influence the results. For instance, if the screening assay seeks activation of certain function controlled by ER that means it is incompetent to detect antagonists.

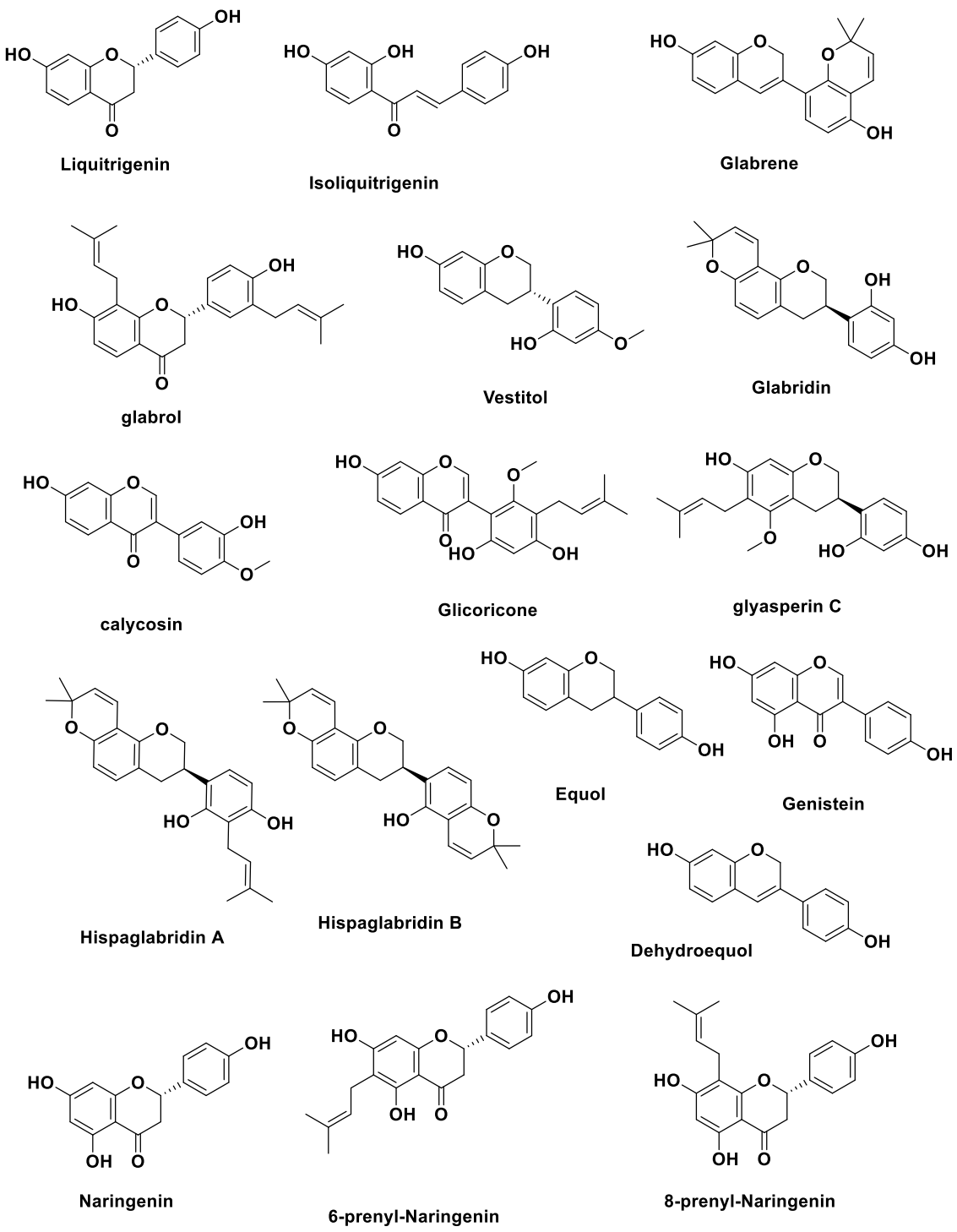


Figure 7. Selected universal known phytoestrogens from licorice and other botanical sources.

Alternatively, the application of new techniques of cheminformatics will shed light onto new phytoestrogens with unique profiles that would not be detected by the classic methods. To demonstrate the utility of ensemble docking for the deconvolution of active compounds in a complex mixture, two species of commonly used licorice, *Glycyrrhiza glabra* and *Glycyrrhiza uralensis*, were selected. Using several computational tools along with crystal structures of two isoforms of human Estrogen Receptors (*h*ERs), phytochemicals from *Glycyrrhiza* were probed for their isoform preference along with their putative functionality as agonists/antagonists of estrogen receptors.

II.2 Results

II.2.1 Ensemble docking

II.2.1.1 Validation

Initially, we have explored the binding site of six distinct *h*ER conformations using sitemap calculation in order to evaluate their diverse characteristics as shown in **Table SI 1**, **Figure 8,** and **Figure 9**. In general, the pocket of *h*ER receptors is highly hydrophobic with few polar residues at the sides of the binding site. Obviously, the active (closed) and the inactive (open) conformations of *h*ERs differ widely in the total surface area. This is modulated by H12 perturbation when it binds to an antagonist or a SERM, which will result in releasing one side of the binding site. One exception is 2P15 that has a large total surface area but it is still active, indicating that *h*ERα can accommodate larger hydrophobic compounds by adopting a unique conformation. On the other hand, *h*ERβ in its active form (4J24) exhibited the least surface area indicating its preference for smaller ligands. Next, we have considered an internal and external validation test as a measure of the accuracy of the docking procedure. The Glide SP docking protocol was reliable enough to reproduce the poses of the co-crystallized ligands with low

RMSD values. We also found that the ensemble docking approach was able to correctly pick the parent crystal structure for each native ligand as shown in **Table SI 2**. Regarding the focused library of known estrogenic compounds, the SP docking scores and the isoform preference are compared to the results of their actual experimental data as summarized in **Table 2**. Ten out of twenty listed compounds are licorice components that have been already tested toward *h*ERs. We found that the active compounds score below −10 kcal/mol generally. This protocol had 85% efficiency in the functional selectivity task (ER (+) for agonist vs ER (−) for the antagonist), but it had lower efficiency in isoform selectivity (65%). In fact, this is expected since it is already established that ER selectivity is hard to gain because of the high similarity between the two isoforms.

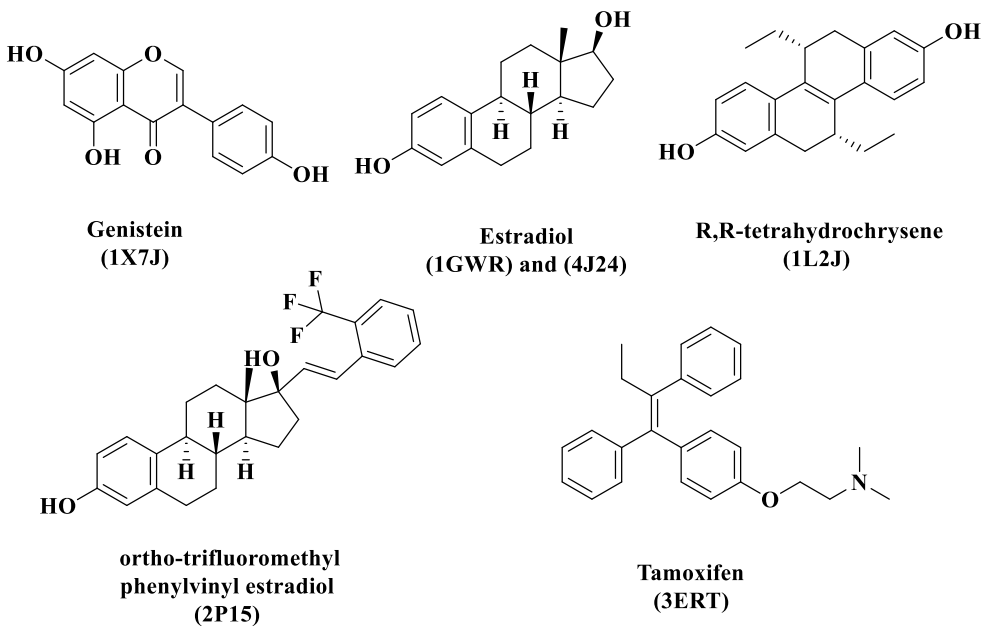


Figure 8. Co-crystallized ligands in the crystal structures used for the ensemble docking.

Table 2. Experimental affinity (RBA in nM) and isoform selectivity toward both *h*ERα and *h*ERβ for the external testing set compared to predicted SP docking score in isoform *h*ER (α and β) ensemble and functional *h*ER ensemble (ER (+): agonist or ER (−): antagonist)

	Experimental data						Prediction					
Ligand Name.	RBA *h* ERα	RBA *h* ERβ	Selectivity index	Prefered ER	*h* ERα pharmacology	*h* ERβ Pharmacology	*h* ERα SP Score	*h* ERβ SP Score	Predicted Selectivity	*h* ER (+)	*h* ER (−)	Predicted Conformation
Estradiol	100	100	1	α	ago	ago	−10.91	−10.81	α	−10.91	−9.81	ER (+)
TFMP-Estradiol	223	NA	NA	α	ago	ago	−14.44	*	α	−14.44	*	ER (+)
Tamoxifen	178	339	1.9	β	Ant	Ant	−11.37	*	α	*	−11.37	ER (−)
R,R-THC	23	144	6	β	ago	Ant	−11.17	−11.92	β	−11.17	−11.92	ER (−)
Genistein	0.021	6.8	324	β	m	m	−10.06	−10.55	β	−10.06	−10.55	ER (−)
Daidzein	0.01	0.04	4	β	m	m	−9.98	−10.51	β	−9.98	−10.51	ER (−)
Equol	0.2	1.66	8	β	ago	ago	−9.86	−10.33	β	−9.86	−10.33	ER (−)
Dehydroequol	0.046	4.3	93	β	ago	ago	−9.93	−10.50	β	−9.93	−10.50	ER (−)
Naringenin	0.08	0.5	6.25	β	ago	ago	−8.94	−10.23	β	−9.12	−10.23	ER (−)
6-Prenyl Naringenin	*	*	*	*	*	*	−9.94	−10.19	β	−9.94	−10.19	ER (−)
8-Prenyl Naringenin	19.46	6.5	0.33	α	ago	ago	−10.98	−11.15	β	−10.98	−11.15	ER (−)
Liquitrigenin (G10)	<0.001	0.013	13	β	m	m	−9.29	−10.23	β	−9.58	−10.23	ER (−)
Isoliquitrigenin (G125)	0.005	0.013	2.6	β	m	m	−9.09	−10.68	β	−9.37	−10.68	ER (−)
Glabrene (38)	0.22	NA	NA	NA	m	m	−10.72	−10.92	β	−10.72	−10.92	ER (−)
Glabrol (G3)	0.0014	0.0032	2.28	β	Ant	Ant	−11.04	−10.34	α	−11.04	−10.34	ER (+)
Vestitol (G128)	0.062	0.039	0.6	α	m	m	−8.87	−9.04	β	−8.87	−9.04	ER (−)
Glabridin (G35)	0.075	0.091	1.2	β	Ant	Ant	−7.70	−9.40	β	*	−9.40	ER (−)
Calycosin (G122)	0.01	0.027	2.7	β	m	m	−8.94	−9.76	β	−8.94	−9.76	ER (−)
Glicoricone (G127)	0.008	0.019	2.4	β	Ant	NA	−10.8	−9.67	α	−10.8	−9.67	ER (−)
Glyasperin C (G175)	0.01	0.111	11.1	β	Ant	NA	−9.50	−10.85	β	−9.50	−10.85	ER (−)

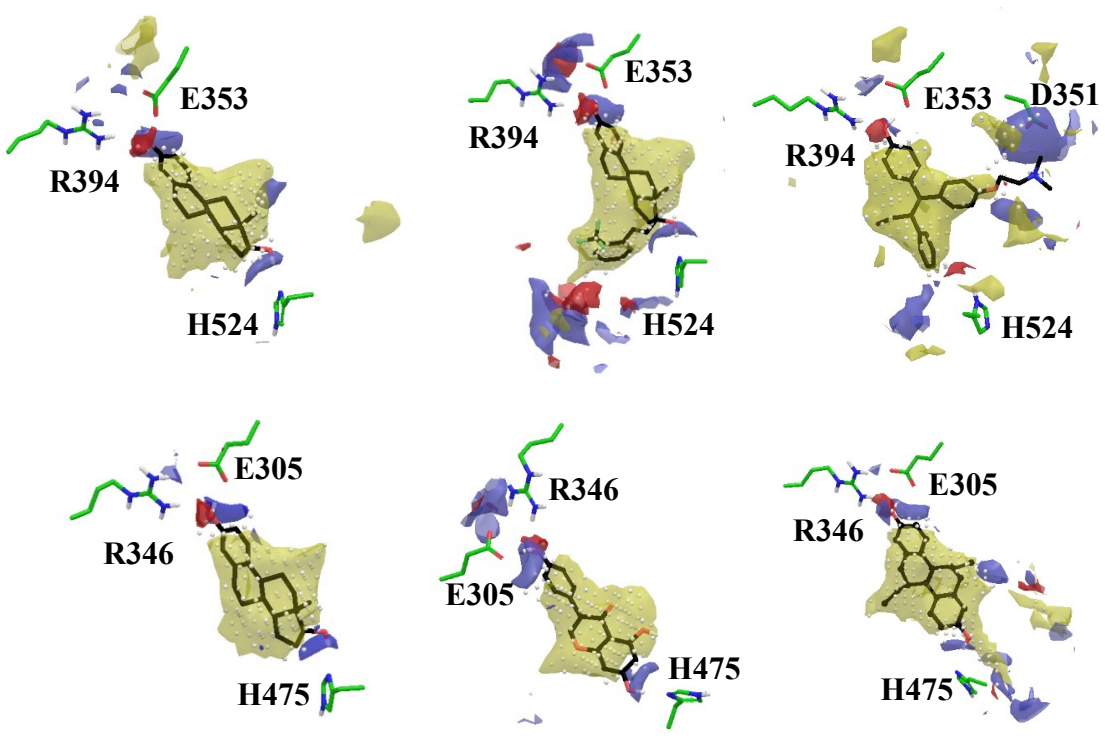


Figure 9. A) Binding sitemaps of *h*ERs with co-crystallized native ligands (**(upper)** *h*ERα PDBs, **left:** 1GWR, **middle:** 2P15, **right:** 3ERT, **(lower)** *h*ERβ PDBs, **left:** 4J24, **lower middle** 1X7J, and **lower right** 1L2J). Hydrogen-bond acceptor sites are indicated by red color, hydrogen-bond donor sites are indicated by purple color and hydrophobic sites are indicated by yellow color. Key residues are shown in green.

II.2.2 Docking of *Glycyrrhiza* compounds library into the ER ensembles

As described in the Experimental, we were interested in the identification of *Glycyrrhiza* compounds that might modulate the estrogenic activity, as we believed that correlation of their estrogenic activity to a couple of compounds is an oversimplification for the plant estrogenicity. Hence, a library of 368 compounds reported from both *Glycyrrhiza* species (*glabra* and *uralensis*) was docked to six *h*ER crystal structures. Apart from the known estrogenic modulators reported from licorice and evaluated in our validation test, this study has uncovered a larger group of phenolic compounds (78 structures), which have the same potential and yet has not

been explored. Firstly, to compare the predicted estrogenic compounds in both species we have plotted the highest glide score among the whole ensemble of six *h*ERs crystal structures for each compound against the preferred crystal structure (**Figure 10**). For *Glycyrrhiza* glabra, 49 compounds out of 177 scored above 10 and 65% of them preferred *h*ER beta isoform.

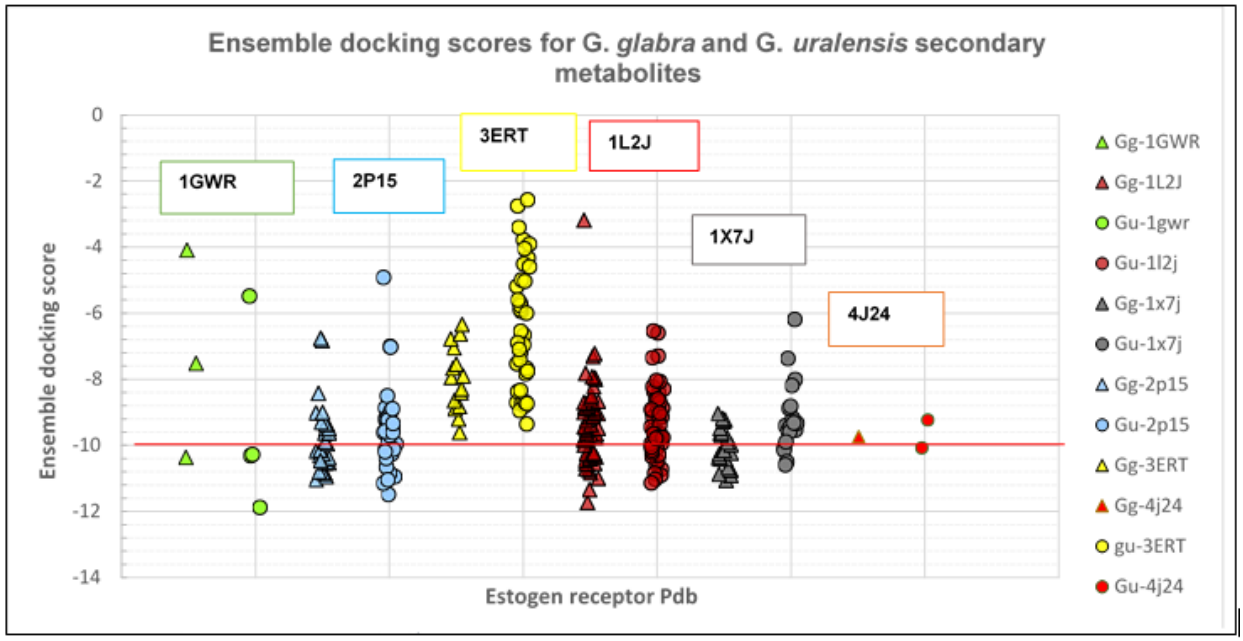


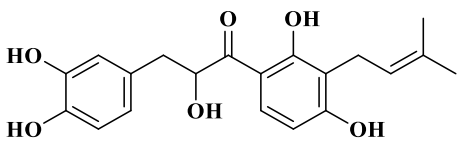
Figure 10. Ensemble SP docking scores in kcal/mol for the secondary metabolites of both *G. glabra* **(triangles)** and *G. uralensis* **(circles)** against six different crystal structures. For *h*ERα: 1GWR in green, 2P15 in blue and 3ERT in yellow. For *h*ERβ: 1L2J in red, 1X7J in grey and 4J24 in orange

For *G. uralensis*, 43 of the compounds out of 191 scored below −10 and 63% of them preferred *h*ERβ crystal structures. Our data suggest the preference of isoform β for both species. The analysis of such results revealed at least 50 compounds in each species that bear the potential of interfering with both *h*ERs α and β in different modes of action. Secondly, for the search of novel phytoSERMS candidates, we also challenged the compounds by following various methods described in the experimental section. The top twenty compounds along with their docking results, which are nominated by this model, belong to several classes as

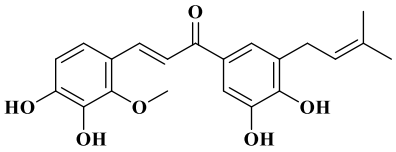
summarized in **Figure 11, Figure 12**, and **Figure 13**.

II.2.3 The qualitative and quantitative estimation of the estrogenic character of licorice secondary metabolites via Estro-model (QSAR model from ADMET Predictor™ software)

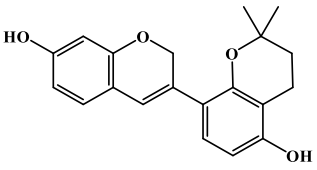
It has been established that the estrogenic activity is one of the complex bimolecular events that involve multiple levels of small molecule-protein-protein-DNA interaction. However, this process is sparked by the ligand (the small molecule-protein) interaction. To enhance our prediction of such an intricate system, we decided to test the licorice components with the orthogonal ligand-based model provided by ADMET Predictor™. Hence, we fed all the collected compounds from both species in the estro_filter model to calculate many atomic and molecular descriptors and eventually gain a consensus vote by the neural network that this model provides for classification. The compounds, which were selected as estrogenic, have been submitted to affinity estimation model (the estro_RBA model). The top 20 nominated estrogenic compound in each species are summarized in **Table SI 7**. Interestingly, many of the top compounds were shared in both models indicating a high chance of interference with this nuclear receptor.



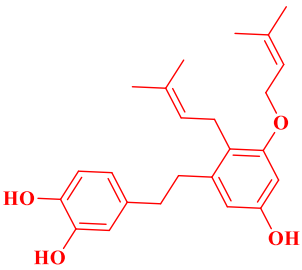
G86
***h*ERα: −10.79**
***h*ERβ (−): −11.35**



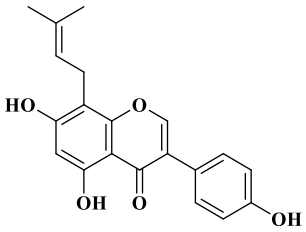
G138
***h*ERα: −10.54**
***h*ERβ (−): −10.74**



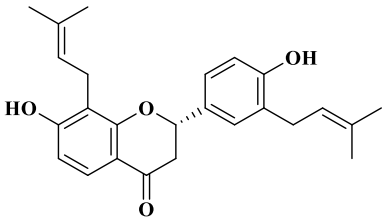
G38
***h*ERα: −10.72**
***h*ERβ (−): −10.92**



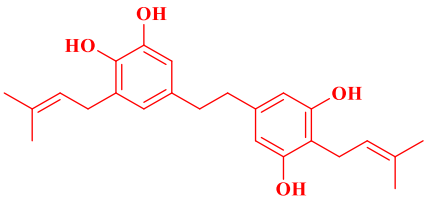
G28
***h*ERα: −10.43**
***h*ERβ (−): −10.54**



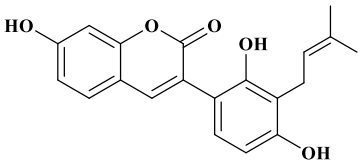
G112
***h*ERα: −10.40**
***h*ERβ (−): −10.34**



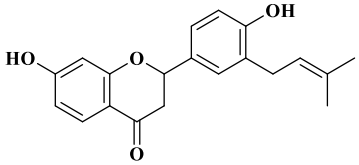
G3
***h*ERα: −11.04**
***h*ERβ (−): −10.34**



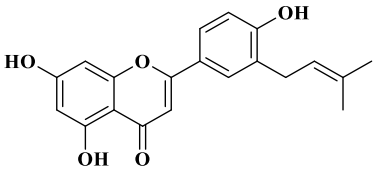
G40
***h*ERα: −10.83**
***h*ERβ (−): −10.43**



G174
***h*ERα: −10.84**
***h*ERβ (−): −10.64**

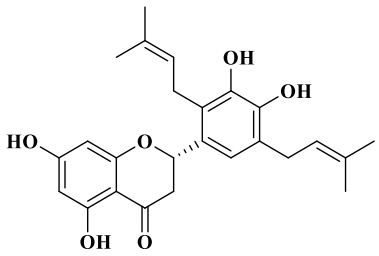


G156
***h*ERα: −10.38**
***h*ERβ (−): −10.34**

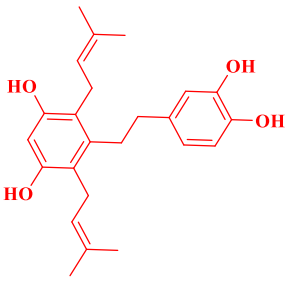


G106
***h*ERα: −10.54**
***h*ERβ (−): −9.31**

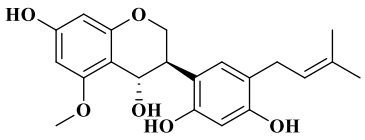
Figure 11. Top scoring compounds of *G. glabra* nominated by ensemble docking (structures **highlighted in red** are: top scoring compounds which belong to DHS class, **identifiers in blue** are those in agreement with top scoring compounds identified by the QSAR model)



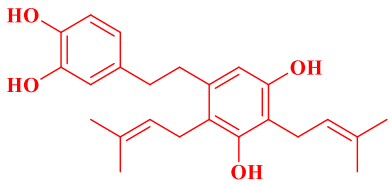
U77
***h*ERα: −11.045**
***h*ERβ (−): −10.474**



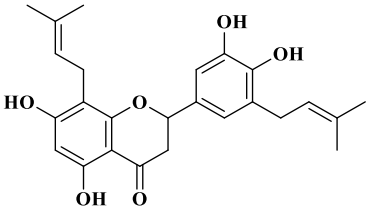
U15
***h*ERα: −11.487**
***h*ERβ (−):−9.652**



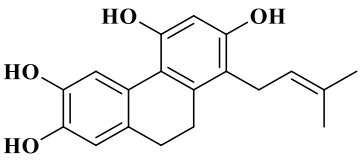
U191
***h*ERα: −10.841**
***h*ERβ (−): −9.011**



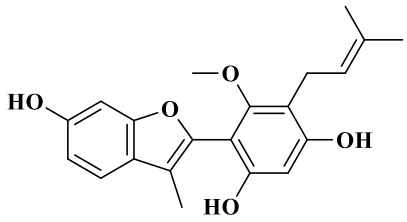
U16
***h*ERα: −8.866**
***h*ERβ (−): −11.35**



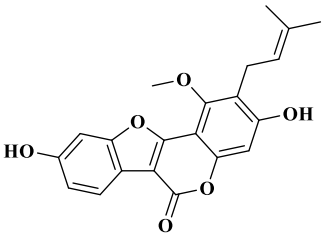
U78
***h*ERα: −10.194**
***h*ERβ (−):−10.061**



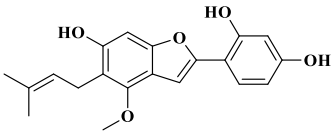
U25
***h*ERα: −9.918**
***h*ERβ: −9.014**



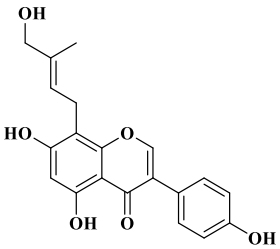
U22
***h*ERα: −10.263**
***h*ERβ (−): −9.862**



U86
***h*ERα: −9.094**
***h*ERβ (−): −10.581**



U23
***h*ERα: −8.246**
***h*ERβ (−):−11.135**



U42
***h*ERα: −10.39**
***h*ERβ (−): −9.039**

Figure 12. Top scoring compounds of *G. uralensis* nominated by ensemble docking (structures **highlighted in red** are: top scoring compounds which belong to DHS class, **identifiers in blue** are those in agreement with top scoring compounds identified by the QSAR model)

	α-ensemble			β-ensemble			α-ensemble			β-ensemble			
	1GWR	3ERT	2p15	1L2J	1X7J	4j24	1GWR	3ERT	2p15	1L2J	1X7J	4j24	
G6	0	0	-11.04	-10.34	0	0	-11.88	-5.63	-10.33	-9.89	-9.07	-10.25	U63
G50	-9.91	0	-10.95	-10.59	-10.74	-9.99	-8.23	-8.11	-11.49	-9.65	0	0	U15
G38	-10.35	0	-10.87	-10.52	-10.92	-9.88	-10.96	-8.18	-11.15	-10.49	-8.64	-10.23	U66
G174	-8.48	-6.55	-10.84	-10.64	-9.47	-7.43	0	-6.48	-11.05	-10.47	0	0	U77
G40	0	-6.19	-10.83	-10.43	0	0	-8.11	-7.42	-10.95	-9.94	-9.51	0	U106
G47	0	-7.23	-10.82	-10.3	-8.55	0	-8.09	-7.42	-10.84	-9.01	0	0	U191
G127	-8.04	0	-10.8	-9.67	0	0	0	-5.34	-10.66	-10.3	0	0	U127
G163	-10.39	-8.21	-10.79	-11.35	-10.84	-11.08	0	-5.98	-10.39	-8.66	-9.04	0	U24
G138	-9.81	-7.23	-10.55	-10.01	-10.75	-10.3	-10.31	-5.76	-9.63	-9.99	-9.63	-9.3	U89
G106	-8.86	-8.26	-10.54	-9.31	0	0	-10.28	-6.88	-10.01	-9.9	-10.23	-10.1	U105
G108	-10.27	-5.71	-10.53	-10.39	-10.42	-10.2	0	-6.78	-10.26	-9.86	0	0	U22
G33	-10.45	-8.51	-10.49	-8.94	-9.82	-9.85	0	-6.25	-10.19	-10.06	0	0	U78
G21	-10.17	-8.45	-10.47	-9.43	-10.4	-10.1	-6.75	-7.68	-10.12	-9.73	-9.24	-9.11	U177
G28	0	-8.27	-10.43	-10.54	-8.41	0	0	-5.69	-10.02	-10.08	0	-7.18	U99
G112	-9.84	-7.55	-10.41	-9.72	-10.34	-9.99	-8.5	-3.91	-9.95	-9.12	0	-8.84	U189
G176	-10.41	-8.34	-9.58	-10.71	0	0	-9.94	-8.56	-9.92	-7.96	-10.09	-9.18	U190
G156	-9.59	-8.86	-10.38	-10.34	0	0	-9.66	-5.23	-9.92	-8.8	-7.77	-9.01	U25
G30	-10.37	-7.44	-10.06	-8.99	-9.43	-9.98	-9.63	-7.21	-9.78	-9.46	0	-10.08	U61
G171	-10.33	-7.38	-10.2	-9.93	-10.88	-9.63	0	-5.32	-9.76	-10.47	-6.49	0	U115
G154	-7.76	-6.86	-10.32	-8.8	0	0	0	-5.31	-9.7	-8.41	0	0	U2
G158	0	-6.51	-9.93	-11.73	0	0	0	-5.36	-8.25	-11.14	0	0	U23
G163	-10.39	-8.21	-10.79	-11.35	-10.84	-11.08	0	-7.45	-8.87	-11.03	0	0	U16
G100	-9.7	-8.13	-9.88	-10.42	-11.06	-9.61	0	-7.54	-8.9	-10.91	0	0	U117
G99	-9.64	-7.06	-9.84	-11	0	0	0	-3.37	-9.5	-10.85	0	0	U175
G38	-10.35	0	-10.87	-10.52	-10.92	-9.88	0	-5.1	-8.41	-10.72	0	0	U88
171	-10.33	-7.38	-10.2	-9.93	-10.88	-9.63	-9.02	-5.97	-9.18	-9.28	-10.59	-9.69	U69
G60	0	0	0	-10.83	0	0	0	-6.08	-9.09	-10.58	-6.05	0	U86
G170	0	-7.99	-10.3	-10.75	-8.46	0	-9.68	-9.44	-9.61	-9.6	-10.49	-9.56	U165
G138	-9.81	-7.23	-10.55	-10.01	-10.75	-10.3	-10.96	-8.18	-11.15	-10.49	-8.64	-10.23	U66
G50	-9.91	0	-10.95	-10.59	-10.74	-9.99	0	-6.48	-11.05	-10.47	0	0	U77
G75	-9.43	0	-10.08	-9.7	-10.74	-9.89	0	-5.32	-9.76	-10.47	-6.49	0	U115
G176	-10.41	-8.34	-9.58	-10.71	0	0	0	-6.07	-8.7	-10.33	0	0	U176
G125	-9.09	-7.34	-8.8	-9.17	-10.68	-9.37	0	-5.34	-10.66	-10.3	0	0	U127
G174	-8.48	-6.55	-10.84	-10.64	-9.47	-7.43	0	-8.17	-7.52	-10.29	0	0	U39
G111	-8.33	-7.7	-10.03	-10.58	-8.8	0	-11.88	-5.63	-10.33	-9.89	-9.07	-10.25	U63
G28	0	-8.27	-10.43	-10.54	-8.41	0	-10.28	-6.88	-10.01	-9.9	-10.231	-10.1	U105
G2	0	0	-9.84	-10.45	0	0	0	-6.18	0	-10.22	0	0	U57
G40	0	-6.19	-10.83	-10.43	0	0	0	-7.78	-8.83	-9.08	-10.15	-8.98	U119
G108	-10.27	-5.71	-10.53	-10.39	-10.42	-10.2	0	-5.73	-7.57	-10.15	0	0	U62
G144	-9.68	-8.61	-8.83	-9.01	-10.42	-9.4	-9.94	-8.56	-9.92	-7.96	-10.09	-9.18	U190

Figure 13. Top 20 compounds **(white color)** by SP docking scores in *h*ERα ensemble compared to all crystal structures to visualize selectivity (*G. glabra:* **upper left**, *G. uralensis* **upper right**). Top 20 compounds by SP docking scores in *h*ERβ ensemble (*G. glabra:*

II.2.4 ADMET Metabolite prediction and estrogenic evaluation via QSAR models

The data revealed a small number of bio-activated metabolites (**Figure 14**). This is exemplified mostly by *O*-demethylation and/or hydroxylation processes. This is not surprising since most of the compounds are oxygenated phenolic structures. Nevertheless, the data suggest an estrogenic character for the majority of these metabolites. This preliminary investigation should be complemented with physiologically based pharmacokinetic models and secondary phase metabolism models to evaluate the effect and time of exposure of each route.

II.3 Discussion:

Licorice (*Glycyrrhiza* spp.) is well known for its wide spectrum of estrogenic activity, which presumably adds to the beneficial effects of this plant. Both species of licorice– *G. glabra* and *G. uralensis* are commonly used in the marketed botanical products without any respect to their soundly different chemical composition. Many studies tried to identify the phytoestrogens found in these plants. Some compounds, which belong to the classes of flavonoids, isoflavonoids or chalcones, such as liquiritigenin, vestitol, calycosin and isoliquiritigenin have demonstrated estrogenic activities. Others, such as glabrene, glabridin, glyasperin C, glabrol, glicoricone, showed antagonism or partial agonism of estrogenic activities. By contrast, compounds like hispaglabridin A and hispaglabridin B have no activity to either estrogen receptors. In fact, licorice plant provides a versatile and powerful SERM toolbox. The advent of ESR1 mutations that lead to (30-50%) resistance of the available therapies call for an urgent search for novel *h*ERα partial antagonists or SERMs of highly favorable kinetics profiles.[67] Paradoxically, in the case of highly estrogen-deprived tissue such as in case of exhaustive and long endocrine treatment of breast cancer or in the case of menopause, recent studies revealed the clinical efficacy of estradiol as an anticancer or antiproliferative agent through *h*ERα.[103]

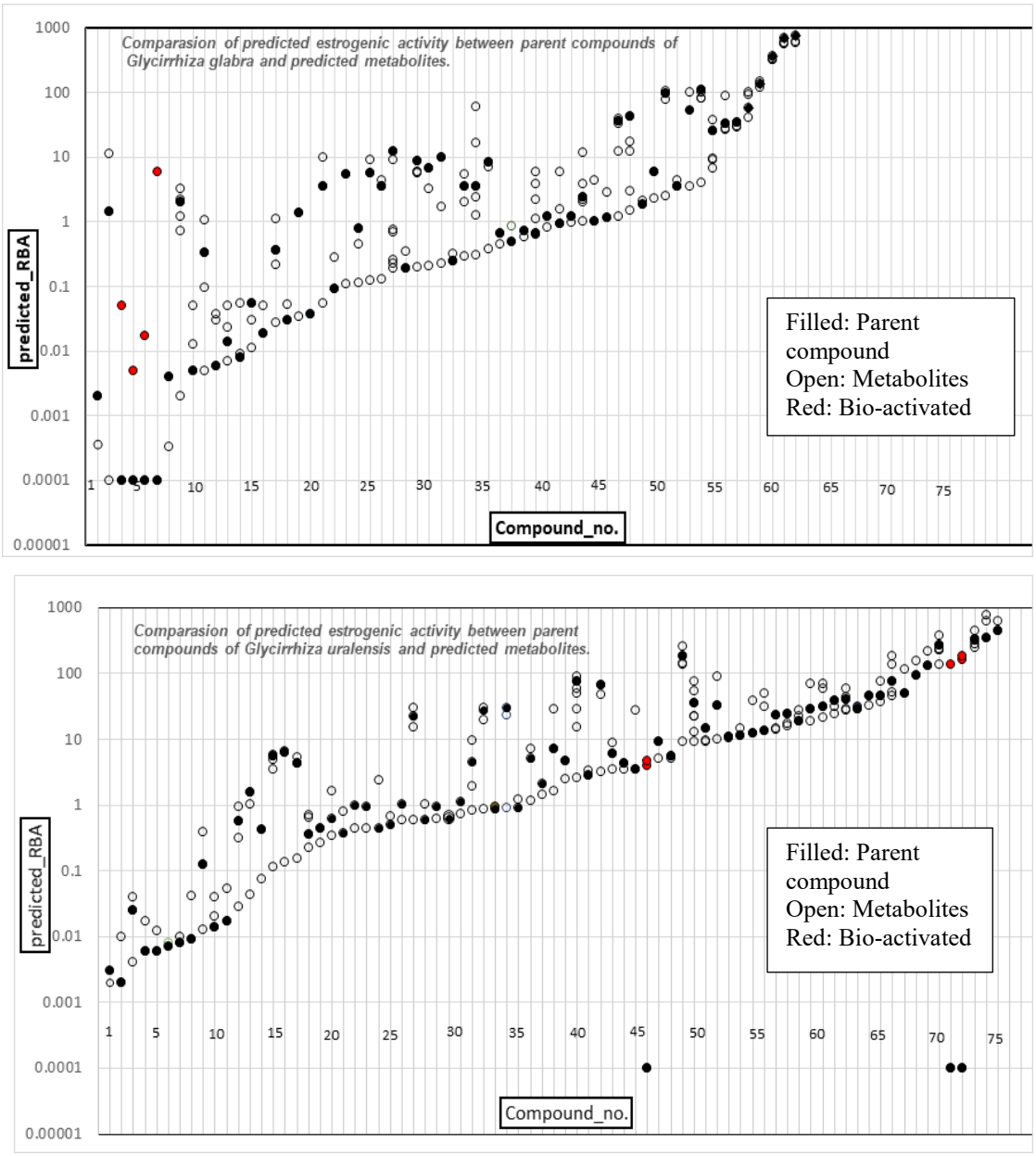


Figure 14. Comparison of ADMET QSAR Predicted RBA for both parent compounds and their putative metabolites originated from *G. uralensis* **(up)** and *G. glabra* **(down)**

As a result, a new-targeted SERM category with *h*ERα partial agonistic effects in cancer tissue but devoid of proliferative activity elsewhere is highly acquired. In this study, we were interested in understanding the estrogenic potential of two licorice species by the application of multiple cheminformatic techniques such as molecular docking and QSAR analysis. In addition, our attempt sought the validation of these tools as a basic approach for cheminformatic-guided investigation of herbal remedy. *h*ERs are nuclear receptors that are known for having a flexible ligand-binding domain. Moreover, the ligand-directed transcriptional activity of *h*ERs is correlated to the conformation of this particular region of the protein. Careful analysis of the conformational space of *h*ERs suggests variable degrees of H12 flexibility leading to different states of open and closed conformation.[75, 104]

Obviously, this suggests variable geometric and electronic features of the binding site in each specific conformation. That was evidenced by our analysis of the LBD pockets using sitemap calculations. To compensate for this flexibility for both *h*ERα and *h*ERβ and to avoid misleading information guided by one rigid receptor we have considered an ensemble docking with multiple protein conformations. To avoid any bias in the ensemble docking protocol and to improve the predictability, the same compounds were also evaluated using a readily available QSAR model from ADMET Predictor™ software for both qualitative and quantitative estimation of estrogenic nature.[105] In the latter approach, the 'hits' which scored above 70% confidence level in ADMET model were included for comparison and confirmation purposes.

II.3.1 Validation

Modeling the estrogenic activity of multicomponent systems such as licorice extracts that are targeting a physiologically versatile target such as *h*ERs is not straightforward.
This situation warrants a careful use of the precise computational methods to address more complicated protein features such as flexibility. In our modeling, we have considered six *h*ER protein structures and we have drawn out four ensembles to target both *h*ER isoform selectivity (α vs β) and conformation preference (active (*h*ER+) vs inactive (*h*ER−)). Cognate ligands were redocked using flexible SP protocol, which was able to reproduce the binding geometry of each ligand. By considering the focused library, the application of ensemble docking allowed us to detect SERM ligands that cannot fit into the agonist conformation but instead they prefer the inactive conformation. In this manner, *Glycyrrhiza* compounds such as (glabridine, glabrol, glicoricone, and glyasperin C) could be detected where they have achieved relatively high scores (−9.40, −10.34, −9.67, and −10.85 kcal/mol respectively) in at least one of the inactive conformations

Moreover, our ensemble model showed high recognition of the isotype-selective ligands by considering the simple SP raw scores. This model could not identify 8-prenyl-naringenin and vestitol as *h*ERα selective. Maybe this is due to the low selectivity index of these two compounds (0.5-0.6). On the other hand, all *h*ERβ selective agonists or partial agonists, which show better selectivity (genistein, daidzein, equol, dehydroequol, liquiritigenin, isoliquiritigenin, calycosin, and glabrene) and the antagonist (glabridin), were classified correctly as *h*ERβ selective. Interestingly, both isoform ensembles were able to rank the known phytoestrogens compounds in agreement with their experimental binding affinities. For instance, previous studies have shown that 8-prenyl-naringenin is the most estrogenically active phytoestrogen.[106]

In addition, *h*ERα agonistic activity declines in the order: 8-prenyl-naringenin> liquitrigenin> isoliquitrigenin> calycosin> vestitol when tested under identical assay conditions.[55, 63] By considering the docking results in the *h*ERα ensemble, we can notice a very similar trend. Moreover,

both binding affinity assays; the relative binding affinity (RBA) and efficacy assays; the concentration at the half maximum efficacy (EC_{50}) toward *h*ERβ showed a higher activity for the isoflavone genistein (13, 6 nM) followed by dehydroequol (4.3, 7.2nM) > equol (1.66, 74 nM) > (0.04, 100 nM).[107] *h*ERβ ensemble SP scores have also similar ranking. Of note, our docking algorithm did not yield any favorable docking pose for either hispaglabridin A or hispaglabridin B, which is in agreement with the previously reported cell-based data.[59, 79]

II.3.2 Estrogenicity of *Glycyrrhiza* secondary metabolites *via h*ER ensembles and estro_filter

Computational tools provide a versatile knowledge about the interactions of small molecules with their target proteins. For this purpose, we have analyzed 368 known secondary metabolites in licorice plant against ensemble *h*ER crystal structures of both isoforms. Our model suggests *h*ERβ selectivity for both *G. glabra* and *G. uralensis* extracts, where more than 60% of the highest scoring compounds (those scored higher than the cut-off −10.0 kcal/mol) preferred *h*ERβ sites. Interestingly, these results are consistent with the recent analysis, which showed that both *G. glabra* and *G. uralensis* extracts have similar potencies and both preferring *h*ERβ.

The top 20 compounds out of 368 are summarized in **Figure 13**. The model was able to identify many known phytoestrogens that are part of licorice in agreement with the reported literature. Among the top 20 compounds, known compounds were present such as glabrene (**G38**) the isoflavene and glycerol (**U86**) the 6-prenyl coumestrol, which are documented for their high experimental binding affinity toward *h*ERα (RBA of 0.22 and 0.11 respectively).[108] As

anticipated, most of the top scoring compounds are tethered with at least one prenyl group. This is due to the hydrophobic nature of the LBD, which will increase the value of the hydrophobic terms in the SP scoring function.

In fact, the experimental data in the validation set supports this prediction, where we noticed multiple prenylated compounds (**G3**, **G35**, **G38**, **G127**, and **U175**) that show sufficient RBA to exert *h*ER activity. However, the prenylation of isoflavonoids is known to change their pharmacology.[109] Although they show high affinity for *h*ER receptors in the binding affinity assay, these compounds show *h*ER inhibition response at least in one specific cell-based assay, a behavior that suggests a SERM activity. This is true for the monoprenylated isoflavones at 8, 6, 5' positions found in glabridine (**G35**), glyasperin C (**U175**) or glicoricone (**G127**) respectively.[55] On the other hand, prenylation at 3' position in flavonoids such as apigenin and liquiritigenin or a di-prenyl at 3' and 5' position such as breviflavone A and B kept their estrogenic character in the range of other known phytoestrogens.[110]

However, in our results, we found several compounds, which belong to the dihydrostilbenoid (DHS) class appended with prenyl groups at various positions, to be an interesting group of compounds that has not been studied for their estrogenic activities. In addition, the model suggests that these prenylated DHS are amenable to occupying available chemical space composed of L346, T347, L354, W383, and L540, hence had the potentials to tightly interact with *h*ERs. Interestingly, some of the bis-prenylated DHS **U15** and **U16** have been reported to be unique to *G. uralensis*. Not only these compounds scored high (< -11 Kcal/mol), but also the position of the prenylation is predicted to change their selectivity, where **U15** preferred *h*ERα in the active form, while **U16** selectively preferred *h*ERβ in the inactive form (**Figure 15** and **Figure 16**). As shown in the figures, the bad interactions between the

prenyl group of **U16** and H524 in the closed form might be the main reason behind this predicted selectivity. Both compounds showed interactions with the key amino acids E353, R394, and H524.

The present computational study revealed interesting features that might be gained with the introduction of prenyl groups onto the stilbenoid and dihydrostilbenoid scaffold. As described in the next chapter, we sought to synthesize this representative class of compounds including various prenylated constitutional isomers to analyze their structure-activity-relationships and their activation against both isoforms of *h*ERs. The number and the position of the prenyl groups might also influence their estrogenic activity.

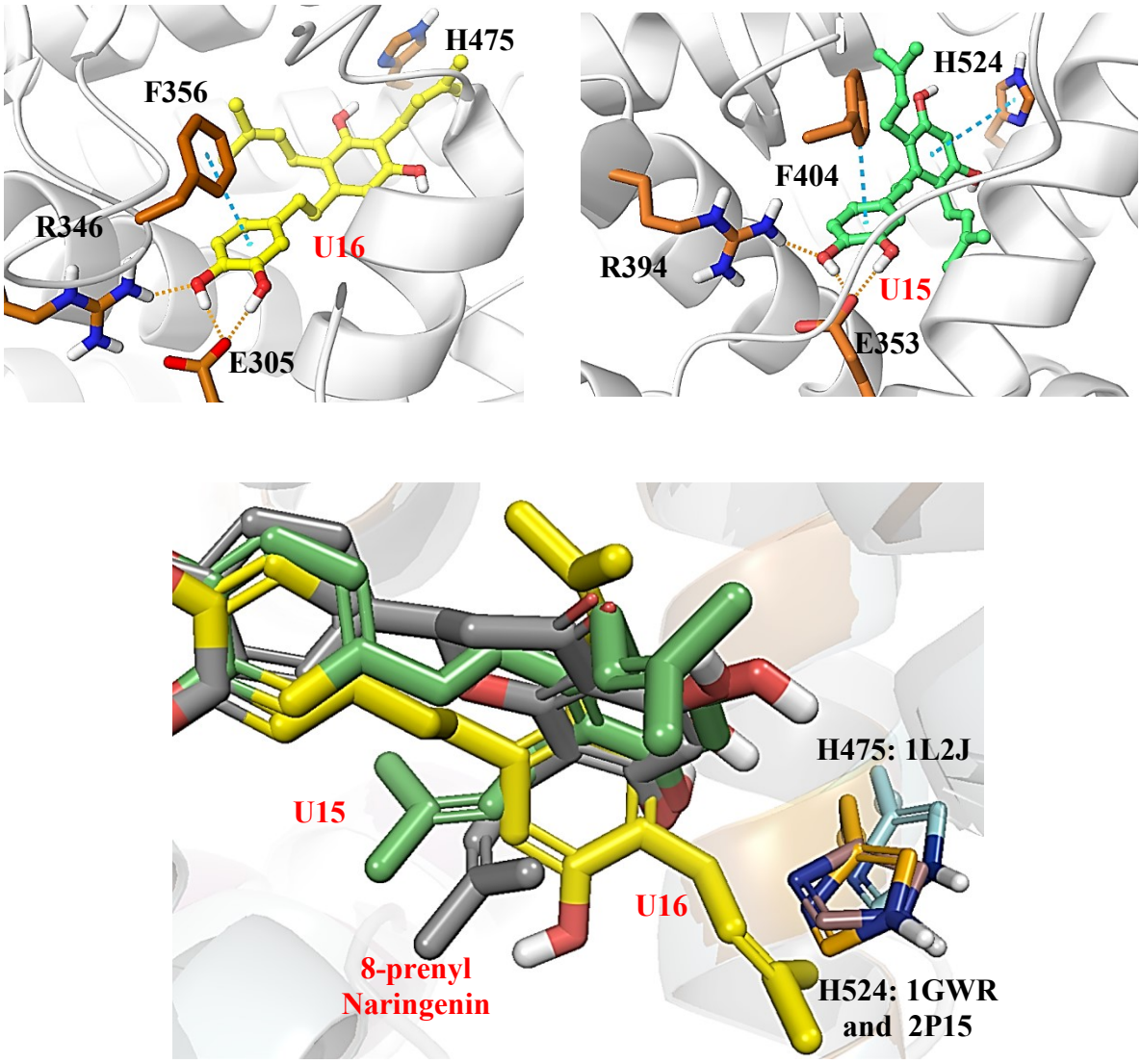


Figure 15. (Up left) Docking pose of **U16** (yellow ball-and-stick) in the LBD of PDB:ID: 1L2J (grey cartoon). **(Up right)** Docking pose of **U15** (green ball-and-stick) in the LBD of PDB:ID: 2P15 (grey cartoon) both showing major interactions with key residues (brown tubes). **(Bottom middle)** The docking poses of **U16** (yellow) and **U15** (green) and 8-prenyl-naringenin (grey) overlaid in PDB structures (1L2J in cyan, 2P15 in orange and 1GWR in light pink), the position of prenyl group in **U15** showing bad interactions with H524 of 2P15 and 1GWR.

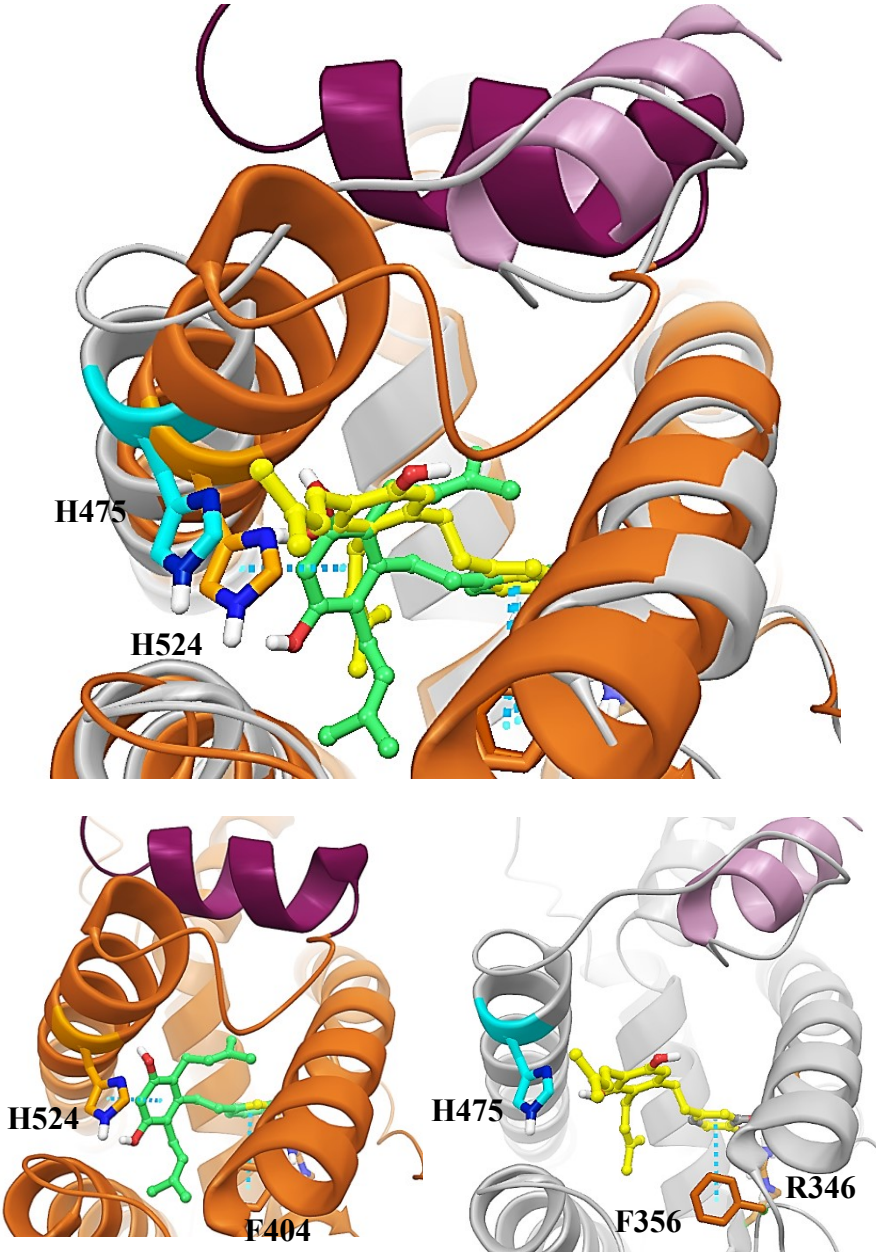


Figure 16. Docking pose of **U15** in PDB 2P15 (green ball-and-stick) and **U16** in PDB 1L2J (yellow ball-and-stick) overlaid with the protein crystal structures **(up)** and showing different positions of H12 (magenta in 2P15 and pink in 1L2J) **(lower left and right)**.

II.3.3 *In silico* evaluation of estrogenic activity for the predicted metabolites

It is now established and supported by literature reports that metabolism changes the parent compounds to a set of metabolites that might have their own spectrum of bioactivities.[111-112] A classic example is a phytoestrogen, daidzein. The intestinal microflora converts daidzein into two different metabolites. One of them is equol, which has stronger estrogenic activity, while the other one is O-desmethylangolensin, which is devoid of estrogenic activity compared to the parent compound, daidzein. It is obvious that inactive chemicals that bear a possible estrogenic metabolite would not be detected in a regular *in vitro* set up.[113] Consequently, it is significant to pre-screen for metabolite estrogenicity.[114] Thus, application of cheminformatic techniques such as QSAR models should be implemented in any early drug discovery programs. In the present study, we utilized an integrated *in silico* approach linking metabolite prediction with qualitative and quantitative assessment of estrogenic activity of both the parent and the predicted metabolites.

ADMET Predictor™ provides both qualitative and quantitative models related to estrogenic activity. Only the compounds classified as estrogenic with "estro_filter" (compounds with a high likelihood to bind to the estrogenic receptors) were submitted to the second model "estro_RBA".[105] The endocrine regression models developed to estimate the numerical values of receptor binding affinity in ADMET Predictor™ were trained using U.S. EPA's Distributed Structure-Searchable Toxicity (DSSTox) database. Competitive binding assays in rats were used as the quantitative measures for building the model. As such, running the calculation for the provided molecule would estimate the degree of binding to the estrogenic receptors expressed as the relative binding affinity %RBA (100%*(IC50 for 17β-estradiol/IC_{50} for the query compound). Higher predicted %RBA value indicates higher affinity. The qualification test of

estrogenicity using ADMET Predictor™ would provide the output data with confidence estimation. The output of the metabolite estrogenic character was further filtered according to the provided confidence ratios. Only those that scored > 70% confidence were further sorted and ranked for their predicted %RBA.

By carrying out this model using our licorice generated libraries, we found that most of the selected estrogenic metabolites with a high level of confidence for estrogenicity are coming from estrogenic parents. While we could not detect a high number of bio-activated compounds, we learned that phase 1 metabolism is not going to inactivate them.

II.4 Experimental

II.4.1 Datasets

The known secondary metabolites of two licorice species, *G. glabra* (179 compounds) and *G. uralensis* (189 compounds) were retrieved from several scientific databases, such as the Dictionary of Natural Products (DNP) and several other publications by a text search in SciFinder Scholar.[115] For the external validation, a focused library of twenty compounds was considered, which included seven known phytoestrogens along with eleven compounds isolated from *Glycyrrhiza* that were already tested for their estrogenic activity. In order to validate the efficiency of incorporating multiple rigid protein crystal structures in our docking procedure, we have considered a diverse group of SERMS and phytoestrogens in the focused library. In addition to the co-crystalized ligands, we have considered *h*ERα selective agonists based on their binding affinity assays (8-prenyl-naringenin, vestitol), or *h*ERβ selective agonists or partial agonists (genistein, daidzein, equol, dehydroequol, liquiritigenin, isoliquiritigenin and calycosin, glabrene), or *h*ERβ selective antagonists (glabridine, glabrol, glicoricone and glyasperin C).

II.4.2 Ligand preparation

The structure libraries were sketched using ChemDraw software and were saved as spatial data files (SDF) using Chem3D. After which, they were imported by Maestro (Schrödinger, LLC) for ligand preparation and data curation. Tautomeric and ionization states were generated at target pH7.4. The chirality of the molecules was determined from the 3D structure input.

II.4.3 Protein preparation

The following protein crystal structures were downloaded from the RCSB protein database. For *h*ERα ensemble, we have considered the following structures: the agonist conformation, estradiol, bound to LBD (PDB:1GWR) [116] , the antagonist conformation or SERM, 4-OH tamoxifen, bound to LBD (PDB: 3ERT), [77] and the hydrophobic agonist conformation, ortho-trifluoromethylphenylvinyl estradiol, bound to LBD (PDB: 2P15).[80] For *h*ER beta ensemble, the structures included the agonist conformation, estradiol, bound to LBD (PDB:4J24) [117], the antagonist conformation, (R,R)-5,11-cis-diethyl-5,6,11,12-tetrahydrochrysene-2,8-diol, bound to LBD (PDB: 1l2J), [118] and the partial agonist conformation, genistein bound to LBD represented by (PDB: 1X7J)[119]. For each protein crystal structure, only one monomeric form was kept and prepared by the protein preparation wizard in Maestro. Hydrogen atoms were added after deleting the original ones. The protein was checked for missing residues and loop segments that were added using Prime. Water molecules with no hydrogen bond were removed. The protonation state and tautomeric state of protic amino acids were adjusted to match a pH 7.4. Possible orientations of side chains were generated and checked in the binding site to match the reported protein-ligand interaction. Finally, the protein-ligand complex was subjected to geometry refinement for hydrogens only, using an OPLS2005

force field. [120]

II.4.4 Sitemap protein binding site analysis

The PDB crystal structures were further studied using sitemap module in Maestro to explore the surface-type characteristic of each LBD.[120] Through the protein analysis panel, the evaluation of a single binding site region was selected to encompass the native ligand in each crystal structure plus 3Å buffer. The binding site examination was set to require at least 15 points with the application of a more restrictive definition of the hydrophobicity and utilizing a standard grid.

II.4.5 Ensemble Docking

For each crystal structure, the grid was calculated using Glide, (Maestro, 10.3.015; Schrodinger).[120] The binding site of each grid was defined by the native ligand with no constraints. The prepared ligands of both species were docked flexibly into the six generated grids of the rigid protein of the ensemble with the help of the virtual screening workflow in Maestro. For each compound, only one docking pose was kept against each crystal structure and scored by Glide SP (standard precision) score. Out of six protein structures, we have mainly considered two ensembles. The first ensemble inclined for active conformation (1GWR, 2P15, and 4J24) or inactive conformation (3ERT, 1X7J, and 1L2J). The highest docking score was considered to assign the preferred conformation in each ensemble. The second ensemble inclined for isoform selectivity. For each compound, only the highest score was kept for each isoform to determine the preferred isoform. To identify the best binders in each species we reviewed the key interactions known for activity and we have computed the simulated binding energy for the top 20 scoring compounds in each isoform ensemble by the molecular mechanics − generalized born surface area model (MM-GBSA). Forty compounds for each species were ranked upon their

MM-GBSA energy allowing the challenge of isoform selectivity to take place. Of note, the top 30 compounds were kept for each species (**Table SI 5** and **SI 6**). Similarly, the top 20 compounds ranked by the predicted RBA values are shown in (**Table SI 7**).

II.4.6 Metabolite prediction and estimation of estrogenic activity via Estro-model

A commercially available software, ADMET Predictor™ 9.0 (Simulations Plus Inc.), provides a friendly interface with a robust prediction environment. Molecular descriptor values are calculated and used to generate independent mathematical models by application of nonlinear machine learning techniques.[105] The built-in trained QSAR models allow the prediction of metabolic oxidation sites mediated by various CYP enzymes. In this work, a pool of primary putative metabolites based on nine cytochrome P450 enzymes (CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP3A4, CYP2A6, CYP2B6, CYP2C8, and CYP2E1) were generated. To launch the metabolite generation in this QSAR package, compounds from licorice were imported as the basic 2D SMILES (simplified molecular-input line-entry system) notations.

Curation of the data included removal of duplicate structures to get the unique compounds for each species as well as sugar hydrolysis using the MedChem Designer module. As such, the number of compounds from *G. glabra* and *G. uralensis* were refined to 146 and 158 respectively. Upon execution of the program with the above parameters, a total of 601 and 708 metabolites were generated for both *glabra* and *uralensis,* respectively. The generated metabolites were used as input for estrogenic activity prediction to analyze the importance of bio-activation as well as for comparison between the two species. ADMET Predictor™ provides both qualitative and quantitative models related to estrogenic activity.

Only the compounds classified as estrogenic with “estro_filter” (compounds with a high likelihood to bind to estrogenic receptors) were submitted to the second model “estro_RBA”.

Consequently, the output of the metabolite estrogenic character was further filtered according to the provided confidence ratios. Only those that scored > 80% confidence were further sorted and ranked for the %RBA.

CHAPTER III
SYNTHESIS OF UNIQUE CHEMICAL ENTITIES AND THEIR ESTROGENIC ACTIVITIES

III.1 Introduction

The previous *in silico* study of licorice against *h*ERs, selected prenylated stilbenoids and DHS as potential components capable to interact differentially with the estrogenic receptors**.** In fact, plant secondary metabolites tethered with isoprenoids appendages provide an immensely rich diversity of biological activities that captured the interest of many scientific communities.[121-123] The vast number of research papers on the chemistry and the biology of many compounds have reflected this impact in the past two decades. From an evolutionary and functional perspective, the interplay between the plants and the surrounding harsh environment has amazingly selected the simplest methods, which coin for diverse effects. The addition of the iconic five carbon units on phenolics or alkaloids systems display the effects of a physicochemical change such as increased hydrophobicity. From a pharmacological point of view, this is usually accompanied by changes in pharmacokinetics and pharmacodynamics of the parent molecule.[112]

Undeniably, the prenylated stilbenoids isolated from a variety of plant sources, have been described as phytoalexins, which show antibacterial, antifungal and antioxidant effects, are indispensable for the plant survival under stressful situations. In fact, these isoprene motives are not exclusive for plant, bacteria or fungi phylogenies, but also they are known to exert vital cellular processes roles in mammals via prenyl transferases.[124] Unsurprisingly, this underpins the

viable crosstalk of prenylated plant secondary metabolites across the living organism. Many prenylated stilbenoids occur in the coniferous species such as *Morus species, Rheum, Abies, Picea, pine, Rheum, and Juniperus* but also found in many other plants.[125] Stilbenoid produced in this line may differ in the number, position, and the type of the prenyls attached (furan and pyran type).[126] Since the prenylated compounds were shown to have profound effects influencing their activity, potency, selectivity, and functionality compared to their parent molecules, the chemical installation of these units is a valued investment.

We have identified the prenylated stilbenoids and prenylated dihydrostilbenoids, **(Figure 17)** found in licorice species as one of its major components that are capable to interact differentially with the estrogenic receptors. To validate our computational data as well as to explore the influential aspects of these appendages to the stilbenoid scaffold, we have envisioned a regiodivergent chemistry that provides multiple isomers in a one-pot reaction. Despite the fact that a regiocontrolled synthesis is preferred, it has proven challenging and multi-step process, usually with low yields. However, in the early stage of pharmacological evaluation of such constitutional isomeric constructs, the straightforward regiodivergent installation of the alkyl group is a viable option. Interesting observations are found in the literature, in how these positional variants may profoundly affect the activity or selectivity of the compounds.

A well-known example is the 8-prenyl naringenin. The presence of the prenyl group at the 8-position converted the flavanone from estrogenically inactive molecule into a highly estrogenic one. The same prenyl group added at 6 position acquired anticancer and androgenic characteristics.[106, 127] On the other hand, the presence of the prenyl group at 8-position transformed the estrogenically active isoflavone or isoflavan structures, genistin and glabridin, into antagonists.[55, 109]

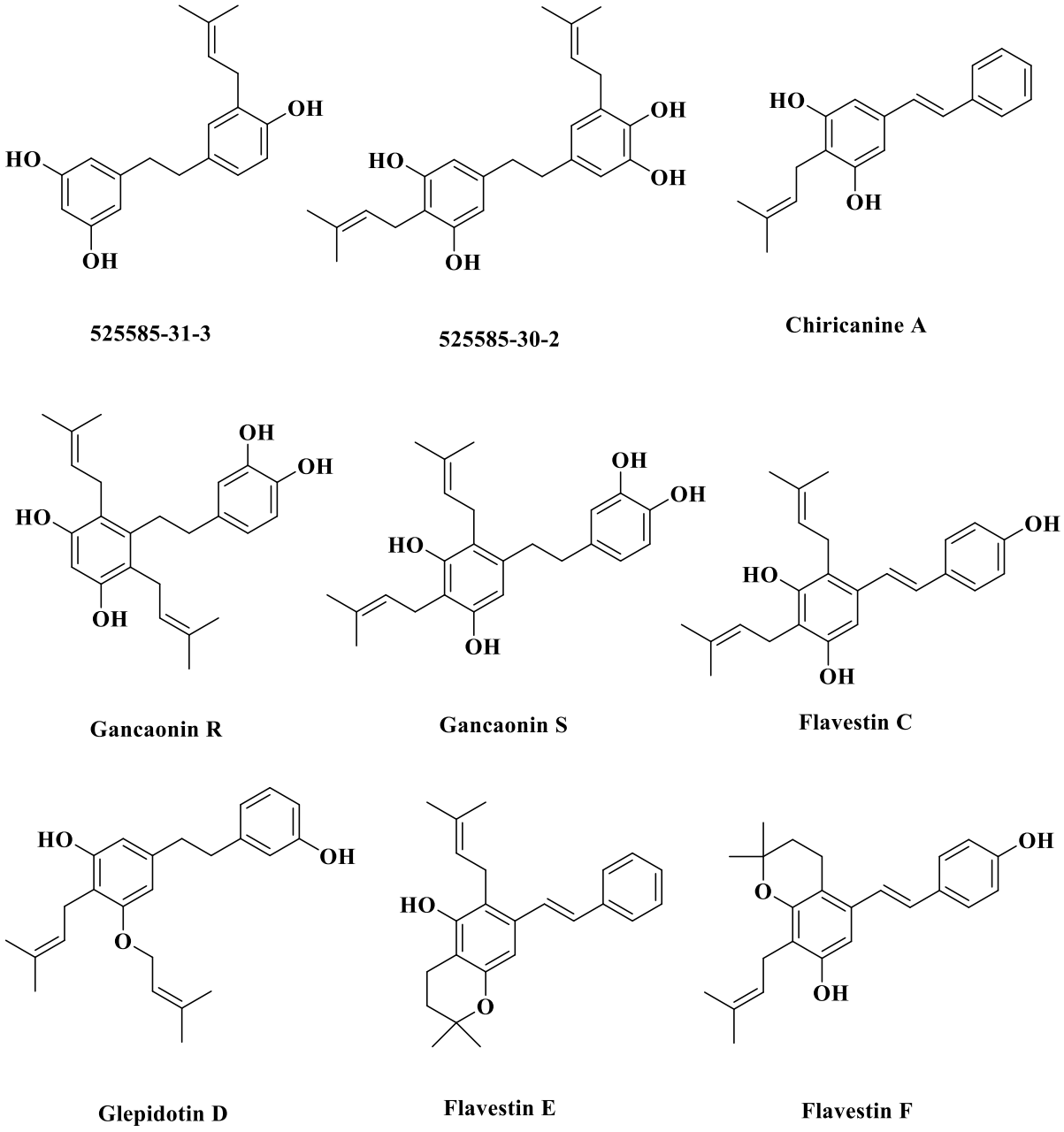


Figure 17. Examples of stilbenoids and DHS reported in different licorice species.

Furthermore, the displacement of the prenyl group from position 3 to 6 in one of the pawhuskin A methylated analog has changed the selectivity of the compound from kappa opioid receptor to a selective delta antagonist.[128] In this study, we report the regiodivergent prenylation of the stilbenoid resveratrol and its dihydro-analog including multiple chemical strategies such as

Friedel-Crafts, ortho-metallation alkylation or the bio-inspired alkylation. During the course of this work, we have identified the plausible analogs in each reaction including major and minor products as well as the new analogs.

III.2 Chemistry

To append/tether a prenyl group on the stilbene scaffold, a flexible synthetic scheme was envisioned to accomplish the synthesis of a diverse set of prenylated stilbenes and dihydrostilbenes starting from commercially available resveratrol.

Access to such analogs is expected to provide ample information on the number of prenyl groups necessary along with the preferred regioisomeric positions related to a specific estrogenic activity along with the isoform selectivity.

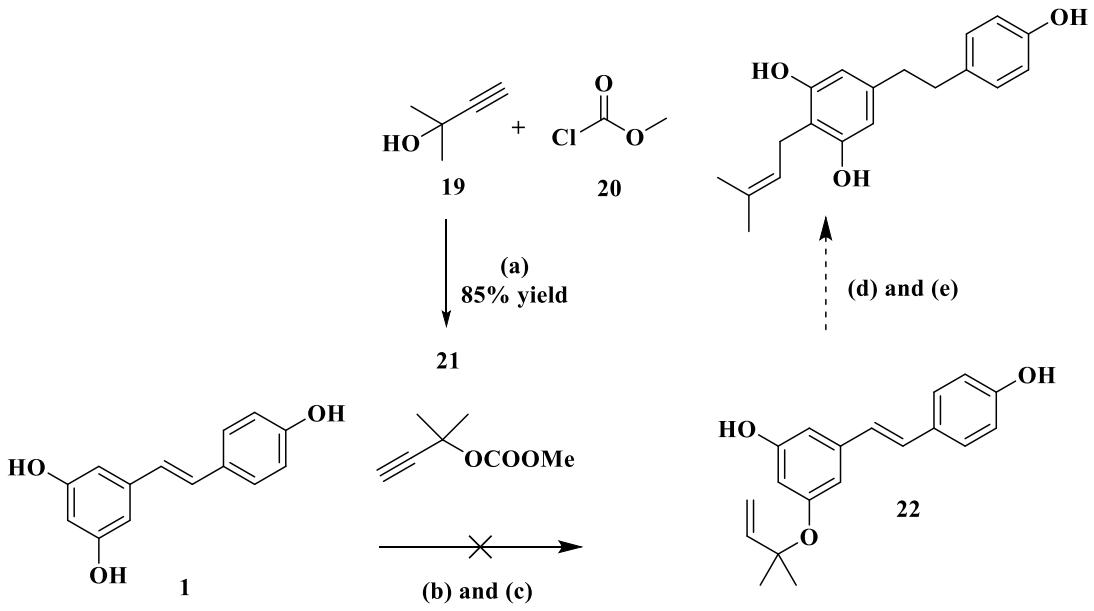
III.2.1 Initial attempts: O-alkylation & Claisen rearrangement and ortho metalation alkylation

The presence of the phenolic group in stilbenes makes them amenable to carbonylation via the carbonate exchange reaction, which could potentially afford several isomeric mono or multiple ether alkyl adducts as precursors for Claisen rearrangement. As shown in **(Scheme 2)**, the readily available propargyl alcohol was converted to the carbonate with Buli and methyl chlorformate. Trans-etherification of the carbonate with the *O*-anion of resveratrol, followed by partial reduction with Lindlarl's catalyst would have resulted in allyl ether.[129] Microwave-assisted Claisen rearrangement of allyl ether would have furnished the prenylated resveratrol. However, we were not able to prepare initial *O*-alkylation with carbonate, which prompted us to search for an alternative synthetic route.

Our second attempt was to employ the nucleophilic addition with ortho metalation, which was inspired by chiricanine A synthesis.[130] The protection of the phenolic moiety in stilbene with methoxymethyl is anticipated to increase the selectivity and stability of the resulting anion under

basic condition. Next, the addition of the prenyl bromide (**24**) will trap the generated anion to furnish the prenylated stilbene. As depicted in **scheme 3 route I**, the hydrogenation of resveratrol under condition (**a**) proceeded with excellent yield (92%).

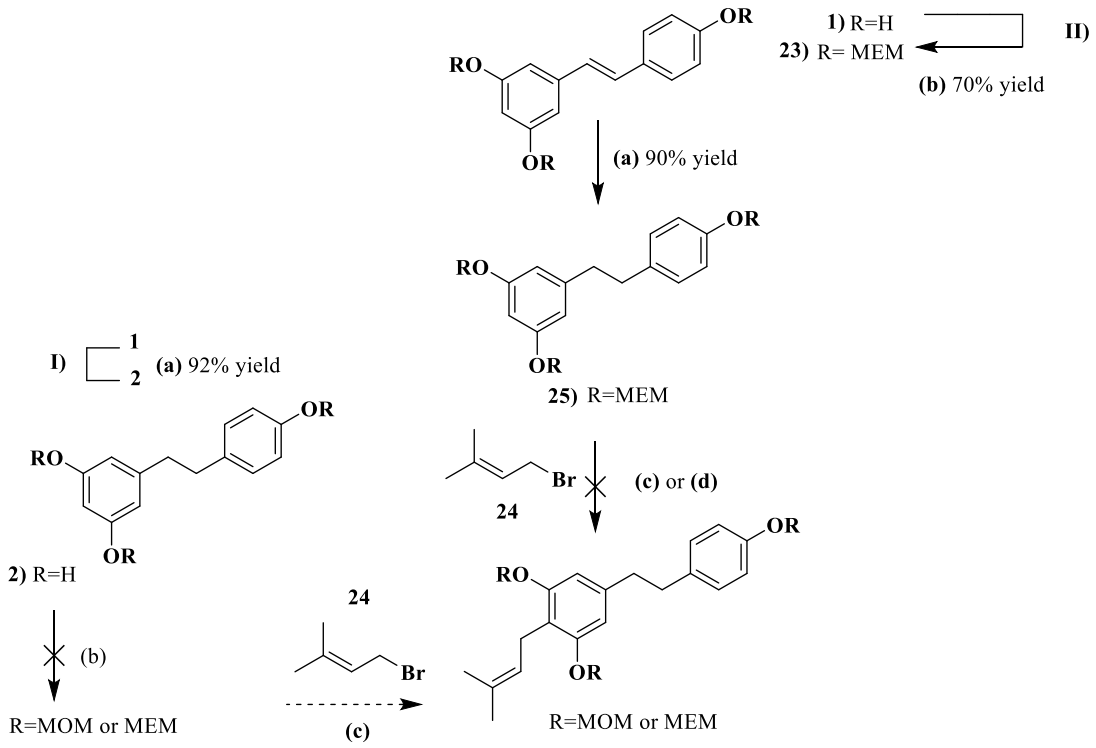
The emerged dihydroresveratrol was treated with methoxymethyl chloride (MOM) in the presence of the hydride anion (NAH) or triethyl amine (TEA) to gain access to the tri-protected product. Unfortunately, our attempts failed; it gave only a slight amount of the required product with NAH and only the mono-protected product was noticed with TEA. Similarly, only trivial amounts of the tri-protected version were detected in the interchange of MOM chloride with methoxyethoxymethyl (MEM) chloride.



Scheme 2. Reagents and conditions: (a) nBuLi, anhyd. THF, 0 ºC (b) $CuCl_2$, DBU, THF at r.t. (c) H_2, Lindlar cat (d) μ-wave, DMF, 180 ºC (e) Pd/H_2, EtOH, r.t.

Alternatively, in **route II**, we started with the stilbene **1** directly to forge the tri-MEM protected product **23** in good yield (70%). Next, Product **23** was subjected to catalytic hydrogenation to afford the intermediate **25** (tri-MEM protected dihydroresveratrol) in 90% yield. Next, the prenyl donor was added under condition **c** or **d** to get the final product.

Unfortunately, the reaction did not proceed in either case.



Scheme 3. Reagents and conditions. (a) Pd/H_2, EtOH, r.t. (b) MOMCl/MEMCl/, HNa/DMF or TEA/DCM, RT, 12 hrs. (c) Prenyl bromide, nBuLi, anhydrous THF, −70 to r.t. °C (d) Prenyl bromide, HNa, anhydrous THF, 0 to r.t. °C.

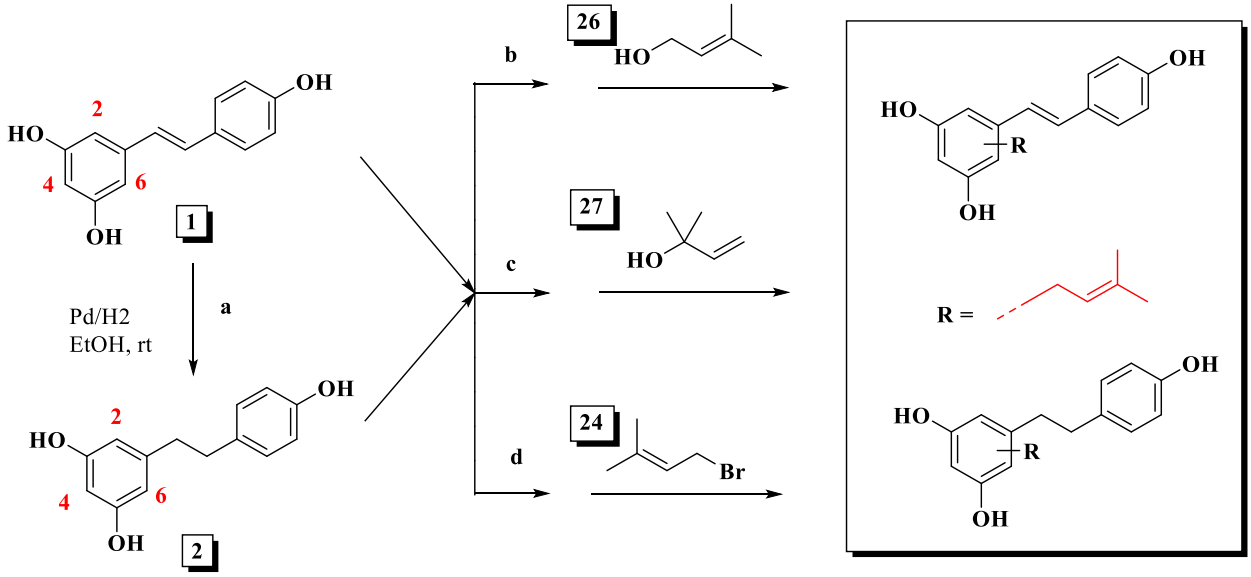
III.2.2 Optimized regiodivergent synthesis of prenylated resveratrol and dihydroresveratrol.

In order to get the target compounds, we have applied both biomimetic and semisynthetic prenylation methods either via electrophilic or nucleophilic additions starting with resveratrol. As depicted in **Scheme 4.** (Condition **b**), heating resveratrol (**1**) or dihydroresveratrol (**2**) with prenyl alcohol (**26**) in acidic conditions led to a regio, divergent installation of the prenyl group at C2, C4 or C6 in different ratios but in low yields. The bio-inspired reaction takes place in acidic aqueous solution through the emerged allylic cation. This reaction resulted in a complex, most diverse set of prenylated analogs along with the corresponding tetrahydropyrano analogs due to harsh, low pH conditions. Formation of the latter analogs is inevitable in this reaction,

where the decrease in time or temperature of the reaction will end up only with the mono-prenylated isomers in low yield. Nevertheless, two new diprenylated (C2 and C4) resveratrol analogs (compounds **M8** and **M9**) were generated under these conditions with a mixed pattern of linear prenyl chain (C2) and cyclized tetrahydropyran products (C4). The generated compounds differ in the mode of cyclization, where compound **M8** dehydrated with C3 hydroxy and compound **M9** is dehydrated with C5 hydroxy as shown by HMBC and NOESY spectra (**Table 4**, **Figure 20,** and **Figure 21**).

Alternatively, the Lewis acid-catalyzed "electrophilic addition" (Friedel-Crafts alkylation) was explored as a milder condition (**c**). As anticipated, the addition of 2,2-dimethylbutenol (**27**) to **1** or **2** at 0 ^{o}C in the presence of $BF_3.OEt_2$ led to the formation of multiple mono- and di-prenylated analogs (with substitution onto positions C2, C4, and C6). This route provided higher yields compared to the previous one and did not lead to cyclization products (tetrahydropyrans moieties). An improved product selectivity was observed as we changed the (substrate: reagent) ratio. Of note, a major C2 monoprenylated product (compound **M2**) was observed at (1:1) ratio. On the other hand, we were pleased to see higher ratios of diprenylated product **M6** (C2, C4) followed by **M7** (C2, C6) preferentially produced at higher ratios as (1:3) or (1:4). As an alternative to electrophilic addition, and to reduce the number of prenylated products, the nucleophilic substitution was explored as outlined in condition **d**. We noticed from the prelude reactions in our initial attempts that we can approach the nucleophilic addition without protection. Thus, direct nucleophilic prenylation of **1** and **2** were investigated under multiple basic conditions including DBU, NaH and basic alumina (Al_2O_3) but they ended up with *O*-prenylation or an inseparable complex mixture.

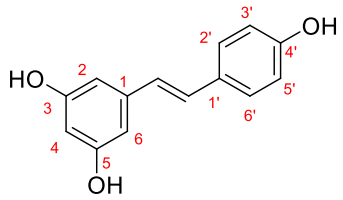
However, deprotonation with *n*-butyl lithium at low temperature followed by the addition of prenyl bromide resulted in mono-prenylated or di-prenylated products.



Scheme 4. Reagents and conditions: (a) Pd/H_2, EtOH, r.t. (b) Prenyl alcohol, EtOH, pH 2.6, 120 °C, 7 hrs. (c) 2,2-Dimethylbutenol, $BF_3.OEt_2$, dioxane, 0 ºC, 12 hrs. (d) Prenyl bromide, nBuLi, anhydrous THF, −40 °C.

Interestingly, this route showed higher selectivity profiles compared to the previous ones as the diprenylated resveratrol **M7** (C2, and C6) emerged in higher amounts. Along with the required *C*-alkylation products, easily separable *O*-alkylated products and cyclized products were also noticed under this condition but in very low yields. However, starting with **2** under these conditions, the mono-tetrahydropyran versions (compounds **M14** and **M15**) and a minor double tetrahydropyran dihydroresveratrol became a major product (compound **M18**) upon purification. Overall, 16 different analogs of both resveratrol (**M2-M10**) and dihydroresveratrol (**M12-M18**) were prepared; five of them were synthesized for the first time (**Figure 18** and **Figure 19**). The overall yields and selectivities are summarized in **Table 3**.

Table 3. Summarized reaction yields and conditions



Route	Starting material	Reagent equiv.	Conditions (°C, h)	Yield%									
				P2	P4	P2ca	P4c	P2, P4	P2, P6	P2, P4c	P2, P4c	P2c, P6	P2c, P6c
B	1	1:2	115, 7h	12%	6%	-	-	-	-	-	-	-	-
B	1	1:2	100, 24h	9%	-	4%	1.5%	-	-	-	-	-	-
B	1	1:2	120, 24h	22%	-	6%	2.5%	4.5%	2%	<1%	<1%	<1%	-
C	1	1:2	r.t., 24h	27%	8%	-	-	9%	3.5%	-	-	-	-
C	1	1:4	r.t., 24h	10%	-	-	-	19%	15%	-	-	-	-
D	1	1:2	−40-25, 24h	28%	-	<1%	-	4%	18%	-	-	3.5%	-
B	2	1:2	115, 7h	12%	4%	-	-	-	-	-	-	-	-
C	2	1:3	r.t., 24h	4%	6%	-	-	20%	5%	-	-	-	-
D	2	1:2	−40-25, 24h	-	-	13%	6%	-	-	-	-	-	8%
D	2	1:2	−40-25 24h	4%	2.5%	-	12%	<1%	<1%	-	-	-	<1%
D	2	1:3	−40-0-25 24h	7%	3.5%	-	-	2%	6%	-	-	<1%	-

a) c is for cyclized

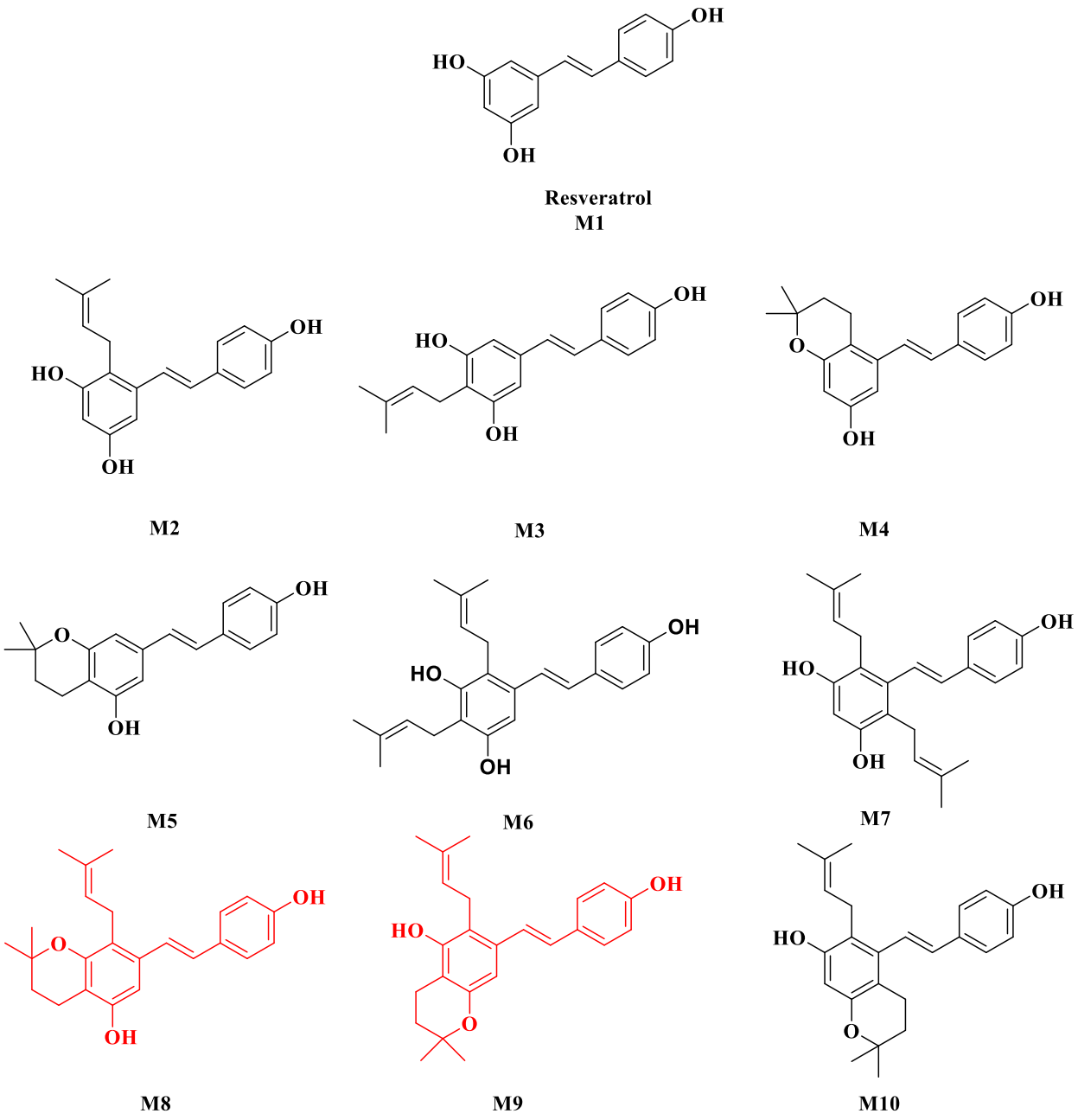


Figure 18. Synthesized resveratrol derivatives **M2-M10**

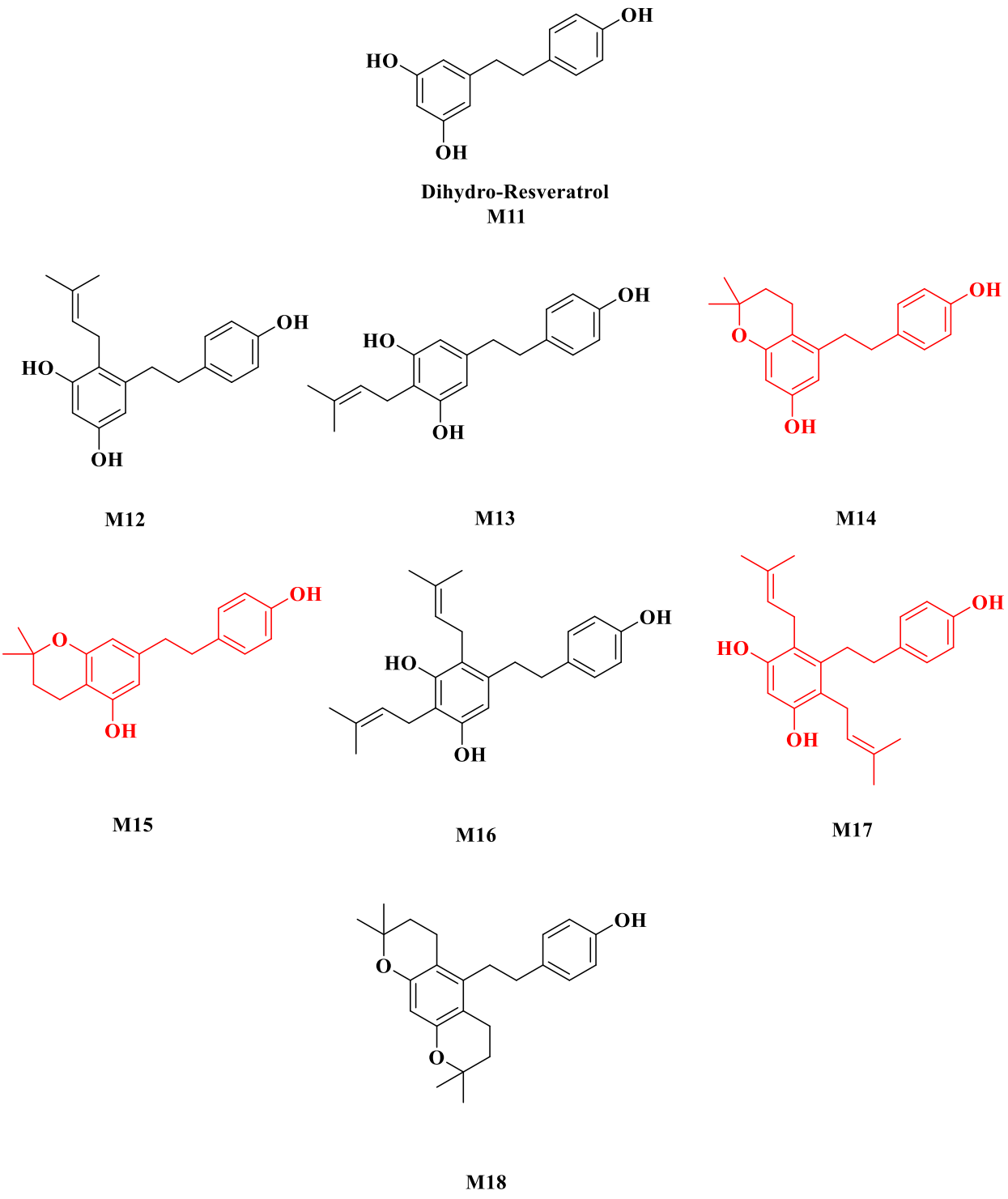


Figure 19. Synthesized dihydro-resveratrol derivatives **M12-M18**

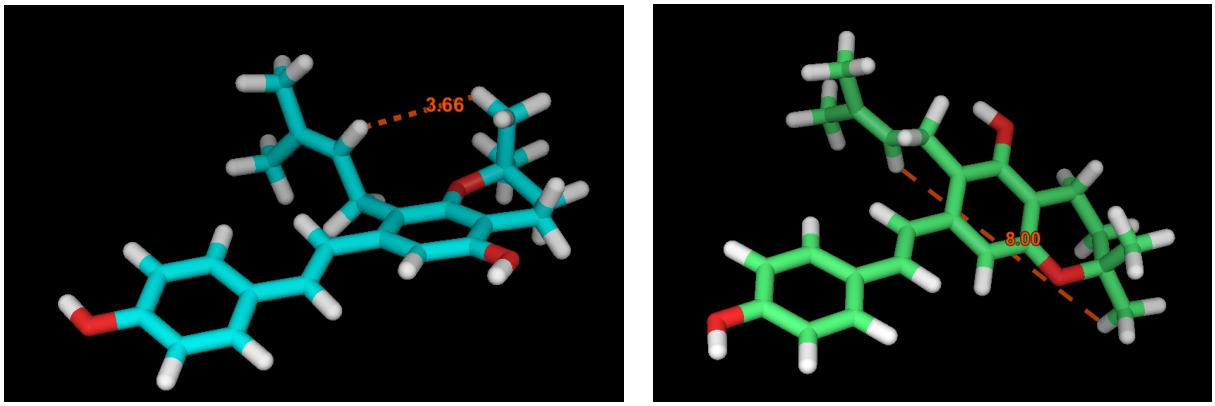
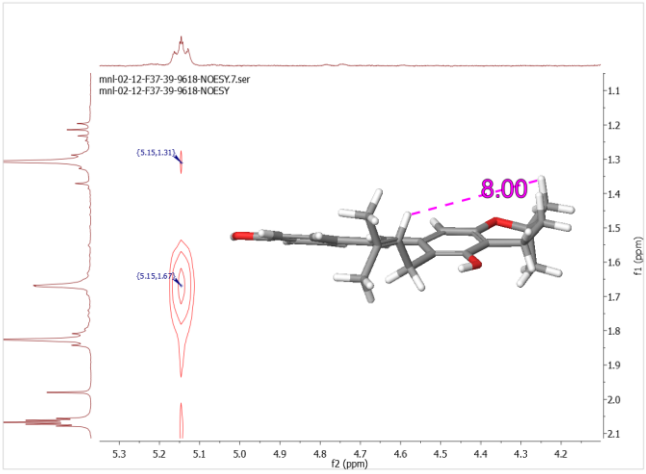
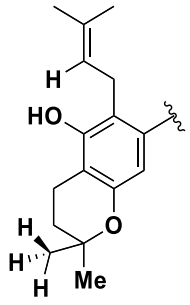
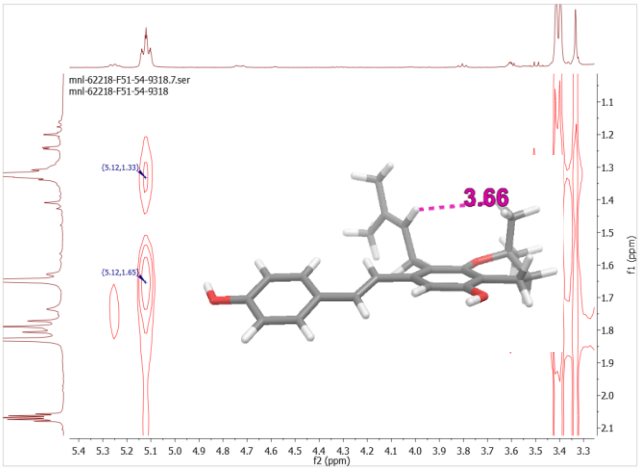


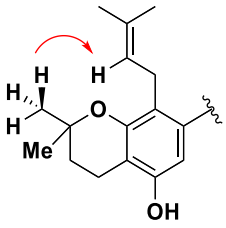
Figure 20. Energy-minimized structures **M8** (cyan tubes) and **M9** (green tubes) with the measured distances between the olefin proton of the straight chain prenyl and the methyl proton of the neighboring THP ring (3.66 and 8 Å, respectively)





M9

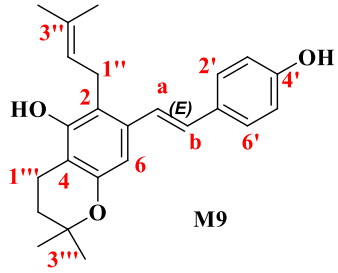




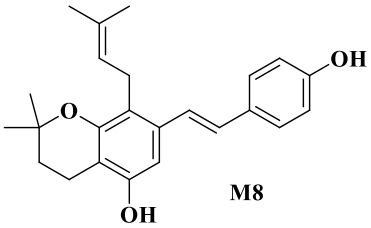
M8

Figure 21. Lack of NOE correlation between olefin proton and the methyl group on the THP ring in **M9** compared to **M8**

Table 4. NMR data for compounds **M8** and **M9**



M9



M8

C/H no	M9			M8		
	δ_C	δ_H, mult. (*J*Hz)	HMBC	δ_C	δ_H, mult. (*J*Hz)	HMBC
1	135.39, C			134.84, C		
2	117.43, C			118.69, C		
3	152.62, C			153.31, C	152.29, C	
4	108.10, C			107.93, C		
5	152.65, C			152.29, C		
6	105.51, CH	6.63 (s, 1H)	a, 2, 4, 5	102.33, CH	6.71 (s, 1H)	a, 2, 4
a	123.69, CH	7.20 (d, J = 16.1 Hz, 1H)	b, 2, 6	124.07, CH	7.24 (d, J = 16.1 Hz, 1H)	6, 2, 1′
b	129.00, CH	6.93 (d, J = 16.1 Hz, 1H)		128.63, CH	6.84 (d, J = 16.3 Hz, 1H)	1, 2′, 6′
1′	129.84, C			129.66, C		
2′	127.75, CH	7.44 (d, J = 8.5 Hz, 2H)	b, 1′, 4′, 5′	127.63, CH	7.40 (d, J = 8.6 Hz, 2H)	6′, 4′
3′	115.48, CH	6.86 (d, J = 8.5 Hz, 2H)		115.38, CH	6.85 (d, J = 8.6 Hz, 2H)	4′
4′	157.13, C			157.21, C		
5′	115.48, CH	6.86 (d, J = 8.5 Hz, 2H)		115.38, CH	6.85 (d, J = 8.6 Hz, 2H)	
6′	127.75, CH	7.44 (d, J = 8.5 Hz, 2H)		127.63, CH	7.40 (d, J = 8.6 Hz, 2H)	
1″	24.49, CH_2	3.49 (d, J = 6.6 Hz, 2H)	1 ,2, 2″	24.11, CH2	3.41 (d, J = 7.0 Hz, 2H)	1, 2, 3, 2″, 3″
2″	124.3, CH	5.14 (t, J = 6.7 Hz, 1H)		124.52, CH	5.12 (t, J = 7.0 Hz, 1H)	
3″	129.66, C			128.72, C		1″, 4″, 5″
4″	17.21, CH_3	1.83 (s, 3H)		17.30, CH3	1.83 (s, 3H)	2″, 3″
5″	25.00, CH_3	1.67 (s, 3H)		25.07, CH3	1.65 (s, 3H)	2″, 3″, 4″
1‴	17.53, CH_2	2.71 (t, J = 6.8 Hz, 2H)	2‴, 3‴, 3, 4	17.27, CH2	2.68 (t, J = 6.9 Hz, 2H)	2‴, 3‴, 4, 5
2‴	32.17, CH_2	1.83 (t, J = 6.8 Hz, 2H)	3‴, 4	31.88, CH2	1.79 (t, J = 6.9 Hz, 2H)	1‴,4 , 3‴
3‴	72.80, C			73.27, C		3‴, 2‴,1‴
4‴	25.99, CH_3	1.30 (s, 6H)	2‴, 3‴	26.16, CH3	1.32 (s, 6H)	5‴, 3‴
5‴	25.99, CH_3	1.30 (s, 6H)		26.16, CH3	1.32 (s, 6H)	4‴, 5‴

III.3 Experimental

III.3.1 Conformational analysis of compounds M8 and M9

The conformational search was applied with MacroModel (Schrodinger, LLC) utilizing advanced search with Low-frequency-Mode and OPLS2005 for energy minimization. The probability of TORS/MOLS steps was set to 0.5 with a maximum number of steps was equal to 1000. The energy window for saving a structure was set to 5.02 kcal/mol to eliminate the higher energy conformations that could be generated. The maximum deviation between conformers was set to a cutoff of 0.5Å to eliminate the redundant structures. The number of structures generated by this method was 1387 structures, out of which 888 were successfully minimized. The sampled conformational populations were assembled based on the Boltzmann potential energy-weighted populations (relative Potential-Energy-OPLS-2005 in KJ/mol at 298.15 K).

III.3.2 General experimental procedures

All the reactions were monitored using thin layer chromatography. All the chemicals and reagents were purchased from Sigma Aldrich (MO, USA). Agilent 630 FTIR (Agilent Technologies, CA, USA) was used to record the IR spectra. ^{1}H and ^{13}C NMR spectral data were recorded on Bruker (400 and 500 MHz; Bruker AU III, MA, USA) spectrometers and chemical shifts were expressed as p.p.m. Agilent UHPLC 6200 series (Agilent Technologies) was used to acquire Mass spectra. Column chromatography was performed on (Merck, MA, USA) silica gel 60 and Sephadex LH20 (sigma). Flash silica gel (40μm, 60 Å, J. T. Baker) and reversed phase RP-C18 silica (Polarbond, J.T. Baker). Preparative-TLC plates were used for extra purification (20 cm × 20 cm, 500μm).

III.3.3 Chemistry

Condition b: Biomimetic prenylation of stilbenoids and dihydrostilbenoids using prenyl alcohol

The stirred solution of resveratrol (500 mg or 2g) or dihydroresveratrol (250 mg or 500 mg) in ethanol was adjusted to pH 2.6 using 500 mg of citric acid dissolved in 5 ml of water,then prenyl alcohol (2 or 4 equivalents) was added. The reaction mixture was heated up to 120 °C gradually in a sealed vessel using silicon oil bath for 7 or 24 hours. The reaction mixture was quenched with aqueous $NaHCO_3$ and the water layer was extracted three times (15 ml) with ethyl acetate. The combined organic layer was dried over $MgSO_4$ and concentrated under reduced pressure. Column chromatography (SiO_2, EtOAc−Hex = 15:85 to 50:50) provided compounds **M2-M10** and **M12-M15**.

Condition c: Prenylation of stilbenoids and dihydrostilbenoids *via* electrophilic addition (Friedel-Crafts alkylation)

To an ice cooled solution of resveratrol or dihydroresveratrol in dry dioxane was slowly added $BF_3.Et_2O$ followed by (2-methyl-but-3-en-2-ol). The reaction mixture was stirred at r.t. overnight. The reaction mixture was evaporated under vacuum and partitioned between H_2O (35 ml) and EtOAc (3x15). The combined organic layers were washed with H_2O (1x15), dried over MgSO4, and concentrated under reduced pressure. Flash chromatography (SiO_2, Acetone-Hex = 5:95 to 75:25) provided compounds **M2, M3, M6, M7, M12, M13, M16** and **M17**.

Condition d: **Prenylation of stilbenoids and dihydrostilbenoids via nucleophilic substitution using prenyl bromide and n-BuLi**

To a stirred solution of resveratrol (300mg) or dihydroreveratrol (1g, 750mg) in dry THF at −40 °C was added dropwise n-BuLi. After 20 min the reaction was allowed to reach −20 °C, at which 3,3-dimethylallyl bromide was added slowly and left to stir for 3 hrs. Then the reaction temperature was raised gradually to the room temperature and was stirred for an additional 12 hrs. The reaction mixture was quenched with saturated aqueous NH_4Cl and the water layer was extracted three times EtOAc (15 ml). The combined organic layers were dried over Na_2SO_4 and concentrated under with reduced pressure. Column chromatography (SiO_2, EtOAc−Hex = 15:85 to 50:50) provided compounds **M2**, **M3**, **M6**, **M7**, **M10** and **M12-M18**.

III.3.4 Spectral data

M2
IR: 3339.7, 1587.8, 1353.0, 1444.3, 1244.9, 1131.2, 821.9. **^{1}H NMR** (400 MHz, MeOD):7.33 (d, *J* = 8.6 Hz, 2H, H-2', H-6'), 7.14 (d, *J* = 16.1 Hz, 1H, H-α), 6.84 (d, *J* = 16.1 Hz, 1H, H-β), 6.78 (d, *J* = 8.7 Hz, 2H, H-3', H-5'), 6.57 (d, *J* = 2.4 Hz, 1H, H-6), 6.25 (d, *J* = 2.4 Hz, 1H, H-4), 5.13 (m, 1H, H-2"), 3.39 (d, *J* = 6.8, 2H, H-1"), 1.81 (s, 3H, H-4"), 1.69 (s, 3H, H-5"). **^{13}C NMR** (100 MHz, MeOD) δ 156.85 (C-4'), 155.64 (C-5), 155.33 (C-3), 138.29 (C-1), 129.50 (C-1'), 129.31 (C-β), 129.00 (C-3"), 127.30 (C-2', C-6'), 124.28 (C-2'), 123.93 (C-α), 117.69 (C-2), 115.08 (C-3', C-5'), 102.88 (C-6), 101.20 (C-4), 24.56 (C-4"), 23.84 (C-1"), 16.74 (C-5"). **HRMS (ESI)** calcd for $C_{19}H_{20}O_3$ (MH^+), 297.1485; found 297.1484.

M3
IR: 3281.9, 1509.6, 1425.7, 1161.1, 1041.8, 965.4. **^{1}H NMR** (400 MHz, MeOD) δ 7.34 (d, *J* = 8.6 Hz, 2H), 6.92 (d, *J* = 16.3 Hz, 1H), 6.76 (d, *J* = 16.3 Hz, 1H), 6.78 (d, *J* = 8.6 Hz, 2H), 6.48 (s, 1H), 5.26 (m, 1H), 3.30 (d, *J* = 7.1 Hz, 2H), 1.78 (s, 3H), 1.68 (s, 3H). **^{13}C NMR** (100 MHz, MeOD) δ 156.75, 155.62, 155.24, 138.38, 129.58, 129.47, 129.10, 127.39, 124.22, 123.97, 117.86, 115.14, 103.04, 101.29, 24.64, 23.91, 16.83. **HRMS (ESI)** calcd for $C_{19}H_{20}O_3$ (MH^+), 297.1485; found 297.1486

M4
IR: 3339.7, 1582.3, 1442.5, 1166.7, 1019.4, 836.8. **^{1}H NMR** (400 MHz, acetone-d_6) δ 7.47 (d, *J* = 8.6 Hz, 2H), 7.18 (d, *J* = 16.1 Hz, 1H), 6.96 (d, *J* = 16.1 Hz, 1H), 6.86 (d, *J* = 8.6 Hz, 2H), 6.71 (d, *J* = 2.4 Hz, 1H), 6.20 (d, *J* = 2.4 Hz, 1H), 2.78 (t, *J* = 6.8 Hz, 2H) 1.82 (t, *J* = 6.8 Hz, 2H), 1.30 (s, 6H). **^{13}C NMR** (100 MHz, acetone-d_6) δ 157.31, 156.30, 155.05, 137.95, 129.79, 129.37, 127.96, 122.87, 115.53, 110.21, 104.11, 103.06, 73.06, 32.73, 25.98, 19.51. **HRMS (ESI)** calcd

for $C_{19}H_{20}O_3$ (MH^+), 297.1485; found 297.1488.

M5

IR: 3308.0, 1547.8, 1511.4, 1422.0, 1116.3, 1053.0, 834.9. **^{1}H NMR** (400 MHz, acetone-d_6) δ 7.42 (d, J = 8.6 Hz, 2H), 7.00 (d, J = 16.3 Hz, 1H), 6.86 (d, J = 16.3 Hz, 1H), 6.85 (d, J = 8.5 Hz, 2H), 6.60 (d, J = 1.4 Hz, 1H), 6.49 (d, J = 1.3 Hz, 1H), 2.67 (t, J = 6.8 Hz, 2H), 1.80 (t, J = 6.8 Hz, 2H), 1.31 (s, 6H). **^{13}C NMR** (100 MHz, acetone-d_6) δ 157.13, 155.62, 155.16, 136.88, 129.21, 127.72, 127.72, 127.45, 125.93, 115.51, 115.51, 108.03, 106.50, 103.89, 73.40, 32.07, 26.07, 26.07, 17.05. **HRMS (ESI)** calcd for $C_{19}H_{20}O_3$ (MH^+), 297.1485; found 297.1487.

M6

IR: 3369.5, 1578.5, 1431.3, 1168.5, 1034.3, 838.1. **^{1}H NMR** (500 MHz, acetone-d_6) δ 7.39 (d, *J* = 7.9 Hz, 2H), 7.20 (d, *J* = 16.0 Hz, 1H), 6.85 (d, *J* = 8.5 Hz, 2H), 6.79 (d, *J* = 16.1 Hz, 1H), 6.76 (s, 1H), 5.26 (t, *J* = 7.0 Hz, 1H), 5.13 (t, *J* = 6.1 Hz, 1H), 3.48 (d, *J* = 6.6 Hz, 2H), 3.44 (d, *J* = 6.9 Hz, 2H), 1.82 (s, 3H), 1.79 (s, 3H) 1.67 (s, 6H). **^{13}C NMR** (125 MHz, acetone-d_6) δ 157.11, 153.57, 153.36, 135.19, 130.85, 129.97, 129.56, 128.79, 127.66, 127.66, 124.26, 124.11, 123.11, 118.07, 115.52, 115.52, 114.66, 104.12, 25.03, 24.99, 24.70, 22.56, 17.24, 17.09. **HRMS (ESI)** calcd for $C_{24}H_{28}O_3$ (MH^+), 365.2111; found 365.2110.

M7

IR: 3317.3, 1589.7, 1442.5, 1151.7, 1086.5, 903.5. **^{1}H NMR** (500 MHz, acetone-d_6) δ 7.38 (d, *J* = 8.5 Hz, 2H), 6.98 (d, *J* = 16.7 Hz, 1H), 6.87 (d, *J* = 8.4 Hz, 2H), 6.44 (d, *J* = 16.6 Hz, 1H), 6.45 (s, 1H), 5.20 (dt, *J* = 8.4, 4.2 Hz, 2H), 3.35 (d, *J* = 4.7 Hz, 4H), 1.65 (s, 6H), 1.63 (s, 6H). **^{13}C NMR** (100 MHz, acetone-d_6) δ 157.06, 153.44, 153.44, 139.55, 133.30, 129.53, 128.95, 128.95, 127.46, 127.46, 124.90, 124.90, 124.21, 117.55, 117.55, 115.45, 115.45, 101.43, 25.95, 25.95, 25.08, 25.08, 17.31, 17.31. **HRMS (ESI)** calcd for $C_{24}H_{28}O_3$ (MH^+), 365.2111; found 365.2111

M8

IR: 3261.4, 1578.5, 1422.0, 1159.2, 1072.2, 821.9. **^{1}H NMR** (400 MHz, acetone-d_6) δ 7.40 (d, *J* = 8.6 Hz, 2H), 7.24 (d, *J* = 16.1 Hz, 1H), 6.85 (d, *J* = 8.6 Hz, 2H), 6.84 (d, *J* = 16.3 Hz, 1H), 6.71 (s, 1H), 5.12 (t, *J* = 7.0 Hz, 1H), 3.41 (d, *J* = 7.0 Hz, 2H), 2.68 (t, *J* = 6.9 Hz, 2H), 1.83 (s, 3H), 1.79 (t, *J* = 6.9 Hz, 2H), 1.65 (s, 3H), 1.32 (s, 6H). **^{13}C NMR** (100 MHz, acetone-d_6) δ 157.08, 153.31, 152.27, 134.89, 129.61, 128.96, 128.53, 127.65, 124.48, 124.04, 118.85, 115.52, 107.93, 102.33, 73.36, 31.84, 26.13, 25.06, 24.04, 17.34, 17.27. **HRMS (ESI)** calcd for $C_{24}H_{28}O_3$ (MH^+), 365.2111; found 365.2110.

M9

IR: 3380.2, 15559.9, 1429.4, 1157.3, 1094.0, 838.7. **^{1}H NMR** (500 MHz, acetone-d_6) δ 7.44 (d, *J* = 8.5 Hz, 2H), 7.20 (d, *J* = 16.1 Hz, 1H), 6.93 (d, *J* = 16.1 Hz, 1H), 6.86 (d, *J* = 8.5 Hz, 2H), 6.63 (s, 1H), 5.14 (t, *J* = 6.7 Hz, 1H), 3.49 (d, *J* = 6.6 Hz, 2H), 2.71 (t, *J* = 6.8 Hz, 2H), 1.83 (t, *J* = 6.8 Hz, 2H), 1.83 (s, 3H), 1.67 (s, 3H), 1.30 (s, 6H). **^{13}C NMR** (125 MHz, acetone-d_6) δ 157.13, 152.63, 135.79, 129.79, 129.63, 129.02, 127.73, 127.73, 124.31, 123.72, 117.45, 115.50, 115.50, 108.11, 105.46, 72.88, 72.88, 32.16, 25.98, 25.98,25.00, 24.50, 17.55, 17.25. **HRMS (ESI)** calcd for $C_{24}H_{28}O_3$ (MH^+), 365.2111; found 365.2108

M10
IR: 3360.2, 1559.9, 1429.4, 1157.3, 1094.0, 838.7. **^{1}H NMR** (400 MHz, acetone-d_6) δ 7.41 (d, *J* = 8.6 Hz, 2H), 6.96 (d, *J* = 16.6 Hz, 1H), 6.87 (d, *J* = 8.6 Hz, 2H), 6.53 (d, *J* = 16.7 Hz, 1H), 6.27 (s, 1H), 5.22 (tdd, *J* = 5.4, 2.9, 1.4 Hz, 1H), 3.38 (d, *J* = 6.7 Hz, 2H), 2.67 (t, *J* = 6.7 Hz, 2H), 1.74 (t, *J* = 6.7 Hz, 2H), 1.66 (d, *J* = 1.5 Hz, 3H), 1.65 (d, *J* = 1.5 Hz, 3H), 1.29 (s, 6H). **^{13}C NMR** (100 MHz, acetone-d_6) δ 157.18, 154.08, 152.54, 138.50, 133.60, 129.36, 129.06, 127.52, 127.52, 124.84, 123.53, 118.34, 115.45, 115.45, 110.40, 102.36, 72.81, 33.10, 26.11, 26.1, 1 25.75, 25.02, 21.25, 17.27. **HRMS (ESI)** calcd for $C_{24}H_{28}O_3$ (MH^+), 365.2111; found 365.2130.

M12
IR: 3295.0, 1593.4, 1448.1, 1131.2, 11131.2, 779.0. **^{1}H NMR** (400 MHz, acetone-d_6) δ 7.07 (d, J = 8.5 Hz, 2H), 6.77 (d, *J* = 8.4 Hz, 2H), 6.30 (d, *J* = 2.5 Hz, 1H), 6.26 (d, *J* = 2.4 Hz, 1H), 5.13 (tdd, *J* = 5.4, 2.8, 1.5 Hz, 1H), 3.31 (d, *J* = 6.6 Hz, 2H), 2.75 (s, 4H), 1.74 (d, *J* = 1.3 Hz, 3H), 1.66 (d, *J* = 1.5 Hz, 3H). **^{13}C NMR** (100 MHz, MeOD) δ 155.65, 155.11, 154.96, 142.01, 133.03, 129.16, 128.88, 128.88, 124.58, 117.64, 114.64, 114.64, 107.23, 99.95, 36.59, 35.44, 24.54, 23.88, 16.71. **HRMS (ESI)** calcd for $C_{19}H_{22}O_3$ (MH^+), 299.1642; found 299.1643

M13
IR: 3233.5, 1589.7, 1438.8, 1115.5, 1034.3, 980.3. **^{1}H NMR** (400 MHz, acetone-d_6) δ 7.03 (d, *J* = 8.4 Hz, 2H), 6.75 (d, *J* = 8.5 Hz, 2H), 6.28 (s, 2H), 5.31 (ddt, *J* = 7.2, 5.8, 1.4 Hz, 1H), 3.34 (d, *J* = 7.2 Hz, 1H), 2.81-2.63 (m, 4H), 1.77 (d, *J* = 1.3 Hz, 3H), 1.65 (d, *J* = 1.4 Hz, 3H). **^{13}C NMR** (100 MHz, acetone-d_6) δ 155.80, 155.80, 155.40, 140.47, 132.75, 129.53, 129.21, 129.21, 123.80, 115.01, 115.01, 112.21, 106.93, 106.93, 37.85, 36.69, 25.07, 22.02, 17.05. **HRMS (ESI)** calcd for $C_{19}H_{22}O_3$ (MH^+), 299.1642; found 299.1642

M14
IR: 3285.8, 1513.3, 1444.3, 1136.8, 1012.0, 834.9. **^{1}H NMR** (400 MHz, MeOD) δ 6.95 (d, J = 8.5 Hz, 2H), 6.68 (d, J = 8.5 Hz, 2H), 6.23 (d, J = 2.5 Hz, 1H), 6.06 (d, J = 2.5 Hz, 1H), 2.77 – 2.69 (m, 4H), 2.48 (t, J = 6.8 Hz, 2H), 1.72 (t, J = 6.8 Hz, 1H) 1.25 (s, 6H). **^{13}C NMR** (100 MHz, MeOD) δ 155.56, 155.04, 154.53, 141.47, 132.68, 129.04, 129.04, 114.63, 114.63, 110.50, 108.01, 101.25, 73.09, 35.84, 34.67, 32.76, 25.53, 25.53, 18.87. **HRMS (ESI)** calcd for $C_{19}H_{22}O_3$ (MH^+), 299.1642; found 299.16.14.

M15
IR: 3360.33, 1514.68, 1120.1, 1057.36, 830.69. **^{1}H NMR** (500 MHz, acetone-d_6) 7.05 (d, J = 8.4 Hz, 2H), 6.77 (d, J = 8.4 Hz, 2H), 6.30 (d, J = 1.6 Hz, 1H), 6.19 (d, J = 1.5 Hz, 0H), 2.82 – 2.73 (m, 2H), 2.72-2.66 (m, 1H), 2.63 (t, J = 6.8 Hz, 2H), 1.76 (t, J = 6.8 Hz, 2H), 1.28 (s, 6H). **^{13}C NMR** (125 MHz, acetone-d_6 δ 155.40, 155.27, 154.85, 141.01, 132.79, 129.25, 129.25, 115.06, 115.06, 108.40, 106.19, 105.99, 73.21, 38.01, 36.72, 32.20, 26.15, 26.15, 16.86. **HRMS (ESI)** calcd for $C_{19}H_{22}O_3$ (MH+), 299.1642; found 299.1642.

M16
IR: 3390.0, 1591.6, 1425.7, 1170.4, 1041.8, 827.5. **^{1}H NMR** (400 MHz, acetone-d_6) 7.05 (d, J = 8.5 Hz, 2H), 6.77 (d, J = 8.5 Hz, 1H), 6.35 (s, 1H), 5.26 (dddd, J = 7.1, 5.6, 2.9, 1.5 Hz, 1H), 5.09 (tq, J = 5.0, 1.5 Hz, 1H), 3.40 (d, J = 7.3 Hz, 2H), 3.34 (d, J = 6.6 Hz, 2H), 2.72 (s, 4H), 1.78 (d, J = 1.3 Hz, 3H), 1.75 (d, J = 1.3 Hz, 3H), 1.67 (d, J = 1.5 Hz, 6H). **^{13}C NMR** (100 MHz, acetone-

d$_6$) δ 155.48, 153.49, 153.41, 138.57, 132.91, 130.57, 130.12, 129.13, 129.13, 124.50, 123.45, 117.63, 115.06, 115.06, 112.79, 108.19, 36.76, 35.55, 25.04, 24.96, 24.72, 22.39, 17.21, 17.07. **HRMS (ESI)** calcd for $C_{24}H_{30}O_3$ (MH^+), 367.2268; found 367.2267

M17

IR: 3546.6, 1513.3, 1436.9, 1149.9, 825.6. **^{1}H NMR** (400 MHz, acetone-*d*$_6$) 7.11 (d, J = 8.4 Hz, 2H), 6.80 (d, J = 8.4 Hz, 2H), 6.39 (s, 1H), 5.17-5.10 (m, 2H), 3.37 (d, J = 6.6 Hz, 4H), 2.85 – 2.74 (m, 2H), 2.68 (dd, J = 11.5, 5.1 Hz, 2H), 1.75 (s, 6H), 1.65 (s, 6H). **^{13}C NMR** (100 MHz, acetone-*d*$_6$) δ 155.51, 153.56, 153.56, 140.31, 133.24, 129.13, 129.13 129.02, 129.02, 125.27, 125.27, 117.69, 117.69, 115.18, 115.18, 100.71, 36.19, 32.04, 25.05, 25.05, 24.77, 24.77, 17.33, 17.33. **HRMS (ESI)** calcd for $C_{24}H_{30}O_3$ (MH^+), 367.2268; found 367.2260.

M18

IR: 3349.0, 1584.1, 1459.3, 1157.3, 1105.2, 827.5. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.08 (d, J = 8.4 Hz, 2H), 6.79 (d, *J* = 8.4 Hz, 1H), 6.22 (s, 1H), 2.86-2.79 (m, 2H), 2.73 – 2.67 (m, 2H), 2.65 (t, *J* = 6.8 Hz, 2H), 1.80 (t, *J* = 6.8 Hz, 2H), 1.31 (s, 12H). **^{13}C NMR** (100 MHz, CDCl3) δ 153.85, 153.25, 153.25, 138.75, 134.20, 129.44, 129.44, 115.23, 115.23, 111.39, 111.39 103.55, 73.14, 73.14, 34.31, 33.29, 33.29, 31.18, 26.70, 26.70, 26.70, 26.70, 19.77, 19.77. **HRMS (ESI)** calcd for $C_{24}H_{30}O_3$ (MH^+), 367.2268; found 367.2270

CHAPTER IV
ACTIVATION OF PXR BY LICORICE COMPOUNDS

IV.1 Background

The common perception that a label of 'natural' ensures safety, the unpleasant episodes and side effects associated with pharmaceutical (synthetic) drugs, along with the long history of usage of botanicals as medicines have resulted in an upsurge of herbal medicines in Western healthcare. Passage of the education act further fueled the utility of botanicals in drugs, dietary supplements, cosmetics or personal care products. As a result, herbal ingredients became part of various formularies and the trend shows a strong inclination of consumers to self-medicate with botanicals.[26] Due to this rapid growth of the market, consumer use in many cases has outpaced adequate scientific understanding of the products, opening the door for potential adverse effects.

Although the efficacy of some botanical drugs has been documented, there is a concern regarding the perceived safety of herbal products, particularly with respect to the knowledge of botanical drug interaction potential with conventional prescription drugs and its clinical significance. Moreover, herb-drug and induced-herb-drug interactions have generally been inadequately studied. With the burgeoning use of botanicals, some reports of serious drug interactions are appearing in the literature. [131-135] A further risk for consumers, which will exacerbate the problems associated with botanical ingredients, is that self-medication might delay or prevent a patient from seeking appropriate medical treatment. Botanical drugs comprise a plethora of bioactive constituents and often exist as very complex mixtures. Several reports indicate that herbal constituents could alter the pharmacokinetics of prescription drugs *via* direct

or inductive pathways.[13, 28, 136] This is critically important when it comes to drugs of a narrow therapeutic index such as immunosuppressants or chemotherapeutic agents or with those intended for a chronic or life-threatening diseases. The incidence of such events has been reported and led to therapeutic failure.

For instance, St. John's Wort concomitantly used with cyclosporine (immunosuppressant) or contraceptives led to the failure of therapy. Other reports described the inefficacy of efavirenz (antiviral) consumed with Ginkgo supplements. Direct modulation of metabolizing enzymes is one route recognized for the pharmacokinetics interactions between conventional drugs and other xenobiotics. This includes CYPs inhibition, both as transient and irreversible inhibition. However, these models could not explain the inductive effects acquired by consumption of some herbs. Later on, researchers have identified the pivotal role played by nuclear receptors in regulating the gene-expression of drug metabolizing enzymes and transporters.[137-138] The Pregnane X Receptor (PXR), as one of the newly defined nuclear receptors, is considered as the master regulator of drug metabolizing enzymes. This includes induction of cytochrome P450s and multiple other phase-I and phase-II enzymes and transporters.

Several botanicals and their constituents have been recognized as inducers of PXR such a *St. John's Wort, Kava Kava and Gugulipids, Gan Cao*.[13] Further investigations revealed a wide variety of molecules responsible for initiating PXR machinery, which indicates the broad substrate acceptability. In conformity with the above, herbal constituents can interact with PXR ligand binding domain and have the ability to modulate principal metabolic enzyme and some of the energy expenditure checkpoints.[139] This situation has raised a huge campaign inclined to unveil the potential risk/benefit of intentional long-term herbal consumption.

IV.1.1 The role of Pregnane X receptor

One major discovery in the nuclear receptors field has occurred in 1998. The discovery of PXR has outlined a new era into understanding the regulatory mechanisms behind xenobiotic detoxification as a primary defense mechanism in our bodies. Initially, it was classified as one of the orphan receptors "a receptor without a characterized endogenous ligand", but soon it was found to bind a family of diverse chemical subtypes.[140] This includes a set of endobiotics such as the pregnane steroids as well as other therapeutic drugs. This nuclear receptor (NR1I2) belongs to the nuclear receptor subfamily 1 group I, which also includes the vitamin D receptor (VDR) and the constitutive androstane receptor (CAR). Similar to the other nuclear receptors, it functions as a ligand-induced transcriptional factor, where the binding of a ligand can trigger a conformational change of the protein that will switch on a cascade of events leading to a controlled and specific genetic induction.[137]

It is well established that PXR has a paramount role as a xenosensor, where it regulates the expression of many metabolic enzymes and transporters responsible for drug or "xenobiotics" disposition. In addition, it has viable but less studied roles in endobiotics synthesis, metabolism and homeostasis including bile acids, lipids, glucose, bilirubin, vitamins and other steroid hormones. This highlights the diverse and crucial roles of PXR implicated into both drug metabolism (efficacy, toxicity, drug interactions, and drug resistance) and disease (metabolic syndrome, cancer, and inflammation).[141-142] In fact, PXR ligand-activation is paradoxically important, in a way to sense these xenobiotics and to dispose them. Alternatively, they may influence the systemic or tissue exposure of the other drugs in case of concurrent consumption with the PXR activators. In fact, PXR versatile role is reflected by its vast control upon many metabolic enzymes and transporters. So far, it is shown to control the target genes of phase-I

CYP450 (CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP3A1, CYP3A4, CYP3A5, CYP3A7, CYP4F12, CYP24, and CYP27A1), phase II uridine diphosphate (UDP)-glucuronosyltransferases (UGT1A1, UGT1A3, UGT1A4, UGT1A6, and UGT1A9), sulfotransferases (Sult2a1), glutathione S-transferases (GSTA2, GSTA4). In addition to the carboxylesterases and phase III P-glycoprotein (MDR1/ABCB1), multidrug resistance-associated protein 1 (ABCC1), multidrug resistance-associated protein 2 (ABCC2), multidrug resistance-associated protein 3 (ABCC3), and organic anion transporting polypeptide 2 (OATP2).[13]

However, induction of PXR is not the ultimate event. Also, it is not necessarily translated into the genetic pool of susceptible enzymes and transporters. Instead, responses are selective and gradient in nature, which rather come as a result of multiple other dimensions including specific tissue, coregulators, epigenetics, and post-translational modifications, specific response elements, and other NRs.[143] In addition to its roles as a xenobiotic sensor, many studies have unveiled the kinetics behind its uncanonical roles. Hence providing the evidence of its function as a mediator of chemo-genetics interplay, which delineates the etiology of the environmentally induced diseases such as the cardiometabolic disease.

This can be explained by upregulation or inhibition of vital checkpoints proteins (enzymes or factors) in specific diseases networks. One example is the hyperglycemic effects corroborated with PXR induction, where PXR induction was found to downregulate the glucose transporter 2 and other vital kinases related to gluconeogenesis (glucokinase, dehydrogenase kinase isozyme 2, and phosphoenolpyruvate carboxykinase 1).[142, 144] Moreover, PXR polymorphism, and PXR variants have been correlated to multiple diseases and malignancy such as Crone's disease, colon cancer and adenocarcinoma.[138]

4.A.2 The Pregnane X receptor structure

PXR shares the common features of other NRs required for the transcriptional machinery. It consists of five main domains among which the DNA-binding domain (DBD) and the ligand-binding domain (LBD) are the most important. The former recognizes specific response elements that flank the genes controlled by PXR. The ligand-binding domain mainly comprises the binding site for ligands in addition to the protein-protein surface interaction. Nevertheless, PXR is unique among other NRs by having a large spherical and flexible binding pocket.

Gene induction is triggered by the ligand-binding event, which will induce a conformational change that will release corepressors already bound to PXR in the AF-2 region. After ligand activation, PXR heterodimerization with retinoid X receptor (RXRα) will take place, which will further stabilize the coactivator binding in the AF-2 region and selectively recognizes DNA response elements of specific genes.[145] The adopted structural features upon ligand binding dictate the following genetic responses by stabilizing and recruiting different coregulators and factors and vice versa.

PXR has a large and flexible hydrophobic pocket wrapped between 12 α-helixes and a short region of 5 β-strands. Compared to other NRs, PXR has a wider surface of β-sheets since it has an extra two β-strands. The ligand binding site volume can vary between apo and ligand-bound crystal structure from 1280-1600 $Å^3$, which is substantially larger than any other NRs. The region comprised of residues 198 and 212 can be completely disordered to fit very large structures such as Rifampicin. This further explains its promiscuous behavior and the lack of direct antagonists through its ligand-binding site. Thus, it can accommodate a wide variety of structures with varying molecular weights (300-800 Daltons). The diversity of the amino acids lining this pocket explain the hospitality of such receptor (ten hydrophobic, four polar and four

charged amino acids).[146-149] Compliance with diversity is essentially one feature that is required for its function as a xenobiotic sensor, which allows the detection of a versatile set of chemicals that may differ in size or electronic features.[150]

However, aside from this sophisticated behavior, several computational studies have deduced global features as the primary requirements needed for PXR activation. For instance, larger hydrophobic molecules have a better chance to interact with PXR; this includes a molecular weight, which is larger than 300 Daltons and hydrophobic features that allow multiple hydrophobic interactions and π-π interactions with a list of recognized hydrophobic amino acids mainly, those occupying the aromatic sub-pocket surrounded by F288 and W299. Further studies recognized S247, Q289, and H407 as indispensable residues for the PXR induction.[149, 151-153] Thus, at least one hydrogen bond interaction with one of the key amino acids is considered as a basic requirement for PXR activation.

The existence of multiple crystal structures of PXR with different ligands has allowed the recognition of the characteristic flexible side chains at the binding site. The AF-2 region is stabilized via ligand binding, especially by the direct interaction with S247. This region recognizes a family of peptides with the LXXLL motif, which is further stabilized by a charge clamp interaction with conserved charged amino acids K259 and E427 (**Figure 22**).[149] Although many inducers of PXR have been identified, antagonists were more resilient.[142] However, more recently, the indirect allosteric modulation was recognized as one pathway for PXR antagonism.[154]

The coregulator site at the AF-2 region constitutes the plausible site for this modulation. This key intervention will prevent the coactivators from binding. Ketoconazole among other azoles is the representative ligand for PXR antagonism.[155] Other naturally occurring and plant-

derived compounds have been discovered, viz. coumestrol, sesamine and resveratrol.[156-157]

IV.1.2 An overview of the applied computational techniques to predict PXR activation

It is increasingly important to screen for drug interaction early in the process of drug development to eliminate hits with a high propensity for inducing adverse effects through the drug metabolizing enzymes. In fact, computational methods provide a significant research axis for pre-screening of hit libraries. In the realm of drug metabolism and excretion, PXR has the major influence upon expression of the vital CYPs enzymes and transporters.[158-159] However, indirect PXR screening is very challenging. It has species-specific selectivity, where animal models cannot exclusively tell us about the possible interaction that could happen. For instance, rifampicin, which is a potent PXR activator in human, does not induce PXR in rodents.

Moreover, PXR *in vitro* models suffer from high variability among different cell lines and can lead to contradicting results. The reasons behind this could be related to the cell viability or the presence or absence of different transporter machinery that will affect the permeabilities and concentrations of the tested compound. Moreover, the origin of the cell-line plays a fundamental role in the expressed corepressors/coactivator ratios. For instance, Hela or HepG2 (cancer cell line) have higher corepressor concentrations compared to normal hepatocytes. Hence, extrapolation should be highly dependent on the targeted population. Application of *in silico* models of both structure-based and ligand-based methods not only would provide alternative routes but also would fine-tune the screening of the existing methods.[160]This will salvage some of the costs and would keep the more expensive screenings methods such as hepatocyte cell-lines or the transgenic mice models (expressing human PXR) for those highly alerting compounds.

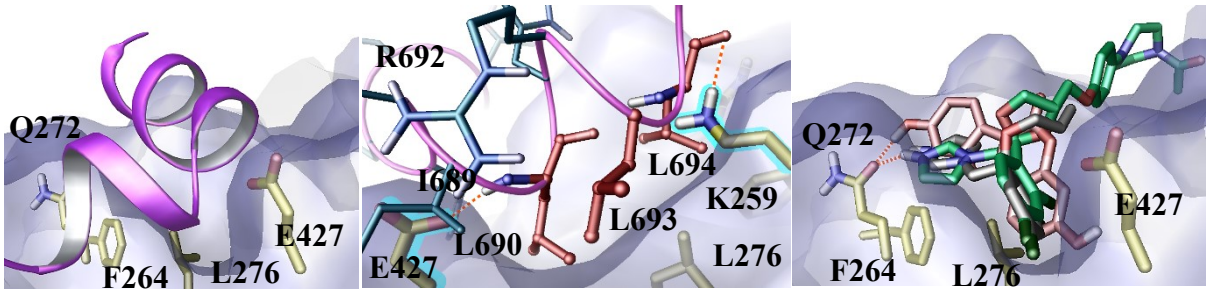


Figure 22. (Left) Interaction of SRC-1 (the coactivator peptide in a purple cartoon) with the hydrophobic groove in the 1NRL crystal structure. **(Middle)** the LXXLL motif in SRC-1 is buried in the groove with the charge clamp lock with K259 and E427. **(Right)** Antagonist ketoconazole (green sticks) and fluconazole (grey sticks) and coumestrol (pink sticks) docked to the SRC-1 site

However, the challenge is not inevitable when applying *in silico* screening. The promiscuity and flexibility of PXR represent the major obstacle. Nevertheless, previous studies have encouraged the application of these approaches coupled to *in vitro* methods. Most of the studies have focused on the ligand-based approaches. For example, one study has applied machine-learning methods to predict PXR active ligands and the best predictive ability has reached 80% accuracy by applying the support vector machine (SVM) method.[161] Other QSAR models have employed the pharmacophore modeling, which comprised of hydrophobic and hydrogen bonding features similar to the features found in PXR crystal structures.[162] Predictions based on the LBP and docking have performed well, where one study has used Gold scores to classify the compounds from the ToxCast database. They were able to correctly identify 8 out of 12 agonists and 7 out of 16 PXR non-agonist.[160]

Taken together, the application of pre-clinical screening of PXR will significantly reduce the costs and can enhance our knowledge about the behavior of such compounds in the presence of a very challenging defense mechanism. There is no straightforward method that can give the perfect answer. Nevertheless, combining *in silico* and *in vitro* screens together will enhance the

performance of these models. In this study, we have utilized the structure-based approach screening coupled to *in vitro* analysis to screen a library of *Glycyrrhiza* compounds. We envisaged that docking (compared to the other mentioned *in silico* methods) would give us direct insights for ligand interaction with PXR; a method, which will cope better with the highly variable chemical space of PXR activators. Particularly, the number and the class of compounds preconceived in the training set for ligand-based methods are usually insufficient to avoid the bias. In addition, ligand-based and machine learning approaches are more susceptible to the inhomogeneity that is typically encountered within PXR *in vitro* data.

IV.1.3 *Glycyrrhiza* and Herb-drug interaction: An *in silico* approach

Induction of PXR is identified as one of the major mechanisms for triggering HDI. Multiple clinically significant pharmacokinetic interactions between herbs and prescription drugs have been reported via PXR. Direct CYP/transporter inhibition is another vital mechanism, although not mutually exclusive, but rather a competitive route under in vivo conditions.[13, 28, 136] Licorice herb has been used for centuries in traditional medicine in different parts of the world, where it has been found in 85% of the traditional Chinese medicinal prescriptions. Moreover, it is increasingly used in the US and Europe as a standardized total extract to alleviate multiple ailments, including postmenopausal symptoms. It is also consumed for its chemoprevention and hepatoprotective properties.

Recently, there is more scientific and preclinical experimental evidence, which confirmed the efficacy of these ethnobotanical remedies.[17, 63, 163] However, these commercialized freely accessible herbal supplements have infiltrated the health systems without a rigorous understanding of its possible side effects. This situation has triggered an urgent call for more rigorous studies of HDI.[26] In addition, natural products are indispensable resources for NCEs,

where the optimization of their pharmacokinetics and their drug metabolism properties are at the top priority for their success.[9] Recent studies have raised the concerns for licorice induced PXR activity, and they pointed out to its glycyrrhizin component as the source of induction.[164-165] However, only weak PXR activation has been detected with the *in vitro* reporter gene assays for glycyrrhizin as compared to rifampicin (the positive control).

Moreover, *Glycyrrhiza* has been prioritized by the FDA as one of the high-risk herbal constituents for inducing the adverse effects.[166] The objective of this study is to screen the components of different *Glycyrrhizae* species against potential PXR induction or CYP inhibition by application of multiple computational methods. The components with high interaction potentials will be further confirmed with cell-based *in vitro* assays for the assessment of PXR induction and the elucidation of the susceptible target genes. The activation of PXR will be determined in HepG2 cells that are transiently transfected with the expression vector of PXR and the reporter plasmid.

By the application of such methods, we anticipate bypassing some of the inconveniences coupled to the direct *in vitro* analysis, including interference of endpoint measurements and the inadequate supply of the purified diverse component systems. To fill the gaps into our knowledge of the possible HDI that could be induced by the secondary metabolites of licorice, we have utilized a tandem scheme of both *in silico* and *in vitro* risk assessment models. To roll out the privileged ingredients with a high propensity for triggering HDI, two clinically relevant *Glycyrrhiza* species (*G. glabra* and *G. uralensis*) have been docked to the LBD of PXR crystal structure to prioritize their potential interaction. Furthermore, *in silico* risk assessment of CYP enzyme inhibition models commercially available in ADMET predictor have been applied to the unglycosylated structures of several secondary metabolites reported in both species.

IV.2 Results

IV.2.1 Docking

The top docking scores for licorice compounds are summarized in **Tables SI 8** and **SI 9**. Redocking of the cognate ligand in both crystal structures has reproduced the binding mode in the native protein structure with relatively good precision (RMSD for SRL and hyperforin were found to be 1.0751Å and 1.1987Å, respectively). The docking scores for a group of experimentally active compounds were used as a control group to estimate the cut-off value **(Figure 23 and Table 5)**. Thus, both SP and XP glide scores were evaluated for a set of active known compounds (highly active, moderate and weak) into two different crystal structures, 1NRL and 1M13. The XP docking scores against 1NRL were capable of reflecting the experimental gradient activity of the control group and thus they were used for our screening purposes.

The calculated scores of the known compounds in the control group showed that moderately active ones scored in the range of (−9.0 to −9.5 kcal/mol), while high activators scored below (−10 kcal/mol). However, earlier studies showed that potent PXR induction correlates successfully with a transient transfection assay when the compound (<10 μM) achieves (70%) maximum induction relative to (10 μM) rifampicin.[167] Thus, in order to reflect these observations onto our docking results, we decided to consider the compounds with glide scores < −9.5 kcal/mol. In this present study, the docking scores against the LBD suggested a high number of licorice components having the potential of inducing PXR activation. According to our predefined threshold of PXR agonist, around 60 phenolic compounds in *G. glabra* and 50 compounds in *G. uralensis* have scored below −9.5 kcal/mole. Interestingly, most of these compounds belong to the isoflavonoid classes with a higher order of prenyl groups. Moreover,

all the prenylated dihydro stilbenoids reported in both licorice species scored very high, due to favorable ligand-protein interactions.

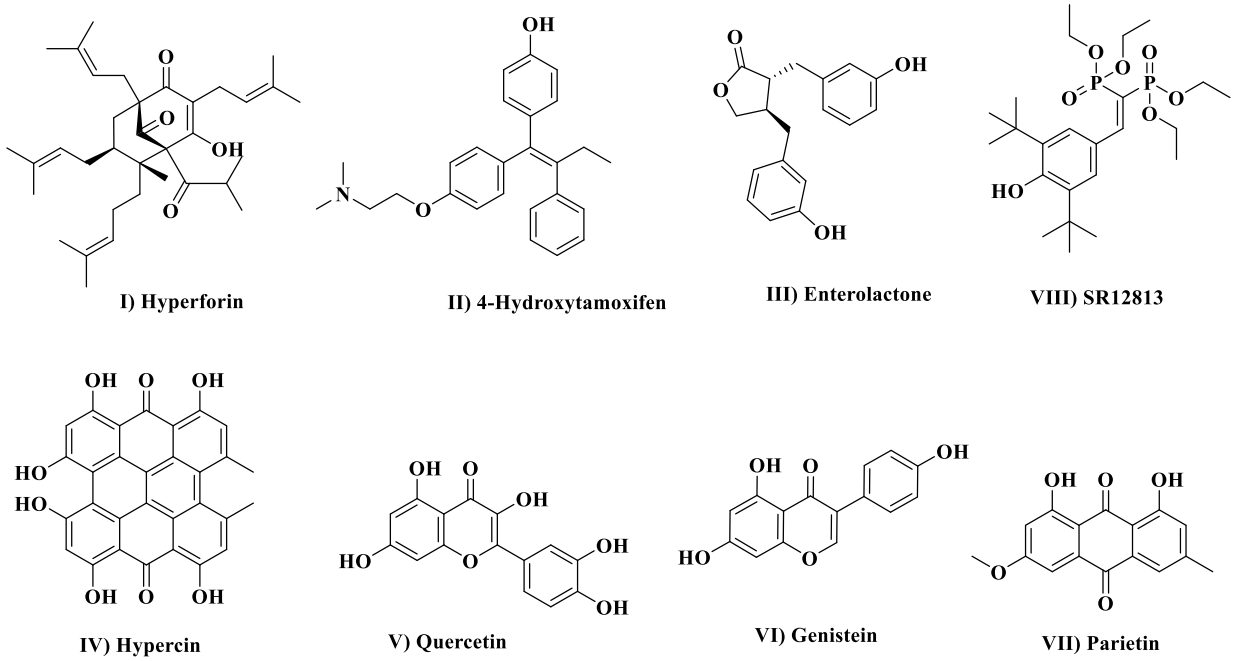


Figure 23. Naturally occurring PXR inducers and 1NRL native ligand SR12813

The visual inspection of the isoflavonoids docking poses revealed the vital hydrophobic and π-π interactions of both prenyl groups or their tetrahydropyran (THP) counterparts (the cyclized prenyl appendages found in many top-ranked compounds) with the hydrophobic sub-pocket comprised of F288 and W299 (**Figure 24)**. Representative examples of each category are shown in

Figure 25 (flavonoids and isoflavonoids), **Figure 26** (DHS), and **Figure 27** (other miscellaneous components). In addition, several compounds were also able to interact with at least one key residue in the LBD *via* hydrogen bonding. On the other hand, the docking poses of the prenylated stilbenoids did not show a certain pattern of interaction since they have a more freely rotatable scaffold and can adopt a variety of conformations. Other prenylated

miscellaneous compounds classes (coumarins, chalcones, and pterocarpans) have shown the similar potential of interacting with PXR. Examples of the docking pose of some representative compounds interacting with PXR LBD is shown in **Figure 28**.

Table 5. SP and XP glide scores (kcal/mol) for a set of known PXR active compounds (highly active, moderate and weak)

No.	Name	SP-Score 1NRL	XP-Score 1NRL	SP-Score 1M13	XP-Score 1M13	Activity
I	Hyperforin	−8.44	−8.15	−9.60	−11.71	[a]High activator
II	4-OH-Tamoxifen	−9.82	−10.54	−9.17	−10.15	[a]High activator
III	Enterolactone	−7.99	−9.34	−7.22	−8.71	[b]Moderate activator
IV	Hypericin	NA	NA	−7.55	−10.02	[b]Moderate activator
V	Quercetin	−6.93	−8.70	−6.76	−9.50	[c] Weak activator
VI	Genistein	−6.66	−8.10	−7.72	−6.58	[c] Weak activator
VII	Physcion	−6.94	−7.71	−6.54	−7.22	[c] Weak activator
VIII	SR12813	−10.82	−12.34	−8.14	−9.68	[a,d]High activator

a) >70% maximum induction relative to 10 µM Rifampicin according to transient transfection assay.
b) 30-70% maximum induction relative to 10 µM Rifampicin according to transient transfection assay.
c) <30% maximum induction relative to 10 µM Rifampicin according to transient transfection assay.
d) One of the most potent activators with Kd value of 40 nm.

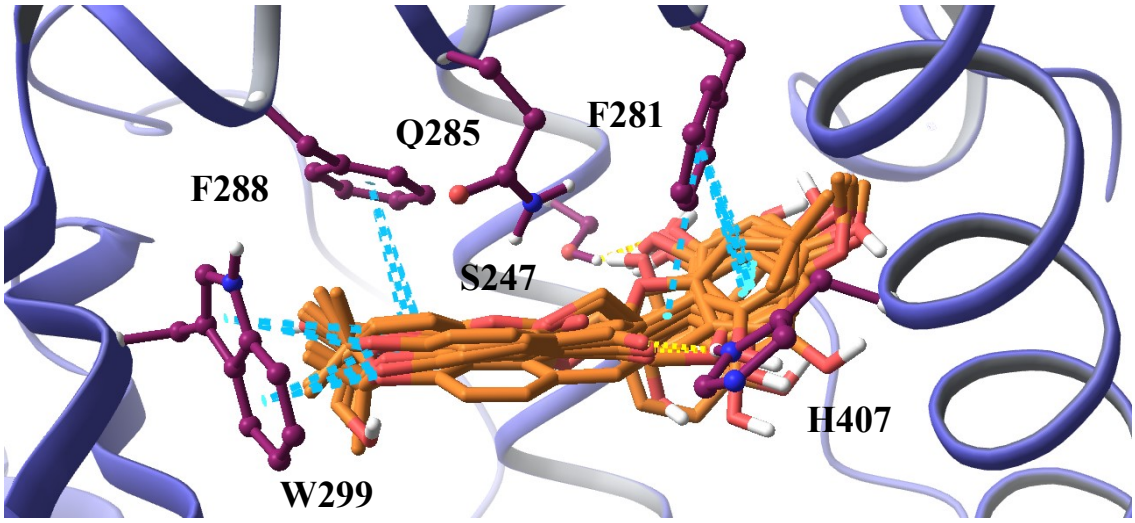


Figure 24. Docking pose (PDB:1NRL) of a number of representative isoflavonoids and flavonoids compounds (orange sticks) in the top scoring list of *G. glabra* and *G. uralensis* showing similar preferable π-π interactions (blue dotted lines) with W299 and F288 (magenta balls and sticks) with their dihydro THP or phenolic groups.

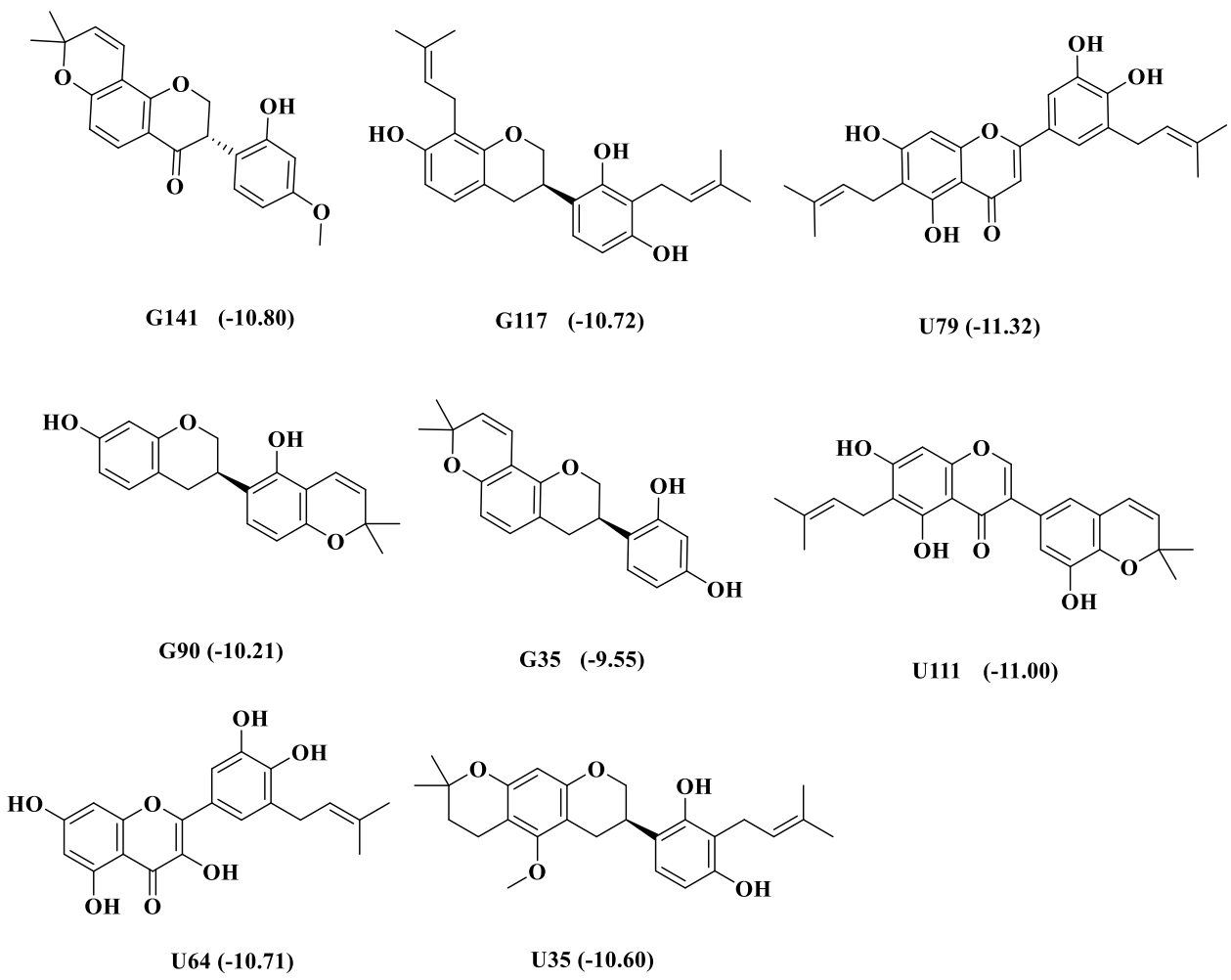


Figure 25. Group 1 (flavonoids and isoflavonoids) of top-scoring compounds from both *G. glabra* and *G. uralensis* species against PXR LBD (PDB:1NRL)

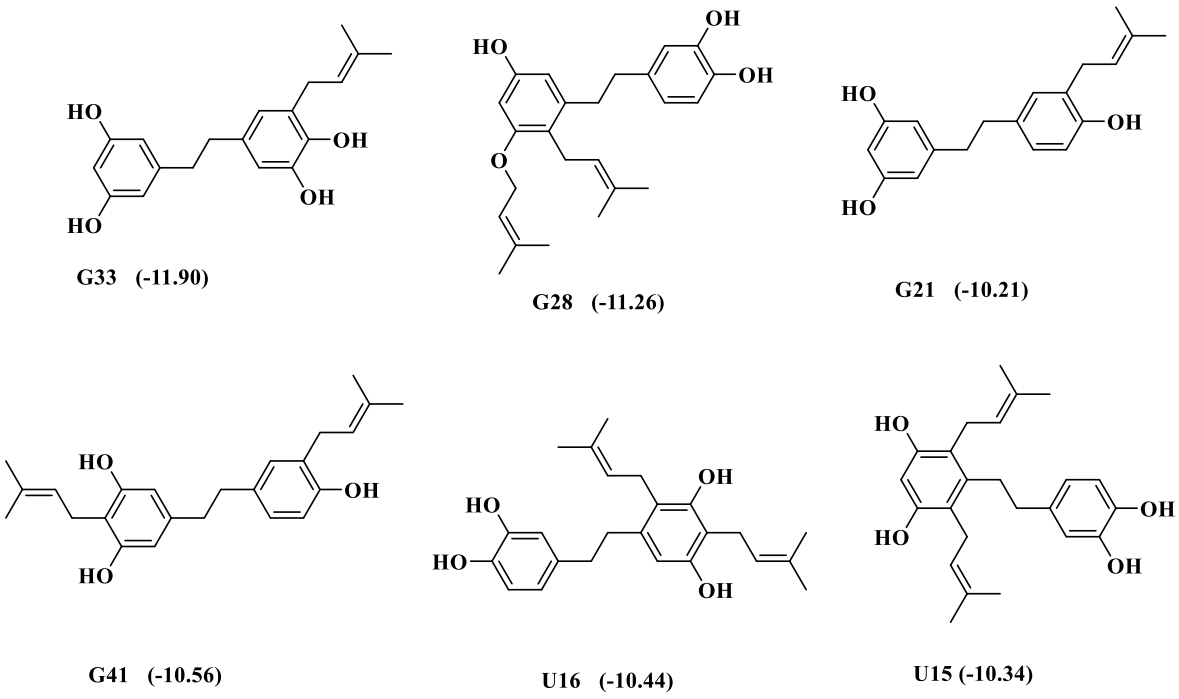


Figure 26. Group 2 (dihydrostilbenoids) of top-scoring compounds from both *G. glabra* and *G. uralensis* species against PXR LBD (PDB: 1NRL)

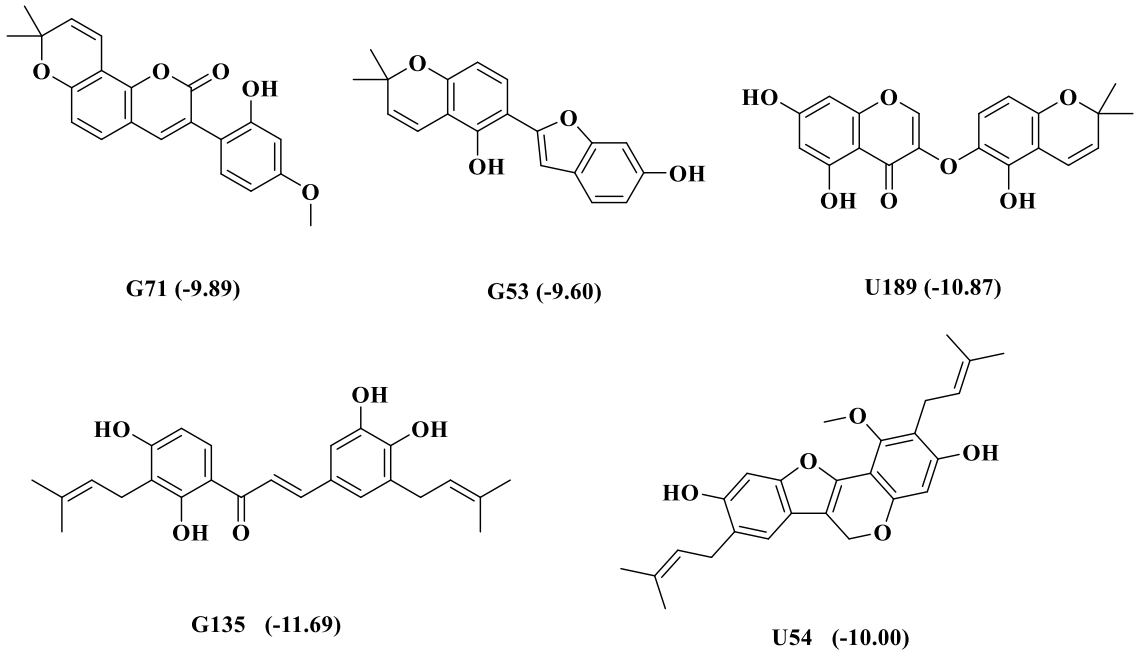


Figure 27. Group 3 (miscellaneous) of top-scoring compounds from both *G. glabra* and *G. uralensis* species against PXR LBD (PDB: 1NRL)

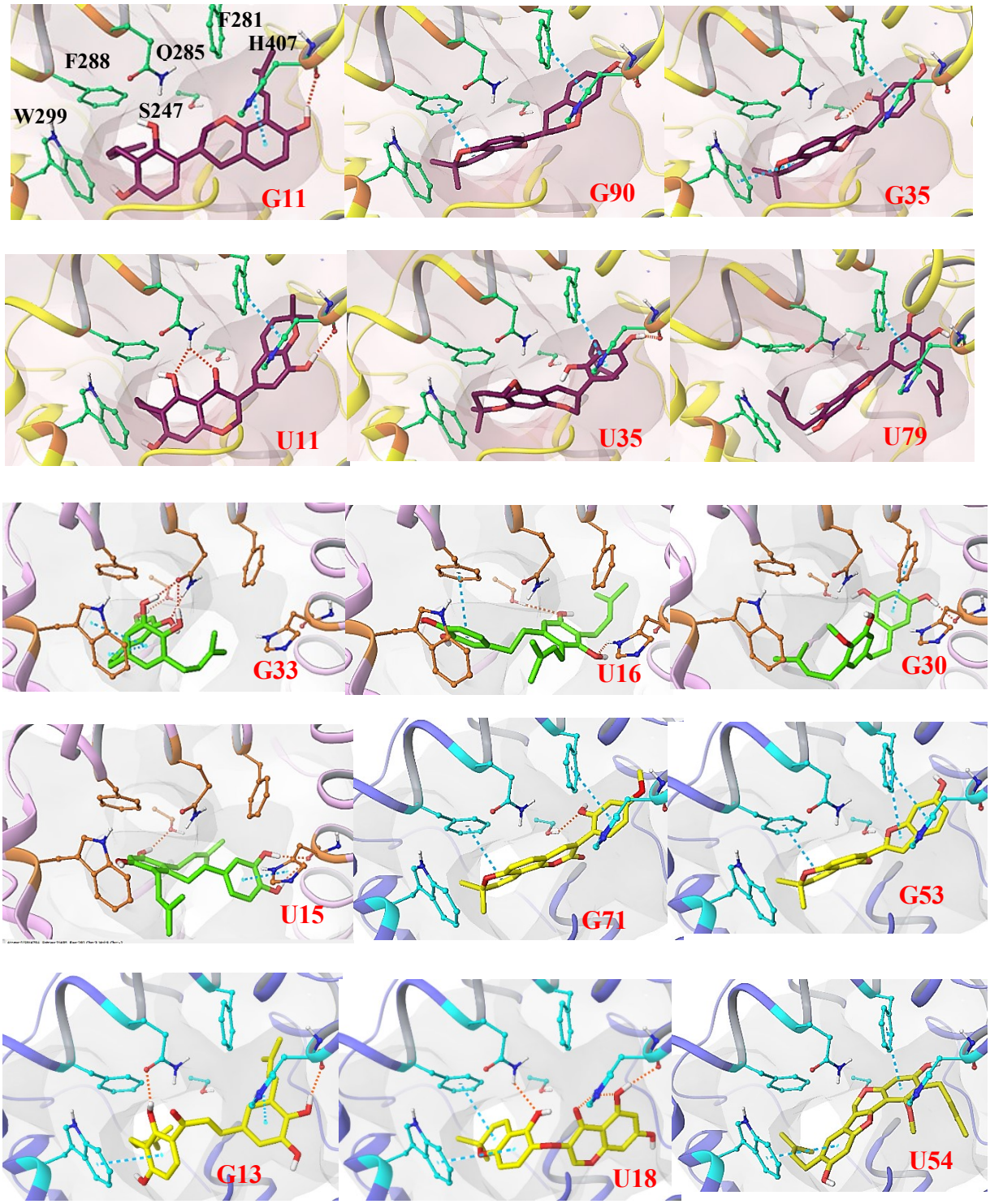


Figure 28. Docking pose of representative top scoring compounds in PDB: 1NRL binding site (group 1 in magenta sticks, group 2 in green sticks and group 3 in yellow sticks)

IV.2.2 *In vitro* testing

Based on *in silico* assessment of licorice secondary metabolites, prenylated stilbenoids have been identified as one of the top scoring components with a high propensity for PXR activation. A library of 18 stilbenoids rich with diverse prenylation patterns (0.78-25 μM), where incubated with HepG2 cells transfected with the pSG5-hPXR (25 μM) and the PCR5 (25 μM) plasmid DNA with and without the positive control rifampicin (3.125-25 μM)

After incubation for 24 hours with the pretreated cells, resveratrol (parent compound) and several compounds of both the stilbenoids and the DHS series (**M7-10** and **M13-18**) were able to induce PXR with multiple fold induction in a dose-dependent manner (**Figure 29, Figure 30** and **Table 6**). Interestingly, compounds **M9**-**M10**, **M15**-**M18** showed strong activity, where they were evenly or slightly more efficacious than the positive control, rifampicin. The maximum fold induction surpassed the positive control (3.8) for compound **M10** (4.37 folds), **M15-18** (4.22-5.25 folds) at the same concentration (12.5 μM). Moreover, potent PXR induction was observed at the sub-micromolar range for compounds **M9**, **10**, **15**, **17** and **18**, which maintained at least two-fold induction at low concentration as 0.78 μM.

Furthermore, the significant increase of CYP3A4 mRNA expression of the compounds **M1**, **7**, **10**, **12**, **15**, **17**, and **18** validates the PXR activation assay (**Table 7** and **Figure 31**). Compounds **M7**, **M10**, and **M15** were able to increase the mRNA levels up to 6 folds with **M10** having the most potent activity. It maintained the 6 fold induction at 1μM concentration. Additionally, to exclude any antagonist allosteric modulation of PXR, known for resveratrol, select compounds, viz **M7**, **10**, **12**, **15**, **17**, **18** including resveratrol were subjected to competitive experiments with rifampicin. None of these compounds were able to antagonize the rifampicin activated PXR in the range of 10.0-0.6 μM.

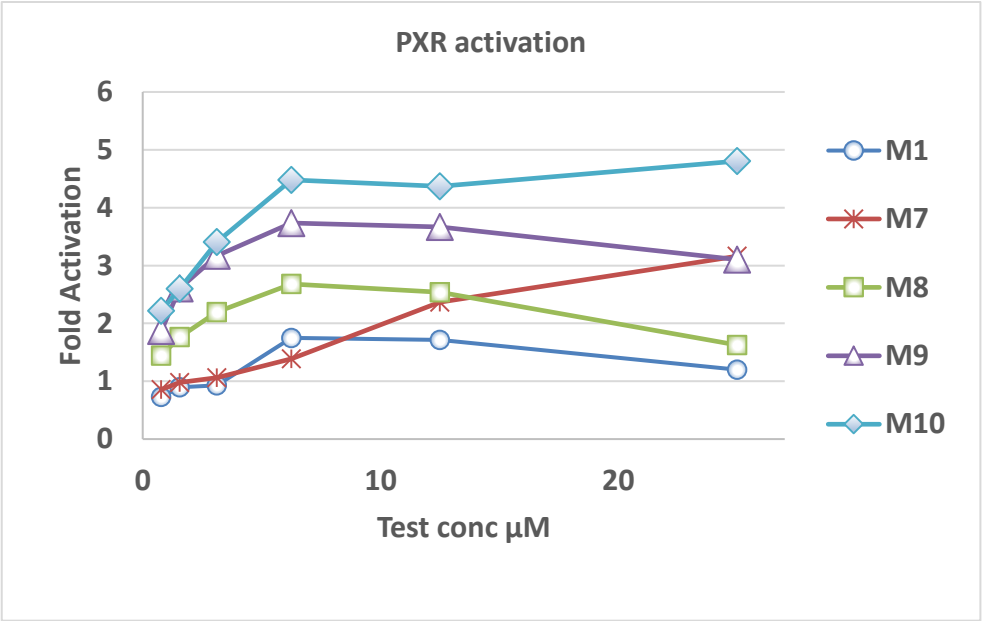


Figure 29. PXR fold induction in transfected HepG2 cells for stilbenoids derivatives **M1**, **7**, **8**, **9** and **10**. The compounds were tested at concentrations of 25, 12.5, 6.25, 3.13, 1.56 and 0.78 µM.

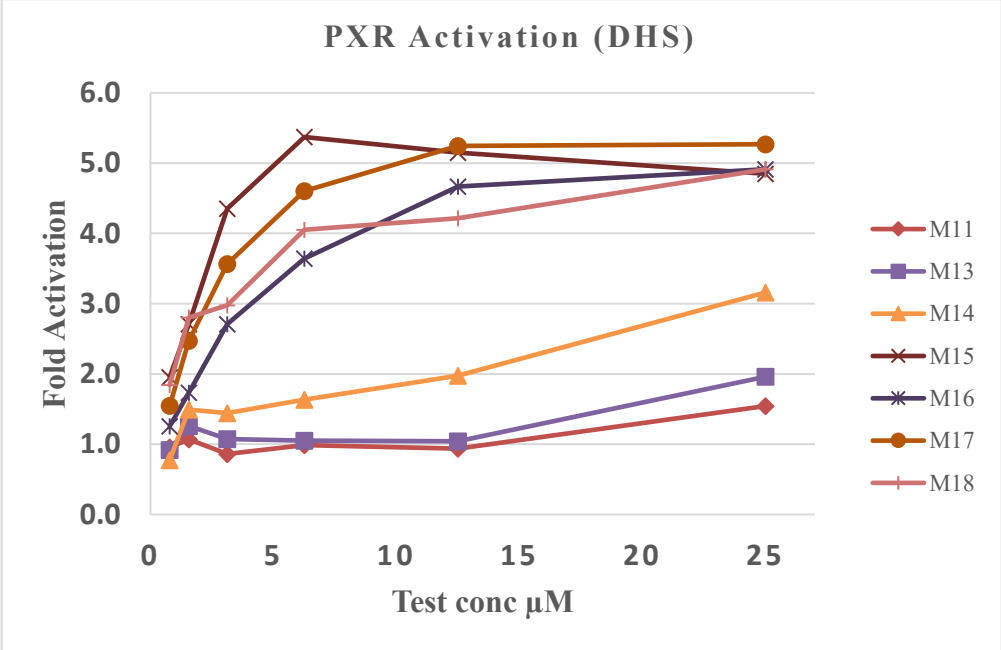


Figure 30. PXR fold induction in transfected HepG2 cells for DHS derivatives **M11**, and **M13-M18**.

Table 6. PXR activation in HepG2 cells treated with prenylated stilbenoids and DHS for 24 hr. The data is included as the means ± standard deviation of three independent experiments

Compounds (μM)	Fold induction					
	25	**12.5**	**6.25**	**3.13**	**1.56**	**0.78**
M1	1.20 ± 0.02	1.71 ± 0.14	1.75 ± 0.01	0.92 ± 0.05	0.90 ± 0.12	0.73 ± 0.01
M7	3.16 ± 0.08	2.37 ± 0.01	1.39 ± 0.14	1.06 ± 0.08	0.98 ± 0.08	0.86 ± 0.01
M8	1.62 ± 0.03	2.54 ± 0.07	2.68 ± 0.01	2.20 ± 0.03	1.76 ± 0.01	1.44 ± 0.04
M9	3.10 ± 0.00	3.67 ± 0.16	3.74 ± 0.27	3.17 ± 0.22	2.60 ± 0.17	1.87 ± 0.13
M10	4.80 ± 0.01	4.37 ± 0.44	4.48 ± 0.08	3.40 ± 0.05	2.60 ± 0.59	2.22 ± 0.21
M11	1.54 ± 0.07	0.94 ± 0.07	0.99 ± 0.02	0.86 ± 0.07	1.07 ± 0.02	0.96 ± 0.11
M13	1.96 ± 0.21	1.04 ± 0.15	1.05 ± .07	1.07 ± 0.10	1.26 ± 0.02	0.92 ± 0.06
M14	3.16 ± 0.14	1.98 ± 0.16	1.64 ± 0.22	1.44 ± 0.15	1.49 ± 0.07	0.77 ± 0.09
M15	4.85 ± 0.39	5.15 ± 0.78	5.37 ± 0.05	4.35 ± 0.26	2.71 ± 0.02	1.95 ± 0.13
M16	4.91 ± 0.15	4.67 ± 0.51	3.64 ± 0.11	2.71 ± 0.01	1.73 ± 0.22	1.26 ± 0.11
M17	5.27 ± 0.28	5.24 ± 0.25	4.60 ± 0.04	3.56 ± 0.05	2.47 ± 0.20	1.55 ± 0.20
M18	4.91 ± 0.32	4.22 ± 0.15	4.05 ± 0.12	2.98 ± 0.12	2.80 ± 0.18	1.84 ± 0.16

Table 7. Increase in the mRNA expression of CYP3A4 in HepG2 cells by the synthesized prenylated stilbenoid and DHS derivatives at three different concentrations.

Compounds (μM)	Fold induction		
	10	3	1
M1	1.66 ± 0.03	1.76 ± 0.20	2.79 ± 0.20
M7	6.37 ± .053	4.96 ± 0.31	4.64 ± 0.21
M10	3.96 ± 0.34	5.65 ± 0.29	6.62 ± 0.13
M12	2.38 ± 0.33	3.46 ± 0.23	2.54 ± 0.22
M15	6.21 ± 0.79	4.39 ± 0.37	2.34 ± 0.40
M17	2.15 ± 0.05	1.87 ± 0.10	1.37 ± 016
M18	1.9 ± 0.10	1.56 ± 0.17	NA

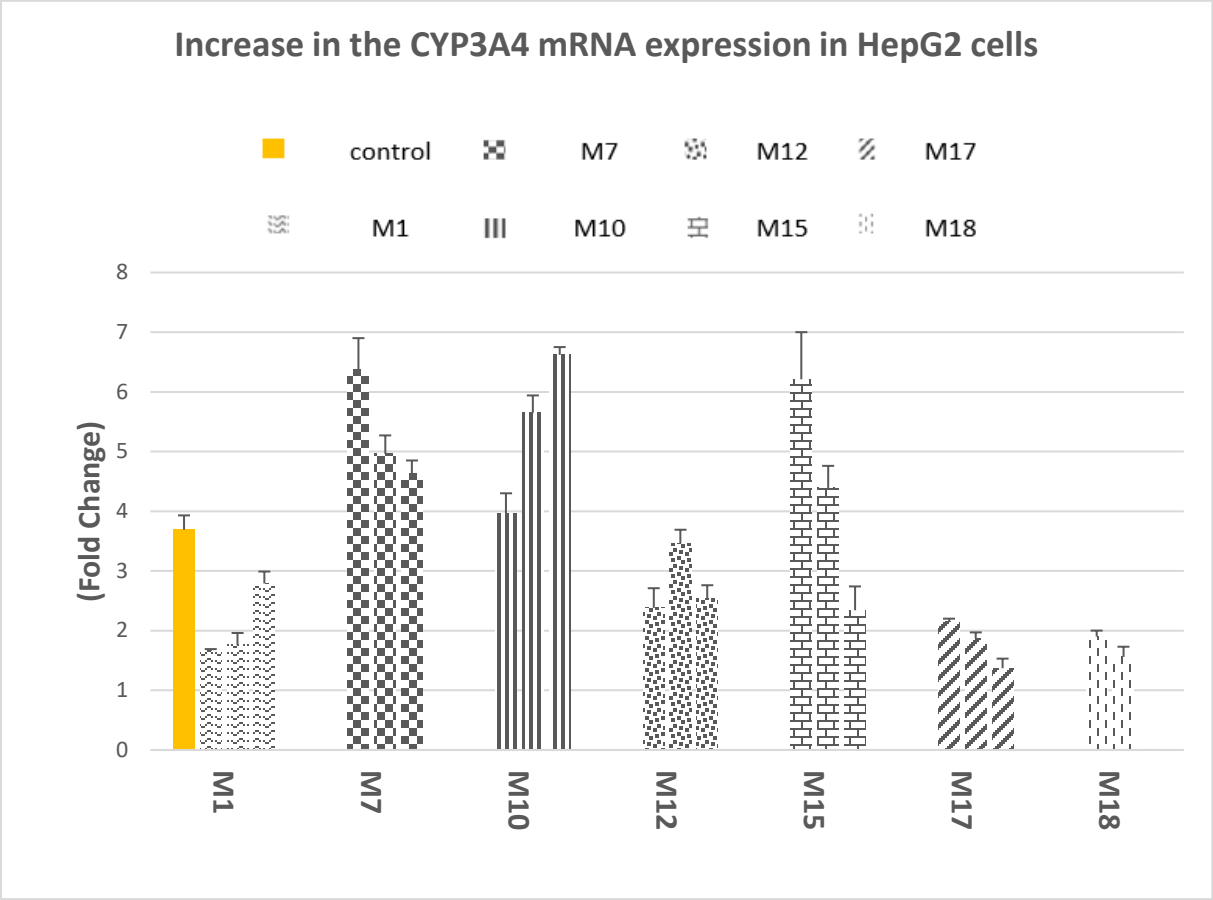


Figure 31. Increase in the mRNA expression of CYP3A4 in HepG2 cells by the synthesized prenylated stilbenoid and DHS derivatives. The compounds were tested at concentrations of 10 (left), 3 (middle), and 1 (right) μM. Rifampicin (10μM, yellow) was used as a positive control. The data is represented as mean ± standard deviation calculated from three independent experiment.

IV.2.3 *In silico* risk assessment of CYP enzyme inhibition models commercially available in ADMET predictor

Phytochemicals from both *G. glabra* and *G. uralensis* have been evaluated through QSAR models utilizing commercially available software product, ADMET Predictor™ 9.0 (Simulations Plus Inc.) for potential inhibitors of CY1P450s. These models include general inhibition models for five recombinant CYPs (1A2, 2C9, 2C19, 2D6, and 3A4). As a result, we have classified and compared the putative components in licorice that might influence CYP inhibition at the level of the enzymes. The analysis unveiled the risk potential upon licorice

consumption amongst two major CYPs, 1A2 and 3A4, where many compounds in both species (**Table 8**) were shown to qualify for CYP inhibition with high confidence level (> 79%).

Table 8. Compounds predicted to have CYP enzyme inhibition with a high confidence level in five different CYP inhibition models (CYP3A4, CYP1A2, CYP2C9, CYP2D6, and CYP2C19) available in ADMET predictor

	G. glabra	***G. uralensis***
CYP3A4_Inh	**1 (80%), 28 (80%), 29 (80%), 67 (80%), 83 (80%), 108 (80%), 140 (80%), 146 (80%), 158 (80%), 170 (80%)**	**15 (80%), 16 (80%), 17 (80%), 20 (80%), 32 (80%), 51 (80%), 53 (80%), 55 (80%), 56 (80%), 61 (80%), 66 (80%), 77 (80%), 79 (80%), 85 (80%), 111 (80%), 121 (80%), 122 (80%), 171 (80%), 186 (80%)**
CYP1A2_Inh	**1 (84%), 8 (95%), 9 (95%), 17 (95%), 18 (95%), 20 (84%), 24 (95%), 31 (79%), 34 (95%), 36 (95%), 43 (84%), 45 (95%), 48 (95%), 53 (95%), 54 (95%), 55 (95%) , 56 (95%), 62 (95%), 71 (95%), 72 (95%), 74 (95%), 76 (95%), 77 (95%), 88 (95%), 92 (84%), 95 (95%), 101 (84%), 104 (95%), 106 (84%), 120 (95%), 122 (95%), 130 (95%), 136 (79%), 139 (95%), 140 (95%), 147 (84%), 148 (84%), 149 (95%), 160 (95%), 166 (95%)**	**2 (84%), 3 (95%), 5 (79%), 6 (84%), 38 (95%), 39 (79%), 42 (95%), 60 (95%), 61 (79%), 64 (84%), 66 (79%), 70 (95%), 85 (79%), 91 (84%), 108 (95%), 111 (79%), 114 (95%), 118 (95%), 124 (95%), 148 (95%), 149 (95%), 157 (84%), 158 (95%), 159 (95%), 161 (95%), 162 (95%), 179 (79%)**
CYP2C9_Inh	**19 (77%), 22 (77%), 37 (77%), 52 (77%), 54 (77%), 112 (77%), 129 (77%)**	**5 (77%), 14 (77%), 19 (77%), 20 (77%), 22 (77%), 24 (77%), 47 (77%), 54 (77%), 57 (77%), 70 (77%), 83 (77%), 88 (77%), 113 (77%), 119 (77%), 133 (77%), 149 (77%), 152 (77%), 161 (77%), 164 (77%), 180 (77%), 181 (77%)**
CYP2D6_Inh	**94 (70%)**	**50 (70%), 54 (70%)**
CYP2C19_Inh	**21 (78%)**	**94 (77%)**

IV.3 Discussion

Earlier reports showed weak PXR activation related to its glycyrrhizin component. Many manufacturers provide deglycyrrhizinated licorice (DGL licorice) supplements, which are labeled to be safe and effective since it avoids the hypertensive crises that could be triggered by glycyrrhizin. However, this does not mean that these DGL supplements are void of HDI. In this manner, underestimation of a plant material, which can potentially produce hundreds of compounds, might be responsible for potentiating HDI. PXR plays a pivotal role in the ADME properties of the co-ingested conventional drugs that might change its metabolism, distribution,

and elimination.

To better understand the pharmacokinetic interaction potential of the licorice plant constituents, we have utilized a dual approach by applying integrated *in silico* structure-based PXR screening followed by *in vitro* testing. The molecular docking into PXR LBD can capture the global features, which might later turn on the activation cascade of the target genes. This approach is anticipated to increase the efficiency of the *in vitro* screening for preclinical studies. In this work, we have revealed multiple licorice components that essentially possessed high docking scores together with excellent interaction with the key amino acids such as S247, W299, and H407. A large group of isoflavonoids and stilbenoids common in licorice attained such features similar to the known activators. All of the top ranking ligands are prenylated with at least one prenyl group, which indicated the highly favorable hydrophobic interactions for PXR. This initial result also suggests the vital role of these prenylation patterns as a fingerprint for PXR recognition. It also demonstrates the bi-faceted nature that these prenylated patterns might bear, since they were shown in previous reports to increase the activity, selectivity, and the bioavailability as compared to their unprenylated congeners.

However, they appear to be a source of liability and should be dealt with caution. In this study, we sought to explore the validity of our *in silico* predictions by the application of *in vitro* screening of a home-designed resveratrol and dihydroresveratrol prenylation products archetypal of stilbenoids and dihydrostilbenoids present into different licorice species. Interestingly, these experiments manifested the ability of these compounds to induce PXR as predicted. Given the fact that PXR activation is highly sensitive for the hydrophobic features, one can surmise that the addition of prenyl groups can enhance the interactions with the LBD. The *in vitro* screening has supported this postulate with some of the compounds far more active than the unprenylated

counterparts. Some rudimentary SAR could be withdrawn from these results. Both unprenylated compounds, resveratrol, and DHR, had very weak but comparable activity against PXR at 25 μM concentration in accord with their low docking scores, with resveratrol having better ability in maintaining this activity at lower concentrations (6.25 μM). The abolished rigidity in the dihydroresveratrol **M11** had dramatically affected the pose prediction in the LBD as compared to resveratrol **M1**, which adopted more cis-like conformation to allow multiple polar interactions (**Figure 32**).

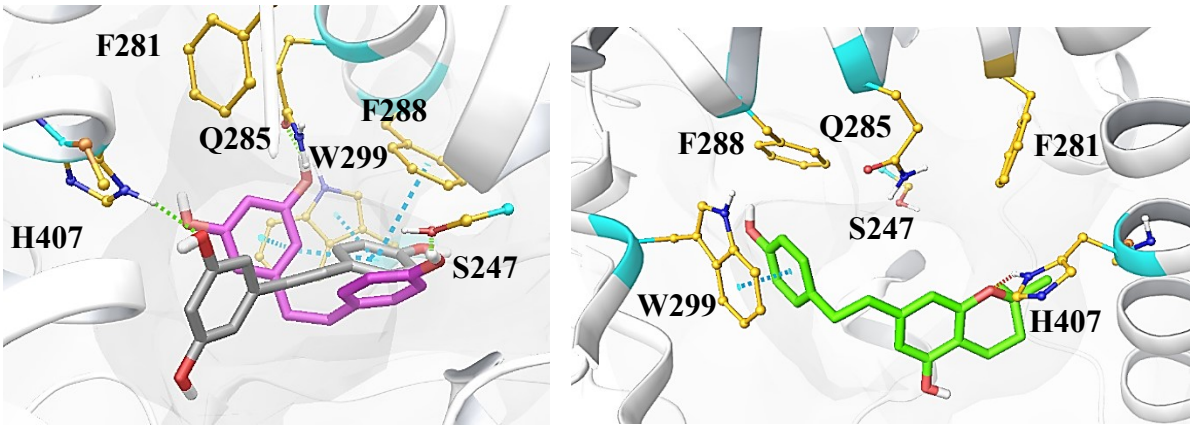


Figure 32. (Left) The docking pose of **M1** (grey sticks) and **M11** (purple sticks) in PDB:1NRL (white cartoon). **(Right)** The docking pose of M15 (green sticks) in the same crystal structure. π-π interactions are shown in blue dotted lines and H-bonds in red dotted lines.

Nevertheless, these interactions failed to increase the docking score (~ –6.5) or the activity of the compound as compared to the trans-resveratrol, most likely due to the increased entropy experienced with the increase in the number of the rotatable bonds. None of the mono-prenylated **M1** derivatives at positions C2 and C4 (as a straight chain or the cyclized THP), namely **M2**, **M3**, **M4**, and **M5** were able to induce PXR. The docking scores (–7.43, –8.563,

–9.922, –8.262 respectively) generally indicated a lower chance of activation but failed to sense the slightly active ones, which is acceptable as we indicated in the validation experiment. Compounds **M3** (arachidin 2) and its cyclized congener **M4** showed cell toxicity in accord with previous reports. In contrast, all the mono-prenylated DHS (**M12**, **M13**, **M14**, and **M15**) were active at the 25 μM concentration.

Interestingly, the more constrained mono-prenylated DHS (the cyclized versions) **M14** and **M15** were able to show salient multiple fold activation in the reporter gene assay (up to 4 folds at 25 μM). Moreover, **M15** showed at least 6-fold induction in CYP3A4 mRNA expression levels, which has further supported the PXR induction observed in the reporter gene assay. According to our cut-off value, these compounds are classified, as moderate activators. Although they were not well differentiated as compared to their resveratrol cyclized mono-prenylated mates, we noticed that the direction of the docking poses of these compounds were opposite (**Figure 32, Right**). We assumed that the free phenyl moiety of DHS (unprenylated side), which resides in the hydrophobic sub-pocket of both F288 and W299 acquire firm π-π interactions compared to the benzo-tetrahydropyran side (with the less aromatic character) of the stilbenoid posed at the same site. Of note, this might not be the case for some of the natural compounds found in licorice where they have a benzo-dehydropyran moiety, which pertains aromatic character in both ring systems. We have also examined the double prenylated derivatives represented with different patterns. Those compounds have their prenylation in a straight chain (**M6**, **M7**, **M16**, and **M17**), mixed patterns of straight prenyl and cyclized (**M8**, **M9**, **M10**) and the double cyclized (**M18**). Amazingly, as predicted by their higher scores, all the double prenylated compounds are PXR actives as verified by their 4-5 fold induction in the reporter assay. In addition, these compounds activated the luciferase reporter gene at very low

concentrations indicating a higher potency. Moreover, all the tested double prenylated compounds were also capable of inducing CYP3A4 mRNA expression.

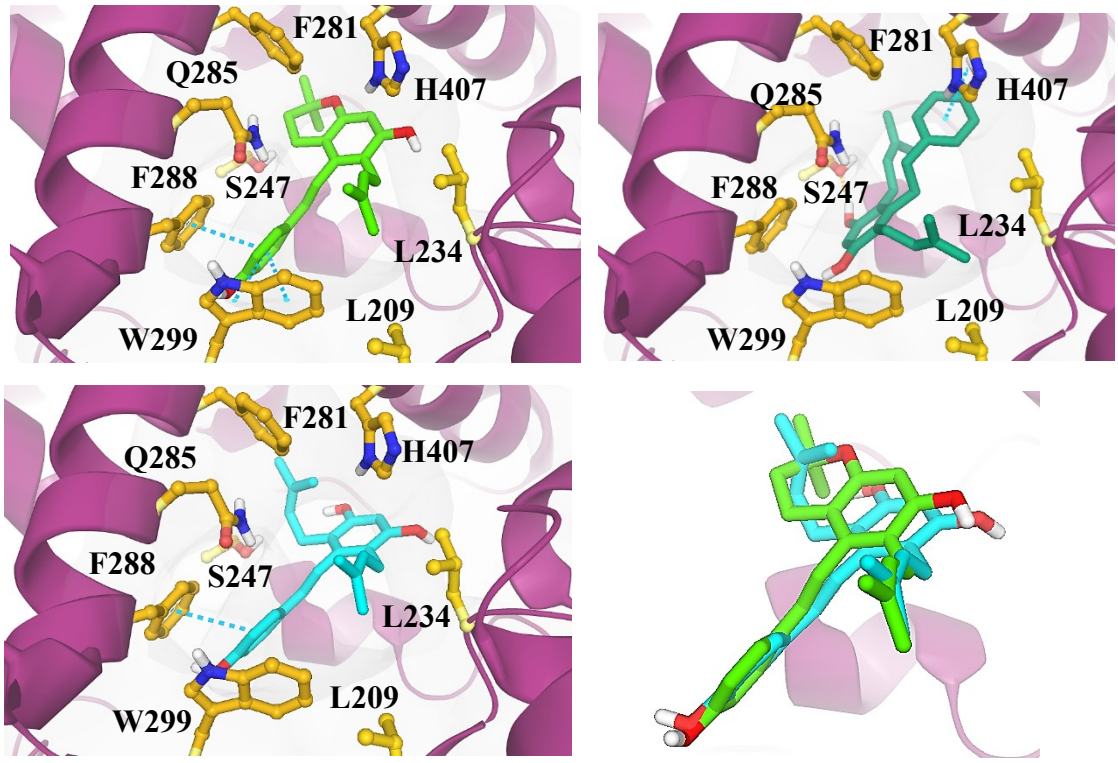


Figure 33. Docking pose of **M10** in green tubes **(Left up)**, **M7** in aqua blue **(left down)** and **M17** in bluish green tubes **(Right up)** in PDB: 1NRL (maroon cartoon) showing π-π bonds (blue dotted line) with W299 and P288 (orange ball and sticks). **(Right down) M7** and **M10** overlaid docking pose.

However, high levels of mRNA expression are noticed for the tested compounds **M7**, **M10** and **M15** (up to 6 folds). **M7** and **M10** have secured π-π interactions with the key residues F288 and W299 as shown in **Figure 33**. These compounds are prenylated at the same positions, but the cyclization of one of the prenyl group (**M10**) did not affect the activity. In contrast, the rigidity of the compound seems to have a prominent role at these positions (based on mRNA levels), which is apparent by comparing **M7** and **M10** (stilbenoids) with **M17** (DHS).

It is also of vital importance to screen for allosteric antagonist activity since some of the

compounds might have a preferred interaction and selectivity toward the AF-2 region rather than the LBD especially in the presence of other agonists. Ketoconazole is the prototype of an allosteric PXR modifier, which was proposed to interact with the AF-2 region precluding the coactivator from binding. Other natural products were also described such as coumestrol and resveratrol. However, for resveratrol, controversial conclusions were drawn from different studies about its allosteric antagonistic character. To evaluate the ability of resveratrol and the other tested compounds to act as pure agonists or allosteric antagonists, selected representative compounds of different prenylation patterns were screened in the reporter assay together with rifampicin. Our findings revealed that resveratrol and the other derivatives failed to antagonize the rifampicin induced PXR in HepG2 cells at the tested concentration range, suggesting that these stilbenoids including DHS might be binding at allosteric site.

These tested synthesized compounds resonated with the DHS present in the licorice species except that the latter have the piceatannol as a precursor rather than resveratrol. It is well-established that piceatannol is a direct metabolite of resveratrol with one extra hydroxy group permitting an extra hydrogen bond. Hence, they have achieved higher scores in the initial screening. Some of the alerting "*in silico* identified" compounds found in licorice are species-specific. For example, the diprenylated DHS **U15** and **U16** found in *Glycyrrhiza uralensis*. Others belong to the cycloprenylated isoflavenes, which are species-specific markers of *G. glabra*. Moreover, multiple moderately active compounds may synergistically activate PXR and dispose of a higher risk for HDI.

IV.4 Experimental

IV.4.1 Docking experiments

Both proteins crystal structures PDB: ID 1NRL and 1M13 were prepared using the

Protein Preparation Wizard (PrepWizard, Schrodinger 10.3). For both structures, hydrogen atoms were added after the deletion of original ones, and the bond orders were assigned. The protein structure was checked and the missing residues or loop segments were added using Prime. Non-hydrogen- bonding water molecules were removed. Only chains B and D for 1NRL and chain A for 1M13 were included in the current study. The protonation and tautomeric states of H, Q, S, T, and E were adjusted to match a pH of 7.4. Possible orientations of side chain were generated and checked. Active site water molecules beyond 5.0 Å from the ligand were deleted. Then, the protein-ligand complex was subjected to geometry refinement using an OPLS2005 force field, and restrained minimization. The convergence of heavy atoms was set to an RMSD of 0.3 Å. Docking grids for both 1NRL and 1M13 were generated with the default settings in Glide using the co-crystalized ligands to define the center of the grid box (20 x 20 x 20Å). No constraints were included during grid generation. The generated databases of both *G. glabra* and *G. uralensis* (described in Chapter II) in addition to the generated library of positive controls (**Table 5**), which includes known potent ligands (hyperforin and 4-hydroxytamoxifen) as well as a moderate inducer (hypericine and enterolactone) and the weak activators (genistein and physcion) were prepared using LigPrep module. Initially, the 2D structures of these molecules were sketched using ChemDraw, converted to 3D structures, and then saved as SD files before they were exported to Maestro.

After 3D structures were loaded, OPLS2005 force field and charges were used in all ligand preparation steps. Possible protonation states and ionization states were explored for each ligand using Epik at a pH of 7.4. Chirality was determined from 3D structures and compared with reported literature. Tautomeric states were generated for chemical groups with possible prototrophic tautomerism. Only the lowest energy conformer was kept for each ligand. For

validation and comparison purposes, the positive controls were docked in the generated grids of both 1NRL and 1M13 using flexible ligand protocol of both SP and XP glide docking. For screening of the generated licorice libraries, they were docked into 1NRL using XP flexible docking. The compounds, which scored below −9.5 kcal/mol, were kept for further analysis.

IV.4.2 PXR activation

IV.4.3 Reporter gene assay

PXR activation by the synthesized compounds was measured in HepG2 cell lines transiently transfected with 25 μg of pSG5-hPXR (the expression vector was provided by Dr. Steven Kliewer, University of Texas Southwestern Medical Center) and 25 μg of PCR5 plasmid DNA (the reporter plasmid was provided by Dr. Christopher Liddle, University of Sydney) by electroporation as described earlier. The 96-well plates were seeded with cells at a density of 50,000 cells per well. When the cells reach > 90% confluency after 24 hrs, the synthesized compounds and the positive control were added at 25, 12.5, 6.25, 3.13, 1.56 and 0.78 μg/mL. Subsequently, the media was aspirated after incubation with the test compounds for 48 hrs, and 40 μL of luciferase reagent (Promega Corporation, Madison, WI, USA) was added to each well. After which, the luminescence was measured on Spectramax M5 plate reader (Molecular Devices, Sunnyvale, CA, USA). The activation of luciferase in the treated cells was calculated in comparison with vehicle-treated cells.

IV.4.4 RT-PCR analysis of CYP3A4

CYP3A4 primers qHsaCID0012316 (both the forward and reverse primers) were purchased from Bio-Rad Laboratories (Hercules, CA, USA). HepG2 cells transfected (> 90% confluency) as described above, were implanted into 6-well plate (5×10^5 per well) and they were allowed to grow for 24 hrs. Subsequently, cells were treated with three different

concentrations of test samples (10, 3 and 1 μM; 10μM for rifampicin (the positive control) for 48 h. The RNA was extracted using the Quick-start protocol (Qiagen®) after washing the cells with PBS. The concentration of RNA was calculated using a BioTek Take3 plate, which uses the A260/A280 ratio to assess the purity. The template used in BioRad's iScript Reverse Transcription kit was RNA equal to 0.5μg/ml for initial strand complementary DNA synthesis. The resultant complementary DNA strand was fed as the template into BioRad's iTaq Universal SYBR Green supermix. RT-PCR was performed in a 96-well plate employing a CFX connect real-time PCR detector system (Bio-Rad). The housekeeping gene HPRT1 (qHsaCID0016375) was used to normalize the quantification of expressed CYP3A4 genes. The fold induction in mRNA expression post-treatment was measured in comparison with the vehicle as described earlier.[168]

CHAPTER V
CONCLUSIONS AND FUTURE DIRECTIONS

The biodiversity offered by NPs and the successful records achieved in the drug discovery based on NPs underline their significant roles in this area. Marked therapeutics of anticancer, anti-infective, lipid-lowering, antimalarial, and immunosuppressant are either NPs or NPs derivatives. Despite the vast developments in the analytical and screening techniques, the activity and risk assessment of plant-based therapy is still challenging. The extensive challenges of an NP-based discovery made them less present in the pharmaceutical industry paradigms. However, the decline in the vital drugs registered or approved by the FDA has changed their outlook for screening programs. Although it is an extremely intricate situation, yet cheminformatic techniques can offer a variety of solution models to deal with the complexity of NPs. Of note, the application of medicinal chemistry roles for filtration of virtual libraries would avoid most of the NPs structural features since they break many regular rules practiced in other disciplines.

However, the unique structural features of NPs evolved to recognize natural target motives in living systems. The application of *in silico* tools to search for these untapped NPs might unveil unprecedented activities. Moreover, the increase in consumption of herbal supplements to alleviate various ailments have outpaced the understanding of their safety or efficacy. This situation has created a knowledge gap in the herbal remedy assessments and revealed the deficiency for a fit of purpose methodology that could deal with complex mixtures

of an herbal supplement. For instance, the preference of a reputable herbal remedy to alleviate menopausal symptoms is encouraged by the failure of therapeutic estrogens to demonstrate adequate safety under different dosage forms, route of administration, or formulations. This trend is further stimulated by the pertained polypharmacology and the perceived safety of herbal treatments. One characteristic plant in this category is licorice. Similar to other herbal remedies, licorice offers digestive relief, antioxidant, anti-inflammatory, antiviral, antibacterial, and chemopreventive properties in addition to its estrogenic nature. However, the rise in licorice consumption and self-medication together with prescription drugs and/or over the counter medication is a major concern for metabolic disruption as well as HDI. In particular, licorice is consumed as an adjuvant therapy together with conventional drugs, and it is present in 70% of the TCM herbal prescriptions, where it is described as a guide drug. Moreover, many medical practitioners believe that it has powerful chemopreventive and antitumor effects. Whereas, some of these co-administered drugs are narrow-indexed therapeutics, which are commonly consumed by the targeted population of elderly women for chemoprevention in a high-risk of breast cancer or with anticancer agents themselves. Hence, there is a great potential of unrecognized HDI that could be attributed to other improper causes, which often raise questions with no specific rational answers.

To evaluate the applicability of cheminformatic methods in such a situation, we have utilized a platform of ensemble docking and QSAR techniques to study its estrogenic activity. Our results allocated a multitude of potential estrogenic components found in licorice diverse phenolic metabolome. Many of them are present with different prenylation patterns that could alter both *h*ER isoforms. The connection of the plant estrogenic activity to a single entity is an oversimplification, which undermines its safety. Deconvolution of the phenolic metabolome of

both *G. glabra* and *G. uralensis* revealed around 50 unique compounds in each species that potentially could alter the *h*ER pathways in different manners of agonism or antagonism, which ultimately might lead to endocrine disruption by its chronic use.

This situation implies the significance of characterization and purification when needed for potential phytoalexins or phytoSERMs that might be involved in such activity. The application of these techniques was successfully validated by the identification of the already known *h*ER active compounds from licorice. After receiving favorable results thus far, we decided to focus onto one of the top scoring components based on both ensemble docking and QSAR analysis that have not been yet rigorously explored. The DHS, tethered with different prenylation patterns, showed favorable interactions in both *h*ERs structures that were dependent on the position of the prenyl group. To further analyze the effect of the prenylation pattern on DHS activity, we aimed to synthesize a small library of prenylated DHS archetypal of those found in licorice. Fortunately, a diverse set of prenylation patterns could be achieved in one pot reaction as we described in chapter III. For initial screening purposes, we have found that the divergent installation of the

prenyl group is convenient to provide sufficient quantities for biological testing. Two series of stilbenoids (**M1-M10**) and DHS (**M11-M18**) were synthesized. Multiple structures were synthesized for the same time (**M8**, **M9**, **M14**, **M15**, **M17**, and **M18**). Having this library of compounds in hand, the *in vitro* analysis of estrogenic activity is ongoing and the results will be reported in the near future to verify what these patterns might reveal as a unique SERM activity. These results would also help us to comprehend the estrogenic activity of licorice plant by validation of these methods. Particularly, some of these diprenylated DHS are species-specific for *G. uralensis*.

Licorice plant was described to have high potential for exerting HDI. However, different mechanisms could be involved in this behavior. Having the induced herb-drug interaction of licorice supplements *via* PXR less studied; we ventured to analyze the potential PXR activation by the application of *in silico* tools. We have coupled the molecular docking experiments with the *in vitro* testing to verify PXR agonists. We have identified many compounds, which are potentially capable of triggering PXR; some of these compounds belong to a specific licorice species. Interestingly enough, the prenylated DHS was one of the identified alerting classes for PXR activation. As a proof of concept, we have further validated our approach by screening the DHS synthesized library with the cell-based *in vitro* luciferase reporter gene assay. As a result, this study has unveiled prenylated stilbenoids and DHS as a potential source of PXR activation. Furthermore, these results were confirmed by the CYP3A4 mRNA expression, which was susceptible to 6 fold activation by multiple tested compounds, especially those of cyclized and diprenylated ones similar to glabridin-like metabolites. This situation raised some questions about higher prenylation patterns, ubiquitous in nature or intended for lead optimization, as a PXR phenotypic alerts.

To verify other mechanisms of HDI, we have screened the licorice components against multiple available QSAR CYP inhibition models. This initial analysis has provided us with a list of potentially problematic compounds that might participate in the precipitation of some risks upon licorice consumption such as HDIs. Two out of five CYPs, namely CYP3A4 and CYP1A2 were predicted as a major concern for direct CYP inhibition in both licorice species under investigation. However, *G. uralensis* is predicted to have a higher burden against CYP2C9. It is important to note that these predictions are void of pharmacokinetics predictions, and further physiologically based models are required to evaluate the in vivo reliability of the predicted

inhibition.

Taken together, the application of cheminformatic tools can provide us with immense information that could be the key to untangle the complex mixtures of herbal remedies and their activity or safety profiles. Furthermore, there is a continuous need to find new drugs for reluctant ailments or resistant therapeutics. NPs are sustainable resources that can supply us with compelling untapped structural activities by the application of the appropriate techniques.

Finally, with the versatile plausible mechanisms of *h*ER interruption that could be exerted by the phenolic content of licorice, in addition to the different HDI mechanism discussed in this research, we find many concerns regarding licorice plant as a safe surrogate of estrogen. Further validation models of endocrine and metabolic disruption should be evaluated. Moreover, our research results trigger many questions: Is *h*ERβ antagonism capable of turning on proliferation effects via deletion of its negative feedbacks against *h*ERα? Does higher prenylation patterns found in many licorice phenolic metabolome exert antagonism against *h*ER-β similar to its capacity of counteracting *h*ERα? And if it does what would be the end result in tissues populated with both *h*ERs isoforms? Which, if any, of those predicted compounds have preferable SERM characteristics? Have our bodies already developed the tools to recognize these prenylation patterns and made them self-limiting? Is it the induction rather than inhibition route of HDI to take over in *in vivo* setting? Faced with all these overwhelming questions, we have more endeavors to explore and find the answers in the future.

REFERENCES

1. Ganesan, A., The impact of natural products upon modern drug discovery. *Current opinion in chemical biology* **2008,** *12* (3), 306-317.
2. Koehn, F. E.; Carter, G. T., The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* **2005,** *4* (3), 206-20.
3. Plenge, R., Disciplined approach to drug discovery and early development. *Sci. Transl. Med.* **2016,** *8* (349), 15.
4. Fang, J.; Liu, C.; Wang, Q.; Lin, P.; Cheng, F., In silico polypharmacology of natural products. *Brief Bioinform* **2018,** *19* (6), 1153-1171.
5. Ciardiello, J. J.; Stewart, H. L.; Sore, H. F.; Galloway, W.; Spring, D. R., A novel complexity-to-diversity strategy for the diversity-oriented synthesis of structurally diverse and complex macrocycles from quinine. *Bioorg Med Chem* **2017,** *25* (11), 2825-2843.
6. Pammolli, F.; Magazzini, L.; Riccaboni, M., The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov* **2011,** *10* (6), 428-38.
7. Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A., An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov* **2015,** *14* (7), 475-86.
8. Ogbourne, S. M.; Parsons, P. G., The value of nature's natural product library for the discovery of New Chemical Entities: the discovery of ingenol mebutate. *Fitoterapia* **2014,** *98*, 36-44.
9. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* **2016,** *79* (3), 629-61.
10. Fukushima, M., Resilience in scientific research: Understanding how natural product research rebounded in an adverse situation. *Science as Culture* **2016,** *25* (2), 167-192.
11. Laraia, L.; Waldmann, H., Natural product inspired compound collections: evolutionary principle, chemical synthesis, phenotypic screening, and target identification. *Drug Discovery Today: Technologies* **2017,** *23*, 75-82.
12. Rodrigues, T.; Reker, D.; Schneider, P.; Schneider, G., Counting on natural products for drug design. *Nat Chem* **2016,** *8* (6), 531-41.
13. Hogle, B. C.; Guan, X.; Folan, M. M.; Xie, W., PXR as a mediator of herb-drug interaction. *J Food Drug Anal* **2018,** *26* (2S), S26-S31.
14. Smith, T.; Kawa, K.; Eckl, V., Herbal Supplement Sales in US Increase 7.7% in 2016. *herbalgram* **2017,** (115), 56-64.
15. Geller, S. E.; Shulman, L. P.; van Breemen, R. B.; Banuvar, S.; Zhou, Y.; Epstein, G.; Hedayat, S.; Nikolic, D.; Krause, E. C.; Piersen, C. E.; Bolton, J. L.; Pauli, G. F.; Farnsworth, N. R., Safety and efficacy of black cohosh and red clover for the management of vasomotor symptoms: a randomized controlled trial. *Menopause* **2009,** *16* (6), 1156-66.
16. Gurley, B. J.; Yates, C. R.; Markowitz, J. S., "...Not Intended to Diagnose, Treat, Cure or Prevent Any Disease." 25 Years of Botanical Dietary Supplement Research and the

Lessons Learned. *Clin Pharmacol Ther* **2018,** *104* (3), 470-483.
17. Dietz, B. M.; Hajirahimkhan, A.; Dunlap, T. L.; Bolton, J. L., Botanicals and their bioactive phytochemicals for women's health. *Pharmacological reviews* **2016,** *68* (4), 1026-1073.
18. Baell, J. B., Feeling Nature's PAINS: Natural Products, Natural Product Drugs, and Pan Assay Interference Compounds (PAINS). *J Nat Prod* **2016,** *79* (3), 616-28.
19. Melagraki, G.; Ntougkos, E.; Papadopoulou, D.; Rinotas, V.; Leonis, G.; Douni, E.; Afantitis, A.; Kollias, G., In Silico Discovery of Plant-Origin Natural Product Inhibitors of Tumor Necrosis Factor (TNF) and Receptor Activator of NF-kappaB Ligand (RANKL). *Front Pharmacol* **2018,** *9*, 800.
20. Do, Q. T.; Lamy, C.; Renimel, I.; Sauvan, N.; Andre, P.; Himbert, F.; Morin-Allory, L.; Bernard, P., Reverse pharmacognosy: identifying biological properties for plants by means of their molecule constituents: application to meranzin. *Planta Med* **2007,** *73* (12), 1235-40.
21. Medema, M. H.; Fischbach, M. A., Computational approaches to natural product discovery. *Nat Chem Biol* **2015,** *11* (9), 639-48.
22. Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J., The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* **2015,** *14* (2), 111-29.
23. Abagyan, R.; Kufareva, I., The flexible pocketome engine for structural chemogenomics. *Methods Mol Biol* **2009,** *575*, 249-79.
24. Breinbauer, R.; Vetter, I. R.; Waldmann, H., From Protein Domains to Drug Candidates–Natural Products as Guiding Principles in the Design and Synthesis of Compound Libraries. *Angew. Chem. Int. Ed.* **2002,** *41*, 2878-2890.
25. Johnson, E. J.; Gonzalez-Perez, V.; Tian, D. D.; Lin, Y. S.; Unadkat, J. D.; Rettie, A. E.; Shen, D. D.; McCune, J. S.; Paine, M. F., Selection of Priority Natural Products for Evaluation as Potential Precipitants of Natural Product-Drug Interactions: A NaPDI Center Recommended Approach. *Drug Metab Dispos* **2018,** *46* (7), 1046-1052.
26. Grimstein, M.; Huang, S. M., A regulatory science viewpoint on botanical-drug interactions. *J Food Drug Anal* **2018,** *26* (2S), S12-S25.
27. Grollman, A. P.; Marcus, D. M., Global hazards of herbal remedies: lessons from Aristolochia: The lesson from the health hazards of Aristolochia should lead to more research into the safety and efficacy of medicinal plants. *EMBO Rep* **2016,** *17* (5), 619-25.
28. Sprouse, A. A.; van Breemen, R. B., Pharmacokinetic Interactions between Drugs and Botanical Dietary Supplements. *Drug Metab Dispos* **2016,** *44* (2), 162-71.
29. Ekins, S.; Andreyev, S.; Ryabov, A.; Kirillov, E.; Rakhmatulin, E. A.; Bugrim, A.; Nikolskaya, T., Computational prediction of human drug metabolism. *Expert Opin Drug Metab Toxicol* **2005,** *1* (1), 1-21.
30. Kirchmair, J.; Goller, A. H.; Lang, D.; Kunze, J.; Testa, B.; Wilson, I. D.; Glen, R. C.; Schneider, G., Predicting drug metabolism: experiment and/or computation? *Nat Rev Drug Discov* **2015,** *14* (6), 387-404.
31. Raunio, H.; Kuusisto, M.; Juvonen, R. O.; Pentikainen, O. T., Modeling of interactions between xenobiotics and cytochrome P450 (CYP) enzymes. *Front Pharmacol* **2015,** *6*, 123.
32. Ekins, S.; Kortagere, S.; Iyer, M.; Reschly, E. J.; Lill, M. A.; Redinbo, M. R.; Krasowski, M. D., Challenges predicting ligand-receptor interactions of promiscuous proteins: the

nuclear receptor PXR. *PLoS Comput Biol* **2009,** *5* (12), e1000594.
33. Clark, R. D., Predicting mammalian metabolism and toxicity of pesticides in silico. *Pest Manag Sci* **2018**.
34. Kufareva, I.; Chen, Y.-C.; Ilatovskiy, A. V.; Abagyan, R., Compound activity prediction using models of binding pockets or ligand properties in 3D. *Curr Top Med Chem* **2012,** *12* (17), 1869–1882.
35. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **2004,** *3* (11), 935-49.
36. Elokely, K. M.; Doerksen, R. J., Docking challenge: protein sampling and molecular docking performance. *J Chem Inf Model* **2013,** *53* (8), 1934-45.
37. Politi, R.; Convertino, M.; Popov, K.; Dokholyan, N. V.; Tropsha, A., Docking and Scoring with Target-Specific Pose Classifier Succeeds in Native-Like Pose Identification But Not Binding Affinity Prediction in the CSAR 2014 Benchmark Exercise. *J Chem Inf Model* **2016,** *56* (6), 1032-41.
38. Sotriffer, C. A., Protein–Ligand Docking: From Basic Principles to Advanced Applications. In *In Silico Drug Discovery and Design*, CRC Press: 2015; pp 164-197.
39. Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D., Molecular docking and structure-based drug design strategies. *Molecules* **2015,** *20* (7), 13384-421.
40. Mixing Pharmacophore Modeling and Classical QSAR Analysis as Powerful Tool for Lead Discovery. In *Virtual Screening*, Taha, M., Ed. Intech: 2012.
41. Lam, P.; Abagyan, R.; Totrov, M., Ligand-biased ensemble receptor docking (LigBEnD): a hybrid ligand/receptor structure-based approach. *J Comput Aided Mol Des* **2018,** *32*, 187–198.
42. Andricopulo, A. D.; Ferreira, L. L., *Chemoinformatics Approaches to Structure-and Ligand-Based Drug Design*. Frontiers Media SA: 2019.
43. Karkanis, A.; Martins, N.; Petropoulos, S. A.; Ferreira, I. C. F. R., Phytochemical composition, health effects, and crop management of liquorice (Glycyrrhiza glabra L.): A medicinal plant. *Food Reviews International* **2016,** *34* (2), 182-203.
44. Hosseinzadeh, H.; Nassiri-Asl, M., Pharmacological Effects of Glycyrrhiza spp. and Its Bioactive Constituents: Update and Review. *Phytother Res* **2015,** *29* (12), 1868-86.
45. Shakeri, A.; Akhtari, J.; Soheili, V.; Taghizadeh, S. F.; Sahebkar, A.; Shaddel, R.; Asili, J., Identification and biological activity of the volatile compounds of Glycyrrhiza triphylla Fisch. & CA Mey. *Microbial pathogenesis* **2017,** *109*, 39-44.
46. Wong, Y. F.; Cacciola, F.; Fermas, S.; Riga, S.; James, D.; Manzin, V.; Bonnet, B.; Marriott, P. J.; Dugo, P.; Mondello, L., Untargeted profiling of Glycyrrhiza glabra extract with comprehensive two-dimensional liquid chromatography-mass spectrometry using multi-segmented shift gradients in the second dimension: Expanding the metabolic coverage. *Electrophoresis* **2018,** *39* (15), 1993-2000.
47. Khalaf, I.; Vlase, L.; Ivanescu, B.; Lazar, D.; Corciova, A., HPLC analysis of polyphenolic compounds phytoestrogens and sterols from Glycyrrhiza glabra L. tincture. *Stud. Univ. Babes-Bolyai, Chem.* **2012,** *57* (2), 113-118.
48. Liao, W. C.; Lin, Y.-H.; Chang, T.-M.; Huang, W.-Y., Identification of two licorice species, Glycyrrhiza uralensis and Glycyrrhiza glabra, based on separation and identification of their bioactive components. *Food chemistry* **2012,** *132* (4), 2188-2193.
49. Frommenwiler, D. A.; Maire-Widmer, V.; Upton, R.; Nichols, J.; Heubl, G.; Reich, E.,

Qualitative and quantitative characterization of two licorice root species (Glycyrrhiza glabra L. and Glycyrrhiza uralensis Fisch.) by HPTLC, validated by HPLC and DNA sequencing. *JPC-Journal of Planar Chromatography-Modern TLC* **2017,** *30* (6), 467-473.

50. Meng, X.; Li, H.; Song, F.; Liu, C.; Liu, Z.; Liu, S., Studies on triterpenoids and flavones in Glycyrrhiza uralensis Fisch. by HPLC-ESI-MSn and FT-ICR-MSn. *Chin. J. Chem.* **2009,** *27* (2), 299-305.
51. Qiao, X.; Liu, C.-F.; Ji, S.; Lin, X.-H.; Guo, D.-A.; Ye, M., Simultaneous Determination of Five Minor Coumarins and Flavonoids in Glycyrrhiza uralensis by Solid-Phase Extraction and High-Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry. *Planta Med.* **2014,** *80* (2/3), 237-242.
52. Rizzato, G.; Scalabrin, E.; Radaelli, M.; Capodaglio, G.; Piccolo, O., A new exploration of licorice metabolome. *Food Chem* **2017,** *221*, 959-968.
53. Farag, M. A.; Porzel, A.; Wessjohann, L. A., Comparative metabolite profiling and fingerprinting of medicinal licorice roots using a multiplex approach of GC-MS, LC-MS and 1D NMR techniques. *Phytochemistry (Elsevier)* **2012,** *76*, 60-72.
54. Song, W.; Qiao, X.; Chen, K.; Wang, Y.; Ji, S.; Feng, J.; Li, K.; Lin, Y.; Ye, M., Biosynthesis-based quantitative analysis of 151 secondary metabolites of Licorice to differentiate medicinal Glycyrrhiza species and their hybrids. *Analytical chemistry* **2017,** *89* (5), 3146-3153.
55. Boonmuen, N.; Gong, P.; Ali, Z.; Chittiboyina, A. G.; Khan, I.; Doerge, D. R.; Helferich, W. G.; Carlson, K. E.; Martin, T.; Piyachaturawat, P.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S., Licorice root components in dietary supplements are selective estrogen receptor modulators with a spectrum of estrogenic and anti-estrogenic activities. *Steroids* **2016,** *105*, 42-9.
56. Oseni, T.; Patel, R.; Pyle, J.; Jordan, V. C., Selective estrogen receptor modulators and phytoestrogens. *Planta Med* **2008,** *74* (13), 1656-65.
57. Anandhi Senthilkumar, H.; Fata, J. E.; Kennelly, E. J., Phytoestrogens: The current state of research emphasizing breast pathophysiology. *Phytother Res* **2018**.
58. Moreira, A. C.; Silva, A. M.; Santos, M. S.; Sardao, V. A., Phytoestrogens as alternative hormone replacement therapy in menopause: What is real, what is unknown. *The Journal of steroid biochemistry and molecular biology* **2014,** *143*, 61-71.
59. Somjen, D.; Knoll, E.; Vaya, J.; Stern, N.; Tamir, S., Estrogen-like activity of licorice root constituents: glabridin and glabrene, in vascular tissues in vitro and in vivo. *J Steroid Biochem Mol Biol* **2004,** *91* (3), 147-55.
60. Tamir, S.; Eizenberg, M.; Somjen, D.; Stern, N.; Shelach, R.; Kaye, A.; Vaya, J., Estrogenic and Antiproliferative Properties of Glabridin from Licorice in Human Breast Cancer Cells. *Cancer Res.* **2000,** *60*, 5704–5709.
61. Hu, C.; Liu, H.; Du, J.; Mo, B.; Qi, H.; Wang, X.; Ye, S.; Li, Z., Estrogenic activities of extracts of Chinese licorice (Glycyrrhiza uralensis) root in MCF-7 breast cancer cells. *J Steroid Biochem Mol Biol* **2009,** *113* (3-5), 209-16.
62. Hajirahimkhan, A.; Mbachu, O.; Simmler, C.; Ellis, S. G.; Dong, H.; Nikolic, D.; Lankin, D. C.; van Breemen, R. B.; Chen, S. N.; Pauli, G. F.; Dietz, B. M.; Bolton, J. L., Estrogen Receptor (ER) Subtype Selectivity Identifies 8-Prenylapigenin as an ERbeta Agonist from Glycyrrhiza inflata and Highlights the Importance of Chemical and Biological Authentication. *J Nat Prod* **2018,** *81* (4), 966-975.

63. Hajirahimkhan, A.; Simmler, C.; Yuan, Y.; Anderson, J. R.; Chen, S. N.; Nikolic, D.; Dietz, B. M.; Pauli, G. F.; van Breemen, R. B.; Bolton, J. L., Evaluation of estrogenic activity of licorice species in comparison with hops used in botanicals for menopausal symptoms. *PLoS One* **2013,** *8* (7), e67947.
64. Katzenellenbogen, B.; Katzenellenbogen, J., defining the S in SERMs. *Science* **2002,** *259*, 2380-2381.
65. Traboulsi, T.; El Ezzy, M.; Gleason, J. L.; Mader, S., Antiestrogens: structure-activity relationships and use in breast cancer treatment. *J Mol Endocrinol* **2017,** *58* (1), R15-R31.
66. Patel, H. K.; Bihani, T., Selective estrogen receptor modulators (SERMs) and selective estrogen receptor degraders (SERDs) in cancer treatment. *Pharmacol Ther* **2018,** *186*, 1-24.
67. Katzenellenbogen, J. A.; Mayne, C. G.; Katzenellenbogen, B. S.; Greene, G. L.; Chandarlapaty, S., Structural underpinnings of oestrogen receptor mutations in endocrine therapy resistance. *Nat Rev Cancer* **2018,** *18* (6), 377-388.
68. Wu, A. H.; Yu, M. C.; Tseng, C. C.; Pike, M. C., Epidemiology of soy exposures and breast cancer risk. *Br J Cancer* **2008,** *98* (1), 9-14.
69. Moore, T. R.; Franks, R. B.; Fox, C., Review of efficacy of complementary and alternative medicine treatments for menopausal symptoms. *Journal of midwifery & women's health* **2017,** *62* (3), 286-297.
70. Rietjens, I.; Louisse, J.; Beekmann, K., The potential health effects of dietary phytoestrogens. *Br J Pharmacol* **2017,** *174* (11), 1263-1280.
71. Chen, H. H.; Chen, S. P.; Zheng, Q. L.; Nie, S. P.; Li, W. J.; Hu, X. J.; Xie, M. Y., Genistein Promotes Proliferation of Human Cervical Cancer Cells Through Estrogen Receptor-Mediated PI3K/Akt-NF-kappaB Pathway. *J Cancer* **2018,** *9* (2), 288-295.
72. Sanchez-Borrego, R.; Navarro, M. C.; Llaneza, P.; Hormigo, A.; Duran, M.; Mendoza, N., Efficacy and safety of a phyto-SERM as an alternative to hormone therapy. *Climacteric* **2015,** *18* (3), 350-7.
73. Stulikova, K.; Karabin, M.; Nespor, J.; Dostalek, P., Therapeutic Perspectives of 8-Prenylnaringenin, a Potent Phytoestrogen from Hops. *Molecules* **2018,** *23* (3).
74. Patisaul, H. B., Endocrine disruption by dietary phyto-oestrogens: impact on dimorphic sexual systems and behaviours. *Proc Nutr Soc* **2017,** *76* (2), 130-144.
75. Burris, T. P.; Solt, L. A.; Wang, Y.; Crumbley, C.; Banerjee, S.; Griffett, K.; Lundasen, T.; Hughes, T.; Kojetin, D. J., Nuclear receptors and their selective pharmacologic modulators. *Pharmacol Rev* **2013,** *65* (2), 710-78.
76. Warner, M.; Huang, B.; Gustafsson, J. A., Estrogen Receptor beta as a Pharmaceutical Target. *Trends Pharmacol Sci* **2017,** *38* (1), 92-99.
77. Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L., The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998,** *95* (7), 927-937.
78. Katzenellenbogen, J. A.; Muthyala, R.; Katzenellenbogen, B. S., Nature of the ligand-binding pocket of estrogen receptor a and b: The search for subtypeselective ligands and implications for the prediction of estrogenic activity. *Pure and Applied Chemistry* **2003,** *75*, 2397–2403.
79. Schans, M. G. M. v. d.; Ritschel, T.; Bovee, T. F. H.; Sanders, M. G.; PieterdeWaard; Gruppen, H.; Vincken, J.-P., Involvement of a Hydrophobic Pocket and Helix 11 in

Determining the Modes of Action of Prenylated Flavonoids and Isoflavonoids in the Human Estrogen Receptor. *ChemBioChem* **2015,** *16*, 2668 –2677.
80. Nettles, K. W.; Bruning, J. B.; Gil, G.; O'Neill, E. E.; Nowak, J.; Hughs, A.; Kim, Y.; DeSombre, E. R.; Dilis, R.; Hanson, R. N., Structural plasticity in the oestrogen receptor ligand-binding domain. *EMBO reports* **2007,** *8* (6), 563-568.
81. Ng, H. W.; Perkins, R.; Tong, W.; Hong, H., Versatility or promiscuity: The estrogen receptors, control of ligand selectivity and an update on subtype selective ligands. *International journal of environmental research and public health* **2014,** *11* (9), 8709-8742.
82. De Angelis, M.; Stossi, F.; Waibel, M.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A., Isocoumarins as estrogen receptor beta selective ligands: Isomers of isoflavone phytoestrogens and their metabolites. *Bioorg Med Chem* **2005,** *13* (23), 6529-42.
83. Balaji, B.; Ramanathan, M., Prediction of estrogen receptor beta ligands potency and selectivity by docking and MM-GBSA scoring methods using three different scaffolds. *J Enzyme Inhib Med Chem* **2012,** *27* (6), 832-44.
84. Naafs, M., Clinical Pharmacodynamics of Endocrine Disruptors: A Historic Perspective. Endocrinol Diabetes Res 4: 1. *of* **2017,** *13*, 6-7.
85. Ruiz, P.; Sack, A.; Wampole, M.; Bobst, S.; Vracko, M., Integration of in silico methods and computational systems biology to explore endocrine-disrupting chemical binding with nuclear hormone receptors. *Chemosphere* **2017,** *178*, 99-109.
86. Andersson, A. M.; Bay, K.; Frederiksen, H.; Skakkebaek, N. E., Endocrine disrupters: we need research, biomonitoring and action. *Andrology* **2016,** *4* (4), 556-60.
87. Judson, R. S.; Magpantay, F. M.; Chickarmane, V.; Haskell, C.; Tania, N.; Taylor, J.; Xia, M.; Huang, R.; Rotroff, D. M.; Filer, D. L.; Houck, K. A.; Martin, M. T.; Sipes, N.; Richard, A. M.; Mansouri, K.; Setzer, R. W.; Knudsen, T. B.; Crofton, K. M.; Thomas, R. S., Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor. *Toxicol Sci* **2015,** *148* (1), 137-54.
88. Kolšek, K.; Mavri, J.; Sollner Dolenc, M.; Gobec, S.; Turk, S., Endocrine Disruptome An Open Source Prediction Tool for Assessing Endocrine Disruption Potential through Nuclear Receptor Binding. ACS Publications: 2014.
89. Ekins, S.; Bugrim, A.; Brovold, L.; Kirillov, E.; Nikolsky, Y.; Rakhmatulin, E.; Sorokina, S.; Ryabov, A.; Serebryiskaya, T.; Melnikov, A.; Metz, J.; Nikolskaya, T., Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. *Xenobiotica* **2006,** *36* (10-11), 877-901.
90. Alam, S.; Khan, F., Virtual screening, Docking, ADMET and System Pharmacology studies on Garcinia caged Xanthone derivatives for Anticancer activity. *Sci Rep* **2018,** *8* (1), 5524.
91. Roncaglioni, A.; Piclin, N.; Pintore, M.; Benfenati, E., Binary classification models for endocrine disrupter effects mediated through the estrogen receptor. *SAR and QSAR in Environmental Research* **2008,** *19* (7-8), 697-733.
92. Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y., Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *Journal of computer-aided molecular design* **2011,** *25* (6), 533-554.

93. Hecker, M.; Hollert, H., Endocrine disruptor screening: regulatory perspectives and needs. *Environmental Sciences Europe* **2011,** *23* (1), 15.
94. Mansouri, K.; Abdelaziz, A.; Rybacka, A.; Roncaglioni, A.; Tropsha, A.; Varnek, A.; Zakharov, A.; Worth, A.; Richard, A. M.; Grulke, C. M., CERAPP: collaborative estrogen receptor activity prediction project. *Environmental health perspectives* **2016,** *124* (7), 1023-1033.
95. Lavecchia, A.; Cerchia, C., In silico methods to address polypharmacology: current status, applications and future perspectives. *Drug Discov Today* **2016,** *21* (2), 288-98.
96. Ellingson, S. R.; Miao, Y.; Baudry, J.; Smith, J. C., Multi-conformer ensemble docking to difficult protein targets. *J Phys Chem B* **2015,** *119* (3), 1026-34.
97. Tian, S.; Sun, H.; Pan, P.; Li, D.; Zhen, X.; Li, Y.; Hou, T., Assessing an ensemble docking-based virtual screening strategy for kinase targets by considering protein flexibility. *J Chem Inf Model* **2014,** *54* (10), 2664-79.
98. Park, S.-J.; Kufareva, I.; Abagyan, R., Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles. *J Comput Aided Mol Des* **2010,** *24*, 459–471.
99. Zhang, Q.; Ye, M., Chemical analysis of the Chinese herbal medicine Gan-Cao (licorice). *J Chromatogr A* **2009,** *1216* (11), 1954-69.
100. Li, G.; Nikolic, D.; van Breemen, R. B., Identification and Chemical Standardization of Licorice Raw Materials and Dietary Supplements Using UHPLC-MS/MS. *J Agric Food Chem* **2016**.
101. Jiang, Y.; Gong, P.; Madak-Erdogan, Z.; Martin, T.; Jeyakumar, M.; Carlson, K.; Khan, I.; Smillie, T. J.; Chittiboyina, A. G.; Rotte, S. C., Mechanisms enforcing the estrogen receptor β selectivity of botanical estrogens. *The FASEB Journal* **2013,** *27* (11), 4406-4418.
102. Mersereau, J. E.; Levy, N.; Staub, R. E.; Baggett, S.; Zogric, T.; Chow, S.; Ricke, W. A.; Tagliaferri, M.; Cohen, I.; Bjeldanes, L. F., Liquiritigenin is a plant-derived highly selective estrogen receptor β agonist. *Molecular and cellular endocrinology* **2008,** *283* (1-2), 49-57.
103. Yamamoto, T.; Sakamoto, C.; Tachiwana, H.; Kumabe, M.; Matsui, T.; Yamashita, T.; Shinagawa, M.; Ochiai, K.; Saitoh, N.; Nakao, M., Endocrine therapy-resistant breast cancer model cells are inhibited by soybean glyceollin I through Eleanor non-coding RNA. *Sci Rep* **2018,** *8* (1), 15202.
104. Chen, Q.; Tan, H.; Yu, H.; Shi, W., Activation of steroid hormone receptors: Shed light on the in silico evaluation of endocrine disrupting chemicals. *Science of The Total Environment* **2018,** *631*, 27-39.
105. *ADMET Predictor*, 9; Simulations Plus, Inc.: Lancaster, CA, USA.
106. Schaefer, O.; Hümpel, M.; Fritzemeier, K.-H.; Bohlmann, R.; Schleuning, W.-D., 8-Prenyl naringenin is a potent ERα selective phytoestrogen present in hops and beer. *The Journal of Steroid Biochemistry and Molecular Biology* **2003,** *84* (2-3), 359-360.
107. Muthyala, R. S.; Ju, Y. H.; Sheng, S.; Williams, L. D.; Doerge, D. R.; Katzenellenbogen, B. S.; Helferich, W. G.; Katzenellenbogen, J. A., Equol, a natural estrogenic metabolite from soy isoflavones: convenient preparation and resolution of R- and S-equols and their differing binding and biological activity through estrogen receptors alpha and beta. *Bioorg Med Chem* **2004,** *12* (6), 1559-67.
108. Nomura, T.; Fukai, T.; Akiyama, T., Chemistry of phenolic compounds of licorice

(Glycyrrhiza species) and their estrogenic and cytotoxic activities. *Pure Appl. Chem.* **2002,** *74*, 1199–1206.
109. Kretzschmar, G.; Zierau, O.; Wober, J.; Tischer, S.; Metz, P.; Vollmer, G., Prenylation has a compound specific effect on the estrogenicity of naringenin and genistein. *J Steroid Biochem Mol Biol* **2010,** *118* (1-2), 1-6.
110. Yap, S. P.; Shen, P.; Butler, M. S.; Gong, Y.; Loy, C. J.; Yong, E. L., New estrogenic prenylflavone from Epimedium brevicornum inhibits the growth of breast cancer cells. *Planta Med* **2005,** *71* (2), 114-9.
111. Walle, T., Absorption and metabolism of flavonoids. *Free Radic Biol Med* **2004,** *36* (7), 829-37.
112. van de Schans, M. G.; Bovee, T. F.; Stoopen, G. M.; Lorist, M.; Gruppen, H.; Vincken, J. P., Prenylation and Backbone Structure of Flavonoids and Isoflavonoids from Licorice and Hop Influence Their Phase I and II Metabolism. *J Agric Food Chem* **2015,** *63* (49), 10628-40.
113. Hwang, C. S.; Kwak, H. S.; Lim, H. J.; Lee, S. H.; Kang, Y. S.; Choe, T. B.; Hur, H. G.; Han, K. O., Isoflavone metabolites and their in vitro dual functions: they can act as an estrogenic agonist or antagonist depending on the estrogen concentration. *J Steroid Biochem Mol Biol* **2006,** *101* (4-5), 246-53.
114. Pinto, C. L.; Mansouri, K.; Judson, R.; Browne, P., Prediction of Estrogenic Bioactivity of Environmental Chemical Metabolites. *Chem Res Toxicol* **2016,** *29* (9), 1410-27.
115. Chapman, H., Dictionary of Natural Products on DVD (23: 1). CRC Press, Taylor & Francis Group, URL: http://dnp. chemnetbase. com: 2014.
116. Wärnmark, A.; Treuter, E.; Gustafsson, J.-Å.; Hubbard, R. E.; Brzozowski, A. M.; Pike, A. C., Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor α. *Journal of Biological Chemistry* **2002,** *277* (24), 21862-21868.
117. Fuchs, S.; Nguyen, H. D.; Phan, T. T.; Burton, M. F.; Nieto, L.; de Vries-van Leeuwen, I. J.; Schmidt, A.; Goodarzifard, M.; Agten, S. M.; Rose, R., Proline primed helix length as a modulator of the nuclear receptor–coactivator interaction. *Journal of the American Chemical Society* **2013,** *135* (11), 4364-4371.
118. Shiau, A. K.; Barstad, D.; Radek, J. T.; Meyers, M. J.; Nettles, K. W.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A.; Agard, D. A.; Greene, G. L., Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. *Nature Structural and Molecular Biology* **2002,** *9* (5), 359.
119. Manas, E. S.; Xu, Z. B.; Unwalla, R. J.; Somers, W. S., Understanding the selectivity of genistein for human estrogen receptor-β using X-ray crystallography and computational methods. *Structure* **2004,** *12* (12), 2197-2207.
120. *Schrödinger Release 2015-3: Maestro*, 2015-3; Schrödinger, LLC: New York, NY, 2015.
121. Šmejkal, K., Cytotoxic potential of C-prenylated flavonoids. *Phytochemistry Reviews* **2013,** *13* (1), 245-275.
122. Vickery, C. R.; La Clair, J. J.; Burkart, M. D.; Noel, J. P., Harvesting the biosynthetic machineries that cultivate a variety of indispensable plant natural products. *Curr Opin Chem Biol* **2016,** *31*, 66-73.
123. Kagiyama, I.; Kato, H.; Nehira, T.; Frisvad, J. C.; Sherman, D. H.; Williams, R. M.; Tsukamoto, S., Taichunamides: Prenylated Indole Alkaloids from Aspergillus taichungensis (IBT 19404). *Angew Chem Int Ed Engl* **2016,** *55* (3), 1128-32.

124. Akinwumi, B. C.; Bordun, K. M.; Anderson, H. D., Biological Activities of Stilbenoids. *Int J Mol Sci* **2018,** *19* (3).
125. Chrząścik, I., Analysis of Biologically Active Stilbene Derivatives. *Critical Reviews in Analytical Chemistry* **2009,** *39* (2), 70-80.
126. Niesen, D. B.; Hessler, C.; Seeram, N. P., Beyond resveratrol: A review of natural stilbenoids identified from 2009–2013. *J. Berry Res.* **2013,** *3*, 181–196.
127. Helle, J.; Kraker, K.; Bader, M. I.; Keiler, A. M.; Zierau, O.; Vollmer, G.; Welsh, J.; Kretzschmar, G., Assessment of the proliferative capacity of the flavanones 8-prenylnaringenin, 6-(1.1-dimethylallyl)naringenin and naringenin in MCF-7 cells and the rat mammary gland. *Mol Cell Endocrinol* **2014,** *392* (1-2), 125-35.
128. Gardner, K. D.; Wiemer, D. F., Selective Prenylation of Protected Phenols for Synthesis of Pawhuskin A Analogues. *J Org Chem* **2016,** *81* (4), 1585-92.
129. Nicolaou, K.; Lister, T.; Denton, R. M.; Gelin, C. F., Total synthesis of artochamins F, H, I, and J through cascade reactions. *Tetrahedron* **2008,** *64* (21), 4736-4757.
130. Park, B. H.; Lee, H. J.; Lee, Y. R., Total synthesis of chiricanine A, arahypin-1, trans-arachidin-2, trans-arachidin-3, and arahypin-5 from peanut seeds. *J Nat Prod* **2011,** *74* (4), 644-9.
131. Choi, J. G.; Eom, S. M.; Kim, J.; Kim, S. H.; Huh, E.; Kim, H.; Lee, Y.; Lee, H.; Oh, M. S., A Comprehensive Review of Recent Studies on Herb-Drug Interaction: A Focus on Pharmacodynamic Interaction. *J Altern Complement Med* **2016,** *22* (4), 262-79.
132. Yuan, Y.; Yang, H.; Kong, L.; Li, Y.; Li, P.; Zhang, H.; Ruan, J., Interaction between rhein acyl glucuronide and methotrexate based on human organic anion transporters. *Chem Biol Interact* **2017,** *277*, 79-84.
133. Alsanad, S. M.; Williamson, E. M.; Howard, R. L., Cancer patients at risk of herb/food supplement-drug interactions: a systematic review. *Phytother Res* **2014,** *28* (12), 1749-55.
134. Mouly, S.; Lloret-Linares, C.; Sellier, P. O.; Sene, D.; Bergmann, J. F., Is the clinical relevance of drug-food and drug-herb interactions limited to grapefruit juice and Saint-John's Wort? *Pharmacol Res* **2017,** *118*, 82-92.
135. Ghosh, N.; Ghosh, R. C.; Kundu, A.; Mandal, S. C., Herb and Drug Interaction. **2018**, 467-490.
136. Cui, Z.; Kang, H.; Tang, K.; Liu, Q.; Cao, Z.; Zhu, R., Screening Ingredients from Herbs against Pregnane X Receptor in the Study of Inductive Herb-Drug Interactions: Combining Pharmacophore and Docking-Based Rank Aggregation. *Biomed Res Int* **2015,** *2015*, 657159.
137. Tirona, R. G.; Kim, R. B., Nuclear receptors and drug disposition gene regulation. *J Pharm Sci* **2005,** *94* (6), 1169-86.
138. Brewer, C. T.; Chen, T., PXR variants: the impact on drug metabolism and therapeutic responses. *Acta Pharm Sin B* **2016,** *6* (5), 441-449.
139. Xu, C.; Huang, M.; Bi, H., PXR- and CAR-mediated herbal effect on human diseases. *Biochim Biophys Acta* **2016,** *1859* (9), 1121-1129.
140. Yan, J.; Xie, W., A brief history of the discovery of PXR and CAR as xenobiotic receptors. *Acta Pharm Sin B* **2016,** *6* (5), 450-452.
141. Zhou, C., Novel functions of PXR in cardiometabolic disease. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **2016,** *1859* (9), 1112-1120.
142. Banerjee, M.; Robbins, D.; Chen, T., Targeting xenobiotic receptors PXR and CAR in human diseases. *Drug discovery today* **2015,** *20* (5), 618-628.

143. Wei, Y.; Tang, C.; Sant, V.; Li, S.; Poloyac, S. M.; Xie, W., A Molecular Aspect in the Regulation of Drug Metabolism: Does PXR-Induced Enzyme Expression Always Lead to Functional Changes in Drug Metabolism? *Curr Pharmacol Rep* **2016,** *2* (4), 187-192.
144. Hassani-Nezhad-Gashti, F.; Rysä, J.; Kummu, O.; Näpänkangas, J.; Buler, M.; Karpale, M.; Hukkanen, J.; Hakkola, J., Activation of nuclear receptor PXR impairs glucose tolerance and dysregulates GLUT2 expression and subcellular localization in liver. *Biochemical pharmacology* **2018,** *148*, 253-264.
145. Chai, S. C.; Cherian, M. T.; Wang, Y.-M.; Chen, T., Small-molecule modulators of PXR and CAR. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **2016,** *1859* (9), 1141-1154.
146. Watkins, R. E.; Maglich, J. M.; Moore, L. B.; Wisely, G. B.; Noble, S. M.; Davis-Searles, P. R.; Lambert, M. H.; Kliewer, S. A.; Redinbo, M. R., 2.1 Å Crystal Structure of Human PXR in Complex with the St. John's Wort Compound Hyperforin. *Biochemistry* **2003,** *42* (6), 1430-1438.
147. Khan, J. A.; Camac, D. M.; Low, S.; Tebben, A. J.; Wensel, D. L.; Wright, M. C.; Su, J.; Jenny, V.; Gupta, R. D.; Ruzanov, M., Developing adnectins that target SRC co-activator binding to PXR: a structural approach toward understanding promiscuity of PXR. *Journal of molecular biology* **2015,** *427* (4), 924-942.
148. Motta, S.; Callea, L.; Tagliabue, S. G.; Bonati, L., Exploring the PXR ligand binding mechanism with advanced Molecular Dynamics methods. *Scientific reports* **2018,** *8* (1), 16207.
149. Watkins, R. E.; Davis-Searles, P. R.; Lambert, M. H.; Redinbo, M. R., Coactivator Binding Promotes the Specific Interaction Between Ligand and the Pregnane X Receptor. *Journal of Molecular Biology* **2003,** *331* (4), 815-828.
150. Ngan, C.-H.; Beglov, D.; Rudnitskaya, A. N.; Kozakov, D.; Waxman, D. J.; Vajda, S., The structural basis of pregnane X receptor binding promiscuity. *Biochemistry* **2009,** *48* (48), 11572-11581.
151. Honkakoski, P.; Sueyoshi, T.; Negishi, M., Drug-activated nuclear receptors CAR and PXR. *Annals of medicine* **2003,** *35* (3), 172-182.
152. Östberg, T.; Bertilsson, G.; Jendeberg, L.; Berkenstam, A.; Uppenberg, J., Identification of residues in the PXR ligand binding domain critical for species specific and constitutive activation. *European journal of biochemistry* **2002,** *269* (19), 4896-4904.
153. Banerjee, M.; Chai, S. C.; Wu, J.; Robbins, D.; Chen, T., Tryptophan 299 is a conserved residue of human pregnane X receptor critical for the functional consequence of ligand binding. *Biochemical pharmacology* **2016,** *104*, 131-138.
154. Ekins, S.; Chang, C.; Mani, S.; Krasowski, M. D.; Reschly, E. J.; Iyer, M.; Kholodovych, V.; Ai, N.; Welsh, W. J.; Sinz, M.; Swaan, P. W.; Patel, R.; Bachmann, K., Human pregnane X receptor antagonists and agonists define molecular requirements for different binding sites. *Mol Pharmacol* **2007,** *72* (3), 592-603.
155. Gao, Y. D.; Olson, S. H.; Balkovec, J. M.; Zhu, Y.; Royo, I.; Yabut, J.; Evers, R.; Tan, E. Y.; Tang, W.; Hartley, D. P.; Mosley, R. T., Attenuating pregnane X receptor (PXR) activation: a molecular modelling approach. *Xenobiotica* **2007,** *37* (2), 124-38.
156. Wang, H.; Li, H.; Moore, L. B.; Johnson, M. D.; Maglich, J. M.; Goodwin, B.; Ittoop, O. R.; Wisely, B.; Creech, K.; Parks, D. J.; Collins, J. L.; Willson, T. M.; Kalpana, G. V.; Venkatesh, M.; Xie, W.; Cho, S. Y.; Roboz, J.; Redinbo, M.; Moore, J. T.; Mani, S., The phytoestrogen coumestrol is a naturally occurring antagonist of the human pregnane X

receptor. *Mol Endocrinol* **2008,** *22* (4), 838-57.
157. Smutny, T.; Pavek, P., Resveratrol as an Inhibitor of Pregnane X Receptor (PXR): Another Lesson in PXR Antagonism. *Journal of Pharmacological Sciences* **2014,** *126* (2), 177-178.
158. Xiao, L.; Nickbarg, E.; Wang, W.; Thomas, A.; Ziebell, M.; Prosise, W. W.; Lesburg, C. A.; Taremi, S. S.; Gerlach, V. L.; Le, H. V.; Cheng, K. C., Evaluation of in vitro PXR-based assays and in silico modeling approaches for understanding the binding of a structurally diverse set of drugs to PXR. *Biochem Pharmacol* **2011,** *81* (5), 669-79.
159. Matter, H.; Anger, L. T.; Giegerich, C.; Güssregen, S.; Hessler, G.; Baringhaus, K.-H., Development of in silico filters to predict activation of the pregnane X receptor (PXR) by structurally diverse drug-like molecules. *Bioorganic & medicinal chemistry* **2012,** *20* (18), 5352-5365.
160. Kortagere, S.; Krasowski, M. D.; Reschly, E. J.; Venkatesh, M.; Mani, S.; Ekins, S., Evaluation of computational docking to identify pregnane X receptor agonists in the ToxCast database. *Environ Health Perspect* **2010,** *118* (10), 1412-7.
161. Han, L.; Ma, X.; Lin, H.; Jia, J.; Zhu, F.; Xue, Y.; Li, Z.; Cao, Z.; Ji, Z.; Chen, Y., A support vector machines approach for virtual screening of active compounds of single and multiple mechanisms from large libraries at an improved hit-rate and enrichment factor. *Journal of Molecular Graphics and Modelling* **2008,** *26* (8), 1276-1286.
162. Chen, C.-N.; Shih, Y.-H.; Ding, Y.-L.; Leong, M. K., Predicting activation of the promiscuous human pregnane X receptor by pharmacophore ensemble/support vector machine approach. *Chemical research in toxicology* **2011,** *24* (10), 1765-1778.
163. Yang, R.; Wang, L.-q.; Yuan, B.-c.; Liu, Y., The pharmacological activities of licorice. *Planta medica* **2015,** *81* (18), 1654-1669.
164. Feng, X.; Ding, L.; Qiu, F., Potential drug interactions associated with glycyrrhizin and glycyrrhetinic acid. *Drug metabolism reviews* **2015,** *47* (2), 229-238.
165. Wang, Y.-G.; Zhou, J.-M.; Ma, Z.-C.; Li, H.; Liang, Q.-D.; Tan, H.-L.; Xiao, C.-R.; Zhang, B.-L.; Gao, Y., Pregnane X receptor mediated-transcription regulation of CYP3A by glycyrrhizin: a possible mechanism for its hepatoprotective property against lithocholic acid-induced injury. *Chemico-biological interactions* **2012,** *200* (1), 11-20.
166. Johnson, E. J.; Gonzalez-Perez, V.; Tian, D.-D.; Lin, Y. S.; Unadkat, J. D.; Rettie, A. E.; Shen, D. D.; McCune, J. S.; Paine, M. F., Selection of Priority Natural Products for Evaluation as Potential Precipitants of Natural Product–Drug Interactions: A NaPDI Center Recommended Approach. *Drug Metabolism and Disposition* **2018,** *46* (7), 1046-1052.
167. Jacobs, M. N.; Nolan, G. T.; Hood, S. R., Lignans, bacteriocides and organochlorine compounds activate the human pregnane X receptor (PXR). *Toxicol Appl Pharmacol* **2005,** *209* (2), 123-33.
168. Manda, V. K.; Ibrahim, M. A.; Dale, O. R.; Kumarihamy, M.; Cutler, S. J.; Khan, I. A.; Walker, L. A.; Muhammad, I.; Khan, S. I., Modulation of CYPs, P-gp, and PXR by Eschscholzia californica (California poppy) and its alkaloids. *Planta medica* **2016,** *82* (06), 551-558.
169. Fukai, T.; Cai, B.-S.; Maruno, K.; Miyakawa, Y.; Konishi, M.; Nomura, T., "Phenolic constituents of Glycyrrhiza species". An isoprenylated flavanone from Glycyrrhiza glabra and rec-assay of licorice phenols. *Phytochemistry* **1998,** *49* (7), 2005-2013.
170. Li, K.; Ji, S.; Song, W.; Kuang, Y.; Lin, Y.; Tang, S.; Cui, Z.; Qiao, X.; Yu, S.; Ye, M.,

Glycybridins A-K, bioactive phenolic compounds from Glycyrrhiza glabra. *J. Nat. Prod.* **2017,** *80* (2), 334-346.
171. Kitagawa, I.; Chen, W. Z.; Hori, K.; Harada, E.; Yasuda, N.; Yoshikawa, M.; Ren, J., Chemical Studies of Chinese Licorice-Roots. I. Elucidation of five new flavonoid constituents from the roots of Glycyrrhiza glabra L. collected in Xinjiang. *Chem. Pharm. Bull.* **1994,** *42* (5), 1056-62.
172. Bhardwaj, D. K.; Murari, R.; Seshadri, T. R.; Singh, R., Liqcoumarin, a novel coumarin from Glycyrrhiza glabra. *Phytochemistry* **1976,** *15* (7), 1182-3.
173. Siracusa, L.; Saija, A.; Cristani, M.; Cimino, F.; D'Arrigo, M.; Trombetta, D.; Rao, F.; Ruberto, G., Phytocomplexes from liquorice (Glycyrrhiza glabra L.) leaves - Chemical characterization and evaluation of their antioxidant, anti-genotoxic and anti-inflammatory activity. *Fitoterapia* **2011,** *82* (4), 546-556.
174. Denisova, S.; Galkin, E.; Murinov, Y. I., Isolation and GC-MS determination of flavonoids from Glycyrrhiza glabra root. *Chemistry of natural compounds* **2006,** *42* (3), 285-289.
175. Canonica, L.; Danielli, B.; Russo, G.; Bonati, A., Triterpenes from Glycyrrhiza glabra IV. 18α-Hydroxyglycyrrhetic acid. *Gazz Chim Ital* **1967,** *97*, 769-786.
176. Elgamal, M.; Fayez, M., Structure of glabric acid A: further triterpenoid constituent of glycyrrhiza glabra L. *Acta Chimica Academiae Scientarium Hungaricae* **1968,** *58* (1), 75.
177. Canonica, L.; Danielli, B.; Manitto, P.; Russo, G.; Bombardelli, E.; Bonati, A., Glycyrrhiza glabra triterpenes VIL 24-Hydroxyliquiritic acid (30, 24-dihydroxy-11-oxo-olean-12-en-29-oic acid) and liquiridolic acid (3β, 21α, 24-trihydroxy-olean-12-en-29-oic acid). *Gazz Chim Ital* **1968,** *98*, 712-728.
178. Canonica, L.; Danieli, B.; Manitto, P.; Russo, G., Triterpenes of Glycyrrhiza glabra. III. Structure of isogabrolide. *Gazzetta Chimica Italiana* **1966,** *96* (6), 843-51.
179. Wang, Y.; Yang, L.; He, Y. Q.; Wang, C. H.; Welbeck, E. W.; Bligh, S. A.; Wang, Z. T., Characterization of fifty-one flavonoids in a Chinese herbal prescription Longdan Xiegan Decoction by high-performance liquid chromatography coupled to electrospray ionization tandem mass spectrometry and photodiode array detection. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2008,** *22* (12), 1767-1778.
180. Li, J.-R.; Wang, Y.-Q.; Deng, Z.-Z., Two new compounds from Glycyrrhiza glabra. *J. Asian Nat. Prod. Res.* **2005,** *7* (4), 677-680.
181. Biondi, D. M.; Rocco, C.; Ruberto, G., Dihydrostilbene derivatives from Glycyrrhiza g labra leaves. *Journal of natural products* **2005,** *68* (7), 1099-1102.
182. Asada, Y.; Li, W.; Yoshikawa, T., The first prenylated biaurone, licoagrone from hairy root cultures of Glycyrrhiza glabra. *Phytochemistry* **1999,** *50* (6), 1015-1019.
183. Hayashi, H.; Yasuma, M.; Hiraoka, N.; Ikeshiro, Y.; Yamamoto, H.; Yeşilada, E.; Sezik, E.; Honda, G.; Tabata, M., Flavonoid variation in the leaves of Glycyrrhiza glabra. *Phytochemistry* **1996,** *42* (3), 701-704.
184. Hayashi, H.; Hanaoka, S.; Tanaka, S.; Fukui, H.; Tabata, M., Glycyrrhetinic acid 24-hydroxylase activity in microsomes of cultured licorice cells. *Phytochemistry* **1993,** *34* (5), 1303-1307.
185. Elgamal, M. H. A.; El-Tawil, B. A. H., Constituents of local plants. XVIII. 28-Hydroxyglycyrrhetic acid, a new triterpenoid isolated from the roots of Glycyrrhiza glabra. *Planta Med.* **1975,** *27* (2), 159-63.

186. Ofir, R.; Tamir, S.; Khatib, S.; Vaya, J., Inhibition of serotonin re-uptake by licorice constituents. *Journal of molecular neuroscience* **2003,** *20* (2), 135-140.
187. Chin, Y.-W.; Jung, H.-A.; Liu, Y.; Su, B.-N.; Castoro, J. A.; Keller, W. J.; Pereira, M. A.; Kinghorn, A. D., Anti-oxidant Constituents of the Roots and Stolons of Licorice (Glycyrrhiza glabra). *J. Agric. Food Chem.* **2007,** *55* (12), 4691-4697.
188. Kuroda, M.; Mimaki, Y.; Honda, S.; Tanaka, H.; Yokota, S.; Mae, T., Phenolics from Glycyrrhiza glabra roots and their PPAR-gamma ligand-binding activity. *Bioorg Med Chem* **2010,** *18* (2), 962-70.
189. Suman, A.; Ali, M.; Alam, P., New prenylated isoflavanones from the roots of Glycyrrhiza glabra. *Chem. Nat. Compd.* **2009,** *45* (4), 487-491.
190. Li, W.; Asada, Y.; Yoshikawa, T., Flavonoid constituents from Glycyrrhiza glabra hairy root cultures. *Phytochemistry* **2000,** *55* (5), 447-456.
191. Bogatkina, V. F.; Murav'ev, I. A.; Stepanova, E. F.; Kir'yalov, N. P., Triterpine compounds from the above ground part of a licorice plant *Khimiya Prirodnykh Soedinenii* **1975,** *11* (1), 101-102.
192. Bharadwaj, D. K.; Murari, R.; Seshadri, T. R.; Singh, R., Occurrence of 2-methylisoflavones in Glycyrrhiza glabra. *Phytochemistry* **1976**.
193. Mitscher, L. A.; Park, Y. H.; Clark, D.; Beal, J. L., Antimicrobial agents from higher plants. Antimicrobial isoflavanoids and related substances from Glycyrrhiza glabra L. var. typica. *J. Nat. Prod.* **1980,** *43* (2), 259-69.
194. Fukai, T.; Tantai, L.; Nomura, T., Isoprenoid-substituted flavonoids from Glycyrrhiza glabra. *Phytochemistry* **1996,** *43* (2), 531-532.
195. Montoro, P.; Maldini, M.; Russo, M.; Postorino, S.; Piacente, S.; Pizza, C., Metabolic profiling of roots of liquorice (Glycyrrhiza glabra) from different geographical areas by ESI/MS/MS and determination of major metabolites by LC-ESI/MS and LC-ESI/MS/MS. *Journal of pharmaceutical and biomedical analysis* **2011,** *54* (3), 535-544.
196. Canonica, L.; Russo, G.; Bonati, A., Triterpenes of Glycyrrhiza glabra. I. Two new lactones with an oleanane structure. *Gazz Chim Ital* **1966,** *96*, 772-785.
197. Kim, H. K.; Park, Y.; Kim, H. N.; Choi, B. H.; Jeong, H. G.; Lee, D. G.; Hahm, K.-S., Antimicrobial mechanism of β-glycyrrhetinic acid isolated from licorice, Glycyrrhiza glabra. *Biotechnology letters* **2002,** *24* (22), 1899-1902.
198. Li, W.; Asada, Y.; Yoshikawa, T., Antimicrobial flavonoids from Glycyrrhiza glabra hairy root cultures. *Planta Med* **1998,** *64* (8), 746-7.
199. Chen, J.-J.; Cheng, M.-J.; Shu, C.-W.; Sung, P.-J.; Lim, Y.-P.; Cheng, L.-Y.; Wang, S.-L.; Chen, L.-C., A New Chalcone and Antioxidant Constituents of Glycyrrhiza glabra. *Chem. Nat. Compd.* **2017,** *53* (4), 632-634.
200. Sadaf, N.; Erum, S.; Khalida, B.; Barlow, D., Ligand-based screening of chemical constituents of Glycyrrhiza glabra in search of inhibitors of xanthine oxidase. *International Journal of Biology and Biotechnology* **2012,** *9* (3), 257-262.
201. Litvinenko, V.; Nadezhina, T., flavonoids of the above-ground part of Glycyrrhiza glabra L. *Rastitel Resursy* **1972**.
202. Bhardwaj, D.; Singh, R., 'Glyzaglabrin', a new isoflavone from Glycyrrhiza glabra. *Current science* **1977**.
203. Biondi, D. M.; Rocco, C.; Ruberto, G., New Dihydrostilbene Derivatives from the Leaves of Glycyrrhiza g labra and Evaluation of Their Antioxidant Activity. *Journal of Natural Products* **2003,** *66* (4), 477-480.

204. Canonica, L.; Danielli, B.; Manitto, P.; Russo, G.; Bombardelli, E., Glycyrrhiza glabra triterpenes V. Glycyrrhetol and 21α-hydroxyisoglabrolide. *Gazz Chim Ital* **1967,** *97*, 1347-1358.
205. Elgamal, M.; Hady, F. A.; Hanna, A.; Mahran, G.; Duddeck, H., A further contribution to the triterpenoid constituents of Glycyrrhiza glabra L. *Zeitschrift für Naturforschung C* **1990,** *45* (9-10), 937-941.
206. Birari, R. B.; Gupta, S.; Mohan, C. G.; Bhutani, K. K., Antiobesity and lipid lowering effects of Glycyrrhiza chalcones: experimental and computational studies. *Phytomedicine* **2011,** *18* (8-9), 795-801.
207. Kinoshita, T.; Tamura, Y.; Mizutani, K., The isolation and structure elucidation of minor isoflavonoids from licorice of Glycyrrhiza glabra origin. *Chem. Pharm. Bull.* **2005,** *53* (7), 847-849.
208. Rafi, M. M.; Vastano, B. C.; Zhu, N.; Ho, C.-T.; Ghai, G.; Rosen, R. T.; Gallo, M. A.; DiPaola, R. S., Novel polyphenol molecule isolated from licorice root (Glycrrhiza glabra) induces apoptosis, G2/M cell cycle arrest, and Bcl-2 phosphorylation in tumor cell lines. *Journal of agricultural and food chemistry* **2002,** *50* (4), 677-684.
209. Hoton-Dorge, M., Identification of some flavonoid aglycone extracts of Glycyrrhiza glabra roots *Journal de Pharmacie de Belgique* **1974,** *29* (6), 560-72.
210. Simmler, C.; Anderson, J. R.; Gauthier, L.; Lankin, D. C.; McAlpine, J. B.; Chen, S.-N.; Pauli, G. F., Metabolite profiling and classification of DNA-authenticated licorice botanicals. *Journal of natural products* **2015,** *78* (8), 2007-2022.
211. Ji, S.; Li, Z.; Song, W.; Wang, Y.; Liang, W.; Li, K.; Tang, S.; Wang, Q.; Qiao, X.; Zhou, D.; Yu, S.; Ye, M., Bioactive Constituents of Glycyrrhiza uralensis (Licorice): Discovery of the Effective Components of a Traditional Herbal Medicine. *J Nat Prod* **2016,** *79* (2), 281-92.
212. Kuroda, M.; Mimaki, Y.; Honda, S.; Tanaka, H.; Yokota, S.; Mae, T., Phenolics from Glycyrrhiza glabra roots and their PPAR-γ ligand-binding activity. *Bioorg. Med. Chem.* **2010,** *18* (2), 962-970.
213. Shul'ts, E.; Petrova, T.; Shakirov, M.; Chernyak, E.; Tolstikov, G., Flavonoids of Roots ofGlycyrrhiza uralensisGrowing in Siberia. *Chemistry of Natural Compounds* **2000,** *36* (4), 362-368.
214. Wei, G. H.; Yang, X. Y.; Zhang, J. W.; Gao, J. M.; Ma, Y. Q.; Fu, Y. Y.; Wang, P., Rhizobialide: a new stearolactone produced by Mesorhizobium sp. CCNWGX022, a rhizobial endophyte from Glycyrrhiza uralensis. *Chemistry & biodiversity* **2007,** *4* (5), 893-898.
215. Li, S.; Li, W.; Wang, Y.; Asada, Y.; Koike, K., Prenylflavonoids from Glycyrrhiza uralensis and their protein tyrosine phosphatase-1B inhibitory activities. *Bioorg. Med. Chem. Lett.* **2010,** *20* (18), 5398-5401.
216. Ryu, Y. B.; Kim, J. H.; Park, S.-J.; Chang, J. S.; Rho, M.-C.; Bae, K.-H.; Park, K. H.; Lee, W. S., Inhibition of neuraminidase activity by polyphenol compounds isolated from the roots of Glycyrrhiza uralensis. *Bioorg. Med. Chem. Lett.* **2010,** *20* (3), 971-974.
217. Jia, S.; Ma, C.; Li, Y.; Hao, J., Glycosides of phenolic acid and flavonoids from the leaves of Glycyrrhiza uralensis Ficsh. *Yao xue xue bao= Acta pharmaceutica Sinica* **1992,** *27* (6), 441-444.
218. Han, Y. N.; Chung, M. S., A pyrrolo-pyrimidine alkaloid fromGlycyrrhiza uralensis. *Archives of Pharmacal Research* **1990,** *13* (1), 103-104.

219. Hatano, T.; Takagi, M.; Ito, H.; Yoshida, T., Acylated flavonoid glycosides and accompanying phenolics from licorice. *Phytochemistry* **1998,** *47* (2), 287-293.
220. Yahara, S.; Nishioka, I., Flavonoid glucosides from licorice. *Phytochemistry* **1984,** *23* (9), 2108-9.
221. KITAGAWA, I.; ZHOU, J. L.; SAKAGAMI, M.; UCHIDA, E.; YOSHIKAWA, M., Licorice-saponins F3, G2, H2, J2, and K2, five new oleanene-triterpene oligoglycosides from the root of Glycyrrhiza uralensis. *Chemical and pharmaceutical bulletin* **1991,** *39* (1), 244-246.
222. Gafner, S.; Bergeron, C.; Villinski, J. R.; Godejohann, M.; Kessler, P.; Cardellina, J. H.; Ferreira, D.; Feghali, K.; Grenier, D., Isoflavonoids and coumarins from Glycyrrhiza uralensis: antibacterial activity against oral pathogens and conversion of isoflavans into isoflavan-quinones during purification. *Journal of natural products* **2011,** *74* (12), 2514-2519.
223. Fukai, T.; WANG, Q.-H.; Nomura, T., Four new prenylated flavonoids from aerial parts of Glycyrrhiza uralensis. *Heterocycles* **1989,** *29* (7), 1369-1378.
224. Zheng, Y.-F.; Qi, L.-W.; Cui, X.-B.; Peng, G.-P.; Peng, Y.-B.; Ren, M.-T.; Cheng, X.-L.; Li, P., Oleanane-type triterpene glucuronides from the roots of Glycyrrhiza uralensis Fischer. *Planta medica* **2010,** *76* (13), 1457-1463.
225. Zhu, D.; Song, G.; Jian, F.; Chang, X.; Guo, W., Chemical constituents of Glycyrrhiza uralensis Fisch - structures of isolicoflavonol and glycycoumarin. *Huaxue Xuebao* **1984,** *42* (10), 1080-4.
226. Shu, Y.; Zhao, Y.; Zhang, R., Isolation and structural identification of triterpene sapogenins from Glycyrrhiza uralensis Fisch. *Yaoxue Xuebao* **1985,** *20* (3), 193-7.
227. Gao, C.; Qiao, L.; Jia, Q.; Zhang, Z., Study and identification of the chemical structure of a new component in Glycyrrhiza uralensis Fisch by NMR method. *Bopuxue Zazhi* **1990,** *7* (1), 11-16.
228. Fukai, T.; Nishizawa, J.; Yokoyama, M.; Nomura, T., Phenolic constituents of Glycyrrhiza species. Five new isoprenoid-substituted flavonoids, kanzonols F-J, from Glycyrrhiza uralensis. *Heterocycles* **1993,** *36* (11), 2565-76.
229. Song, W.; Si, L.; Ji, S.; Wang, H.; Fang, X. M.; Yu, L. Y.; Li, R. Y.; Liang, L. N.; Zhou, D.; Ye, M., Uralsaponins M-Y, antiviral triterpenoid saponins from the roots of Glycyrrhiza uralensis. *J Nat Prod* **2014,** *77* (7), 1632-43.
230. Kitagawa, I.; Hori, K.; Uchida, E.; Chen, W. Z.; Yoshikawa, M.; Ren, J., Saponin and sapogenol. L. On the constituents of the roots of Glycyrrhiza uralensis Fischer from Xinjiang, China. Chemical structures of licorice-saponin L3 and isoliquiritin apioside. *Chem. Pharm. Bull.* **1993,** *41* (9), 1567-72.
231. Fukai, T.; Nishizawa, J.; Nomura, T., Variations in the chemical shift of the 5-hydroxyl proton of isoflavones; two isoflavones from licorice. *Phytochemistry* **1994,** *36* (1), 225-228.
232. He, J.; Chen, L.; Heber, D.; Shi, W.; Lu, Q.-Y., Antibacterial Compounds from Glycyrrhiza u ralensis. *Journal of natural products* **2006,** *69* (1), 121-124.
233. Shibano, M.; Henmi, A.; Matsumoto, Y.; Kusano, G.; Miyase, T.; Hatakeyama, Y., Studies on the index compounds for HPLC analysis of Glycyrrhiza uralensis. *Heterocycles* **1997,** *10* (45), 2053-2060.
234. Jia, S.; Liu, D.; Zheng, X.; Zhang, Y.; Li, Y., Two new isoprenyl flavonoids from the leaves of Glycyrrhiza uralensis Fisch. *Yao xue xue bao= Acta pharmaceutica Sinica*

1993, *28* (1), 28-31.
235. Jia, S.; Liu, D.; Wang, H.; Suo, Z., Isolation and identification of gancaonin P-3'-methylether from the leaves of Glycyrrhiza uralensis Fisch. *Yao xue xue bao= Acta pharmaceutica Sinica* **1993,** *28* (8), 623-625.
236. Tao, W.-W.; Duan, J.-A.; Yang, N.-Y.; Tang, Y.-P.; Liu, M.-Z.; Qian, Y.-F., Antithrombotic phenolic compounds from Glycyrrhiza uralensis. *Fitoterapia* **2012,** *83* (2), 422-425.
237. Ye, R.; Fan, Y.-H.; Ma, C.-M., Identification and enrichment of α-glucosidase-inhibiting dihydrostilbene and flavonoids from Glycyrrhiza uralensis leaves. *Journal of agricultural and food chemistry* **2017,** *65* (2), 510-515.
238. Kitagawa, I.; Zhou, J. L.; Sakagami, M.; Taniyama, T.; Yoshikawa, M., Licorice-saponins A3, B2, C2, D3, and E2, five new oleanene-type triterpene oligoglycosides from Chinese Glycyrrhizae radix. *Chemical and Pharmaceutical Bulletin* **1988,** *36* (9), 3710-3713.
239. Yuldashev, M., Flavonoids of the epigeal part ofGlycyrrhiza uralensis. *Chemistry of natural compounds* **1998,** *34* (4), 508-509.
240. Wang, J.; Zhang, J.; Gao, W.; Wang, Q.; Yin, S.; Liu, H.; Man, S., Identification of triterpenoids and flavonoids, step-wise aeration treatment as well as antioxidant capacity of Glycyrrhiza uralensis Fisch. cell. *Industrial crops and products* **2013,** *49*, 675-681.
241. Fukai, T.; Wang, Q. H.; Takayama, M.; Nomura, T., Structures of five new prenylated flavonoids, gancaonins L, M, N, O, and P from aerial parts of Glycyrrhiza uralensis. *Heterocycles* **1990,** *31* (2), 373-82.
242. Han, Y. N.; Chung, M. S.; Kim, T. H.; Han, B. H., Two tetrahydroquinoline alkaloids fromGlycyrrhiza uralensis. *Archives of Pharmacal Research* **1990,** *13* (1), 101-102.
243. Villinski, J. R.; Bergeron, C.; Cannistra, J. C.; Gloer, J. B.; Coleman, C. M.; Ferreira, D.; Azelmat, J.; Grenier, D.; Gafner, S., Pyrano-isoflavans from Glycyrrhiza uralensis with antibacterial activity against Streptococcus mutans and Porphyromonas gingivalis. *Journal of natural products* **2014,** *77* (3), 521-526.
244. Fukai, T.; Marumo, A.; Kaitou, K.; Kanda, T.; Terada, S.; Nomura, T., Anti-Helicobacter pylori flavonoids from licorice extract. *Life sciences* **2002,** *71* (12), 1449-1463.
245. Feng, K.-P.; Chen, R.-D.; Li, J.-H.; Tao, X.-Y.; Liu, J.-M.; Zhang, M.; Dai, J.-G., Flavonoids from the cultured cells of Glycyrrhiza uralensis. *Journal of Asian natural products research* **2016,** *18* (3), 253-259.
246. Fukai, T.; Nishizawa, J.; Yokoyama, M.; Tantai, L.; Nomura, T., Five new isoprenoid-substituted flavonoids, kanzonols M, P and R, from two Glycyrrhiza species. *Heterocycles* **1994,** *38* (5), 1089-1098.
247. Kim, H. J.; Seo, S. H.; Lee, B.-g.; Lee, Y. S., Identification of tyrosinase inhibitors from Glycyrrhiza uralensis. *Planta medica* **2005,** *71* (08), 785-787.
248. Eerdunbayaer; Orabi, M. A.; Aoyama, H.; Kuroda, T.; Hatano, T., Structures of two new flavonoids and effects of licorice phenolics on vancomycin-resistant Enterococcus species. *Molecules* **2014,** *19* (4), 3883-97.
249. Qiao, X.; Song, W.; Ji, S.; Wang, Q.; Guo, D. A.; Ye, M., Separation and characterization of phenolic compounds and triterpenoid saponins in licorice (Glycyrrhiza uralensis) using mobile phase-dependent reversed-phasexreversed-phase comprehensive two-dimensional liquid chromatography coupled with mass spectrometry. *J Chromatogr A* **2015,** *1402*, 36-45.

250. Kuroda, M.; Mimaki, Y.; Sashida, Y.; Mae, T.; Kishida, H.; Nishiyama, T.; Tsukagawa, M.; Konishi, E.; Takahashi, K.; Kawada, T.; Nakagawa, K.; Kitahara, M., Phenolics with PPAR-γ ligand-Binding activity obtained from licorice (Glycyrrhiza uralensis Roots) and ameliorative effects of glycyrin on genetically diabetic KK-Ay mice. *Bioorg. Med. Chem. Lett.* **2003,** *13* (24), 4267-4272.

APPENDIX I.
List of SI tables

SI 1. Sitemap surface type criteria

	ACCPTR [A]	DONOR [B]	PHIL [C]	PHOB [D]	METAL	SURF [E]
1GWR	128.45	185.70	318.23	247.90	0	654.24
2P15	191.43	397.62	587.40	340.33	0	1039.75
3ERT	188.00	427.37	602.56	299.20	0	1496.72
4J24	87.60	204.93	276.69	223.40	0	564.67
1X7J	96.15	172.19	284.55	266.36	0	628.82
1L2J	116.13	288.99	421.65	394.48	0	1244.09

(a) Hydrogen bond acceptor (b) Hydrogen bond donor (c) Hydrophilic (d) Hydrophobic (e) Total surface

SI 2. SP docking scores (kcal/mol) of redocking and crossdocking of co-crystallized ligands

		*h*ERα			*h*ERβ		RMSD
Ligand/PDB	**1GWR**	**2P15**	**3ERT**	**1X7J**	**4J24**	**1L2J**	
1GWR-Nativelig	**−10.91**	**−10.78**	**−8.65**	**−9.81**	**−10.81**	**−9.42**	**0**
2P15-Nativelig	**NA**	**−14.44**	**NA**	**NA**	**NA**	**NA**	**0.51**
3ERT-Nativelig	**NA**	**−10.77**	**−11.37**	**NA**	**NA**	**NA**	**1.90**
1X7J-Nativelig	**−10.06**	**−9.81**	**−9.38**	**−10.55**	**−9.43**	**−9.43**	**0.29**
4J24-Nativelig	**−10.91**	**−10.78**	**−8.65**	**−9.81**	**−10.81**	**−9.42**	**0**
1L2J-Nativelig	**−11.09**	**−11.17**	**−9.36**	**−11.08**	**−11.33**	**−11.92**	**0.40**

SI 3. Compounds from *Glycyrrhiza glabra*

No.	Name (*G. glabra*)	M. Formula	CAS	Class	Ref.
G1	Prenyllicoflavone A= Licoflavone B	$C_{25}H_{26}O_4$	91433-17-9	diprenyl-flavone	169
G2	3-Hydroxyglabrol	$C_{25}H_{28}O_5$	74148-41-7	Flavanone-3-ol	169-170
G3	Glabrol	$C_{25}H_{28}O_4$	59870-65-4	diprenylflavanone	169-170
G4	6''-O-Acetylliquiritin	$C_{23}H_{24}O_{10}$	166531-17-5	Flavanone-glucoside	DNP
G5	Liquiritoside =Liquiritin	$C_{21}H_{22}O_9$	551-15-5	flavanone glycoside	171
G6	Liqcoumarin	$C_{12}H_{10}O_4$	36695-19-9	Coumarin	172
G7	Glucoliquiritin apioside	$C_{32}H_{40}O_{18}$	157226-47-6	flavanone glycoside	46, 171
G8	Pinocembroside	$C_{21}H_{22}O_9$	75829-43-5P	flavanone glycoside	173
G9	Pallidiflorin	$C_{16}H_{12}O_4$	133086-79-0	isoflavone	174
G10	Liquiritigenin	$C_{15}H_{12}O_4$	578-86-9	Flavanone	170
G11	17991-67-2	$C_{30}H_{46}O_5$	17991-67-2	saponin	175
G12	Glabric acid	$C_{30}H_{46}O_5$	22327-86-2	saponin	176
G13	24-Hydroxyliquiritic acid	$C_{30}H_{46}O_5$	20528-69-2	saponin	177
G14	Isoglabrolid	$C_{30}H_{44}O_4$	10376-64-4	saponin	178
G15	Liquiritin apioside	$C_{26}H_{30}O_{13}$	74639-14-8	flavanone glycoside	171
G16	Glabroside	$C_{26}H_{30}O_{13}$	152246-80-5	flavanone glycoside	179
G17	Eicosanyl (E)-caffeate	$C_{29}H_{48}O_4$	28593-90-0	miscellaneous	DNP
G18	Glychionide B	$C_{22}H_{20}O_{11}$	51059-44-0	flavone glycoside	180
G19	862389-76-2	$C_{21}H_{24}O_4$	862389-76-2	prenyl stillbenoid	173
G20	Isoglabranin	$C_{20}H_{20}O_4$	55051-77-9	6-prenylflavanone	181
G21	525585-31-3	$C_{19}H_{22}O_3$	525585-31-3	prenyl stillbenoid	173
G22	kanzonol D	$C_{25}H_{24}O_4$	155233-20-8	prenyl-flavone	182
G23	3,3',5'-Trihydroxy-4-methoxybibenzyl	$C_{15}H_{16}O_4$	60640-97-3	stillbenoid	173
G24	7-O-Methylgenistein or Prunetin	$C_{16}H_{12}O_5$	552-59-0	isoflavone	183
G25	98063-17-3	$C_{31}H_{48}O_4$	98063-17-3	saponin	DNP
G26	24-Hydroxyglycyrrhetinic acid	$C_{30}H_{46}O_5$	52911-55-4	saponin	184
G27	28-Hydroxyglycyrrhetic acid	$C_{30}H_{46}O_5$	56061-86-0	saponin	185
G28	Uralstilbene	$C_{24}H_{30}O_4$	677709-69-2	diprenyl-stilbene	DNP
G29	139101-68-1	$C_{21}H_{26}O_9$	139101-68-1	stillbenoid glycoside	DNP
G30	525585-32-4	$C_{20}H_{24}O_4$	525585-32-4	prenyl stillbenoid	173
G31	2'-O-Methylglabridin	$C_{21}H_{22}O_4$	211099-37-5	pyran-isoflavan	186
G32	4'-O-Methylglabridin	$C_{21}H_{22}O_4$	68978-09-6	pyrano-isoflavane	170
G33	525585-33-5	$C_{19}H_{22}O_4$	525585-33-5	prenyl stillbenoid	173
G34	Isoglycyrol	$C_{21}H_{18}O_6$	23013-86-7	pyran-coumestan	54
G35	Glabridin	$C_{20}H_{20}O_4$	59870-68-7	pyrano-isoflavane	170, 187
G36	Gancaonin F	$C_{21}H_{16}O_6$	126716-33-4	miscellaneous	169
G37	Licoflavone A	$C_{20}H_{18}O_4$	61153-77-3P	prenyl-flavone	171, 188

G38	**Glabrene**	**$C_{20}H_{18}O_4$**	**60008-03-9**	**pyrano-Isoflavene**	**170, 188**
G39	**Arachidic alcohol**	**$C_{20}H_{42}O_1$**	**629-96-9**	**miscellaneous**	**DNP**
G40	**525585-30-2**	**$C_{24}H_{30}O_4$**	**525585-30-2**	**diprenyl stillbenoid**	**173**
G41	**525585-29-9**	**$C_{24}H_{30}O_3$**	**525585-29-9**	**diprenyl stillbenoid**	**173**
G42	**1201180-35-9**	**$C_{25}H_{28}O_4$**	**1201180-35-9**	**prenyl-pyrano-isoflavanone**	**189**
G43	**Shinflavanone**	**$C_{25}H_{26}O_4$**	**157414-03-4**	**pyrano-prenyl-flavanone**	**170-171**
G44	**Licoagrocarpin**	**$C_{21}H_{22}O_4$**	**202815-29-0**	**prenyl-pterocarpan**	**170**
G45	**Xambioona**	**$C_{25}H_{24}O_4$**	**82345-36-6**	**dipyrano-flavanone**	**188**
G46	**Glycyrrhizin**	**$C_{42}H_{62}O_{16}$**	**1405-86-3**	**saponin glycoside**	**171**
G47	**Glycycoumarin**	**$C_{21}H_{20}O_6$**	**94805-82-0P**	**coumarin**	**46**
G48	**Glyinflanin B**	**$C_{20}H_{18}O_5$**	**142750-23-0**	**miscellaneous**	**190**
G49	**3,4-Didehydroglabridin**	**$C_{20}H_{18}O_4$**	**214283-58-6**	**pyrano-isoflavene**	**170**
G50	**Glabrocoumarin**	**$C_{20}H_{16}O_5$**	**866021-47-8**	**Coumarin**	**188**
G51	**Glabrone =Eurycarpin B**	**$C_{20}H_{16}O_5$**	**60008-02-8**	**pyrano-Isoflavone**	**170**
G52	**56262-33-0**	**$C_{30}H_{46}O_4$**	**56262-33-0**	**saponin**	**191**
G53	**Glabrocoumarone B =Glyinflanin H**	**$C_{19}H_{16}O_4$**	**164123-54-0**	**pyrano-1-benzofuran**	**170**
G54	**Glyzarin**	**$C_{18}H_{14}O_4$**	**62820-28-4**	**isoflavone**	**170**
G55	**2-Methyl-7-acetoxyisoflavone**	**$C_{18}H_{14}O_4$**	**3211-63-0**	**isoflavone**	**192**
G56	**7-Methoxy-2-methylisoflavone**	**$C_{17}H_{14}O_3$**	**19725-44-1**	**isoflavone**	**192**
G57	**7-Hydroxy-2-methylisoflavone**	**$C_{16}H_{12}O_3$**	**2859-88-3**	**isoflavone**	**192**
G58	**11-Deoxoglycyrrhetinic acid**	**$C_{30}H_{48}O_3$**	**564-16-9**	**saponin**	**53**
G59	**14884-88-9**	**$C_{30}H_{46}O_3$**	**14884-88-9**	**saponin**	**DNP**
G60	**Glyinflanin A**	**$C_{25}H_{28}O_5$**	**142542-83-4**	**propanedione**	**190**
G61	**Hispaglabridin A**	**$C_{25}H_{28}O_4$**	**68978-03-0**	**prenyl-pyrano-Isoflavane**	**170**
G62	**944257-60-7**	**$C_{25}H_{26}O_4$**	**944257-60-7**	**dipyrano-chalcone**	**187**
G63	**Hispaglabridin B**	**$C_{25}H_{26}O_4$**	**68978-02-9**	**dipyrano-Isoflavane**	**170**
G64	**Licoagroside B**	**$C_{18}H_{24}O_{12}$**	**325144-72-7**	**miscellaneous**	**190**
G65	**3'-Methoxyglabradin**	**$C_{21}H_{22}O_5$**	**74046-05-2**	**pyran-isoflavan**	**193**
G66	**944257-63-0**	**$C_{65}H_{108}O_8$**	**944257-63-0**	**neolignan lipid esters**	**187**
G67	**Kanzonol T**	**$C_{25}H_{26}O_7$**	**181476-22-2**	**pyrano-prenyl-isoflavone**	**194**
G68	**Kanzonol Z**	**$C_{25}H_{26}O_5$**	**220860-37-7**	**pyrano-prenyl-Flavanone-3-ol**	**170**
G69	**Kanzonol V**	**$C_{24}H_{24}O_4$**	**184584-65-4**	**prenyl-pyrano-benzofuran**	**169**
G70	**Isoglycycoumarin**	**$C_{21}H_{20}O_6$**	**117038-82-1**	**coumarin**	**DNP**
G71	**FCO69-H**	**$C_{21}H_{18}O_5$**	**FCO69-H**	**prenyl-isoflavene**	**DNP**
G72	**Kanzonol B**	**$C_{20}H_{18}O_4$**	**155233-19-5**	**pyrano-chalcone**	**170**
G73	**Kanzonol W**	**$C_{20}H_{16}O_5$**	**184584-82-5**	**pyrano-Arylcoumarin**	**170**
G74	**isoderon**	**$C_{20}H_{16}O_5$**	**121747-89-5**	**pyrano-isoflavone**	**DNP**
G75	**Glabrocoumarone A**	**$C_{19}H_{16}O_4$**	**178330-48-8**	**pyrano-1-benzofuran**	**170**
G76	**944257-62-9**	**$C_{63}H_{104}O_8$**	**944257-62-9**	**neolignan lipid esters**	**187**
G77	**944257-61-8**	**$C_{61}H_{100}O_8$**	**944257-61-8**	**neolignan lipid esters**	**187**

G78	1201180-37-1	$C_{36}H_{58}O_7$	1201180-37-1	saponin glycosides	189
G79	Glycyrrhetol	$C_{30}H_{48}O_3$	14226-18-7	saponin	195
G80	Liquiritic acid	$C_{30}H_{46}O_4$	10379-72-3	saponin	196
G81	18α-Glycyrrhetinic acid	$C_{30}H_{46}O_4$	1449-05-4	saponin	197
G82	18β-Glycyrrhetinic acid= Enoxolone	$C_{30}H_{46}O_4$	471-53-4	saponin	46
G83	Kanzonol Y	$C_{25}H_{30}O_5$	184584-87-0	diprenyldihydrochalcone	170
G84	Licoagrodin	$C_{45}H_{44}O_9$	325144-66-9	miscellaneous	190
G85	Licoagrochalcone D	$C_{21}H_{22}O_5$	325144-69-2	chalcone	190
G86	Licochalcone A	$C_{21}H_{22}O_4$	58749-22-7	prenylated Chalcone	171
G87	Licoagrodione	$C_{20}H_{20}O_6$	220860-12-8	miscellaneous	198
G88	Methoxyphaseollin	$C_{21}H_{20}O_5$	157226-48-7	pyrano-pterocarpin	171
G89	Licoagrochalcone B	$C_{21}H_{20}O_4$	325144-67-0	pyrano-chalcone	199
G90	phaseolinisoflavin	$C_{20}H_{20}O_4$	40323-57-7	pyran-isoflavan	193
G91	Isoglabrone	$C_{20}H_{16}O_5$	1115012-65-1	pyrano-Isoflavone	170
G92	Licochalcone B	$C_{16}H_{14}O_5$	58749-23-8P	chalcone	188
G93	Deoxyglabrolide	$C_{30}H_{46}O_3$	10379-62-1	saponin	199
G94	Licoricidin	$C_{26}H_{32}O_5$	30508-27-1	diprenyl-isoflavan	200
G95	Licoagroside A	$C_{23}H_{24}O_{12}$	30508-27-1	isoflavonoids glycosides	190
G96	Licocoumarin A	$C_{25}H_{26}O_5$	222034-74-4	coumarin	DNP
G97	Kanzonol R	$C_{22}H_{26}O_5$	156250-73-6	prenyl-Isoflavan	169
G98	Glyasperin K	$C_{22}H_{24}O_6$	156162-03-7	prenyl-isoflavanone	DNP
G99	licoagrochalcone C	$C_{21}H_{22}O_5$	325144-68-1	chalcone	190
G100	Morachalcone A	$C_{20}H_{20}O_5$	76472-88-3P	prenyl-chalcone	188
G101	Isolicoflavonol	$C_{20}H_{18}O_6$	94805-83-1	prenyl-3-OH-flavone	DNP
G102	Isomucronulatol	$C_{17}H_{18}O_5$	64474-51-7	isoflavan	200
G103	Folerogenin	$C_{16}H_{14}O_6$	35815-06-6	3-OH-flavanone	201
G104	glyzaglabrin	$C_{16}H_{10}O_6$	65242-64-0	isoflavone	202
G105	Liquiridiolic acid	$C_{30}H_{48}O_5$	20528-70-5	saponin	177
G106	Licoflavanone =3'-Prenylnaringenin	$C_{20}H_{20}O_5$	119240-82-3	prenyl-flavanone	173
G107	6-Prenylnaringenin	$C_{20}H_{20}O_5$	68236-13-5	prenyl-flavanone	169
G108	Gancaonin L	$C_{20}H_{18}O_6$	129145-50-2P	prenyl-Isoflavone	188
G109	Licoagrochalcone A	$C_{20}H_{20}O_4$	202815-28-9	prenyl-Chalcone	170
G110	Licoagroaurone	$C_{20}H_{18}O_5$	325144-70-5	prenyl-Benzofuranone	190
G111	Erythrinin B =Wighteone	$C_{20}H_{18}O_5$	51225-30-0	prenyl isoflavone	203
G112	8-Prenylgenistein =Lupiwighteone	$C_{20}H_{18}O_5$	104691-86-3	prenyl-Isoflavone	170
G113	163121-02-6	$C_{15}H_{12}O_4$	163121-02-6	chalcone	DNP
G114	18184-25-3	$C_{30}H_{44}O_5$	18184-25-3	saponin	204
G115	10401-33-9	$C_{30}H_{44}O_4$	10401-33-9	saponin	205
G116	Licuraside	$C_{26}H_{30}O_{13}$	29913-71-1	chalcone glycoside	171
G117	Kanzonol X =Tenuifolin B	$C_{25}H_{30}O_4$	182745-37-5	diprenyl-Isoflavane	170

G118	**Isoliquiritin**	**$C_{21}H_{22}O_9$**	**5041-81-6**	**chalcone glycoside**	**171**
G119	**Neoisoliquiritigenin**	**$C_{21}H_{22}O_9$**	**59122-93-9**	**chalcone glycoside**	**171**
G120	**Glychionide A**	**$C_{21}H_{18}O_{11}$**	**119152-50-0**	**flavone glycoside**	**180**
G121	**Licochalcone C**	**$C_{21}H_{22}O_4$**	**144506-14-9**	**prenyl chalcone**	**199**
G122	**Calycosin**	**$C_{16}H_{12}O_5$**	**20575-57-9**	**isoflavone**	**55**
G123	**(−)-Naringenin**	**$C_{15}H_{12}O_5$**	**480-41-1**	**Flavanone**	**173**
G124	**Formononetin=Biochanin B**	**$C_{16}H_{12}O_4$**	**485-72-3**	**Isoflavone**	**47**
G125	**Isoliquiritigen**	**$C_{15}H_{12}O_4$**	**961-29-5**	**Chalcone**	**170**
G126	**Kumatakenin B**	**$C_{15}H_{10}O_4$**	**2196-14-7**	**flavone**	**206**
G127	**Glicoricone**	**$C_{21}H_{20}O_6$**	**161099-37-2**	**prenyl-isoflavone**	**DNP**
G128	**R)-(−)-Vestitol**	**$C_{16}H_{16}O_4$**	**35878-41-2P**	**isoflavan**	**188**
G129	**4',7-Dihydroxyflavone 7-D-glucoside**	**$C_{21}H_{20}O_9$**	**20633-86-7**	**flavone-glycoside**	**54**
G130	**1217305-76-4**	**$C_{21}H_{20}O_5$**	**1217305-76-4**	**benzaldehyde-pyrano-isoflavan**	**188**
G131	**1217305-78-6**	**$C_{20}H_{20}O_5$**	**1217305-78-6**	**prenyl-Flavanone-3-ol**	**188**
G132	**201157-06-4**	**$C_{16}H_{16}O_5$**	**201157-06-4**	**isoflavan**	**188**
G133	**Echinatin =Retrochalcone**	**$C_{16}H_{14}O_4$**	**34221-41-5P**	**chalcone**	**171, 188**
G134	**131319-67-0**	**$C_{15}H_{14}O_5$**	**131319-67-0**	**dihydrochalcone**	**188**
G135	**938190-29-5**	**$C_{25}H_{28}O_5$**	**938190-29-5**	**diprenyl-chalcone**	**188**
G136	**151135-83-0**	**$C_{25}H_{26}O_5$**	**151135-83-0**	**prenyl-pyrano-chalcone**	**188**
G137	**Euchrenone a_5**	**$C_{25}H_{26}O_4$**	**125140-20-7**	**pyrano-prenyl-flavanone**	**188**
G138	**905708-40-9**	**$C_{21}H_{22}O_6$**	**905708-40-9**	**prenyl-chalcone**	**188**
G139	**3'-Hydroxy-4'-O-methylglabridin**	**$C_{21}H_{22}O_5$**	**175554-11-7**	**pyrano-isoflavane**	**188**
G140	**938190-32-0**	**$C_{21}H_{20}O_5$**	**938190-32-0**	**prenyl-isoflavone**	**188**
G141	**Glabroisoflavanone B**	**$C_{21}H_{20}O_5$**	**866021-46-7**	**isoflavanone**	**207**
G142	**Glabroisoflavanone A**	**$C_{20}H_{18}O_5$**	**204700-00-5**	**isoflavanone**	**207**
G143	**400900-10-9**	**$C_{21}H_{24}O_{10}$**	**400900-10-9**	**chalcone glycoside**	**208**
G144	**131319-67-0**	**$C_{15}H_{14}O_5$**	**131319-67-0**	**dihydrochalcone**	**188**
G145	**Rocymosin B**	**$C_{21}H_{24}O_{10}$**	**150036-01-4**	**chalcone glycoside**	**DNP**
G146	**Glabraisoflavanone B**	**$C_{30}H_{38}O_5$**	**1201428-07-0**	**prenyl flavanone**	**189**
G147	**p-Hydroxychalcone**	**$C_{15}H_{12}O_2$**	**20426-12-4**	**chalcone**	**209**
G148	**Shinpterocarpin**	**$C_{20}H_{18}O_4$**	**157414-04-5**	**pyrano-pterocarpan**	**188**
G149	**Cordifolin**	**$C_{16}H_{14}O_5$**	**53219-84-4**	**chalcone**	**DNP**
G150	**Licoagrone**	**$C_{45}H_{42}O_{10}$**	**228099-12-5**	**miscellaneous**	**DNP**
G151	**3-Methyl-3-hepten-2-one**	**$C_8H_{14}O_1$**	**39899-08-6**	**miscellaneous**	**DNP**
G152	**8-Prenylphaseollinisoflavan**	**$C_{25}H_{28}O_4$**	**175554-12-8**	**prenyl-pyrano-isoflavane**	**170**
G153	**DHY67-Q**	**$C_{25}H_{28}O_6$**	**DHY67-Q**	**Diprenyl-isoflavanone**	**DNP**
G154	**Glyasperin B**	**$C_{21}H_{22}O_6$**	**142488-54-8**	**prenyl-isoflavanone**	**DNP**
G155	**135432-48-3**	**$C_{26}H_{30}O_{13}$**	**135432-48-3**	**flavanone glycoside**	**210**
G156	**Abyssinone II =3'-Prenylliquiritigenin**	**$C_{20}H_{20}O_4$**	**77263-08-2**	**prenyl-flavanone**	**170**

G157	**Isoviolanthin**	**$C_{27}H_{30}O_{14}$**	**40788-84-9**	**flavone C-glycoside**	**211**
G158	**1217888-51-1**	**$C_{25}H_{30}O_{6}$**	**1217888-51-1**	**miscellaneous**	**212**
G159	**4''-Hydroxyglabridin**	**$C_{20}H_{20}O_{5}$**	**938190-33-1P**	**pyrano-isoflavan**	**188**
G160	**O-Methylshinpterocarpin**	**$C_{21}H_{20}O_{4}$**	**157479-38-4P**	**pyrano-pterocarpan**	**171, 188**
G161	**938190-31-9**	**$C_{14}H_{16}O_{4}$**	**938190-31-9**		**188**
G162	**1217888-54-4**	**$C_{14}H_{16}O_{3}$**	**1217888-54-4**		**188**
G163	**938190-35-3**	**$C_{20}H_{22}O_{6}$**	**938190-35-3**	**prenyl-dihydrochalcone**	**188**
G164	**905708-41-0**	**$C_{21}H_{20}O_{6}$**	**905708-41-0**	**pyrano-isoflavanone-3-ol**	**188**
G165	**Ononin 6''-O-acetate**	**$C_{24}H_{24}O_{10}$**	**120727-10-8**	**isoflavone-glucoside**	**NCNPR**
G166	**Isoliquiritigenin-4'-Me-ether**	**$C_{16}H_{14}O_{4}$**	**32274-67-2**	**chalcone**	**NCNPR**
G167	**demethylvestitol**	**$C_{15}H_{14}O_{4}$**	**ZA17-Gg-64190-84-7**	**isoflavan**	**NCNPR**
G168	**Resorcinol**	**$C_{6}H_{6}O_{2}$**	**Gg2-ZA13**	**phenolic**	**NCNPR**
G169	**ZA15-Gg**	**$C_{21}H_{22}O_{7}$**	**ZA15-Gg**	**arylcoumarine**	**NCNPR**
G170	**ZA19-Gg**	**$C_{20}H_{18}O_{6}$**	**ZA19-Gg**	**Prenyl-isoflavone**	**NCNPR**
G171	**Erythbidin A**	**$C_{20}H_{20}O_{4}$**	**210050-83-2**	**pyran-isoflavan**	**NCNPR**
G172	**ZA20-Gg**	**$C_{27}H_{26}O_{9}$**	**ZA20-Gg**	**miscellaneous**	**NCNPR**
G173	**ZA41-Gg**	**$C_{36}H_{38}O_{16}$**	**ZA41-Gg**	**miscellaneous**	**NCNPR**
G174	**ZA21-Gg**	**$C_{20}H_{18}O_{5}$**	**ZA21-Gg**	**arylcoumarine**	**NCNPR**
G175	**GDY88-C**	**$C_{25}H_{28}O_{6}$**	**GDY88-C**	**diprenylisoflavanone**	**DNP**
G176	**Erybacin B**	**$C_{19}H_{18}O_{5}$**	**1314877-89-8**		**170**
G177	**Parvisoflavone A**	**$C_{20}H_{16}O_{6}$**	**50277-01-5**	**pyran-isoflavone**	**170**

SI 4. Compounds from *Glycyrrhiza uralensis*

No.	Name (*G. uralensis*)	M. Formula	CAS	Class	Ref
U1	Licorice glycoside C2	$C_{36}H_{38}O_{16}$	202657-55-4	flavanone glycoside	DNP
U2	342433-16-3	$C_{22}H_{24}O_9$	342433-16-3	flavanone glycoside	213
U3	Rhizobialide	$C_{18}H_{32}O_3$	928213-29-0	stearolactone	214
U4	Neoliquiritin	$C_{21}H_{22}O_9$	5088-75-5	flavanone glycoside	50
U5	Cyclolicocoumarone	$C_{20}H_{20}O_5$	1253641-17-6	pyrano-benzofuran	215
U6	Kumatakenin	$C_{17}H_{14}O_6$	3301-49-3	flavone-3-ol	216
U7	Uralenneoside	$C_{12}H_{14}O_8$	143986-30-5	miscellaneous	217
U8	Licorice glycoside D2	$C_{35}H_{36}O_{15}$	202657-65-6	flavanone glycoside	53
U9	76884-47-4	$C_8H_{10}N_2O_1$	76884-47-4	pyrrolo-pyrimidine	218
U10	Licorice glycoside D1	$C_{35}H_{36}O_{15}$	202657-63-4	flavanone glycoside	53
U11	Licorice glycoside E	$C_{35}H_{35}N_1O_{14}$	202657-66-7	miscellaneous	219
U12	Liquiritigenin 7,4'-diglucoside	$C_{27}H_{32}O_{14}$	93446-18-5	flavanone glycoside	220
U13	134250-13-8	$C_{42}H_{64}O_{16}$	134250-13-8	saponin glycosides	221
U14	Glycycarpan	$C_{26}H_{32}O_6$	1346768-08-8	prenyl-pyran-pterocarpan	222
U15	Gancaonin R	$C_{24}H_{30}O_4$	134958-53-5	diprenyl-dihydrostillbenoid	169
U16	Gancaonin S	$C_{24}H_{30}O_4$	134958-54-6	diprenyl-dihydrostillbenoid	169
U17	Gancaonin U	$C_{24}H_{28}O_4$	134958-56-8	diprenyl-Dihydrophenancerene	169
U18	Glycyrin	$C_{22}H_{22}O_6$	66056-18-6	coumarin	216
U19	Gancaonin I	$C_{21}H_{22}O_5$	126716-36-7	prenylated-benzofuran	211
U20	Gancaonin D	$C_{21}H_{20}O_7$	124596-88-9	prenyl-isoflavone	223
U21	Licoricesaponin G_2	$C_{42}H_{62}O_{17}$	118441-84-2	saponin glycoside	50
U22	Glycybenzofuran	$C_{21}H_{22}O_5$	1253641-15-4	prenylated-benzofuran	215
U23	Licocoumarone	$C_{20}H_{20}O_5$	118524-14-4	prenylated-benzofuran	215
U24	Gancaonin C	$C_{20}H_{18}O_6$	124596-87-8	prenyl-isoflavone	223
U25	Gancaonin V	$C_{19}H_{20}O_4$	134958-57-9	prenyl-Dihydrophenancerene	169
U26	Uralsaponin A	$C_{42}H_{62}O_{16}$	103000-77-7	saponin glycoside	224
U27	Licorice glycoside C1	$C_{36}H_{38}O_{16}$	202657-31-6	flavanone glycoside	225
U28	24-Hydroxyglycyrrhetic acid	$C_{31}H_{48}O_5$	18184-26-4	saponin	226
U29	Glyuranolide	$C_{31}H_{44}O_6$	123914-44-3	saponin-lactone	227
U30	Licoricesaponin A_3	$C_{48}H_{72}O_{21}$	118325-22-7	saponin glycoside	50
U31	Glypallidifloric acid	$C_{30}H_{46}O_3$	17991-81-0	saponin	DNP
U32	Kanzonol L	$C_{30}H_{32}O_6$	156281-31-1	diprenyl-pyrano-isoflavone	DNP
U33	Licobichalcone	$C_{32}H_{26}O_{10}$	637338-06-8	biflavonoid	DNP
U34	Kanzonol I	$C_{27}H_{32}O_5$	152546-94-6	prenyl-pyran-isoflavan	228
U35	Kanzonol H	$C_{26}H_{32}O_5$	152511-46-1	prenyl-pyran-isoflavan	228
U36	Kanzonol J	$C_{26}H_{30}O_5$	152511-47-2	dipyrano-isoflavan	228
U37	Kanzonol F	$C_{26}H_{28}O_5$	152511-44-9	prenyl-pyran-pterocarpan	228

U38	Kanzonol M	$C_{23}H_{26}O_6$	156250-70-3	prenyl-isoflavan	DNP
U39	Kanzonol N	$C_{22}H_{24}O_6$	156250-71-4	prenyl-isoflavan	DNP
U40	Uralsaponin B	$C_{42}H_{62}O_{16}$	105038-43-5	saponin glycoside	211
U41	Glicophenone	$C_{20}H_{22}O_6$	303175-66-8		211
U42	Glycyrrhizol B	$C_{21}H_{18}O_5$	877373-01-8	pterocarpan	DNP
U43	Licoricesaponin B_2	$C_{42}H_{64}O_{15}$	118536-86-0	saponin glycoside	229
U44	22-Dehydroxyuralsaponin C	$C_{42}H_{64}O_{15}$	134250-15-0	saponin glycoside	224
U45	Licoricesaponin K2	$C_{42}H_{62}O_{16}$	135815-61-1	saponin glycoside	230
U46	134449-15-3	$C_{42}H_{62}O_{15}$	134449-15-3	saponin glycoside	229
U47	Licoricesaponin C_2	$C_{42}H_{62}O_{15}$	118525-49-8	saponin glycoside	53
U48	Uralsaponin W	$C_{42}H_{62}O_{15}$	1616062-88-4	saponin glycoside	229
U49	18β-Glycyrrhetic acid methyl ester	$C_{31}H_{48}O_4$	1477-44-7	saponin	226
U50	Licorisoflavan A	$C_{27}H_{34}O_5$	129314-37-0	diprenyl-isoflavan	216
U51	Kanzonol G	$C_{26}H_{30}O_6$	152511-45-0	diprenyl-isoflavanone	228
U52	1-Methoxyficifolinol	$C_{26}H_{30}O_5$	129280-35-9	diprenyl-pterocarpan	228
U53	Kanzonol K	$C_{26}H_{28}O_6$	156281-30-0	diprenyl-isoflavan	231
U54	Glycyrrhizol A	$C_{26}H_{28}O_5$	877373-00-7	pterocarpan	232
U55	Isoangustone A	$C_{25}H_{26}O_6$	129280-34-8	diprenylated-isoflavone	211
U56	Glyurallin B	$C_{25}H_{26}O_6$	199331-53-8	diprenylated-isoflavone	233
U57	Gancaonin Q	$C_{25}H_{26}O_5$	134958-52-4	diprenyl-flavone	169
U58	22β-Acetoxyglycyrrhizin	$C_{44}H_{64}O_{18}$	938042-17-2	saponin glycoside	211
U59	150853-98-8	$C_{21}H_{20}O_7$	150853-98-8	prenyl-flavone	234
U60	gancaonin P-3'-methyl ether	$C_{21}H_{20}O_7$	151776-21-5	prenyl-3-OH-flavone	235
U61	Uralene	$C_{21}H_{20}O_7$	150853-99-9	prenyl-flavone	234
U62	Glycryrurol	$C_{21}H_{20}O_7$	1415339-39-7	miscellaneous	236
U63	Licoleafol	$C_{20}H_{20}O_7$	677709-68-1	prenyl flavanone	237
U64	Uralenol	$C_{20}H_{18}O_7$	139163-15-8	prenyl-isoflavone-3-ol	211
U65	Gancaonin P	$C_{20}H_{18}O_7$	129145-54-6	prenyl-flavone-3-ol	169
U66	Neouralenol	$C_{20}H_{18}O_7$	C20 H18 O7	prenyl-3-OH-flavone	DNP
U67	119418-01-8	$C_{44}H_{64}O_{16}$	119418-01-8	saponin glycoside	238
U68	Phaseol	$C_{20}H_{16}O_5$	88478-02-8	coumestan	DNP
U69	Isotrifoliol	$C_{16}H_{10}O_6$	329319-08-6	coumestan	236
U70	Galangin	$C_{15}H_{10}O_5$	548-83-4	3-OH-flavone	239
U71	Licorice glycoside A	$C_{36}H_{38}O_{16}$	202657-28-1	chalcone-glycoside	53
U72	Licorice glycoside B	$C_{35}H_{36}O_{15}$	202657-29-2	chalcone-glycoside	240
U73	166531-18-6	$C_{31}H_{44}O_5$	166531-18-6	saponin	DNP
U74	24-Hydroxyglabrolide	$C_{30}H_{44}O_5$	98063-18-4	saponin-lactone	226
U75	Isoliquiritin apioside= Neolicuroside	$C_{26}H_{30}O_{13}$	120926-46-7	chalcone glycoside	216
U76	118536-87-1	$C_{50}H_{76}O_{21}$	118536-87-1	saponin	238
U77	Sigmoidin A	$C_{25}H_{28}O_6$	87746-48-3	diprenyl-flavanone	DNP

U78	Gancaonin E	$C_{25}H_{28}O_6$	124596-89-0	diprenyl-flavanone	169
U79	782503-66-6	$C_{25}H_{26}O_6$	782503-66-6	diprenyl-flavone	DNP
U80	Vestitol 7-O-glucoside	$C_{22}H_{26}O_9$	202533-15-1	flavan glycoside	DNP
U81	Licoricone	$C_{22}H_{22}O_6$	51847-92-8	prenyl-isoflavone	211
U82	952481-52-6	$C_{21}H_{20}O_9$	952481-52-6	flavone glycoside	DNP
U83	Topazolin	$C_{21}H_{20}O_6$	109605-79-0	prenyl-flavone-3-ol	211
U84	Gancaonin N	$C_{21}H_{20}O_6$	129145-52-4	prenyl-isoflavone	241
U85	Gancaonin B	$C_{21}H_{20}O_6$	124596-86-7	prenyl-isoflavone	DNP
U86	Glycyrol	$C_{21}H_{18}O_6$	23013-84-5	prenyl-coumestan	216
U87	Licoflavonol	$C_{20}H_{18}O_6$	60197-60-6	prenyl-flavone-3-ol	216
U88	Gancaonin O	$C_{20}H_{18}O_6$	129145-53-5	prenyl-flavone	169
U89	AlloLicoisoflavone b	$C_{20}H_{16}O_6$	117204-81-6	pyrano-isoflavone	211
U90	60169-66-6	$C_{11}H_{15}N_1$	60169-66-6	tetrahydroquinoline	242
U91	28971-03-1	$C_{10}H_{13}N_1$	28971-03-1	tetrahydroquinoline	242
U92	Uralsaponin C	$C_{42}H_{64}O_{16}$	1262326-46-4	saponin glycoside	211
U93	166531-19-7	$C_{32}H_{48}O_6$	166531-19-7	saponin	DNP
U94	Uralenolide	$C_{30}H_{44}O_4$	111150-27-7	saponin	DNP
U95	Licoriquinone A	$C_{27}H_{32}O_6$	1346768-10-2	Isoflavan-Quinone	222
U96	Licoriquinone B	$C_{26}H_{30}O_6$	1346768-14-6	Isoflavan-Quinone	243
U97	Uralsaponin F	$C_{44}H_{64}O_{19}$	1208004-79-8	saponin glycoside	211
U98	Edudiol	$C_{21}H_{22}O_5$	63343-94-2	prenyl-pterocarpan	222
U99	1-Methoxyphaseollidin	$C_{21}H_{22}O_5$	65428-13-9	prenyl-pterocarpan	244
U100	Gancaonin A	$C_{21}H_{20}O_5$	27762-99-8	prenyl-isoflavone	223
U101	Gancaonin G	$C_{21}H_{20}O_5$	126716-34-5	prenyl-isoflavone	243
U102	Gancaonin M	$C_{21}H_{20}O_5$	129145-51-3	prenyl-isoflavone	241
U103	646508-72-7	$C_{43}H_{62}O_{17}$	646508-72-7	saponin glycoside	DNP
U104	Glyurallin A	$C_{21}H_{22}O_5$	213130-81-5	pterocarpan	DNP
U105	Isobavachalcone	$C_{20}H_{20}O_4$	20784-50-3	prenyl-chalcone	245
U106	Licoisoflavone a	$C_{20}H_{18}O_6$	66056-19-7	prenyl-isoflavone	211
U107	Uralsaponin E	$C_{42}H_{60}O_{17}$	1262489-45-1	saponin glycoside	224
U108	Kanzonol Q	$C_{15}H_{16}O_4$	17053-75-7	miscellaneous	246
U109	Uralsaponin D	$C_{42}H_{58}O_{18}$	1262489-44-0	1262489-44-0	DNP
U110	2',4',5,7-Tetrahydroxy-3',8-diprenylisoflavanone	$C_{25}H_{28}O_6$	141846-47-1	diprenyl-isoflavanone	247
U111	Gancaonin H	$C_{25}H_{24}O_6$	126716-35-6	prenyl-pyrano-isoflavone	DNP
U112	Kanzonol P	$C_{22}H_{24}O_5$	156250-72-5	prenyl-pterocarpan	228
U113	1680176-18-4	$C_{20}H_{20}O_5$	1680176-18-4	prenyl-benzofuran	248
U114	1177433-19-0	$C_{21}H_{18}O_6$	1177433-19-0	coumestan	248
U115	Demethylglycyrol	$C_{20}H_{16}O_6$	1680176-22-0	coumestan	249
U117	1680176-21-9	$C_{19}H_{18}O_5$	1680176-21-9	prenyl-benzofuran	248
U118	Semilicoisoflavone B	$C_{20}H_{16}O_6$	129280-33-7	pyrano-isoflavone	211
U119	Isokaempferide	$C_{16}H_{12}O_6$	1592-70-7	flavone-3ol	215

U120	6468-37-7	$C_{16}H_{12}O_4$	6468-37-7	Coumarin	248
U121	6,8-Diprenylorobol	$C_{25}H_{26}O_6$	66777-70-6	diprenyl-isoflavone	248
U122	6,8-Diprenylgenistein	$C_{25}H_{26}O_5$	51225-28-6	diprenyl isoflavone	211
U123	Glyasperin D	$C_{22}H_{26}O_5$	142561-10-2	prenyl-isoflavan	216
U124	23013-88-9	$C_{23}H_{22}O_6$	23013-88-9	prenyl-coumestan	216
U125	1680176-23-1	$C_{22}H_{22}O_6$	1680176-23-1	prenyl-isoflavone	216
U126	1680176-20-8	$C_{21}H_{20}O_6$	1680176-20-8	coumarine	216
U127	1680176-17-3	$C_{21}H_{20}O_6$	1680176-17-3	prenyl-isoflavone	216
U128	Licoarylcoumarin	$C_{21}H_{20}O_6$	125709-31-1	prenyl-coumarin	211
U129	Uralsaponin X	$C_{50}H_{74}O_{22}$	1616062-89-5	saponin glycoside	229
U130	Uralsaponin U	$C_{42}H_{62}O_{17}$	1616062-86-2	saponin glycoside	229
U131	Uralsaponin N	$C_{42}H_{62}O_{17}$	1616062-79-3	saponin glycoside	229
U132	Licoricesaponin H_2	$C_{42}H_{62}O_{16}$	118441-85-3	saponin glycoside	230
U133	Uralsaponin V	$C_{42}H_{62}O_{15}$	1616062-87-3	saponin glycoside	229
U134	Uralsaponin O	$C_{42}H_{60}O_{16}$	1616062-80-6	saponin glycoside	229
U135	Licoricesaponin E_2	$C_{42}H_{60}O_{16}$	119417-96-8	saponin glycoside	230
U136	Araboglycyrrhizin	$C_{41}H_{62}O_{14}$	1622142-49-7	saponin glycoside	216
U137	22β-Acetoxyglycrrhetaldehyde	$C_{32}H_{48}O_5$	1614257-31-6	saponin	229
U138	Uralsaponin S	$C_{48}H_{74}O_{20}$	1616062-84-0	saponin glycoside	229
U139	Uralsaponin R	$C_{48}H_{74}O_{20}$	1616062-83-9	saponin glycoside	229
U140	Uralsaponin T	$C_{48}H_{74}O_{19}$	1616062-85-1	saponin glycoside	229
U141	Uralsaponin Y	$C_{48}H_{70}O_{20}$	1616062-90-8	saponin glycoside	229
U142	Uralsaponin Q	$C_{47}H_{72}O_{19}$	1616062-82-8	saponin glycoside	229
U143	Uralsaponin M	$C_{44}H_{64}O_{18}$	1616062-78-2	saponin glycoside	229
U144	Uralsaponin P	$C_{42}H_{64}O_{16}$	1616062-81-7	saponin glycoside	229
U145	Licoricesaponin J_2	$C_{42}H_{64}O_{16}$	938042-18-3	saponin glycoside	211
U146	Isoliquiritigenin 4,4'-di-O-glucopyranoside	$C_{27}H_{32}O_{14}$	69262-36-8	chalcone glycoside	211
U147	Glycyuralin B	$C_{21}H_{22}O_5$	1879910-23-2	prenyl-pterocarpan	211
U148	Ononin=Formononetin 7-O-glucoside	$C_{22}H_{22}O_9$	486-62-4	isoflavone glycoside	211
U149	Apigenin 4'-O-glucoside	$C_{21}H_{20}O_{10}$	20486-34-4	flavone glycoside	211
U150	Sophoraflavone B	$C_{21}H_{20}O_9$	22052-75-1	flavone glycoside	211
U151	Daidzin	$C_{21}H_{20}O_9$	552-66-9	isoflavone glycoside	211
U152	Glycyuralin E	$C_{21}H_{22}O_6$	1879910-26-5	prenylated-benzofuran	211
U153	Glycyuralin F	$C_{20}H_{20}O_7$	1879910-27-6	hydroxyprenyl-isoflavone	211
U154	Syringic acid glucoside	$C_{15}H_{20}O_{10}$	33228-65-8	phenolic-glycoside	211
U155	Luteone	$C_{20}H_{18}O_6$	41743-56-0	prenyl isoflavone	211
U156	Vicenin= Apigenin 6,8-di-C-glucoside	$C_{27}H_{30}O_{15}$	23666-13-9	flavone-glycoside	220
U157	Homobutein	$C_{16}H_{14}O_5$	34000-39-0	chalcone	211

U158	**Pratensein**	**$C_{16}H_{12}O_6$**	**2284-31-3**	**isoflavone**	**211**
U159	**Isomedicarpin**	**$C_{16}H_{14}O_4$**	**74560-05-7**	**pterocarpan**	**211**
U160	**7-Methoxy-2',4'-dihydroxyisoflavone**	**$C_{16}H_{12}O_5$**	**7622-53-9**	**isoflavone**	**211**
U161	**Apigenin 7-O-methyl ether**	**$C_{16}H_{12}O_5$**	**437-64-9**	**flavone**	**211**
U162	**Biochanin A**	**$C_{16}H_{12}O_5$**	**491-80-5**	**isoflavone**	**211**
U163	**Isoviolanthin**	**$C_{27}H_{30}O_{14}$**	**40788-84-9**	**flavone C-glycoside**	**211**
U164	**Kaempferol**	**$C_{15}H_{10}O_6$**	**520-18-3**	**flavone-3-ol**	**211**
U165	**Daidzein**	**$C_{15}H_{10}O_4$**	**486-66-8**	**isoflavone**	**211**
U166	**4-Methoxy-2H-pyran-2-one**	**$C_6H_6O_3$**	**100047-51-6**	**simple lactone**	**211**
U167	**Schaftoside**	**$C_{26}H_{28}O_{14}$**	**51938-32-0**	**flavone C-glycoside**	**211**
U168	**Isoschaftoside**	**$C_{26}H_{28}O_{14}$**	**52012-29-0**	**flavone C-glycoside**	**211**
U169	**Glycyuralin A**	**$C_{26}H_{34}O_6$**	**1879910-22-1**	**prenyl-pyran-isoflavan**	**211**
U170	**Glycyuralin C**	**$C_{26}H_{32}O_6$**	**1879910-24-3**	**hydroxyprenyl-pyran-isoflavan**	**211**
U171	**Angustone a**	**$C_{25}H_{26}O_6$**	**90686-13-8**	**diprenyl-isoflavone**	**211**
U173	**no trivial name**	**$C_{43}H_{62}O_{16}$**	**1874227-82-3**	**saponin glycoside**	**211**
U174	**Dehydroglyasperin D**	**$C_{22}H_{24}O_5$**	**517885-72-2**	**prenyl-dehydroisoflavan**	**211**
U175	**Glyasperin C**	**$C_{21}H_{24}O_5$**	**142474-53-1**	**prenyl-isoflavan**	**211**
U176	**Dehydroglyasperin C**	**$C_{21}H_{22}O_5$**	**199331-35-6**	**prenyl-dehydroisoflavan**	**250**
U177	**Glicophenone**	**$C_{20}H_{22}O_6$**	**303175-66-8**		**211**
U178	**7-O-Methylluteone**	**$C_{21}H_{20}O_6$**	**122290-50-0**	**prenylated-isoflavone**	**211**
U179	**1307578-72-8**	**$C_{21}H_{20}O_6$**	**1307578-72-8**	**prenyl-flavone-3-ol**	**211**
U180	**no trivial name**	**$C_{42}H_{62}O_{15}$**	**1874227-83-4**	**saponin glycoside**	**211**
U181	**Glyurallin a**	**$C_{21}H_{20}O_5$**	**199331-36-7**	**prenyl pterocarpan**	**211**
U182	**Hirtellanine**	**$C_{21}H_{18}O_6$**	**1369378-20-0**	**pyrano-isoflavone**	**211**
U183	**11b-Hydroxy-11b,1-dihydromedicarpin**	**$C_{16}H_{16}O_5$**	**210537-04-5**	**pterocarpan!**	**211**
U184	**1622179-42-3**	**$C_{33}H_{38}O_{19}$**	**1622179-42-3**	**flavone C-glycoside**	**211**
U185	**Glycyroside**	**$C_{27}H_{30}O_{13}$**	**125310-04-5**	**isoflavone glycoside**	**211**
U186	**2'-Hydroxy-isolupalbigenin**	**$C_{25}H_{26}O_6$**	**121747-94-2**	**diprenyl-isoflavone**	**211**
U187	**Isolupalbigenin**	**$C_{25}H_{26}O_5$**	**162616-70-8**	**diprenyl-isoflavone**	**211**
U189	**ZA13-Gu**	**$C_{20}H_{16}O7$**	**ZA13-Gu**	**miscellaneous**	**NCNPR**
U190	**Conferol A**	**$C_{16}H_{16}O_5$**	**1172120-53-4**	**isoflavanol**	**NCNPR**
U191	**ZA9-Gu**	**$C_{21}H_{24}O_6$**	**ZA9-Gu**	**prenyl-isoflavanol**	**NCNPR**

SI 5. *Glycyrrhiza glabra* top scoring compounds sorted by MM-GBSA

Rank	No.	Isoform preference	X-ray crystal structure	SP docking score	MM-GBSA energy
1	G163	α	2p15	−10.79	−105.92
2	G28	α	2p15	−10.43	−105.86
3	G106	α	2p15	−10.54	−105.34
4	G3	α	2p15	−11.04	−105.11
5	G40	α	2p15	−10.83	−104.52
6	G138	α	2p15	−10.55	−101.07
7	G156	α	2p15	−10.38	−100.18
8	G112	α	2p15	−10.41	−97.60
9	G38	α	2p15	−10.87	−96.96
10	G174	α	2p15	−10.84	−96.31
11	G163	β	1L2J	−11.35	−96.15
12	G33	α	2p15	−10.49	−95.93
13	G108	α	2p15	−10.53	−95.76
14	G2	β	1L2J	−10.45	−95.35
15	G28	β	1L2J	−10.54	−94.89
16	G21	α	2p15	−10.47	−94.64
17	G50	α	2p15	−10.95	−94.43
18	G127	α	2p15	−10.80	−93.82
19	G158	β	1L2J	−11.73	−93.19
20	G47	α	2p15	−10.82	−93.11
21	G30	α	2p15	−10.37	−93.06
22	G40	β	1L2J	−10.43	−91.87
23	G170	α	2p15	−10.30	−90.75
24	G154	α	2p15	−10.32	−90.73
25	G75	β	1X7J	−10.74	−90.66
26	G38	β	1X7J	−10.92	−90.10
27	G111	β	1L2J	−10.58	−89.09
28	G99	β	1L2J	−11.00	−88.50
29	G100	β	1X7J	−11.06	−86.56
30	G171	α	1GWR	−10.33	−85.01

SI 6. *Glycyrrhiza uralensis* top scoring compounds sorted by MM−GBSA

Rank	No.	selectivity	X-ray Crystal structure	SP docking score	MM-GBSA energy
1	U77	α	2p15	−11.05	−122.91
2	U15	α	2p15	−11.49	−111.57
3	U16	β	1L2J	−11.03	−111.36
4	U191	α	2p15	−10.84	−105.61
5	U78	α	2p15	−10.19	−105.11
6	U25	α	2p15	−9.92	−100.44
7	U22	α	2p15	−10.26	−100.04
8	U86	β	1L2J	−10.58	−98.73
9	U23	β	1L2J	−11.14	−98.39
10	U24	α	2p15	−10.39	−98.18
11	U176	β	1L2J	−10.33	−97.53
12	U57	β	1L2J	−10.22	−96.81
13	U66	α	2p15	−11.15	−95.79
14	U63	α	1GWR	−11.88	−95.75
15	U127	α	2p15	−10.66	−93.42
16	U175	β	1L2J	−10.85	−92.92
17	U117	β	1L2J	−10.91	−92.15
18	U115	β	1L2J	−10.47	−91.79
19	U177	α	2p15	−10.12	−91.71
20	U62	β	1L2J	−10.15	−91.59
21	U77	β	1L2J	−10.47	−91.24
22	U88	β	1L2J	−10.72	−91.07
23	U119	β	1X7J	−10.15	−89.28
24	U127	β	1L2J	−10.30	−88.71
25	U105	β	1X7J	−10.23	−88.66
26	U189	α	2p15	−9.95	−86.80
27	U105	α	1GWR	−10.28	−84.87
28	U165	β	1X7J	−10.49	−84.09
29	U69	β	1X7J	−10.59	−83.22
30	U61	α	2p15	−9.78	−82.03

SI 7. Predicted RBA for top 20 compounds in both *G. glabra* and *G. uralensis* that scored (>70%) confidence in estro_filter as calculated in ADMET Predictor™.

Rank	No.	Estro_Filter	Estro_RBA	Rank	No.	Estro_Filter	Estro_RBA
1	**G94**	**Toxic (97%)**	**758.16**	**1**	**U35**	**Toxic (97%)**	**330.74**
2	**G117**	**Toxic (97%)**	**699.12**	**2**	**U169**	**Toxic (89%)**	**268.66**
3	**G41**	**Toxic (97%)**	**364.00**	**3**	**U175**	**Toxic (97%)**	**184.89**
4	**G152**	**Toxic (97%)**	**138.98**	**4**	**U170**	**Toxic (76%)**	**133.40**
5	**G96**	**Toxic (97%)**	**108.12**	**5**	**U147**	**Toxic (73%)**	**93.40**
6	**G61**	**Toxic (97%)**	**98.37**	**6**	**U176**	**Toxic (97%)**	**75.16**
7	**G97**	**Toxic (97%)**	**56.62**	**7**	**U123**	**Toxic (97%)**	**75.05**
8	**G49**	**Toxic (97%)**	**42.985**	**8**	**U191**	**Toxic (97%)**	**46.29**
9	**G175**	**Toxic (76%)**	**37.22**	**9**	**U15**	**Toxic (97%)**	**46.17**
10	**G21**	**Toxic (97%)**	**36.00**	**10**	**U22**	**Toxic (97%)**	**39.45**
11	**G40**	**Toxic (97%)**	**34.13**	**11**	**U16**	**Toxic (97%)**	**39.10**
12	**G69**	**Toxic (89%)**	**33.32**	**12**	**U110**	**Toxic (76%)**	**38.06**
13	**G38**	**Toxic (97%)**	**25.32**	**13**	**U174**	**Toxic (89%)**	**36.04**
14	**G28**	**Toxic (76%)**	**16.33**	**14**	**U98**	**Toxic (97%)**	**31.45**
15	**G174**	**Toxic (97%)**	**12.44**	**15**	**U187**	**Toxic (73%)**	**31.26**
16	**G3**	**Toxic (73%)**	**10.55**	**16**	**U117**	**Toxic (97%)**	**29.87**
17	**G159**	**Toxic (97%)**	**10.034**	**17**	**U104**	**Toxic (76%)**	**29.17**
18	**G47**	**Toxic (97%)**	**8.77**	**18**	**U99**	**Toxic (97%)**	**28.60**
19	**G19**	**Toxic (97%)**	**8.40**	**19**	**U23**	**Toxic (97%)**	**27.93**
20	**G73**	**Toxic (97%)**	**6.88**	**20**	**U186**	**Toxic (84%)**	**24.72**

SI 8. *Glycyrrhiza glabra* top-scoring compounds against PXR (1NRL) crystal structure sorted by XP score in kcal/mol

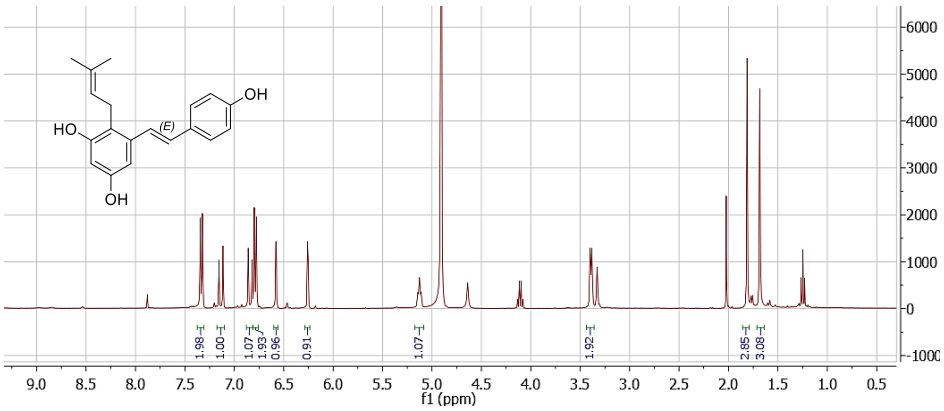
Rank	No.	Xp Score	Glide emodel	Rank	No.	Xp Score	Glide emodel
1	G158	−13.36	−82.20	33	G106	−10.06	−67.56
2	G136	−12.59	−57.69	34	G101	−10.02	−66.64
3	G146	−12.53	−72.39	35	G159	−9.98	−45.66
4	G172	−12.52	−70.72	36	G159	−9.98	−9.98
5	G83	−12.33	−88.91	37	G137	−9.98	−9.98
6	G163	−12.18	−75.52	38	G68	−9.97	−9.97
7	G96	−11.95	−80.85	39	G140	−9.94	−9.94
8	G33	−11.91	−66.11	40	G94	−9.91	−9.91
9	G135	−11.69	−74.52	41	G110	−9.96	−9.96
10	G60	−11.58	−87.77	42	G170	−9.89	−9.89
11	G1	−11.54	−41.99	43	G97	−9.88	−9.88
12	G28	−11.26	−74.99	44	G30	−9.84	−9.84
13	G40	−11.09	−69.09	45	G71	−9.83	−9.83
14	G42	−10.99	−72.37	46	G134	−9.82	−9.82
15	G164	−10.99	−60.34	47	G100	−9.84	−9.84
16	G152	−10.91	−35.20	48	G131	−9.74	−9.74
17	G141	−10.80	−62.74	49	G19	−9.70	−9.70
18	G99	−10.87	−74.75	50	G111	−9.69	−9.69
19	G64	−10.72	−80.71	51	G32	−9.69	−9.69
20	G117	−10.72	−58.66	52	G65	−9.67	−9.67
21	G41	−10.56	−67.82	53	G62	−9.65	−9.65
22	G2	−10.47	−65.61	54	G53	−9.60	−9.60
23	G48	−10.42	−71.42	55	G102	−9.60	−9.60
24	G109	−10.46	−77.76	56	G138	−9.63	−9.63
25	G61	−10.28	−20.10	57	G132	−9.58	−9.58
26	G98	−10.25	−68.26	58	G139	−9.56	−9.56
27	G108	−10.24	−66.42	59	G35	−9.55	−9.55
28	G21	−10.21	−57.29	60	G88	−9.54	−9.54
29	G90	−10.21	−41.13	61	G91	−9.54	−9.54
30	G142	−10.21	−64.10	62	G69	−9.52	−9.52
31	G144	−10.12	−63.78	63	G154	−9.50	−9.50
32	G3	−10.07	−63.40				

SI 9. *Glycyrrhiza uralensis* top-scoring compounds against PXR (1NRL) crystal structure sorted by XP score in kcal/mol

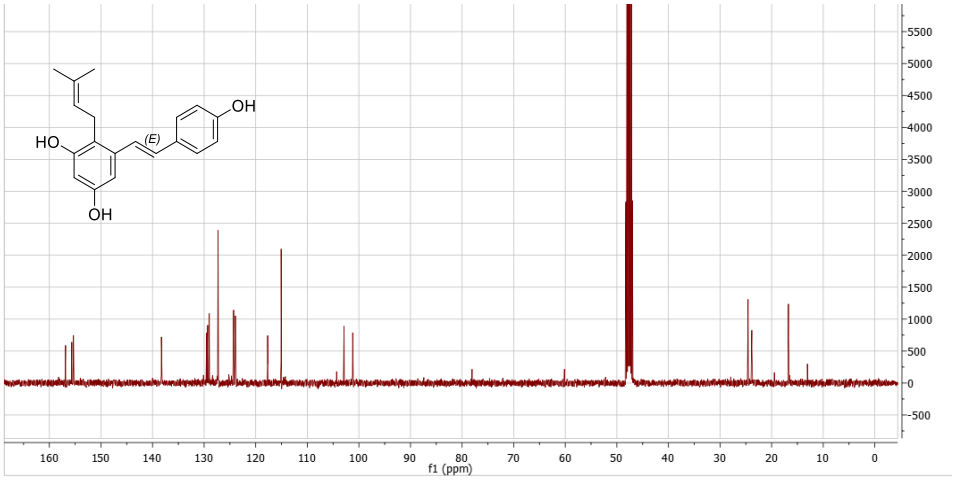
Rank	No	XP score	Emodel score	Rank	No	XP score	Emodel score
1	**U77**	**−12.26**	**−61.88**	**29**	**U54**	**−10.07**	**−53.88**
2	**U186**	**−12.18**	**−75.86**	**30**	**U34**	**−10.04**	**−52.24**
3	**U121**	**−11.72**	**−62.73**	**31**	**U128**	**−10.04**	**−61.43**
4	**U79**	**−11.32**	**−34.46**	**32**	**U106**	**−10.03**	**−67.06**
5	**U52**	**−11.16**	**−20.64**	**33**	**U191**	**−10.00**	**−66.84**
6	**U187**	**−11.06**	**−73.77**	**34**	**U88**	**−9.99**	**−70.15**
7	**U111**	**−11.00**	**−16.83**	**35**	**U105**	**−10.03**	**−58.39**
8	**U51**	**−10.88**	**−63.85**	**36**	**U60**	**−9.95**	**−73.45**
9	**U189**	**−10.87**	**−54.86**	**37**	**U117**	**−9.92**	**−60.62**
10	**U57**	**−10.72**	**−44.33**	**38**	**U178**	**−9.86**	**−66.60**
11	**U64**	**−10.71**	**−73.32**	**39**	**U18**	**−9.85**	**−68.63**
12	**U55**	**−10.69**	**−69.68**	**40**	**U84**	**−9.83**	**−65.02**
13	**U171**	**−10.67**	**−34.90**	**41**	**U155**	**−9.78**	**−63.13**
14	**U78**	**−10.66**	**−63.91**	**42**	**U81**	**−9.77**	**−61.65**
15	**U35**	**−10.60**	**−43.01**	**43**	**U126**	**−9.76**	**−67.29**
16	**U110**	**−10.60**	**−58.17**	**44**	**U36**	**−9.76**	**−52.58**
17	**U85**	**−10.53**	**−63.83**	**45**	**U189**	**−9.76**	**−70.68**
18	**U59**	**−10.46**	**−75.87**	**46**	**U153**	**−9.71**	**−67.16**
19	**U16**	**−10.45**	**−73.07**	**47**	**U170**	**−9.70**	**−10.11**
20	**U38**	**−10.47**	**−80.33**	**48**	**U19**	**−9.67**	**−64.20**
21	**U65**	**−10.40**	**−70.29**	**49**	**U174**	**−9.66**	**−73.89**
22	**U152**	**−10.37**	**−70.27**	**50**	**U179**	**−9.66**	**−68.80**
23	**U15**	**−10.34**	**−73.23**	**51**	**U50**	**−9.65**	**−19.40**
24	**U41**	**−10.44**	**−78.57**	**52**	**U63**	**−9.64**	**−70.95**
25	**U17**	**−10.21**	**−61.27**	**53**	**U41**	**−9.79**	**−76.19**
26	**U169**	**−10.21**	**−4.63**	**54**	**U87**	**−9.53**	**−68.63**
27	**U39**	**−10.27**	**−78.75**	**55**	**U25**	**−9.52**	**−53.86**
28	**U122**	**−10.20**	**−52.22**	**56**	**U178**	**−9.51**	**−65.99**

APPENDIX II.
SPECTRAL DATA

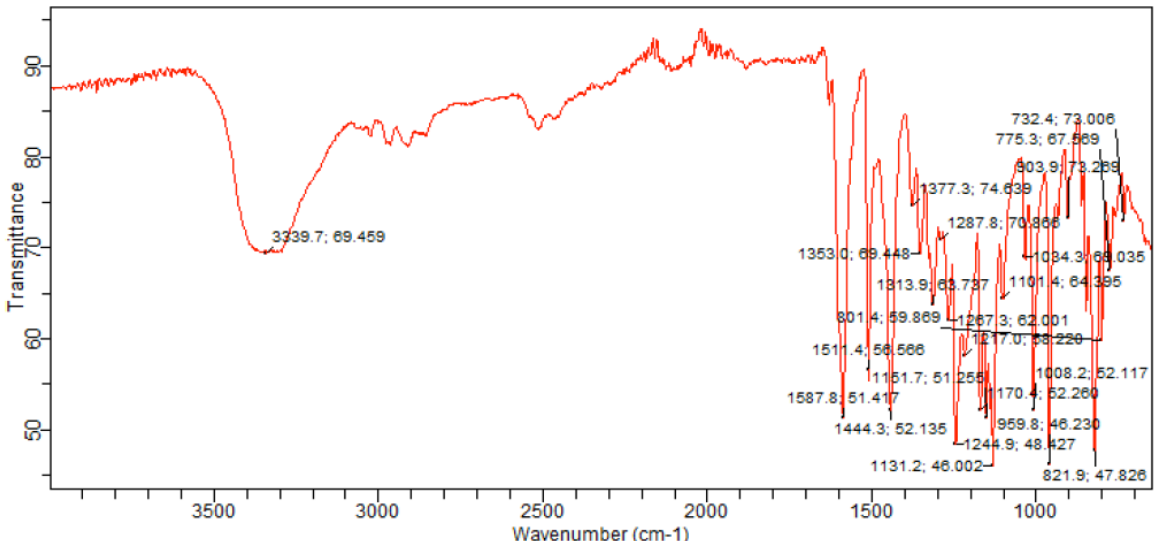
Resveratrol prenylation products (M2-M10)



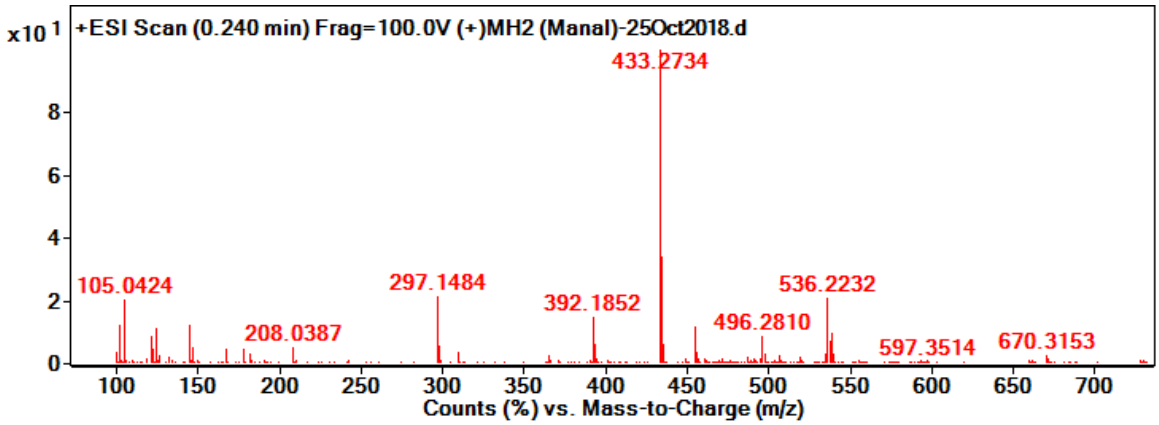
^{1}H NMR (MeOD, 400MHz) spectrum of M2.



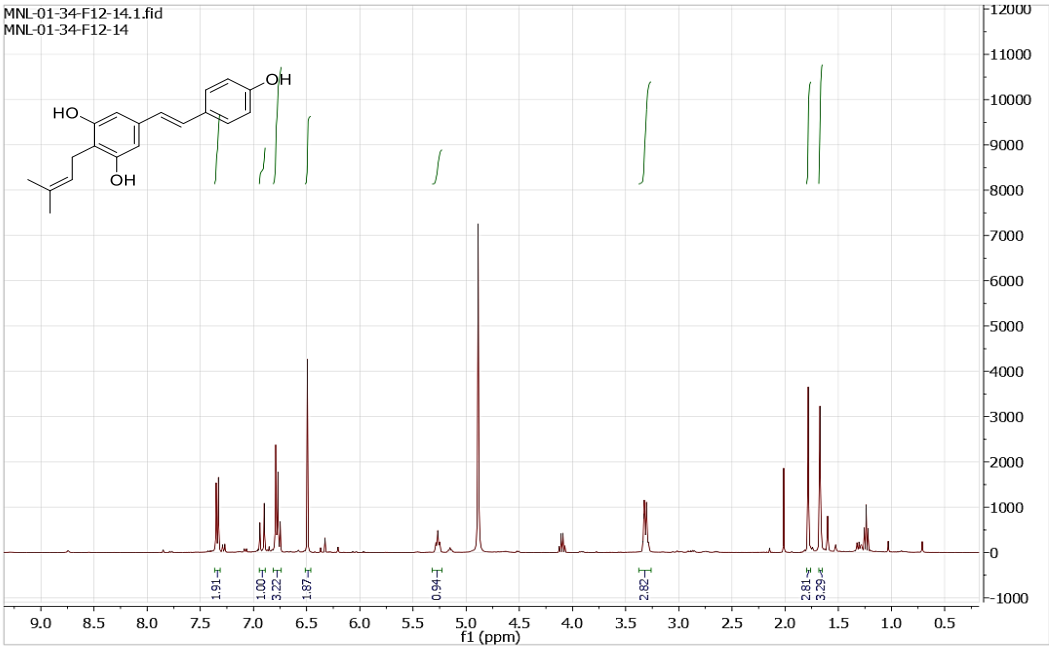
^{13}C NMR (MeOD, 100MHz) spectrum of compound M2



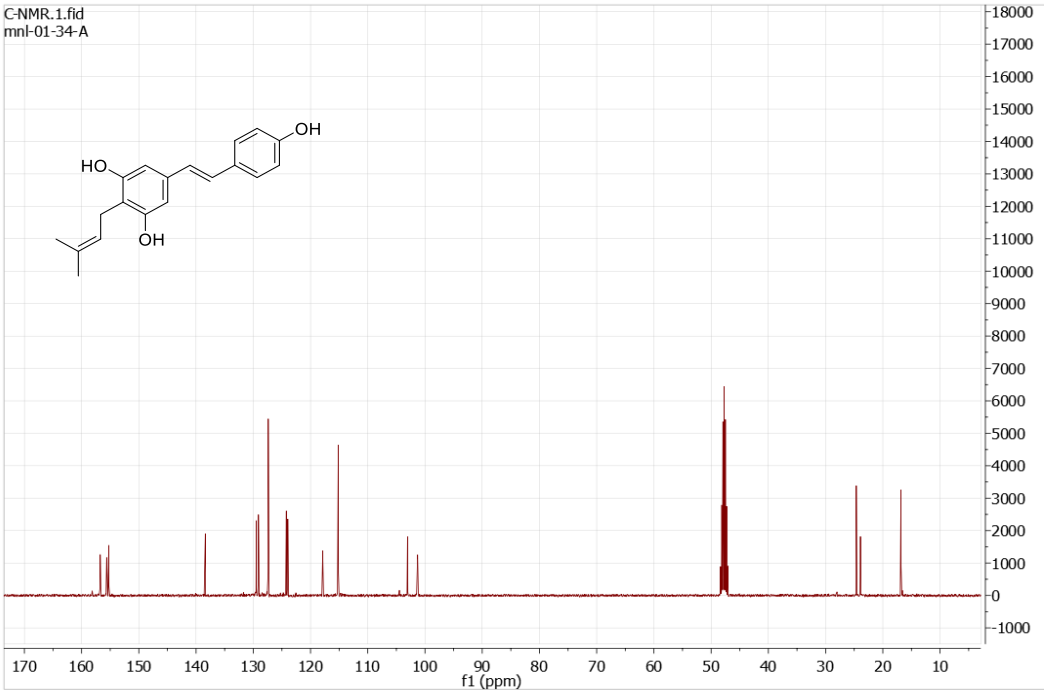
IR spectrum of compound M2



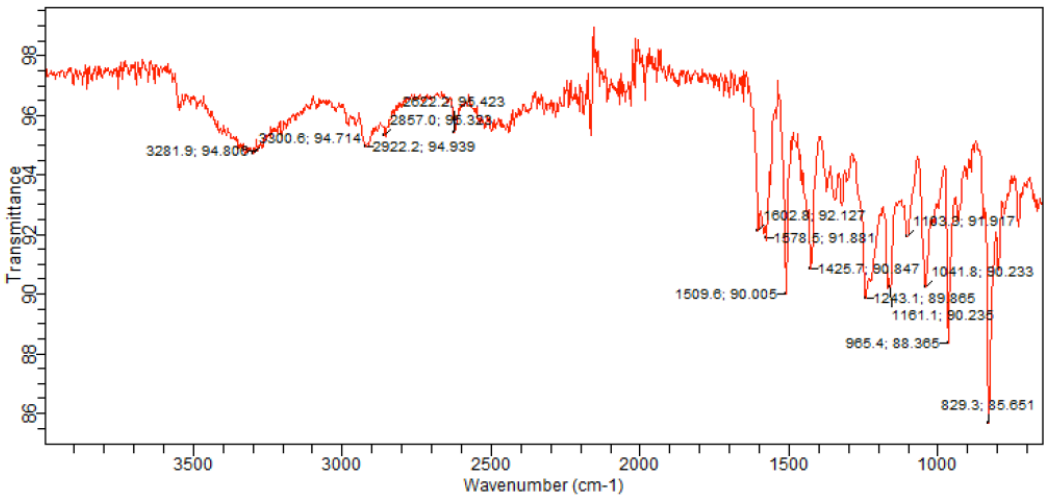
ESI-HRMS of compound M2



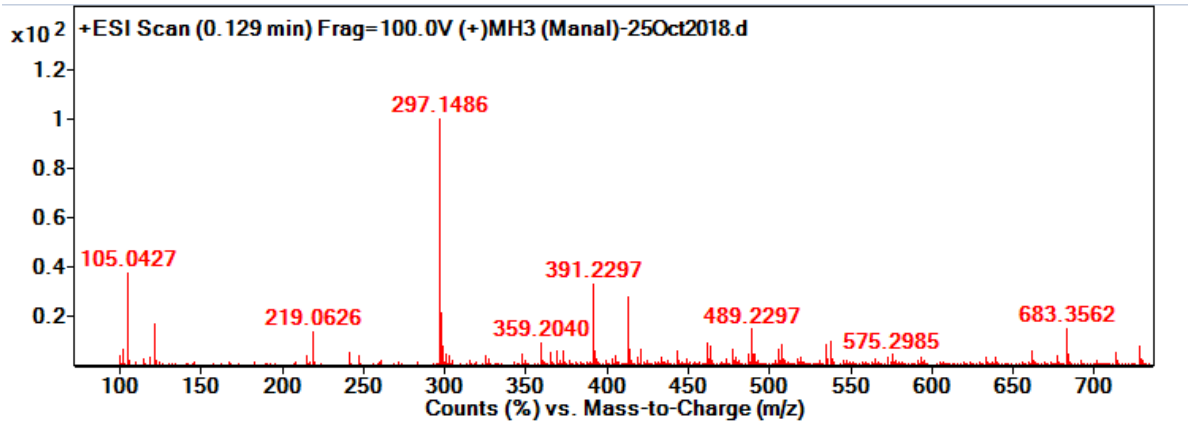
^{1}H NMR (MeOD, 400MHz) spectrum of M3



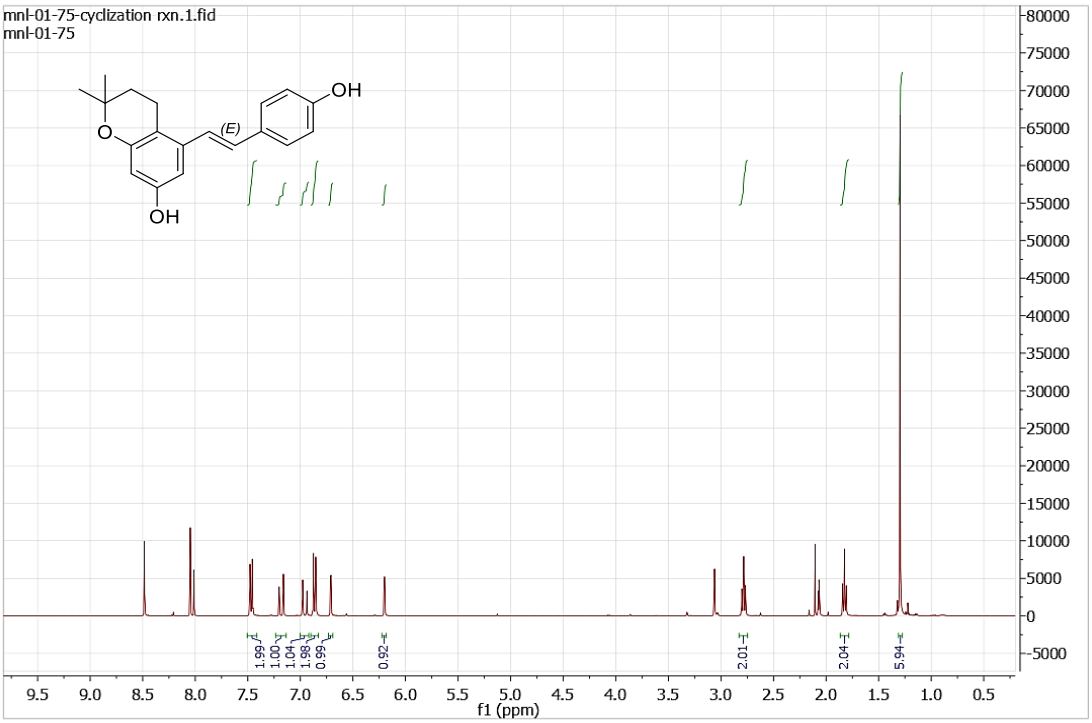
^{13}C NMR (MeOD, 100MHz) spectrum of compound M3



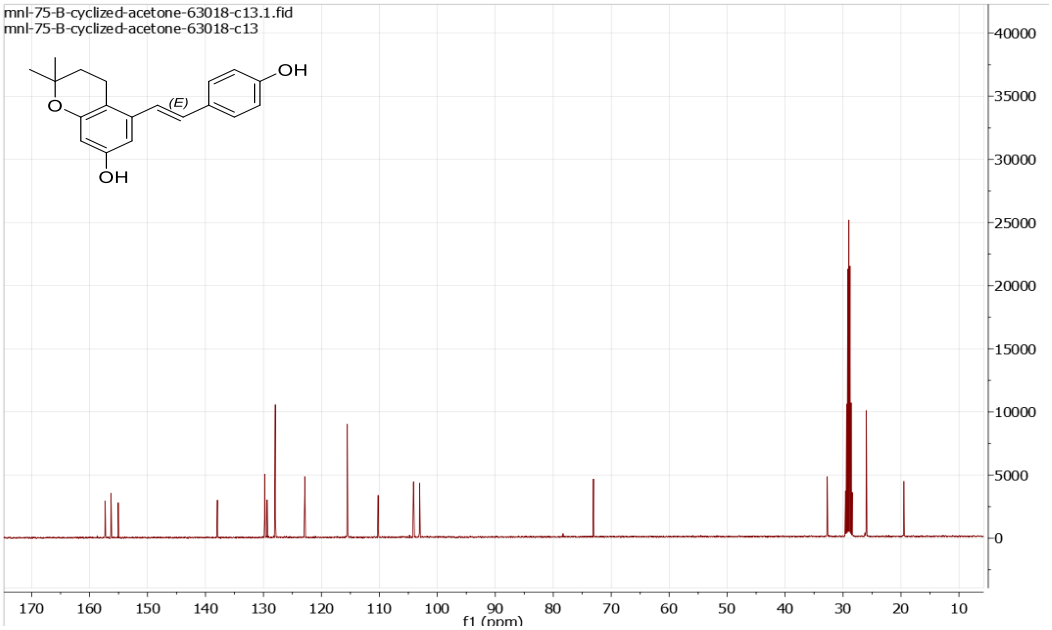
IR spectrum of compound M3



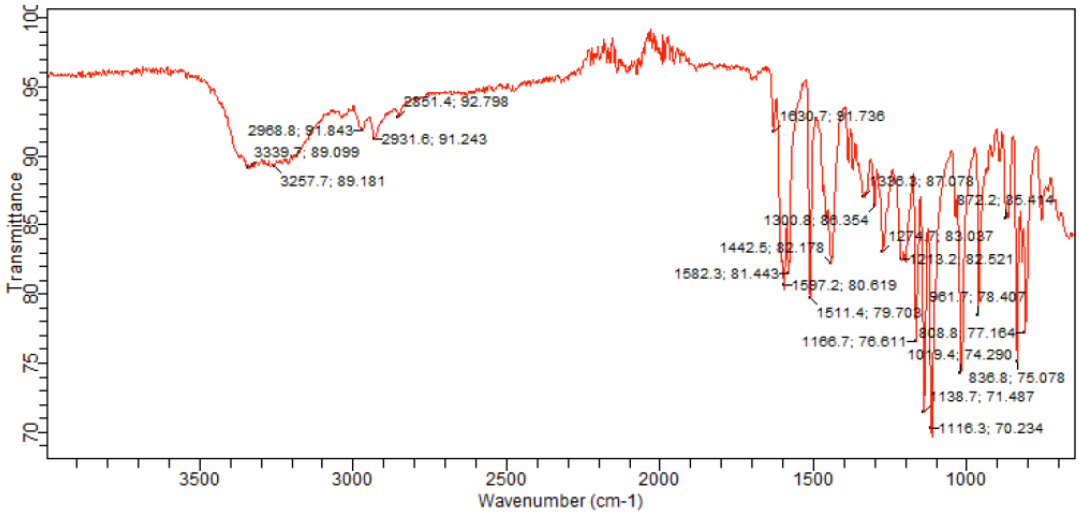
ESI-HRMS of compound M3



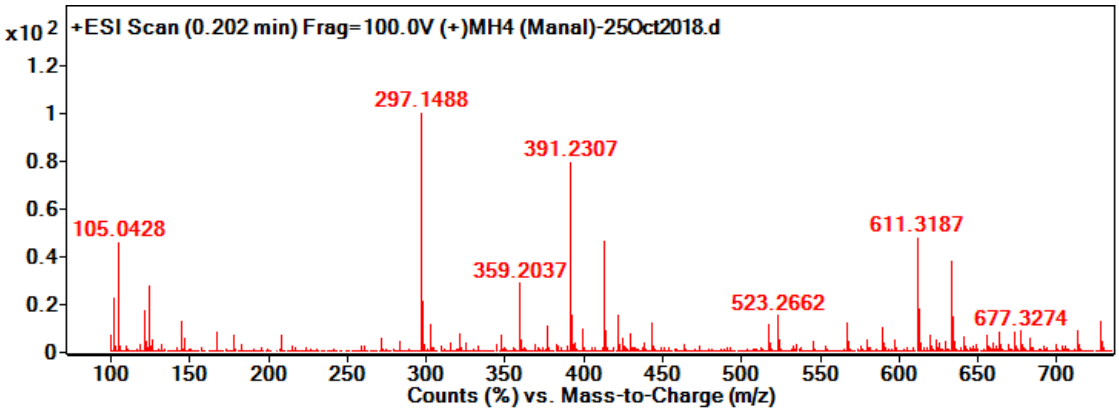
^{1}H NMR (acetone-*d6*, 400 MHz) spectrum of M4



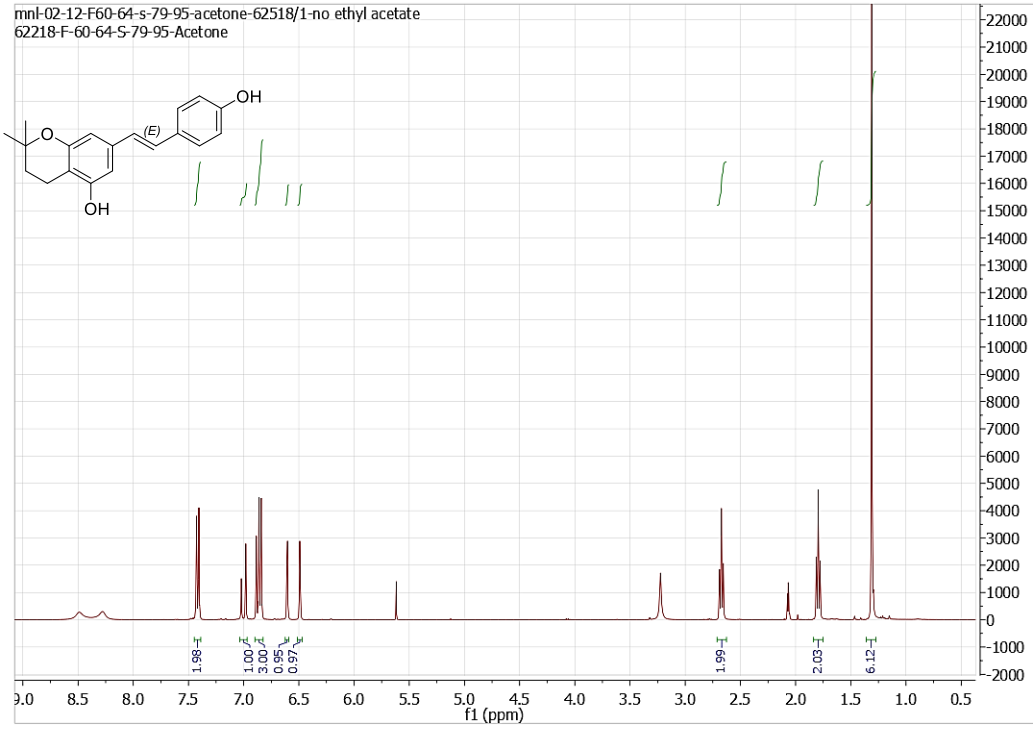
^{13}C NMR (acetone-*d6*, 100 MHz) spectrum of compound M4



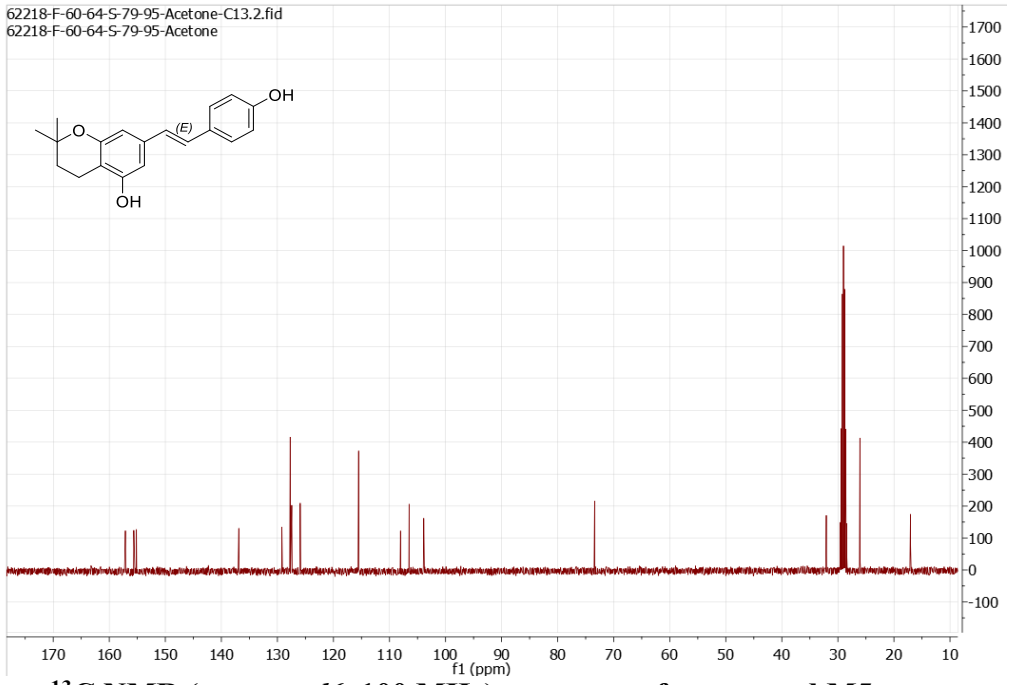
IR spectrum of compound M4



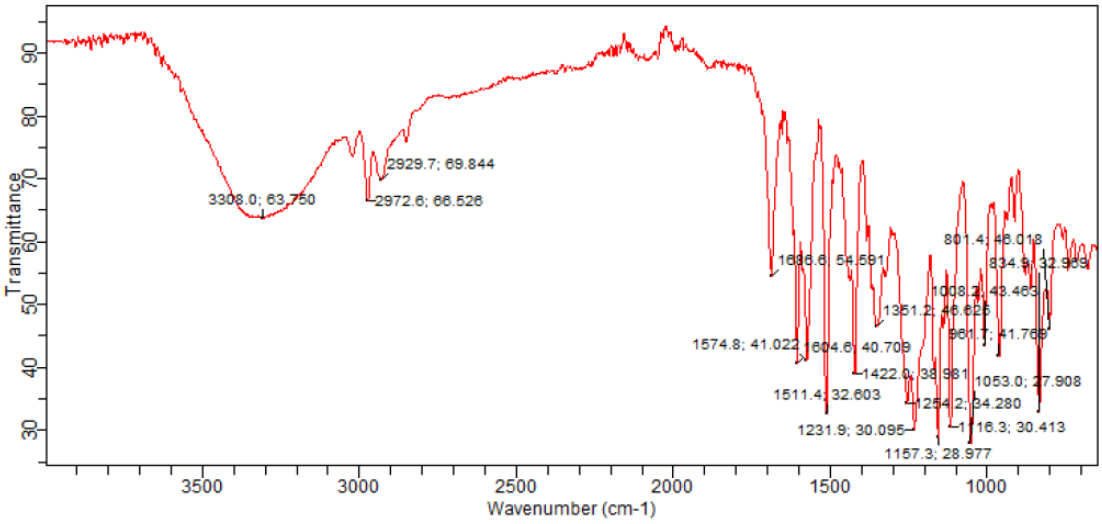
ESI-HRMS of compound M4



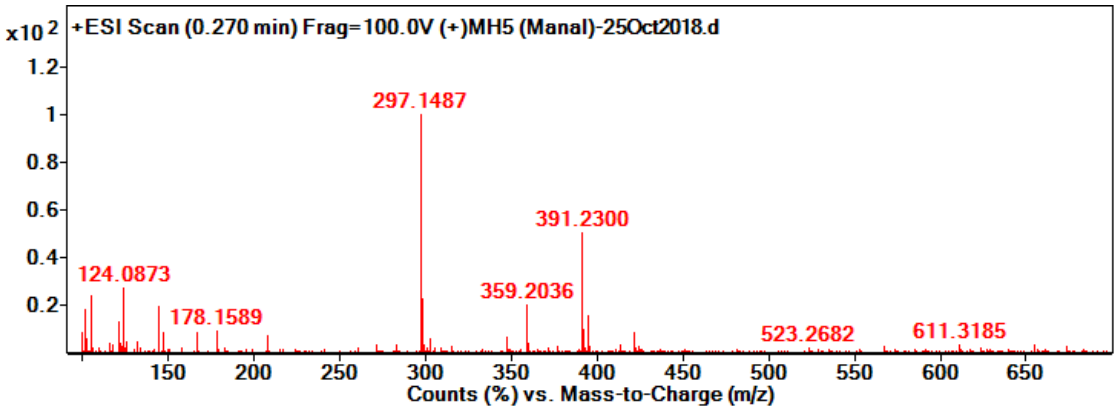
^{1}H NMR (acetone-*d6*, 400 MHz) spectrum of M5



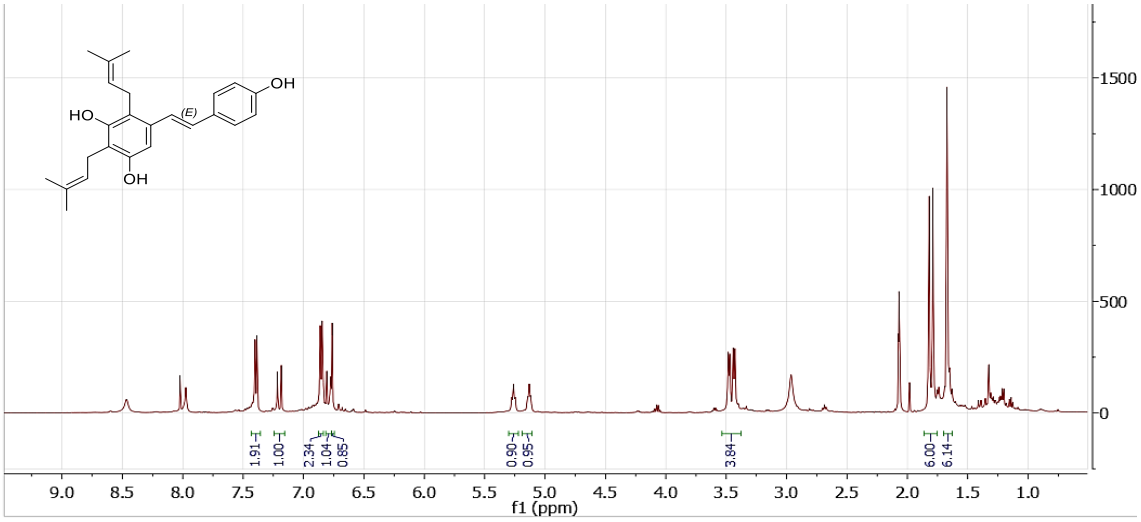
^{13}C NMR (acetone-*d6*, 100 MHz) spectrum of compound M5



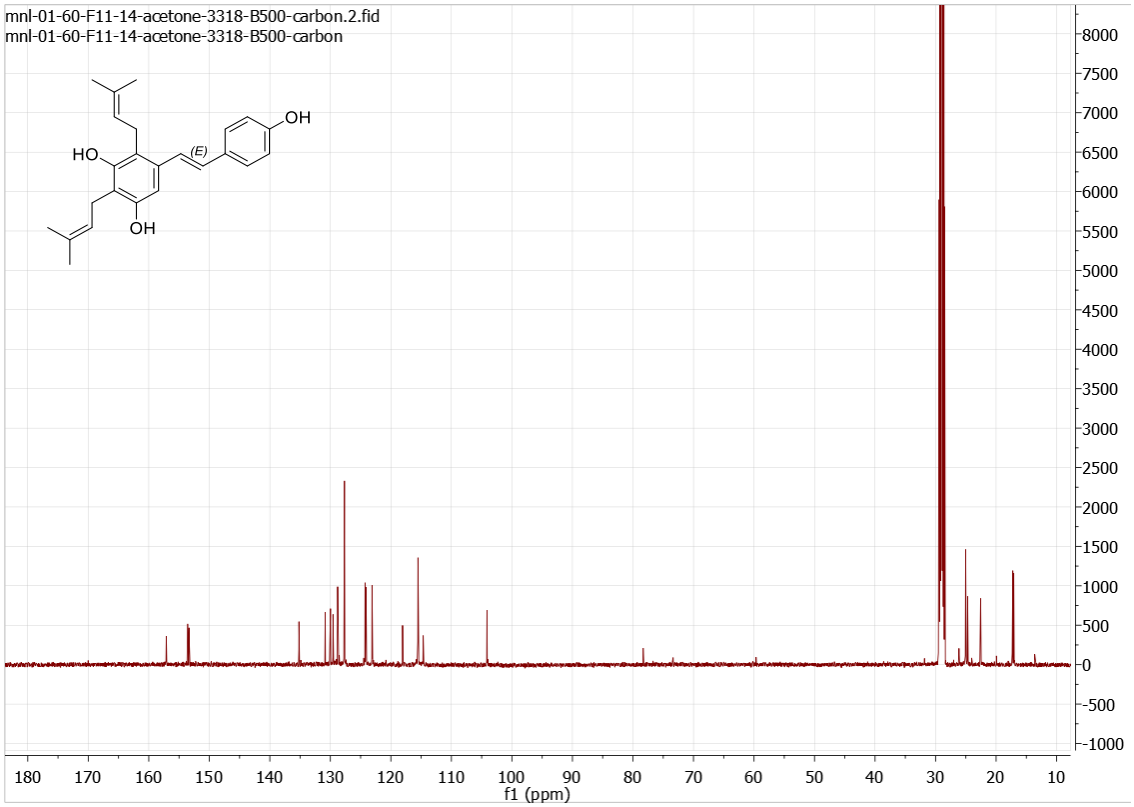
IR spectrum of compound M5



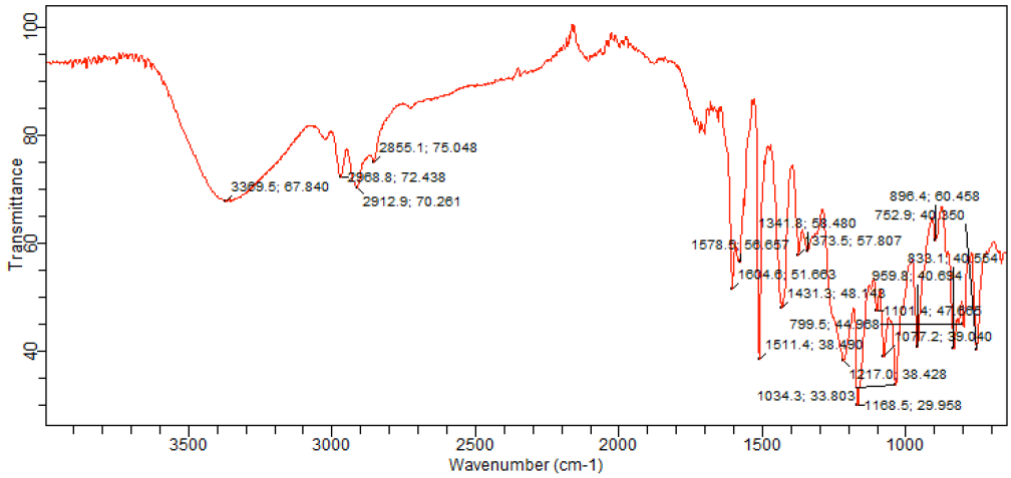
ESI-HRMS of compound M5



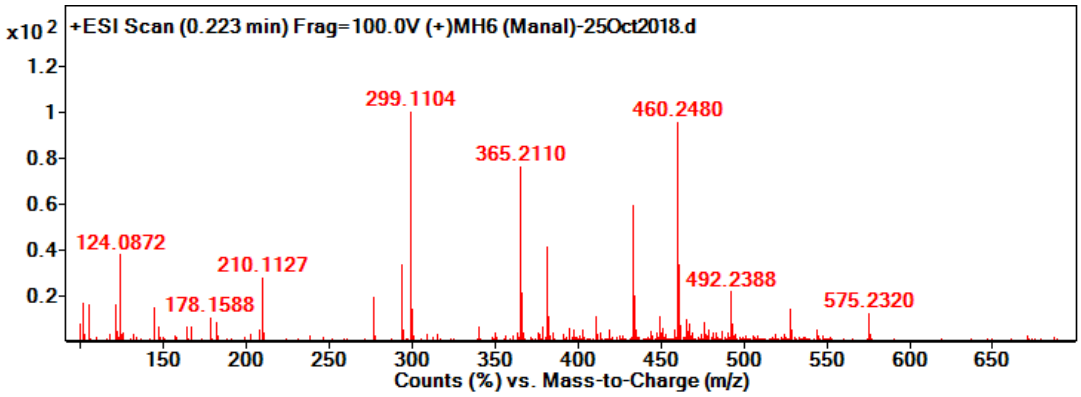
^{1}H NMR (acetone-*d6*, 500 MHz) spectrum of M6



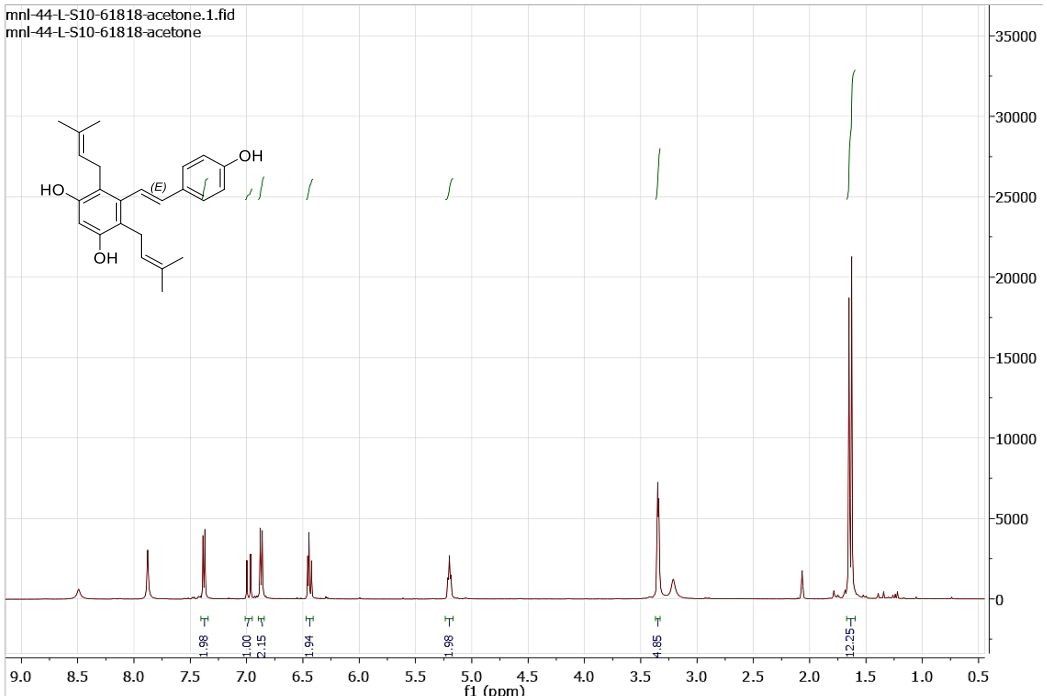
^{13}C NMR (acetone-*d6*, 125 MHz) spectrum of compound M6



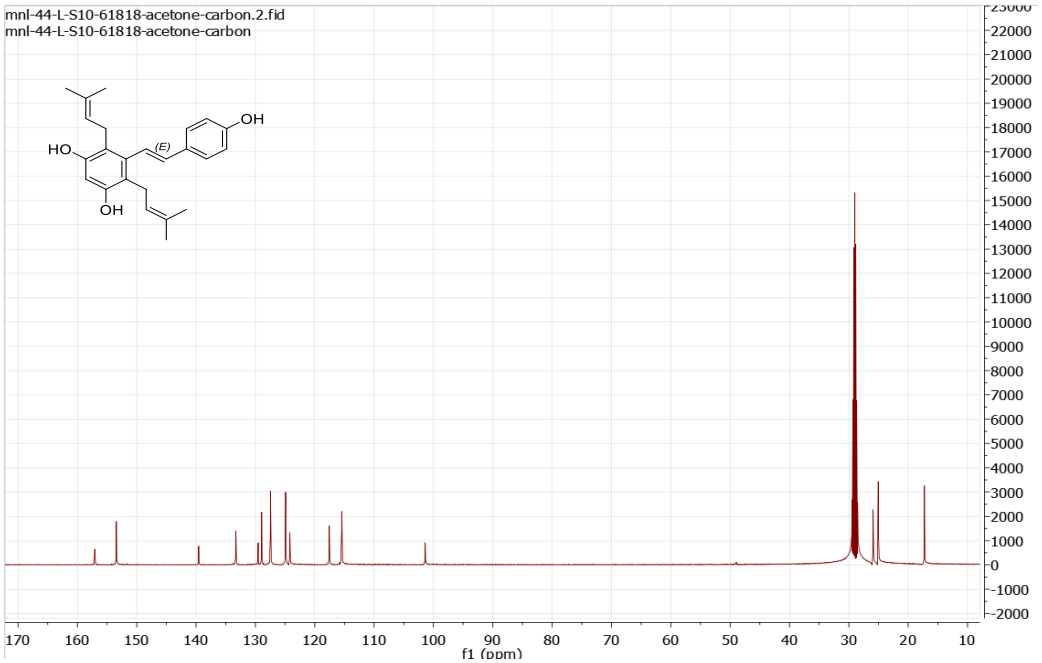
IR spectrum of compound M6



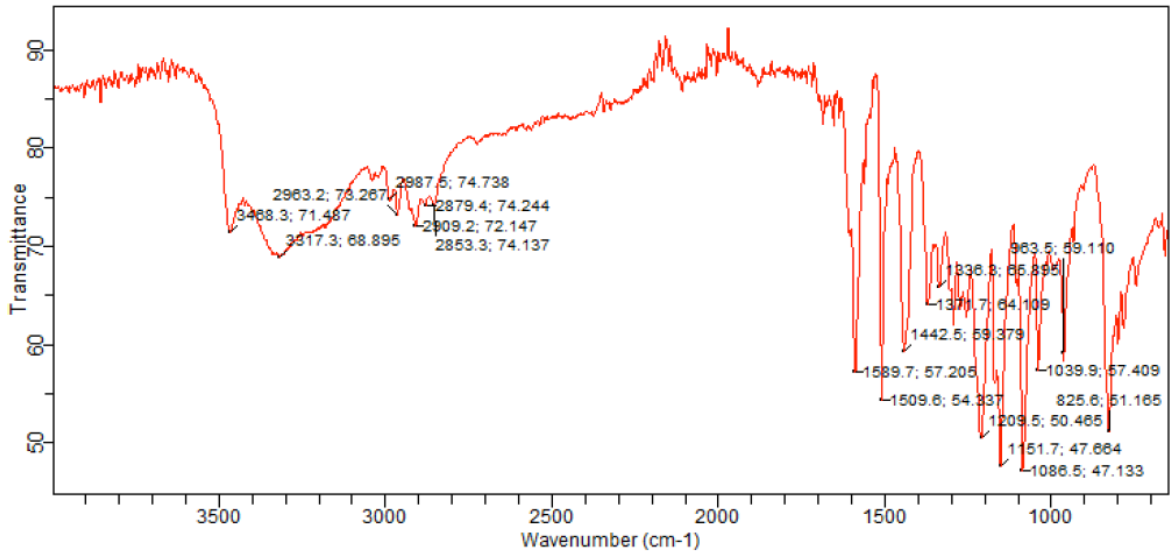
ESI-HRMS of compound M6



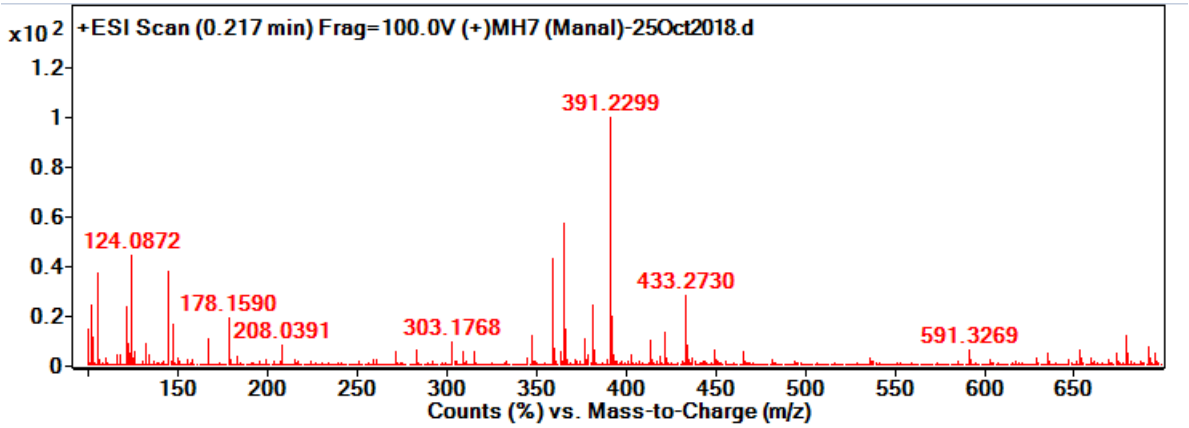
^{1}H NMR (acetone-*d6*, 500 MHz) spectrum of compound M7



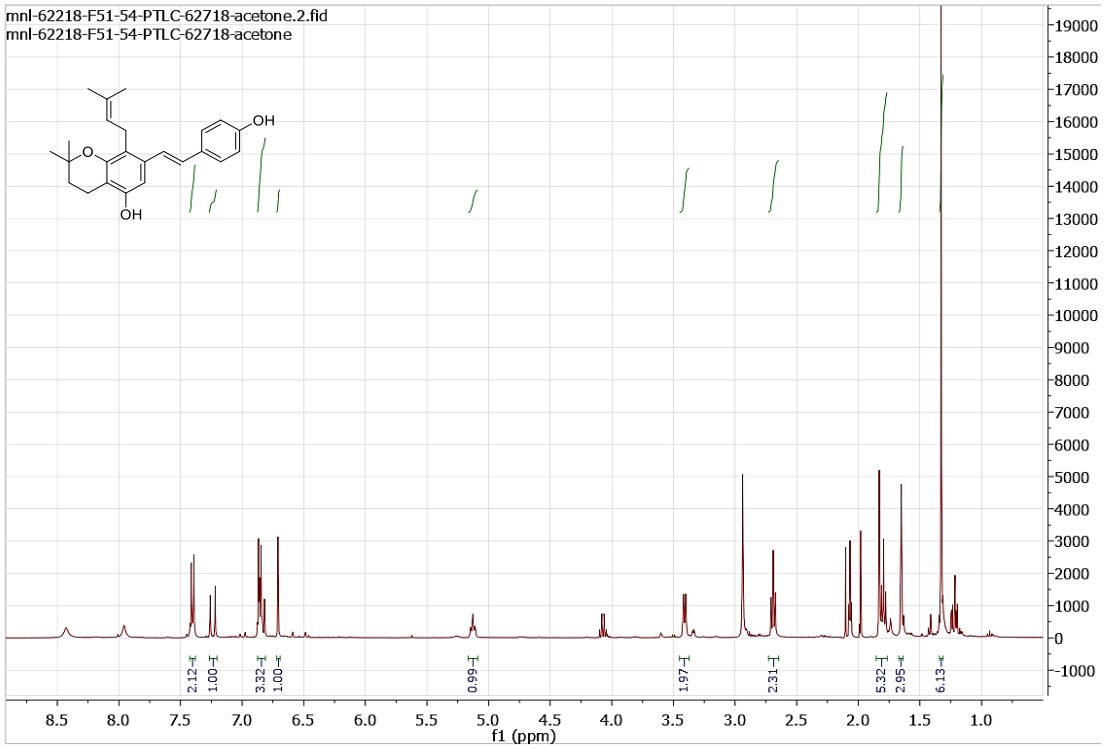
^{13}C NMR (acetone-*d6*, 125 MHz) spectrum of compound M7



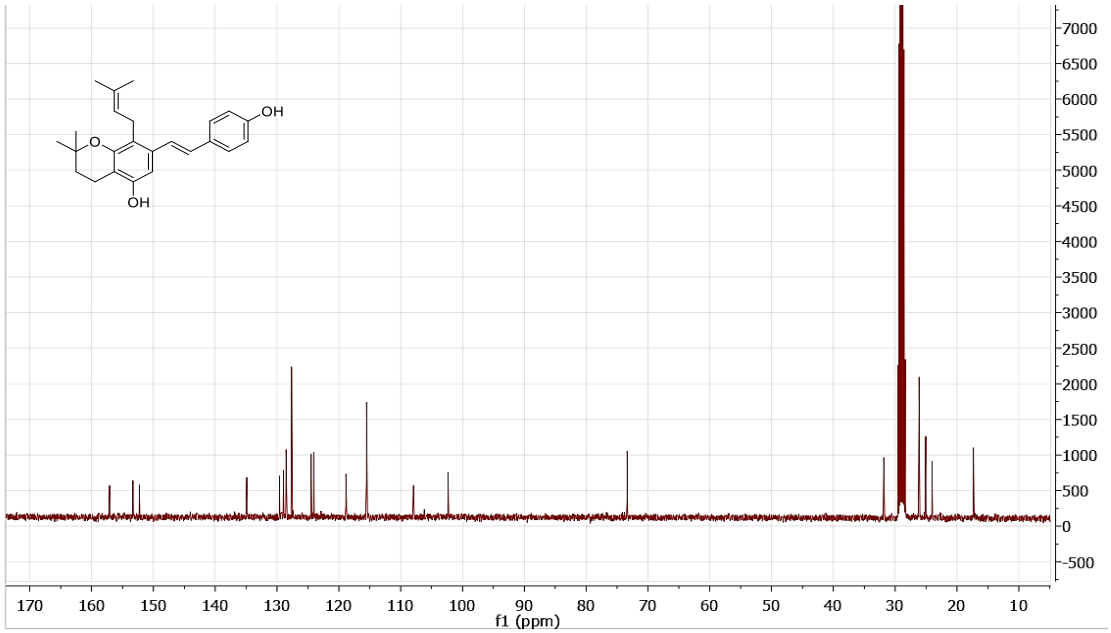
IR spectrum of compound M7



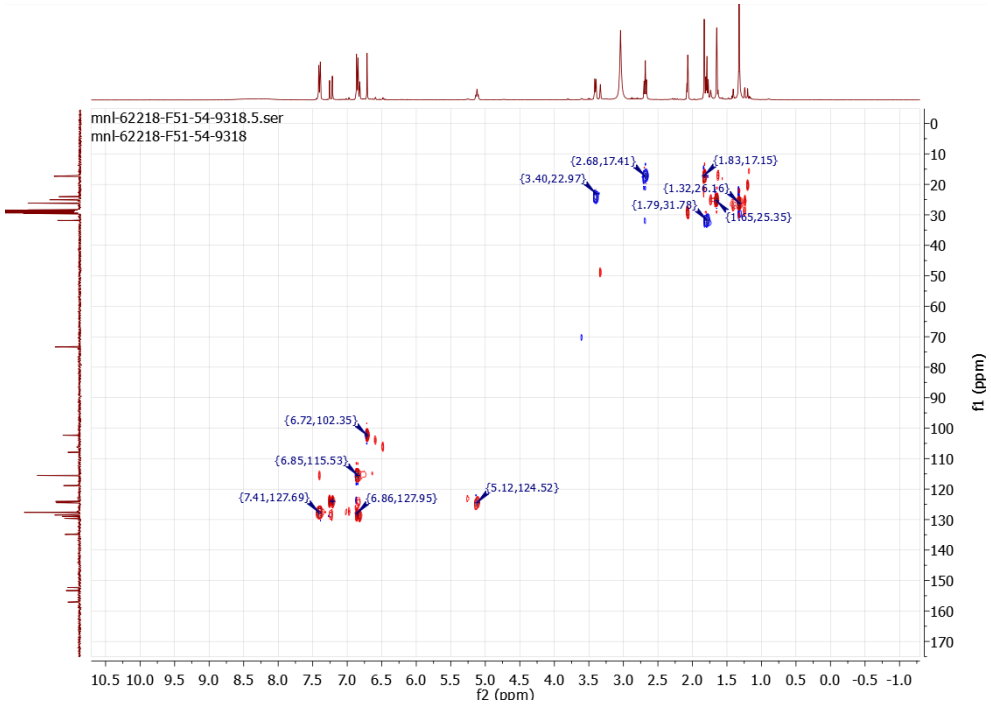
ESI-HRMS of compound M7



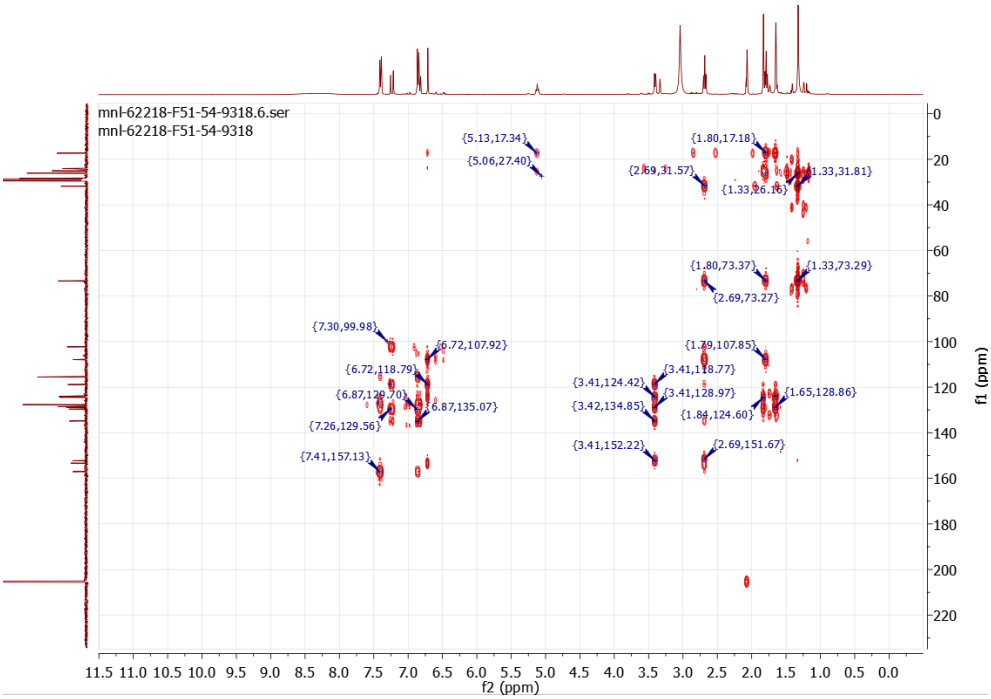
^{1}H NMR (acetone-*d6*, 400 MHz) spectrum of M8



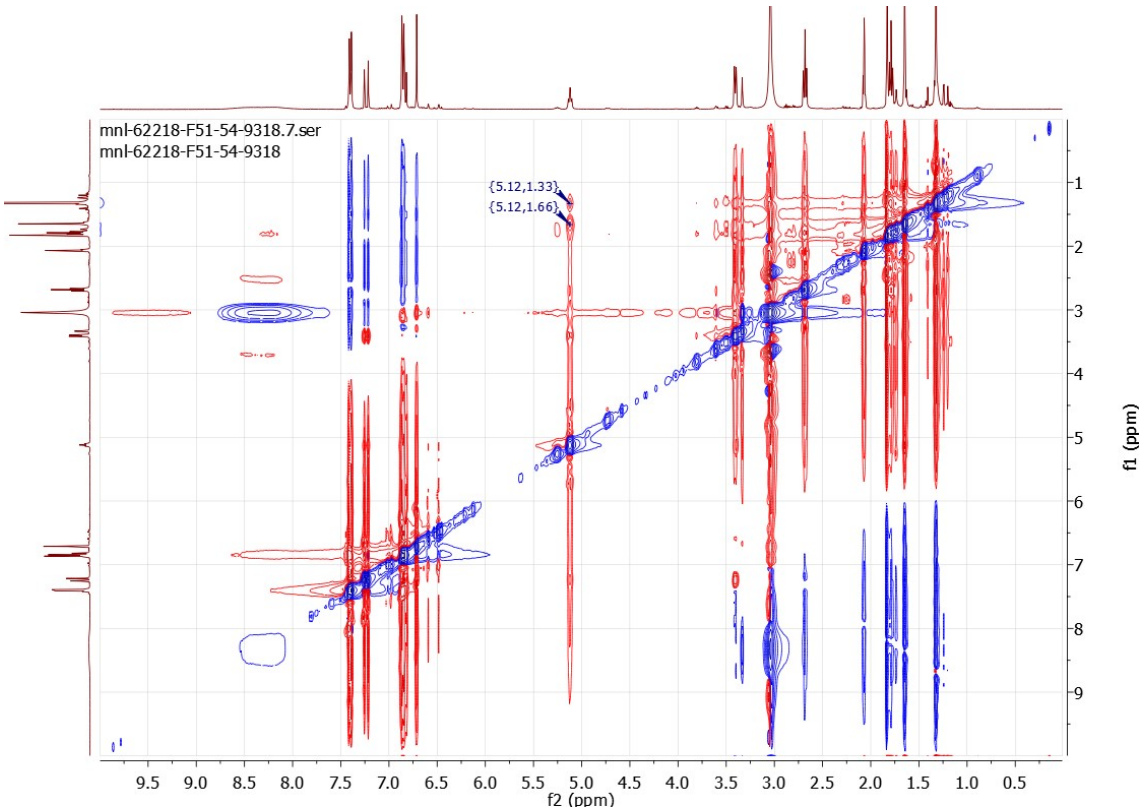
^{13}C NMR (acetone-*d6*, 100 MHz) spectrum of compound M8



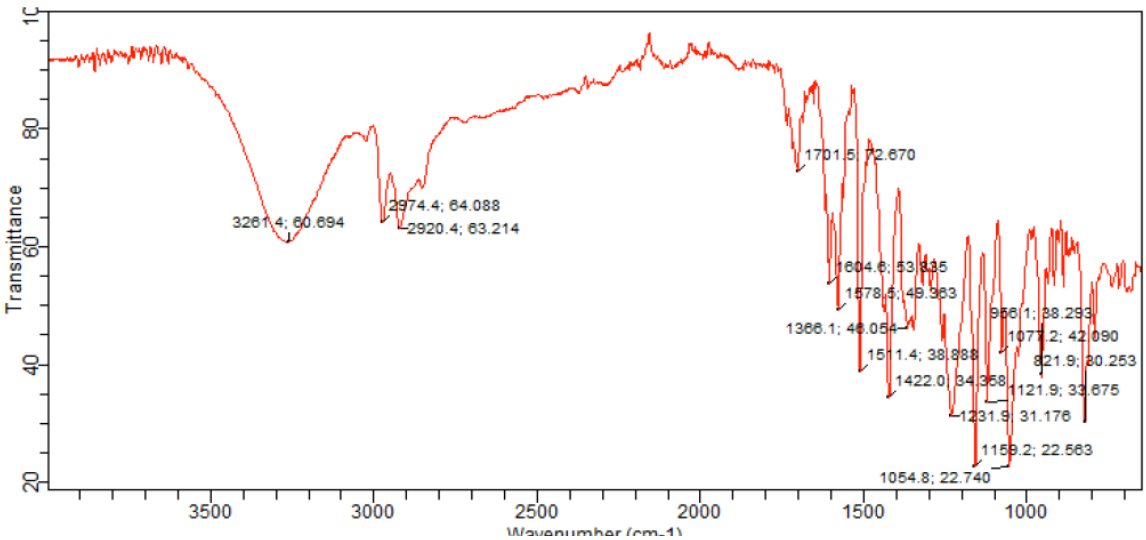
HSQC spectrum of compound M8



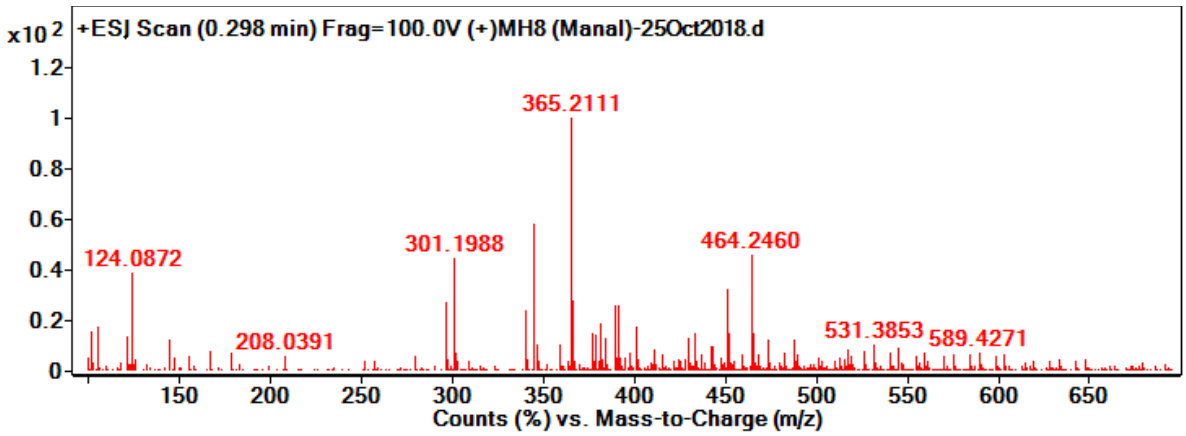
HMBC spectrum of compound M8



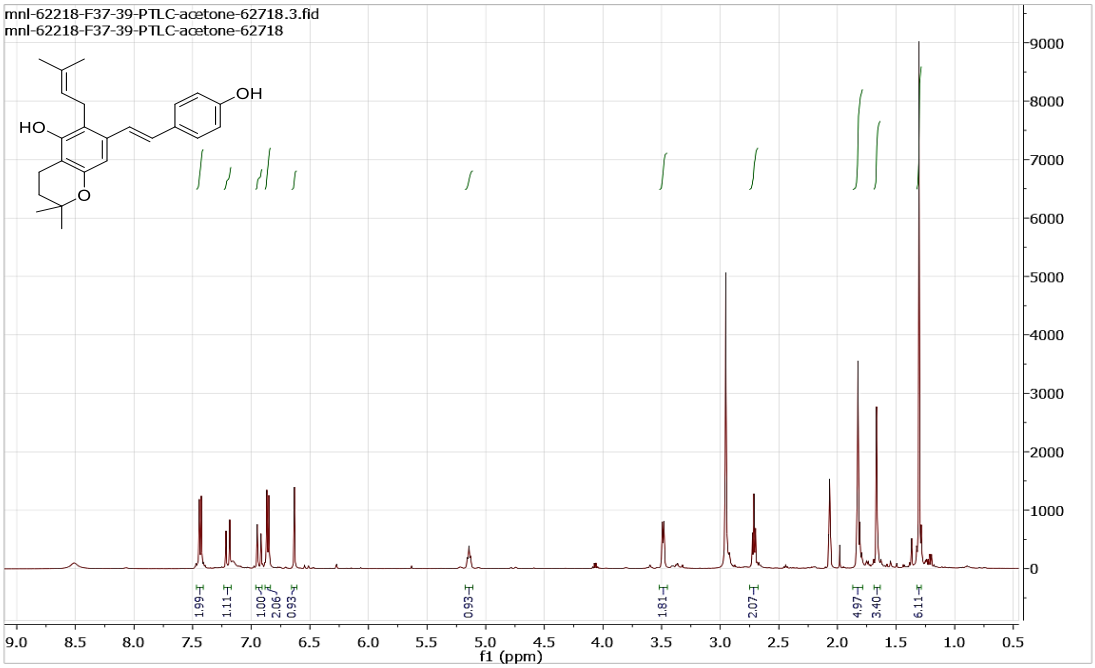
NOESY spectrum of compound M8



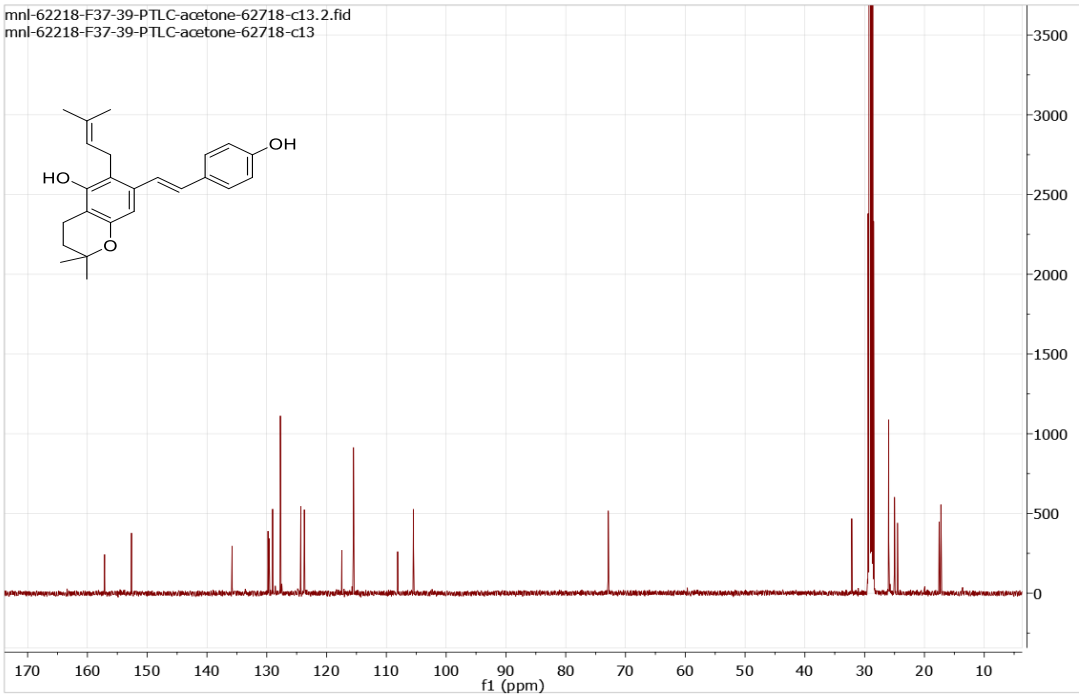
IR spectrum of compound M8



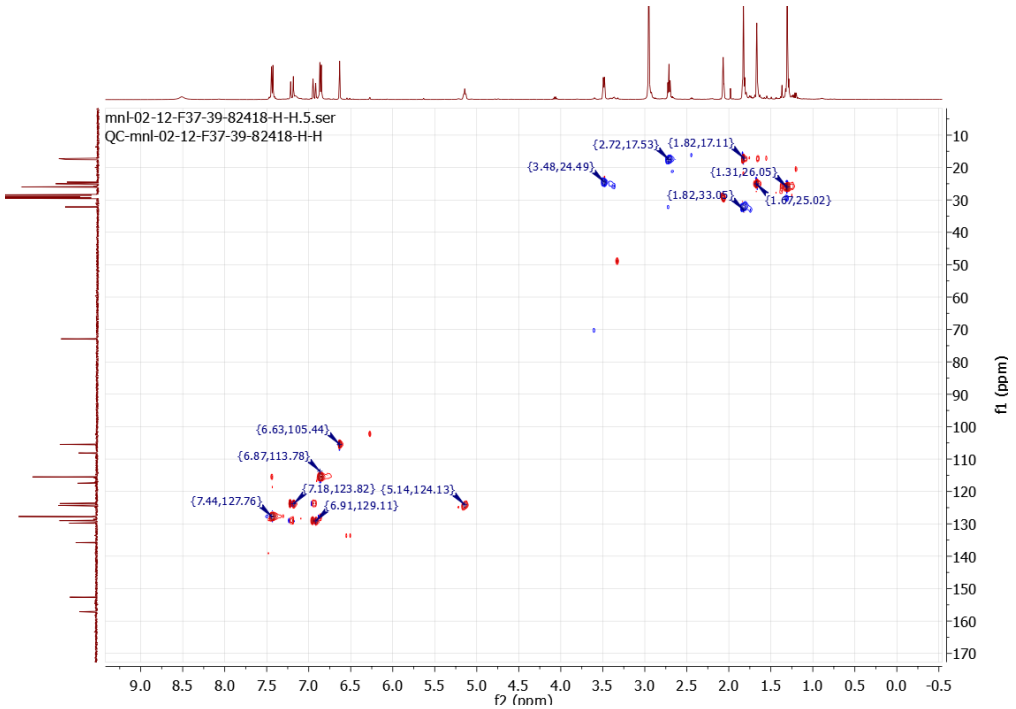
ESI-HRMS of compound M8



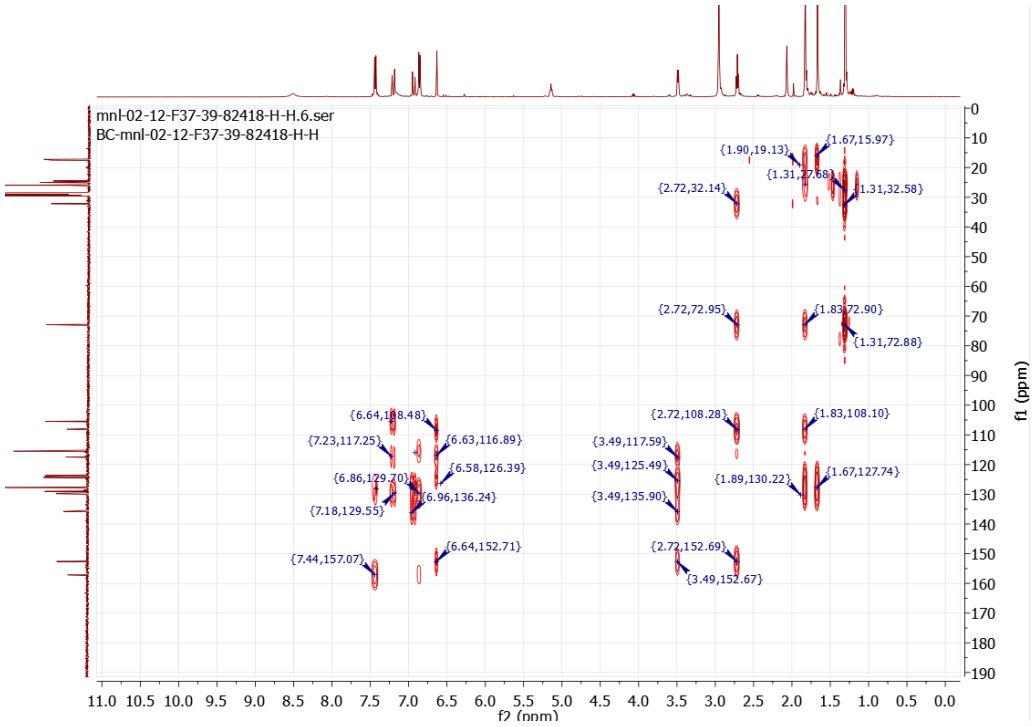
^{1}H NMR (acetone-*d6*, 500 MHz)spectrum of M9



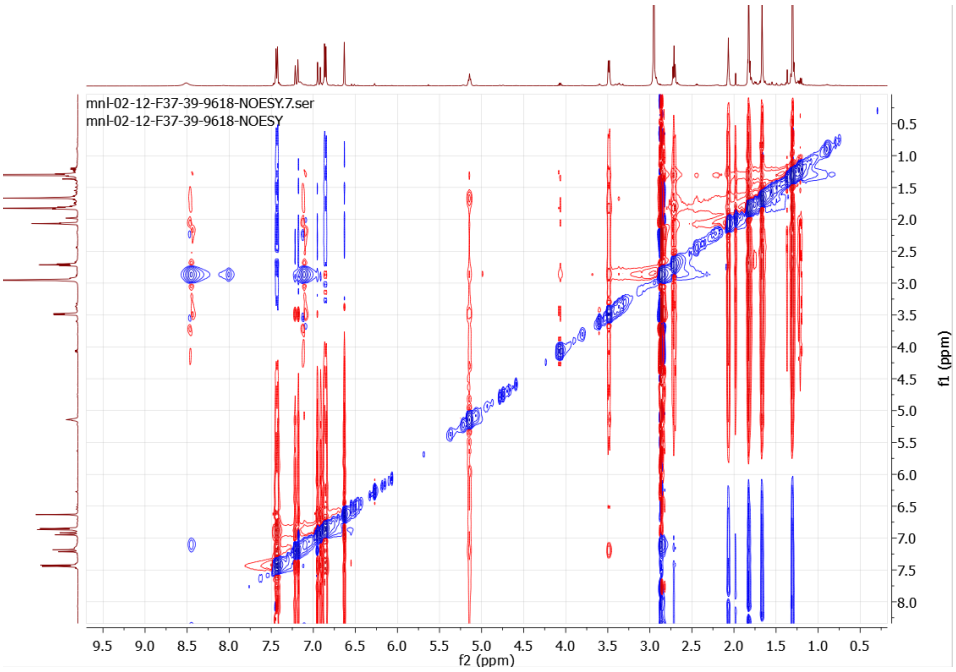
^{13}C NMR (acetone-*d6*, 125 MHz) spectrum of compound M9



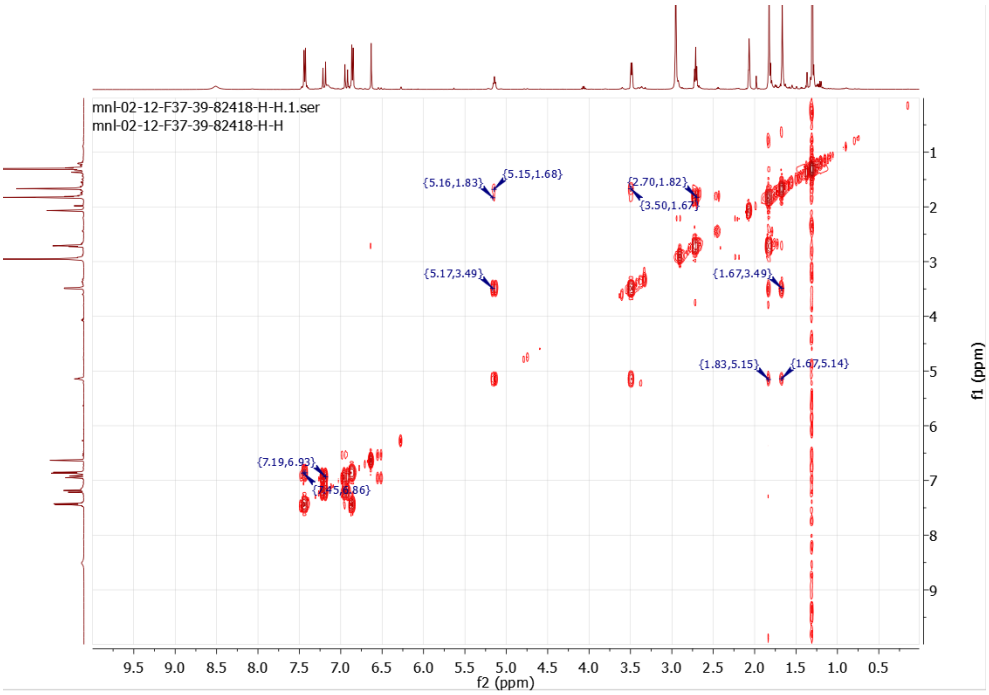
HSQC spectrum of compound M9



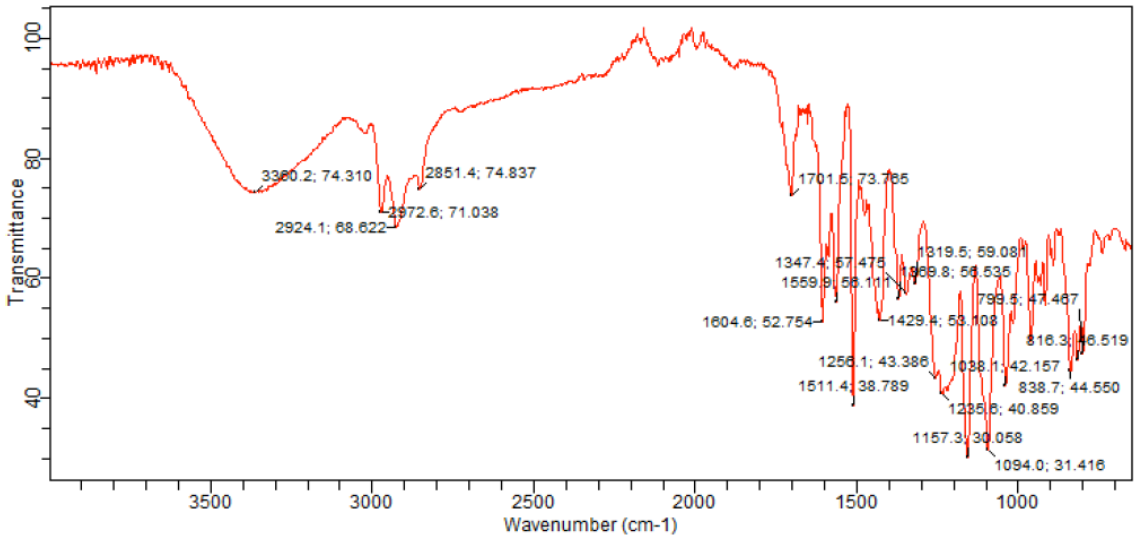
HMBC spectrum of compound M9



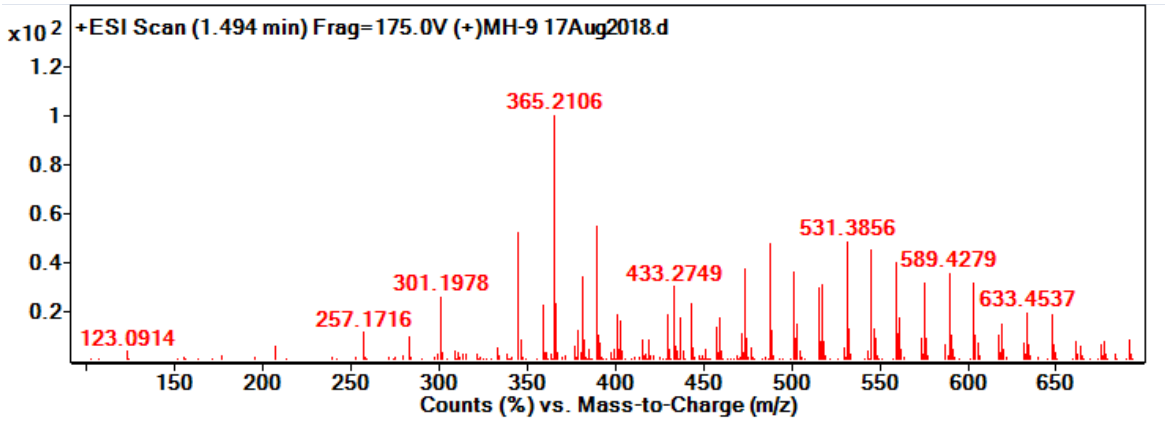
NOESY spectrum of compound M9



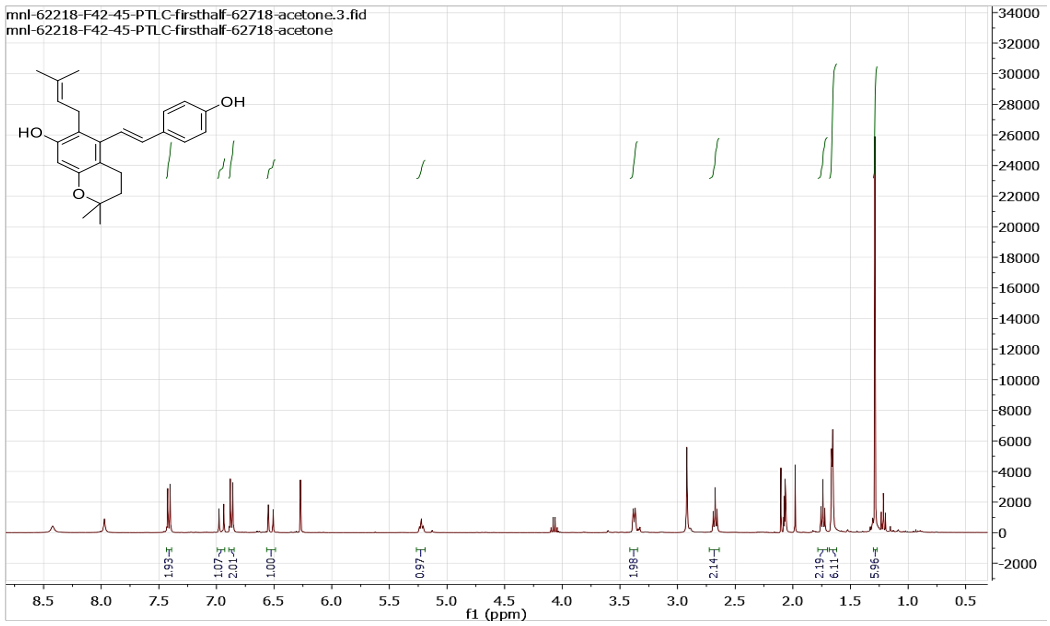
COSY spectrum of compound M9



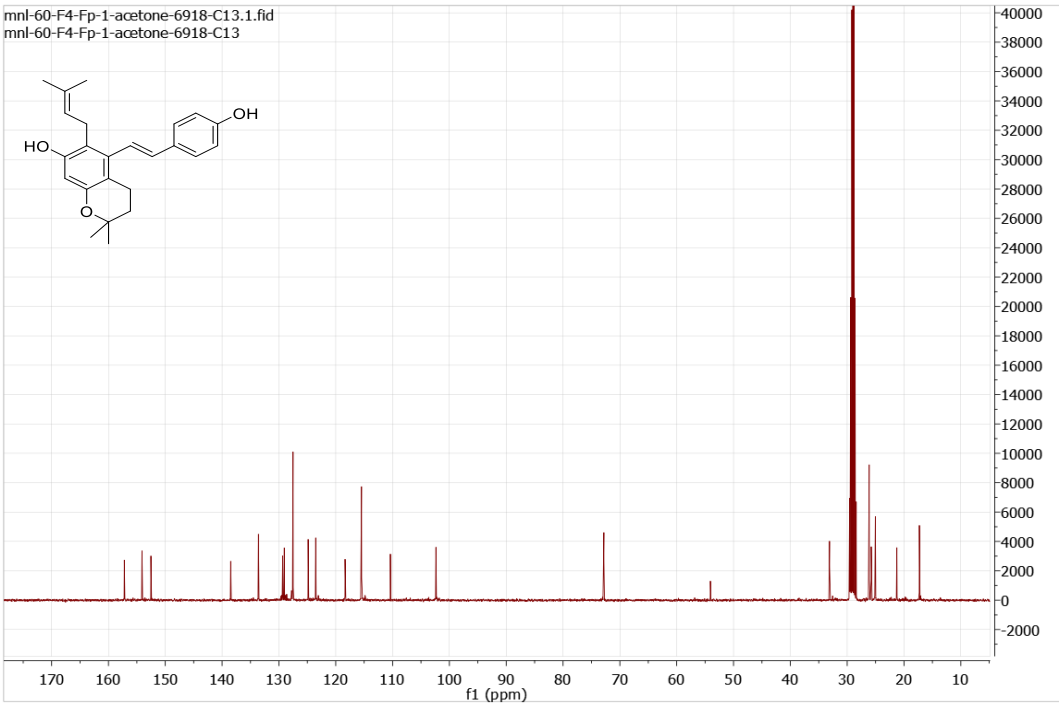
IR spectrum of compound M9



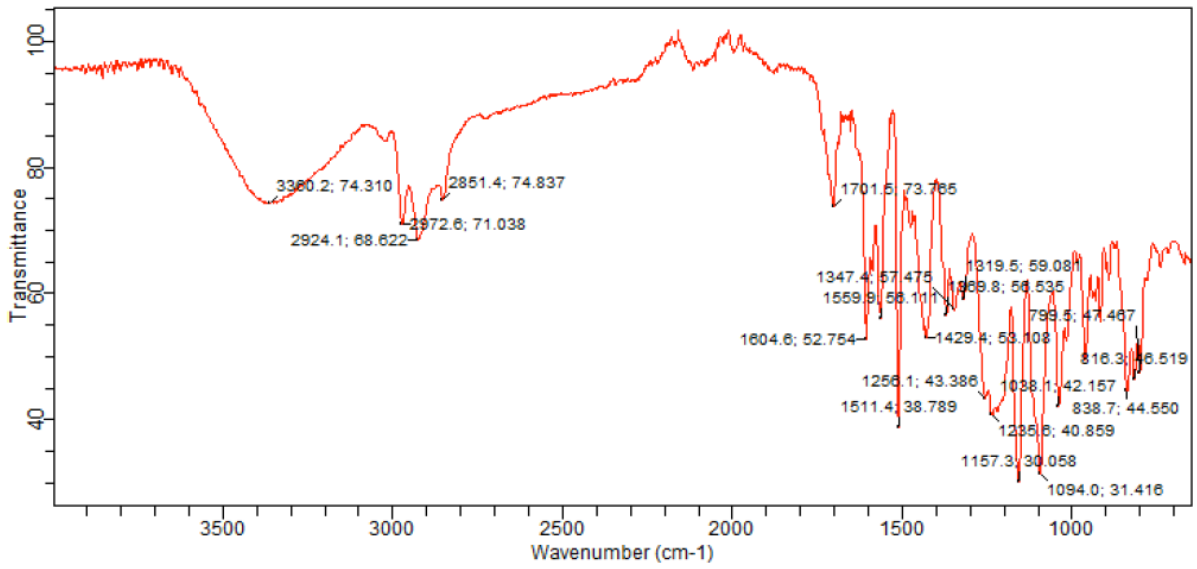
ESI-HRMS of compound M9



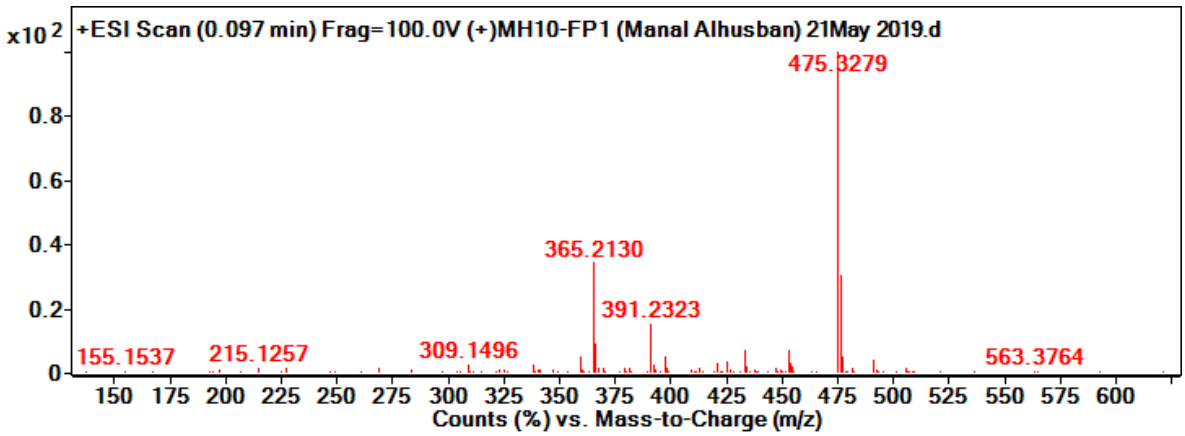
^{1}H NMR (acetone-*d6*, 400 MHz) spectrum of M10



^{13}C NMR (acetone-*d6*, 100 MHz) spectrum of compound M10

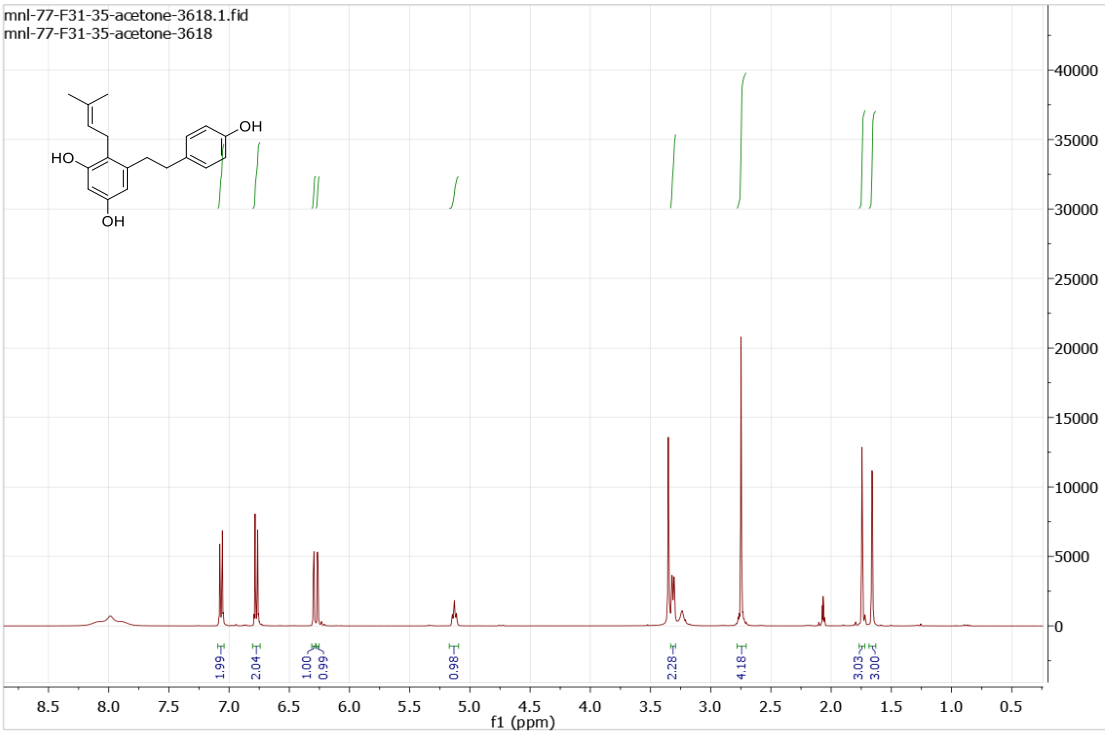


IR spectrum of compound M10

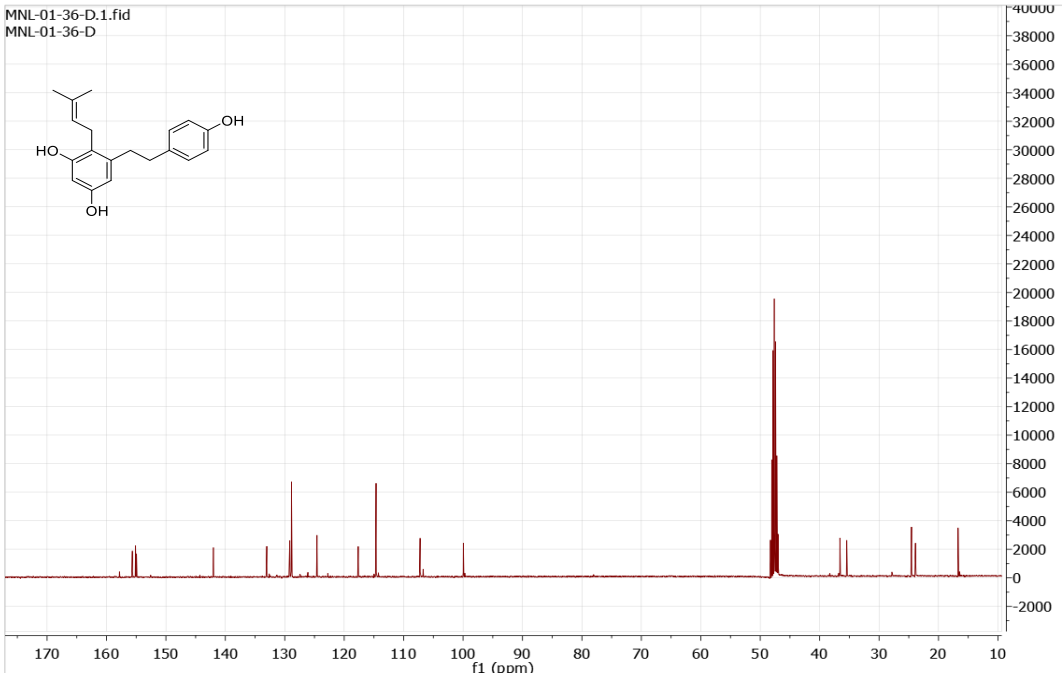


ESI-HRMS of compound M9

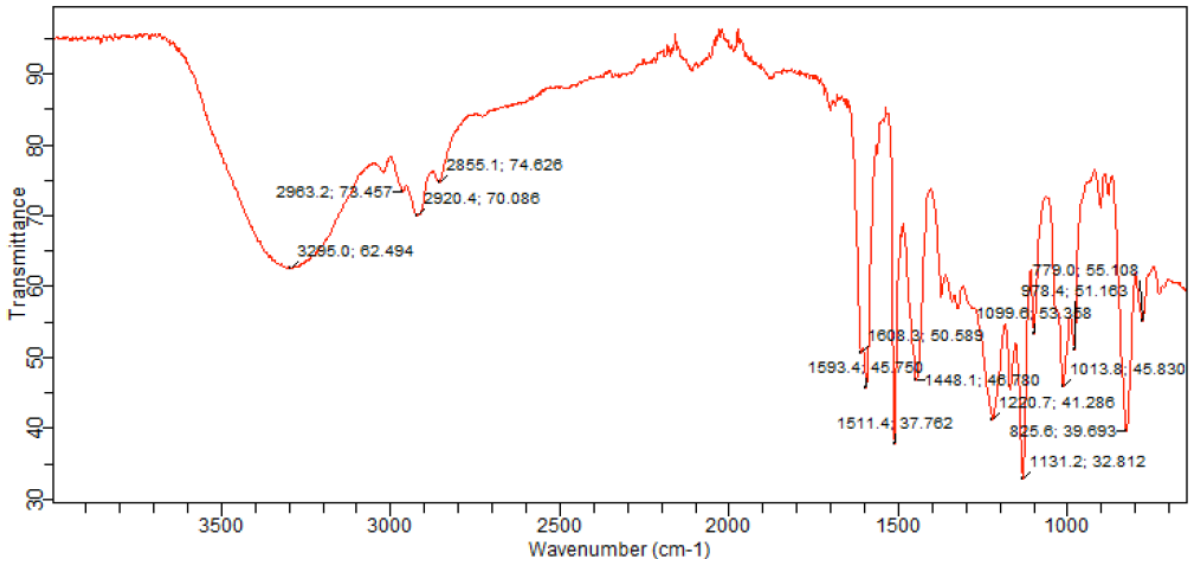
DHS prenylation products (M12-M18)



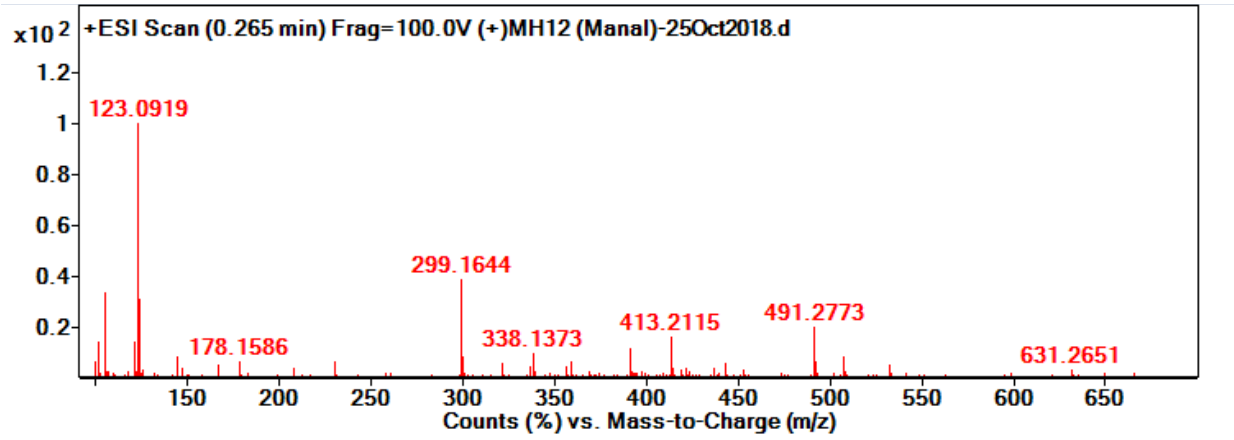
^{1}H NMR (acetone-*d6*, 400 MHz) spectrum of M12



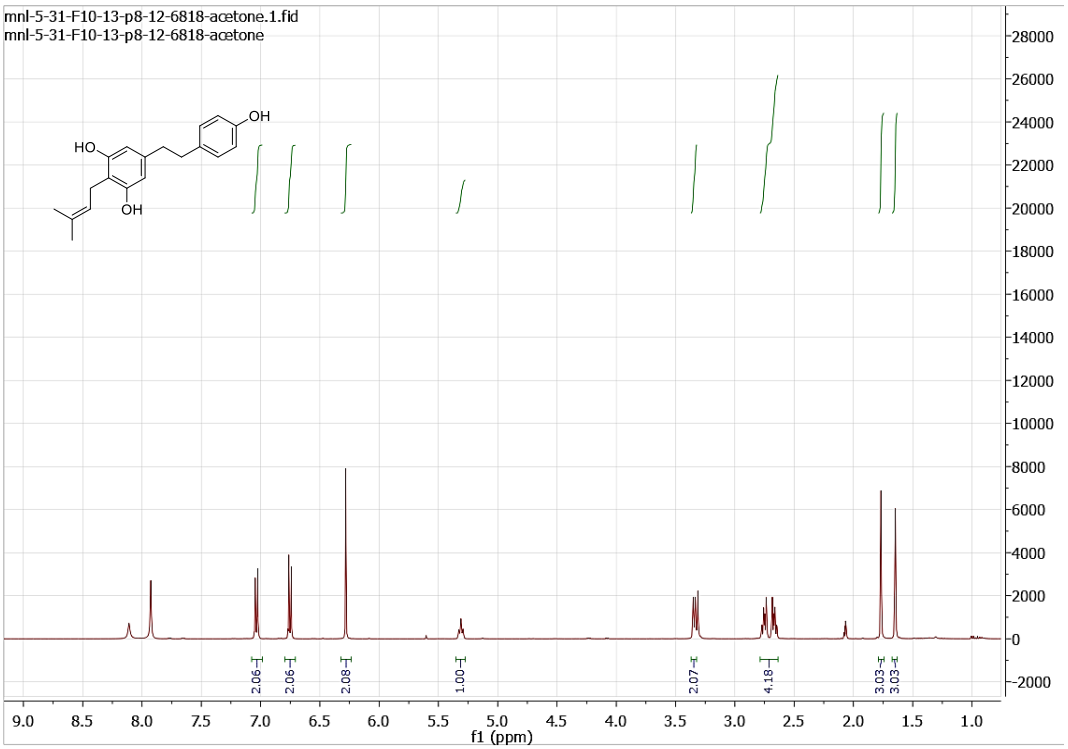
^{13}C NMR (MeOD, 100 MHz) spectrum of compound M12



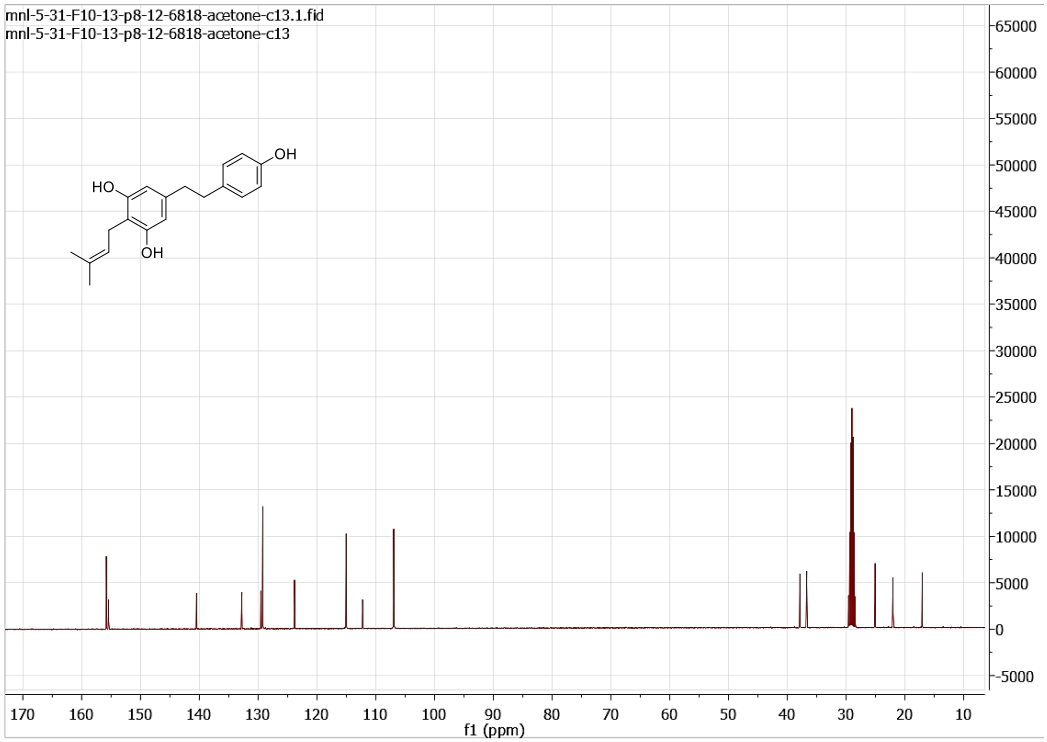
IR spectrum of compound M12



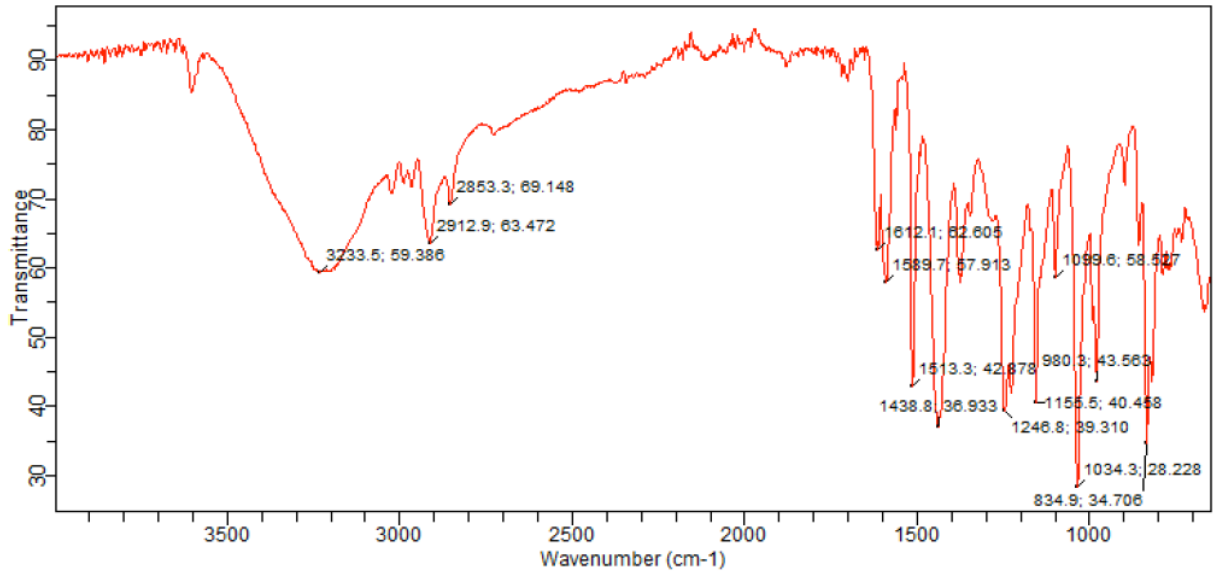
ESI-HRMS of compound M12



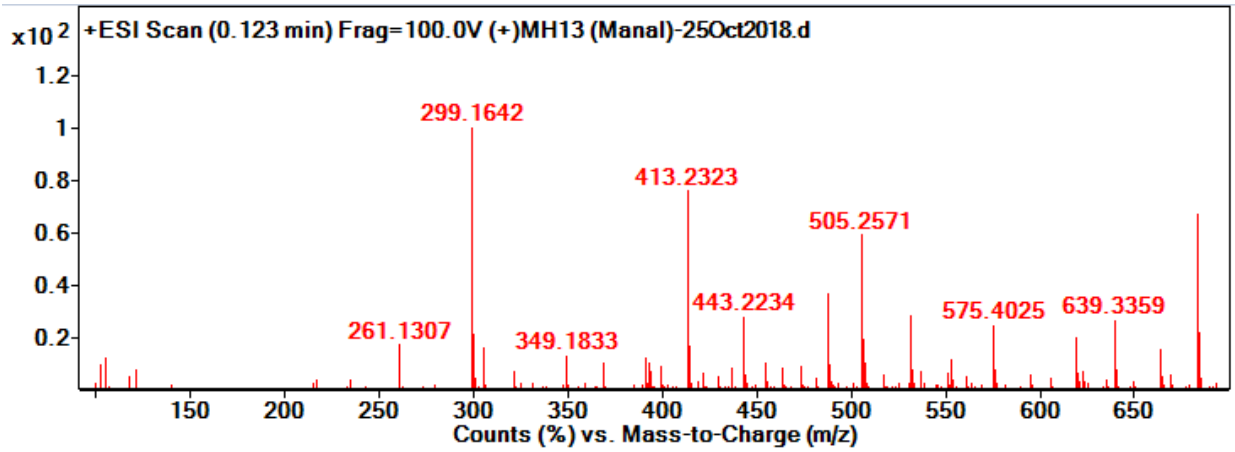
^{1}H NMR (acetone-*d6*, 100 MHz) spectrum of M13



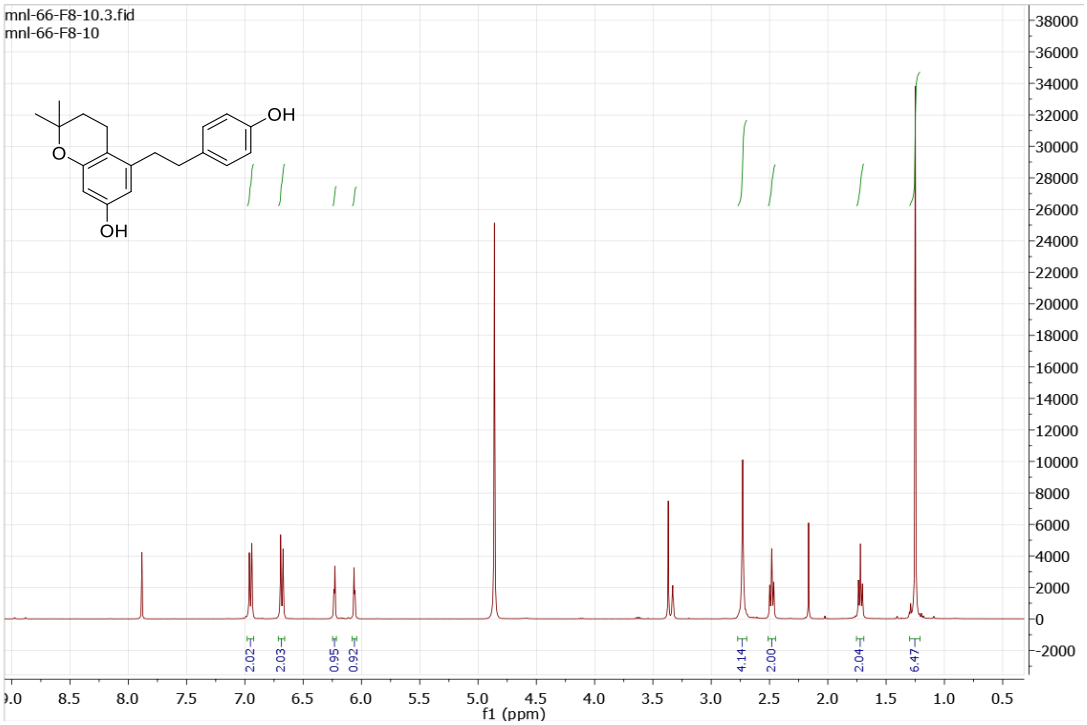
^{13}C NMR (acetone-*d6*, 100 MHz) spectrum of M13



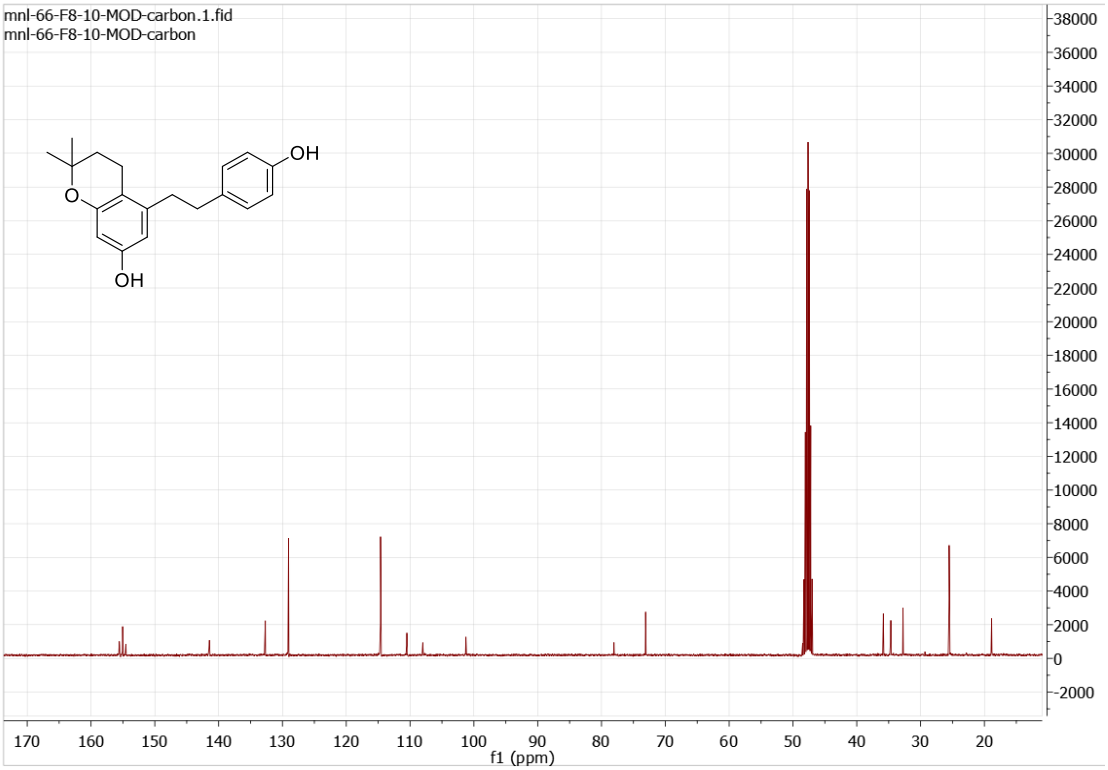
IR spectrum of compound M13



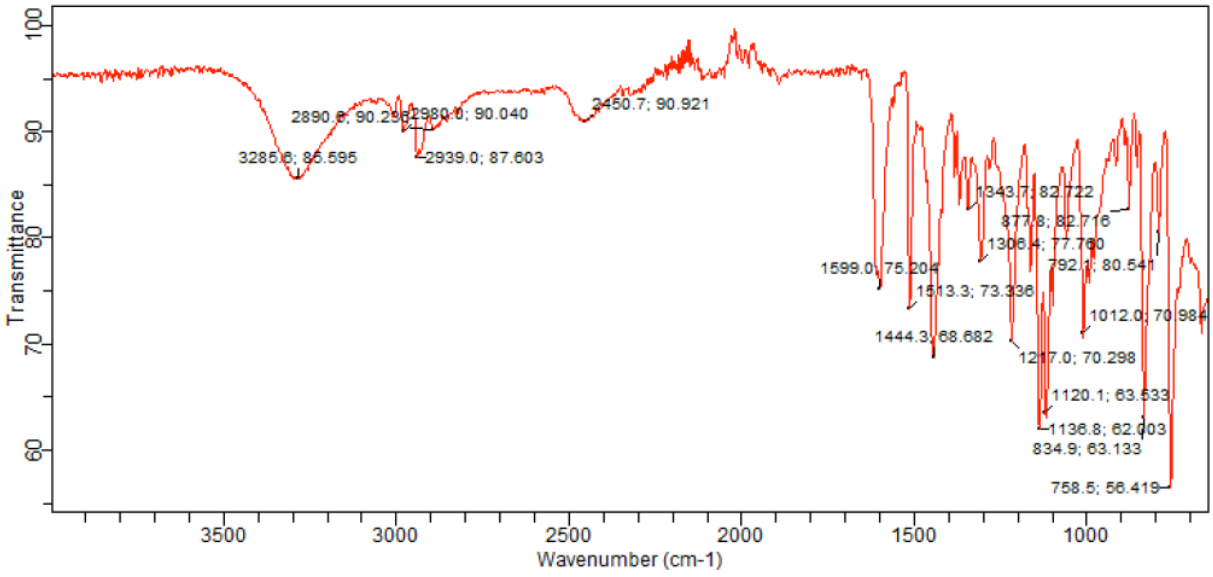
ESI-HRMS of compound M13



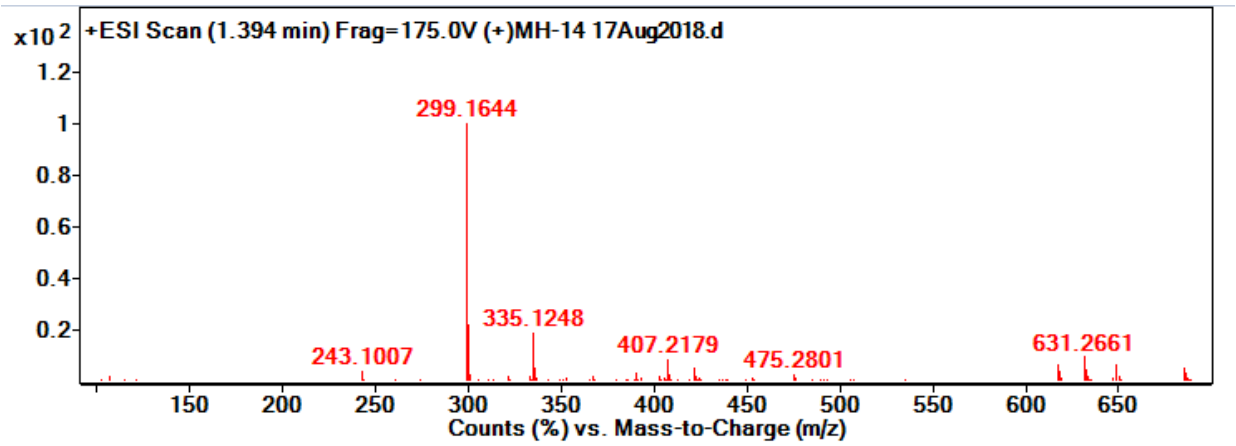
^{1}H NMR (MeOD, 400MHz) spectrum of compound M14



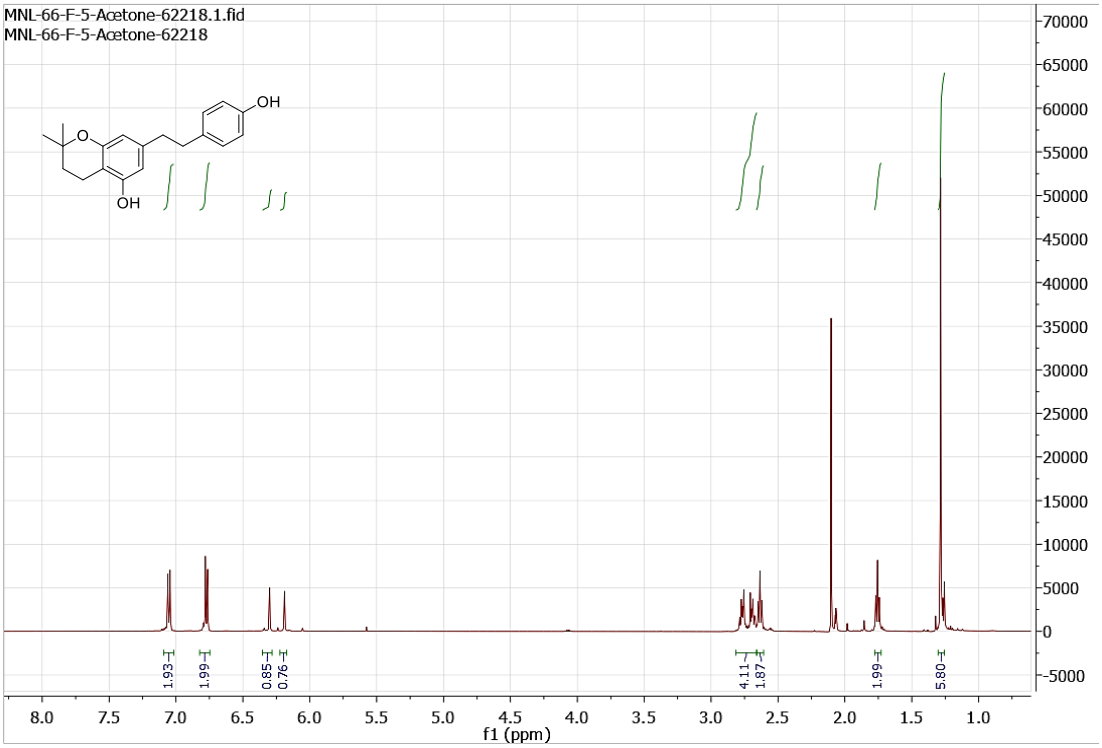
^{13}C NMR (MeOD, 100MHz) spectrum of compound M14



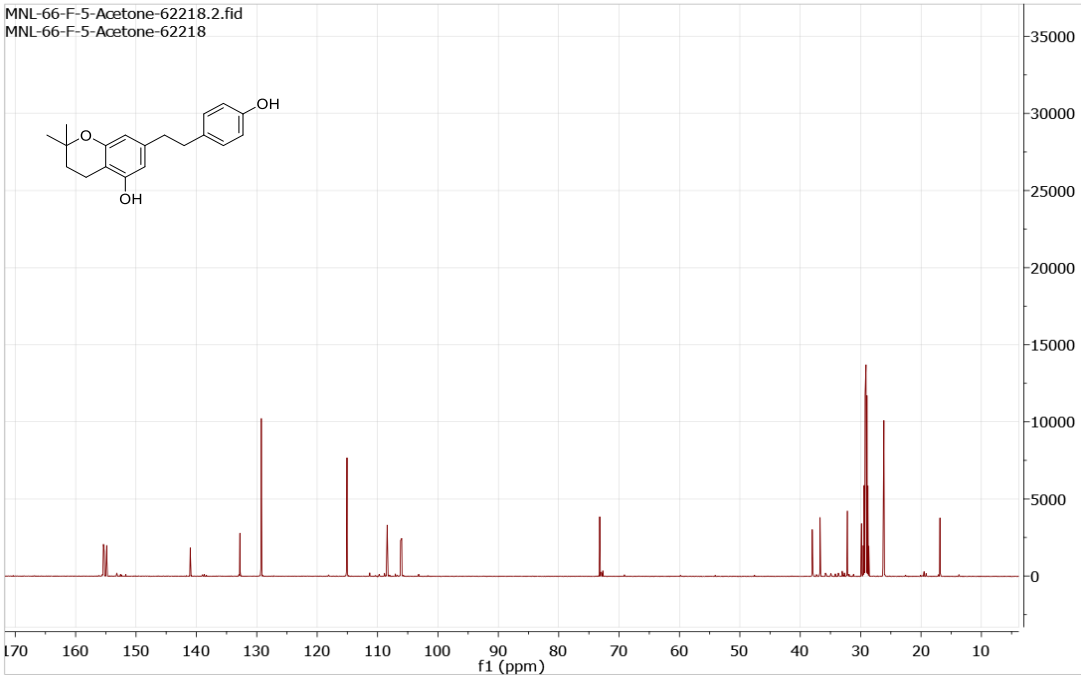
IR spectrum of compound M14



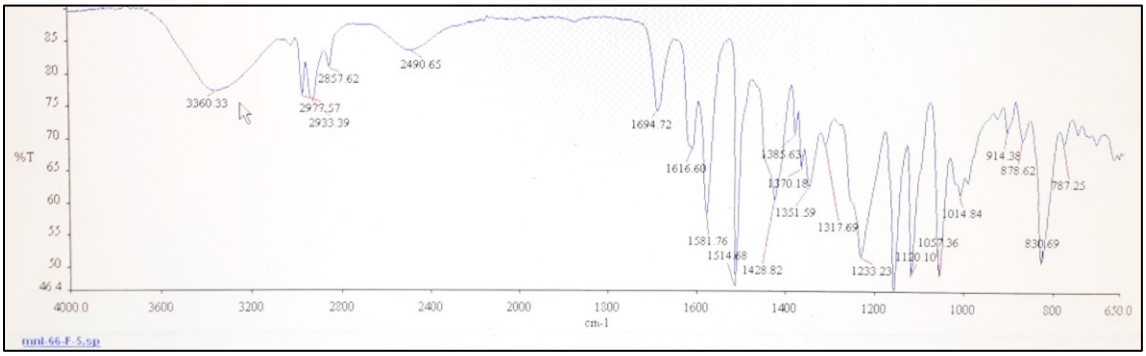
ESI-HRMS of compound M14



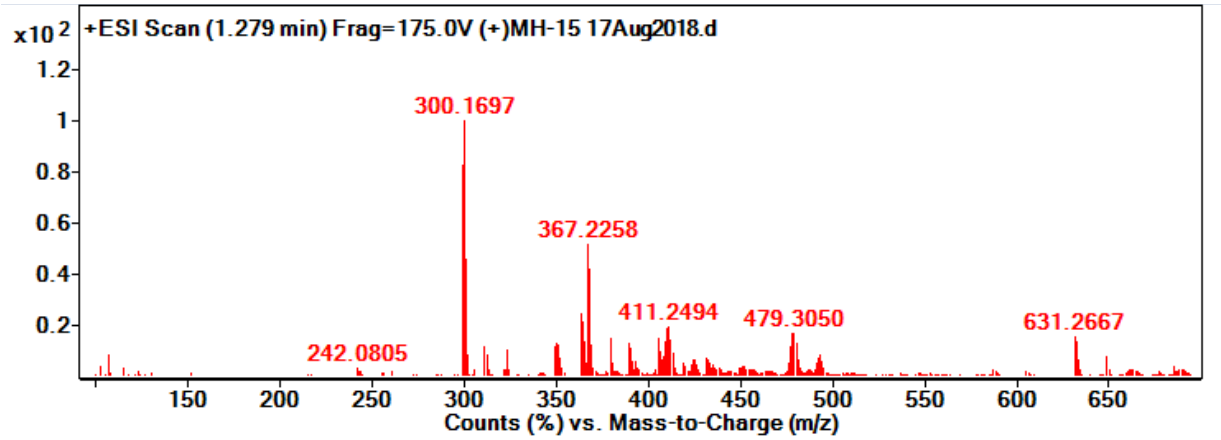
^{1}H NMR (acetone-*d6*, 500 MHz) spectrum of compound M15



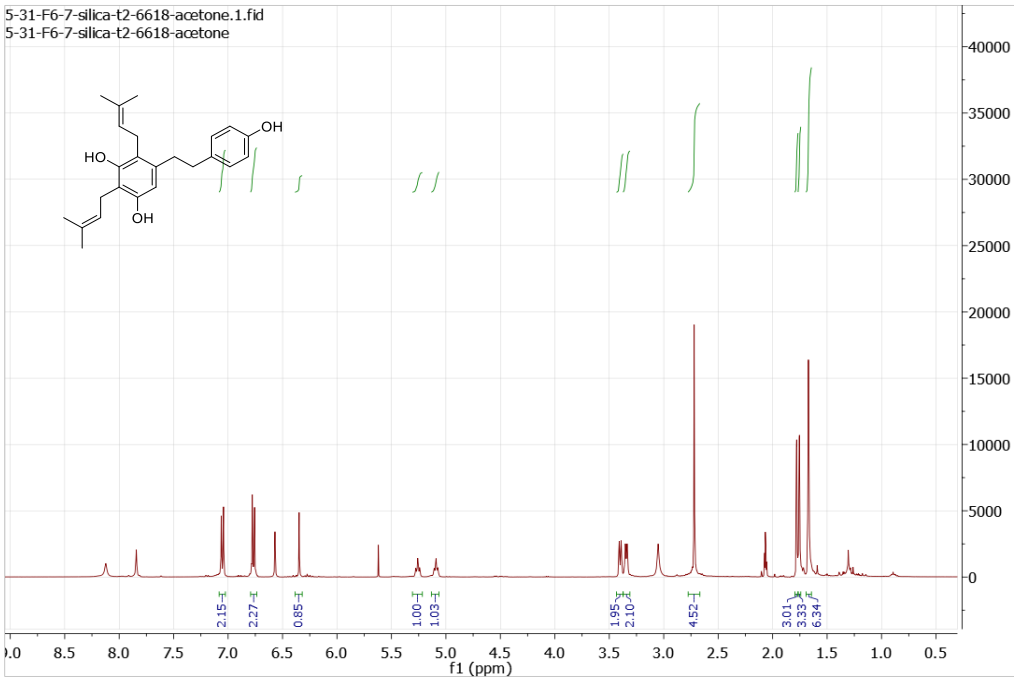
^{13}C NMR (acetone-*d6*, 125 MHz) spectrum of compound M15



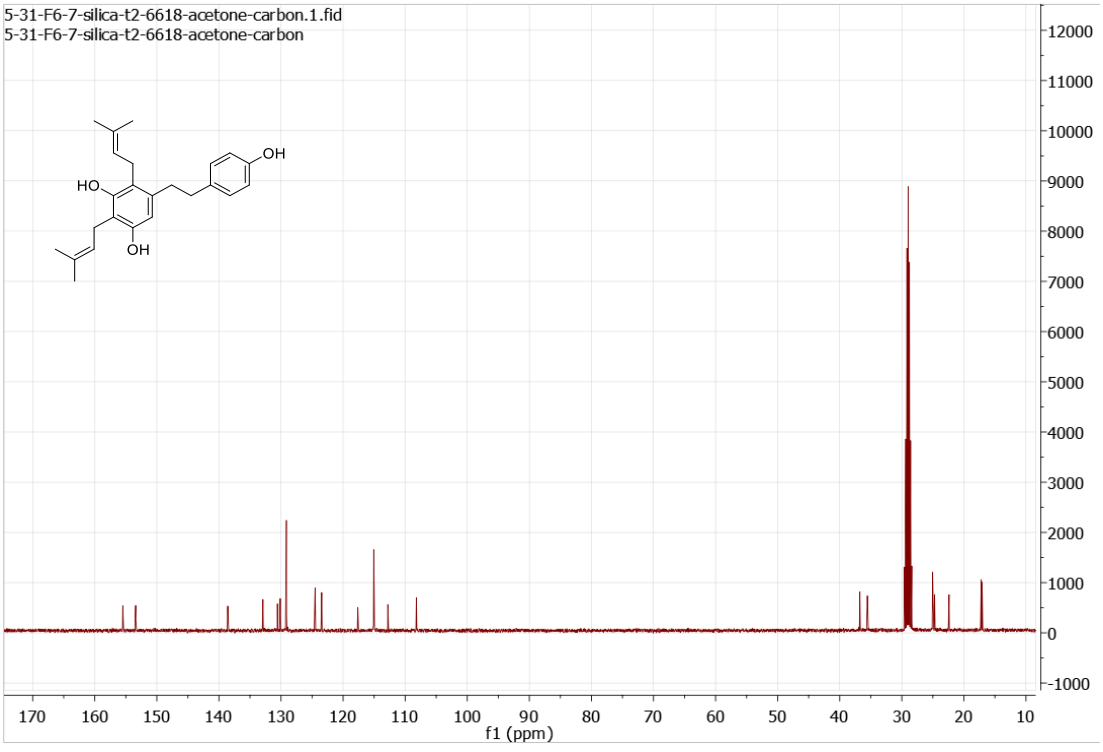
IR spectrum of compound M15



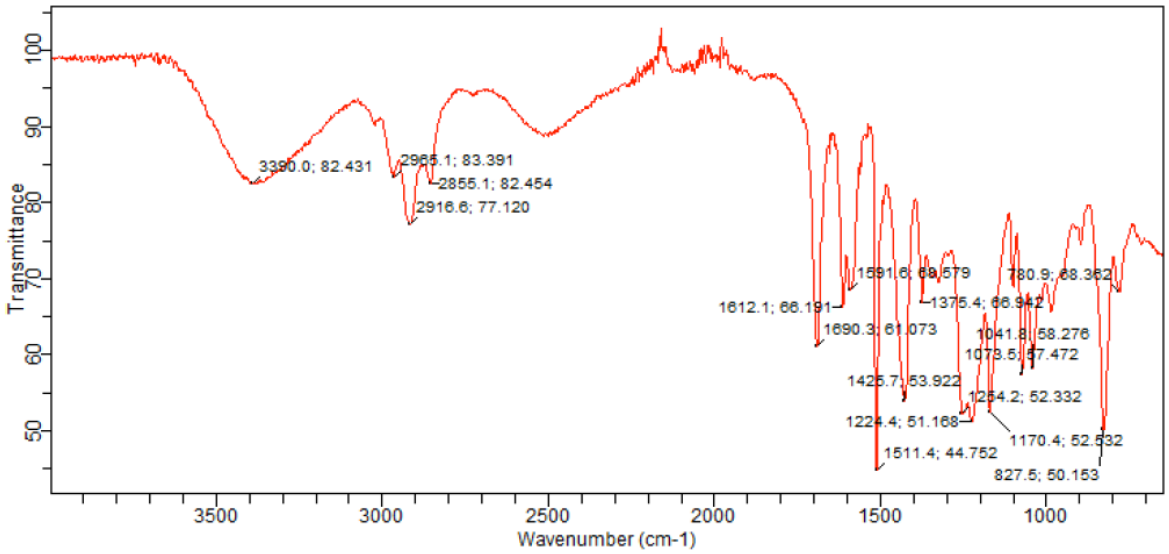
ESI-HRMS of compound M15



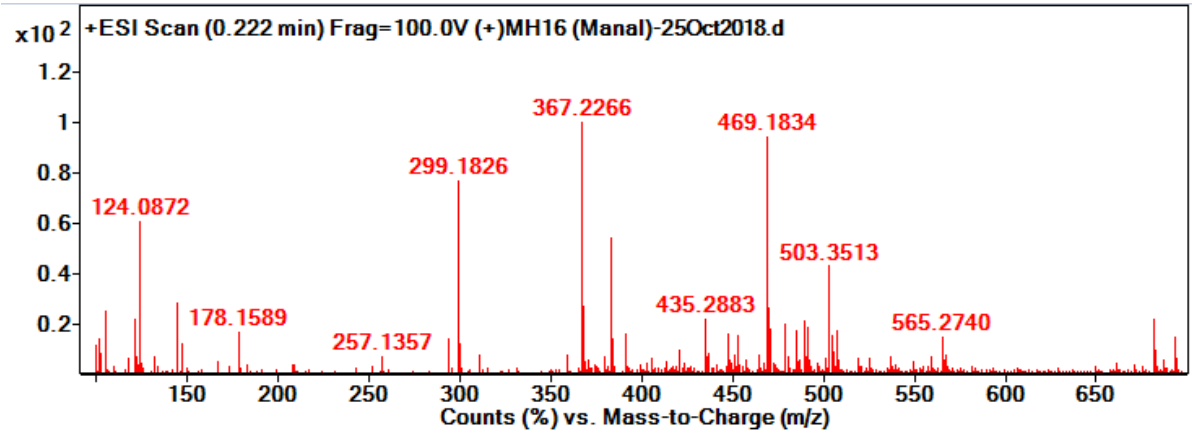
^{1}H NMR (acetone-*d6*, 400 MHz) spectrum of M16



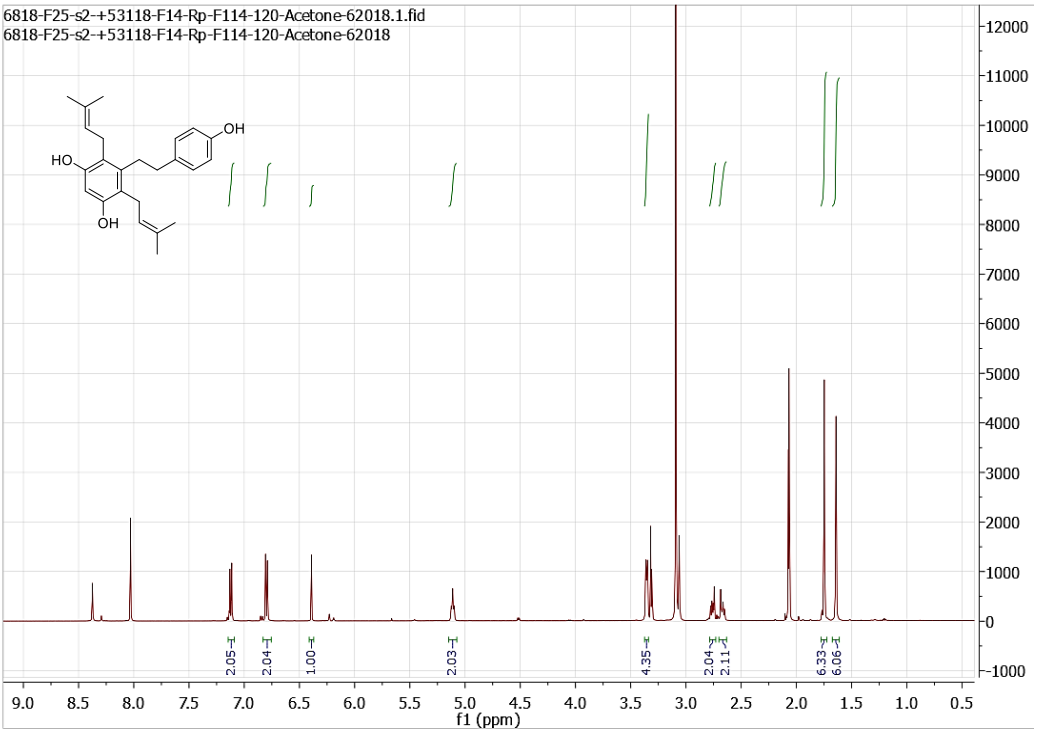
^{13}C NMR (acetone-*d6*, 100 MHz) spectrum of compound M16



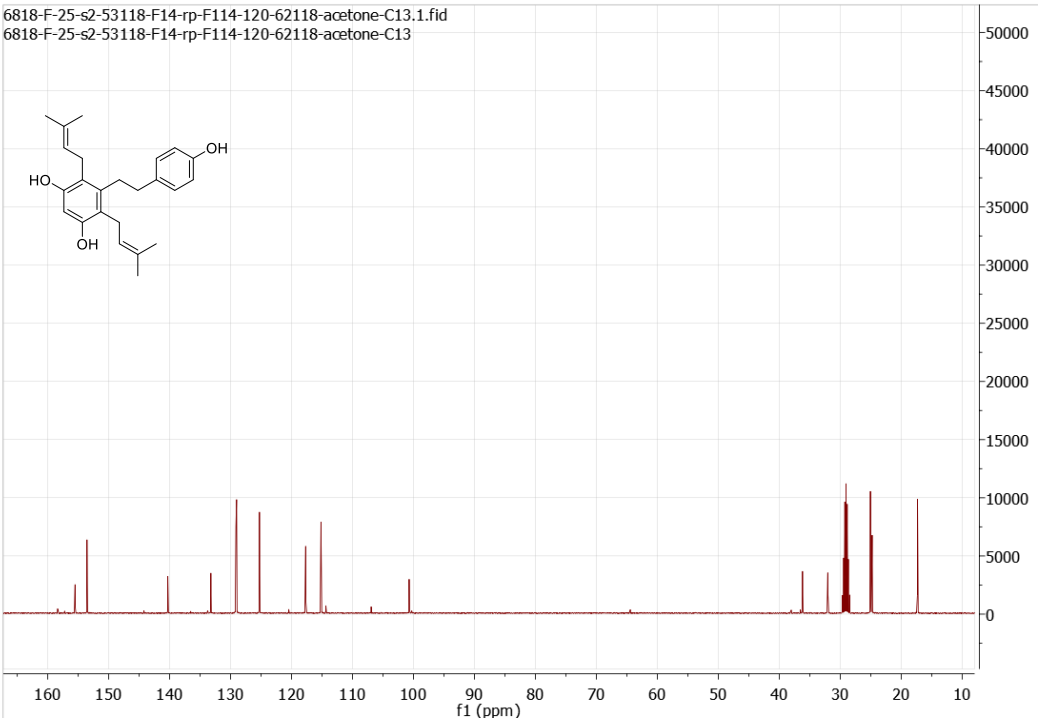
IR spectrum of compound M16



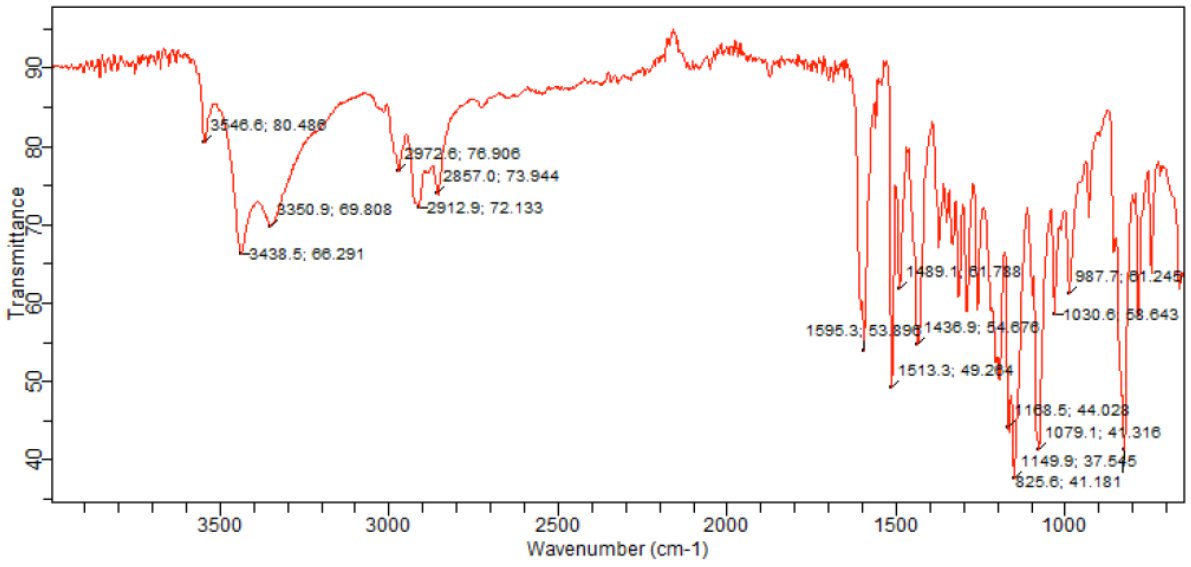
ESI-HRMS of compound M16



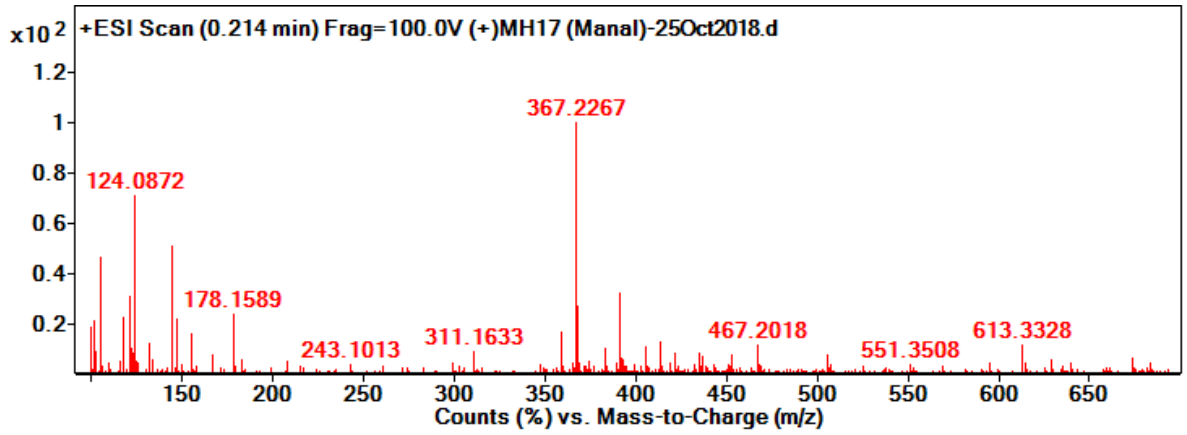
^{1}H NMR (acetone-*d6*, 400 MHz) spectrum of M17



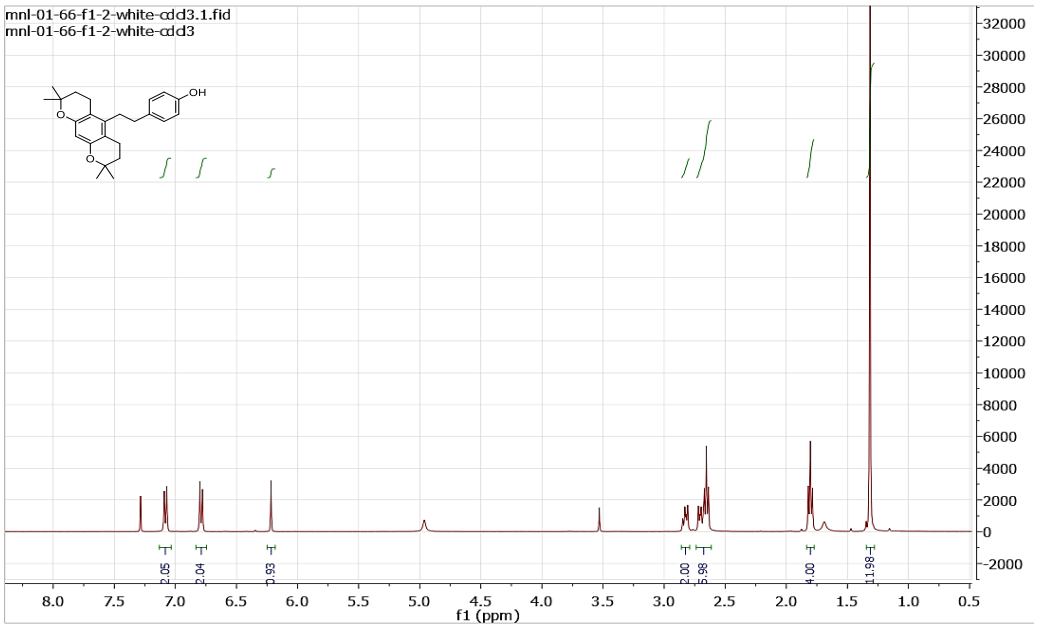
^{13}C NMR (acetone-*d6*, 100 MHz) spectrum of compound M17



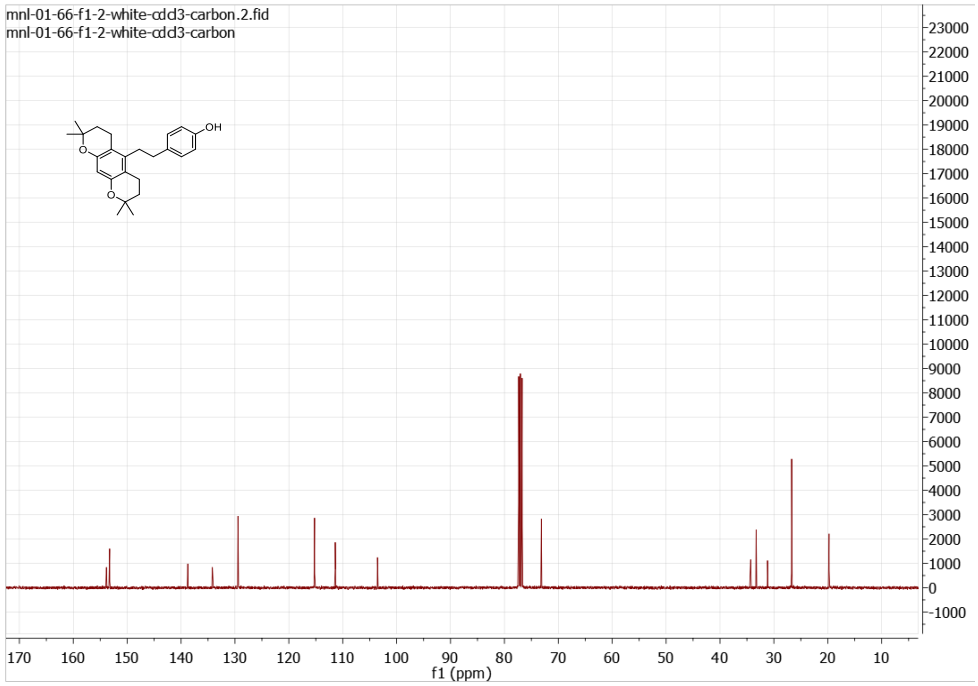
IR spectrum of compound M17



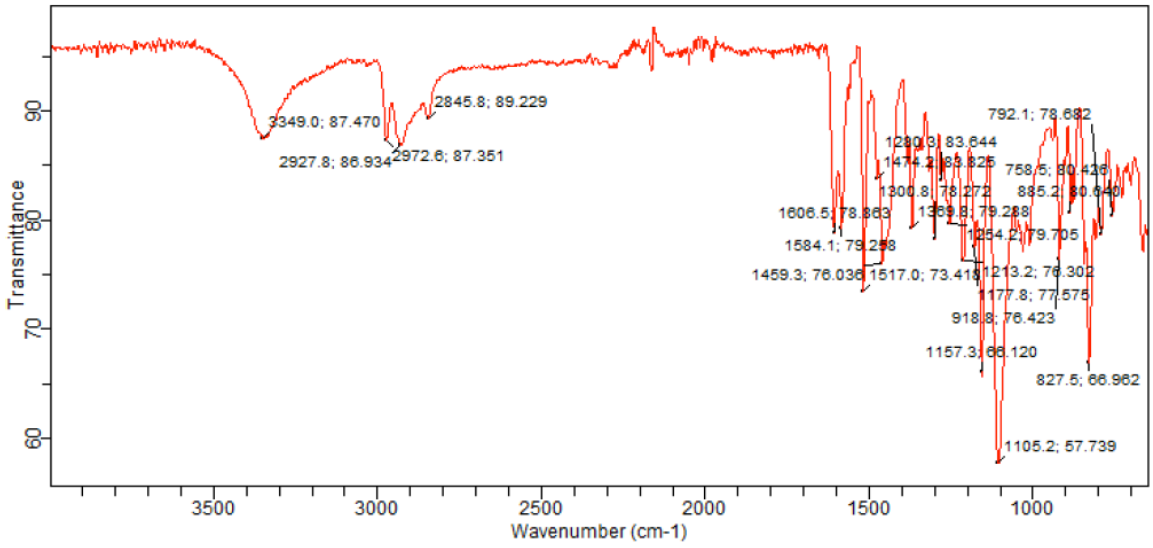
ESI-HRMS of compound M17



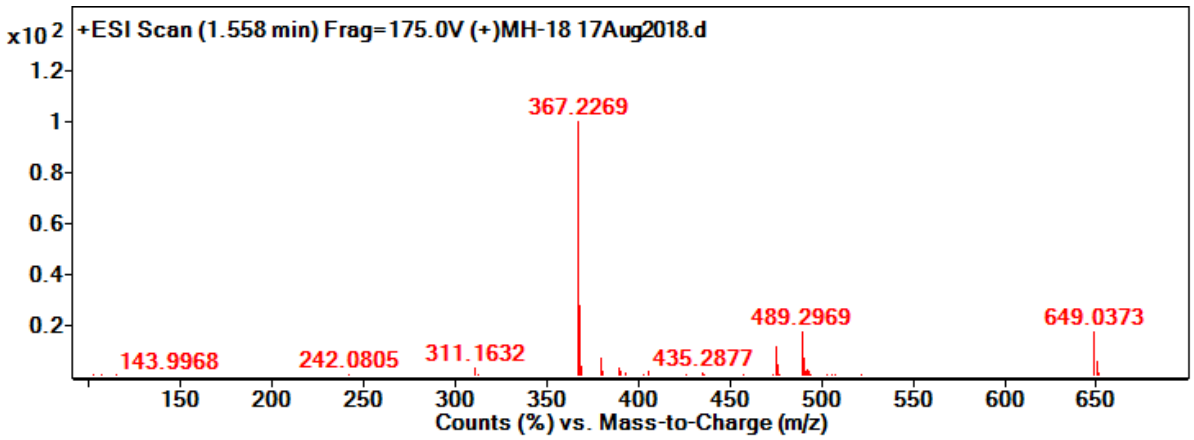
^{1}H NMR ($CDCl_3$, 400 MHz) spectrum of compound M18



^{13}C NMR ($CDCl_3$, 100 MHz) spectrum of compound M18



IR spectrum of compound M18



ESI-HRMS of compound M18

VITA

MANAL ALHUSBAN

Education

2015-present: Doctor of Philosophy Pharmaceutical Sciences (Pharmacognosy), University of Mississippi

2012-2014: Master of Science (Pharmaceutical Sciences), University of Jordan

2002-2007: Bachelors of Pharmacy, University of Jordan

Experience

2016-present: Research Assistant-Pharmacognosy, University of Mississippi

2010-2014: Pharmacist, Ministry of Health-Jordan

2010-2010: Pharmacist, Jordan University Hospital

Skills

Analytical techniques: NMR, HRMS, HPLC, Column Chromatography, Size-exclusion Chromatography, Spectroscopic techniques (FTIR)

Modelling and analysis software: Schrodinger software, PyMOL, Discovery Studio, AutoDock

Publications

1. Haider, S., Alhusban, M., Chaurasiya, N., Tekwani, B., Chittiboyina, A., Khan, IA. Isoform selectivity of harmine-conjugated 1, 2, 3-triazoles against human monoamine oxidase. *Future medicinal chemistry*, 2018. Vol. 10, no. 12, 1435-48.

Oral Presentation

1. *In silico* screening of *Bulbine* genus reveals potential PXR activators, MALTO meeting, The University of Tennessee, Health Sciences Center (2019).

Poster presentation

1. Alhusban M., Chittiboyina AG. and Khan IA. (2019). *Cheminformatic Approach for Deconvolution of Active Compounds in a Complex Mixture - PhytoSERMs in Licorice*. (19th Oxford ICSB meeting)

2. Alhusban M., Zulfiqar A., Khan S. Chittiboyina AG., and Khan IA. (2018). *Computational*

Discovery of potential PXR activators from natural sources: The case of anthraquinones. (18th Oxford ICSB meeting)

3. Haider S., Alhusban M., Chaurasiya N., Tekwani B., Chittiboyina A., and Khan IA. (2018). *Isoform selectivity of harmine-conjugated 1,2,3-triazoles against human monoamine oxidase.* (2018 ASP Annual Meeting)

Awards

1. Best Poster, 2nd place in 19th Annual International Conference on Science of Botanicals ICSB annual meeting (2019)

2. Best Graduate Research Scholar award From the American Chemical Society Ole Miss Local Chapter (2019)