Development of an R-language tool to enhance *in silico* drug discovery from ethnopharmacologically used plant sources: The example of androgenetic alopecia.

Georgios D. Chrysochoou

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy.

FEBRUARY 2023

DEDICATION

This PhD thesis is dedicated to the memory of my beloved father Demetrios Chrysochoou.

ACKNOWLEDGMENTS

There are many people whom I would like to thank for their support.

First and foremost, my supervisor Professor Mark Cronin. I thank him for trusting me, for giving me the opportunity to expand my skills and my limits as my project evolved through these years. I wouldn't have made it this far without his guidance, knowledge and patience.

I would like to express my deep gratitude to my supervisor Dr. Judith Madden for her academic guidance and invaluable input on my research. A special thanks goes to Dr. Steve Enoch for his advice and inspiration throughout this project.

I would also like to thank my wife Nicki for her understanding, patience and encouragement. Finally, but definitely not last, I would like to thank my two sons Demetrios and Stephanos for being the best children a parent would wish for and apologise for not being with them as much as I wanted during my studies.

Abstract

Herbal medicines have been, are, and will always be a major asset in drug discovery. A freely available, user-friendly suite of computation tools has been developed to enhance a variety of *in silico* drug discovery processes. The investigation focused on discovering novel leads for the treatment of androgenetic alopecia (AGA) from natural sources already used ethnopharmacologically.

A set of twenty-two R code snippets (termed 'Tool Services') was created. Tool Services focus on collecting, manipulating, and analysing data from a variety of sources. These sources include general information, ethnopharmacology, chemistry, pharmacology, targets, diseases, pathways and predictive QSAR modelling data.

Sixty-nine plants with established use in AGA were studied and their 2,157 phytochemical ingredients recorded. Taxonomically, more than a third of these plants belong to four families that share many similarities in terms of DNA and phytochemical content.

Structural similarity studies on 34 phytochemicals chosen based on their frequency of occurrence in the plants revealed similarities between them and with UV-protectants, vascular protectants and anti-inflammatory agents.

Seven drugs currently marketed as monotherapy of AGA were structurally compared against our phytochemicals and were also assessed for drug-drug interactions, side effects, adverse drug reactions, and their metabolic fate. During this phase of the study, studies of alopecia as a side effect and as an adverse drug reaction of drugs were also prepared.

A study on 48 targets revealed a strong relation with pathways that are implicated in hair follicle growth and development. More than half of these genes were linked to diseases such as hypotrichosis. Finally, the actual binding sites of these targets and the binding affinities of chemicals for these targets were revealed.

Undoubtedly, the androgen receptor (AR) is one of the most studied target in AGA. A QSAR classification model was built for AR using 206 active and 1600 inactive compounds in terms of AR antagonism, utilising both Random Forest and Naïve Bayes algorithms.

Table of Contents

List of abbreviations

ADR	Adverse Drug Reaction
APC	APC regulator of WNT signalling pathway
AR	Androgen Receptor
ATP	Adenosine Triphosphate
AUC	Area Under Curve
BAMBI	BMP and activin membrane bound inhibitor
BDNF	Brain-Derived Neurotrophic Factor
BMP	Bone Morphogenetic Protein
BR	Biological Response
CADD	Computer-Aided Drug Discovery
CAMD	Computer-Aided Molecular Design
CHRM	Cholinergic Receptor Muscarinic
CSNK	Casein Kinase
CYP	Cytochrome P450
DDI	Drug-Drug Interaction
DHEA-S	Dihydroepiandrosterone
DHH	Desert Hedgehog
DHT	Dihydrotestosterone
DKK	Dickkopf WNT signalling pathway
DP	Dermal papilla
EC50	half-maximal effective concentration
ED50	Median effective dose
EDA	Ectodysplasin
EGF	Epidermal Growth Factor
EOL	Encyclopedia of Life
ERK	Extracellular Regulated MAP Kinase
FDA	Food and Drug Administration
FGF	Fibroblast Growth Factor
FGFR	Fibroblast Growth Factor Receptor
GSK	Glycogen Synthase Kinase
HCD	High Conformational Dependence
HF	Hair Follicle
HFSC	Hair Follicle Stem Cells
HGF	Hepatocyte Growth Factor
HH	Hedgehog
HOMO	Highest Occupied Molecular Orbital
IC50	half-maximal inhibitory concentration
ICD	Intermediate Conformational Dependence
IGF	Insulin-like Growth Factor
IHH	Indian Hedgehog
IL	Interleukin

IRS	Inner Root Sheath
JNK	c-Jun N-terminal kinase
KCNJ	Potassium inwardly rectifying channel
Kd	dissociation constant
KGF	Keratinocyte Growth Factor
Ki	inhibitor constant
KIT	Tyrosine Kinase Receptor
KITL	KIT Ligand
LCD	Low Conformational Dependence
LD	Linkage Disequilibrium
LD50	Median lethal dose
LEF	Lymphoid Enhancer Factor
LGR	Leucine-rich repeat containing G-protein coupled eceptor
LH	Lutenizing Hormone
LUMO	Lowest Unoccupied Molecular Orbital
MAPK	Mitogen-Activated Protein Kinase
MCS	Maximum Common Substructure
MHRA	Medicines and Healthcare products Regulatory Agency
MITF	Microphthalmia Transcription Factor
MLR	Multiple Linear Regression
mRNA	messenger RNA
MSX	Msh homeobox
NCD	No Conformational Dependence
ORS	Outer Root Sheath
PAX	Paired Box
PCA	Principal Component analysis
PGE	Prostaglandin E
Pgp	P-Glycoprotein
PI3K	Phosphatidylinositol Kinase
PKB	Serine/Threonine Kinase
PLCG	Phospholipase C Gamma
PLS	Partial Least Square analysis
PSA	Prostate Specific Antigen
PTCH	Patched Gene
PTGFR	Prostaglandin F Receptor
PTGS	Prostaglandin Synthase
QSAR	Quantitative Structure-Activity Relationship
R2	Coefficient of determination
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RSPO	R-Spondin
sd	standard deviation
SGK	Serum/Glycocorticoid regulated Kinase
SHBG	Sex Hormone Binding Globulin

SHH	Sonic Hedgehog
SMARTS	SMILES Arbitrary Target Specification
SMILES	Simplified Molecular-Input Line-Entry System
SMO	Smoothened, frizzled class receptor
SNP	Single Nucleotide Polymorphisms
SOX	SRY-box Transcription Factor
SRD5A	Steroid 5 Alpha-Reductase
STAT	Signal Transducer and Activator of Transcription
SVM	Support Vector Machines
TAB	TGF-beta activated kinase binding protein
TAK	TGF-beta activated kinase
TCF/LEF	T-Cell Factor/Lymphoid Enhancer Factor
TGF	Tissue Growth Factor
TNF	Tumor Necrosis Factor
VDR	Vitamin D Receptor
VEGF	Vascular Endothelial Growth Factor
XIAP	X-linked inhibitor of apoptosis

CHAPTER 1: Thesis overview

Drug development is a time and money consuming process which requires the participation of scientists from different disciplines. A way to reduce the substantial cost and lengthy production time, is to accelerate each individual process before drug candidates enter pre-clinical and clinical trials. The usual strategy for selecting drug candidates is by screening at least 1-10 million organic compounds with drug-like properties using a focused approach. *In silico* rational drug design has attracted the interest of researchers in the last decades as it aims to reduce time and cost of these processes through synthesising molecules, biologically testing them, improving their characteristics and re-synthesising and testing these improved versions of the molecules.

Androgenetic alopecia (AGA) is the most common form of hair loss and thinning in men and women. Regardless of age or sex, AGA has a substantial negative impact in people's quality of life and psychology. The aetiology of AGA is polygenic and multifactorial with the metabolism of androgens and interactions with their receptors apparently playing a very significant role in its pathogenesis. The purpose of treatment of AGA is to prevent hair loss, reverse miniaturisation and promote hair regrowth. At the moment, no therapy can restore all hair lost and most therapies require large periods of uninterrupted administration until visible results are observed and sustained. This is a possible reason why people have turned to alternative therapies, with herbal medicines being their preferred approach (Braun et al., 2010). The use of herbal medicines has undoubtedly increased worldwide over time (Ekor, 2014). The main reasons are the ineffectiveness of conventional drugs to treat some diseases and people's perception that herbal medicines are safer to use. Most herbal medicines consist of hundreds of constituent compounds which makes their study more complicated in terms of pharmacology, toxicology and drug safety (Farah et al., 2000). In the last few decades, there has been considerable effort to overcome these problems but, as of the current time, the lack of consistent scientific knowledge persists.

Herbal medicines are a huge asset in pharmaceutical science and with the utilisation of contemporary computing power as well as the development of all related sciences, there is a

great opportunity to reduce time and cost of drug discovery from such sources. The main advantage of using natural products is that phytochemicals provide a wider spectrum of diverse structures in comparison to synthetic chemistry (Harvey and Waterman, 1998), offering the opportunity to find novel lead structures. Hence, many plants which have been used successfully in ethnopharmacology may provide the basis to elucidate novel therapeutic mechanisms of action for disorders where current therapies are still ineffective.

The desire to develop therapies from natural products poses several challenges. In terms of ethnopharmacological research, there is a need to address the problems of accuracy in plant nomenclature and taxonomy which are crucial with regard to medicinal plant identification. Efficient and accurate identification of natural products will promote the reproducibility, recording and prediction of ethnopharmacological parameters of importance in drug discovery. There is also a need to ensure the safety of phytochemicals from an early stage in the development process and to determine the targets and biochemical pathways associated to the disease.

1.1 Chapters' overview

Chapters 1-4 provide the theoretical backbone of this thesis. Chapter 1 provides an overview of the thesis and it's aims and objectives. Chapter 2 is an in-depth literature review on the anatomy, embryology and physiology of the hair follicle. Chapter 3 is a study of types of non-cicatricial alopecia and provides information on the epidemiology, etiopathogenesis and pharmacological management of androgenetic alopecia. The chapter continues with an analysis of the main signalling pathways involved in AGA and their link to the anatomical structures discussed in Chapter 2. Chapter 4 briefly explores Computer-Aided Drug Design and mainly focuses on the importance of investing in Quantitative Structure-Activity Relationships as a means to reduce the cost of novel therapies and optimise existing therapies. Chapter 5 contains all the methodologies used throughout the thesis. It also contains individual examples of use for each Tool Service. The results of the application of all Tool Services are presented and discussed in Chapter 6. In Chapter 7, there is an effort to draw conclusions and link them with the theory presented in previous chapters.

1.2 Aim and objectives of the thesis

Aim

The aim of this thesis was to create a computational tool which will enhance a number of individual drug discovery processes and allow us to exploit the wealth of ethnopharmacologically used natural products to propose novel, safe lead compounds for the treatment of androgenetic alopecia.

Objectives

The main objectives of this thesis were to:

- Identify phytochemical constituents from plants ethnopharmacologically used in AGA, with the potential to serve as lead compounds in treating this condition
- Contribute to ethnopharmacological research by automating processes related to plant nomenclature, taxonomy and identification
- Improve structural similarity studies and ensure the safety of phytochemical leads
- Enhance target identification and validation by taking into account, analysing and combining a diverse range of information (ranging from text-mining of published literature and genome-wide association studies to Gene Ontologies and biochemical pathway data)
- Determine pathways and diseases relevant to the targets studied and assess them using Network Pharmacology analyses
- Support the development of validated Quantitative Structure-Activity Relationship (QSAR) models, which enable the prediction of the biological activity and properties of novel leads from phytochemicals, thereby enhancing the efficiency and success rate of drug discovery efforts.

CHAPTER 2: The hair follicle

2.1 Hair

Hair plays a significant role in an individual's aesthetic appearance. Through the ages, hairstyles became a means of discrimination of races, classes, culture and age. The significance of scalp hair can also be seen by the amount of money spent for its care (Twigg and Majima, 2014) and also by the fact that hair disorders affect the psychological balance causing reactions ranging from stress to panic (Peters et al., 2017).

Physiological functions of scalp hair

The physiological functions of scalp hair include:

1. Sensory function: Scalp and body hair are organs of touch (Jönsson et al., 2017)
2. Thermoregulation function: Scalp hair contributes significantly in keeping the head warm during winter months (Tansey and Johnson, 2015)
3. Protective function: Protects from UV rays, trauma and chemical substances (Botchkarev, 2003).
4. Aesthetics: Dense hair is a sign of health and youth

Types of hair

Based on morphological criteria, location, sex, and age, hair is classified in the following categories:

(a) **Lanugo hair**: This is the first hair produced by the foetal hair follicles during life in uterus. This kind of hair is soft, unpigmented and lacks a medulla, and falls out during the 7^{th} and 8^{th} months of gestation (Jung and Chang, 2016). This is the first generation of hair and is replaced at the end of gestation by vellus hair. Lanugo hair can be seen in prematurely born babies and in congenital hypertrichosis lanuginosa which can last a lifetime or be found in alopecia areata foci (Bartosová et al., 1984).

(b) **Vellus hair**: This is the second generation of hair. Hair is soft, without medulla, pigmented or unpigmented and very rarely exceeds 2cm (Bartosová et al., 1984; Rook, 1965).

(c) **Intermediate hair**: In children, vellus scalp hair is replaced by thicker and more pigmented hair without medulla. Intermediate scalp hair is replaced by terminal hair as the person reaches puberty (De Villez, 1986, Miranda et al., 2010).

(d) **Terminal hair**: Terminal hair appear gradually in puberty. This hair is long and thick, contains a medulla and is pigmented. Hair medulla in teenage boys is more developed than in women (Srettabunjong et al., 2016). The first terminal hair to develop before puberty are eyelashes and eyebrows, which are complete in childhood and change only slightly in puberty. Eyelashes have the deepest pigmentation and the biggest diameter (10-20 µm). Eyelashes are flat and curved with a length of approximately 10mm (Bartosová et al., 1984).

Various types of hair are shown in Figure 2.1 below.

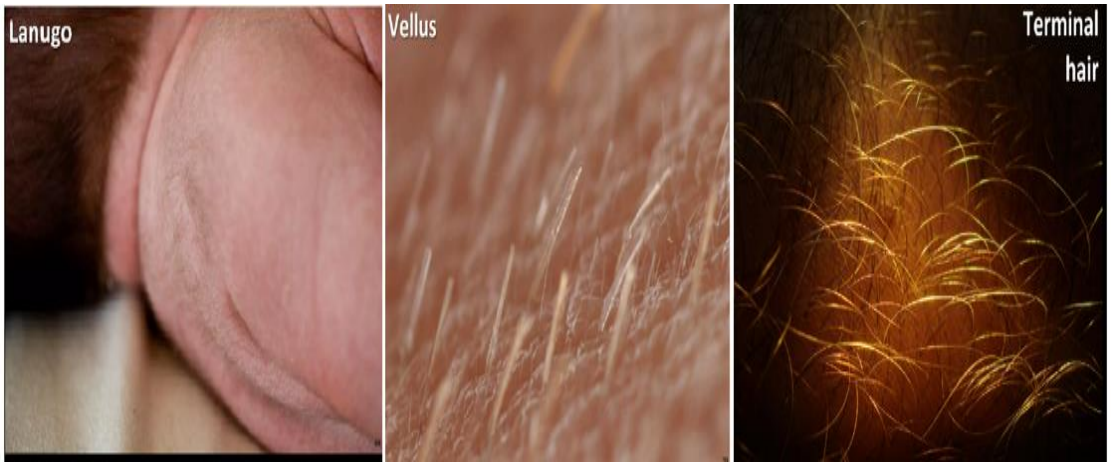


Figure 2.1: The different types of hair

The change from intermediate to terminal hair in the scalp is slow and occurs as the individual reaches puberty. There are changes in colour and shape of hair (e.g. straight hair are replaced by curly hair and darker hair may become lighter). The direction of

growth which is defined in embryogenesis remains unchanged (Bartosová et al., 1984).

Vellus body hair changes into terminal hair before puberty. During puberty the stimulative effect of androgens makes changes more profound. The first hair to develop into terminal hair is pubic and armpit hair as they respond to normal levels of androgens. Armpit hair gradually becomes thicker, and the shaft turns around its axis, reaching a length of 10-60mm (Bartosová et al., 1984).

As age increases terminal hair appear on the face of teenage boys, starting from the upper lip, then on the cheeks, and lastly on their jaw (Bartosová et al., 1984).
Terminal hair on hands and feet and the rest of the body develops at a later stage and is sex-dependent. In men, terminal hair develops on the thighs, belly, glutes, shoulders, back, etc. (Reynolds, 1951). Changes are due to male hormones that act on the growth, diameter and extend of medulla appearance (Ebling, 1987). In women, around 35% of hair follicles will develop to produce at least one hair, while in men, this percentage is 90%.

Hair density

The total number of hairs in a human is around 5 million (with 1 million found on the head). There are no differences in race or sex regarding this number (Montagna and Carlisle, 1990). As the individual ages there is a significant loss of hair. Average hair density is 615 hairs per cm^2 between 20-40 years of age, goes down to 485 hairs per cm^2 in those 30-50 years of age and reaches its lowest of 435 hairs per cm^2 in 80-90 years of age (Ebling, 1987; Spearman, 1977).

Rate of hair growth

Hair growth differs depending on the location of the hair, sex and age. For most of the body hair, there is a daily increase of 0.21-0.39 mm, while in facial hair, this is 0.44 mm, and in scalp hair, around 0.5 mm (Barman et al., 1964). In women, the speed of hair growth is slightly

higher than in men (Saitoh and Uzuka, 1969). In the period just before puberty, hair growth becomes faster in boys than in girls (Pecoraro et al., 1964). Studies in guinea pigs have shown that the speed of hair growth is highly dependent on the phase of activity the hair follicle is in (Johnson, 1977a). Most researchers believe that shaving and washing does not affect the speed of hair growth, as initially suggested (Saitoh and Uzuka, 1969; Lynfield and Macwilliams, 1970).

Some hormones may affect the speed of hair growth in animals (e.g. oestrogens decrease it and thyroxine increases it). In humans the speed of hair growth increases due to androgens (Ebling, 1987). Warm climates tend to increase the rate of hair growth while on the other hand, colder climates decrease it (Johnson, 1977b). Hair growth, in general, is dependent on the overall health of the individual, their endocrine status and their nutrition (Ebling, 1987; Bartosová et al., 1984).

2.2 Anatomy and Embryology of the hair follicle

The hair follicle

The hair follicle (HF) is a unique and complex biological system responsible for the production of hair. Hair is a characteristic of mammals and its primary aim is to provide thermal insulation to the body. Human hairs are developed through a large number of tiny follicles found throughout the body and play an important role in an individual's psychology.

During adult life every HF runs cycles of regression and regeneration of its lower part, which produces the shaft. Many of these molecular phenomena which are responsible for the development of HF also influence their development and life cycle. Molecular and cellular biology research has produced a significant amount of data, which, sooner or later will lead to their clinical application in order to confront hair pathologies. To exploit these data, it is essential to understand HF anatomy and development.

2.2.1 Anatomy of the hair follicle

The human scalp contains around 100,000 HFs and is the densest area of the skin in terms of HFs (400-500 HF per cm^2). HF density is much higher in Caucasians than in African Americans (Sperling, 1999). After the formation of lanugo hair, which is the characteristic of the gestation period, two types of hair are developed. Terminal hairs are more than 60µm in diameter and can have a length of more than 1m (Garcia et al., 2011). Medulla is the innermost region of the hair and is generally amorphous in appearance in humans. The hair bulb of terminal anagen hair is found subcutaneously. On the other hand, vellus hair has a diameter of less than 30µm, does not contain a medulla and never exceeds 2 cm in length.

Hair convexity varies intra-individually and inter-racially. Convex hair shafts are developed by convex HFs and the shape of the cortex defines the shape of the hair. Cross-sections of straight hair are circular, while curly hair tends to be more elliptical (Schlake, 2007). The factors defining the size of hair are unclear, despite the fact that mutations in the EGFR gene may cause the development of curly or wavy hair in mice (Du et al., 2004).

The follicle has two distinct parts:

1) the upper follicle which is composed of the infundibulum and isthmus. Infundibulum extends from the opening of the sebaceous gland to the surface of the skin, and isthmus the lower portion of the upper part of HF is located between the opening of the sebaceous gland and the insertion of the arrector pili muscle. (Erdogan, 2017)
2) The lower follicle is divided into two distinct areas: the suprabulbar and bulb regions.

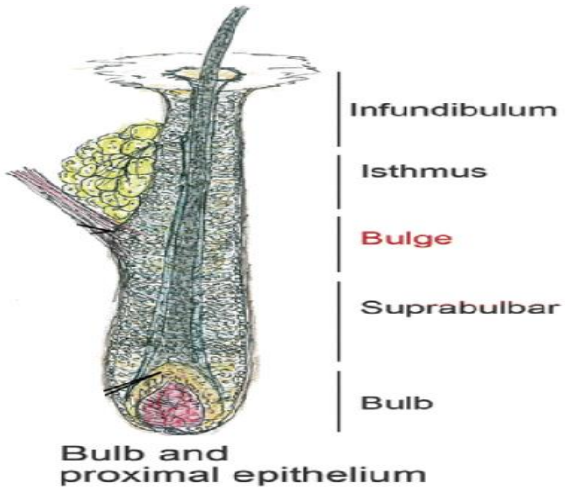


Figure 2.2.1: The upper and lower parts of the hair follicle

Infundibulum

The supraglandular region or infundibulum is the region between the epidermis and the opening of the sebaceous gland (Sperling, 1991). The epithelium of this region is a continuity of the epidermis and can regenerate from its cells. Keratinisation is done in a similar manner to the epidermis with formation of stratum granulosum and stratum corneum. Despite the morphological resemblance to epidermis, infundibulum epithelial cells have a higher mitotic activity and can contribute to its regeneration after impairment.

Isthmus

Isthmus extends between the opening of the sebaceous gland and the insertion of arrector pili muscle. In contrast to infundibulum's epithelium there is a lack of granular layer producing a more solid keratinisation, called "trichilemmal keratinisation" (Pinkus et al., 1981). At the insertion of the arrector pili muscle the fully keratinised interior epithelial layer is disorganised. Above this point there is trichilemmal keratinisation of the outer epithelial cells. Cysts found in this area are called trichilemmal cysts (Brownstein, 1979). The lower part of isthmus contains the HF bulge. The bulge consists of a biochemically distinct population of epithelial cells resembling in characteristics to primordial epithelial cells. They express cytokeratin 15 (Porter et al., 2000), rarely multiply (except for the initial phase of anagen) and can differentiate to a variety of cell types including: epidermal keratinocytes, sebaceous gland cells and at least seven other types of epithelial cells in the lower HF.

Suprabulbular region

The suprabulbar region of the HF is above the hair bulb and below the isthmus. It is comprised of three layers from outermost to innermost: outer root sheath (ORS), inner root sheath (IRS) and hair shaft. In this region the three layers of the inner epithelial cells are: Henle's, Huxley's and the cuticle layer (Ito, 1990). Cells in the periphery of the outer epithelial layer are cylindrical and progressively their conciseness in glycogen is reduced.

Hair bulb

The bulb is the lower part of the HF and contains the follicular dermal papilla and the hair matrix. Hair matrix cells are localised in the lowermost portion of the follicle surrounding the follicular papilla. As matrix cells divide and differentiate, they form a cellular column that develop into the hair shaft and inner root sheath. The rate of their proliferation is one of the highest in body tissues. The cells of the future hair shaft are positioned at the apex of the dermal papilla and will form the medulla (in terminal hairs), cortex, and hair shaft cuticle. At the side of these are the cells which will develop into the inner root sheath: its neighbouring single layer will become the cuticle of the inner root sheath, the next two to four cell layers will become Huxley's layer, and the next single layer will become Henle's layer. The pre-keratogenous zone is situated over Auber's critical line and extends to the region where the IRS and shaft cornify and lose their keratohyalin granules. The actual line of keratinisation changes with each layer of the IRS and shaft: Henle's layer is keratinised first, followed by the IRS cuticle, and lastly Huxley's layer and the hair shaft.

Dermal papilla

The dermal papilla is a structure of mesenchyme origin surrounded by matrix epithelium and consisting of fibroblasts, collagen bundles, a layer rich in mucopolysaccharides, nerve fibers and a single capillary loop. Tissue recombination studies have shown that dermal papillae retain their ability to provoke HF development even when grafted on the epidermis from the sole of feet. These findings suggest that the type of tissue defined during gestation can be altered under specific conditions. Differences in HF dermal papilla volume are due to extracellular matrix volume and cell number (Elliott et al., 1999). Dermal papilla volume determines hair volume.

Dermal papilla synthesises and releases many soluble growth factors which act with a paracrine mechanism on incumbent matrix epithelial cells. For example, keratinocyte growth factor (KGF, also known as FGF7) is synthesised by anagen dermal papillae and its receptor FGFR2 is found in matrix keratinocyte.

Sebaceous gland

Hair sebaceous glands secrete, through a holocrine mechanism, sebum which consists of squalene, cholesterol and esters of cholesterol, wax esters and triglycerides (Harkey, 1993). Cells at the periphery of the gland are small, include a nucleus and are free of fat. They divide and are impelled towards the inner side of the gland where they produce and store fat. After that they lose their subcellular organelles, disintegrate and empty via a squamous epithelium-lined duct into the upper portion of the hair canal. Apart from its role in sebum production, the sebaceous gland is also important in the physiological development of the hair shaft. In mutant asebia mice, which lack a gene responsible for fatty acid metabolism, sebaceous glands are either missing or hypoplastic and the internal epithelium layer adheres on the hair shaft, thus obstructing its exit from the follicle (Stenn, 2001). Considering that bulge matrix cells continue to produce a new hair shaft, this leads to perforation of the follicular epithelium and finally the appearance of scarring alopecia.

2.2.2 Embryology of the hair follicle

Hair types

During gestation HF development starts from the scalp and descends to the rest of the body. Initially, HFs produce lanugo hair which is soft and falls off between the 32nd and 36th week of gestation, even though almost one third of babies are born with it (Dawber, 1988). Most follicles then produce vellus hair. Vellus hair increases the sense of touch on the skin acting as pressure receptors. It should be noted that after birth, there is no further production of HFs in humans.

Molecular level of hair development

Important progress in the understanding of HF development was made with the discovery that respective mammals genes found in *Drosophila* may influence HF development. Homeobox (HOX), hedgehog (HH), patched (PTCH), wingless (WNT), disheveled (DSH), engrailed (EN), NOTCH1 and armadillo-β-catenin (CTNNB) are very important genes in HF

development of vertebrates (Rishikaysh et al., 2014). Because these genes were initially found in *Drosophila*, most of the names given to them describe the phenotype of the respective *Drosophila* mutations.

In human foetuses, HFs develop by a small collection of cells called the epithelial placode. The placode appears in a homogenous epithelium during the 10th week of gestation. Cells in the placode multiply to form a primitive form of the HF. The non-random, symmetrical distribution of HFs is probably defined by Homeobox genes which are known to express in a constant and specific way in space and time during the development (Stelnicki et al., 1998; Noveen et al., 1995). In adult mice, the expression of Homeobox genes recurs in HFs and serves the maintenance of producing a physiological hair shaft. Homeobox-type genes are engrailed, and are responsible for the head to tail type of development. Mice lacking such genes developed HFs on their feet.

Previous studies showed that HF development is dependent on a series of interactions between the mesenchyme and the epithelium (Rendl, Lewis and Fuchs, 2005). An initial signal, the first dermal signal, derived from mesenchyme leads the incumbent epithelium to form the placode (Schmidt-Ullrich and Paus, 2005). The second signal arises from the placode and leads to an accumulation of cells in the underlying mesenchyme forming the dermal papilla. Finally, a signal from the dermal papilla induces the mitotic activity and the differentiation of the placode causing the formation of a mature HF. These mutual signals pass through the basic membrane which undergoes changes of thickness and composition, altering its ability to isolate growth factors and bound proteins, thus modifying the interactions between epithelium and mesenchyme.

At the current time, many regulating molecules have been discovered, however the mechanism by which they develop the HF is still unknown. In a theoretical model, the arrangement in space and the size of the placode are regulated by a dermal signal the characteristics of which vary in different regions of the body. This signal is uniform in all body regions and activates promoters or inhibitors of the HFs in the epithelium from the interaction of which, the totality of the HF spectrum is created. Differences in the activation levels of these molecules may be responsible for regional differences in the size and arrangement of the HFs. Based on this model, several positive or negative regulators of the HF development are originally expressed in the epidermis and are afterwards found in the placode.

One of the primary molecular pathways activated in HF development is the WNT/β-catenin pathway (Andl et al., 2002). β-catenin is a mediator that down-regulates WNT signals. WNT proteins bind on cytoplasmic membrane receptors and through a series of signals inhibit the disintegration of cytoplasmic β-catenin which is then transported to the nucleus where it forms complex with the family of transcription factors TCF/LEF resulting in the expression of down-regulating genes. Naturally, β-catenin is inactive in adult skin but if artificially activated in basal layer cells of adult transgenic mice, it induces *de novo* HF development (Gat et al., 1998). This remarkable finding may be of therapeutic importance, despite the fact that continuous activation of the pathway may develop rare skin tumours.

Ectodysplasin (EDA), a molecule linked to tumour necrosis factor (TNF), and its receptor EDAR are part of another primary pathway for HF development. Mutations on the EDA gene cause X-linked anhidrotic ectodermal dysplasia, a syndrome characterised by a reduced number of HFs and abnormalities of teeth and sweat glands (Kere et al., 1996). The EDAR gene is mutated in autosomal dominant or recessive hypohidrotic ectodermal dysplasia leading to similar phenotypes as with the EDA mutations. In mice, the EDAR gene is expressed in the whole epithelium before the formation of the placode and is then restricted to the placode only (Zhang et al., 2009). On the other hand, the EDA gene is expressed in the whole epithelium both before and after the formation of the placode. Mice with mutations in these genes present similar phenotypes with those appearing in humans.

In contrast to EDA and EDAR genes, which promote HF development, members of the bone morphogenetic proteins family (part of the transforming growth factor –β superfamily) inhibit their formation. Bone Morphogenetic Protein 2 (BMP2) is abundantly expressed in the ectoderm but is then located in the premature placode and the underlying mesenchyme. BMP4 is expressed in the dermal condensate. EDAR and Sonic Hedgehog (SHH) genes are required for the expression of BMP4, a fact which shows that EDAR is activated very early in follicles morphogenesis and is required both for the induction of the placode and the inhibition of placode expansion to neighbouring tissues (Pummila et al., 2007). BMP activity inhibitors, like Noggin, are essential for the physiological development of HFs. Mice lacking Noggin have fewer HFs and appear to have slower development (Botchkarev et al., 1999).

Notch signalling pathway also regulates the type of follicles development. Notch ligand δ1, is expressed naturally in the mesenchyme underlying the placode and seems to promote placode formation, inhibiting its formation in surrounding tissues (Powell et al., 1998).

Another protein secreted in the placode which has a more important role in signalling between the mesenchyme and epithelium is Sonic Hedgehog (SHH). Mice skin lacking SHH have very weak HFs and the dermal papilla are not completely developed. This shows that SHH controls the mitotic activity and size of HFs. Patched 1 (PTC1) receptor of SHH is expressed in stem cells and the underlying dermal papilla showing that SHH may have both autocrine and paracrine induction capabilities, necessary for the formation of the HF and dermal papilla (Millar, 2002).

The primitive follicle

In the next stage of embryonic development of the HF primary epithelium is converted to the primitive follicle. Mesenchyme cells of the sides of the initial follicle are developed in the fibrous sheath and those on the top of the initial follicle are converted to the dermal papilla. '*During hair follicle morphogenesis, melanocyte precursors migrate into developing HFs and give rise to differentiated melanocytes that actively produce and transport pigment into the keratinocytes that form the hair shaft*' (Botchkareva, 2003). Several factors affecting the development of the melanoblasts have been identified. These include SOX10, transcription factor PAX3, the basic transcription factor linked to microphthalmia transcription factor MITF, receptor B of endothelin and its ligand, endothelin3 and tyrosine kinase receptor (KIT) and its ligand KITL (Potter et al., 2006). Experiments in mice showed that KIT and KITL are necessary for the survival, mitotic activity and migration of the melanoblasts. In addition, these factors are necessary for the transport of the melanocytes from dermis to epidermis. Mutations of these genes are responsible for inherited disorders of hair pigmentation in humans. For example, mutations of KIT may cause poliosis of the hair (Richards et al., 2001) and Waardenburg syndrome is caused by mutations in PAX3, MITF, Endothelin-B receptor, endothelin-3 or in SOX10 (Seco et al., 2015).

The primitive bulb

The primitive bulb is formed by the elongation of the primitive follicle, the lower part of which surrounds the dermal papilla and forms a bulb-like structure. Epithelial cells of this structure become the matrix cells of the HF and produces the hair shaft. Moreover, melanocytes migrate to the inner layers of the matrix (Scott et al., 1994). The outer epithelial layer forms two bulges at the back of the HF. The sebaceous gland is formed from the surface bulge. The lower part of the bulge forms the position of adhesion of the arrector pili muscle. As the arrector pili muscle attaches to the back of the HF, its contraction causes a vertical move of the hair shaft so that a larger volume of air can be trapped between the hair and hence provide thermal insulation (Tansey and Johnson, 2015).

Notch-1 a membrane protein and its ligands, Serrate1 and Serrate2, are expressed in the matrix in order to form the hair shaft (Favier et al., 2000). Notch-1 seems to regulate the phenotype of keratinocytes from the time these migrate from the bulb's matrix and differentiate into multiple cell types.

Formation of the mature hair follicle

The central canal in which the hair shaft is placed is formed by the keratinisation of the supraglandular epithelial cells (Veniaminova, 2013). According to Millar et al. (1999), during the formation of the hair shaft, the WNT signal transduction pathway plays a crucial role, and researchers have observed that the ectopic expression of WNT3 can lead to a fragile hair shaft.

Genes in hair shaft keratin contain binding sites for LEF1 (Zhou et al., 1995), which is transposed to the nucleus when the WNT/β-catenin pathway is activated. Other transcription factors regulating hair shaft differentiation include HOXC13 (a homeobox protein) (Potter et al., 2006) and Forkhead box protein N1 (FOXN1). The latter regulates the expression of a gene coding an acidic keratin of the hair (Meier et al., 1999). In a case involving two brothers, who were carriers of a mutated form of FOXN1, both appeared to suffer from immune, hair and nail disorders (Bosticardo et al, 2019).

2.3 Hair Cycle

During a human's lifetime, cells from the skin's basal membrane differentiate and migrate to its surface, from where they fall off and are replaced by proliferating basal layer stem cells (Myung and Ito, 2012). On the other hand, hair follicles regenerate through a circular process of three phases: anagen, catagen and telogen. Approximately a decade ago, researchers described a distinct phase known as exogen, which encompasses the final stage of hair shedding (Higgins et al., 2009).

This cycle of growth and differentiation of the cell, which will eventually produce the hair shaft, is orchestrated by epithelial progenitors called Hair Follicle Stem Cells (HFSC), located in the hair follicle bulge (Paus and Cotsarelis, 1999). HFSC are self-renewing and verstile, possessing the ability to regenerate all epithelial layers of the hair follicle throughouth an individula's lifespan, if required (Blanpain et al., 2004). The discovery of specialised molecular markers (Lyle et al., 1998) demonstrates the heterogeneity of HFSC cell subpopulations and explains the dynamic of their function and behaviour in hair regeneration process, in healing process and their possible implications in carcinogenesis (Rogers et al., 2006). The term bulge was coined in 1876 by Paul Gerson Unna (Oshima et al., 2001), to describe an epithelial swelling at the base of the ORS, observed in foetal histological studies. The original study which indicated the presence of multifaceted epithelial cells in the bulge region used data from kinetic studies to identify epithelial germ cells by labelling them with 3H thymidine. These kinetic studies highlighted HF's ability to undergo slow cellular cycles, and multiply rapidly after receiving an appropriate signal originating from the dermal papilla. Based on these findings, the bulge activation hypothesis was formed (Sennett and Rendl, 2012). In detail, according to this hypothesis, label-retaining cells (LRCs) located on the bulge are activated due to interactions with the underlying dermal papilla and form cells (transit amplifying cells) that migrate to the hair bulb, and resulting in the initiation of the regeneration of the lower part of the hair follicle (anagen phase). According to the same hypothesis, cells in this label retaining region are decomposed in catagen, and cells in the bulge of telogen follicles express specific markers, including keratin 15 (KRT15) and CD34. This showed that the lower part of the anagen follicle comes from KRT15+ HFSCs. Further *ex vivo* studies have shown that these cells can differentiate into all the different types of cells found in the skin. Despite these facts, it is still unclear whether each telogen hair follicle cell can

develop into all epithelial cell types that make up the anagen follicle, or if there are specific subpopulations of cells that differentiate into those types, or if they simply reproduce themselves (Cotsarelis, 2006).

Although it was initially considered that bulge cells can regenerate all epithelial sheaths of the hair follicle and sebaceous gland, subsequent studies gave rise to a new theory regarding follicle regeneration. According to this theory (called hair follicle predetermination hypothesis) (Panteleyev et al., 2001), in mature hair follicles, bulge cells form the outer root sheath that surrounds the IRS, matrix keratinocytes and the shaft, which are derived from secondary hair germ cells (sHG). These cells are deposited after morphogenesis and during the first telogen phase, in the region between the dermal papilla and telogen club hair, where they remain. Indeed, immunohistochemical analyses, using markers such as CD34 and the calcium-bound S100A4 and S100A6 proteins, have identified a population of secondary germ cells, above the dermal papilla, which have the ability to self-renew (the mechanism is still unknown) and are biochemically distinct from the bulge cells. However, during the anagen phase as the follicle sinks into the dermal papilla, both sHG and bulge cells multiply and appear morphologically similar (Ito and Kizawa, 2001).

Recently, it has been shown that sHG cells are activated and multiply at the end of the telogen due to a stimulation from the dermal papilla and before the activation of bulge cells which occurs at the beginning of anagen, supporting hair growth during this period (Greco et al., 2009). It should be noted that distinguishing the end of telogen from the start of the anagen is not an easy task.

Despite the theories that have been postulated, so far no studies have yet proven that the cells in these two regions (sHG and bulge) are preprogrammed to create specific parts of the hair follicle. Bulge cells have the ability to produce follicular, sebaceous gland, and epidermal cells, as proven in reconstitution models. On the other hand, sHG can produce bulge cells when these are destroyed. In the bulge and sHG regions, there are also melanocyte stem cell which in the anagen phase are activated and produce mature melanocytes that give hair its pigmentation (Nishimura et al., 2002).

Hair follicle cycle

Every hair follicle has a part which is recycled and sequentially goes through three phases of evolution (anagen, catagen, telogen). Conversely, the bulge, infundibulum, isthmus, sebaceous gland and hair canal, where the hair shaft develops, are thought to remain constant, as there are no apparent changes in these structures during hair cycle. However, no section remains unchanged if we consider the apoptotic activity and vigorous cell proliferation during the cycle. The changes of the hair follicle from the telogen phase through the six stages of the anagen phase (I-VI) and the eight stages of the catagen phase (I-VIII), follow a rhythmically repeated sequence of events that is controlled by locally acting chemical agents. Many molecular factors have been identified that control the transition from one phase to another. Research is therefore focused on identifying the molecular 'signals' that orchestrate the transition from one phase to the other (Stenn and Paus, 2001). Studies in mice have identified some molecular factors that appear to be associated with growth and development of the hair follicle, without any of them having been proven to be absolutely decisive for the process.

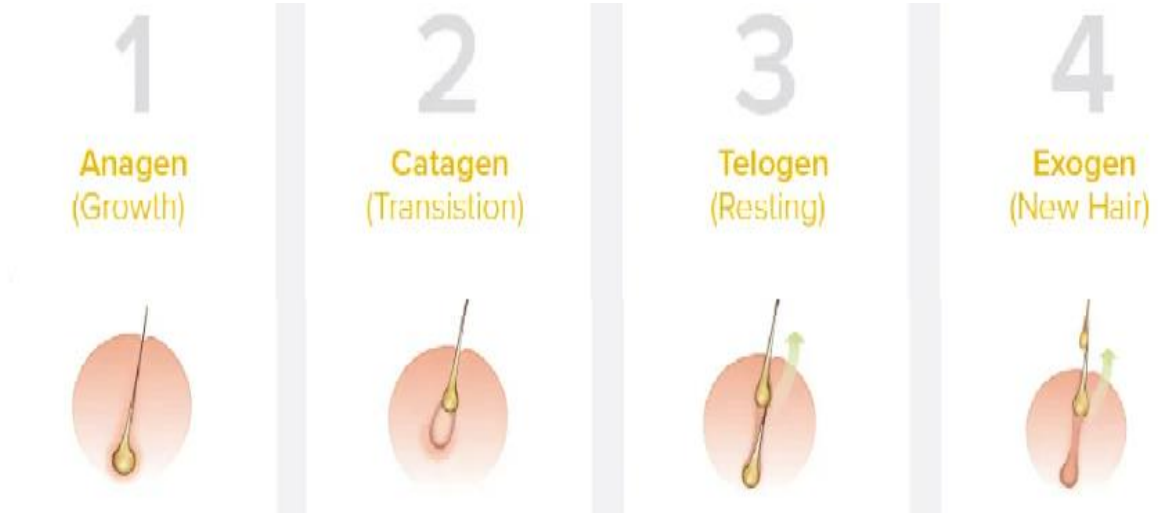


Figure 2.3: A diagram of the different phases of the hair cycle

Anagen phase

With the beginning of the anagen phase starts the process of hair follicle development by multiplication of its multifaceted cells due to a dermal papilla stimulation (Hardy, 1992). Although hair follicle development process seems to be the same during gestation and after the birth, it is not known whether the same molecular signals or the same mediators are involved in both processes.

Two factors that play an important role in the development and cycling of hair follicles are insulin-like growth factor 1 (IGF-1) and fibroblast growth factor 7 (FGF7). Both are produced in the dermal papillae and their receptors are found in the matrix cells (Danilenko et al., 1996). Mice lacking IGF-1 or its receptor have poor hair follicle development. Insulin-like growth factor 1 maintains and enhances follicle growth *in vitro* (Dawber, 1997). Mice lacking FGF7 have normal follicles, but with a non-functional receptor, that is also a receptor for FGF2, causes a profound reduction and ectopic hair follicle formation. Molecules promoting the transition from anagen to catagen phase include FGF5 and EGF growth factors, neurotrophins like BDNF and possibly the p75-neurotrophin receptor, factors of the p53 and TGF family and BMPRIa. Factors elongating the anagen phase include SGK3 and MSX2 (Alonso et al., 2005; Ito et al., 2002).

Hairs on different parts of the body are of different sizes, which are proportional to the duration of their anagen phase. Thus, the scalp follicles remain in anagen phase for 2 to 8 years and produce long hairs, while eyelashes stay only for 2 to 3 months and produce short hair. Ending of the anagen phase, is regulated by FGF5. Mice lacking FGF5 have a prolonged anagen, resulting in the "Angora" phenotype, with hair 50% longer than normal (Sundberg et al., 1997).

It seems that without the proper functioning of the receptors' growth factor system, the anagen phase is prolonged, while exogenous administration of epidermal growth factor (EGF) stops anagen phase in sheep (Hollis et al., 1983) and delays its onset in mice (Nakamura et al., 2013).

Catagen phase

During the catagen phase, follicles decrease in size, reflecting the programmed cell death (apoptosis) of the majority of follicular keratinocytes (Lindner et al., 1997). Additionally, at this stage, melanogenesis stops, and apoptosis occurs in some melanocytes (Slominski et al., 1994). At the end of this phase, the dermal papilla condenses and moves upwards, reaching the bulge. If the dermal papilla does not reach the bulge during catagen, the follicle cycle stops and the hair falls off. This process is controlled by a gene encoding a factor whose inhibition of transcription has the effect of blocking the movement of the dermal papilla and consequently causes permanent alopecia (Ahmad et al., 1998).

Telogen phase

During the telogen phase, the hair shaft matures and may often fall off the follicle during combing or bathing. It is not clear if this is the result of an active, regulated, or passive process occurring at the beginning of the anagen phase when new hair grows (Chuong, 1998). Most people, under normal conditions lose 100 to 150 hairs a day from their scalp. The telogen phase usually lasts 2 to 3 months, before the next hair cycle begins. The percentage of follicles in telogen phase depends on the area of the body (e.g., 5 to 15% of hair follicles on the scalp and 40 to 50% of body hair are in the telogen stage). Increased percentage of follicles in the telogen phase leads to hair loss. Therefore, drugs affecting the percentage of hair follicles in this phase could be considered as potential treatments of alopecia. The transition from telogen to anagen is similar to the activation of embryonic stem cells that initiate the de novo creation of the follicle. The anagen phase is then initiated by signal from factors affecting the WNT and HH pathways, while BMP mediators are involved in the differentiation of the follicle (Lowry et al., 2005; Callahan et al., 2004; Kulessa et al., 2000).

Exogen phase

The process of hair falling off the follicle is described as a separate phase called exogen. It is thought to occur in three possible ways:

1) Irrespective of the onset of the anagen phase (Exogen A) (Rebora and Guarrera, 2004)

2) At some point during the anagen phase of the next cycle and can be detained for 1-2 cycles (Exogen B)

3) Late in the anagen phase (stage VI) of the next cycle (Exogen C).

Since the time of hair apoptosis varies between different follicles, it appears that the exogen phase is independent of the other phases of the cycle (Stenn, 2005). Factors such as cathepsin L appear to facilitate hair apoptosis during this phase (Roth et al., 2000). The importance of the exogen phase for hair follicle biology is increasingly recognised, although more research is needed in regard to the molecular mechanisms regulating this phase cycle. Further research will reveal whether this phase implicates target factors for therapeutic intervention.

CHAPTER 3: Non-cicatricial scalp hair disorders

3.1 Introduction

Scalp hair plays a significant role in the individual's life. Dense hair is a sign of health, attraction, and youth. Hair thinning and/or hair loss can provoke feelings of insecurity, low self-esteem, and decreased acceptance in the work and social environment.

To fully understand diseases related to hair loss, it is necessary to comprehend their physiological development and the hair cycle. Hair development starts in the womb sometime between the ninth and twelfth week of gestation (Muller et al., 1991). Hair units are composed of epidermal follicles produced in the mesenchyme. As mentioned in the previous chapter, the embryo, during weeks 18-20 of intrauterine life, is covered by lanugo hair (Soo and Lorenz, 2004). Lanugo hair is the first generation of anagen hair (growth phase), which is followed by telogen phase (resting phase) and then by the falling of the lanugo during the 7^{th} and 8^{th} months of gestation. At the end of gestation or during the first couple of months in infancy, foetal hair is replaced by slightly coloured hair of about a 2cm length called vellus hair (Harkey, 1993). One of the characteristics of vellus hair is the lack of medulla, the portion of hair's core which provides its strength. Scalp hair gradually convert to darker, medulla-containing, hair called intermediate hair (Miranda et al., 2010). During puberty, intermediate hair is eventually replaced by terminal hair. Terminal hairs contain a medulla, are dense, pigmented, and of greater length than intermediate hairs due to their longest duration of anagen (Narisawa and Kohda, 1996).

The number of hair follicles remains stable after birth but their density is reduced by the skin's expansion. During birth hair follicle density is 1,135 per cm^2 (Inaba and Inaba, 1996), at 3 months old 795 per cm^2 (Inamadar et al., 2014), and at 20 years of age it is 615 per cm^2 (Shehan, 2006). As the years go by, hair shaft diameter increases with the greater increase observed in the first four to six years (Birch et al., 2001).

As mentioned in Section 2.3, a very important characteristic of the hair follicle is its intermittent cycle activity. Resting periods alternate with active periods, in terms of hair growth. Hair cycle includes three stages:

- Anagen phase is the phase of hair growth. It is characterised by intense activity, of the follicle needed to produce the new hair (Alonso and Fuchs, 2006). It is composed of six stages (Oh et al., 2016), starting just before the falling of the telogen hair. The duration of anagen in the scalp is 2-6 years. It is estimated that at any time 80%-85% of all hair are in anagen phase (Burg et al., 2017).
- Catagen phase is a transitional, intermediate phase during which the hair follicle is prepared for the fallout of the hair (Müller-Röver et al., 2001). Catagen lasts for about two weeks and 1% of all scalp hair are in catagen at any time.
- Telogen phase is the resting phase for the hair follicle. During this phase hair growth is suspended and the hair eventually falls. When this occurs, in exogen, a signal is sent to the hair follicle to start developing the anagen hair (Stenn, 2005). The duration of telogen is 3-4 months and at any time 13%-15% of all scalp hair are in telogen phase.

Assuming the scalp contains approximately 100,000 hairs with an average lifespan of 1,000 days, it can be estimated that approximately 100 hairs are lost every day. The majority of patients presenting to General Practitioners (GPs) with concerns over hair loss are due to effluvium and/or decreased density (alopecia). Alopecia is a visible decrease of scalp hair density reflecting a decrease of this density about 30%. The term effluvium indicates a loss of more than 100 hairs per day for a period of 2-4 weeks. Hair loss and thinning may be diffuse, focal or following a specific distribution pattern as a consequence of a non-cicatricial (non-scarring) and/or cicatricial (scarring process). A classification of non-cicatricial alopecia according to the type of lost hair and spread of hair loss is given in Table 3.1a-c:

Table 3.1 a-c: Classification of different types of non-cicatricial alopecia according to type of lost hair and extent of hair loss

Table 3.1.a: Non-cicatricial alopecia. Loss of anagen hair	
DIFFUSE	FOCAL
Anagen effluvium due to chemotherapy	Alopecia areata
Poisoning	Initial stage of androgenetic alopecia
Diffuse alopecia areata	Trichotillomania
	Traction alopecia
	Fungal alopecia
	Loose-anagen hair syndrome

Table 3.1.b: Non-cicatricial alopecia. Loss of telogen hair	
DIFFUSE	FOCAL
Acute telogen effluvium	Androgenetic alopecia
Post-natal alopecia	Frontal alopecia of the neonates
Chronic telogen effluvium	Alopecia due to syphilis
Alopecia areata	
Drug-induced alopecia	
Iron-deficiency alopecia	
Alopecia caused by malnutrition	
Alopecia caused by malabsorption	
Alopecia caused by thyroid diseases	
Alopecia caused by systemic lupus erythematosus	
Alopecia caused by chronic kidney insufficiency	
Alopecia caused by hepatic insufficiency	
Alopecia caused by malignancies	

Table 3.1.c: Non-cicatricial alopecia. Loss of both anagen and telogen hair
Androgenetic alopecia
Alopecia areata
Alopecia due to long-lasting hormone therapy
Alopecia due to chronic infections
Alopecia due to therapy with cytostatic agents

3.2 Diffuse alopecia - Anagen hair loss

Anagen effluvium

Anagen hair loss is considered a pathological loss of hair during the growing phase. Except in the cases of alopecia areata and loose anagen syndrome, this is due to the action of a potent external harmful factor. The majority of cases concern chemotherapeutic drugs and radiotherapy. In each case, a reduction of the metabolic activity of the cells of the inner epithelial sheath occurs followed by an abrupt pause of hair production in anagen, breaking of the hair shaft a few millimeters from the scalp surface and finally the heavy loss of anagen hair a few days or weeks later (Kanwar and Narang, 2013). Loss is nearly total as 80%-85% of all hair is in anagen phase and becomes more obvious 1-2 months after its start. Telogen hair remains in place up to the time they are programmed to fall. Chemotherapy-induced alopecia is usually reversible (Trüeb, 2010). It depends on the dosing and length of treatment (Kanwar

and Narang, 2013). Hair which regrows may have different colour, quality and form, or even lesions which may persist for years.

Some of the chemotherapeutics which have been implicated in the provocation of alopecia are listed in Table 2.2 given below:

Table 2.2: Chemotherpeutics inducing alopecia

Chemotherapeutic drugs implicated in alopecia	
Actinomycin D	Idarubicin
Bleomycine	Iphosphamide
Busulphan	Lomustine
Chlorambucil	Mechlorethamine
Cyclophosphamide	Melphalan
Cytarabine	Methotrexate
Dacarbazine	Mitomycin
Daunomycin	Mitoxantrone
Doxorubicin	Nitrosureas
Etoposide	Procarbazine
Fludarabine	Thiotepa
Fluorouracil	Vinblastin
Hexamethylmelamin	Vincristine

The prevention of chemotherapy-induced alopecia is questionable. For neoplastic diseases where the scalp is a 'pool' of cancer cells or a place for metastases, such as brain tumors, lymphomas, leukemia and malignancies of the haematopoietic system in general, it is wise not to take measures for the prevention of alopecia. A prevention measure which has been tested is a cooling apparatus which reduces scalp temperature and blood flow (Katsimbri, Bamias and Pavlidis, 2000; Komen et al., 2013). Topical application of 2% minoxidil also seems to reduce the length of alopecia but does not prevent its occurrence.

Hair loss and the potential to regrowing hair after radiotherapy is dependent on the type, dosing, length of time between doses and depth of treatment. Hair loss begins 8-10 days after radiotherapy and continues for 2-3 weeks. Regrowth is observed after 8-12 weeks. Permanent alopecia has been observed following radiotherapy. Studies in mice showed that administration of keratinocyte growth factor (KGF) or prostaglandin E_2 (PGE_2) promotes hair growth after radiation therapy (Booth and Potten, 2000; Ali and Singh, 2010). Moreover, topical application of minoxidil may speed this process.

The combination of chemotherapy and radiotherapy may have various results on hair loss (Tosti and Pazagglia, 2007). Administration of chemotherapy with specialised action at some stage of the hair cycle may produce partially synchronised cell populations which, according to their positioning in cell cycle, may increase or reduce the toxicity of the subsequent radiotherapy.

Anagen hair loss may also be observed in cases of mercury poisoning (Kanwar and Narang, 2013). It usually occurs due to long-term industrial exposure but can also occur in cases of consumption of contaminated water or seafood. Hair loss may be the single manifestation or may co-exist with other symptoms such as: neurological disturbances, acrodynia, headache, photophobia and hyperhidrosis.

Diffuse loss of anagen hair has been reported in case of arsenic (As) poisoning (Yu et al., 2018). Arsenic can be found in insecticides, rodenticides, fungicides, wood preservatives and glass structures. Acute arsenic poisoning manifests itself with symptoms of the gastrointestinal (GI) tract, hypotension, cyanosis, dyspnea, delirium, spasms, coma, acute tubular necrosis and hemolysis. Six weeks after an (acute) poisoning event, one can observe the characteristic transverse white line on the nails known as Mees lines.

A total loss of all scalp hair is observed in patients poisoned with thallium (Blain and Kazantzis, 2015). It is accompanied by symptoms of the autonomic and peripheral nervous systems, Mees lines are observed at a later stage. Initial symptoms include: insomnia, pain in hands and feet and intestinal colic. Anagen hair loss can also be observed after exposure to boric acid (Stein et al., 1973). Boric acid is a constituent of pesticides and can be found in many domestic products. Boric acid poisoning manifests with GI tract, central nervous system (CNS), renal and skin symptoms.

Telogen hair loss

Acute diffuse telogen effluvium

Acute loss of telogen hair is a usually self-limiting, non-cicatricial alopecia that occurs 2-3 months after exposure to a harmful factor. A wide spectrum of disorders is suspected as the aetiology: infections accompanied by fever, chronic systemic diseases, surgery, dietary changes, and emotional stress (Grover and Khurana, 2013).

A significant and extensive loss of telogen hair is observed after childbirth (telogen gravidarum). It affects 30%-50% of mothers. During gestation, high levels of circulating oestrogens extend the length of anagen resulting in improved hair appearance. The abrupt decrease of these hormone levels after birth leads hair to enter catagen and telogen phases and hair to fall a few months later (Malkud, 2015).

In each clinical case of acute telogen effluvium, a dramatic diffuse loss of hair is observed, leading to density decrease and thinning of the hair shaft. The prognosis of acute telogen effluvium is good, as it self-limits and is typically restored in 3-6 months' time. However, in a percentage of genetically predisposed women hair thinning may persist due to the activation of androgenetic alopecia. Clinical observations show that topical use of minoxidil may be helpful in prolonging the anagen phase and stimulating the induction of hair growth from telogen to anagen.

Chronic diffuse telogen effluvium

Diffuse loss of hair, which lasts for more than 3 months, is caused by primary chronic telogen effluvium or secondary telogen effluvium. The main reasons are: iron-deficiency, hypothyroidism, hyperthyroidism, systemic lupus erythematosus drugs and, more rarely, syphilis.

1. Alopecia due to iron-deficiency

Iron-deficiency, with or without anaemia, is a common finding during investigations of diffuse alopecia cases. On many occasions, therapy by iron supplements results in the cessation of hair loss and regrowth occurs in a short period of time (Pierre et al., 2010). In other cases, restoring normal iron levels does not result in the end of hair loss. It is evident that in these cases, iron deficiency is an unassociated finding and the primary reason for hair loss is an initiating androgenetic alopecia or chronic telogen effluvium.

Ribonucleotide reductase (RNR), an iron-dependent enzyme, is essential for DNA synthesis. It has been proposed that iron deficiency reduces hair matrix cells. Telogen effluvium is a result of this stoppage of matrix cells proliferation. Prognosis of iron-deficiency alopecia is good and hair is restored about 3 months after achieving proper iron levels.

2. Alopecia due to nutrition disorders

There are several nutritional reasons relating to hair loss. Sudden weight loss may lead to diffuse loss of telogen hairs. Prolonged starvation, and especially marasmus, results in dry, thin, straight and brittle hair. Kwashiorkor disease leads to diffuse telogen effluvium and/or mild forms of reduced hair diameter accompanied by increases in its length (Vandiviere, 1971). Changes in pigmentation are a primary characteristic of Kwashiorkor. Dark hair turns brown or red and brown hair turns blonde. Deficiencies of essential fatty acids leadto loss of telogen and lighter-pigmented hair. Zinc deficiency, both hereditary and acquired, results in dry, thin and brittle hair (Karashima et al., 2012). Hair loss in these cases may direct the practitioner to its early diagnosis.

3. Alopecia due to thyroid diseases

Diffuse loss of telogen hair is observed in both hypothyroidism and hyperthyroidism (Vincent and Yogiraj, 2013). Hair loss may precede other clinical manifestations and is considered a significant indication of early thyroid disease diagnosis. In addition, drugs that cause hypothyroidism, may provoke diffuse telogen effluvium. Diffuse telogen effluvium occurs in almost one third of hypothyroidism patients. In severe hyperthyroidism more than half of patients present with diffuse alopecia.

In hypothyroidism there is an inhibition of cell division in both the epidermis and hair. In a large percentage (almost half) of patients, this inhibition of mitotic activity induces the hair to catagen and delays re-entry of telogen hair to anagen. The mechanism of hair loss in hyperthyroidism is not fully understood at this time.

Therapy with thyroxine leads to decrease in hair loss and regrowth of anagen hair (Messenger, 2000) in a few months, except for the case of an unregulated hypothyroidism in which case hair follicles atrophy. A failure of therapy may indicate co-existing androgenetic alopecia. Hair loss in hyperthyroidism usually usually stops three months after restoration of euthyroidism.

4. Alopecia due to other diseases

Chronic renal insufficiency and haemodialysis cause brittle, diffuse hair, body hair loss including armpit and pubic hair (Hajheydari and Makhlogh, 2008). Chronic hepatic insufficiency, pancreatic diseases and GI diseases related to malabsorption syndrome or malignancies, may lead to chronic diffuse alopecia. Connective tissue diseases, such as systemic lupus erythematosus and dermatomyositis may be accompanied by non-cicatricial diffuse alopecia (Concha and Werth, 2018; Muro, Sugiura and Akiyama, 2016).

5. Drug-induced alopecia

A large number of drugs may interfere with the hair cycle and contribute to hair loss. It should also be noted that the same drug depending on the route of administration may, or may not, cause loss of either anagen or telogen hair (Tosti et al., 1994). For instance, it is common to observe hair loss with administration of etretinate or acitretin, a phenomenon which is dose-dependent. Hair loss is less frequent with isotretinoin. Hair loss is very frequent after discontinuing contraceptive pharmacotherapy, especially of those treatments containing a high ratio of progesterone to oestrogens. In general, hair loss starts 6-12 weeks after the initiation of therapy and increases gradually with continuation of administration. With the exceptions of androgenetic alopecia and cicatricial alopecia, there is regrowth upon discontinuing the causative drug. In some patients a permanent thinning of the hair is observed, and in a few cases, prolonged hair loss occurs. This is the case with retinoids, however the exact mechanism has not yet been defined. Some of the drugs causing diffuse telogen alopecia are listed below in Table 2.3.

Table. 2.3: Drugs implicated in diffuse telogen alopecia

Drugs implicated in diffuse telogen alopecia	
Acitretin	Glibenclamide
Albendazole	Heparin
Alopurinol	Interferon
Amiodarone	Isotretinoin
Amphetamines	Levodopa
Anabolic steroids	Mepacrine
Bromocryptine	Methysergide
Captopril	Metoprolol

Carbimazole	Penicillamine
Chloroquine	Phenytoin
Clofibrate	Propranolol
Colchicine	Propylthiouracil
Contraceptives	Simetidine
Danazole	Sulphasalazine
Enalapril	Testosterone
Etretinate	Warfarin

Chronic telogen effluvium

Chronic telogen effluvium is defined as chronic diffuse loss of telogen hair that lasts for more than six months (Whiting, 1996). It is much less frequent than acute telogen effluvium. Diagnosis is usually made by excluding other aetiologies of diffuse hair loss. It is observed mainly in women and more rarely in men with long hair.

The influence of a harmful factor is rarely found in chronic telogen effluvium. Intense emotional stress has been proposed as a factor for the appearance of chronic telogen effluvium (Rebora, 1997; Grover and Khurana, 2013). The pathogeny of this condition is unclear but it is likely to be caused by a shortening of the anagen phase of the hair cycle. It affects women between 30 and 50 years of age with very dense, long hair. It is often linked to a history of childhood long hair, which may indicate a prolonged anagen phase.

Clinically, a recession of hair in the frontal and temporal areas is observed, without lesions in the parietal area. Prognosis is quite good and the disease resolves after 3-4 months and these women do not lose their hair. In very rare cases this phenomenon may persist for 10 years or more. Despite the high number of lost hairs over a period of many years, these patients have quite dense hair.

3.3 Androgenetic alopecia

Androgenetic alopecia (AGA) is the most common form of hair loss and thinning with a distinctive distribution in men and women. Regardless of age or sex, AGA has a substantial negative impact on people's quality of life and psychology. The aetiology of androgenetic alopecia is polygenic and multifactorial. The involvement of androgens and their interaction with androgen receptors seems to play a very significant role in its pathogenesis. The role of dihydrotestosterone (DHT) and the SRD5A1 and SRD5A2 enzymes, which catalyse the conversion of testosterone to DHT, has also been confirmed as potential pathogenetic mechanisms.

There are significant differences in the clinical appearance of AGA in males and females as they follow different patterns of hair loss. Treatment of mild to moderate AGA in males includes either oral finasteride and/or topical 5% minoxidil solution. Treatment of mild to moderate AGA in females is usually with topical 2% minoxidil solution. The therapeutic potential of antiandrogens and contraceptives is not justified for females with normal androgen levels. Prompt initiation of the above-mentioned therapies may delay the progression of AGA in either male or female patients.

Epidemiology of androgenetic alopecia

AGA affects almost half of the male Caucasian population over 30 years of age (Rhodes et al., 1998). There are ethnic differences in the prevalence of AGA, which is significantly higher (20%-40%) in Caucasians than in other populations (Paik, 2001). Prevalence of AGA is lower and less severe in African Americans, Native Americans, and Asians (Hoffmann and Happle, 2000; Xu et al., 2009; Tang et al., 2000). Asians, in particular, experience the onset of AGA almost a decade later compared to Caucasians (Takashima et al., 1981).

There are other factors that affect AGA. Most studies agree that the prevalence and extent of AGA increase with age (Gan and Sinclair, 2005). Alcohol consumption (Severi et al., 2003), diet, (Pathomvanich et al., 2002) and smoking (Su and Chen, 2007) are three social factors associated with an increased risk of AGA. Another risk factor is family history. Fathers' and

maternal grandfathers' hair loss status seem to play an important role in an individual's hair loss (Chumlea et al., 2004). Pre-pubertal castrated eunuchs are more likely to lose hair under androgenic treatment after 60 years of age than in their 20s. This shows that genetic predisposition, endocrine factors and aging are interdependent in AGA (Hamilton, 1951).

Associations of androgenetic alopecia with other diseases

A strong association of androgenetic alopecia and benign prostatic hyperplasia was described by Oh et al. (Oh et al., 1998). It seems that early onset of male-pattern baldness could be an early marker of prostate growth-associated urinary symptoms (Arias-Santiago et al., 2016). Early onset vertex balding has also been associated with early severe coronary heart disease in both male (Matilainen et al., 2001; Lotufo et al., 2000; Yamada et al., 2013) and female populations (Mansouri et al., 2005). Diseases associated with insulin-resistance such as hypertension, obesity, and dyslipedemias have also been related to AGA (Ahouansou et al., 2007; Hirsso et al., 2006). The incidence of prostate cancer is higher in men with androgenetic alopecia (He et al., 2018).

Aetiology and pathogenesis of androgenetic alopecia

There are approximately 55,000-60,000 hair follicles and approximately 100,000-150,000 hairs on the human scalp. Hair follicles are biologically complete structures including nerves, blood vessels, glands, collagen, and a microscopic muscle called arrector pili (see Figure 3.2 below).

Every follicle includes 1-4 hairs. Hence, hair follicles are structures in which specialised processes take place in order to achieve hair growth.

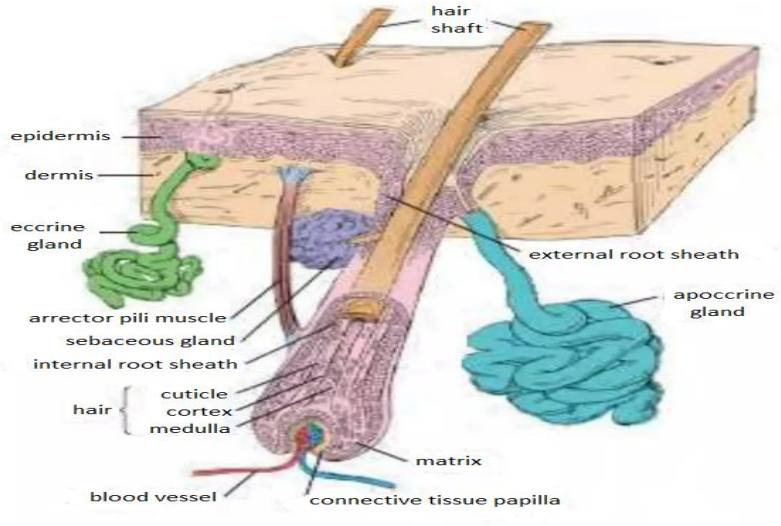


Figure 3.2 An illustration of the hair follicle and its structures

Hair life cycle is asynchronous and consists of three phases:

- **Anagen (Active growth phase)**: Its duration is 2-6 years. Hair follicles at this stage are characterised by intense activity and reside in the epidermal fat
- **Categen (Transition phase)**: Its duration is 2 weeks. At this stage activities and cell division stops. Hair follicles shrink in size, and hair stops to grow.
- **Telogen (Resting phase):** Its duration is 3-4 months. Hair follicles remain idle, and are extruded.

AGA is the consequence of a stepwise conversion of terminal follicles into smaller, almost invisible, vellus-like ones (Otberg, Finner and Shapiro 2007). This is caused by a decrease in the anagen phase duration and an elongation of the duration between hair shedding and anagen regrowth, leading to a reduction of hair present on the scalp.

The role of androgens

The mechanism of androgens' action on follicular cellular processes is unclear, but in most cases it is considered a local effect, (increased expression of androgen receptors, abnormal androgen metabolism) on the scalp's hair follicles. This means that most AGA patients are endocrinologically healthy. Terminal hair follicles are transformed to vellus-like follicles and atrophy. During consecutive follicle cycles, the produced hair gradually becomes shorter in

size and smaller in diameter (Ellis et al., 2002). On the other hand, androgens induce the growth of body and genital hair (Bergfeld, 1988).

Males who were castrated before or during puberty do not develop AGA even if there is a strong family history. Administration of androgen therapy in these individuals may cause AGA only in those predisposed. When androgen therapy is discontinued, baldness does not progress, nor does it reverse (Hamilton, 1942).

Dihydrotestosterone (DHT), an intracellular hormone, promotes the development of androgen-dependent hair (e.g., beard and genital) and the extrusion of non-androgen-dependent hair (e.g., scalp). In males, the most potent androgen is testosterone produced in the testes. In females, the main circulating androgens are androstenedione and dihydroepiandrosterone sulphate (DHEA-S), which are slowly converted in the keratinocyte to DHT with the mediation of 5α-reductase (Rebora, 2004). Both testosterone serum levels and the rate of its conversion to DHT, are much higher in males than females.

Other aetiological factors

In addition to androgens and genetic predisposition, other potential pathogenic factors of AGA are suspected, such as microbial flora, endogenous or exogenous stress, and microinflammation (Trüeb, 2002). At least in one third of AGA cases the damage results due to a microinflammation process, possibly mediated by the spontaneous production of IL-1α (Mahé et al., 2000).

The acute onset of alopecia in individuals with inflammatory diseases of the scalp has suggested a variety of aetiologies including the influence of the inflammatory cells, release of cytokines, presence of growth factors and increased interaction of stromal cells (i.e. fibroblasts, endothelial cells and adnexal keratinocytes) (Bergfeld, 1995).

Based on alopecic scalp biopsies, miniaturisation is frequently associated with perifollicular fibrosis. Perifollicular fibrosis is considered another possible AGA pathogenesis mechanism. It is thought that testosterone induces TGF-β1 (tissue growth factor β1 has a negative growth-regulatory effect on hair follicles *in vitro*) and type I procollagen (which is responsible for the

densely packed collagen bundles around miniaturised hair follicles in perifollicular fibrosis) (Yoo et al., 2006).

3.4 Pharmacological treatment of androgenetic alopecia

A prerequisite for therapy is the confirmation of the diagnosis of androgenetic alopecia. The purpose of treatment, especially in the early stages of androgenetic alopecia, is firstly to prevent hair loss and secondly to reverse miniaturisation and promote hair regrowth (Cranwell, 2016). Patients should be informed that currently, no drug can restore all lost hair (Lee et al., 2013).

Therapy should be administered for a period of 12 months to assess its efficacy (Dinh and Sinclair, 2007). However, in many cases, improvement can be observed earlier. Moreover, therapy should not be discontinued to maintain its efficacy (Fabbrocini et al., 2018). If therapy is discontinued for a few months, the results in terms of hair density are comparable to no therapy at all.

Patients should receive instructions to avoid products or techniques that may damage the scalp or scalp hair, follow a balanced diet and if feasible, stop taking drugs that are linked to hair loss. It is also essential to treat scalp conditions such as psoriasis and seborrheic dermatitis, as they can affect the use of topical therapies (Bergfeld, 2009). Therapy can be systemic, topical or a combination of both.

Males: Systemic therapy

Finasteride

In 1997, oral administration of 1mg/day of finasteride was the first drug to be approved by the FDA for the treatment of androgenetic alopecia (Anastassakis, 2002). According to 2012 European Medicines Agency (EMA) and United States Food and Drug Administration (FDA) guidelines, finasteride is indicated for men 18-41 years old with mild to moderate stages II-IV of androgenetic alopecia, as classified by Hamilton's scale (Hamilton, 1951) (Shapiro and

Kaufman, 2004). Finasteride is an SRD5A2 inhibitor (IC_{50}=25nM) that reduces DHT by up to 60% in serum, prostate and scalp hair (Batrinos, 2014). It is metabolised by CYP3A4 in the liver, and there is no evidence of interactions which are metabolised by the same enzyme such as warfarin, digoxin, theophylline etc. Results show that 1mg/day finasteride, leads to a significant increase of total scalp hair compared to placebo. After 24 months of therapy the increase is 6.2%, after 36 months 8.5%, after 48 months 7.2% and after 60 months 4.3% (Blumeyer et al., 2011). These increases are statistically significant in comparison to placebo.

Other studies have shown that half of the men aged 18-41 who received finasteride for a year, show an improvement in the appearance of the hair. This increased to 66% in patients of the same age who received finasteride for two years. The percentage improvement in men aged 41-60 was about 38% after two years of finasteride treatment (Olsen et al., 2012).

The efficacy of finasteride should be assessed six months after the initiation of treatment, and if effective, it should be continued to maintain its effectiveness. Discontinuing the therapy results in loss of any positive results seen up to that point, within 12 months (Cecchin et al., 2014). Finasteride has a negative impact on prostate specific antigen (PSA) levels. The decrease varies with age ranging from 20% in men aged 26-40 up to 50% in men aged 49-60 (D'Amico and Roehrborn, 2007). Finasteride side effects include a decrease of libido, ejaculation disorders and erectile dysfunction. There are also reported cases of decreased sperm count, azoospermia and decreased sperm motility with finasteride (Samplaski et al., 2013). Regarding its potential to cause prostate cancer, studies have not concluded whether it promotes or inhibits it (Thompson et al., 2003). The relationship between finasteride and prostate cancer is still subject to ongoing research and investigation, and more evidence is needed to determine the precise conditions under which finasteride may contribute to or prevent the occurrence of prostate cancer.

Dutasteride

Dutasteride is a potent inhibitor of both the SRD5A1 and SRD5A2 enzymes (Kaliyadan et al., 2013). It has been shown to be three times more potent than finasteride in inhibiting SRD5A2 and one hundred times more potent in SRD5A1 inhibition (Zhou et al., 2019). Dutasteride results in a 90% reduction of DHT in serum compared to finasteride at 70% (Shanshanwal and

Dhurat, 2017). In scalp hair it reduces DHT by 51%, compared to finasteride which reduces DHT by 41%. Dutasteride 0.5mg/day was approved by the FDA in 2002, to treat benign prostatic hyperplasia (Dhingra and Bhagwat, 2011). Dutasteride 2.5mg/day resulted in a 10% increase in hair growth in more than 56% of men (Levy and Emer, 2013). With regard to side effects, dutasteride decreases sperm volume, libido and can cause gynecomastia (Amory et Al., 2007). In addition, it interacts with other drugs metabolised by the CYP3A4 (Olsen et al., 2006), whilst most interactions are moderate.

Males: Topical therapy

Minoxidil

Minoxidil was initially used orally for arterial hypertension. Its use in AGA was mainly due to the hypertrichosis side effect that occurred in these patients. The use of 2% minoxidil topical solution was approved by the US FDA in 1988, 5% solution in 1997 and 5% foam in 2006 (Rogers and Avram, 2008). Minoxidil exerts its action by its conversion to minoxidil sulphate through metabolism (Buhl et al., 1990). The exact mechanism by which minoxidil acts on scalp hair is not clear (Kumar et al., 2018). Minoxidil sulphate opens the ATP-sensitive potassium channels on cell membranes resulting in vasoconstriction (Shorter et al., 2008). However, vasoconstriction does not seem to be responsible for the increase of scalp hair. Other possible actions on hair follicles are:

(a) The increased expression of vascular endothelial growth factor (VEGF) mRNA (Lachgar et al., 1998)
(b) The activation of prostaglandin synthase-1 (PTGS1) which is a cytoprotective enzyme and hair growth regulator (Nieves and Garza, 2014)
(c) The increased expression of hepatocyte growth factor (HGF) mRNA, which promotes the increase in hair (Yamazaki et al., 1999).

Minoxidil is applied in a quantity of 1ml two times daily on dry scalp skin. It takes about one hour to be absorbed. After 48 weeks of treatment with the 2% solution hair regrowth was evident in about 41% of patients and 58% treated with the 5% solution. Response is faster with the 5% solution.

It should be mentioned that it is possible for minoxidil to cause telogen effluvium 2-8 weeks from initiation of therapy (Malkud, 2015). This temporary hair shedding is caused by minoxidil's promotion of re-entering anagen. This phenomenon is self-limited with the continuation of therapy which should not be disrupted. Minoxidil's effectiveness is higher in men below 40 years of age with recent hair loss in the vertex and fronto-parietal areas. Visible increase of hairs is observed after 8 weeks with its peak being in 4 months from the start of the therapy (Badri and Kumar, 2019). Even 4 years after therapy the number of hairs are higher than that in the beginning of therapy. Side effects are mainly dermatological in nature including: irritative dermatitis, dryness and itching of the scalp (Hagemann et al., 2005).

Finasteride and Minoxidil combination

Both drugs are effective at treating hair loss. Finasteride improves AGA in 62% of patients while minoxidil improves AGA in 58%. The efficacy is almost equivalent in terms of improving AGA, but minoxidil seems to produce results faster, and finasteride produces better results over the long term. When topical minoxidil is combined with oral finasteride results are much better (94.1% efficacy) than using the drugs individually (Hu et al., 2015).

Females: Systematic therapy

Antiandrogens

Antiandrogens act mainly by binding to the androgen receptor (AR) thus not allowing circulating androgens to bind to this receptor (Lai et al., 2012). Several chemicals have been tested which inhibit steroidogenesis and progestogen activity. However, there are insufficient data existing to support the use of antiandrogens (chlormadinone acetate, cyproterone acetate, drospirenone, spironolactone, flutamide) to improve AGA.

Cyproterone acetate can be used in women with clinical or biochemical findings of hyperandrogenism (Dinh and Sinclair, 2007). Cyproterone acetate is a progestin derivative often combined with ethinylestradiol to reinforce antiandrogen action and contraception.

Spironolactone 50-250 mg/day has been used in women with androgenetic alopecia due to hyperandrogenism of adrenal gland aetiology (Olsen, 2001; Sinclair et al., 2005; Rathnayake and Sinclair, 2010). According to these studies hair shedding was reduced and there was regrowth. Postural hypertension, electrolyte imbalances and mastodynia are side effects commonly observed with this drug (Helfer et al.,1988).

Flutamide 62.5-250 mg/day has been given to women with AGA due to hyperandrogenism of adrenal gland aetiology, but results were controversial (Yazdabadi and Sinclair, 2011; Paradisi et al., 2011; Diamanti-Kandarakis, 1999). Oestrogens act immediately as antiandrogens increasing the sex hormone binding globulin (SHBG) which results in a decrease in free testosterone and also inhibits luteinizing hormone (LH), resulting in a decrease of androgen production (Laurent et al., 2016). Even though oestrogens seem to delay AGA, data is inconclusive regarding their efficacy in re-growing hair (Randall, 2008). Administration of antiandrogens and contraceptives to postmenopausal women is not recommended.

Finasteride

Finasteride, in general, is not indicated for female pattern hair loss. Especially in women of reproductive age, finasteride can cause feminisation of male foetuses (Iorizzo et al., 2006). There are individual reports of good results in female pattern hair loss with finasteride 2.5-5 mg/day (Yeon et al., 2011; Trüeb, 2004) but these should be proven in controlled clinical trials.

Females: Topical therapy

Minoxidil

The use of 2% minoxidil solution has been approved by the FDA for the treatment of female pattern hair loss since 1991 (Fabbrocini et al., 2018), as it was proven that it improves the condition and hampers its progression. Although the 5% solution is not approved it has been used for many years. It is indicated for women over 18 years old with mild to moderate female pattern alopecia (stage I and II in Ludwig's scale) (Dinh and Sinclair, 2007; Ludwig, 1977). After 32 weeks of therapy with minoxidil 2% solution regrowth of hair was observed in 55% of women in comparison to 41% with placebo (Jacobs et al., 1993). Clinically, it appears that the efficacy is much higher in women than in men.

Minoxidil and Tretinoin

Minoxidil 2% or 3% solution combined with tretinoin 0.025% or 0.05% or 0.1% once daily has been tested in female pattern hair loss (Kwon et al., 2007; Yoo et al., 2007). When minoxidil 2% is combined with tretinoin 0.05% its absorption is increased three times (Shin et al., 2007). According to a study combination of minoxidil 5% and tretinoin 0.025% seems to increase hair by 20% in more than half of the women with female pattern hair loss (Bazzano et al., 1986). The use of such a regimen also increases the possibility of side effects such as: itching, erythema, folliculitis, eruptive pyogenic granuloma, hypertrichosis and photosensitivity (Baran, 1989; Melkote et al., 2009).

3.5 Pathways in Androgenetic Alopecia

WNT Pathway

The WNT pathway is found in all species and was first identified in the *Drosophila melanogaster* fruit fly through mutations in the Wingless gene, that often led to loss of the fly's wings. (Sharma and Chopra, 1976; Nusse and Varmus, 1982). The intracellular signalling of WNT ligands utilises two different signal transduction pathways: the "canonical pathway", which leads to the stabilisation of β-catenin protein, and the "non-canonical pathway", which is independent of β-catenin. Although these two pathways have been studied separately, there is evidence that they are interconnected. Canonical WNT signalling is the most significant and most studied in the hair follicle.

WNT genes encode a large family of glycoproteins, that function as extracellular signalling molecules, activating β-catenin. β-catenin is typically found in three distinct locations: cell junctions, where it interacts directly with E-cadherin, in the cytoplasm and in the nucleus. In most normal adult cells which do not receive stimuli, the WNT pathway is inactive and this is preserved by the absence of WNT protein and the degradation of β-catenin. When WNT/ β-catenin activation is absent, cytoplasmic β-catenin is phosphorylated by the nuclear protein

axin, by APC, by glycogen synthase kinase 3 (GSK3) and by casein kinase 1 (CSNK1) (Kimelman and Xu, 2006).

Activation of the WNT / β-catenin pathway occurs through the binding of secreted growth factors from the WNT family to Frizzled (FZ) receptors located on the surface of target cells. Up-regulation of WNT ligands expression, increased Fz receptor expression, methylation of secreted FZ-related proteins (sFRPS), AXIN1 mutation and activation of β-catenin associated with phosphorylation of tyrosine are thought to be responsible for the activation of β-catenin (Mo and Cui, 2012). Stabilised β-catenin is transferred from the cytosol to the nucleus where it interacts with the TCF/LEF1 transcription factors. The result is the activation of numerous genes that control diverse biological processes such as: embryonic development, cell processes (differentiation, migration, proliferation, survival), carcinogenesis and tumour suppression. Canonical WNT signalling is most probably tissue and cell dependent as until now there is no common set of target genes reported.

In terms of hair follicle growth and development WNT pathway is implicated in a range of processes. It is required to initiate the development of hair placodes in the overlying epithelium and is also crucial in the catagen to next cycle anagen transition (Das Gupta and Fuchs, 1999). The latter is due to the WNT co-receptors LGR5 and LGR6 binding to WNT agonist R-spondin (RSPO) (Li et al., 2016). The WNT pathway also triggers the expression of its own inhibitors DKK1 and DKK2. Overexpression of these inhibitors leads to a complete blockage of hair development (Andl et al., 2002). In the case of androgenic alopecia, androgens deregulate dermal papilla cells by inhibiting the Canonical WNT signalling pathway, which leads to hair follicle miniaturisation (Leiros, Attorressi and Balaña, 2012).

BMP Pathway

Bone morphogenetic proteins (BMPs) are extracellular proteins that act by binding to membrane receptors. Their molecular signalling pathway is similar to that of other molecules of the TGF-β superfamily. Their activation is achieved through special kinase receptors of serine/threonine and signal transduction from the cell membrane to the nucleus is achieved through Smad proteins.

BMPs bind and activate heterodimeric complexes comprising Type I receptors (subtypes: ALK2, ALK3, ALK6) and Type II receptors (subtypes: BMPR-II, ActR-II, Actr-IIB). These complexes, in turn, phosphorylate SMADs (SMAD 1/5/8). Phosphorylated SMAD proteins create a heterodimeric complex with intermediate SMAD4 and are transported to the nucleus where they act as transcription factors to regulate the transcription of BMP target genes. The description above is a synopsis of the so-called canonical BMP pathway. However, there are also other signalling pathways comprising the non-canonical pathway. BMPs activate the MAPK pathway and through TAK1/TAB1/XIAP lead to the activation of p38 and JNK (Yamaguchi et al., 1995; Guo and Wang, 2009).

The BMP inhibitor Noggin is another essential component of this signalling network as it modulates powerful BMP signals by blocking BMP2, BMP4 and BMP7 binding to the receptor (Zimmerman et al., 1996; Groppe et al., 2002). Noggin is upregulated in the hair germ and dermal papilla at the onset of the hair growth (Plikus et al., 2008). Noggin deficiency brings a significant delay in hair placode induction and its overexpression in the epidermis results in thickening of the epithelium, increased hair density and altered hair types (Plikus et al., 2004). Generally, for hair follicle initiation, BMP activity needs to be downregulated by Noggin, which results in induction of LEF1's expression and maintenance of WNT signalling in the hair germ cells (Samuelov et al., 2015).

While WNT pathway provides a promoting signal, BMP in general has an inhibitory impact on hair follicle development and regeneration, yet keeping a delicate balance for the tissue activation (Ahn, 2013). WNT and BMP pathways interact through a variety of mechanisms. The LEF1/TCF complex and β-catenin form a complex with SMAD4. BMP-dependent MSX2 activation is also dependent on LEF1 and SMAD4 binding on MSX2's promoter (Hussein et al., 2003). BAMBI, a transmembrane receptor activated by the WNT/β-catenin pathway acts as BMP inhibitor (Sekiya et al., 2004). DLX3 ablation has been reported to lead to hair loss through loss of BMP (Michel et al., 2017).

Sonic Hedgehog (SHH) Pathway

The hedgehog (HH) gene was first discovered by Nusslein Volhard and Eric Wieschaus in *Drosophila melanogaster*, during a genetic study of the fruit fly's genetic material, which

presented developmental abnormalities of its wings. The HH gene was found to be one of the many genes that were important in the differences observed between the front and rear wings of *Drosophila*. The name Hedgehog was used because *Drosophila* mutations in the gene made their hair look like those of a hedgehog. HH genes can also be found in other organisms, such as humans, mice, chicken and zebrafish (Roy et al., 2001). In these species, HH proteins are important for the development and morphogenesis of many different tissues.

In mammals there are three different HH subtypes with a wide range of functions that do not overlap: Sonic hedgehog (SHH), Indian hedgehog (IHH) and Desert hedgehog (DHH) (Echelard et al., 1993). SHH is necessary for the formation of many tissues during development including hair follicles. The basic mode of operation of this pathway has been evolutionary conserved as it remains the same from the fly to human. The beginning of signalling occurs as soon as the HH ligand is produced from the corresponding cell producer. Then the HH factor binds to the PTCH receptor of the cell where it acts, which causes the activation of SMO, which acts as its switch signal. SMO transfers the signal from the cellular membrane in the cytoplasm by activating transcription factors, called GLI factors, that regulate gene transcription path targets.

Sonic hedgehog (SHH) has been implicated in hair morphogenesis and regeneration (Biggs and Mikkola, 2014). Treatment of C57/B6 telogenic mice with an SHH agonist, made entry into the anagen phase apparent within seven days of treatment (Harel et al., 2015). Vitamin D receptors (VDRs) are important in controlling the expression of genes of hair follicle cycling including those of the HH pathway (Fawzi et al., 2016). FGF-1, FGF-2 and FGF-10 (Lin et al., 2015) and Noggin (Datta et al., 2009) upregulate SHH expression. Zinc is a component of zinc finger motifs for many transcription factors, which regulate hair growth through HH signalling (Kil et al., 2013). Vismodegib, a HH signalling pathway inhibitor, causes non-scarring alopecia as a side-effect (Rubio-Gonzalez et al., 2018). BMP4 downregulates SHH (Datta et al., 2009).

Fibroblast growth factor (FGF) Pathway

The FGF family contains 22 FGF ligands that signal through 7 isoforms of FGF receptor tyrosine kinases (FGFR), encoded by 4 genes. The signal transduction may go through MAPK, PI3K/PKB, PLCG and STAT intracellular pathways (Ornitz and Itoh, 2015). In the developing skin, FGF

ligands are mainly expressed in the dermal component. Among the four genes of FGF receptors, three are expressed in the skin: FGFR1 mainly in the dermis and FGFR2 and FGFR3 mainly in the epidermis. Mice lacking IIIb isoform of FGFR2, FGFR2IIIb, had reduced number of hair follicles and retardation of hair growth (Petiot et al., 2003). However, for hair follicle initiation, FGFR2IIIb needs to be downregulated in the placodal epithelium as it transmits signalling that favours interfollicular epidermal differentiation (Richardson et al., 2009).

FGF signals function during multiple stages of hair development (Rishikaysh et al., 2014). FGF is believed to promote proliferation in dermal papilla cells by increasing FGF-7 and β-catenin and activating ERK signalling. (Gupta and Carviel, 2016). FGF contributes to hair follicle stem cell activation. Expression of FGF-7 and FGF-10 increase within the dermal papilla during telogen. FGF-7 instructs hair germ to proliferate and initiate a new hair cycle. FGF-7 acts on the dermal papilla to prolong anagen (Sclafani, 2014). FGF-20 mediates dermal fibroblasts condensation, which follows placode formation (Rishikaysh et al., 2014). FGF-9 promotes follicle neogenesis (Anitua et al., 2017).

Telogenic mice treated with topical fibroblasts FGF-1, FGF-2, and FGF-10 induced an anagen phase and prolonged the mature anagen phase, probably by activating β-catenin and SHH (Lin et al., 2015). On the other hand, FGF-5 acts as an inhibitor of hair growth and elongation during anagen and induces catagen (Herman and Herman, 2016).

3.6 Linking theory to practice

Chapters 2 and 3 investigated current data from different sciences regarding AGA. The purpose of this research was to shed light on underlying mechanisms of this pathology and identify possible anatomic structures, targets and biochemical pathways which could be implicated in its pathogenesis. As mentioned in these chapters, AGA is a complex (polygenic and multifactorial) condition, and as long as there is no definitive therapy, people will turn to alternative therapies (e.g ethnopharmacologically used plants). It is our duty as scientists to investigate the safety and effectiveness of such therapies, compare them to current therapies, and attempt to propose novel compounds and mechanisms of action.

In the next chapter, we will discuss Computer-Aided Drug Design (a part of contemporary drug discovery) that uses an extensive range of theoretical and computational techniques to enhance drug development. We will also focus on the role of QSAR models as a means for designing better drugs from phytochemicals.

CHAPTER 4: *In silico* drug discovery

4.1 Computer Aided Drug Design (CADD)

The design and synthesis of new biologically active molecules with the help of computers is a scientific area that has attracted the interest of researchers in recent years, mainly due to the reduction in time and cost, compared to conventional drug design methods (Samudra and Sahinidis, 2013). The latter is based on the logic of 'trial and error', which involves synthesising and testing a large number of molecules for their biological activity. This process is both time consuming and cost-intensive (Verma et al., 2010).

Computer Aided Drug Design (CADD) is currently one of the most commonly used methods for identifying and developing potential lead compounds. It leverages the advancements in computer power and the development of software packages capable of accurately simulating experimental results (Doucet and Weber, 1996). This enables the design of novel pharmaceutical molecules (Zhang, 2011).

CADD includes:

1. designing of new molecules based on mining chemical data,
2. optimising these molecules to enhance their activity and minimise possible side effects (such as toxicity), and
3. designing novel molecules with increased binding affinity to specific receptors to serve as agonists or inhibitors (Bajorath, 2015)

Currently, the cost of bringing a new drug to the market exceeds 2.5 billion US dollars and requires approximately 10-15 years of extensive research (Mullin, 2014). Pharmaceutical companies, following the European Union directive (Directive 63/EU, 2010) to explore *in silico* research before resorting to animal testing, may invest in Quantitative Structure-Activity Relationships (QSAR) predictive models as an alternative to animal testing (Sullivan, Manuppello and Willett, 2014). They rely on these models to reduce the cost of developing new therapies and optimise existing therapeutics (Neeves et al., 2018).

4.2 A brief history of SAR and QSAR

The history of Structure-activity relationship (SAR) dates back to the mid-nineteenth century. Crum Brown and Fraser observed the correlation of 'physiological action' and 'chemical constitution' (Crum Brown and Fraser, 1868). Later, Richet correlated toxicity to aqueous solubility (Richet, 1893), and at the beginning of the 20th century, Overton (1896) and Meyer (1899) observed a linear relation between the anaesthetic potential of a series of organic compounds and their behaviour during separation. In the mid-1930s, Hammett defined the reaction constant r in order to describe the activity of the aromatic system R, as an expression of the constant ratio k (or the equilibrium constant K) and a parameter σ describing the electronic properties of the aromatic substituent X (Hammett, 1937).

The application of statistical methodologies can quantify the relationship between structural characteristics and biological activity (Nendza, 2012). QSAR correlates different biological properties with different physicochemical properties (Hansch analysis) or with the presence or absence of some structural features (Free-Wilson analysis). Hansch and Free-Wilson's work initiated a variety of QSAR applications in medical, biocidal and environmental sciences, amongst others (Hansch et al., 1995).

Quantitative Structure-Activity Relationships (QSARs)

QSARs are amongst the first efforts to develop a methodology for rational design with the primary objective of producing a more focused number of compounds that will be tested for their biological activity (Mavromoustakos et al., 2011). Minimising the randomness and the time-consuming and costly nature experiments for chemical synthesis and biological evaluation of new substances is the essential objective of this effort. Subsequently, QSAR methods result in the development of pharmaceuticals with increased safety to humans.

The ultimate goal of the QSAR technique is to find an equation (e.g. Equation (1)) between a particular property/activity, which is the dependent variable, and the structural characteristics of a series of similar structural molecules.

$$\text{Activity} = a_0 + (a_1 * P_1) + (a_2 * P_2) + \ldots + (a_n * P_n). \qquad \text{Eq (1)}$$

Where parameters P_1-Pn express specific properties of the selected molecules (e.g. lipophilicity) that may correlate with their biological actions. Those parameters are called molecular descriptors and their values are calculated either theoretically or experimentally for each molecule. Coefficients a_0- a_n are calculated with the use of multiple linear regression (MLR). These formulae are always accompanied by statistics such as the correlation coefficient r^2, standard deviation (sd) and Fisher test, on which the reliability of the equation can be assessed.

An essential element of reliable QSAR models is the diligent attention in choosing the data (Grisoni et al., 2018). Ideally the number of molecules should be as high as feasibly possible and the biological activity should be spread over a wide range of values (Hevener et al., 2008).

The values of biological activity (IC_{50}, EC_{50}, Ki, etc.) are converted to their corresponding negative logarithm (molar values can also be used), and therefore, the higher the resulting value, the more active the compound is (Papadatos et al., 2015). There is a plethora of parameters that can be used to create a QSAR model: physicochemical constants, molecular parameters, sub-structural features, topological indices or quantum-derived parameters such as dipole moment, or energy magnitudes such as, the energies of the Lowest Unoccupied Molecular Orbital (E_{LUMO}) and the Highest Occupied Molecular Orbital (E_{HOMO}). A summary of descriptive variables, in terms of their data content, their origin and how they are calculated, is included in Todeschini and Consonni (2008).

4.3 Molecular descriptors

Various statistical techniques are used to describe the mathematical relationship that exists between the activity under study and the physicochemical, quantum and structural descriptors of the compounds (Randic, 1991). The role of molecular descriptors is to encode chemical structures and capture relevant quantitative information (Roy et al., 2015). The calculation of molecular descriptors allows us to identify useful molecular features in our quest to design novel molecular structures with improved characteristics.

In essence, molecular descriptors are mathematical representations of a molecule. They are the result of a mathematical process which converts chemical information to a useful number, e.g in the case of a physico-chemical property like log P. In order for the molecular descriptor to be useful, it must possess the ability to aid in the interpretation of molecular properties and activity, enabling its application in predictive models.

The term molecular descriptor became very popular with the development of QSARs, although they have been around since the time of quantum chemistry and graph theory. The first two molecular descriptors were Platt's index and the Wiener index (Todeschini and Consonni, 2008).

According to their definition, molecular descriptors can be classified in two basic categories: those which are results of experimental measurements such as in lipophilicity, molecular refraction, dipole moment and polarisability and those which arise from the symbolic representation of the molecule and can be classified further depending on the type of the molecular representation.

By the term 'molecular representation' we mean the way by which the molecule is represented symbolically based on a specific procedure and rules. The quantity of chemical information transferred to the molecules' symbolical depiction is dependent on its type (Testa and Kier 1991).

It is a common place that the simplest form of molecular depiction is the molecular formula which gives an account of the type and number of atoms in a molecule. This representation of the molecule is independent of any kind of information related to the molecular structure

and molecular descriptors are termed 0D-descriptors. Examples of 0D- descriptors are: number of atoms, molecular weight and in general all descriptors concerning the molecules composition and its atomic properties. The most significant atomic properties are: atomic weight, atomic charge, van der Waals radius, atomic polarisability and hydrophobic atom constants.

The list with the functional subgroups comprising the molecule can be considered a one-dimensional representation. This type of representation does not require full knowledge of the molecular structure. These descriptions are termed 1D-descriptors and are used in sub-structural searching.

In two-dimensional molecular structure the way atoms are connected in the molecule is depicted, or else, the existence and nature of the chemical bonds. Molecular graphs are an approximation of its 2D-representation and are commonly referred to as the topological representation of the molecule. These molecular descriptors are termed 2D-descriptors. An alternative 2D-molecular representation of the molecule with vast utility in software is the SMILES depiction (Weininger, 1988; Weininger et al., 1989; Weininger, 1990).

With the 3D-molecular representation we take into account its geometrical representation in space and so we are not only considering the connectivity of the atoms but also its overall conformation in space. These descriptors are termed 3D-descriptors and two examples are geometric descriptors and stereochemical descriptors (Guha and Willighagen, 2012).

It is important to highlight that certain descriptors originate from the molecule's graph and are dependent on its properties, making them applicable as both 2D and 3D descriptors. Stereolelectronic, or lattice, representation of a molecule is a description of the molecule related to the molecular properties arising from the distribution and interactions of the molecule's electrons with its surroundings. These descriptors are termed 4D and consist of a lattice of scalar quantities (Todeschini and Consonni, 2009). Other representations of molecular structure include those for bulk representation, which describes the molecule as a natural object with three-dimensional properties such as: surface, volume etc. and stereodynamic representation which is dependent on time as time may change the structural characteristics of the 3D representation such as: flexibility, configuration etc. (Todeschini and Consonni, 2008).

Besides the classification of molecular descriptors based on the way of molecular representation, they can also be classified based on the following criteria:

1. The type of mathematical quantity used to express the descriptor e.g. if it is a scalar quantity, a vector or a two-dimensional table
2. The ability of descriptors to keep their value constant indepedently of specific characteristics of the molecular representation. The minimum independence which is definitely required for all descriptors used is the invariance to molecular numbering. Moreover this requirement, there are certain types of descriptors which are independent of the kind of atoms and molecular bonding, there are descriptors which are independent of spin and/or the molecular conformation. The latter are classified into four categories: No Conformational Dependence (NCD descriptors), Low conformational Dependence (LCD descriptors), Intermediate Conformational Descriptors (ICD descriptors) and High Conformational Descriptors (HCD descriptors) (Gasteiger and Engel, 2006).
3. The ability of a descriptor to avoid equal values for different molecules (degeneracy). There is a scale for degeneracy : no degeneracy (N), low degeneracy (L), intermediate degeneracy (I), and high degeneracy (H) (Todeschini and Consonni, 2009).

An estimation of the degree of degeneracy of molecular descriptors can be calculated by using Shannon's entropy and for QSAR studies is dependent on the molecules to be modelled (Godden et al., 2000). In general, degeneracy of descriptors value is an unwanted characteristic in QSAR studies (Todeschini and Consonni, 2009).

Based on the above criteria, Table 4.3 classifies some of the most common descriptors used in QSAR analyses according to their type.

Table.4.3: Classification of common types of computationally generated molecular descriptors used in QSAR analysis

Descriptor	Representation	Mathematical depiction	Conformational Dependence	Degeneracy level
Molecular weight	0D	Scalar quantity	NCD[1]	H[4]
Topological descriptors	2D	Scalar quantity	NCD	L/I [5]
Surface/Volume descriptors	3D	Scalar quantity	HCD/ICD [2]	L[6]
Interaction energy	4D	Lattice	HCD [3]	N[7]

[1]NCD: No Conformational Dependence,[2]HCD/ICD: High/Intermediate Conformational Dependence, [3]HCD: High Conformational Dependence, [4]H: high degeneracy, [5]L: low/intermediate degeneracy, [6]L: low degeneracy, [7]N: no degeneracy

According to Randic (1996) in order for molecular descriptors to be included in a QSAR analysis the following conditions should be met:

1. They are able to interpret structural changes
2. They are related to at least one property of the structure
3. They should preferably distinguish between isomers
4. They can be applied to local structures
5. They can be generalised to "higher" descriptors
6. They are independent of each other
7. Their calculation is simple and efficient
8. Their value should not be based on experimental properties
9. They are not trivially related to other descriptors
10. They use basic principles of molecular structure
11. They have the correct size dependence, in terms of molecule size
12. They gradually change following the molecules structural change

Of the two types of descriptors used in contemporary QSAR analyses, calculated descriptors have the advantage that they are less time-consuming to obtain. Thus they are cost-effective

in comparison with the experimentally derived ones and moreover they do not exhibit statistical issues, such as noise observed in experimental procedures. However, some assumptions and simplifications in calculated descriptors may lead to errors in their values.

With the help of mathematics, statistics, graph theory, computational chemistry, molecular modelling techniques and developed experimental techniques, a large number of theoretical and experimental descriptors has been defined and used in QSAR analyses. The quest for novel descriptors is still a crucial field of research in which new approaches and propositions on molecular structure coding are added all the time.

4.4 Procedure for the determination of a QSAR equation

The usual procedure followed in a QSAR analysis to determine the equation (shown as a workflow in Figure 4.4) is the following:

1. **Selection of the group of molecules to be tested**

The first step is to select the chemical structures that will be analysed and refined to derive the QSAR equation. An adequate number of analogous compounds with structural differences in the substituents should be selected.

2. **Inclusion of biological activity data**

For each of the molecules examined, information on the observed biological activity that is, of course, related to the particular molecule must be provided/ tested. The conditions for biological data used for drug design are:

a. The compounds should have the same mechanism of action and act on the same receptor.

b. The biological activity should be expressed in figures corresponding to molecular concentrations.

c. The level of biological experiments performed e.g. molecular or cellular organ or system, should be taken into consideration.

d. Biological experiments should be accompanied by information on their reliability and repeatability.

e. It is essential to ensure a well distributed range of biological activity values across the series of molecules.

Biological data are expressed in the form of a negative decimal logarithm so that the figures increase with increasing activity. The logarithmic expression of biological response (logBR) or log 1/C is often used where C is the molecular concentration of the substance that induces a particular biological response e.g. ED_{50}, IC_{50}, LD_{50} or log % response to a particular dose.

3. Calculation of physicochemical parameters

A wide range of electronic, thermodynamic, stereochemical, topological, physicochemical parameters are calculated in a QSAR analysis. It is possible to modify existing parameters and combine others, creating physicochemical combinations.

4. Data analysis and process

In this section, the distribution of physicochemical parameters can be plotted, the selection of new compounds to be introduced into the study, or the use of terms or symbols describing the physicochemical parameters and the processing as specified by the assay program.

5. Derivation of the QSAR equation

As long as the dependent and independent variables have been processed and identified, computation using different statistical methods lead to the extraction of the QSAR equation.

6. Evaluation of the QSAR equation

There are techniques through which it can be ascertained which data should be set within or outside the scope of the analysis (a 'jackknifing' process). One can also perform graphical analysis and cross validation to assess the reliability of the resulting QSAR.

7. Analysis of the equation

In this part of the study, the equation determines the most important physicochemical parameters, the parameters that will not be taken into account, as well as those points in the chemical molecule, which are responsible for each of the predicted biological activities. That is to say, in a nutshell, the nature of the structure-activity relationship is defined.

8. Activity prediction

As long as we have the QSAR relationship, we can now use it to predict the biological activity of various compounds within the applicability domain of this model, i.e '*the response and chemical structure space in which the model makes predictions with a given reliability'* (Netzeva, 2005). By designing a "candidate" chemical structure, the computation based on the QSAR equation gives the degree of predicted activity.

9. Future use of the QSAR equation

When this particular QSAR analysis is completed, the equation that has been derived for the compounds processed can then be used for analysing other chemical structures in the future.

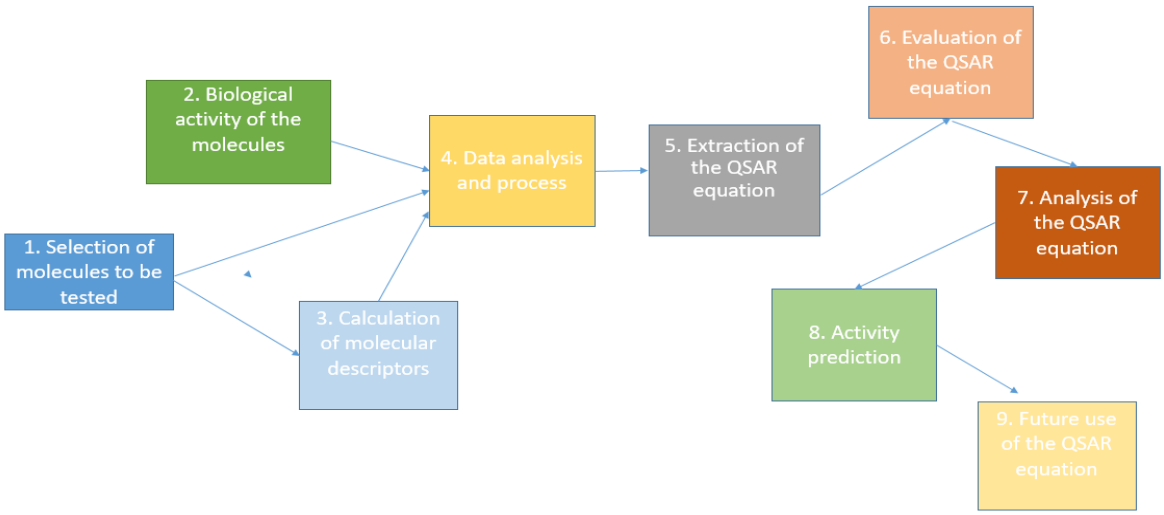


Figure 4.4 shows the usual procedure followed in a QSAR analysis to determine the equation

4.5 Modelling methods

Modelling methods are classified in three major categories:

1. **Linear methods**

The main methods in this category are: Multiple Linear Regression (MLR), Principal Component Analysis (PCA) and Partial Least-Squares analysis (PLS).

2. **Non-linear methods**

The main methods in this category are: neural networks, non-linear regression and Volterra series.

3. **Classification methods**

The main methods for classification are: pattern recognition, cluster analysis and Support Vector Machines (SVM).

The choice of modelling method is dependent on the form of the activity under investigation, the method of descriptors selection, the quality and quantity of available data. The most

commonly used QSAR method is Multiple Linear Regression. It has the advantage of being simpler, requires less time for calculations and gives a clear orientation for the role of every independent variable in relation to the dependent variable.

PCA and PLS analyses are very popular methods and are usually used to decrease the number of input descriptors and to develop correlations between the independent variables.

Principal Component Analysis (PCA) is a powerful multivariate statistical analysis technique widely employed in various domains, including pattern recognition processes (Massart, 1988). Its fundamental purpose is to transform a set of correlated variables into a new set of uncorrelated variables, known as principal components (PCs). This linear transformation enables the extraction of essential information from the original variables.

Each principal component is associated with a specific variance, representing the amount of information it captures from the original dataset. This variance provides a measure of the significance of each PC in describing the underlying patterns within the data.

One of the key advantages of PCA is its ability to reduce the dimensionality of the data. By selecting a subset of principal components that account for the largest proportion of variance, one can effectively reduce the number of variables required to describe the correlation with the dependent variable. This reduction in dimensionality not only simplifies the modeling process but also helps avoid issues related to overfitting and computational complexity.

Additionally, PCA serves as an unsupervised learning technique, enabling the identification of patterns in high-dimensional datasets. By analysing scatter plots of the first few principal components, researchers can gain insights into the inherent structure and relationships present in the data. This capability makes PCA a valuable tool for exploratory data analysis and visualisation, aiding in the discovery of meaningful patterns and trends.

The Partial Least Squares (PLS) method incorporates elements of dimensionality reduction and regression modelling to establish a predictive relationship between input and output variables. More specifically, the initially multidimensional spaces of input and output variables, are linearly projected to subspaces with a smaller dimension. The latent variables coming from the aforementioned subspaces contain all the relevant information of the initial data, excluding the noise and the autocorrelations that may exist.

Basic steps for the creation of reliable QSAR models

The three basic steps for developing reliable QSAR models are selection of the data, statistical modelling and evaluation.

QSAR models aiming for accurate predictions must fulfill the criteria listed below:

1. The models should provide an accurate description of the activity of all the compounds based on which they were deduced.
2. The predictions should be achieved with the lowest number of parameters possible in relation, always, with the degrees of freedom in the data.

The objective of every statistical model is not only to capture correlations but also to uncover meaningful mechanistic relationships. An effective model strives to describe the activity using the fewest possible descriptors. However, a model that incorporates a large number of variables raises concerns about the presence of spurious correlations (Nendza, 2012), which can undermine the reliability and interpretability of the model.

4.6 The role of QSAR in drug discovery from phytochemicals

As the reliability of in-silico models continues to improve, they can increasingly be employed for filling in gaps resulting from missing data. A large amount of phytochemical data regarding activity and toxicity are yet to be explored. In this perspective, QSAR methodology is a major tool that can fill the data gaps more rapidly and cost effectively than traditional methods (Kar and Roy, 2012). Hence, integrating the wealth of ethnopharmacological knowledge with QSAR, opens a new door which allows us to explore plant sources all over the world and possibly discover and design more effective and safe drugs.

4.7 R-language as a platform of *in silico* drug discovery

R-language is a robust programming language and statistical package that finds extensive application across various scientific disciplines for data analysis and statistical modeling purposes. R is a free and open-source software, which ensures its widespread availability. Additionally, the substantial community of R users has contributed to the development of a diverse range of packages, further augmenting its utility. It is noteworthy that many new methods are often initially implemented in R, adding to its appeal among researchers.

One of the key advantages of R is its simplified syntax, which facilitates an easier learning curve, especially for individuals without prior programming experience. In comparison to lower-level languages like C, Java, and Fortran, R's syntax simplicity offers a distinct advantage in terms of accessibility and ease of adoption.

Given its broad usage, open-source nature, extensive package ecosystem, and user-friendly syntax, R has become a preferred choice for data analysis and statistical modeling in numerous scientific disciplines.

R has gained popularity in the field of drug discovery in recent years thanks to its flexibility and capacity for handling immense and diverse amounts of data**.** Target identification, lead optimisation, and drug safety studies are just a few of the drug discovery applications that R has been used in recently. R has also been used in predicting drug response, to find novel targets for therapeutic intervention, and to explore the mechanisms of action of current medications. These facts, in combination with R’s high versatility in tasks such as data visualisation, machine learning, and predictive modelling influenced the decision to implement the analyses and techniques described in this thesis.

CHAPTER 5: Methodologies

Methodology

This chapter describes the various methodologies employed to achieve the aims of this thesis. As mentioned in section 4.7, the platform chosen to achieve this task was R (version 4.2.2). A total of twenty-two computational methods were developed using R-language packages (All code used in this thesis is available in Appendix 1. Appendices are available in the author's github account: https://github.com/giochrysox). The computational methods are referred to as "Tool Services". To develop these Tool Services R code was adopted from R packages, modified and/or created new to provide different services. Tool Services were organised into six categories depending on the type of data used. For each Tool Service, an example is given to demonstrate its use.

5.1 Development of Tool Services to enable the collection and analysis from various sources of information

Section 5.1 contains the methodologies used to create Tool Services 1-6. These Tool Services allow the user to retrieve valuable information from different sources. Tool Services 1 and 2 retrieve and text-mine PubMed articles. Tool Service 3 gives an insight into Genome-wide association studies data. Tool Service 4 reveals patents from the US Patents Office. Tool Service 5 helps us query clinical data from Clinical Trials.gov and Tool Service 6 allows us to use biochemical assay data from PubChem (https://pubchem.ncbi.nlm.nih.gov). An overview of the Tool Services and sources used in this section is given below in Table 5.1 and Figure 5.1, respectively.

Table 5.1: An overview of the Tool Services used in section 5.1

Tool Service	Title	Purpose	Methods	Output	R packages involved	Other comments
General search data						
1	PubMed search	Search PubMed articles for specific terms	User queries PubMed, adjusts the dates of and number of articles to retrieved	Two .csv files one containing the article's: PMID[1], title and abstracts and a second one containing the terms used by PubMed for the query	Code adapted from Amunategui (2015)	PubMed articles can be retrieved without leaving the environment of R
2	PubMed text-mining	Text-mining of PubMed abstracts	User inputs a text query and up to 10,000 PubMed aritcle abstracts are text-mined	Six .csv files containing words, genes, diseases, chemicals and organisms associated to the queried term respectively. A bar plot with the number of articles related to the query	Code combines packages: RISmed (Kovalchik, 2017), tm (Feinerer and Hornik, 2018) and pubmed.mineR (Rani and Ramachandran, 2015)	Gives an indication of chemicals, genes, diseases, and organisms associated the most to a query term in large volumes of PubMed articles
3	GWAS[2]	Query GWAS from 2008 until present	User can explore GWAS by a specfic disease (or trait) or gene	A .csv file containing Genome-wide association studies with all available data contained in the GWAS catalogue	Code was developed in this project	Can give an indication of genetic markers for a disease
4	USPTO[3] patents	Find relevant USPTO patents	Patents can be searched for specific terms	A .txt file containing all patents related to the query in cluding their ID and title	Code adapted from: patentsview (Baker, 2019)	Patents databases contain a wealth of information not available in other scientific databases

5	Clinical trials data	Find relevant Clinical trials data	Clinical trial data from ClinicalTrials.gov are searched for relevant data	A .csv file containing clinical trial information. This data is sorted by study, trial phase, gender and age of the participants in these trials	Code was developed in this project	Clinical trials are important as these are studies conducted on humans and evenwhen researchers do not obtain the outcomes they predicted, trial results can help point scientists in the correct direction.
6	PubChem assays	Search PubChem Bio assays	Bioactivity data from PubChem are searched to identify relevant compounds and targets	A .csv file containing different bioassays with the bioactivity of compounds at a specific target	Code adapted from rpubchem (Guha, 2018)	Bioassays contain valuable information which can be used for further drug design and pharmacodynamic studies

[1]PMID: PubMed ID, [2]GWAS: Genome Wide Association studies, [3]USPTO: United States Patent and Trademark office

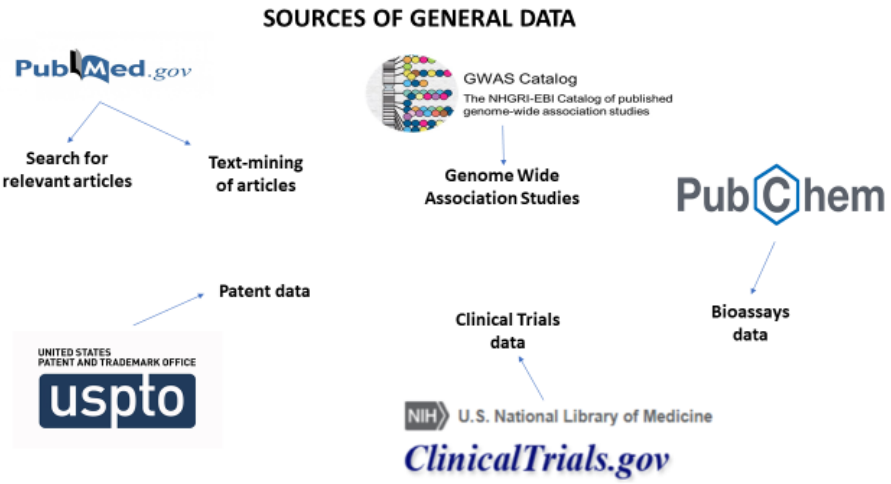


Figure. 5.1: An overview of the sources used in the Tool Services of Section 5.1

Tool Service 1 (PubMed search)

The aim of this Tool Service is to enhance PubMed queries at all stages of this project without leaving the environment of R (Amunategui, 2015). The user can query PubMed, adjust the dates and number of articles to be retrieved. One .csv file is exported containing three columns: the article's PMID, title, and abstract. The other file contains the MeSH Terms used for the query.

Example: If our query terms are: 'androgenetic alopecia' and the search is for all articles since 1950 and we want to limit our search to only the 1,000 most relevant of these, then the service returns two .csv files, one containing three columns with the PMID, title and full abstracts of the 1,000 articles yielded and the other containing the terms used by PubMed for the query.

Tool Service 2 (PubMed text-mining)

In the last years we have seen a remarkable increase in collected data stored in large biomedical databases (Hunter and Cohen, 2006). This dramatic growth has led to inabilty in terms of analysis and interpretation of this data. It is imperative to develop and use a new generation of methods and tools to help researchers mine invaluable nuggets of knowledge out of these data.

Text-mining is a thorough process where the user interacts with a collection of texts by using a set of analytical tools. Text-mining is a process that aims to extract information from data sources through identification and investigation of interesting patterns (Krallinger and Valencia, 2005).

PubMed is, perhaps, the most comprehensive online collection of biomedical research published both in English and other languages. PubMed's collection of thirty million articles is dynamic, with an increasing rate of about 40,000 new abstracts every month. Due to Its vast volume, manual attempts to retrieve information, especially to understand the interrelationships between abstracts, become almost impossible. Automated text-mining's processes help researchers use current data volumes more efficiently and reveal correlations in biomedical data that can be used in diagnosis and therapy (Guin et al., 2019).

The aim of this Tool Service is to quickly text-mine thousands of PubMed to retrieve articles related to our query. Packages RISmed (Kovalchik, 2017), tm (Feinerer and Hornik, 2018), and pubmed.mineR (Rani and Ramachandran, 2015) were employed in this service. The user inputs a text query and up to 10,000 PubMed articles are text-mined to retrieve in .csv files words, genes, diseases, chemicals and organisms associated to the term queried. The user is

also able to retrieve a bar plot with the number of articles related to the query in the range of years that they were published.

Example: If we want to text-mine articles with "*alopecia*" in their title, we initially need to download from PubMed a .txt file containing abstract texts of those articles. This is read in R and yields six .csv files:

1. The first contains 48,941 words and their frequency of occurrence in these article abstracts.
2. The second contains 47 genes mentioned in these articles along with their respective NCBI id.
3. The third contains the 10,000 article PMIDs which were used for the text-mining.
4. The fourth contains 510 diseases also mentioned in these abstracts.
5. The fifth yields 104 chemicals (with their MESH terms).
6. The sixth contains information on the 76 organisms mentioned.

A bar plot of articles containing the term alopecia shows a constant upward trend in the last 70 years.

Tool Service 3 (GWAS)

Genome-wide association studies (GWAS) are studies that use high-throughput genotyping technology, making it possible to genotype hundreds of thousands to millions of single nucleotide polymorphisms (SNPs) (Edenberg and Liu, 2009). This has revealed a vast number of genetic links to diseases. There were two major factors for the evolution of GWAS. The first was the HapMap project (HapMap consortium, 2005), a worldwide initiative to map linkage disequilibrium (LD) patterns in human populations (Reich et al., 2001). The second was the rapid reduction in genotyping costs during the last decade (Hirschhorn, 2005).

The US National Human Genome Research Institute catalogues Genome Wide Association Studies with at least 100,000 genotyped polymorphisms. Since 2007 hundreds of GWAS have been published for over 80 diseases (Hindorff et al., 2012).

Genome-wide association studies (GWAS) assist researchers in identifying genes linked to a specific disease or other trait. This technique examines a large population's entire genome, looking for tiny variations known as single nucleotide polymorphisms. GWAS catalog was downloaded from the European Bioinformatics Institute (https://www.ebi.ac.uk/gwas/downloads). It contains data on Genome Wide Association Studies from year 2008 up to the present day. The file was subset in R to allow queries of genome wide association studies by a specific disease or gene.

Example: In the case of querying the term **'**male-pattern baldness' we yield a.csv file containing 9 GWAS studies with all available data. If we want to query the SRD5A2 gene we can see that this gene is reported in 3 studies.

Tool Service 4 (USPTO Patents)

A patent is the recognition of the legal right of intellectual property of an invention, which secures its depositor a temporary monopoly of this invention. The number of patents in the pharmaceutical industry has increased significantly in the last four decades, and so has the amount of data in patent databases. The United States Patent and Trademark Office (USPTO) database is freely available and contains more than seven million patents.

The aim of this Tool Service is to make United States Patent and Trademark Office (USPTO) patent data available for exploration of 'hidden' knowledge.

United States Patent and Trademark Office (USPTO) patents are searched with package patentsview (Baker, 2019). The user is able to retrieve a .txt file all patents that match the query.

Example: If we query USPTO with the term 'Androgenetic alopecia' we get 9 patents (with their IDs and titles).

Tool Service 5 (Clinical Trials data)

Clinical trials provide high-quality data on specific patient-populations which can be exploited to provide insights into health conditions and treatment effectiveness.

This Tool Service enables users to query and retrieve valuable clinical trials data from www.ClinicalTrials.gov. The resulting .csv files contain information about all trials data for the specific search term plus the same data sorted by study, trial phase, gender and age of the participants in these trials.

Example: If we query ClinicalTrials.gov with the term 'androgenetic alopecia', we get 114 trials. The majority of trials were interventional having a drug as the intervention with participants mainly adults from both genders. The majority of the trials were Phase 2 trials.

Tool Service 6 (PubChem Assays)

The large collection of bioactivity data and molecular target information in PubChem has facilitated research areas such as SAR studies. We employed this Tool Service to establish the connections among chemicals, targets and signalling pathways as it is very helpful in drug design.

This service is dedicated to retrieving PubChem Assays data. It utilises package rpuchem (Guha, 2018) and allows the user to access assay data such as activity of a compound in a specific target.

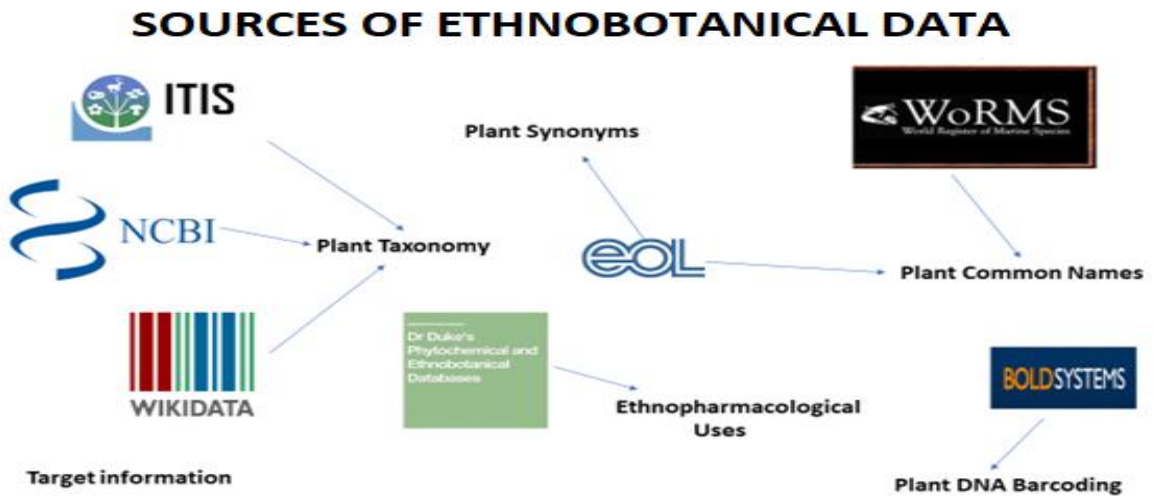
Example: For the query term "*androgenetic alopecia*" treatment, we obtain 40 PubChem assays with detailed data on compounds and their PubChem substance and compound IDs (SID, CID), activity outcomes and corresponding values.

5.2 Development of Tool Services to enable the collection and analysis of ethnobotanical information

This section contains the methodologies used to create Tool Services 7 and 8. These Tool Services allow the user to retrieve and analyse valuable ethnopharmacological information from different sources. Tool service 7 attempts to address the problems of accuracy in plant nomenclature and taxonomy, which are vital with respect to medicinal plant identification. It allows the correction of plant names to the accepted ones, provides plant synonyms, common names and taxonomy from different databases. Tool Service 8 allows the study of evolutionary similarities (phylogenetic comparison) of plants. An overview of the Tool Services and sources used in this section is given below, respectively, in Table 5.2 and Figure 5.2.

Table 5.2: An overview of the Tool Services used in section 5.2

Tool Service	Title	Purpose	Methods	Output	R packages involved	Other comments
Ethnobotanical data						
7	Plant information	Search several ethnobotanical parameters and address the problem of plant nomenclature	User can retrieve plant and phytochemical data, correct plant names to the accepted ones, retrieve plant taxonomy, common names and synonyms	In .csv files: plant and phytochemical data, corrected plant names, plant taxonomy, plant synonyms and plant common names	Code combines packages: Taxonstand (Cayuela, Stein and Oksanen, 2017), rentrez (Winter, 2017), Reol (Banbury and O'Meara, 2014), bold (Chamberlain, 2017)	This Tool Service attempts to address the problems of taxonomy and plant identification
8	Plant phylogeny	Assign a DNA profile for plant identification and analyse evolutionary similarities	User can retrieve FASTA files for identification purposes, allign multiple plant sequences of interest and produce phylogeny diagrams	Multiple .csv files containing plants DNA FASTA, a view of the multiple sequence allignment in the browser and a .png file containing the phylogeny diagram	Code combines packages: bold (Chamberlain, 2017), DECIPHER (Wright, 2018), Biostrings (Pages et al., 2018)	This Tool Service can also be used to compare evolutionary similarities between targets. In this case, FASTA file may be retrieved from NCBI



Source	URL	Brief description
EOL	https://eol.org/	Encyclopedia of Life (EOL) is an online database providing information about all species on Earth.
ITIS	https://www.itis.gov/	Integrated Taxonomic Information System (ITIS) is a comprehensive taxonomic database of species.
WORMS	http://www.marinespecies.org/	World Register of Marine Species (WORMS) is a global taxonomic database for marine organisms.
NCBI	https://www.ncbi.nlm.nih.gov/	The NCBI taxonomy database provides a hierarchical structure that represents the evolutionary relationships and taxonomic classifications of a wide range of organisms, including bacteria, archaea, fungi, plants, and animals.
WIKIDATA	https://www.wikidata.org/	Wikidata serves as a valuable resource for storing and organizing taxonomic information about species, genera, families, and other taxonomic ranks.
DR. DUKES	https://phytochem.nal.usda.gov/	Dr. Duke's Phytochemical and Ethnobotanical Databases provide information on plants, their chemical constituents, and ethnobotanical uses.
BOLD SYSYTEMS	https://v3.boldsystems.org	Barcode of Life Data Systems (BOLD) is an online platform for DNA barcode data of various species.

Figure. 5.2: An overview and brief description of the sources used in Tool Services of Section 5.2

Tool Service 7 (Plant information)

The aim of this Tool Service is to provide a variety of ethnobotanical data on the plants used in this thesis. This service uses packages Taxonstand (Cayuela, Stein and Oksanen, 2017), rentrez (Winter, 2017), Reol (Banbury and O'Meara, 2014), taxize (Chamberlain et al., 2017), bold (Chamberlain, 2017a) and ritis (Chamberlain, 2017b). Using the plants and phytochemicals database (Appendix 1) mainly created using Tool Service 1 (Pubmed search)

and containing data on phytochemicals contained in 69 plants used ethnopharmacologically in AGA (references of the use of each plant in AGA are also available in Appendix 1). Phytochemicals are coded as their PubChem CID, plants as their scientific name and references as their PMID or other reference type e.g. DOI or URL). The user is able to retrieve:

- plant and phytochemical data
- a corrected list of plant names to the accepted name even if the list contains synonyms, unresolved names or the names are spelled incorrectly
- plant common names from 4 databases (EOL, ITIS, NCBI, and WORMS)
- plant synonyms
- taxonomy (from ITIS, NCBI, WIKIDATA)

Example: Using the database of plants and their phytochemicals (Appendix 1) we find that *Acacia concinna* contains at least 11 phytochemicals. *Acacia concinna* is the accepted name for this plant according to The Plant List. We may also find common names, synonyms and the taxonomy of this plant

Tool Service 8 (Plant phylogeny)

The aim of this Tool Service is to provide an analysis of evolutionary similarities between the plants. It can also be used in the same respect to compare targets. It uses packages bold, DECIPHER (Wright, 2016) and Biostrings (Pagès et al., 2018) to retrieve:

-Plants FASTA files (for identification and consistency of research purposes)

-Align multiple sequences of interest

-Plant or target DNA similarities in terms of a phylogeny dendrogram

Example: Initially we run the part of the code which retrieves the plants FASTA files. These are then combined into a .fas file. This file is used in the second part of the code to align the DNA sequences of the plants and then get the multiple sequence alignment (MSA) in our

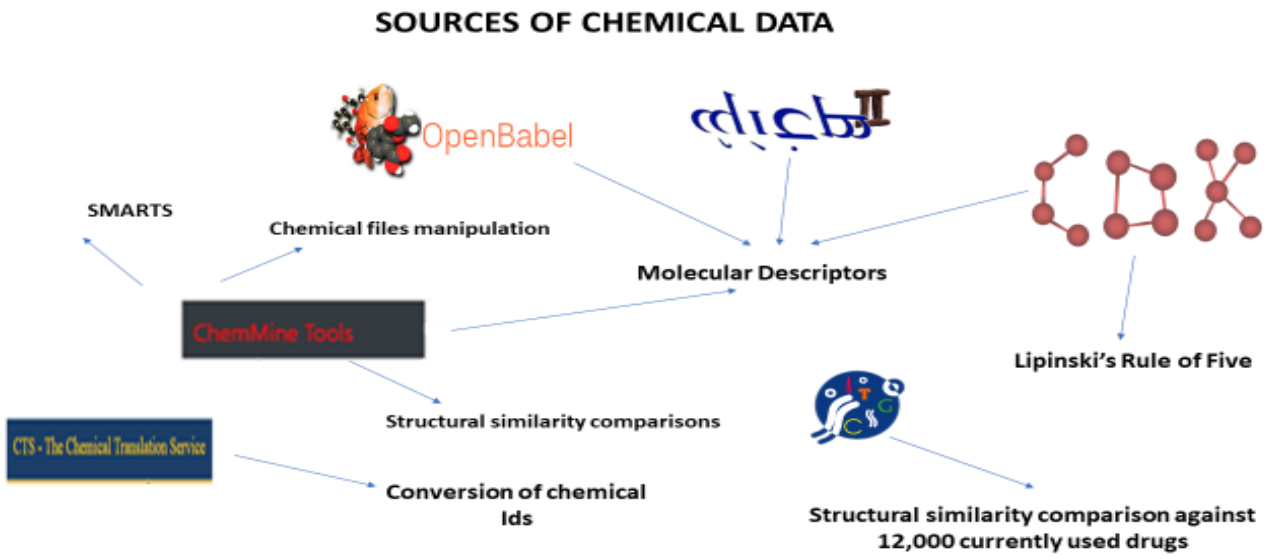
browser. Finally, a phylogenetic tree showing the DNA similarities of the plants is yielded (this can be stored in either .png or .pdf format).

5.3 Development of a Tool Service to enable the collection and analysis of chemical information

Section 5.3 contains the methodologies used to create Tool Service 9. This Tool Service was created to handle most of the computational chemistry tasks involved in this thesis. These range from simple manipulation of different chemical file formats to calculation of molecular descriptors, to structural similarity and drug safety assessments. An overview of the Tool Services and sources used in this section is given below in Table 5.3 and Figure 5.3 respectively.

Table 5.3: An overview of the Tool Services used in section 5.3

Tool Service	Title	Purpose	Methods	Output	R packages involved	Other comments
Chemical data						
9	Chemical data	To manipulate, retrieve and analyse a variety of computational chemistry tasks	Manipulation of different chemical files, translation of different chemical identifiers, structural similarity studies, compare compounds against SMARTS datasets, retrieve molecular descriptors, perform hierarchical clustering	All data is provided in .csv files. In the case of structural comparisons with ChemmineR a .png file of the molecules is produced. In the case of hierarchical clustering a .png file of the cluster dendrogram is produced	Code combines packages: rcdk (Guha et al., 2017), ChemmineR (Cao et al., 2008), fmcsR (Wang et al., 2013)	This Tool Service handles a great number of computational chemistry tasks involved in drug design and safety



Source	URL	Brief description
JoeLib	https://sourceforge.net/projects/joelib/	JoeLib is an open-source cheminformatics library written in Java. It provides a wide range of functionalities for molecular modeling, substructure searching, and chemical file format support
CDK	https://cdk.github.io/	The Chemistry Development (CDK) is an open-source cheminformatics library implemented in Java. It offers a set of algorithms and tools for working with chemical structures, molecular descriptors, 2D/3D rendering, and chemical data analysis.
OpenBabel	http://openbabel.org/	OpenBabel is an open-source chemical toolbox designed to speak the many languages of chemical data. It provides a wide range of utilities for file format interconversion, molecule manipulation, and property calculations.
CTS	https://cts.fiehnlab.ucdavis.edu/	The Chemical Translation Service (CTS) is a web-based platform that enables conversion and translation of chemical identifiers, names, and structures between various formats and databases.
SIMCOMP	https://www.genome.jp/tools/simcomp/	SIMCOMP is a chemical similarity search tool developed by the Kyoto Encyclopedia of Genes and Genomes (KEGG). It allows users to compare the structural similarity of small molecules based on their chemical substructures.
Chemmine Tools	http://chemmine.ucr.edu/	ChemmineTools is a cheminformatics and bioinformatics tool. It provides a set of functions and algorithms for molecular database mining, chemical similarity searching, and other cheminformatics tasks.

Figure. 5.3: An overview and a brief description of the sources used in Tool Service 9

Tool Service 9 (Chemical data)

This service takes advantage of packages rcdk (Guha et al., 2017), ChemmineR (Cao et al., 2008), fmcsR (Wang et al., 2013) and provides the user with the ability to:

- manipulate (download, read, load, write and convert) different chemical file formats
- retrieve physicochemical properties such as: Molecular Formula, Molecular Weight, Canonical SMILES, Isomeric SMILES, InChI, InChIKey, IUPAC Name, XLogP, Exact Mass, Monoisotopic Mass, TPSA etc. of the compounds
- evaluate the drug-likeness of compounds by using Lipinski's rule of five (Ro5)
- compare a compound's structure similarity against any batch of compounds
- translate between most chemical identifiers
- search for structural similarities between any compound and 12,000 currently used drugs, using KEGG's Simcomp service (Hattori et al., 2010)
- compare compounds to a wide variety of SMARTS datasets regarding lead-likeness, promiscuity and other toxicological alerts (e.g skin sensitisation)
- retrieve 2D and 3D molecular descriptors from CDK, Openbabel and JoeLib

Example: From the database of plants and phytochemicals (Appendix 1) we can get the PubChem CIDs of the phytochemicals contained in plants used ethnopharmacologically for alopecia. Using PubChem's download service, we can have a multi-SDF file containing all phytochemicals. The SDF (Structure-Data File) format is a standard file format in cheminformatics that represents chemical structures and associated data, allowing for the storage and exchange of molecular information such as atom coordinates, bond types, and properties in a single file. The new file is checked for inter-molecular similarities and these are recorded. We also investigate them with KEGG's SIMCOMP, a method for structural comparison that analyses the chemical structures of compounds and compares them to KEGG's database of more than 12,000 drugs. With the help of SMARTS datasets compiled by the University of Hamburg's Bioinformatics Centre (Ehrlich and Rarey, 2012), all phytochemicals are studied for the possibility of containing fragments which may be of interest in terms of lead-likeness, skin sensitisation, mutagenicity, carcinogenicity, mitochondrial toxicity etc. Finally, 2D and 3D molecular descriptors for all phytochemicals are retrieved from CDK with the help of package rcdk.

5.4 Development of a Tool Service to enable the collection and analysis of pharmacological data

Early studies of pharmacological data may prevent drug failures at later stages. Serious side effects and ADRs can result in pre- and post-marketing withdrawals respectively. *In silico* studies of pharmacological data may enhance predictions of such phenomena and reveal crucial off-targets early in drug development. Translating the off-target effects of drugs may also be used as a strategy to discover novel indications for already-marketed drugs (drug repurposing). An overview of the Tool Service and definetely the sources used in this section is given below in Table 5.4 and Figure 5.4 respectively.

Table 5.4: An overview of the Tool Services used in section 5.4

Tool Service	Title	Purpose	Methods	Output	R packages involved	Other comments
Pharmacological data						
10	Pharmacological data	To retrieve and analyse various pharmacological parameters	User can explore a series of usefeul parameters such as: indications side effects, ADRs and DDIs	All data is provided in .csv files	Code was developed in this project	A database of ADR cases was compiled from MHRA's Yellow Card Scheme to improve query results

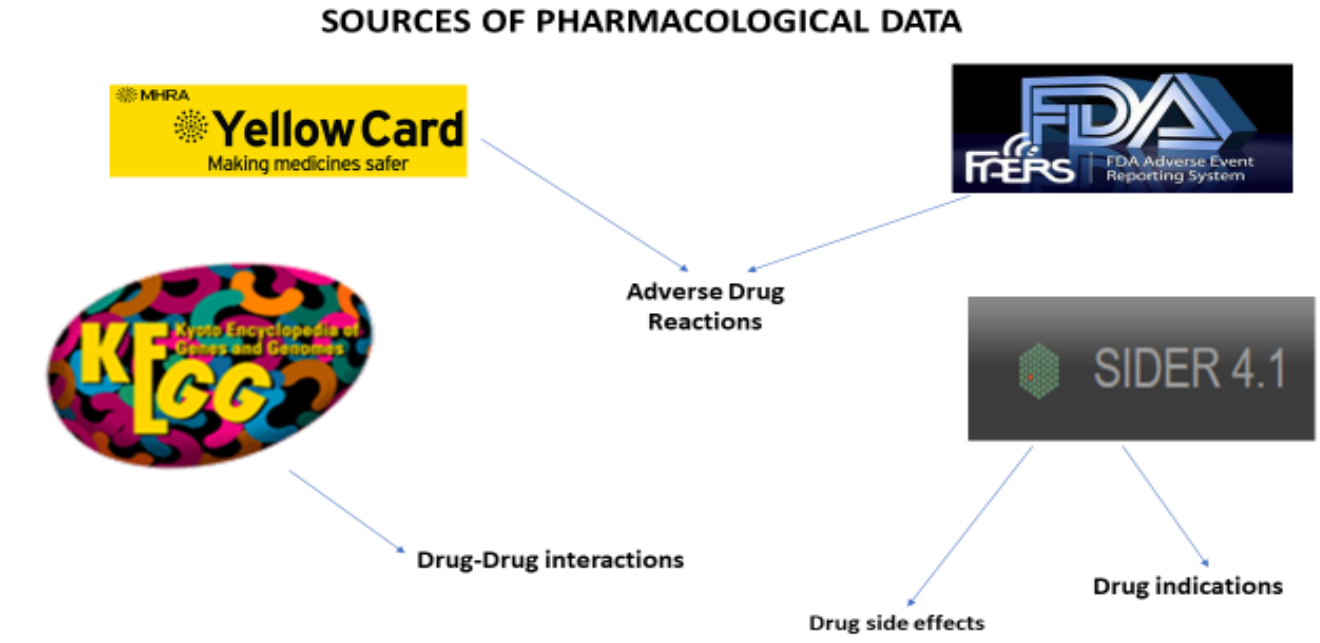


Figure. 5.4: An overview of the sources used in Tool Service 10

Tool Service 10 (Pharmacological data)

This service requires either the drug's PubChem or KEGG ID and retrieves a list of the drug-drug interactions along with their classification according to KEGG's DDI (Kanehisa et al., 2010) database. It also enables the user to retrieve adverse drug reactions by either the US FDA or UK MHRA's Yellow Card scheme (for this purpose a database of more than two million ADR cases was compiled). The user is also able to retrieve data on drug indications, indicated pharmacotherapy, side effects and their frequency of occurrence.

Example: Finasteride and minoxidil are two drugs commonly used in the treatment of alopecia. By using either their PubChem CID or KEGG ID we may retrieve a list of drugs with which concomitant use causes drug-drug interactions along with KEGG's DDI classification.

MHRA's Yellow Card Scheme (MHRA, 2016) contains data on more than 2 million adverse drug reaction (ADR) cases caused by approximately 2,000 drugs. 700 drugs and their respective 7,400 cases of causing alopecia as an ADR were compiled into a .csv file (found in Appendix 10)

Similarly, the user can retrieve ADR data from FDA (FDA Adverse Event Reporting System, 2016). In the case of Propecia® (Finasteride) 310 ADRs were reported. Using SIDER database (Sider DB, 2016), we can also reveal drugs indicated for alopecia, data on alopecia as a drug side effect and the respective side effect frequency of drugs causing it.

5.5 Development of Tool Services to enable the collection and analysis of Targets, Diseases and Pathways data

One of the most crucial steps in the drug discovery process is to identify a potential biological target and understand how it contributes to the disease. A target is a biological component (such as a gene, protein, or RNA) that a medication can attach to in order to cause a physiological change. Targets must have a binding pocket or active site where a prospective medication can bind to. A good target should be safe, effective, and druggable.

Target-based drug research relies heavily on knowledge of the molecular mechanisms underlying the disease. Revealing the causal relationship between a target or a biochemical

pathway and a disease may enhance our chances for better therapeutic results. An overview of the Tool Services and sources used in this section is given below in Table 5.5 and Figure 5.5, respectively.

Table 5.5: An overview of the Tool Services used in section 5.5

Tool Service	Title	Purpose	Methods	Output	R packages involved	Other comments
Targets, Diseases and Pathways data						
11	Target information	To retrieve important information regarding targets	By inputting target's KEGG ID we retrieve data such as the target name, organism, pathways, structure (PDB code) and the DNA and protein sequences of a target in FASTA file.	A. csv file containing all target info	Code adapted from: KEGGREST (Tenenbaum, 2019)	Entry data can be translated with Tool Service 12, PDB code is used in Tool Service 17 to retrieve the target's binding site and FASTA data can be used with Tool Service 8 to reveal evolutionary similarities between targets
12	Target ID Translation	To translate target IDs	User inputs UniProt ID of the target and selects the databases	A .csv file containing the relevant IDs from all databases selected	Code adapted from: Uniprot.ws (Carlson, 2016)	User can choose to translate target IDs from more than one hundred databases
13	BindingDB affinity data	To obtain binding affinity data from Binding DB	Using the UniProt ID of a target of interest and setting an affinity cut-off value we may retrieve a set of ligands for this target	A. csv file containing the target, ligand affinity type (e.g Ki, Kd, IC_{50}) and their respective affinity value	Code was developed in this project	Binding affinity is given in Ki, Kd or IC_{50}
14	EGGNOG's Gene Ontology (GO) terms	To retrieve Gene Ontology Terms for a specific target	Using the EGGNOG ID (can be translated from another target ID with the use of Tool Service 12) of a target of interest we can get all GO Terms associated with this target	A .csv file which contains the GO ID and GO Term associated with the target	Code was developed in this project	Given a list of genes, gene ontology analysis can provide useful information on the biological processes, cellular components and molecular functions that are implicated in a disease.

15	DisGenet-Cytoscape disease data	To study Gene-Disease associations using DisGeNET. To retrieve and visualise Gene-Disease using Cytoscape, and Gene-Pathway networks using Cytoscape and WikiPathways	For Gene-Disease associations using DisGeNET the user can either input a single ID (OMIM ID) or multiple disease IDs (NCBI Concept IDs) or input a list of genes. For Gene-Disease networks we need to input a list of genes. For Gene-Pathway networks by entering a keyword.	.csv files containing Gene-Disease associations from DisGeNET. A set of .jpg files showing the Gene-Disease and Gene-Pathway networks from Cytoscape	Code combines packages: disgenet2r (Piñero et al., 2020), Rcy3 (Gustavsen et al., 2019) and rWikiPathways (Slenter et al., 2018)	Our network was compared against other STRING database disease networks obtained with Cytoscape for: androgenetic alopecia, hypotrichosis, atrichia with papular lesions, familial wooly hair syndrome, prostate cancer, Type 2 Diabetes mellitus and four pathways related to hair follicle growth and development.
16	CYP450, Phase-II enzymes and Transporters Data	To study the metabolic fate of drugs and if possible make an early assessment of their metabolic similarity to phytochemicals	The user can search by either the drug, (enzyme or transporter) and/or metabolic relation (e.g inhibition)	A. csv file containing the drug, enzyme or transporter and the metabolic relation	Code was developed in this project	A fully-referenced database of 891 compounds and their relation to 23 enzymes and transporters was compiled for this Tool Service.
17	Target binding sites	To elucidate the target's binding site	Using the PDB codes retrieved with Tool Service 11 the user is able to retrieve the relevant amino acids comprising the binding site along with their position in the protein chain	Data are provided in a single .csv file	Code adapted from: Bio3d (Grant et al., 2006)	To test the appropriateness of our retrieved data, amino acids consisting the binding sites of different targets were visualised in software PyMOL
18	KEGG Pathways	To retrieve KEGG pathways of interest and highlight all the associated targets we are interested in	The user inputs the KEGG IDs of the pathways and targets	An image in .png file of the pathway of interest. The same image opens in the browser having targets marked with	Code adapted from: KEGGREST (Tenenbaum, 2019)	

				a different colour		
19	Pathway Network Visualisation	To perform pathway analysis and visualisation	The user manually inputs a .txt file containing the genes names their ENSEMBL IDs and NCBI entrez IDs	The resulting animated diagram visualises the relation of genes and the associated pathways .	Code adapted from: PANEV (Palombo et al., 2020)	We need to create a .txt file containing three columns. The first column contains the ENSEMBL IDs of the targets, the second their corresponding NCBI entrez IDs and the last column the actual target names (note that target names can be translated to ENSEMBL and NCBI IDs using Tool Service 12)
20	Comparative Toxicogenomic data	To retrieve associations between chemicals, targets, pathways and diseases	The user inputs the name of the chemical, gene, disease or pathway and retrieves the associations	Data are provided in .csv files	Code was developed in this project	

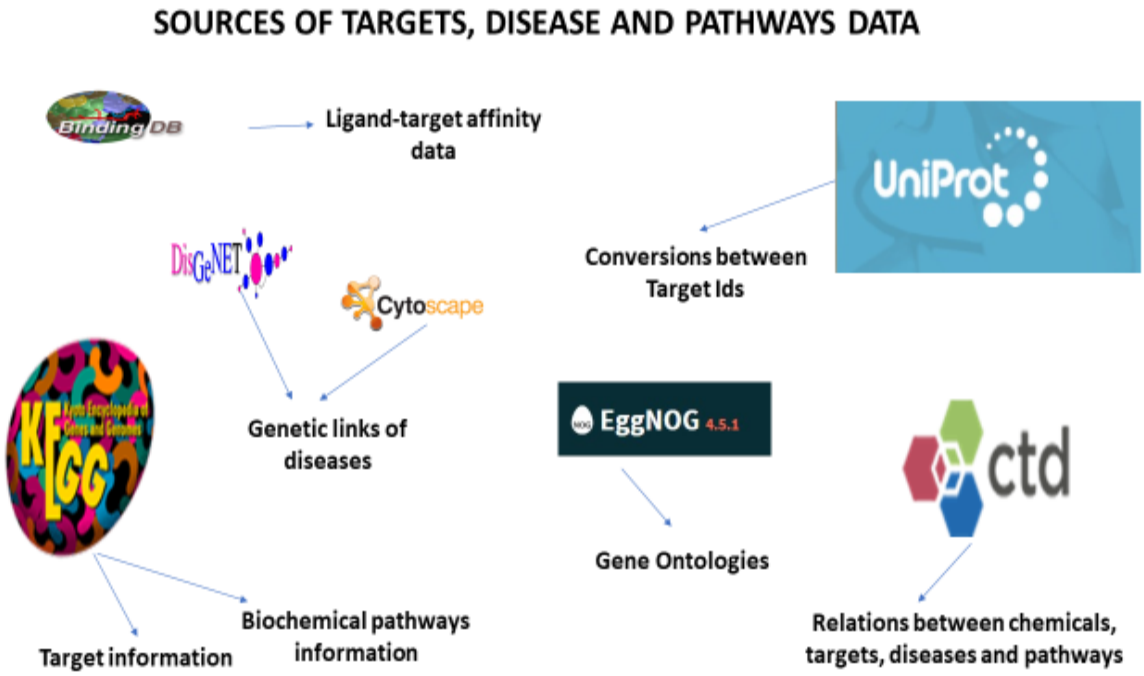


Figure. 5.5: An overview of the sources used in Tool Services in Section 5.5

Tool Service 11 (Target information)

The aim of this Tool Service is to obtain important information regarding studied targets. It uses the KEGGREST package (Tenenbaum, 2019) to provide data such as entry (KEGG ID), target name, organism, pathway, structure (PDB code) and the DNA and protein sequences of a target in FASTA file. All data are recorded in a .csv file.

Example: The androgen receptor (AR) is probably the most important target in androgenetic alopecia. By inputting the target's KEGG ID we may reveal important information from KEGG's REST regarding this target's name, organism, pathways and structure. We may also retrieve in .FASTA files its protein and DNA sequences which can be used to elucidate interspecies target similarities, or similarities between targets implicated in the same pathway or disease.

Tool Service 12 (Target ID Translation)

The aim of this Tool Service is to translate between different target ids. It uses code from package Uniprot.ws (Carlson, 2016) to achieve target ID translations. Can translate IDs from UniProt to one hundred IDs from other databases.

Example: We can easily get all IDs from more than one hundred databases by simply entering a list of their corresponding UniProt ID. In this case we translated the set of the 48 genes chosen to study AGA-related targets by using their IDs.

Tool Service 13 (BindingDB affinity data)

This Tool Service aims to take advantage of the BindingDB's API to get information regarding the binding affinity of various ligands for a target. The user is able to obtain these data by entering the UniProt ID of the actual target.

Example: SHBG is a target implicated in androgenetic alopecia. The UniProt ID for this compound is P04278 and if we set an affinity cutoff = 100 we obtain a list of 58 ligands, along with their affinity type (e.g., Ki, Kd, IC_{50}) and respective affinity value for this target.

Tool Service 14 (EGGNOG's GO terms)

Gene Ontology (GO) terms, are standardized annotations that represent the functional attributes and characteristics of genes and gene products. The Gene Ontology Consortium developed a controlled vocabulary of terms that categorize genes based on their biological processes, molecular functions, and cellular components. GO terms provide a systematic and structured way to describe the functions and attributes of genes, allowing researchers to annotate and classify genes based on their known or predicted roles. Each GO term is represented by a unique identifier and is associated with a specific biological concept.

Enrichment analysis using these terms can help relate genes with pathways by identifying shared biological processes and functions between the genes of interest and the annotated pathways.

This Tool Service takes advantage of the Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (EGGNOG) API to retrieve Gene Ontology (GO) terms and covers the three domains of cellular components, molecular function and biological process.

Example: SRD5A2 is a target implicated in androgenetic alopecia. With this service the user is able to immediately retrieve GO terms for this target.

Tool Service 15 (DisGenet-Cytoscape disease data)

This service takes advantage of packages disgenet2r (Piñero et al., 2020), RCy3 (Gustavsen et al., 2019) and rWikiPathways (Slenter et al., 2018) to provide us with:

- Gene-Disease associations obtained from DisGeNET. We may retrieve and store information from DisGeNET by using either a single or multiple diseases. In case of multiple diseases, we may also plot the resulting data to visualise common genes. We may also retrieve and store disease data for a set of genes.

- Gene-Disease networks retrieval and visualisation in Cytoscape

- Pathway-Genes networks retrieval and visualisation in Cytoscape

Example: In this case we used the chosen 48 genes to retrieve and store gene-disease associations from DisGeNET. The same genes were used to create a network in Cytoscape.

Tool Service 16 (CYP450, Phase-II enzymes and Transporters Data)

The user is able to retrieve data regarding enzymes and transporters which are mostly implicated in drug metabolism. The user can retrieve a list of enzymes and transporters involved in drug metabolism either by choosing the PubChem CID of a compound or by selecting an enzyme or transporter and retrieving a list of implicated compounds.

Example: The user can retrieve compounds that interact with CYP3A4 or P-glycoprotein (Pgp), or search by the compound's PubChem CID to check its interaction with the main enzymes and transporters implicated in drug metabolism.

Tool Service 17 (Target binding sites)

This service uses the Bio3d package (Grant et al., 2006), which reveals the possible binding sites of a target. Additionally, it can be used to reveal allosteric binding sites that potentially regulate the activity of the receptor.

Example: To retrieve the binding sites of AR, we input the PDB code of the target, which is already retrieved with Tool Service 11. The resulting .csv file contains the amino acids and their positions in the protein.

Tool Service 18 (KEGG Pathways)

This service reveals KEGG pathways of interest and highlights all the associated targets we are studying in each case. This is a very important step in our effort to distinguish those targets and pathways implicated in disease process. By utilising R-package KEGGREST we can retrieve:

- a .png image file of the pathway of interest
- an image in our browser of the pathway, where we can mark the targets being investigated with a different colour.

Example: WNT signalling pathway is a pathway of interest in AGA. Using this service, the user can retrieve an image (.png file) of the pathway marked with a different colour for

some of the targets of interest (in this case LGR4, CCND1, DKK1, CTNNB1, SMAD4, WNT10A, WNT10B).

Tool Service 19 (Pathway Network Visualisation)

R- package PANEV allows us to perform pathway analysis '*taking into account both upstream and downstream dependent network of functional related genes, from 2 to n degrees of interaction and generating n .txt files (one for each level of interaction) containing the candidate genes and the related pathways highlighted*' (Palombo et al., 2020). A diagram visualisation is also output along with the tabular format of the results

Example: We create a .txt file in Notepad containing all 48 genes and we also include their corresponding NCBI and ENSEMBL IDs (translated using Tool Service 12). To create the network, we manually input the KEGG IDs for the pathways we want to search (in this case we have included WNT signalling, NOTCH signalling and Hedgehog signalling pathways). The resulting diagram visualises the relationships of genes and three crucial pathways in hair growth and development.

Tool Service 20 (Comparative Toxicogenomic data)

This service takes advantage of CTD data files to retrieve valuable relations between chemicals, targets, pathways and diseases.

Example: A simple query for chemicals implicated in androgenetic alopecia shows 2,136 entries. AGA is implicated in 88 pathways and is linked to 1039 GO Terms and 13 genes according to the Comparative Toxicogenomics database.

5.6 Development of Tool Services to enable the collection and analysis of QSAR data

An overview of the Tool Services used in this section is given below in Table 5.6.

Table 5.5: An overview of the Tool Services used in section 5.6

Tool Service	Title	Purpose	Methods	Output	R packages involved	Other comments
QSAR data						
21	QSAR-DB predictions	To predict properties	The user inputs SMILES of a compound	Two .csv files are produced. Contains the predicted value for the compound's Melting Point and the other the predicted values for Abraham's solvation coefficients	Code was developed in this project	
22	QSAR Classification	To assess if phytochemicals can bind or not to the androgen receptor.	The user inputs the .csv file containing the compounds, their calculated molecular descriptors and column of activity (active, inactive)	A summary of the two models' classification statistics (Accuracy, Sensitivity, Specificity, AUC and F1 score) and an ROC curve, gives an indication of which model performs best. k-folds cross validation is also available with an average of AUC for all folds used.	Code adapted from: (Seal, 2014)	Uses Random forest and Naïve Bayes algorithms

Tool Service 21: **(QSAR-DB predictions)**

This service is taking advantage of QSARDB's API (Ruusmann, Sild and Maran, 2015) to make predictions for properties such as a compound's melting point, and Abrahams solvent coefficients and descriptors. The Abraham model is based on the linear free energy relationship (LFER):

$$\text{Log SP} = c + eE + aA + bB + vV$$

where: **log SP** is the solute property in a given system, **E, S, A, B, and V** are effective parameters for the polarizability, polarity, hydrogen-bond acidity and basicity, and volume of the solute molecules, respectively and **c, e, a, b, v** are the Abrahams solvent coefficients.

This equation is widely used for the prediction of drug molecules partitioning between blood and select body organs (Bradley et al., 2015).

Example: If we use the SMILES annotation for myristic acid [**CCCCCCCCCCCCCC(=O)O**] we get a predicted value for its melting point of 51.5 ^{o}C. Similarly using the same annotation, we get the values for Abrahams descriptors and Abraham's solvent coefficients.

	Abraham's descriptors					
Phytochemical	L	V	B	A	S	E
myristic acid	7.731	2.029	0.604	0.46	0.79	0.402
	Abraham's solvent coefficients					
Phytochemical	v	b	a	s	e	c
myristic acid	4.15	-4.411	-0.359	-0.43	0.487	0.055

Tool Service 22 (QSAR Classification)

This Tool Service is also an adaptation of code by Seal (Seal, 2015). It uses a dataset (SMILES and corresponding classes) as input. SMILES are parsed and using rcdk package fingerprints (such as estate, maccs, Klekota and Roth, PubChem etc.) which are generated to be used as the independent variables. As in the previous service, data are split into training and test sets containing random compounds in a proportion set by the user and can be modelled with three different algorithms (Random forest, Naïve Bayes and Support Vector Machines). A summary of the three models' classification statistics (Accuracy, Sensitivity, Specificity, AUC and F1 score) and an ROC curve, gives an indication of which model performs best. Finally, k-folds cross validation is also available with an average of AUC for all folds used.

5.7 Summary of Methodologies and workflow followed in this thesis

This chapter presented the methodologies and sources of information used to develop the twenty-two Tool Services employed in this thesis. Examples of each Tool Service were given to demonstrate their functionality. Depending on the project, future users may use them individually or by combining them according to preference.

To achieve the aims set in this thesis we followed the workflow described below. The main results of its application are given in Chapter 6.

Workflow

Tool Service 1 was used to collect all articles used in this thesis. It enabled the compilation of a fully-referenced database of 69 plants used ethnopharmacologically in AGA and their respective 2,157 phytochemical constituents. It was also used to collect PubMed information on current therapeutics used in AGA as monotherapy.

Tool Services 2-6 were used to text-mine PubMed articles and retrieve patent, GWAS, clinical trial and biochemical assay data related to AGA.

Plant data from the database was explored using Tool Services 7 and 8. Phytochemicals underwent computational analyses including structural similarity assesments, drug development investigations and safety evaluations using Tool Service 9. Current therapeutics were researched for various pharmacological parameters using Tool Service 10. Data from Tool Services 2-4 was evaluated and a decision was made on which targets should be studied. Tool Services 11-20 provides valuable information on these targets and the opportunity to study the links between these targets, relevant diseases and biochemical pathways. Finally using Tool Services 21 and 22 we managed to deduce predictions regarding the partitioning of phytochemicals between blood and select body organs and to create a validated QSAR classification model for their ability to bind the androgen receptor (AR).

The workflow of the Tool Services as applied in this thesis is illustrated in Figure. 5.7 below.

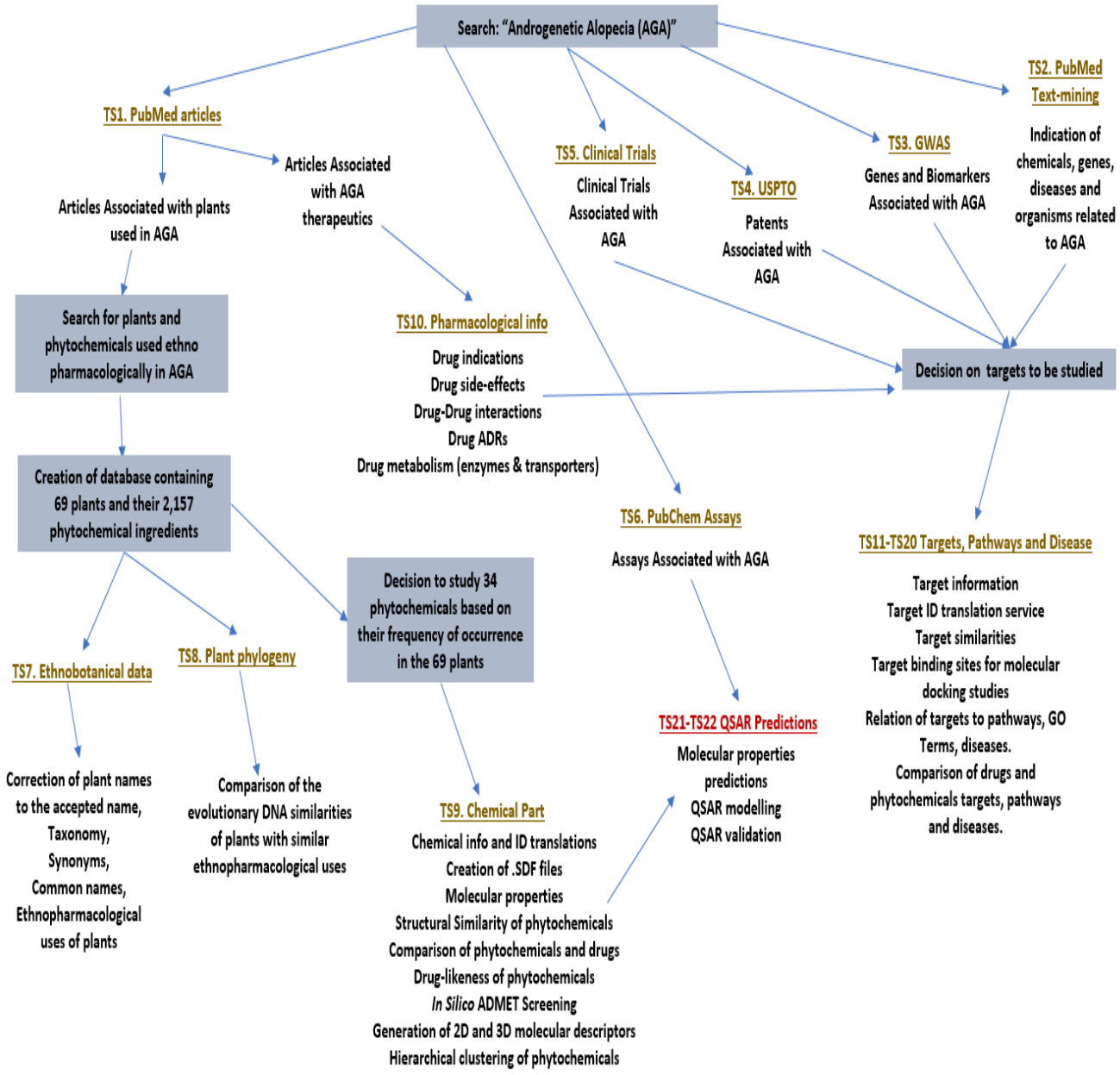


Figure 5.7: A diagram of the workflow of Tool Services as these were applied in this thesis.

CHAPTER 6: Results and discussion

6.1 Results from the application of Tool Services 1-6

6.1.1 Results from the application of Tool Service 1 (PubMed search)

Tool Service 1 was used multiple times to enable PubMed queries in R, at all stages of this project. To establish a robust and well-referenced database of plants and their corresponding phytochemicals, a meticulous methodology was employed within the framework of this PhD thesis. The initial step involved conducting a targeted search in PubMed, utilizing the keyword "Androgenetic alopecia" to retrieve pertinent articles. Tool Service 1, our specialized tool for PubMed retrieval, was instrumental in facilitating this search process. Subsequently, Tool Service 2 was employed to perform text-mining on the retrieved articles, with a primary focus on identifying plant species co-mentioned in the context of androgenetic alopecia. These identified plant species were diligently recorded for subsequent analysis.

To ensure a comprehensive investigation, manual literature searches were meticulously conducted, encompassing esteemed sources such as Google Scholar and Scopus, in addition to PubMed. This approach enabled an exhaustive review of the existing literature pertaining to plants ethnopharmacologically used in the treatment of androgenetic alopecia, ensuring a comprehensive and up-to-date understanding of the subject matter.

The compilation process yielded a dataset comprising 69 plant species intricately associated with androgenetic alopecia. To procure information on the specific phytochemical constituents present in these plants, additional PubMed searches and text-mining procedures were conducted. Each search was tailored to incorporate the accepted name, common names, and synonyms of the individual plant species. To facilitate this aspect, Tool Service 7 was employed, effectively enabling the retrieval of relevant information on the common names and synonyms associated with the 69 plant species. Moreover, exhaustive manual literature searches were meticulously performed across various pertinent sources to encompass the complete spectrum of available phytochemical information for each plant species.

The final outcome of this comprehensive research endeavor is a fully-referenced database that contains detailed information on 69 plant species ethnopharmacologically used in the treatment of androgenetic alopecia, along with their respective 2,157 phytochemical constituents. This meticulously compiled and referenced database serves as a valuable resource for further analysis and exploration within the domain of androgenetic alopecia research. (Appendix 1).

As it is impossible to record all of the queries made to produce the database we present here the articles retrieved (using Tool Service 1) for the search term: 'androgenetic alopecia plants'.

Tool Service 1 was applied to identify all articles with the key words "androgenetic alopecia plants" in their abstract that were published between 1950 and 2021. The search was performed on 3/12/2021. The output was a .csv file containing 298 article PMIDs, titles and abstracts.

The abstract of one of these articles is shown in Table 6.1.1. The full list of results obtained with Tool Service 1 is provided in Appendix 2.

Table 6.1.1 An illustrative example of an abstract relating to the search term "androgenetic alopecia plants" retrieved from PubMed by Tool Service 1

PMID	Title	Abstract
34959442	Application of medicinal plants in several dermatovenerological entities.	A large number of people suffer from alopecia or hair loss worldwide. Drug-based therapies using minoxidil and finasteride for the treatment of alopecia are available but they have shown various side effects in patients. Thus the use of new therapeutic approaches using bioactive products to reduce the risk of anti-hair-loss medications has been emphasised. Natural products have been used since ancient times and have been proven safe with few side effects. Several studies have demonstrated the use of plants and their extracts to promote hair growth. Moreover, commercial products based on these natural ingredients have been developed for the treatment of alopecia. Several clinical animal and cell-based studies have been conducted to determine the anti-alopecia effects of plant-derived biochemicals. This review is a collective study of phytochemicals with anti-alopecia effects focusing mainly on the mechanisms underlying their hair-growth-promoting effects.

6.1.2 Results from the application of Tool Service 2 (PubMed text-mining)

A tool written in R code was adopted to gain information on androgenetic alopecia. As mentioned in Chapter 5 (relating to the methodology of Tool Service 2), text-mining PubMed for the term “androgenetic alopecia” we initially obtain a bar plot of articles produced for this term for the last 70 years shown in Figure 6.1.2.

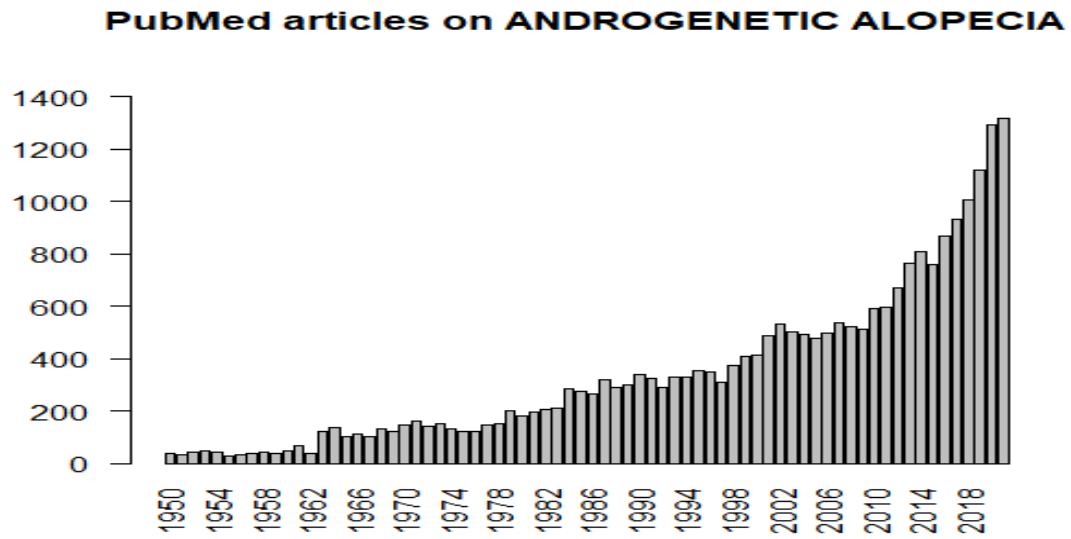


Figure 6.1.2: A barplot showing the number of PubMed articles containing the keyword: ’androgenetic alopecia’ over the last seven decades

The analysis of Figure 6.1.2 reveals a consistent growth in the number of publications related to "androgenetic alopecia" in PubMed since 1950. Notably, there has been a significant acceleration in the publication rate over the past decade. This upward trend in publications may be attributed to the general scientific output across all fields during this period or even indicate a heightened scientific interest in androgenetic alopecia within the scientific community. Whichever is true, there is a vast amount of data which awaits to be exploited.

Six .csv files (Appendix 2) were also retrieved containing information on the frequency of about 50,000 terms, the actual 10,000 PMIDs of the articles used in text-mining, 30 genes, 60 chemicals (along with their MeSH terms), 81 diseases (with their corresponding MeSH terms)

and 16 organisms all co-mentioned in these articles with our search term ‘androgenetic alopecia’. For economy of space only the top 10 entries of these data are presented in Table 6.1.2.

Table 6.1.2 Shows the first 10 entries in each of the six files produced with Tool Service 2

WORDS	
Word	Frequency
hair	11783
alopecia	9795
patients	5911
loss	3882
treatment	3784
scalp	3241
university	2621
information	2364
results	2343
dermatology	2275

GENES	
Gene symbol	NCBI Gene Id
EDA2R	60401
VEGF	7422
PTGDR	5729
AR	367
CTNNB1	1499
DKK1	2294
GRH	2796
INS	3630
SOD	6647
SHBG	6462

PUBMED ARTICLES	
	PMID
1	28349362
2	34741573
3	28294070
4	35044013
5	24566563
6	29595184
7	28396101
8	31677111
9	31415838
10	23974579

CHEMICALS	
Chemicals	MeSH Term
DHT	MESH: D013196
minoxidil	MESH:D008914
finasteride	MESH:D018120
dutasteride	MESH:D000068538
testosterone	MESH:D013739
caffeine	MESH:D002110
ketoconazole	MESH:D007654

DISEASES	
Disease	MeSH Term
baldness	MESH:D000505
pigmentation	MESH:D010859
hyperandrogenism	MESH:D017588
hirsutism	MESH:D006628
anxiety	MESH:D001007
depression	MESH:D000275
Endocrinological diseases	MESH:D004700

SPECIES	
Species	NCBI Taxonomy Id
Homo sapiens	9606
Serenoa repens	4722
Salvia romarinus	39367
SARS-CoV-2	2697049
Mus musculus	10090
Camelia sinensis	4442
Citrullus colocynthis	252529

prostaglandins	MESH:D011453	PCOS	MESH:D011085	*Cuscuta reflexa*	4129
cyproterone acetate	MESH:D017373	hyperprolactinaemia	MESH:D002640	*Abies alba*	45372
spironolactone	MESH:D013148	adrenal hyperplasia	MESH:D000312	*Meleagris gallopavo*	9103

6.1.3 Results from the application of Tool Service 3 (GWAS)

Data in GWAS dataset downloaded from the US National Human Genome Research Institute were manipulated in R to retrieve nine studies (Brockschmidt et al., 2011; Pirastu et al., 2017; Hillmer et al., 2008; Richards et al., 2008; Li et al., 2012; Pickrell et al., 2016; Hagenaars et al., 2017; Yap et al., 2018, Heilmann-Heimbach et al., 2017) relating to androgenetic alopecia. Results of these GWAS were combined to reveal all up to date genes and SNPs implicated in the disease. The p-values derived from the GWAS analysis provide statistical evidence for the association between genes and androgenetic alopecia (AGA). The presence of extremely low p-values indicates a high level of statistical significance, further supporting the biologically plausible involvement of the androgen receptor in AGA. Table 6.1.3 shows the genes implicated in AGA according to the nine Genome-wide association studies with the lowest p-value. The full table with all the results is available in Appendix 3.

Table 6.1.3 The genes with the lowest p-value implicated in androgenetic alopecia according to nine Genome Wide Association studies.

PMID OF GWAS	REGION	REPORTED GENE(S)	SNPs	p-Value
27182965	Xq12	EDA2R, AR	rs200644307	1.00E-247
30573740	Xq12	OPHN1	rs1475417	4.00E-178
28196072	Xq12	AR	rs12558842	5.00E-178
28196072	Xq12	EDA2R, HEPH	rs73221556	5.00E-178
28196072	Xq12	OPHN1	rs5919427	5.00E-178
29146897	20p11.22	PAX1	rs11087368	1.00E-105
22693459	Xq12	AR	rs2497938	2.00E-91
27182965	20p11.22	FOXA2, PAX1	rs201563	3.00E-81
28196072	7p21.1	HDAC9	rs71530654	5.00E-70

6.1.4 Results from the application of Tool Service 4 (USPTO patents)

Tool Service 4 facilitated the retrieval of valuable data from this resource. A query of the USPTO with the term 'androgenetic alopecia' revealed 9 patents (with their IDs and titles). The resulting files can be found in Appendix 4. By using the patent number, the actual .pdf files containing the patent can be retrieved. These patents are summarised in Table 6.1.4.

The assessment of patents targeting specific factors implicated in androgenetic alopecia (AGA) has revealed several interesting findings.

First, four of the treatments (Patents: US6281241, US8758993, US9561224, and US9855204) target the androgen receptor (AR). The androgen receptor plays a crucial role in AGA as it mediates the effects of androgens on hair follicles. Targeting the AR is a logical approach for AGA treatment, as it aims to inhibit the binding of androgens and subsequently reduce their negative effects on hair growth.

Second, two of the treatments (Patents US6281241 and US8067470B2) target the enzyme 5-alpha reductase (SRD5A). SRD5A is responsible for converting testosterone into dihydrotestosterone (DHT), which is the primary androgen involved in AGA. By inhibiting this enzyme, these treatments aim to reduce the production of DHT and mitigate its damaging effects on hair follicles.

Lastly, one treatment (Patent: US8691518) targets minoxidil sulphotransferase (SULT1A1), an enzyme associated with the inter-individual responsiveness to minoxidil therapy. Minoxidil is a widely used topical medication for AGA, and its effectiveness varies among individuals. SULT1A1 is involved in the metabolism of minoxidil, and variations in its activity can affect the response to the treatment. Targeting this enzyme may help optimize the use of minoxidil therapy by addressing individual differences in drug metabolism.

Overall, these results indicate that the retrieved patent data focuses on targeting key factors involved in the condition. By targeting the androgen receptor, 5-alpha reductase, and minoxidil metabolism, these patents demonstrate an effort to develop more effective and personalized approaches to treat AGA.

Table 6.1.4 USPTO Patents retrieved for the term 'androgenetic alopecia' using Tool Service 4

	Patent number	Patent title
1	10,143,639	Use of adelmidrol and other topical or oral cannabinomimetic or aliamide mast cell inhibitors to treat dermatoheliosis, seborrheic keratoses, and androgenetic alopecia
2	9,855,204	Methods and compositions for treating androgenetic alopecia
3	9,561,224	Methods and compositions for treating androgenetic alopecia
4	8,758,993	Systems and methods for predicting response to anti-androgen therapy for the treatment of androgenetic alopecia
5	8,691,518	Systems and methods for predicting response to minoxidil for the treatment of androgenetic alopecia
6	8,067,470	Linoleic acid preparations for the topical treatment of male and female pattern androgenetic alopecia, age-related alopecia, and keratosis pilaris
7	6,953,571	Cosmetic or pharmaceutical composition for topical use to prevent or differ androgenetic alopecia
8	6,281,241	Use of melatonin for the treatment of androgenetic alopecia
9	5,094,857	Treatment of acne or androgenetic alopecia by topical administration of ethisterone

6.1.5 Results from the application of Tool Service 5 (Clinical trials data)

Tool Service 5 allows the user to query and retrieve valuable clinical trials data from ClinicalTrials.gov. The resulting .csv files (found in Appendix 5) contain information about all trial data for the specific search term plus the same data sorted by study, trial phase, gender and age of the participants in these trials.

ClinicalTrials.gov was queried with the term androgenetic alopecia revealing 122 relevant clinical trials. The majority of clinical trials were interventional i.e. having a drug as the intervention with participants mainly adults from both genders. The majority of the trials were Phase 2. The results are shown in Figure 6.1.5 below.

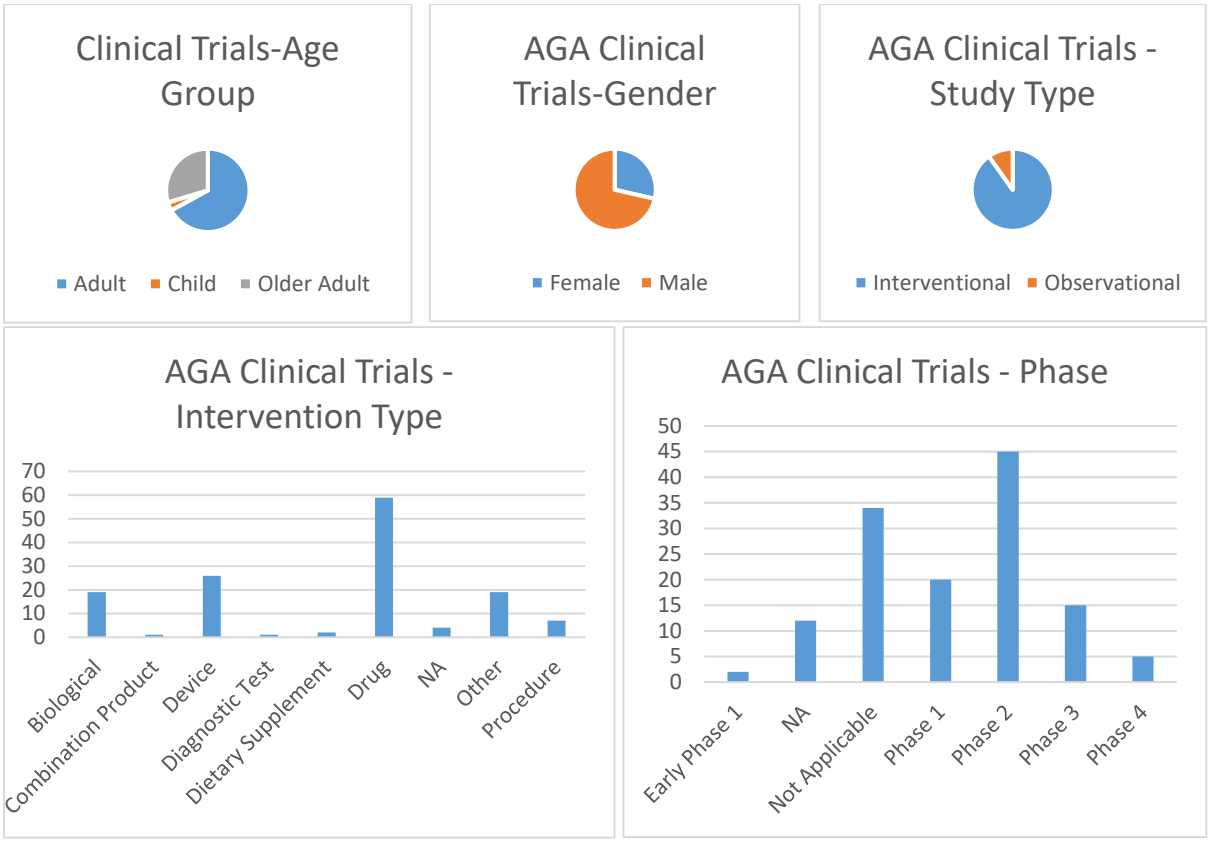


Figure 6.1.5: 122 clinical trials on AGA analysed by: age group, gender, study type, intervention type and Phase.

6.1.6 Results from the application of Tool Service 6 (PubChem assays)

Tool Service 6 is dedicated to retrieving PubChem Assays data. It allows the user to obtain PubChem assay data such as activity of a compound in a specific target.

For the query term “androgenetic alopecia treatment”, information on 40 PubChem assays, with detailed data on 16 active compounds in three targets (SRD5A2, THRB and AR) and their PubChem IDs (SID, CID), activity outcome and values were obtained. The results (available in Appendix 6) are shown in the Table 6.1.6.

Table 6.1.6 Pubchem bioassays retrieved using Tool Service 6 for the term ‘androgenetic alopecia treatment’.

Target	PUBCHEM CID	PUBCHEM ACTIVITY OUTCOME	*Standard Type	Standard Relation	Standard Value	Standard Units
SRD5A2	57363	Active	IC_{50}	=	25	nM
AR	132981	Active	IC_{50}	=	16	nM
AR	132981	Active	IC_{50}	=	40	nM
AR	132981	Active	IC_{50}	=	1000	nM
SRD5A2	5280933	Active	IC_{50}	=	14000	nM
THRB	9819350	Active	Ki	=	0.5	nM
THRB	9819350	Active	Ki	=	0.7	nM
THRB	9847869	Active	Ki	=	1	nM
THRB	9847869	Active	Ki	=	20	nM
THRB	10274474	Active	Ki	=	0.4	nM
THRB	10274474	Active	Ki	=	3	nM
THRB	10274474	Active	EC_{50}	=	0.5	nM
THRB	10276774	Active	Ki	=	0.5	nM
THRB	10276774	Active	Ki	=	2.1	nM
THRB	10279835	Active	Ki	=	2	nM
THRB	10279835	Active	Ki	=	126	nM
THRB	10322653	Active	Ki	=	0.51	nM
THRB	10322653	Active	Ki	=	8	nM
AR	11609033	Active	IC_{50}	=	60	nM
AR	11609033	Active	IC_{50}	=	12	nM
AR	11738282	Active	IC_{50}	=	20	nM
AR	11738282	Active	IC_{50}	=	45	nM
AR	11738282	Active	IC_{50}	=	1660	nM
AR	24180647	Active	IC_{50}	=	39.9	nM
AR	24180647	Active	IC_{50}	=	0.02	nM
AR	24180647	Active	IC_{50}	=	4450	nM
AR	25093227	Active	IC_{50}	=	37	nM
AR	25093227	Active	IC_{50}	=	20.1	nM
AR	25093227	Active	IC_{50}	=	3250	nM
AR	25093229	Active	IC_{50}	=	26	nM
AR	25093229	Active	IC_{50}	=	90	nM
AR	25093230	Active	IC_{50}	=	1599	nM
AR	25093231	Active	IC_{50}	=	37	nM
AR	25093231	Active	IC_{50}	=	43	nM

*In PubChem bioassays, the column termed "Standard Relation" refers to a field that indicates the relationship between the reported value of the biological activity and the expected or standard value. It provides information on how the measured activity compares to the expected activity. The values in this column typically include symbols such as "=", ">", "<", or "~" to denote equality, greater than, less than, or approximate values, respectively.

Summary of the results retrieved with Tool Services 1-6:

The application of Tool Services 1 and 2 (PubMed search and PubMed text-mining), yielded valuable information on the use of plants in the ethnopharmacological treatment of androgenetic alopecia (AGA). By retrieving relevant articles, a comprehensive database consisting of 69 plants and their 2,157 phytochemical constituents was compiled, providing insights into potential therapeutic compounds for AGA. Text-mining of AGA articles further prioritized chemicals and targets associated with AGA.

Tool Service 3 (GWAS analysis) shed light on the genetic factors underlying AGA. Through the combined analysis of nine Genome-wide association studies, genes implicated in AGA were identified. Notably, the androgen receptor (AR) showed a remarkably low p-value, reinforcing its significant role in AGA and emphasising the importance of genetic factors in the development and progression of the condition.

Tool Service 4 (USPTO patents) retrieved a limited number of patents related to AGA, suggesting a relatively low volume of patented inventions in this field. Patent data revealed treatments that focus on inhibiting the androgen receptor (AR) and the enzyme 5-alpha reductase (SRD5A) to reduce the negative effects of androgens on hair follicles. Additionally, a patent targeting minoxidil sulphotransferase (SULT1A1) aims to optimize the use of minoxidil therapy by addressing individual differences in drug metabolism.

Tool Service 5 (Clinical trials data) revealed the existence of 122 relevant clinical trials focused on AGA. The majority of these trials are interventional, indicating active exploration of new treatments and therapies for AGA. The distribution of trials across different parameters, such as age group, gender, study type, intervention type, and trial phase, provides a comprehensive overview of the clinical research landscape in AGA, highlighting diverse approaches and potential interventions under investigation.

Tool Service 6 (PubChem assays) retrieved bioassay data relevant to AGA treatment. These results offer insights into specific targets and compounds that have demonstrated activity in AGA-related assays. Notably, compounds targeting SRD5A2 and AR showed promising activity, suggesting potential therapeutic avenues for AGA treatment.

Overall, results from the application of Tool Services 1-6 provide a comprehensive understanding of AGA, encompassing ethnopharmacology, genetic factors, patents, clinical trials, and bioassay data. This multidimensional approach enhances our knowledge of AGA and facilitates the identification of potential therapeutic targets and interventions for further research and development.

6.2 Results from the application of Tool Services 7-8

6.2.1 Results from the application of Tool Service 7 (Plant information)

Correcting plant nomenclature

Tool Service 7 was able to correct the names of the 69 plants selected (Appendix 7):

61 plants were already associated with their accepted names, six were synonyms (a synonym, according to the International Code of Botanical Nomenclature, is a name considered to apply to the same taxon as the accepted name) and two were unresolved according to The Plant List. The corrected names are reported in Table 6.2.1. The corrected list of plant names thus contained 67 accepted names and only 2 unresolved names.

Table 6.2.1.a The eight plant names that needed correction

Input name	Taxonomic status	Corrected genus	Corrected species	New Taxonomic status
Cnidium officinale	Synonym	*Ligusticum*	*officinale*	Unresolved
Cynomorium songaricum	Synonym	*Cynomorium*	*coccineum*	Accepted
Illicium anisatum	Unresolved	*Illicium*	*anisatum*	Unresolved
Ocimum sanctum	Synonym	*Ocimum*	*tenuiflorum*	Accepted
Pueraria thomsonii	Synonym	*Pueraria*	*montana*	Accepted
Salvia rosmarinus	Synonym	*Rosmarinus*	*officinalis*	Accepted

Solanum nigrum	Synonym	*Solanum*	*americanum*	Accepted
Tamus communis	Synonym	*Dioscorea*	*communis*	Accepted

Plant taxonomy

Tool Service 7 identified the data from NCBI and ITIS regarding the taxonomy of the 69 plants (see Appendix 7). In terms of their taxonomy, families with the highest occurrence and respective plants were:

-**Fabaceae**: (8/69 plants): *Acacia concinna, Albizia julibrissin, Clitoria ternatea, Cullen corylifolium, Pueraria montana var. chinensis, Sophora flavescens, Trifolium pratense.*

-**Asteraceae**: (7/69 plants): *Artemisia abrotanum, Bidens tripartita, Carthamus tinctorius, Chrysanthemum morifolium, chrysanthemum zawadskii, Eclipta prostrata, Wedelia chinensis.*

-**Apiaceae**: (5/69 plants): *Ammi majus, Angelica dahurica, Angelica gigas, Angelica sinensis, Ligusticum officinale.*

-**Lamiaceae**: (4/69 plants): *Leonurus sibiricus, Ocimum teniflorum, Salvia miltiorrhiza, Scutelaria baicalensis.*

-**Zingiberaceae**: (3/69 plants): *Alpinia zerumbet, Curcuma longa, Zingiber officinalis*

-**Solanaceae**: (3/69 plants): *Lycium chinense, Lycopersicon esculentum, Solanum nigrum*

Common names of plants

For 52 of the 69 plants' common names were retrieved from four sources (EOL, ITIS, NCBI, WORMS) using Tool Service 7. These are reported in Table 6.2.1b. Full results are available in Appendix 7.

Table 6.2.1.b An indicative list of plant and their common names identified using Tool Service 7.

Plant Name	**Indicative common names**
Acacia concinna	No common name was retrieved
Albizzia julibrissin	silktree, mimosa, powderpuff tree
Allium cepa	onion, globe onion, garden onion
Allium sativum	cultivated garlic
Aloe vera	Barbados aloe, Indian aloe, Mediterranean aloe
Alpinia zerumbet	shellplant, shellflower, shell ginger
Ammi majus	large bullwort, Bishopsweed, Lace Flower
Andrographis paniculata	Kalmegh
Angelica dahurica	Dahuriankarhunputki

Plant Synonyms

Synonyms for 68 of our plants were retrieved from EOL and ITIS. Full results can be found in Appendix 7. An example of plant synonyms for plant *Acacia concinna* is shown in Table 6.2.1 c.

Table 6.2.1 c.: An example of the synonyms retrieved for the plant *Acacia concinna* using Tool Service 7

Plant Name	Synonyms
Acacia concinna	*Mimosa rugata Lam.| Acacia concinna (Willd.) DC.| Mimosa concinna Willd.| Acacia quisumbingii Merr.| Acacia rugata (Lam.) Merr.| Acacia hooperiana| Acacia concinna var. rugata (Benth.)Baker| Acacia hooperiana var. glabriuscula Miq.| Acacia hooperiana Miq.| Acacia hooperiana var. subcuneata Miq.| Acacia philippinarium Benth.| Acacia poilanei Gagnep.| Acacia polycephala DC.| Guilandina microphylla DC.| Nygae sylvarum-minimae Rumph.*

6.2.2 Results from the application of Tool Service 8 (Plant phylogeny)

Phylogenetic comparison

DNA barcodes are a biological tool which can be used to enhance our understanding of plant systematics, taxonomy and species phylogeny. In the last decade, plant markers such as trnH-psbA, ITS2, rbcL and matK have been widely used to provide accurate information on plant identification and phylogenetic comparisons. Tool Service 8 was used to run the part of code which retrieves plants FASTA files from BOLD Systems database (https://v3.boldsystems.org). These are then combined into an .fas file (Appendix 8). Using the second part of the code, multiple plant DNA sequences are aligned (results available in both .FASTA and .html) and a phylogeny tree showing the evolutionary similarities of the plants is shown in Figure 6.2.2.

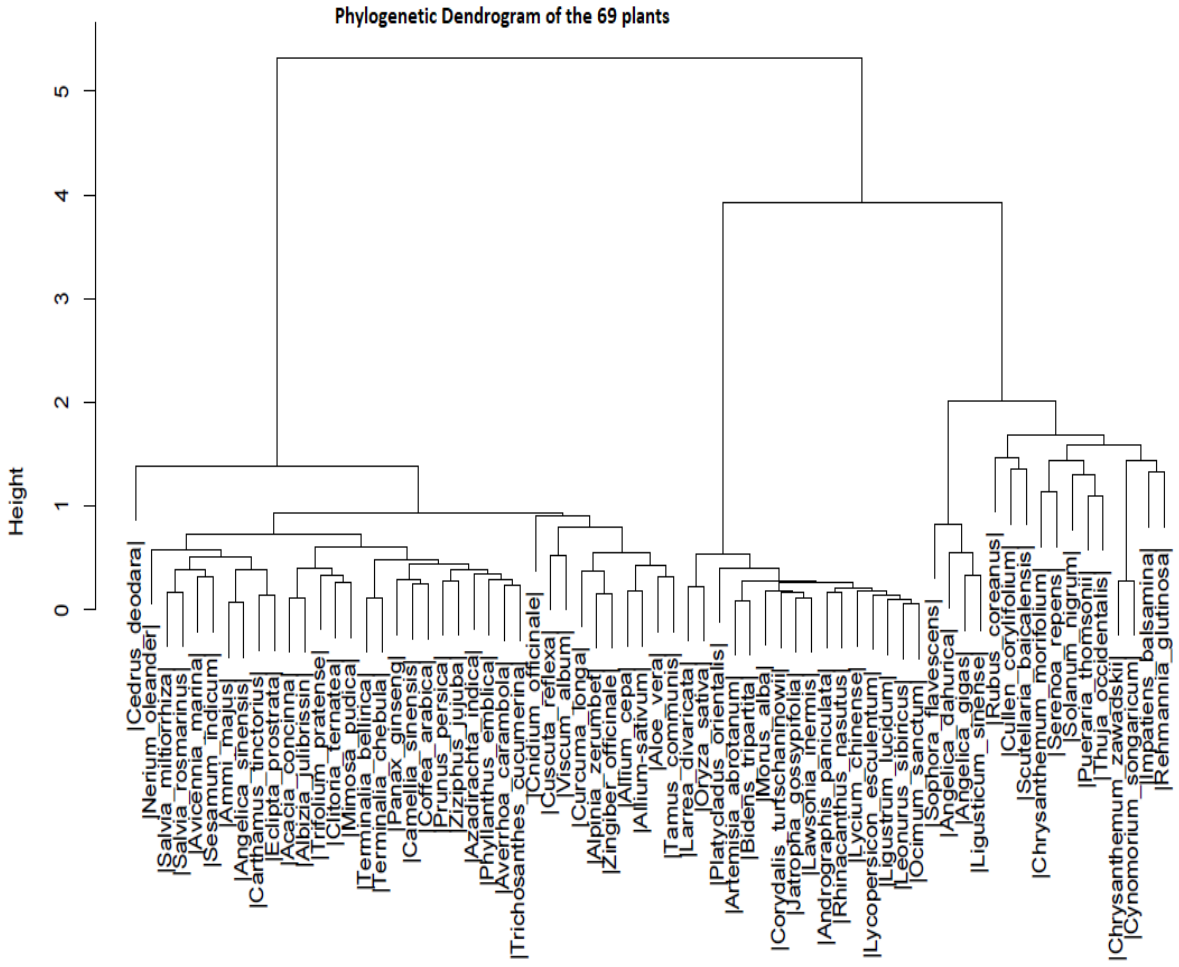


Figure 6.2.2 A phylogenetic diagram of the 69 plants produced with Tool Service 8.

An analysis of the phylogenetic diagram confirms the evolutionary similarities between plants of the same taxonomic groups. This combined to the similarity of taxonomic groups suggests plants have evolved in a way that related species may produce similar phytochemicals.

Summary of the results retrieved with Tool Services 7-8

The results obtained from Tool Service 7 (Plant Information) offer valuable insights into the significance of the identified plants in relation to androgenetic alopecia. One crucial aspect of this tool is the accurate correction of plant names, which ensures the use of proper taxonomic information for subsequent analysis. By rectifying any inaccuracies or inconsistencies in the names of the plants, researchers can confidently rely on the taxonomic data to explore their characteristics and potential therapeutic benefits. Moreover, the analysis goes beyond name correction by shedding light on the families that prominently occur among the selected plants. This exploration expands our understanding of the diversity within the plant kingdom and provides a broader context for assessing their potential medicinal properties. By identifying the families with a higher occurrence, researchers gain insights into the prevalence of certain plant groups associated with androgenetic alopecia, guiding further investigations into their specific attributes and potential contributions to treatment.

Another significant outcome of Tool Service 7, is the retrieval of common names and synonyms for the identified plants. This process attempts to effectively bridge the gap between traditional knowledge and scientific research. The inclusion of common names and synonyms enriches the understanding of these plants by establishing connections with their historical usage and traditional medicinal practices.

Moving on to Tool Service 8 (Plant Phylogeny), it further enhances our understanding of the identified plants and their implications for androgenetic alopecia. By utilising DNA barcodes and conducting phylogenetic comparisons, this tool uncovers the evolutionary relationships and similarities among the plants. The generated phylogenetic diagram serves as a visual representation of these relationships, confirming that plants within the same taxonomic groups exhibit closer evolutionary connections. This finding suggests shared genetic characteristics and phytochemical profiles among related species, providing valuable insights into their potential efficacy in treating androgenetic alopecia.

Understanding the evolutionary relationships of plants is crucial when investigating their therapeutic properties. By recognising that related species within the same taxonomic group tend to share similar phytochemical compounds, researchers can focus their attention on specific taxonomic groups that have demonstrated promising properties. This knowledge helps narrow down the search for effective plant-based treatments and streamlines the exploration of potential candidates for further study.

When combined with the taxonomic information obtained from Tool Service 7, the phylogenetic analysis offered by Tool Service 8 provides a comprehensive perspective on the identified plants. This integrated approach allows researchers to identify candidate plants with similar genetic backgrounds and explore their shared characteristics and medicinal potential more effectively. By considering both the taxonomic relationships and evolutionary connections among the plants, researchers can make informed decisions regarding which plant groups to prioritize in their investigations, ultimately advancing our understanding of androgenetic alopecia and opening new avenues for therapeutic interventions.

6.3 Results from the application of Tool Service 9 (Chemical data)

As the number of phytochemicals contained in the 69 plants was quite large (2,157 phytochemicals), a decision was made to study only those with the highest frequency occurrence in the 69 plants. Thus, a set of 34 phytochemicals was selected. A full list of the 34 phytochemicals, along with the plants they are found in, is given in Appendix 9. Table 6.3.1 below shows three of them.

Table 6.3.1 Three of the phytochemicals selected for analysis and the number of plants they are found in.

Phytochemical	Pubchem CID	Found in # plants	Plant names
beta-sitosterol	222284	17	*Cynomorium songaricum, Curcuma longa, Tamus communis, Panax ginseng, Morus alba, Viscum album, Sophora flavescens, Lycium chinense, Eclipta prostrata, Rubus coreanus, Zingiber officinale, Ligusticum sinense, Solanum nigrum, Sesamum indicum, Salvia miltiorrhiza, Angelica dahurica, Carthamus tinctorius*
palmitic acid	985	16	*Cynomorium songaricum, Terminalia chebula, Carthamus tinctorius, Oryza sativa, Zingiber officinale, Cnidium officinale, Viscum album, Panax ginseng, Curcuma longa, Eclipta prostrata, Ziziphus jujuba, Bidens tripartita, Lycium chinense, Chrysanthemum zawadskii, Andrographis paniculata, Albizia julibrissin*
ferulic acid	445858	15	*Zingiber officinale, Cnidium officinale, Angelica sinensis, Salvia miltiorrhiza, Allium sativum, Ligusticum sinense, Rubus coreanus, Solanum nigrum, Ziziphus jujuba, Scutellaria baicalensis, Lycium chinense, Carthamus tinctorius, Chrysanthemum morifolium, Sesamum indicum, Morus alba*

Using Tool Service 9, the 34 phytochemicals listed in Appendix 7 were compared in terms of structural similarity. Combining the structural similarity files for the 34 phytochemicals, an atom pair distance matrix (Appendix 9) containing theTanimoto similarities between them was produced. Two chemical structures are considered similar if their Tanimoto coefficient is higher than 0.85. Those pairs of phytochemicals in the matrix with a Tanimoto coefficient > 0.85 are shown in Figure 6.3.1.

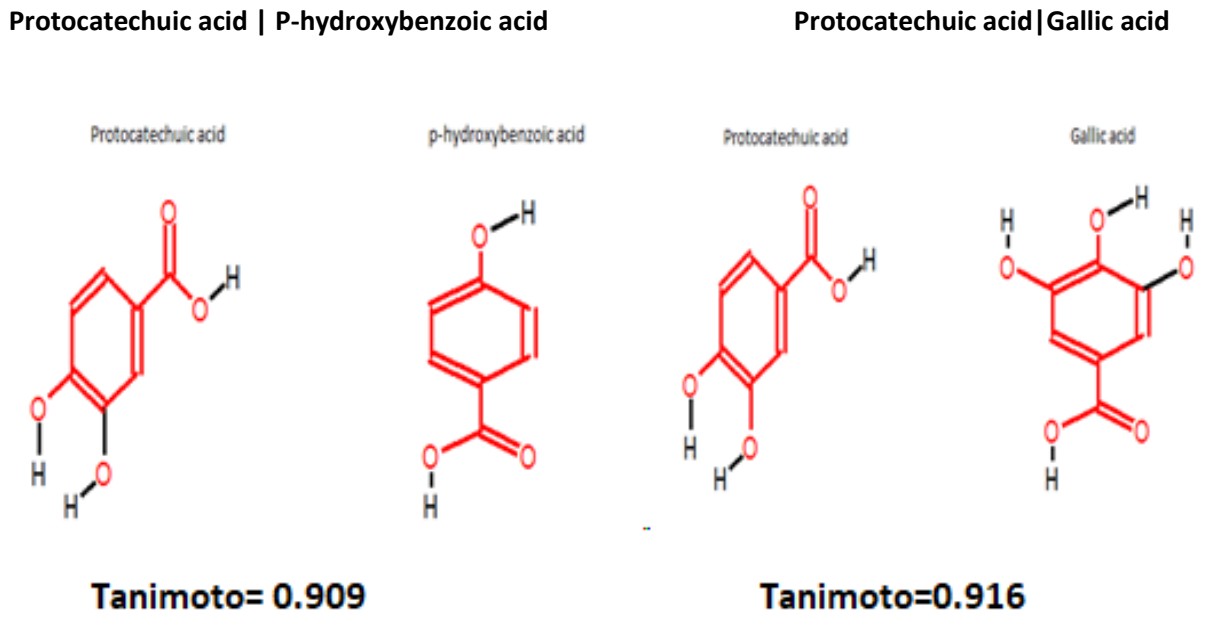
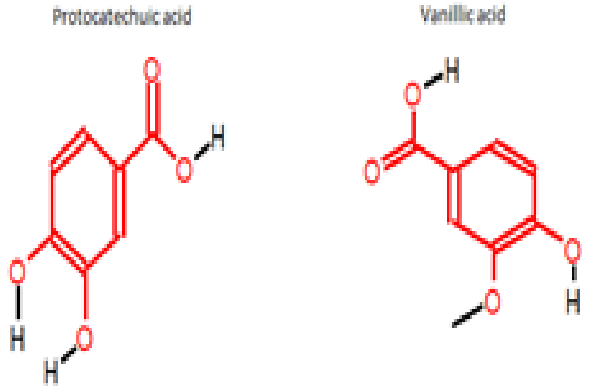


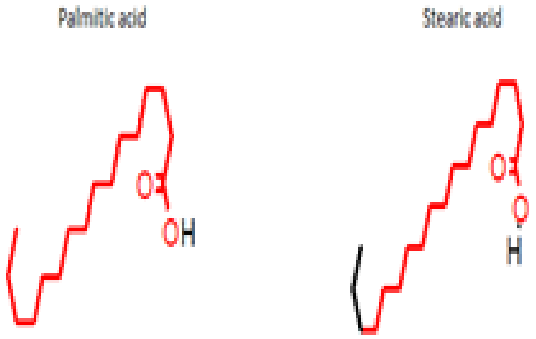
Figure 6.3.1. Pairs of the 34 phytochemicals with a Tanimoto coefficient >0.85

Protocatechuic acid|Vanillic acid



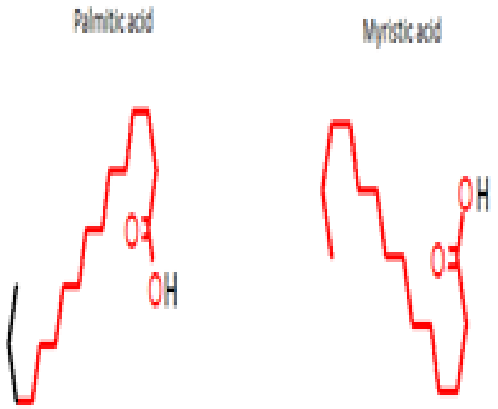
Tanimoto=0.916

Palmitic acid|Stearic acid



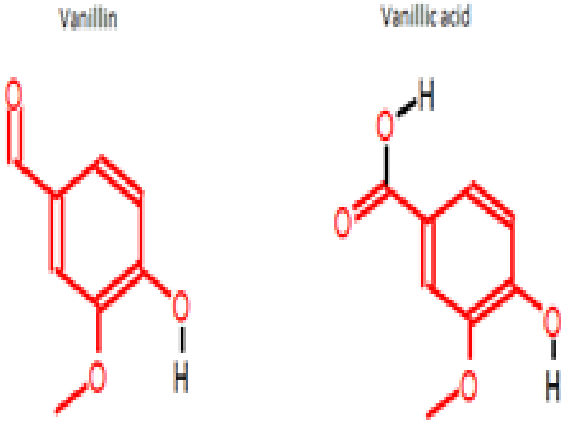
Tanimoto=0.900

Palmitic acid|Myristic acid



Tanimoto=0.888

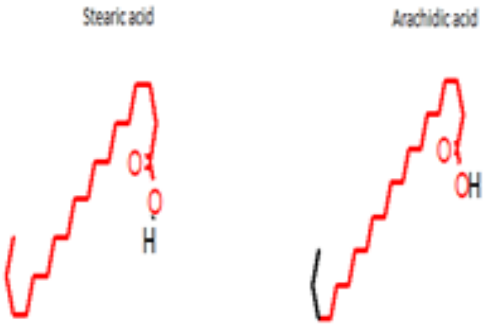
Vanillin|Vanillic acid



Tanimoto=0.916

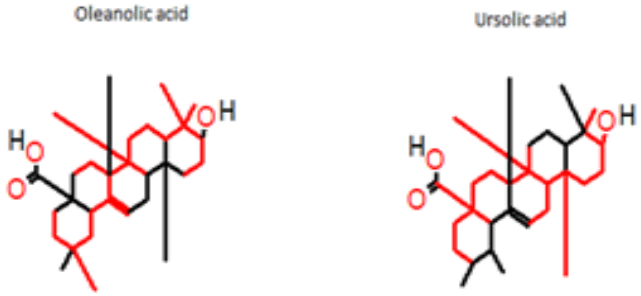
Figure 6.3.1. Pairs of the 34 phytochemicals with a Tanimoto coefficient >0.85 (continued)

Stearic acid | Arachidic acid



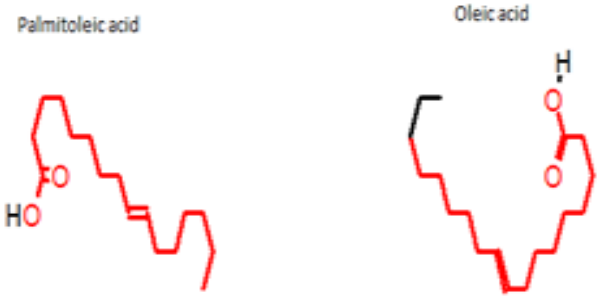
Tanimoto=0.909

Oleanolic acid|Ursolic acid



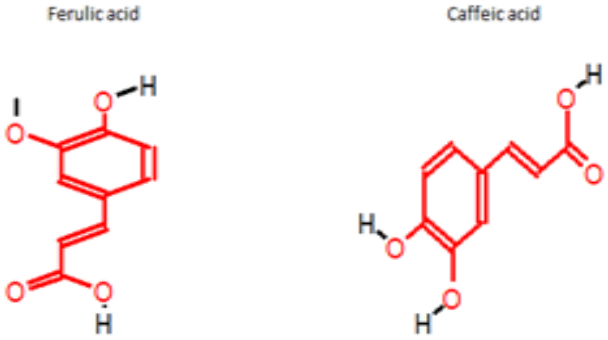
Tanimoto=0.941

Palmitoleic acid| Oleic acid



Tanimoto=0.900

Ferulic acid | Caffeic acid

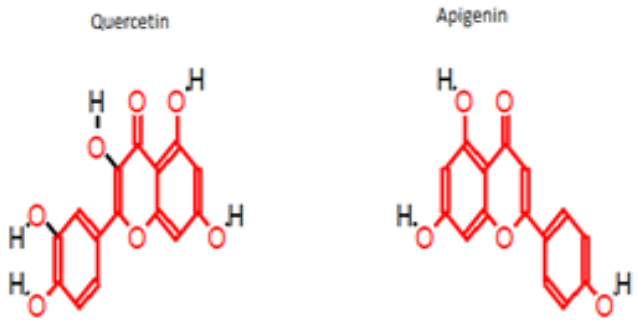


Tanimoto=0.928

Quercetin |Apigenin

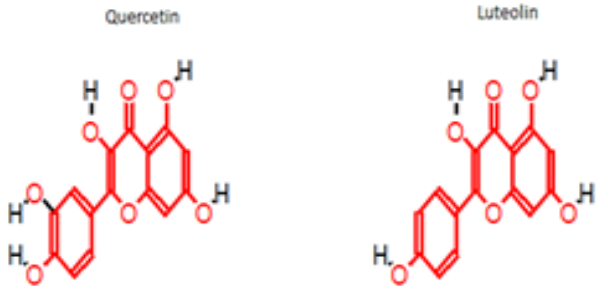
Quercetin|Luteolin

Figure 6.3.1. Pairs of the 34 phytochemicals with a Tanimoto coefficient >0.85 (continued)



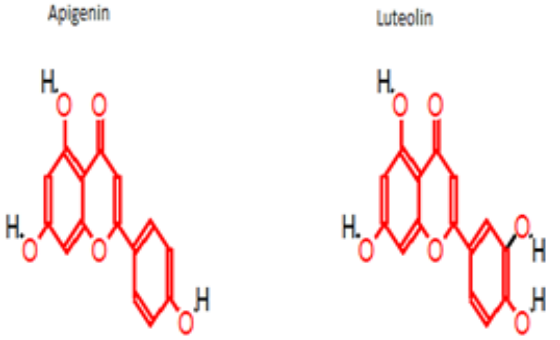
Tanimoto=0.909

Apigenin|Luteolin

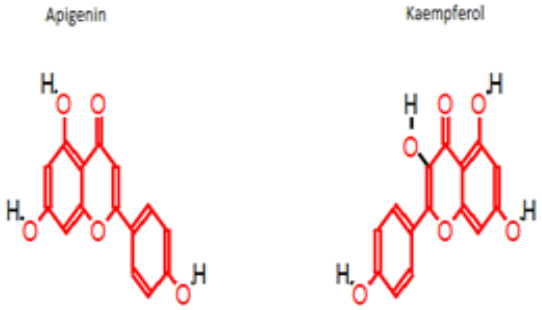


Tanimoto=0.954

Apigenin|Kaempferol

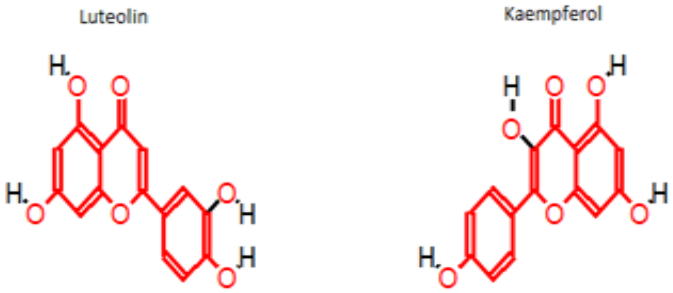


Tanimoto=0.952



Tanimoto=0.952

Luteolin|Kaempferol



Tanimoto=0.909

Figure 6.3.1. Pairs of the 34 phytochemicals with a Tanimoto coefficient >0.85 (continued)

Clustering of the phytochemicals

The purpose of clustering is to make sense of and extract value from a dataset. This logical grouping of data is a common strategy before attempting to analyse it. Hierarchical clustering of chemicals can be used to observe similarities of chemicals in a cluster such as in their activity, their safety or binding affinity for a target.

This part of Tool Service 9 hierarchically clustered our phytochemicals. A distance matrix of all-against-all compound distances was created by deducting the similarity metric (Tanimoto coefficient Tc) from one, and this was provided as the input (1 - Tc). Next, the resulting distance matrix was used to create a hierarchy of moleculess, ranging from most to least similar. The output is presented as a cluster dendrogram. The clustering analysis based on Tanimoto similarity of the 34 chemicals is shown in Figure 6.3.2.

From analysing the dendrogram we can designate two big clusters. The first cluster from top is consisted mainly of phenolic acids and flavonoids while the second cluster contains mainly fatty acids and terpenoids.

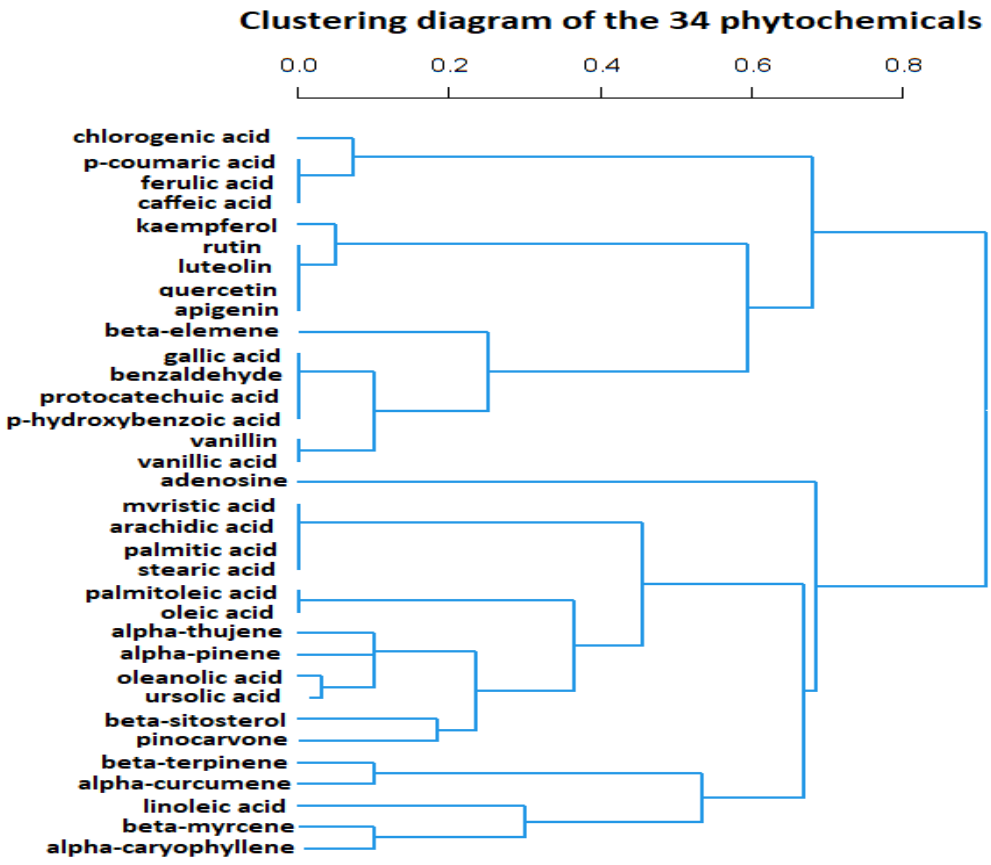


Figure 6.3.2. A hierarchical clustering diagram of the 34 phytochemicals

Lipinski's Rule of Five (RO5)

Lipinski (1997) suggested four parameters for the identification of orally bioavailable drug-like molecules based on an analysis of existing drugs. The rule of five (RO5) got its name from the cutoff values of these parameters which are either five or a multiple of five. The rule states that a molecule is less likely to have sufficient oral absorption (or permeation) in cases that:

- It has more than 5 groups acting as hydrogen bond donors
- It has more than 10 groups acting as hydrogen bond acceptors
- Its molecular weight is more than 500 Da
- Log P is greater than 5.

Most phytochemicals had no Ro5 violations. Those chemicals with >1 violations were: palmitic acid, stearic acid, arachidic acid, myristic acid, palmitoleic acid, oleic acid, linoleic acid and rutin. Full results are shown in file Ruleof5.csv (Appendix 9).

SMARTS patterns

Daylight SMARTS is a 'language' used to describe and write molecular patterns. SMARTS is a fast and efficient way to query long molecular datasets for specific substructures. SMARTS can identify substructures that may pose different risks during drug development while searching for leads in chemical libraries.

The structures of 34 phytochemicals were loaded as an SDF to Tool Service 9 and were compared against different sets of SMARTS. The results showed that none of the 34 phytochemicals belong to either Hann's unsuitable leads or Hann's unsuitable natural products (Hann et al., 1999). On the other hand, applying Walters and Murcko's (2002) drug likeness SMARTS alerts, it seems that benzaldehyde fails this drug-likeness test as it contains an aldehyde group (Walters_smarts_2) and four phytochemicals (palmitic acid, stearic acid, eicosanoic acid, myristic acid) due to their long aliphatic chain they contain.

In terms of skin sensitisation several phytochemicals contain structures considered to cause skin sensitisation when compared against Enoch's SMARTS (Enoch, Madden and Cronin, 2008). These phytochemicals can be seen in Table 6.3.2.

Table 6.3.2 Phytochemicals which may cause skin sensitisation

Skin sensitisation SMARTS	**Phytochemicals containing the structure**
Enoch_10	benzaldehyde, vanillin
Enoch_20	ferulic acid, caffeic acid, coumaric acid, chlorogenic acid
Enoch_21	Pincarvone
Enoch_22	ferulic acid, caffeic acid, coumaric acid, chlorogenic acid
Enoch_31	protocatechuic acid, gallic acid, caffeic acid, chlorogenic acid, quercetin, luteolin, rutin
Enoch_33	vanillin, vanillic acid, ferulic acid
Enoch_36	gallic acid, quercetin, apigenin, luteolin, rutin, kaempferol

In order to assess the risk of the phytochemicals causing drug-induced human liver injury, they were compared against Liu et al. (2015) SMARTS, with none having a similar structure. Similar results were retrieved by using Madden et al. (2014) alerts for mitochondrial toxicity.

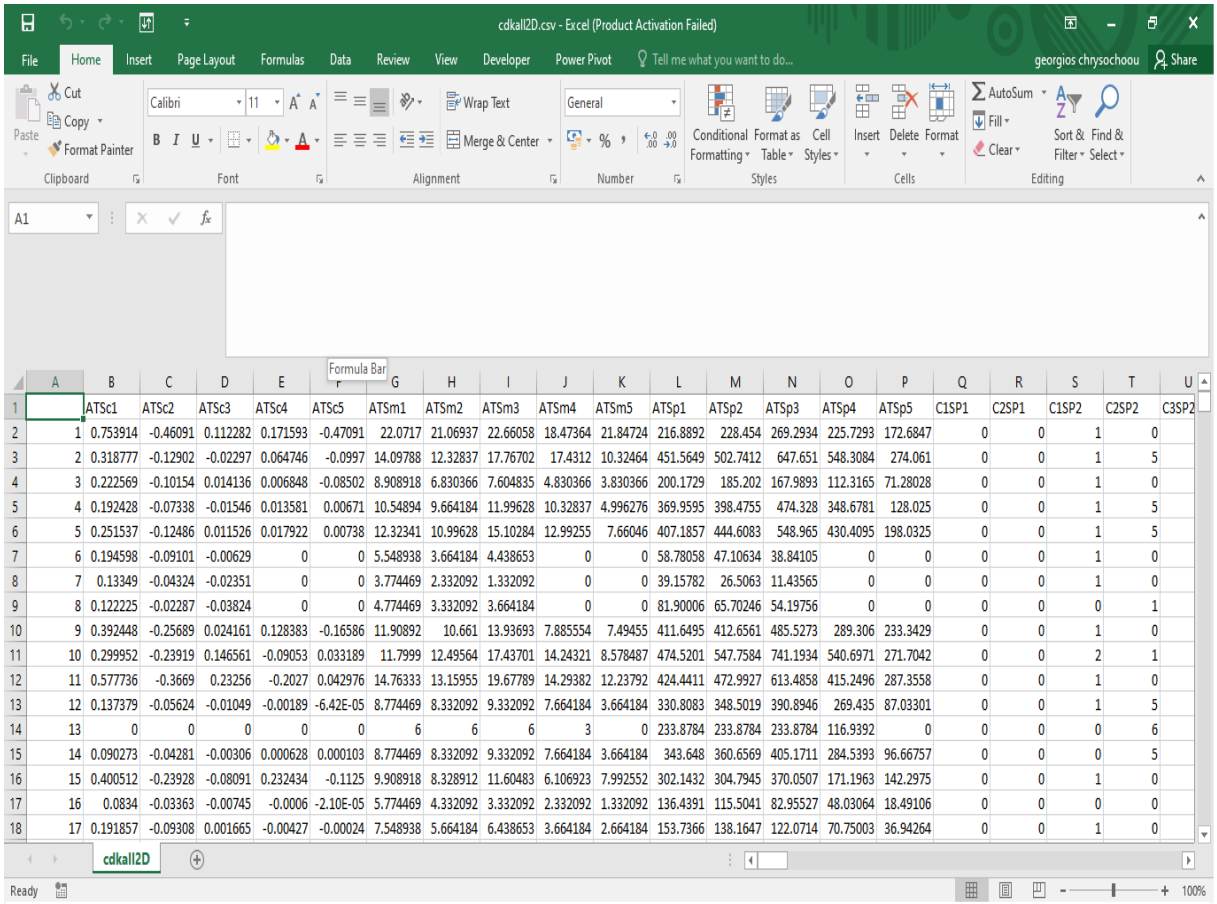
Mutagenicity is another issue to avoid in drug discovery. Kazius, McGuire and Bursi (2005) identified 37 structural alerts. Only adenosine had a similar structure (aromatic amine) which is considered to induce mutagenicity. Benigni and Bossa (2008) have identified 35 structural alerts regarding carcinogenicity. Phytochemicals benzaldehyde and vanillin, were found to have a simple aldehyde group and could be potential carcinogens. It should be noted that no results were found for some alerts as these could not be parsed by smartSearchOB function of the ChemmineOB package used in this part of the code. These include the following alerts: Hann 32 and 53, Liu 6, Madden 8 and 9, Kazius 8 and 25, and Benigni&Bossa 18,19 and 20. This could most probably be attributed to the limitation of the package in handling complex SMARTS patterns and unsupported features such as various levels of nested substructures and conditional expressions.

CDK 2D and 3D molecular descriptors

CDK is a tool for molecular descriptor calculations written in Java. Using Tool Service 9 we may also retrieve 2D and 3D molecular descriptors from package rcdk.

The resulting 2D and 3D molecular descriptors and a list describing the calculated CDK descriptors can be found in Appendix 9. The snapshots of the two excel files produced are shown in Figures 6.3.3 and 6.3.4 below.

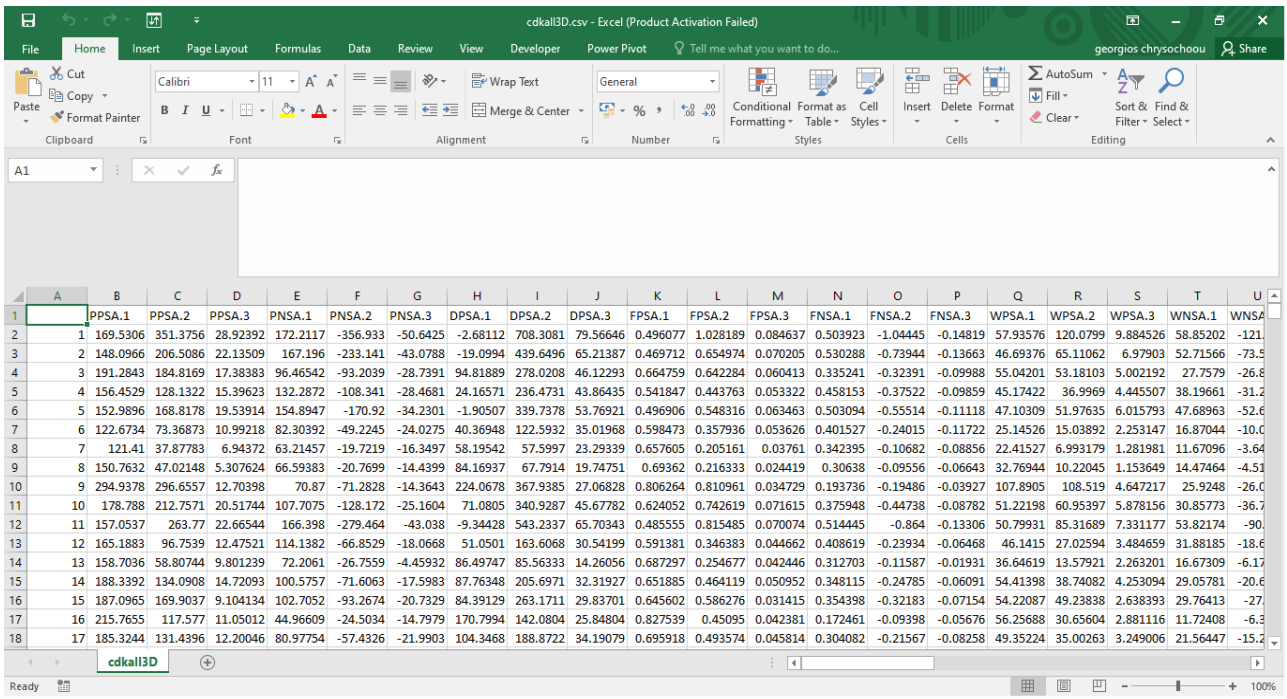
2D-CDK molecular descriptors



		ATSc1	ATSc2	ATSc3	ATSc4	ATSc5	ATSm1	ATSm2	ATSm3	ATSm4	ATSm5	ATSp1	ATSp2	ATSp3	ATSp4	ATSp5	C1SP1	C2SP1	C1SP2	C2SP2	C3SP2
2	1	0.753914	-0.46091	0.112282	0.171593	-0.47091	22.0717	21.06937	22.66058	18.47364	21.84724	216.8892	228.454	269.2934	225.7293	172.6847	0	0	1	0	
3	2	0.318777	-0.12902	-0.02297	0.064746	-0.0997	14.09788	12.32837	17.76702	17.4312	10.32464	451.5649	502.7412	647.651	548.3084	274.061	0	0	1	5	
4	3	0.222569	-0.10154	0.014136	0.006848	-0.08502	8.908918	6.830366	7.604835	4.830366	3.830366	200.1729	185.202	167.9893	112.3165	71.28028	0	0	1	0	
5	4	0.192428	-0.07338	-0.01546	0.013581	0.00671	10.54894	9.664184	11.99628	10.32837	4.996276	369.9595	398.4755	474.328	348.6781	128.025	0	0	1	5	
6	5	0.251537	-0.12486	0.011526	0.017922	0.00738	12.32341	10.99628	15.10284	12.99255	7.66046	407.1857	444.6083	548.965	430.4095	198.0325	0	0	1	5	
7	6	0.194598	-0.09101	-0.00629	0	0	5.548938	3.664184	4.438653	0	0	58.78058	47.10634	38.84105	0	0	0	0	1	0	
8	7	0.13349	-0.04324	-0.02351	0	0	3.774469	2.332092	1.332092	0	0	39.15782	26.5063	11.43565	0	0	0	0	1	0	
9	8	0.122225	-0.02287	-0.03824	0	0	4.774469	3.332092	3.664184	0	0	81.90006	65.70246	54.19756	0	0	0	0	0	1	
10	9	0.392448	-0.25689	0.024161	0.128383	-0.16586	11.90892	10.661	13.93693	7.885554	7.49455	411.6495	412.6561	485.5273	289.306	233.3429	0	0	1	0	
11	10	0.299952	-0.23919	0.146561	-0.09053	0.033189	11.7999	12.49564	17.43701	14.24321	8.578487	474.5201	547.7584	741.1934	540.6971	271.7042	0	0	2	1	
12	11	0.577736	-0.3669	0.23256	-0.2027	0.042976	14.76333	13.15955	19.67789	14.29382	12.23792	424.4411	472.9927	613.4858	415.2496	287.3558	0	0	1	0	
13	12	0.137379	-0.05624	-0.01049	-0.00189	-6.42E-05	8.774469	8.332092	9.332092	7.664184	3.664184	330.8083	348.5019	390.8946	269.435	87.03301	0	0	1	5	
14	13	0	0	0	0	0	6	6	6	3	0	233.8784	233.8784	233.8784	116.9392	0	0	0	0	6	
15	14	0.090273	-0.04281	-0.00306	0.000628	0.000103	8.774469	8.332092	9.332092	7.664184	3.664184	343.648	360.6569	405.1711	284.5393	96.66757	0	0	0	5	
16	15	0.400512	-0.23928	-0.08091	0.232434	-0.1125	9.908918	8.328912	11.60483	6.106923	7.992552	302.1432	304.7945	370.0507	171.1963	142.2975	0	0	1	0	
17	16	0.0834	-0.03363	-0.00745	-0.0006	-2.10E-05	5.774469	4.332092	3.332092	2.332092	1.332092	136.4391	115.5041	82.95527	48.03064	18.49106	0	0	0	0	
18	17	0.191857	-0.09308	0.001665	-0.00427	-0.00024	7.548938	5.664184	6.438653	3.664184	2.664184	153.7366	138.1647	122.0714	70.75003	36.94264	0	0	1	0	

Figure 6.3.3: A snapshot of the 279 2D molecular descriptors calculated with Tool Service 9 from CDK

3D-CDK molecular descriptors



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		PPSA.1	PPSA.2	PPSA.3	PNSA.1	PNSA.2	PNSA.3	DPSA.1	DPSA.2	DPSA.3	FPSA.1	FPSA.2	FPSA.3	FNSA.1	FNSA.2	FNSA.3	WPSA.1	WPSA.2	WPSA.3	WNSA.1	WNSA
2	1	169.5306	351.3756	28.92392	172.2117	-356.933	-50.6425	-2.68112	708.3081	79.56646	0.496077	1.028189	0.084637	0.503923	-1.04445	-0.14819	57.93576	120.0799	9.884526	58.85202	-121.
3	2	148.0966	206.5086	22.13509	167.196	-233.141	-43.0788	-19.0994	439.6496	65.21387	0.469712	0.654974	0.070205	0.530288	-0.73944	-0.13663	46.69376	65.11062	6.97903	52.71566	-73.5
4	3	191.2843	184.8169	17.38383	96.46542	-93.2039	-28.7391	94.81889	278.0208	46.12293	0.664759	0.642284	0.060413	0.335241	-0.32391	-0.09988	55.04201	53.18103	5.002192	27.7579	-26.8
5	4	156.4529	128.1322	15.39623	132.2872	-108.341	-28.4681	24.16571	236.4731	43.86435	0.541847	0.443763	0.053322	0.458153	-0.37522	-0.09859	45.17422	36.9969	4.445507	38.19661	-31.2
6	5	152.9896	168.8178	19.53914	154.8947	-170.92	-34.2301	-1.90507	339.7378	53.76921	0.496906	0.548316	0.063463	0.503094	-0.55514	-0.11118	47.10309	51.97635	6.015793	47.68963	-52.6
7	6	122.6734	73.36873	10.99218	82.30392	-49.2245	-24.0275	40.36948	122.5932	35.01968	0.598473	0.357936	0.053626	0.401527	-0.24015	-0.11722	25.14526	15.03892	2.253147	16.87044	-10.0
8	7	121.41	37.87783	6.94372	63.21457	-19.7219	-16.3497	58.19542	57.5997	23.29339	0.657605	0.205161	0.03761	0.342395	-0.10682	-0.08856	22.41527	6.993179	1.281981	11.67096	-3.64
9	8	150.7632	47.02148	5.307624	66.59383	-20.7699	-14.4399	84.16937	67.7914	19.74751	0.69362	0.216333	0.024419	0.30638	-0.09556	-0.06643	32.76944	10.22045	1.153649	14.47464	-4.51
10	9	294.9378	296.6557	12.70398	70.87	-71.2828	-14.3643	224.0678	367.9385	27.06828	0.806264	0.810961	0.034729	0.193736	-0.19486	-0.03927	107.8905	108.519	4.647217	25.9248	-26.0
11	10	178.788	212.7571	20.51744	107.7075	-128.172	-25.1604	71.0805	340.9287	45.67782	0.624052	0.742619	0.071615	0.375948	-0.44738	-0.08782	51.22198	60.95397	5.878156	30.85773	-36.7
12	11	157.0537	263.77	22.66544	166.398	-279.464	-43.038	-9.34428	543.2337	65.70343	0.485555	0.815485	0.070074	0.514445	-0.864	-0.13306	50.79931	85.31689	7.331177	53.82174	-90.
13	12	165.1883	96.7539	12.47521	114.1382	-66.8529	-18.0668	51.0501	163.6068	30.54199	0.591381	0.346383	0.044662	0.408619	-0.23934	-0.06468	46.1415	27.02594	3.484659	31.88185	-18.6
14	13	158.7036	58.80744	9.801239	72.2061	-26.7559	-4.45932	86.49747	85.56333	14.26056	0.687297	0.254677	0.042446	0.312703	-0.11587	-0.01931	36.64619	13.57921	2.263201	16.67309	-6.17
15	14	188.3392	134.0908	14.72093	100.5757	-71.6063	-17.5983	87.76348	205.6971	32.31927	0.651885	0.464119	0.050952	0.348115	-0.24785	-0.06091	54.41398	38.74082	4.253094	29.05781	-20.6
16	15	187.0965	169.9037	9.104134	102.7052	-93.2674	-20.7329	84.39129	263.1711	29.83701	0.645602	0.586276	0.031415	0.354398	-0.32183	-0.07154	54.22087	49.23838	2.638393	29.76413	-27.
17	16	215.7655	117.577	11.05012	44.96609	-24.5034	-14.7979	170.7994	142.0804	25.84804	0.827539	0.45095	0.042381	0.172461	-0.09398	-0.05676	56.25688	30.65604	2.881116	11.72408	-6.3
18	17	185.3244	131.4396	12.20046	80.97754	-57.4326	-21.9903	104.3468	188.8722	34.19079	0.695918	0.493574	0.045814	0.304082	-0.21567	-0.08258	49.35224	35.00263	3.249006	21.56447	-15.2

Figure 6.3.4: A snapshot of the 66 3D molecular descriptors calculated with Tool Service 9 from CDK

SIMCOMP results

SIMCOMP (https://www.genome.jp/tools/simcomp) is a chemical structure search server provided by KEGG. It does so by allowing the computation of the maximum common subgraph for pairs of chemicals. Tool Service 9 exploits SIMCOMP and KEGG's DRUG database to allow the comparison of a chemical against more than 12,000 drugs currently contained in their database.

The user inputs the KEGG ID (can be translated through the same Tool Service from other chemical databases IDs e.g from PubChem CID) of the chemical and the number of similar drugs to be retrieved. The output is a .csv file containing the KEGG IDs of the similar drugs retrieved and the Tanimoto similarity coefficients for the phytochemical-drug pairs.

At this point, it is crucial to highlight that the inclusion criteria for phytochemical-drug pairs were based on the Tanimoto coefficient, to ensure a comprehensive analysis. Specifically, all pairs with a Tanimoto coefficient greater than 0.3 were documented. Notably, within this

dataset, certain pairs exhibited higher levels of structural resemblance, with Tanimoto coefficients surpassing 0.8, indicating a high similarity. Additionally, pairs demonstrating Tanimoto coefficient values ranging from 0.6 to 0.8 can be considered as exhibiting moderate similarity. On the other hand, pairs falling within the range of 0.3 to 0.6 display low similarity.

After applying Tool Service 9 for all 34 phytochemicals the results were classified into five major categories:

A. Structural similarity to UV protectants

Three phytochemicals resemble 4-aminobenzoic acid (ATC: D02BA01). These are: 4-hydroxybenzoic acid, protocatechuic acid and gallic acid (shown in Figure. 6.3.5 below).

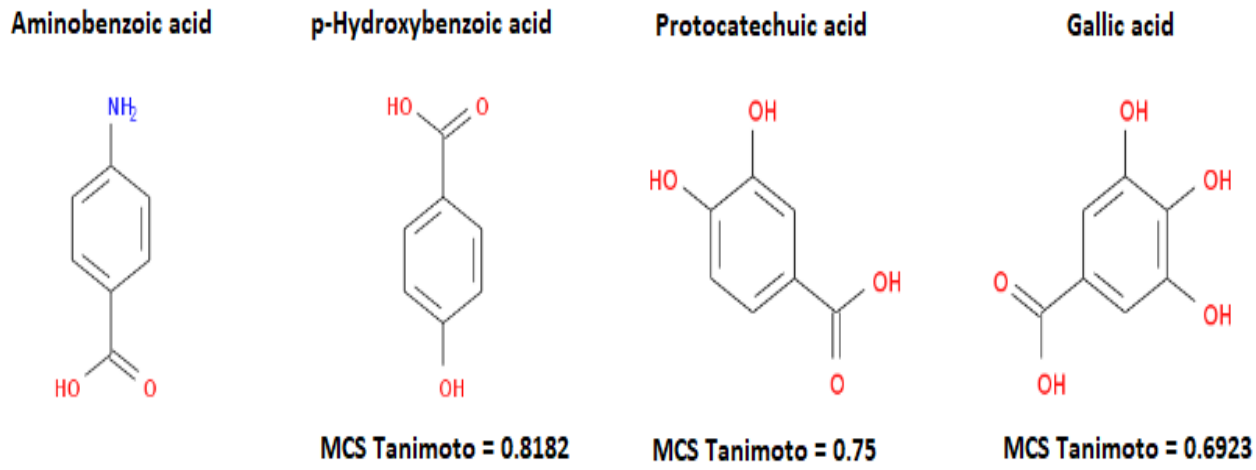


Figure. 6.3.5: The structural similarity of aminobenzoic acid to three phytochemicals chosen

Chlorogenic acid structurally resembles three chemicals considered to have a protective effect against UV radiation damage. These are in ascending order of Maximum Common Substructure (MCS) Tanimoto similarity: amiloxate, octinoxate (ATC: D02BA02) and homosalate (Figure. 6.3.6).

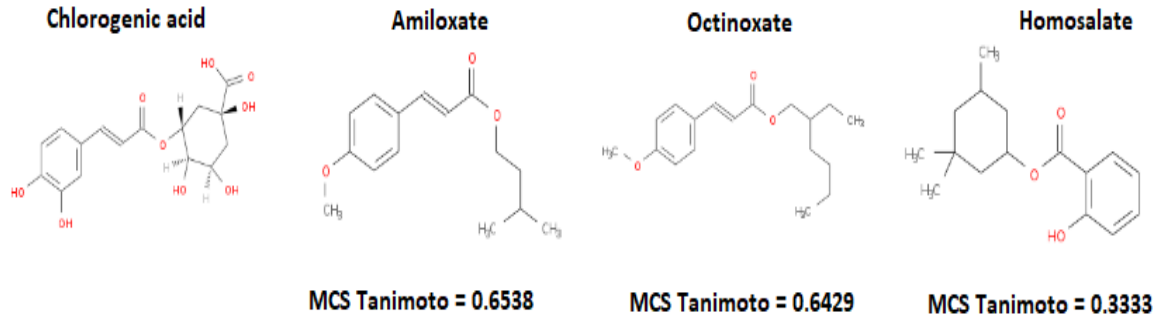


Figure. 6.3.6: The structural similarity of the phytochemical chlorogenic acid to three chemicals having a protective effect against UV radiation

Bornelone is a UV-absorber used in cosmetics. Its structure is quite similar to that of the phytochemical alpha-caryophyllene (Figure. 6.3.7).

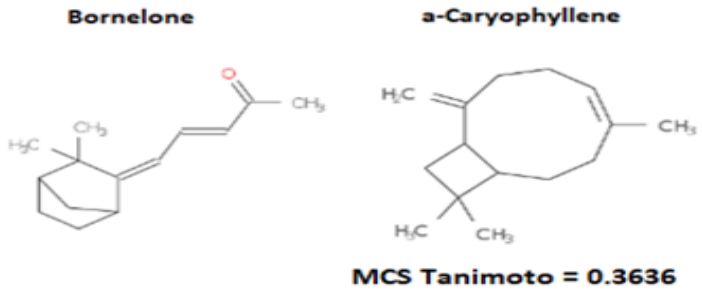


Figure. 6.3.7: The structural similarity of the phytochemical alpha-caryophyllene to the UV-absorber bornelone

From the results above it can easily be deduced that many of the phytochemicals which are present in ethnopharmacologically-used plants, resemble in structure various compounds used as UV protectants.

UV radiation is thought to have a role in hair cycle synchronisation, telogen effluvium and appears to adversely influence androgenetic alopecia (Piérard-Franchimont et al., 2002).

B. Structural similarity to dermatologicals of the ATC: D11AX group

Four phytochemicals resemble potassium aminobenzoate (ATC: D11AX23). These are: vanillic acid, p-hydroxybenzoic acid, gallic acid and protocatechuic acid (Figure. 6.3.8).

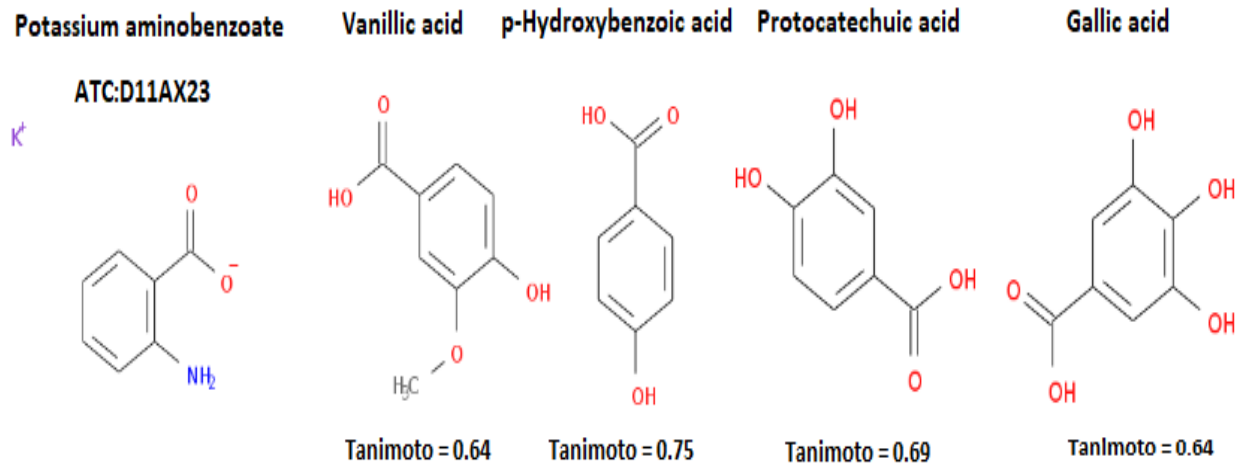


Figure. 6.3.8: The structural similarity of four phytochemicals to the dermatological potassium aminobenzoate (ATC: D11AX23).

Another three phytochemicals resemble gamolenic acid (ATC: D11AX02) in structure. These are: stearic acid, linoleic acid and oleic acid (Figure. 6.3.9).

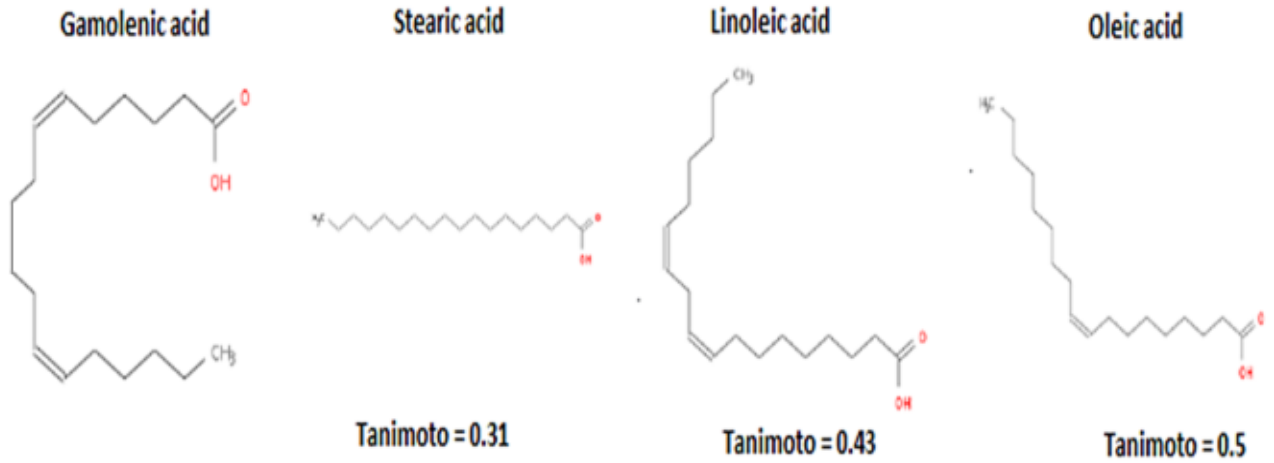


Figure. 6.3.9: The structural similarity of three phytochemicals to the dermatological gamolenic acid (ATC: D11AX02).

Finally, as shown in Figure. 6.3.10, beta-sitosterol resembles deoxycholic acid in structure (ATC: D11AX24)

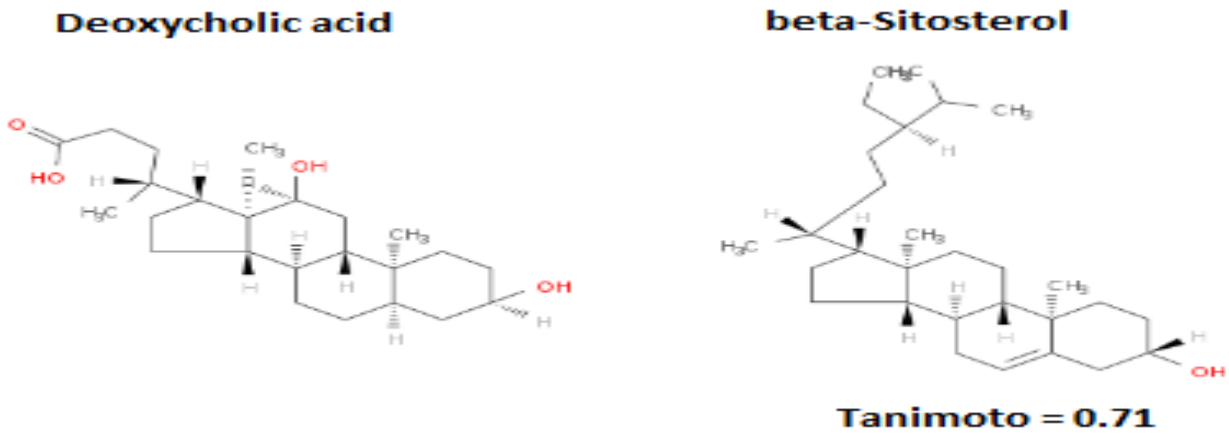


Figure. 6.3.10: The structural similarity of the phytochemical beta-sitosterol to the dermatological deoxycholic acid (ATC: D11AX24).

C. Phytochemicals structurally resembling vascular protectants

Rutin resembles the four vascular protectants: monoxerutin, diosmin, hesperidin and esculin (Figure. 6.3.11).

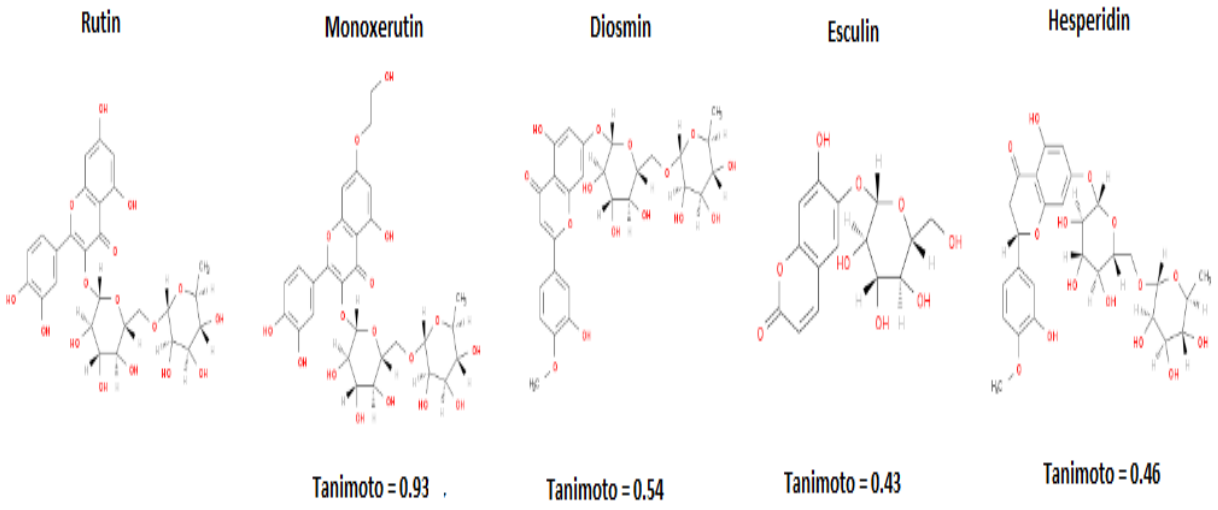


Figure. 6.3.11: The structural similarity of the phytochemical rutin to four vascular protectants.

On the other hand, quercetin, kaempferol and luteolin resemble two other vascular protectants: leucocianidol and flavodic acid (Figures 6.3.12 and 6.3.13).

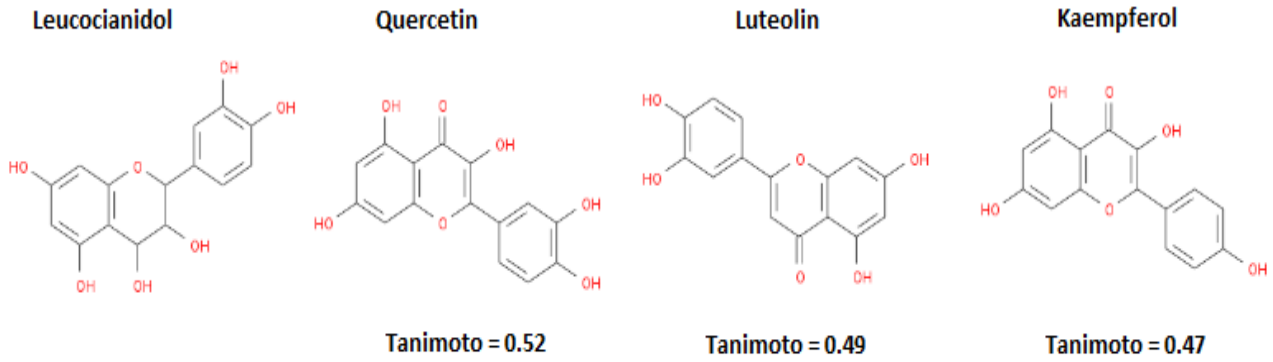


Figure. 6.3.12: The structural similarity of three phytochemicals to the vascular protectant leucocianidol

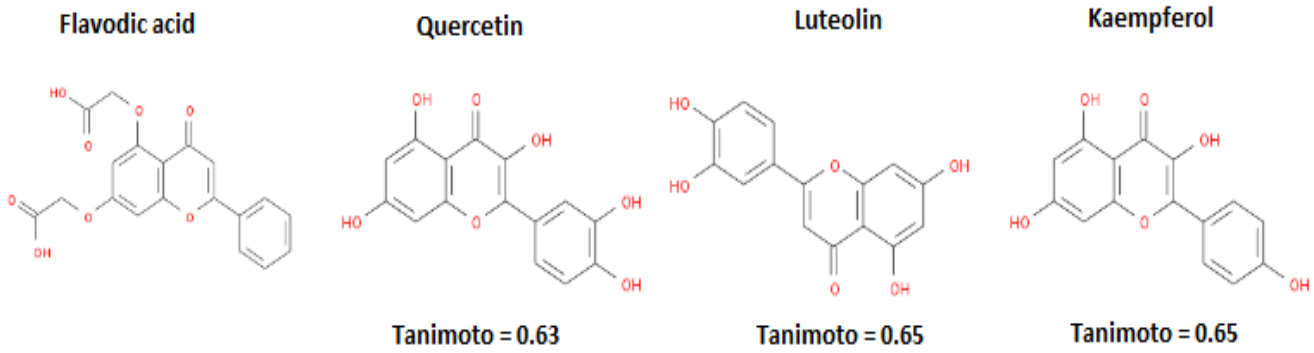


Figure. 6.3.13: The structural similarity of three phytochemicals to the vascular protectant leucocianidol

From the results above one may assume that there is a link of vascular protectants with flavonoid phytochemicals of plants used ethnopharmacologically as anti-alopecics in AGA.

Dermal microcirculation is essential for hair maintenance and insufficient blood supply can lead to hair follicle diseases (Bassino et al., 2020). VEGF release increases follicle and hair size, while a blockade of VEGF-mediated angiogenesis leads to impaired hair growth (Bassino et al., 2016). A review of the literature shows that both the phytochemical Quercetin and the

drug Hesperidin have the potential to induce VEGF release (Bassino et al., 2016; Oak et al., 2006).

D. Phytochemicals structurally resembling anti-inflammatory drugs

Oleanolic acid and ursolic acid are two phytochemicals which resemble two compounds with anti-inflammatory actions like bardoxolone and diammonium glycyrrhizinate (Figures 6.3.14 and 6.3.15).

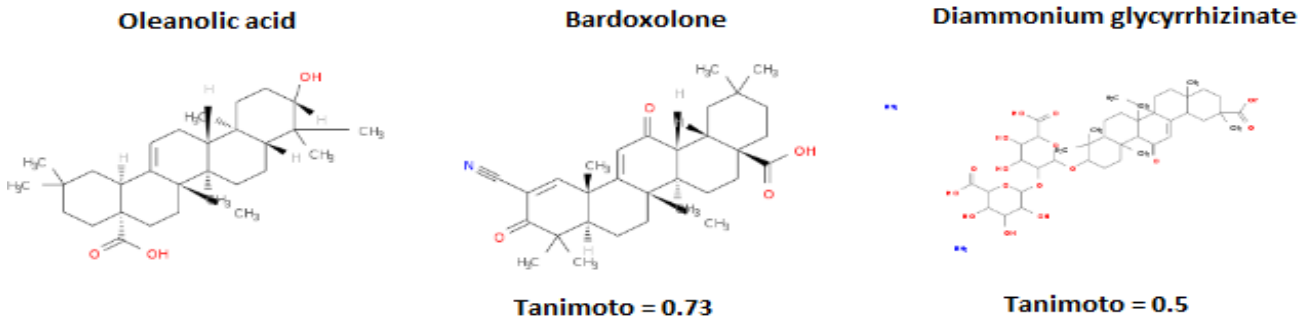


Figure. 6.3.14: The structural similarity of the phytochemical oleanolic acid to the anti-inflammatory compounds bardoxolone and diammonium glycyrrhizinate

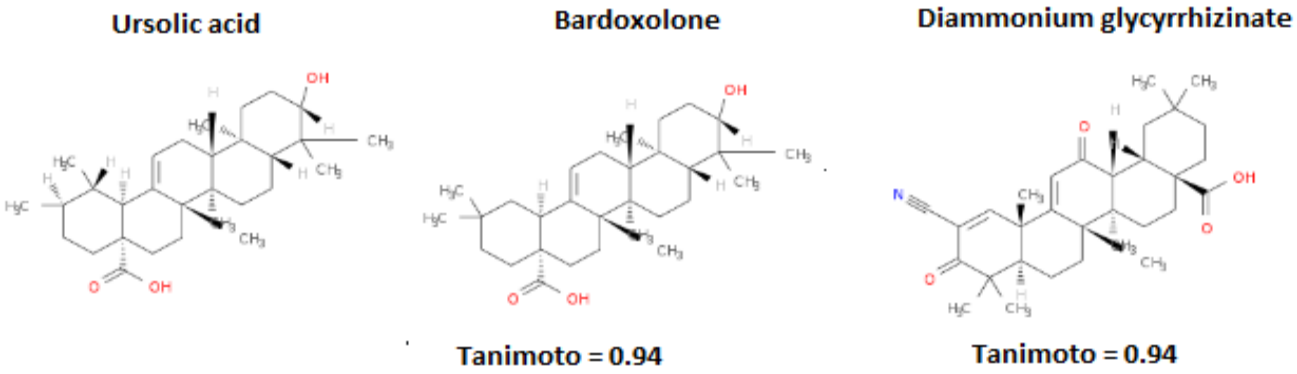


Figure. 6.3.15: The structural similarity of the phytochemical ursolic acid to the anti-inflammatory compounds bardoxolone and diammonium glycyrrhizinate

Beta-sitosterol structurally resembles the anti-inflammatory drug epihydrocholesterin and alpha-curcumene has a similar structure to fluretofen (Figures 6.3.16 and 6.3.17).

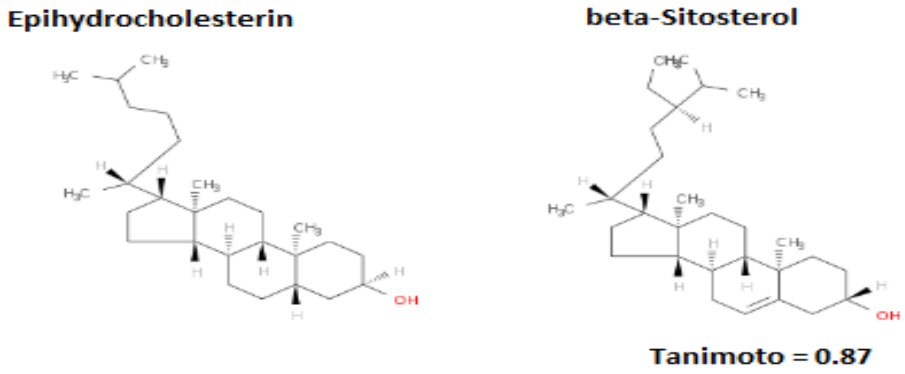


Figure. 6.3.16: The structural similarity of the phytochemical beta-sitosterol to the anti-inflammatory compound epidihydrocholesterin

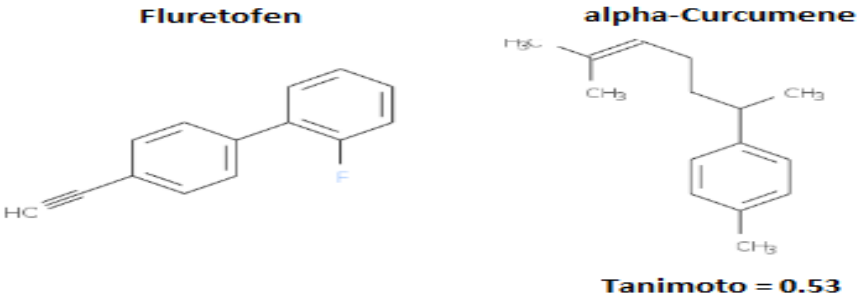


Figure. 6.3.17: The structural similarity of the phytochemical alpha-curcumene to the anti-inflammatory compound fluretofen

The implication of various activators of inflammation in the aetiology of androgenetic alopecia has been extensively researched in the literature (Jaworsky et al., 1992; Harmon and Nevins, 1993; Piérard, 1996). The similarity of several phytochemicals contained in the plants under investigation may suggest a role of anti-inflammatory agents in confronting the microinflammation observed in almost a third of androgenetic alopecia patients (Mahé, 2000). An assessment, using Tool Service 20, of the phytochemicals: ursolic acid, oleanolic acid, beta-sitosterol and the anti-inflammatory agents bardoxolone and diammonium glycyrrhizinate showed one common target. This target was prostaglandin-endoperoxide synthase 2 (PTGS2) or most commonly known as cyclooxygenase-2 (COX-2). Indeed, studies suggest that increased expression of COX-2 in the scalp of individuals with androgenetic alopecia leads to elevated levels of PGD2, which, in turn, can suppress hair follicle growth and induce premature catagen phase in hair follicles, resulting in hair thinning and eventual hair loss (Garza et al., 2012; Nieves and Garza, 2014).

E. Phytochemicals structurally resembling antineoplastic drugs

Phytochemical adenosine shows a structural resemblance to four antineoplastic drugs. These include the nucleoside analogues thioinosine, cladribine (ATC: L01BB04), clofarabine (L01BB06) and fludarabine (L01BB05) (Figure. 6.3.18).

Nucleoside analogues interfere with DNA and RNA synthesis, leading to inhibition of cell growth and proliferation. These drugs are commonly used in the treatment of hematological malignancies such as leukemia and lymphomas, where they target rapidly dividing cancer cells.

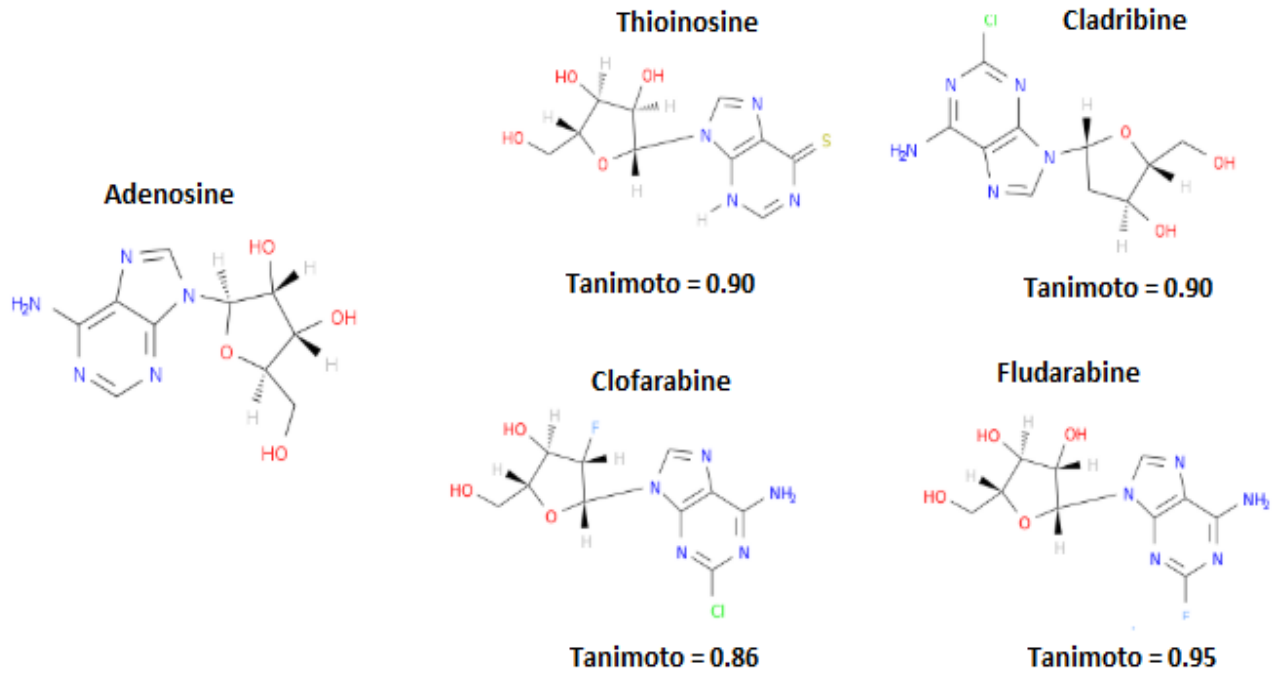


Figure. 6.3.18: The structural similarity of the phytochemical adenosine to four anti-neoplastic drugs

Three of our phytochemicals (quercetin, kaempferol and apigenin) structurally resemble two other antineoplastic drugs: genistein and idronoxil (shown in Figure. 6.3.19).

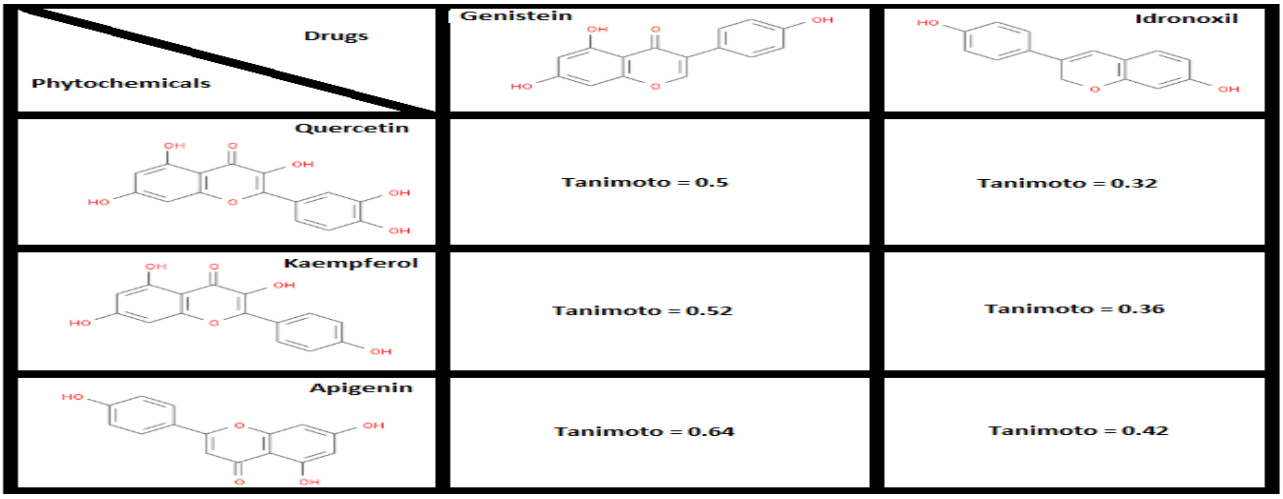


Figure. 6.3.19: The structural similarity of three phytochemicals to the anti-neoplastics genistein and idronoxil

p-Coumaric acid and ferulic acid resemble to the antineoplastic drug racemetyrosine (Figure. 6.3.20).

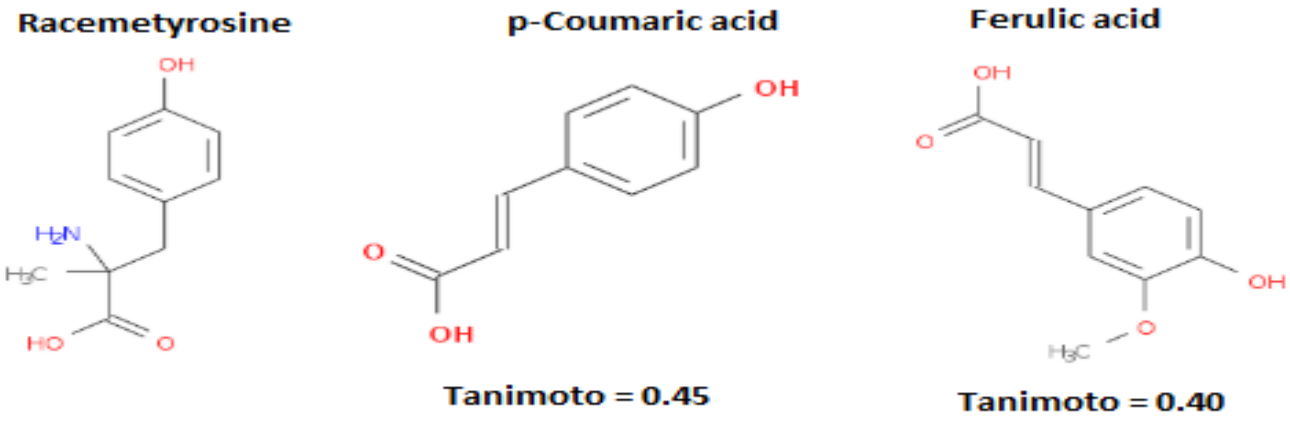


Figure. 6.3.20: The structural similarity of the two phytochemicals with the anti-neoplastic drug racemetyrosine

Ursolic acid and oleanolic acid are structurally similar to the antineoplastic drug bardoxolone (Figure. 6.3.21). Both the two phytochemicals and bardoxolone, exhibit potential similarities in their abilty to to induce apoptosis in cancer cells (Gai et al.,2016; Zhu, Huang and Wu, 2015; Wang et.al, 2015).

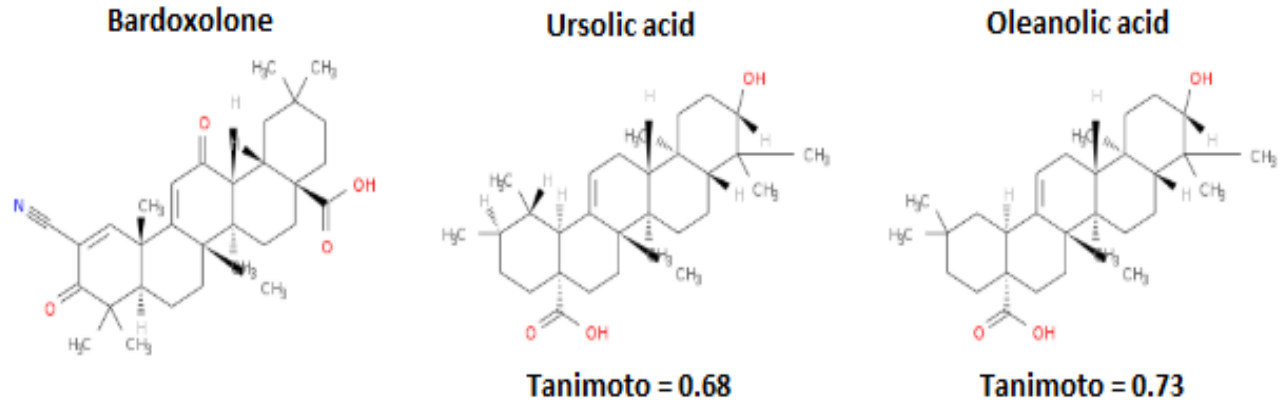


Figure. 6.3.21: The structural similarity of the two phytochemicals with the anti-neoplastic drug bardoxolone

Another phytochemical that resembles an antineoplastic drug is stearic acid and safingol (Figure. 6.3.22). Safingol, a sphingolipid analogue, has been investigated for its anticancer properties. It has shown promising results in inhibiting cancer cell growth and inducing apoptosis. Interestingly, stearic acid has also demonstrated activity in cancer cells, specifically in human breast cancer cells, where it preferentially induces apoptosis (Evans et al., 2009). Furthermore, studies have shown that safingol significantly inhibits aggressive mammary tumor growth (Acharya et al., 2019). While more research is needed to establish a direct6 connection between the structural similarity of stearic acid and safingol and their specific anticancer activities, these findings suggest that the structural features shared between these compounds may contribute to their ability to affect cancer cell behavior, including apoptosis induction and tumor growth inhibition.

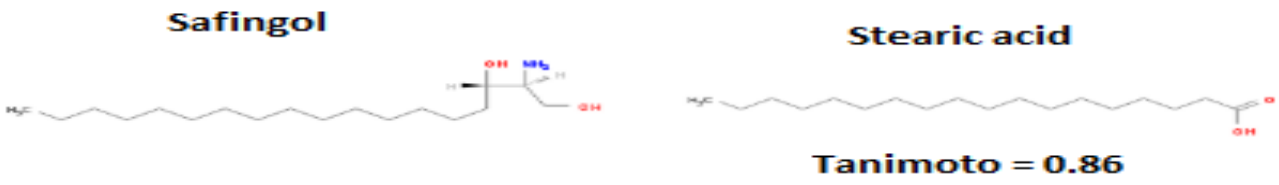


Figure. 6.3.22: The structural similarity of the phytochemical stearic acid with the anti-neoplastic drug safingol

Summary of the results retrieved with Tool Service 9

The application of Tool Service 9 (Chemical data) yielded valuable insights into the phytochemical composition of the 69 identified plants and their significance in relation to androgenetic alopecia. Out of the 2,157 phytochemicals identified in the plants, a set of 34 phytochemicals was selected for further analysis. Using the Tanimoto coefficient, pairs of phytochemicals with structural similarities above 0.85 were identified, indicating potential functional similarities. Clustering based on structural similarities revealed two major groups: one composed of phenolic acids and flavonoids, and the other predominantly consisting of fatty acids and terpenoids. These clusters provide meaningful groupings that aid in identifying phytochemicals contributing to the therapeutic properties of androgenetic alopecia.

The selected phytochemicals were also assessed for their drug-likeness using Lipinski's Rule of Five (RO5), revealing most of them to be suitable as orally bioavailable drug-like molecules. However, a few phytochemicals exhibited violations of certain RO5 parameters, suggesting challenges in oral absorption or permeation. SMARTS patterns were employed to identify specific substructures within the phytochemicals, revealing potential concerns related to drug-likeness or skin sensitisation for some compounds.

Using the Chemistry Development Kit (CDK), 2D and 3D molecular descriptors were calculated, providing quantitative representations of various chemical properties and aiding in the identification of compounds with desirable attributes for treating androgenetic alopecia.

Furthermore, SIMCOMP analysis compared the identified phytochemicals with over 12,000 drugs in the KEGG DRUG database, leading to several interesting findings. Some phytochemicals showed structural similarity to UV protectants, suggesting a potential link between protection against UV radiation damage and the treatment of androgenetic alopecia. The identification of phytochemicals similar to vascular protectants suggests a role in improving blood circulation to hair follicles. Additionally, the structural resemblance to anti-inflammatory and antineoplastic drugs raises intriguing possibilities for addressing inflammation and exploring potential therapeutic mechanisms.

In summary, the results obtained from Tool Service 9 and the SIMCOMP analysis provide comprehensive insights into the phytochemicals present in the identified plants and their potential therapeutic significance for androgenetic alopecia. These findings guide further research and the development of effective treatments based on the identified phytochemicals.

6.4 Results from the application of Tool Service 10 (Pharmacological data)

A literature search of currently used drugs in AGA shows that seven drugs are currently marketed as monotherapy for AGA. These are: biotin, minoxidil, finasteride, bimatoprost, dutasteride, carpronium chloride and alfatradiol (Figure 6.4).

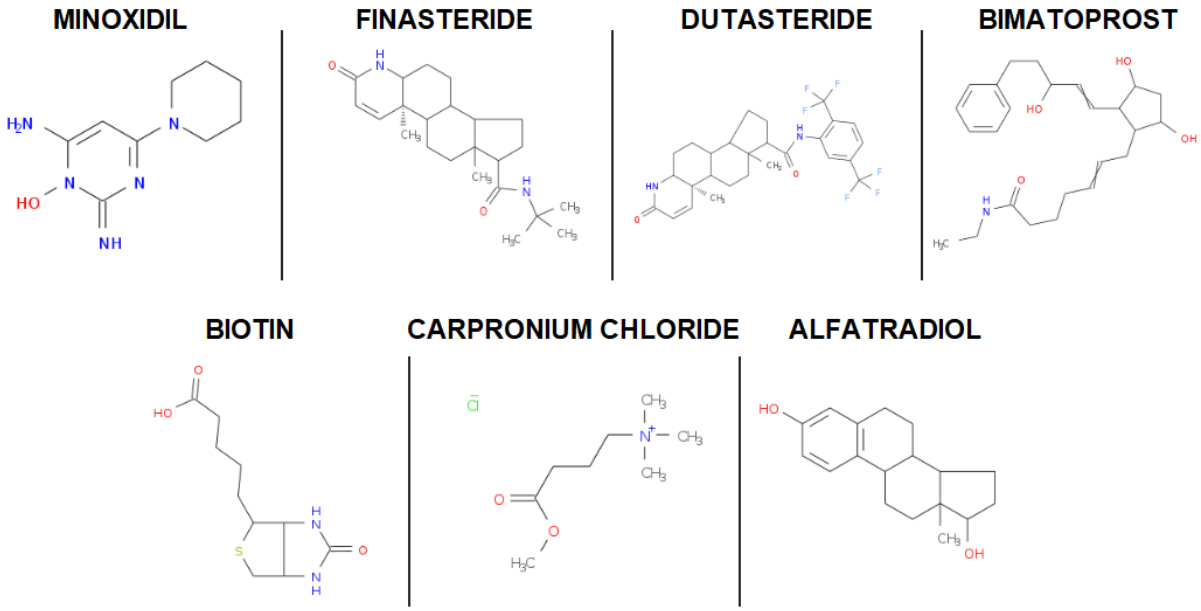
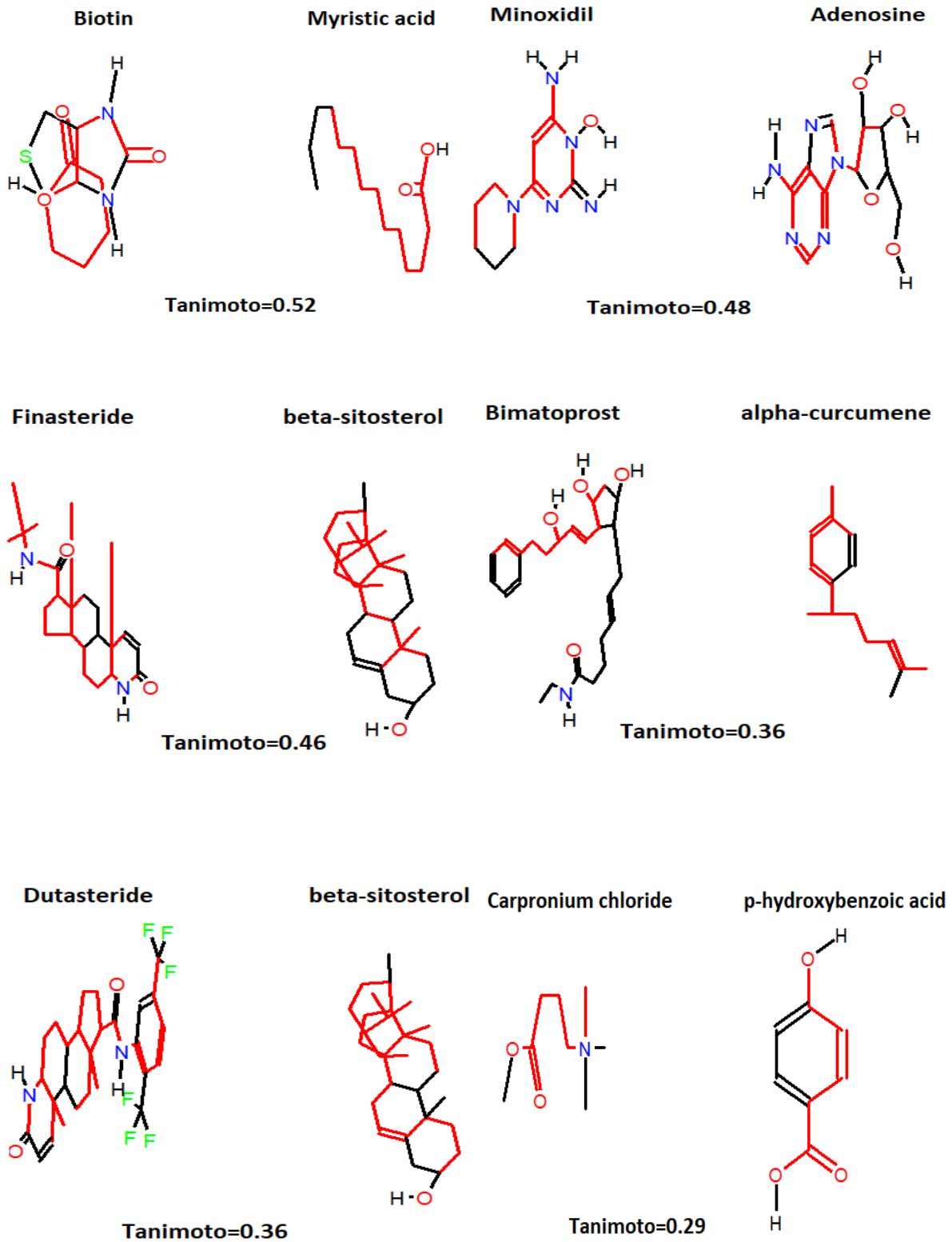


Figure 6.4: The chemical structures of drugs currently used as monotherapy for androgenetic alopecia

Structural comparison of currently used drugs in androgenetic alopecia with our phytochemicals

Using Tool Service 9 we structurally compared all 34 phytochemicals against the seven drugs mentioned above. This was done in order to investigate whether the effectiveness of the

plants (and their constituents) could be attributed to the structural similarity with drugs already marketed for this indication. Phytochemicals with the highest structural similarity to the seven currently used monotherapies are shown below in Figure 6.4.1.



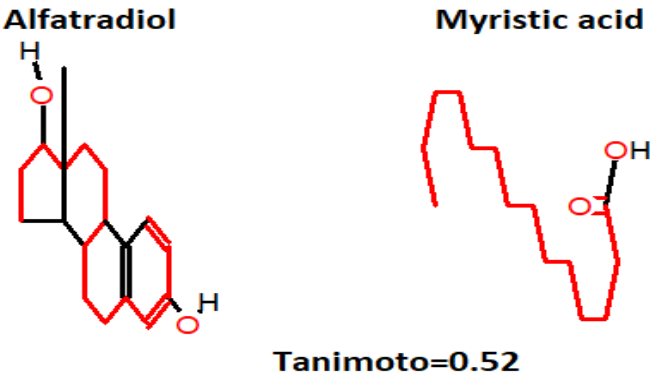


Figure. 6.4.1: Phytochemicals with the highest structural similarity to the seven currently used monotherapies in AGA

The pairs with the highest structural similarity (Tanimoto coefficient of 0.52) were those of the phytochemical myristic acid and the drugs biotin and alfatrdiol respectively. All results of the comparison of the 34 phytochemicals against the 7 drugs are available in Appendix 10.

Drug-Drug Interactions of currently used drugs in androgenetic alopecia.

The likelihood that a drug will experience clinically significant interactions with concurrent medications can determine whether it will be a commercial success or failure, and in the worst case scenario could even result in the product being pulled off the market. Therefore, providing drug-drug interactions data early in the process is a crucial component for successful drug development.

No drug-drug interactions were found for alfatrdiol, bimatoprost and biotin.

Carpronium chloride interacts with nine drugs. The interaction of carpronium chloride with four drugs (ethanol, erythromycin, adrenaline and cyclosporine) is due to their implication of CYP3A4.

Dutasteride interacts with 129 drugs. About two thirds of these interactions are due to the involvement of either CYP3A4 or CYP3A5.

Finasteride interacts with 28 drugs. 17 cases involve CYP3A4 and one case involves CYP2C9.

Minoxidil interacts with 85 drugs and all cases are unclassified. Full results of the drug-drug interactions involving the seven drugs currently used in AGA can be found in Appendix 10.

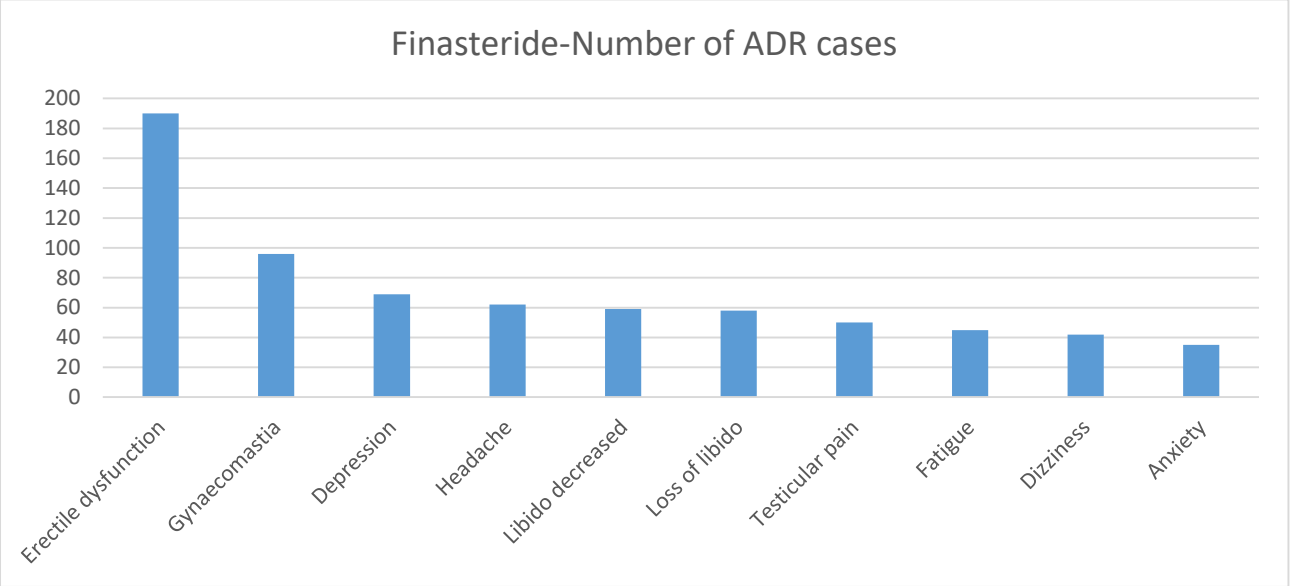
Adverse Drug Reactions of currently used anti-alopecia drugs

Tool Service 10 was used to retrieve cases of ADR for the seven drugs currently used in the treatment of AGA from the MHRA's Yellow Card Scheme. No results concerning adverse drug reactions were found for alfatradiol and carpronium chloride.

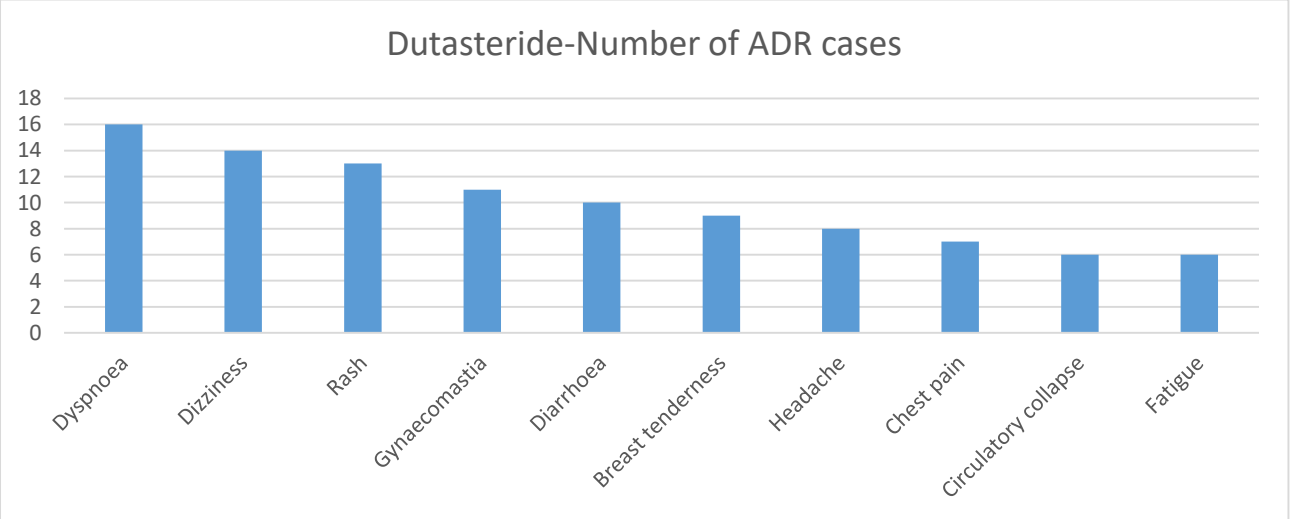
An analysis of all the cases currently reported in MHRA's Yellow Card Scheme, did reveal only one notable pattern amongst the currently used drugs as monotherapy in AGA. The common adverse drug reaction (ADR) of gynecomastia, the enlargement of male breast tissue, has been reported with the use of dutasteride and finasteride. The underlying mechanism for this ADR is believed to be related to the drugs' effects on the hormonal balance and the modulation of androgen signalling pathways. They both work by inhibiting the SRD5A enzyme responsible for converting testosterone into dihydrotestosterone (DHT), a more potent androgen. Gynecomastia is believed to occur due to an imbalance in the ratio of estrogens (female sex hormones) to androgens (male sex hormones) in the body. DHT plays a role in the regulation of estrogen levels, and its inhibition by dutasteride and finasteride can disrupt this balance. This disruption may lead to an increase in estrogen levels relative to androgens, potentially contributing to the development of gynecomastia.

While the precise mechanisms by which dutasteride and finasteride contribute to gynecomastia are not fully understood, it is believed that the modulation of androgen receptors and the alteration of estrogen and androgen ratios play a role. The interaction of these drugs with hormone receptors and their impact on the hormonal milieu may contribute to the development of breast tissue enlargement.

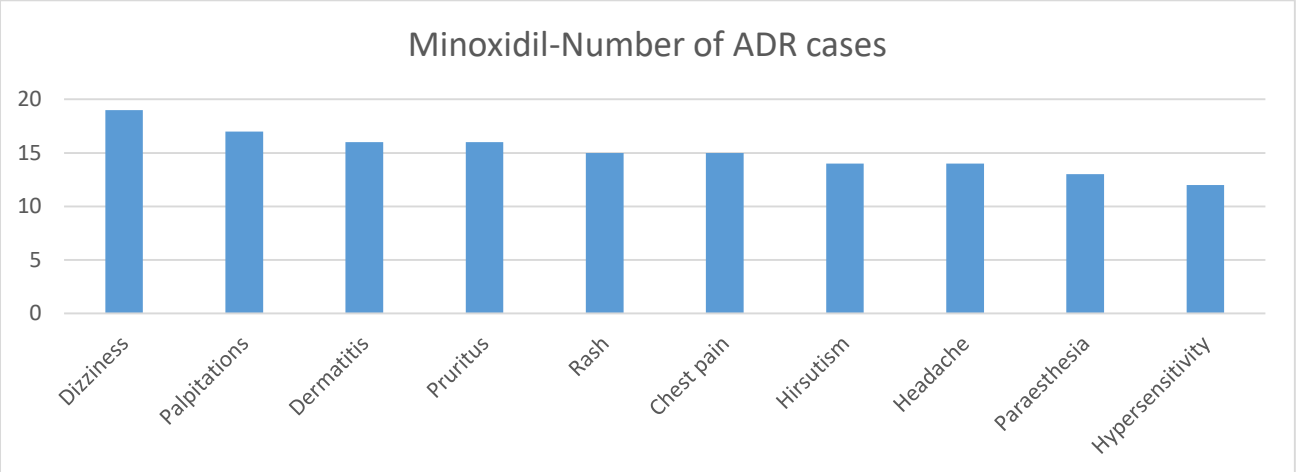
Finasteride



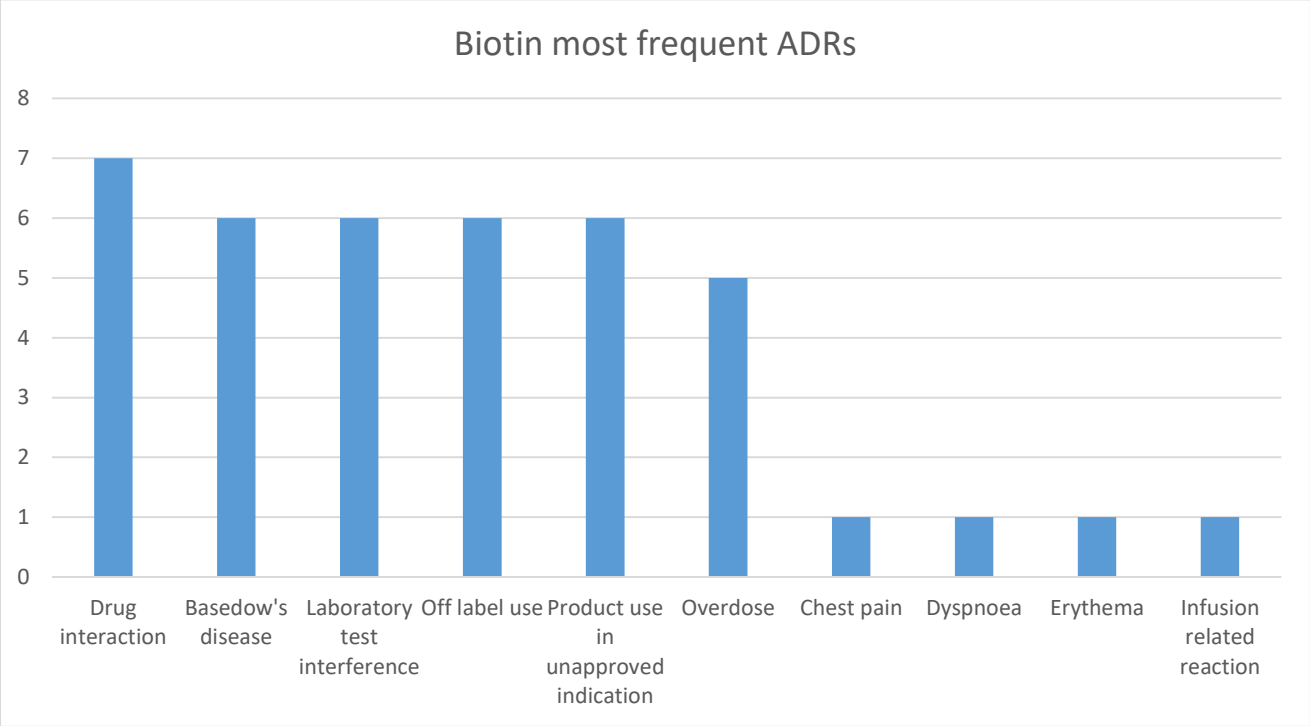
Dutasteride



Minoxidil



Biotin



Bimatoprost

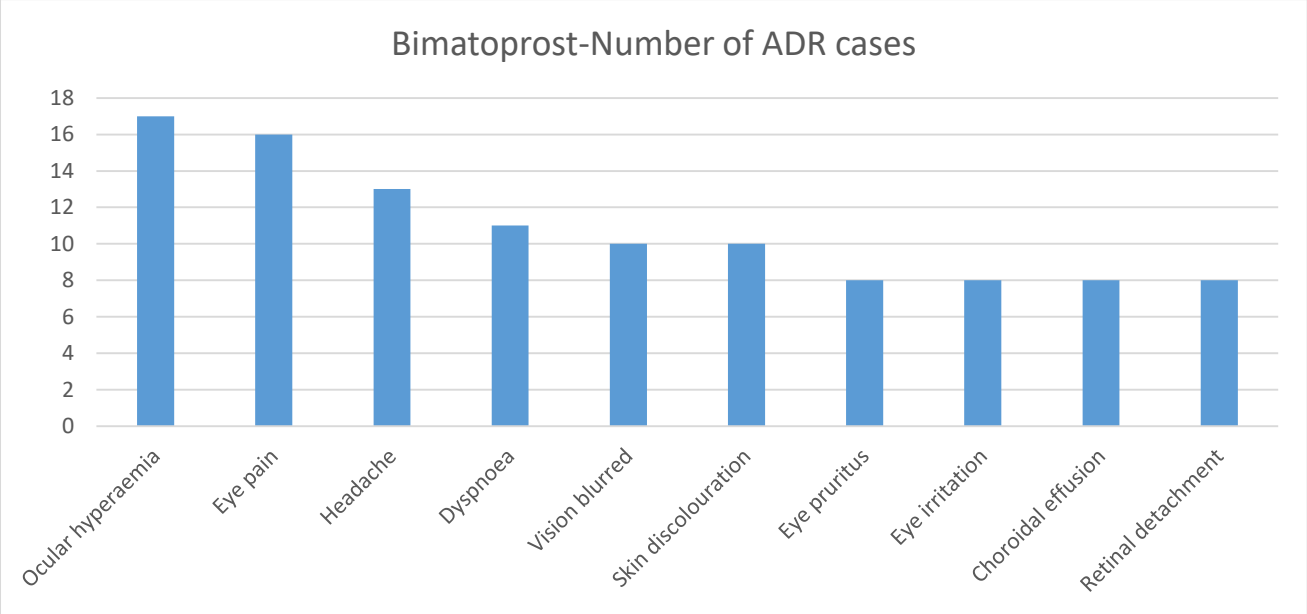


Figure 6.4.2: Adverse drug reactions caused by five of the currently used drugs in AGA

Alopecia as a side effect

'Both drug and side effects are often governed by interaction of the drug molecule with many, often unrelated, targets' (Schmidt et al., 2014). As a result, *in silico* off-target profiling has emerged as a key tool in the early stages of drug discovery for identifying potential off-target liabilities, which are more frequently safety relevant but can sometimes be proven to be advantageous. These methods make use of vast quantities of biochemical data from various sources to optimise unfavourable side effects (or favorable effects that can be exploited)) in pharmaceutical research.

Using the SIDER database, drugs which have alopecia as a side effect and their frequency of occurence were established.

The rate of side effects occurring is semi-quantified as follows:

>= 1/10 are classified as very common,

>=1/100 and < 1/10 are classified as common (or frequent)

>=1/1000 and < 1/100 are classified as uncommon (or infrequent)

>= 1/10,000 as rare.

The table below shows the results for the drugs commonly or very commonly causing alopecia as a side effect.

Table 6.4.1: Drugs commonly or very commonly causing alopecia as a side effect

STITCH CID	Drug Name	ATC code	Frequency of Side Effect
CID100005372	Tacrolimus	D11AH01, L04AD02	Common
CID100000699	Estrogen	G03C	Common
CID100054373	Octreotide	H01CB02	Common
CID100071616	Voriconazole	J02AC03	Common
CID100060787	Saquinavir	J05AE01	Common
CID100003877	Lamivudine	J05AF05	Common
CID145375808	Sofosbuvir	J05AP08, J05AP51, J05AP55, J05AP56	Common
CID100005394	Temozolomide	L01AX03	Common
CID100119182	Clofarabine	L01BB06	Common

CID100003461	Gemcitabine	L01BC05	Common
CID100001805	5-Azacytidine	L01BC07	Common
CID100005291	Imatinib	L01XE01	Common
CID100123631	Gefitinib	L01XE02	Common
CID106442177	Everolimus	L01XE10, L04AA18	Common
CID100644241	Nilotinib	L01XE08	Common
CID100003902	Letrozole	L02BG04	Common
CID100060198	Exemestane	L02BG06	Common
CID105479141	Mifamurtide	L03AX15	Common
CID100003899	Leflunomide	L04AA13	Common
CID100216326	Lenalidomide	L04AX04	Common
CID100034312	Oxcarbazepine	N03AF02	Common
CID100004158	Methylphenidate	N06BA04	Common
CID100060613	Cidofovir	J05AB12	Very Common
CID100005064	Ribavirin	J05AP01	Very Common
CID110324367	Boceprevir	J05AP03	Very Common
CID100004053	Melphalan	L01AA03	Very Common
CID100005394	Temozolomide	L01AX03	Very Common
CID100004666	Paclitaxel	L01CD01, L01CD03	Very Common
CID100004609	Oxaliplatin	L01XA03	Very Common
CID100208908	Lapatinib	L01XE07	Very Common
CID144462760	Dabrafenib	L01XE23	Very Common
CID100005515	Topotecan	L01XX17	Very Common
CID100001690	Doxorubicin	L01DB01	Very Common

As was expected the group with the most drugs having alopecia as a side effect was the antineoplastics (ATC: L01). Alopecia is very common or common in the most of these drugs. Another interesting fact is that six of them belong to the protein kinase inhibitors (ATC: L01XE) group.

The second largest group of drugs having alopecia as a common or very common side effect are the direct acting antivirals (ATC: J05A). Six drugs (saquinavir, lamivudine, sofosbuvir, cidofovir, ribavirin and boceprevir) belong to this category.

Using Tool Service 20 we assessed the genes related to the drugs causing alopecia very commonly. Nine out of ten drugs (data for boceprovir was not available) had caspase 3

(CASP3) in common and eight out of ten drugs had the BCL2 apoptosis regulator (BCL2) and the BCL2 associated X, apoptosis regulator (BAX) in common. Indeed, all three of them could be potential 'off-targets' as they are considered critical factors in hair loss (Zhu et al., 2018). Results are vailable in Appendix 10.

Alopecia as an ADR

Drug-induced alopecia is becoming a frequent phenomenon. The main drug classes which provoke hair loss are the: antineoplastics, retinoids, contraceptives, anticoagulant drugs, immunosuppresants and Selective Serotonin Reuptake Inhibitor (SSRI) antidepressants.

Table 6.4.2 below contains the first 10 drugs (in regard to their frequency of reported cases of alopecia induction in MHRA's Yellow Card Scheme) and their respective ATC codes. Results of recorded cases for 100 drugs most frequently causing alopecia as an ADR from the MHRA's Yellow Card Scheme can be found in Appendix 10. Table 6.4.2 below presents ten drugs with the highest frequency of such cases.

Table 6.4.2: 10 drugs with the most recorded cases of causing alopecia as an ADR in MHRA's Yellow Card Scheme

PubChem CID	Drug Name	Frequency of cases	ATC Coding
13109	levonorgestrel	240	G03AC03-G03AD01
54454	simvastatin	192	C10AA01
3121	valproic acid	191	N03AG01
4594	omeprazole	179	A02BC01
5991	ethinyl estradiol	176	G03CA01-L02AA03
3386	fluoxetine	147	N06AB03
6230	norethisterone	126	G03AC01-G03DC02
60823	atorvastatine	112	C10AA05

43815	paroxetine	90	N06AB05
2249	atenolol	88	C07AB03

An investigation of the drugs listed in Appendix 10 indicates that almost half of them can be further categorised in the following ten drug classes:

1. **Sex hormones and modulators of the genital system (ATC: G03)**

Eleven drugs (levonorgestrel, ethinyl estradiol, norethisterone, etonogestrel, estradiol, desogestrel, danazol, tibolone, drosperinone, norgestrel and ethynodiol) out of the hundred drugs belong to this drug class. The combined number of alopecia cases caused by these eleven drugs in MHRA was 879 (12% of cases).

2. **Antiepileptics (ATC: N03)**

Eight drugs (valproic acid, lamotrigine, carbamazepine, gabapentin, pregabalin, topirmate, levetiracetam and Vvgabatrin) out of the hundred drugs belong to the antiepileptics class. The combined number of alopecia cases caused by these four drugs in MHRA was 469 (6.4% of cases).

3. **Psychoanaleptics (ATC: N06)**

Eight drugs (fluoxetine, paroxetine, sertraline, citalopram, venlafaxine, methylphenidate, amitriptyline, mirtazapine) out of the hundred drugs belong to the psychoanaleptics class. The combined number of alopecia cases caused by these four drugs in MHRA was 477 (6.5% of cases).

4. **Non-steroidal anti-inflammatory and antirheumatic drugs (ATC: N06)**

Seven drugs (diclofenac, naproxen, celecoxib, rofecoxib, piroxicam, ibuprofen and penicillamine) out of the hundred drugs belong to this class. The combined number of alopecia cases caused by these seven drugs in MHRA was 228 (3.1% of cases).

5. **Antibacterials for systemic use (ATC: J01)**

Six drugs (pefloxacin mesylate, ciprofloxacin, trimethoprim, amoxicillin, minocycline and sulfamethoxazole) out of the hundred drugs belong to this class. The combined number of alopecia cases caused by these six drugs in MHRA was 157 (2.1% of cases).

6. **Lipid modifying agents, plain (ATC: C10)**

Five drugs (simvastatin, atorvastatin, bezafibrate, pravastatin and ezetimibe) out of the hundred drugs belong to this class. The combined number of alopecia cases caused by these five drugs in MHRA was 374 (5.1% of cases).

7. **Agents acting on the renin-angiotensin system (ATC: C09)**

Five drugs (ramipril, enalapril, lisinopril, captopril and losartan) out of the hundred drugs belong to this class. The combined number of alopecia cases caused by these five drugs in MHRA was 299 (4% of cases).

8. **Drugs for acid related disorders (ATC: A02)**

Four drugs (omeprazole, cimetidine, lansoprazole and ranitidine) out of the hundred drugs belong to this class. The combined number of alopecia cases caused by these four drugs in MHRA was 386 (5.2% of cases).

9. **Beta-blocking agents (ATC: C07)**

Four drugs (atenolol, propanolol, bisprolol and timolol) out of the hundred drugs belong to this class. The combined number of alopecia cases caused by these four drugs in MHRA was 180 (2.4% of cases).

10. **Antipsychotics (ATC: N05)**

Four drugs (lithium, aripiprazole, risperidone and quetiapine) out of the hundred drugs belong to this class. The combined number of alopecia cases caused by these four drugs in MHRA was 82 (1.1% of cases).

A review of the literature on the drugs reported in the ten classes immediately above shows that most of these drugs have already data regarding their potential to induce hair loss although this is not always communicated to patients (through patient information leaflets) and healthcare providers.

Table 6.4.3 below shows a list of the drugs along with references of their alopecia inducing potential.

Table 6.4.3: Drugs which have a potential to induce alopecia according to the literature and provoke many ADR cases of alopecia.

Drug Name	Reference of drug-induced alopecia
Amitriptyline	Warnock et al. (1991)
Aripiprazole	Bilgin et al. (2015)
Atenolol	Trüeb (2008)
Atorvastatine	Segal (2002)
Bezafibrate	Trüeb (2008)
Captopril	Tosti and Pazzaglia (2007)
Carbamazepine	Shuper et al. (1985)
Cimetidine	Tosti et al. (1994); Trüeb (2008)
Citalopram	Gupta and Masand (2000)
Danazol	Trüeb (2008)
Diclofenac	Jain, Joshi and Dass (2012)
Enalapril	Tosti and Pazzaglia (2007)
Ethynodiol	Carruthers (1968)
Etonogestrel	Lee and Ewer (2006)
Fluoxetine	Gupta and Major (1991)
Gabapentin	Picard et al. (1997)
Ibuprofen	Meyer (1979)
Lamotrigine	Patrizi et al. (2005)
Levetiracetam	Zou, Hong and Zhou (2014)
Levonorgestrel	Paterson et al. (2007)
Lisinopril	Kataria et al. (2017)
Lithium	Orwin (1983); Yassa and Ananth (1983); Muniz, Salem and Director (1982)
Methylphenidate	Ardic and Ercan (2017); Sayin and Turkoglu (2016)
Mirtazapine	Osman and McCauley (2016)
Naproxen	Barter (1989)
Norethisterone	Trüeb (2008)
Norgestrel	Sinclair et al. (2011)
Omeprazole	Borum and Cannava (1997)
Paroxetine	Zalsman, Sever and Munitz (1999)
Penicillamine	Jain, Joshi and Dass (2012)
Piroxicam	D'Arcy and Gerber (1987)
Propanolol	Tosti and Pazzaglia (2007)
Quetiapine	McLean and Harrison-Woolrych (2007) ; Yazici and Percinel (2015)
Ramipril	Tosti and Pazzaglia (2007)
Ranitidine	Trüeb (2008)
Risperidone	Karabulut and Cakir (2016)
Sertraline	Bourgeois (1996)
Tibolone	Trüeb (2008)
Timolol	Tosti and Pazzaglia (2007)
Topiramate	Chuang et al. (2002); Chen et al. (2015)
Valproic acid	Yilmaz, Tasdemir and Paksu (2009)
Venlafaxine	Pitchot and Ansseau (2001)
Vigabatrin	Lampl et al. (1996)

Using Tool Service 20 we assessed the genes related to the drugs causing alopecia as an ADR more frequently than others (ATC: G03). Eight out of nine drugs (data for drosperinone and ethynidiol was not available) had progesterone receptor (PGR) in common, seven out of nine drugs had the estrogen receptor (ESR1) in common and six out of nine drugs had the androgen (AR) in common. These three genes could be potential 'off-targets' as the complex process of steroid metabolism in hair follicles is likely to involve interactions between steroid hormones and their receptors as well as a relationship between the signalling pathways for oestrogen, progesterone, and androgen receptors. (Yip, Rufaut and Sinclair, 2011). Results are available in Appendix 10.

Summary of the results retrieved Tool Service 10 (Pharmacological data)

In order to gain insights into the efficacy of the phytochemicals, we conducted comprehensive studies to assess the structural similarity between the seven currently used monotherapies for androgenetic alopecia (AGA) and the selected phytochemicals. Using Tool Service 9, which provides chemical data, we performed a thorough analysis. Unfortunately, the results did not indicate any significant structural similarity between the phytochemicals and the AGA drugs, except for myristic acid, which showed the highest similarity (Tanimoto coefficient=0.52) with biotin and alfatradiol.

To ensure successful drug development, it is crucial to evaluate potential drug-drug interactions. With this in mind, we carefully examined the seven AGA drugs to assess the likelihood of interactions. Among the notable findings, it was observed that drug interactions involving carpronium chloride, dutasteride, and finasteride were primarily associated with the involvement of the enzyme CYP3A4.

Furthermore, an analysis of adverse drug reactions (ADRs) cases for the seven drugs, utilising data from the MHRA's Yellow Card Scheme, revealed a notable pattern associated with dutasteride and finasteride—gynecomastia, the enlargement of male breast tissue. These drugs exert their effects by inhibiting the SRD5A enzyme responsible for converting testosterone into dihydrotestosterone (DHT), a potent androgen. The disturbance in estrogen

to androgen levels caused by the inhibition of DHT is believed to contribute to the development of gynecomastia.

In addition to examining the specific AGA drugs, analysing drugs causing alopecia as a side effect and as an ADR provided valuable insights into the significance of androgenetic alopecia in relation to these medications. This analysis suggests potential ‘off- targets’ and mechanisms involved in drug-induced hair loss, providing valuable knowledge for drug development. By understanding the underlying mechanisms and employing strategies to prevent or mitigate drug-induced hair loss, patient care can be significantly improved.

6.5 Results from the application of Tool Services 11-20

The goal of target identification is to increase the likelihood of discovering a new drug. At this point, researchers can concentrate on particular targets that may be involved in the pathogenesis of the disease before screening their potential candidates. In this thesis, to assist the decision on the targets to be studied we took into account data from:

- Genome Wide Association Studies (using data retrieved with Tool Service 3)
- text-mining of relevant articles (using data retrieved with Tool Service 2)
- GO Terms relevant to hair processes (using data retrieved with Tool Services 15 and 20)
- CTD (using data retrieved with Tool Service 20)

The selection of these genes was based on specific criteria. Firstly, we prioritized genes with the lowest p-values from Genome-Wide Association Studies. Additionally, we considered genes that were frequently co-mentioned with the term 'androgenetic alopecia' in PubMed text-mining. Furthermore, we included genes associated with relevant Gene Ontology (GO) Terms related to hair processes. Lastly, we incorporated genes from the Chemical Toxicogenomics Database (CTD) that are recognized as biomarkers or contribute to the disease's aetiology. By employing these criteria, we aimed to ensure the inclusion of genes with the highest relevance and significance in our investigation. These targets are shown in Table 6.5.1.

Table 6.5.1: The 48 AGA targets chosen to be studied in this thesis

	Gene symbol	Protein name		Gene symbol	Protein name
1	**AKT1**	RAC-alpha serine/threonine-protein kinase	25	**LIPH**	Lipase member H
2	**ALX4**	Homeobox protein aristaless-like 4	26	**LPAR6**	Lysophosphatidic acid receptor 6
3	**AR**	Androgen receptor	27	**LSS**	Lanosterol synthase
4	**BCL2**	Apoptosis regulator Bcl-2	28	**MAPT**	Microtubule-associated protein tau
5	**BRD4**	Bromodomain-containing protein 4	29	**MTHFR**	Methylenetetrahydrofolate reductase
6	**CCND1**	G1/S-specific cyclin-D1	30	**NFKB1**	Nuclear factor NF-kappa-B p105 subunit
7	**CRH**	Corticoliberin	31	**NOTCH1**	Neurogenic locus notch homolog protein 1
8	**CTNNB1**	Catenin beta-1	32	**PTCH1**	Protein patched homolog 1
9	**DKK1**	Dickkopf-related protein 1	33	**PTGS2**	Prostaglandin G/H synthase 2
10	**DSG4**	Desmoglein-4	34	**RELA**	Transcription factor p65
11	**EBF1**	Transcription factor COE1	35	**RUNX1**	Runt-related transcription factor 1
12	**EDAR**	Tumor necrosis factor receptor superfamily member EDAR	36	**SHBG**	Sex hormone-binding globulin
13	**FGF5**	Fibroblast growth factor 5	37	**SHH**	Sonic hedgehog protein
14	**FGF7**	Fibroblast growth factor 7	38	**SMAD4**	Mothers against decapentaplegic homolog 4
15	**HDAC9**	Histone deacetylase 9	39	**SNAI1**	Zinc finger protein SNAI1
16	**HR**	Lysine-specific demethylase hairless	40	**SOX9**	Transcription factor SOX-9
17	**IGF1**	Insulin-like growth factor I	41	**SRD5A2**	3-oxo-5-alpha-steroid4-dehydrogenase 2
18	**INS**	Insulin	42	**TGFB2**	Transforming growth factor beta-2 proprotein
19	**IRF4**	Interferon regulatory factor 4	43	**TNF**	Tumor necrosis factor
20	**KLK3**	Prostate-specific antigen	44	**TP63**	Tumor protein 63
21	**KRT74**	Keratin, type II cytoskeletal 74	45	**TWIST2**	Twist-related protein 2
22	**KRT75**	Keratin, type II cytoskeletal 75	46	**VDR**	Vitamin D3 receptor
23	**KRT86**	Keratin, type II cuticular Hb6	47	**WNT10A**	Protein Wnt-10a
24	**LGR4**	Leucine-rich repeat-containing G-protein coupled receptor 4	48	**WNT10B**	Protein Wnt-10b

6.5.1 Results from the application of Tool Service 11 (Target information)

The 48 targets were initially assessed for information using Tool Service 11. The resulting .csv files contains KEGG info on these targets. Table 6.5.2 shows the information retrieved with Tool Service 11 for gene AKT1. Results for all 48 targets can be found in Appendix 11.

Table 6.5.2: Target information generated from the use of Tool Service 11 for gene AKT1

ENTRY	207
NAME	(RefSeq) AKT serine/threonine kinase 1
SYMBOL	AKT1, AKT, PKB, PKB-ALPHA, PRKBA, RAC, RAC-ALPHA
ORGANISM	Homo sapiens (human)
X1	NCBI-GeneID: 207
X2	NCBI-ProteinID: NP_001014431
X3	OMIM: 164730
X4	HGNC: 391
X5	Ensembl: ENSG00000142208
X6	Vega: OTTHUMG00000170795
X7	Pharos: P31749(Tchem)
X8	UniProt: P31749 B0LPE5
X9	PDB: 1H10 1UNP 1UNQ 1UNR 2UVM 2UZR 2UZS 3CQU 3CQW 3MV5 3MVH 3O96 3OCB 3OW4 3QKK 3QKL 3QKM 4EJN 4EKK 4EKL 4GV1 5KCV 6BUU 6CCY 6HHF 6HHG 6HHH 6HHI 6HHJ 6NPZ 6S9W 6S9X
AASeq	MSDVAIVKEGWLHKRGEYIKTWRPRYFLLKNDGTFIGYKERPQDVDQREAPLNNFSVAQCQLMKTERPRP NTFIIRCLQWTTVIERTFHVETPEEREEWTTAIQTVADGLKKQEEEEMDFRSGSPSDNSGAEEMEVSLAKPK HRVTMNEFEYLKLLGKGTFGKVILVKEKATGRYYAMKILKKEVIVAKDEVAHTLTENRVLQNSRHPFLTALKY SFQTHDRLCFVMEYANGGELFFHLSRERVFSEDRARFYGAEIVSALDYLHSEKNVVYRDLKLENLMLDKDGH IKITDFGLCKEGIKDGATMKTFCGTPEYLAPEVLEDNDYGRAVDWWGLGVVMYEMMCGRLPFYNQDHEK LFELILMEEIRFPRTLGPEAKSLLSGLLKKDPKQRLGGGSEDAKEIMQHRFFAGIVWQHVYEKKLSPPFKPQVT SETDTRYFDEEFTAQMITITPPDQDDSMECVDSERRPHFPQFSYSASGTA
NTSeq	ATGAGCGACGTGGCTATTGTGAAGGAGGGTTGGCTGCACAAACGAGGGGAGTACATCAAGACCTGGC GGCCACGCTACTTCCTCCTCAAGAATGATGGCACCTTCATTGGCTACAAGGAGCGGCCGCAGGATGTG GACCAACGTGAGGCTCCCCTCAACAACTTCTCTGTGGCGCAGTGCCAGCTGATGAAGACGGAGCGGCC CCGGCCCAACACCTTCATCATCCGCTGCCTGCAGTGGACCACTGTCATCGAACGCACCTTCCATGTGGA GACTCCTGAGGAGCGGGAGGAGTGGACAACCGCCATCCAGACTGTGGCTGACGGCCTCAAGAAGCAG GAGGAGGAGGAGATGGACTTCCGGTCGGGCTCACCCAGTGACAACTCAGGGGCTGAAGAGATGGAG GTGTCCCTGGCCAAGCCCAAGCACCGCGTGACCATGAACGAGTTTGAGTACCTGAAGCTGCTGGGCAA GGGCACTTTCGGCAAGGTGATCCTGGTGAAGGAGAAGGCCACAGGCCGCTACTACGCCATGAAGATC CTCAAGAAGGAAGTCATCGTGGCCAAGGACGAGGTGGCCCACACACTCACCGAGAACCGCGTCCTGCA GAACTCCAGGCACCCCTTCCTCACAGCCCTGAAGTACTCTTTCCAGACCCACGACCGCCTCTGCTTTGTC ATGGAGTACGCCAACGGGGGCGAGCTGTTCTTCCACCTGTCCCGGGAGCGTGTGTTCTCCGAGGACCG GGCCCGCTTCTATGGCGCTGAGATTGTGTCAGCCCTGGACTACCTGCACTCGGAGAAGAACGTGGTGT ACCGGGACCTCAAGCTGGAGAACCTCATGCTGGACAAGGACGGGCACATTAAGATCACAGACTTCGGG CTGTGCAAGGAGGGGATCAAGGACGGTGCCACCATGAAGACCTTTTGCGGCACACCTGAGTACCTGGC CCCCGAGGTGCTGGAGGACAATGACTACGGCCGTGCAGTGGACTGGTGGGGGCTGGGCGTGGTCATG TACGAGATGATGTGCGGTCGCCTGCCCTTCTACAACCAGGACCATGAGAAGCTTTTTGAGCTCATCCTC ATGGAGGAGATCCGCTTCCCGCGCACGCTTGGTCCCGAGGCCAAGTCCTTGCTTTCAGGGCTGCTCAAG AAGGACCCCAAGCAGAGGCTTGGCGGGGGCTCCGAGGACGCCAAGGAGATCATGCAGCATCGCTTCT TTGCCGGTATCGTGTGGCAGCACGTGTACGAGAAGAAGCTCAGCCCACCCTTCAAGCCCCAGGTCACG TCGGAGACTGACACCAGGTATTTTGATGAGGAGTTCACGGCCCAGATGATCACCATCACACCACCTGAC CAAGATGACAGCATGGAGTGTGTGGACAGCGAGCGCAGGCCCCACTTCCCCCAGTTCTCCTACTCGGC CAGCGGCACGGCCTGA

6.5.2 Results from the application of Tool Service 12 (Target ID translation)

Using Tool Service 12, IDs of all 48 targets were translated to obtain KEGG, EGGNOG, ENSEMBL and PDB IDs to be used in the next steps (Appendix 12). Table 6.5.3 shows the data retrieved for targets AKT1 and ALX4.

Table 6.5.3: An example of two targets (AKT1 and ALX4) translated into UniProt, KEGG, EGGNOG, ENSEMBL, PDB IDs and Gene Name

UniProt ID	KEGG ID	EGGNOG ID	ENSEMBL ID	PDB ID	Gene Name
P31749	hsa:207	KOG0690	ENSG00000142208	1H10, 1UNP, 1UNQ, 1UNR, 2UVM, 2UZR, 2UZS, 3CQU, 3CQW, 3MV5, 3MVH, 3O96, 3OCB, 3OW4, 3QKK, 3QKL, 3QKM, 4EJN, 4EKK, 4EKL, 4GV1, 5KCV, 6BUU, 6CCY, 6HHF, 6HHG, 6HHH, 6HHI, 6HHJ, 6NPZ, 6S9W, 6S9X	AKT1
Q9H161	hsa:60529	KOG0490	ENSG00000052850	2M0C	ALX4

6.5.3 Results from the application of Tool Service 13 (Binding DB affinity data)

For each of the 48 targets mentioned above (see Table 6.5.1) a .csv file with data from BindingDB was retrieved using Tool Service 13 (see Appendix 13), regarding the active chemicals on these targets and their respective affinity for these targets. For example, for target CTNNB1 (UniProt ID: P35222) the chemicals with proven affinity according to BindingDB are listed in Table 6.5.4.

Table 6.5.4: Chemicals with binding affinity for CTNNB1, from BindingDB

BindingDB ID	Pubchem CID	SMILES	Affinity Type	Affinity (nM)
BDBM50436054	73347465	CC[C@H](C)[C@@H]1NC(=O)[C@](C)(CCC\C=C/CCC[C@](C)(NC(=O)[C@H](CC(O)=O)NC(=O)[C@H](CC(C)C)NC1=O)C(=O)N[C@@H](Cc1cnc[nH]1)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CCCN=C(N)N)C(=O)N[C@@H](CCCN=C(N)N)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@@H](CCCN=C(N)N)C(N)=O)NC(=O)[C@H](CCCN=C(N)N)NC(=O)[C@@H]1CCCN1C(=O)[C@H](Cc1c[nH]c2ccccc12)NC(=O)[C@H](CCCN=C(N)N)NC(=O)[C@H](CCCN=C(N)N)NC(C)=O	Kd	53
BDBM50436055	91899923	CC[C@H](C)[C@@H]1NC(=O)[C@](C)(CCC\C=C/CCC[C@](C)(NC(=O)[C@H](CC(O)=O)NC(=O)[C@H](CC(C)C)NC1=O)C(=O)N[C@@H](Cc1cnc[nH]1)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CCCN=C(N)N)C(=O)N[C@@H](CCCN=C(N)N)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@@H](CCCN=C(N)N)C(N)=O)NC(=O)[C@H](CCC(N)=O)NC(=O)[C@@H]1CCCN1C(=O)[C@H](Cc1c[nH]c2ccccc12)NC(=O)[C@H](CCCN=C(N)N)NC(=O)[C@H](CCCN=C(N)N)NC(C)=O	Kd	13
BDBM50436056	91899924	CC[C@H](C)[C@@H]1NC(=O)[C@](C)(CCC\C=C/CCC[C@](C)(NC(=O)[C@H](CC(O)=O)NC(=O)[C@H](CC(C)C)NC1=O)C(=O)N[C@@H](Cc1cnc[nH]1)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CCC(N)=O)C(=O)N[C@@H](CCCN=C(N)N)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CCSC)C(=O)N[C@@H](CCCCN)C(O)=O)NC(=O)[C@H](CCC(O)=O)NC(=O)[C@@H]1CCCN1C(=O)[C@H](CC(N)=O)NC(=O)[C@H](CCC(O)=O)NC(=O)[C@@H](N)CCC(O)=O	Kd	60

.

6.5.4 Results from the application of Tool Service 14 (EGGNOG's GO Terms)

EGGNOG GO Terms was initially used to reveal targets related to hair processes and when the choice of the 48 targets (see Table 6.5.1) was made, it was used again to assess the link of AGA-related signalling pathways with these 48 targets. The results of the latter are shown in Table 6.5.5. All results are available in Appendix 14.

Table.6.5.5: GO Terms relating a number of the 48 genes to important biochemical pathways in AGA

GO Term	Process	GENES
WNT		
GO:0016055	Wnt receptor signalling pathway	ALX4|DKK1|DSG4|FG5|INS|NFKB1|SHH|SNAI1|VDR|WNT10A|WNT10B
GO:0030111	Regulation of Wnt receptor signalling pathway	ALX4|DKK1|DSG4|FGF5|NFKB1|SHH|SNAI1|SOX9|TGFB2|VDR|WNT10A|WNT10B
GO:0030178	Negative regulation of Wnt receptor signalling pathway	ALX4|DKK1|DSG4|FGF5|NFKB1|SHH|SNAI1|SOX9|TGFB2|WNT10A|WNT10B
GO:0060070	Canonical Wnt receptor signalling pathway	ALX4|DSG4|FGF5|INS|SHH|SNAI1|VDR|WNT10A|WNT10B
GO:0060828	Regulation of canonical Wnt receptor signalling pathway	ALX4|DKK1|DSG4|FGF5|NFKB1|SHH|SNAI1|SOX9|TGFB2|VDR|WNT10A|WNT10B

BMP		
GO:0030509	BMP signalling pathway	DSG4|FGF5|SMAD4|TGFB2|WNT10A
GO:0030510	Regulation of BMP signalling pathway	DKK1|SHH|SMAD4|TGFB2|VDR|WNT10A
GO:0030513	Positive regulation of BMP signalling pathway	DKK1|SHH|SMAD4|TGFB2|VDR|WNT10A
GO:0030514	Negative regulation of BMP signalling pathway	DKK1|SHH|SMAD4|TGFB2|VDR|WNT10A
NOTCH		
GO:0007219	Notch signalling pathway	DSG4|NFKB1|SNAI1|TGFB2|VDR
GO:0008593	Regulation of Notch signalling pathway	ALX4|DSG4|FGF5|NFKB1|TGFB2|TP63|WNT10A
GO:0045747	Positive regulation of Notch signalling pathway	ALX4|FGF5|NFKB1|TGFB2|TP63|WNT10A

The findings obtained from applying Tool Service 14 to the selected 48 genes support the literature findings discussed in Section 2.2.2. Specifically, our results reveal that fourteen of the chosen genes are associated with the WNT pathway, which plays a critical role in various processes related to hair follicle growth and development. Furthermore, eight of the selected genes are involved in the BMP signalling pathway, which is responsible for the differentiation of hair follicles, while nine genes are implicated in the Notch signalling pathway, which is involved in regulating hair follicle development.

6.5.5 Results from the application of Tool Service 15 (DisGeNET-Cytoscape data)

Genome-wide association studies have identified numerous disease-susceptibility genes. As knowledge of gene–disease associations accumulates, it is becoming increasingly important to translate this knowledge into clinical practice and produce more effective therapies. One of the largest publicly accessible collections of genes and variants linked to human diseases is found on the discovery platform DisGeNET (https://www.disgenet.org). Tool Service 15 was used to retrieve gene-disease associations for AGA and a number of diseases which are thought to be genetically or phenotypically related to AGA. The resulting genes for each disease were compared to the 48 genes chosen in this thesis.

Gene similarity to the 48 chosen genes (see Table 6.5.1) was assessed by comparing them to the genes linked to several similar diseases in DisGeNET (results are available in Appendix 15) like:

- AGA,

-diseases with similar phenotypes to AGA like Hypotrichosis (1,2,3,4,6,7,8,

14), atrichia with papular lesions and Loose anagen hair syndrome

-diseases which have been linked to AGA in literature (Prostate cancer and Type 2 diabetes mellitus).

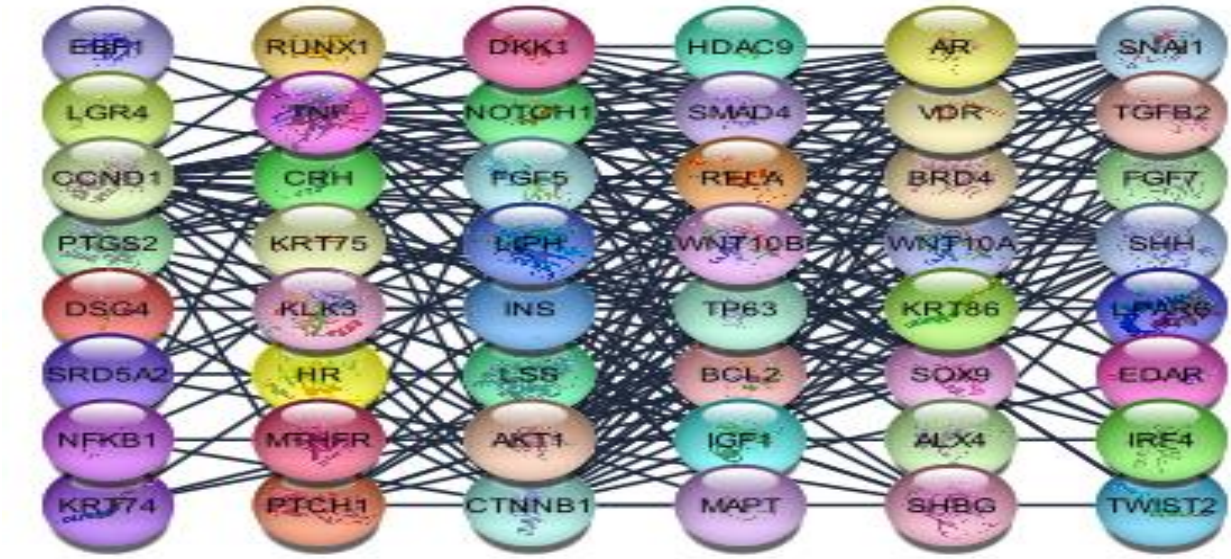
Results are summarised in Table 6.5.6.

Table 6.5.6: A summary of the comparison of the 48 genes chosen and gene sets contained in DisGeNET for diseases related to AGA

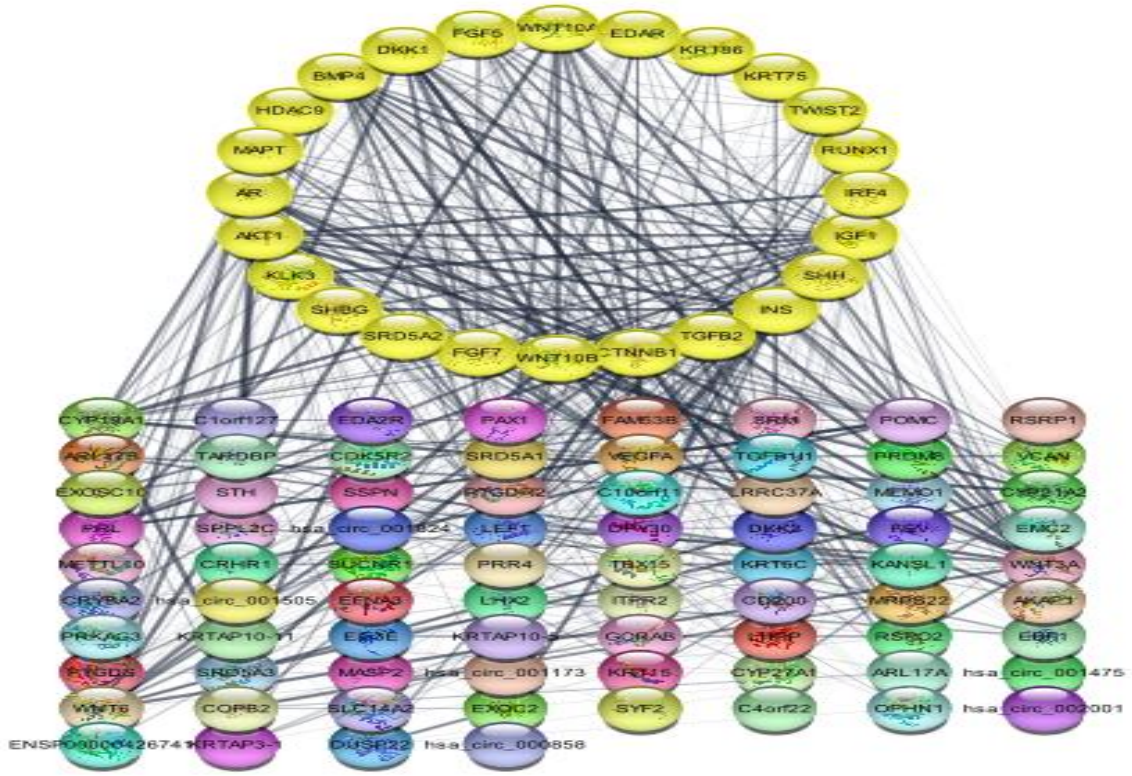
OMIM ID	Disease	Common genes with the 48 genes chosen
109200	Androgenetic alopecia	AR,SRD5A2,CRH,BRD4,MTHFR,HR,VDR
605389	Hypotrichosis 1	LPAR6,LIPH,LSS,DSG4
146520	Hypotrichosis 2	KRT74
613981	Hypotrichosis 3	KRT74
146550	Hypotrichosis 4	HR
607903	Hypotrichosis 6	DSG4
604379	Hypotrichosis 7	LIPH
278150	Hypotrichosis 8	LPAR6
618275	Hypotrichosis 14	LSS
209500	Atrichia with papular lesions	HR
600628	Loose anagen hair syndrome	KRT75
176807	Prostate cancer	AR,CTNNB1,CCND1,KLK3,AKT1,BRD4,IGF1,MTHFR,PTGS2,BCL2, SHBG,SRD5A2,VDR,WNT10B
125853	Type 2 Diabetes Mellitus	INS,BCL2,TNF,SHBG,NFKB1,RELA

Using RCy3 a network of the genes chosen was created and visually compared against STRING database disease networks (for the diseases mentioned above) retrieved in Cytoscape. The results are shown in Figure 6.5.1.

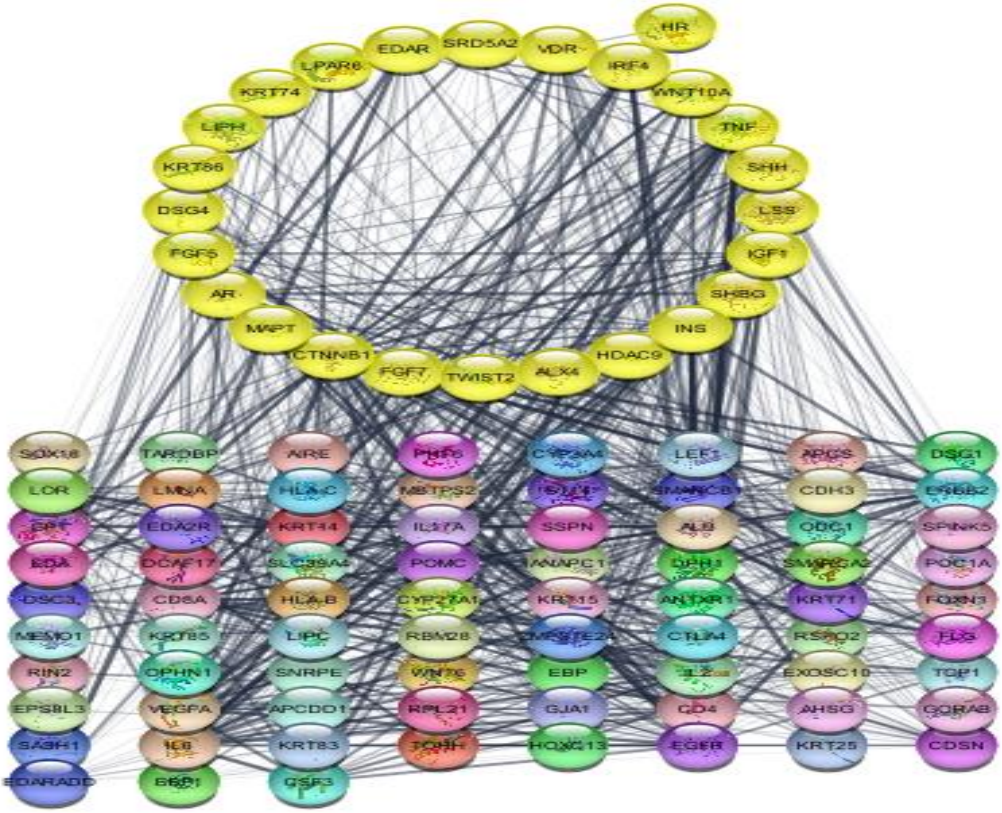
The Network of the 48 chosen genes



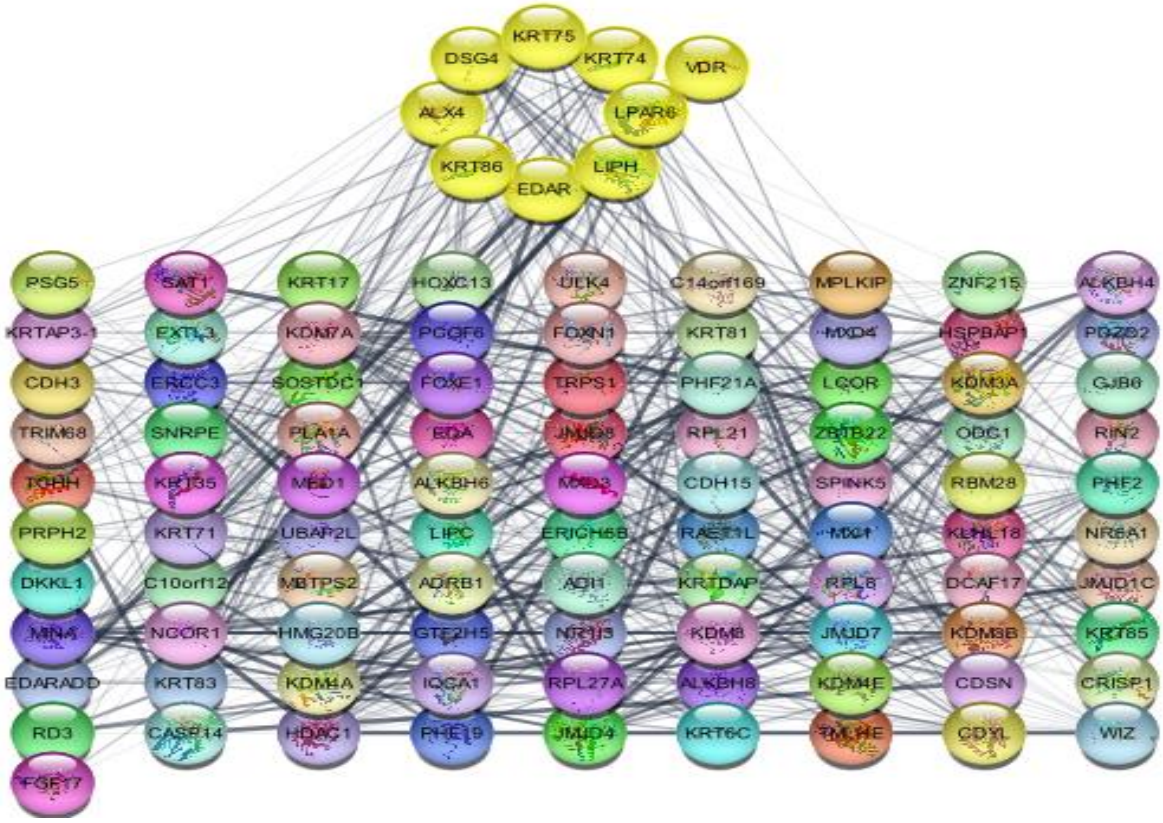
Androgenetic alopecia implicated genes according to STRING database



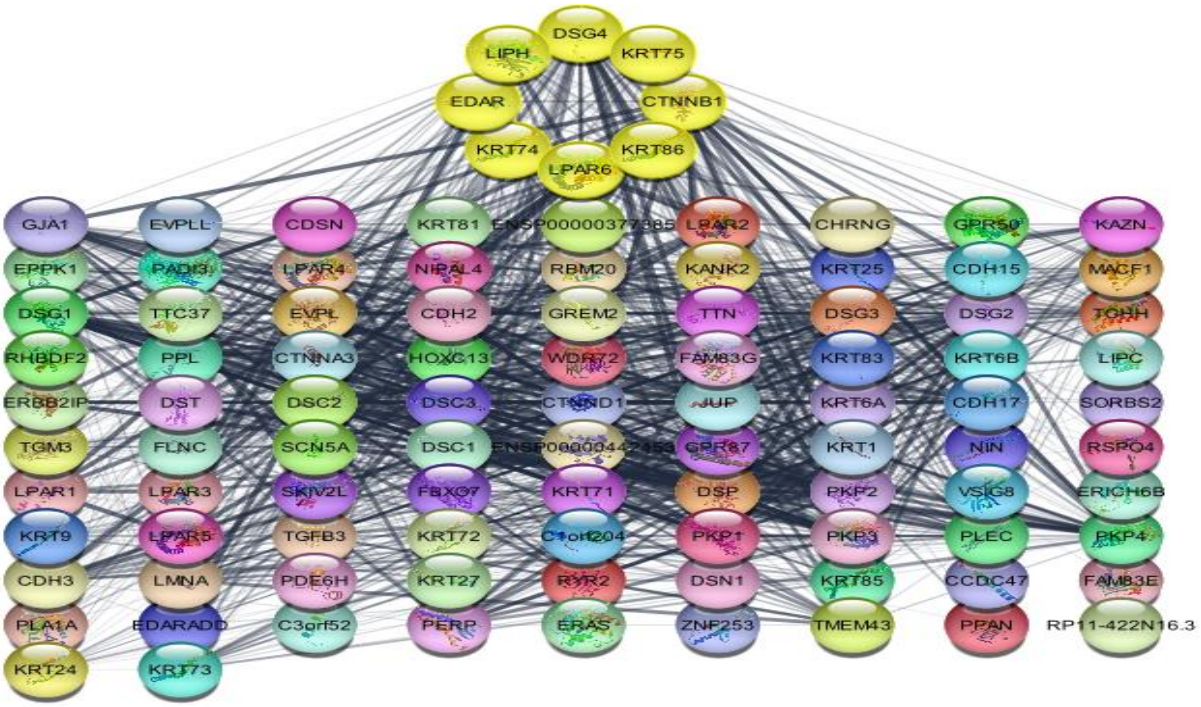
Hypotrichoses genes implicated genes according to STRING database



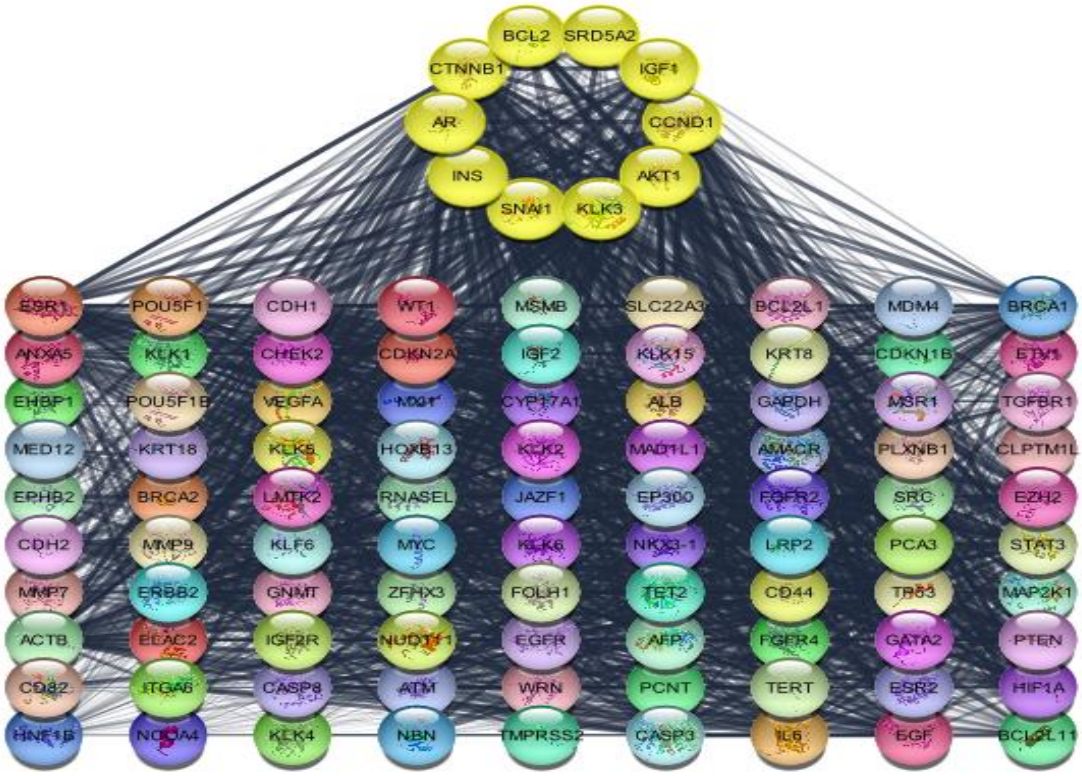
Atrichia with papular lesions implicated genes according to STRING database



Familial woolly hair syndrome implicated genes according to STRING database



Prostate neoplasm implicated genes according to STRING database



Type 2 Diabetes Mellitus implicated genes according to STRING database

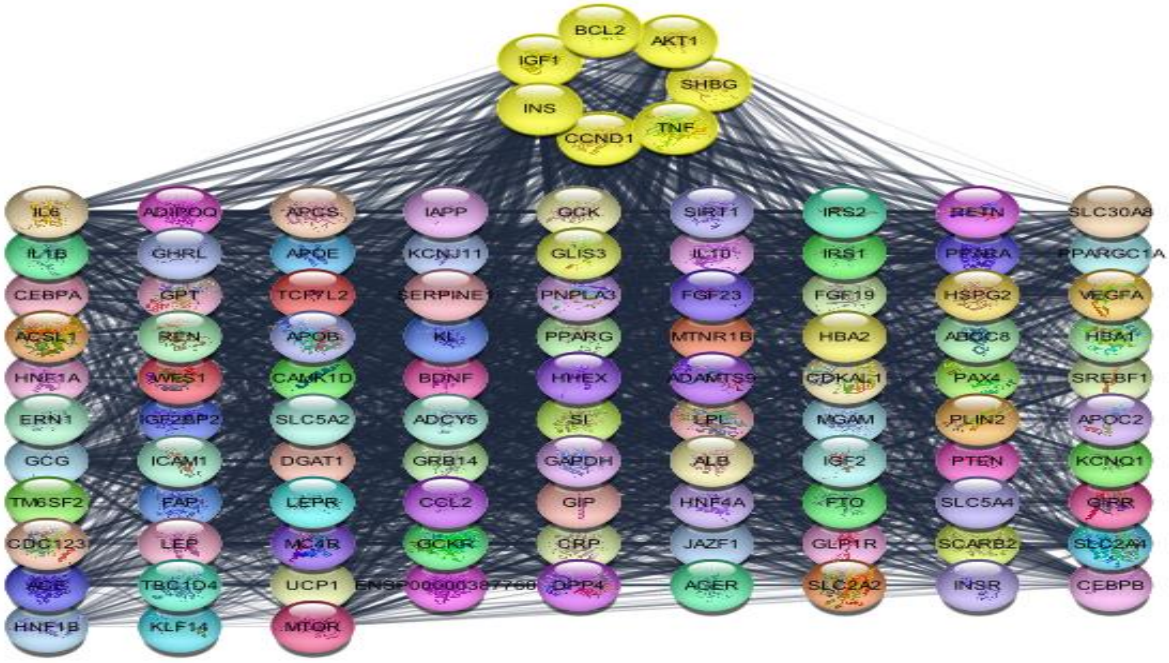


Figure 6.5.1: Comparison of the network of 48 genes against STRING database networks of similar diseases. Common genes are shown in a circle of each diagram. All diagrams were retrieved in R with RCy3

The comparison of our network of the 48 genes chosen against different STRING disease networks revealed that our network shared 24 out of 48 genes with the AGA network, 25 out of 48 genes with the hypotrichosis network, 9 out of 48 genes with the atrichia network, 8 out of 48 genes with the wooly hair network, 10 out of 48 genes with the prostate neoplasms network, and 7 out of 48 genes with the diabetes mellitus network.

These results indicate that a significant number of the chosen genes are overlapping with the disease networks related to the conditions mentioned. This suggests that the selected genes may play important roles in the pathogenesis or regulation of these diseases. The high number of shared genes between our gene network and the disease networks for androgenetic alopecia (AGA) and hypotrichoses indicates a strong relevance of the selected genes to these hair-related conditions. Additionally, the presence of a subset of the selected genes in the disease networks of atrichia, wooly hair, prostate neoplasms, and diabetes mellitus suggests potential connections or shared molecular mechanisms between these diseases and the genes of interest. These findings support the notion that the chosen genes

are promising candidates for further investigation and may provide insights into the underlying molecular pathways and potential therapeutic targets for AGA.

6.5.6 Results from the application of Tool Service 16 (CYP450, Phase II enzymes and transporters data)

Tool Service 16 was used to obtain information on the enzymes and transporters involved in the metabolism of the seven drugs currently used in alopecia (see Section 6.4). No information was found regarding the metabolism of Carpronium chloride, Alfatradiol, Bimatoprost and Biotin. Information retrieved for Finasteride, Minoxidil and Dutasteride is summarised in Table 6.5.7.

Table 6.5.7: Results on the metabolic fate of three of the seven drugs currently marketed for AGA

Drug	PubChem CID	Relation	Metabolism	PMID
Finasteride	57363	Substrate	3A4	11996015
Finasteride	57363	Substrate	3A4	8654202
Finasteride	57363	Inhibitor	GST	19077407
Minoxidil	4201	Substrate	Sulphotransferases	1349030
Minoxidil	4201	Substrate	Sulphotransferases	8423770
Minoxidil	4201	Substrate	Sulphotransferases	2390100
Minoxidil	4201	Substrate	Sulphotransferases	9566733
Minoxidil	4201	Substrate	Sulphotransferases	9037254
Minoxidil	4201	Substrate	Sulphotransferases	2230218
Minoxidil	4201	Substrate	Sulphotransferases	3480782
Minoxidil	4201	Substrate	Sulphotransferases	6958263
Minoxidil	4201	Substrate	Sulphotransferases	7755612
Dutasteride	6918296	Substrate	3A5	34690761

From Table 6.5.7 it can be seen that Finasteride is metabolised by CYP3A4 and glutathione s-transferase, Dutasteride is metabolised by CYP3A5 while minoxidil by sulphotransferases (SULT1A1 and SULT1A3).

Using Tool Service 20 we retrieved the genes implicated in drug metabolism for 30 of the 34 phytochemicals (no data was available for arachidic acid, alpha-thujene, alpha-curcumene

and pinocarvone). Results are available in Appendix 16. Table 6.5.8 below shows the results for five phytochemicals and Sulphotransferase 1A1 (SULT1A1).

Table. 6.5.8: The five phytochemicals implicated in the metabolism of SULT1A1

ChemicalName	PubChem CID	GeneSymbol	PubMedIds
p-hydroxybenzoic acid	135	SULT1A1	16308312
apigenin	5280443	SULT1A1	17293380
caffeic acid	689043	SULT1A1	17433394
gallic acid	370	SULT1A1	16308312
p-coumaric acid	637542	SULT1A1	16308312

The active metabolite of minoxidil that causes the drug's vasodilator and hair-growing effects is minoxidil sulphate. Sulphotransferase enzymes are expressed in the ORS of the HF and sulphonate minoxidil for hair growth. In some patients, the low response rate to topical minoxidil was found to be predicted by the high inter-subject variability in follicular sulphotransferase. Topical tretinoin appears to increase hair by 20% and triple the absorption of minoxidil (as was mentioned in Section 3.4). According to a Sharma et al. (2019) study in less than five days after topical tretinoin application, non-responders to minoxidil become responders. This is probably because tretinoin affects the expression of follicular sulphotransferase.

In order to explore the potential of the five phytochemicals to elicit similar effects as tretinoin does, we conducted molecular docking studies to assess the impact of minoxidil, tretinoin, and the five phytochemicals (refer to Table 6.5.8) on SULT1A1. The aim was to determine if these compounds could exhibit comparable actions to tretinoin, specifically regarding their interaction with SULT1A1.. 3D-structure of the target SULT1A1 (PDB:1LS6) was downloaded from RSCB (https://www.rscb.org). Molecular docking was performed using PyRx v0.8. All 3D-SDF files were downloaded from PubChem compound database. Both ligands and the target were prepared using the PyRx v.0.8 GUI. The 3D modelled protein was imported into PyRx software, which generated a PDBQT file of the protein structure with all polar hydrogens included. All ligand bonds were considered rotatable. Following the completion of the docking

study, the ligand docked pose with the least binding energy was chosen. The binding pattern of docked complexes bonding details, and length was studied using the Discovery Studio Visualizer v16.1.0.15350.

The results of the molecular docking and binding study (see Table. 6.5.9 and Figure 6.5.2 below) showed that from the five tested phytochemicals, only apigenin had a binding affinity comparable to tretinoin's for SULT1A1 (the rest of the four phytochemicals had binding affinities lower than minoxidil) and could be further investigated for its possible role in increasing the response to minoxidil therapy in non-responding patients.

Table 6.5.9: The results of the molecular docking for the five phytochemicals, minoxidil and tretinoin

ChemicalName	PubChem CID	Binding affinity(Kcal/mol)
p-hydroxybenzoic acid	135	-4.3
apigenin	5280443	-6
caffeic acid	689043	-4.6
gallic acid	370	-4.2
p-coumaric acid	637542	-5.1
minoxidil	4201	-5
tretinoin	444795	-5.7

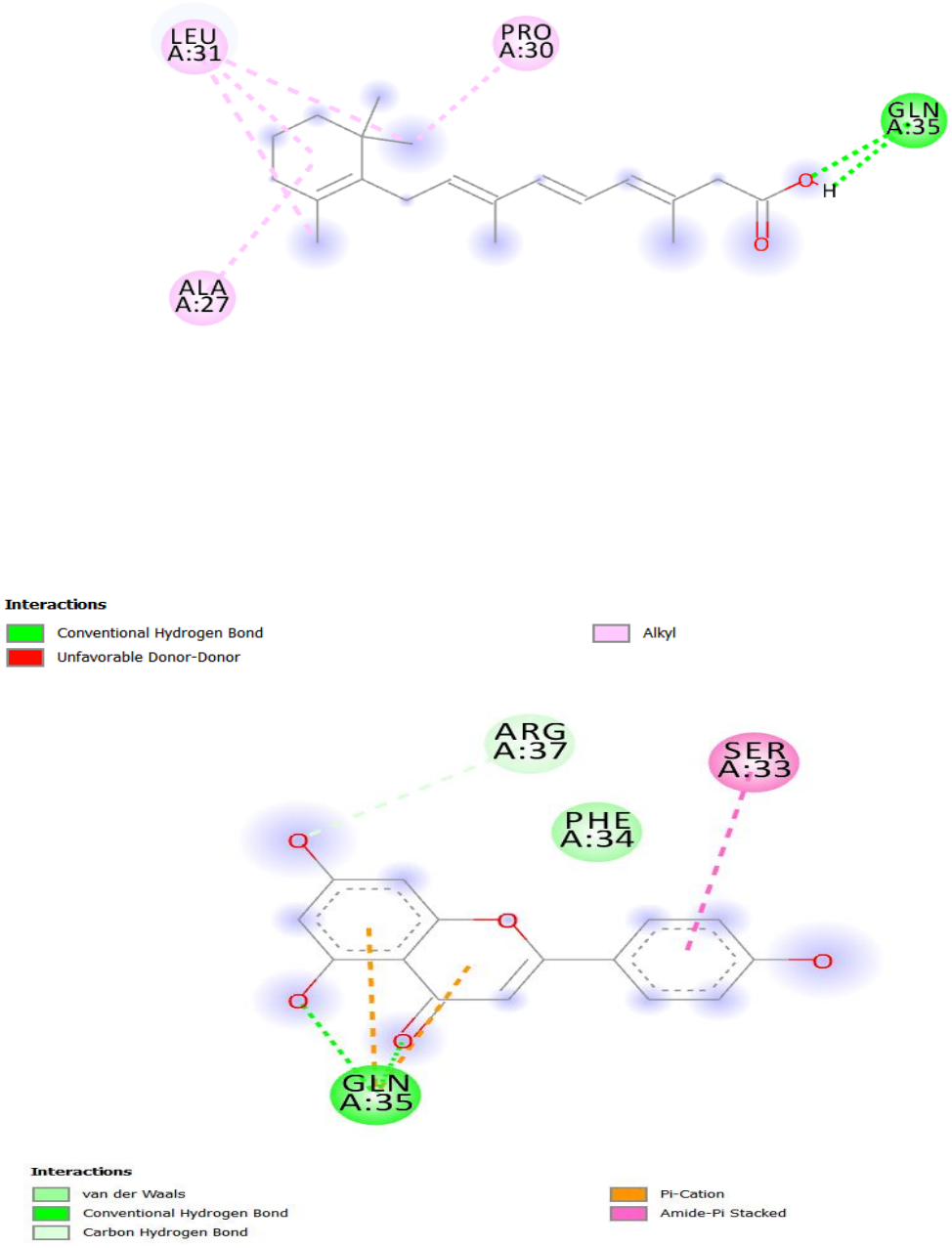


Figure. 6.5.2: The binding of tretinoin (top) and apigenin (bottom) to SULT1A1. (Discovery Studio Visualizer was used for these illustration).

6.5.7 Results from the application of Tool Service 17 (Target binding sites)

Tool service 17 was used to quickly reveal the possible binding sites of the 48 targets selected (see Table 6.5.1). To retrieve the binding sites, PDB codes (already retrieved in a previous step with Tool service 11) of each target were input. The resulting .csv files contain a list of the amino acids, their position in the protein and the specific chain of the protein (and are

available in Appendix 17). For genes: DSG4, EDAR, FGF5, FGF7, HDAC9, HR, KRT74, KRT75, KRT86, TWIST2, WNT10A AND WNT10B this was not possible as there were no valid PDB files found.

Molecular docking

To test the appropriateness of our retrieved data and also assess the binding affinity of both the 7 drugs and 34 phytochemicals for a given target (in this case AR), we performed molecular docking analysis using PyRx v0.8. The 3D structure of AR (PDB ID: 1E3G) was downloaded from RSCB. All 3D-SDF files were downloaded from PubChem compound database except for the 3D-SDF files of stearic acid and arachidic acid which were downloaded from the Human Metabolome Database (https://hmdb.ca). Both ligands and the target were prepared using the PyRx V.0.8 GUI. The 3D modelled protein was imported into PyRx software, which generated a PDBQT file of the protein structure with all polar hydrogens included. All ligand bonds were considered rotatable. Following the completion of the docking study, the ligand docked pose with the least binding energy was chosen. The binding pattern of docked complexes bond details, and bond length was studied using the Discovery Studio Visualizer v16.1.0.15350.

The results show that most of the drugs currently indicated for AGA (minoxidil, finasteride, dutasteride and alfatradiol) exhibit a very high binding interaction with the binding site of AR. Ten of the phytochemicals (apigenin, quercetin, kaempferol, beta-sitosterol, chlorogenic acid, rutin, adenosine, luteolin, oleanolic acid and ursolic acid) showed similar binding energies and can be potential leads for the design of novel AGA treatments.

The compounds were ranked based on the binding energy of their pose with the lowest binding energy and can be found in Appendix 17.

An analysis of the molecule-target interactions shows that both minoxidil and kaempferol could form hydrogen-bond interactions with ASN-705 and GLY-464 residues located in the substrate-binding site (Figure 6.5.2 a and b).

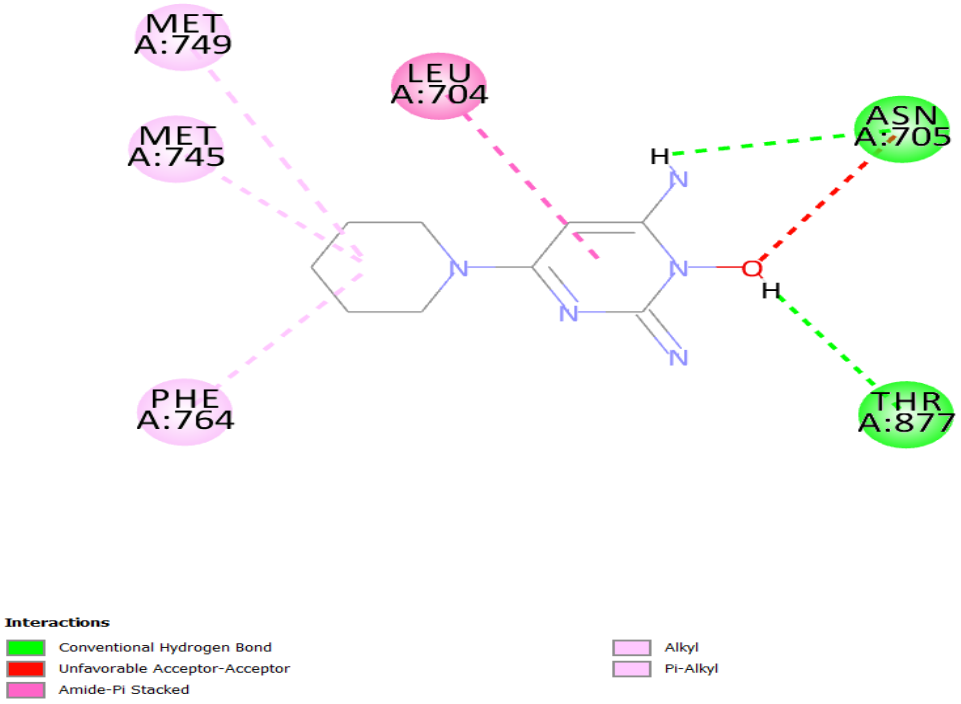


Figure 6.5.2.a Binding of minoxidil to the binding site of the androgen receptor. (Discovery Studio Visualizer was used for this illustration).

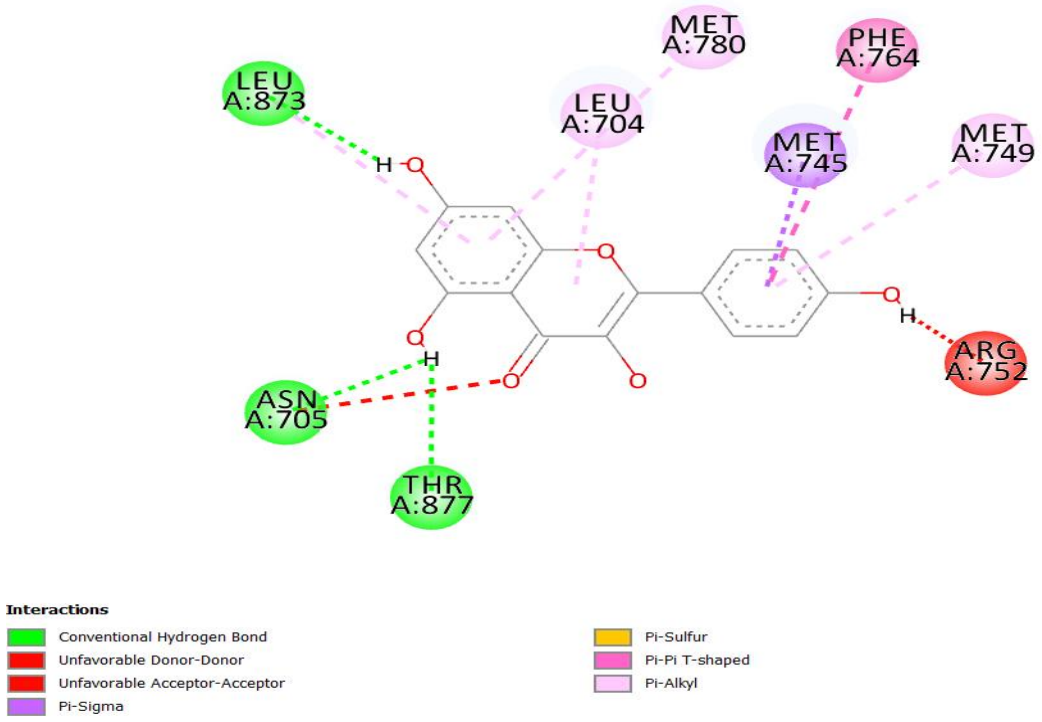


Figure 6.5.2.b Binding of kaempferol to the binding site of the androgen receptor. (Discovery Studio Visualizer was used for this illustration).

6.5.8 Results from the application of Tool Service 18 (KEGG pathways)

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, are a vast and highly curated collection of annotated biological pathways. These pathways provide detailed information about the molecular interactions, cellular processes, and functional relationships between genes, proteins, and small molecules involved in various biological systems. Each pathway in the KEGG database is meticulously documented and organized, enabling researchers to explore and understand the complex networks and signalling cascades that govern different biological processes. By utilising tools and services that leverage KEGG pathways, researchers can identify key targets and pathways relevant to specific diseases or biological phenomena, aiding in elucidating molecular mechanisms and the discovery of potential therapeutic interventions.

Tool Service 18 reveals the KEGG pathways of interest and highlights all the associated targets being studied in each case. This is a very important step in our effort to distinguish those targets and pathways implicated in the process of androgenetic alopecia. One of these pathways is the WNT signalling pathway which is the most significant and the most studied in regards to the hair follicle.

By using Tool Service 18, several genes (CCND1, CTNNB1, DKK1, LGR4, SMAD4, WNT10A, WNT10B) that we selected based on their potential implication in AGA were also found to be associated with the Wnt pathway, indicating a potential connection between these genes and the WNT signalling pathway in the context of AGA. The resulting pathway, depicted in Figure 6.5.3 visually represents the relationship between these genes and the Wnt pathway. This finding supports the hypothesis that dysregulation of the WNT pathway may contribute to the development and progression of AGA.

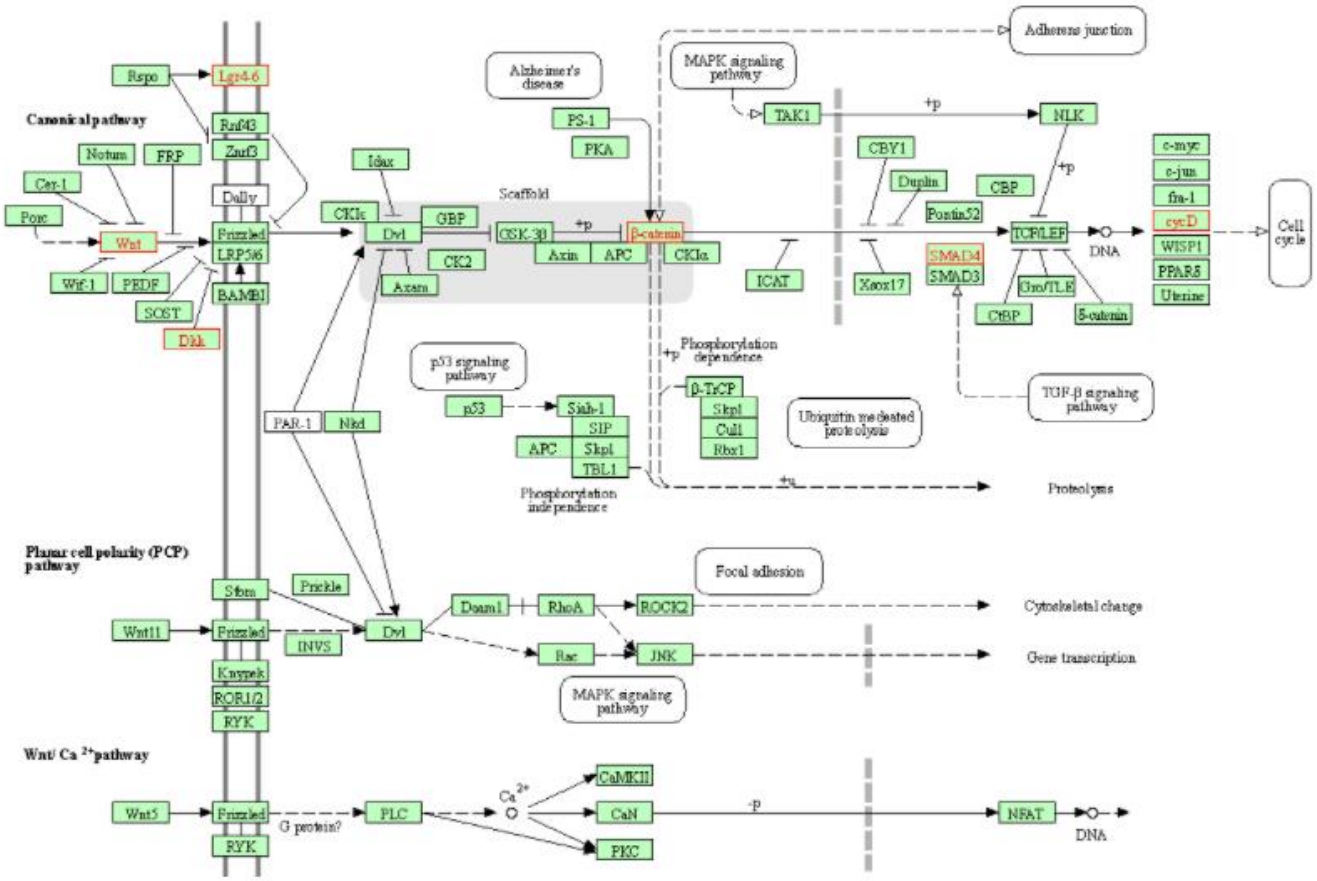


Figure 6.5.3: A diagram of the KEGG Wnt signalling pathway. Seven of the chosen targets (marked with red color) participate in this pathway.

All results retrieved with Tool Service 18 are found in Appendix 18.

6.5.9 Results from the application of Tool Service 19 (Pathway network visualisation)

In order to investigate the underlying biology of a complex phenomenon and ultimately find functional candidate genes affecting the trait under study, it may be helpful to create a network of numerous interconnected pathways of interest.

Tool Service 19 visualises in our browser genes within a network of multiple levels (from 1 to n) of connected upstream and downstream pathways using data from KEGG. The network graph visualisation aids in deciphering a gene cluster's functional profiles. In this case, the results suggest that the identified genes are part of a larger network encompassing various pathways (WNT, Notch and Hedgehog) related to hair development and growth. By visualising this network, it becomes possible to discern the functional profiles and interactions of gene

clusters involved in the studied trait. The network graph visualisation provided by Tool Service 19 aids in unraveling the complex relationships and potential crosstalk between the genes within the chosen pathways.

Figure 6.5.4 shows the resulting network created for all 48 genes chosen and three signalling pathways important in hair development and growth (WNT, Notch and Hedgehog).

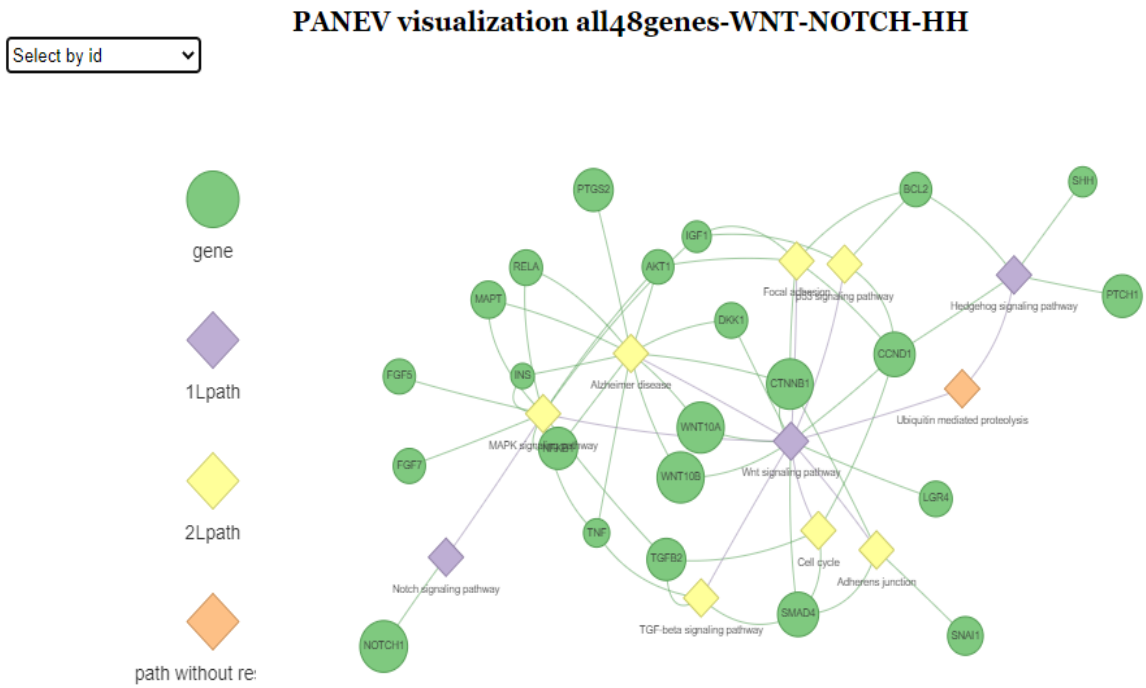


Figure 6.5.4: The network created for all 48 genes chosen and WNT, NOTCH and Hedgehog signalling pathways

All results of Tool Service 19 are found in Appendix 19.

6.5.10 Results from the application of Tool Service 20 (Comparative Toxicogenomics data)

Tool Service 20 was used in a variety of processes throughout this thesis, mainly to confirm results from other Tool Services and help in deciding the targets to be studied. It also provided data linking many of the chosen phytochemicals to multiple targets implicated in the treatment of AGA (according to the Comparative Toxicogenomics database).

Association of phytochemicals to androgenetic alopecia biomarkers

All 34 phytochemicals were tested for their link to genes having direct evidence (according to CTD) of either being a biomarker or have a role in the aetiology of androgenetic alopecia, using Tool Service 20. A list of all chemicals considered to be linked to Comparative Toxicogenomic Database's AGA biomarkers can be found in Appendix 20.

Table 6.5.10 Associations of phytochemicals with genes linked to androgenetic alopecia by CTD.

Phytochemical	Gene	Inference score	PubMed Id
Protocatechuic acid	PARP1	3.54	20561897
Benzaldehyde	AR	3.86	15902657
Gallic acid	PARP1	2.71	20561897
Palmitic acid	PARP1	2.54	20561897
Stearic acid	AR	3.38	15902657
gamma-Terpinene	AR	4.01	15902657
Oleanolic acid	PARP1	2.8	20561897
Ursolic acid	PARP1	2.97	20561897
beta-Sitosterol	PARP1	3.68	20561897
Oleic acid	PARP1	2.53	20561897
Caffeic acid	CRH	3.51	21359208
Chlorogenic acid	ABCC2	7.62	18381794
Chlorogenic acid	AR	7.62	15902650
Chlorogenic acid	SRD5A2	7.62	17136762
Quercetin	ABCC2	1.99	18381794
Quercetin	AR	1.99	15902657
Quercetin	SRD5A2	1.99	21359208
Quercetin	CRH	1.99	20561897
Quercetin	PARP1	1.99	19652058
Quercetin	TNFRSF10	1.99	1338926
Quercetin	VDR	1.99	22466564
Quercetin	ZFP36	1.99	15944294
Apigenin	PARP1	2.53	20561897
Luteolin	ABCC2	4.7	18381794
Luteolin	PARP1	4.7	20561897
Linoleic acid	VDR	3.31	22466564
Rutin	PARP1	3.95	20561897
Kaempferol	AR	6.45	15902657
Kaempferol	PARP1	6.45	20561897
Kaempferol	VDR	6.45	2246654

From the results shown in Table 6.5.8, we can deduce that more than half (18) of the chosen 34 phytochemicals had a link to a biomarker of androgenetic alopecia. These results can be visualised in the following Figure 6.5.5.

Figure. 6.5.5: A diagram of phytochemicals and their links to genes that may be (according to CTD) biomarkers or play a role in the aetiology of androgenetic alopecia

Summary of the results retrieved withTool Services 11-20

The utilisation of Tool Services 11-20 in our study have provided valuable insights into target identification and prioritisation for androgenetic alopecia (AGA) and the discovery of potential new drugs for its treatment. By considering multiple data sources and criteria, our goal was to identify genes that are highly relevant and significant in investigating AGA.

To prioritize genes associated with AGA, we employed factors such as Genome-Wide Association Studies (GWAS) to assess their p-values, indicating their potential involvement in AGA's pathogenesis. Text-mining through Tool Service 2 helped identify genes frequently associated with AGA in scientific literature. Additionally, relevant Gene Ontology (GO) terms related to hair processes were used to further narrow down potential targets. The Chemical

Toxicogenomics Database (CTD) provided genes recognized as biomarkers or contributors to AGA's aetiology.

The amalgamation of these criteria and data from various sources led us to compile a list of 48 AGA targets, which were subjected to analysis using different tools and services. Tool Service 11 provided essential target information, while Tool Service 12 facilitated the translation of target IDs into databases like KEGG, EGGNOG, ENSEMBL, and PDB. Tool Service 13 retrieved data from BindingDB, enabling us to understand the interactions between chemicals and targets, thereby identifying potential drug candidates. The findings obtained from Tool Service 14 aligned with the literature findings discussed in Section 2.2.2, revealing that fourteen of the selected genes are associated with the WNT pathway, which plays a critical role in hair follicle growth and development. Moreover, eight genes were found to be involved in the BMP signalling pathway responsible for hair follicle differentiation, and nine genes were implicated in the Notch signalling pathway regulating hair follicle development.

Tool Service 15 allowed us to compare the chosen genes with genes in DisGeNET and Cytoscape, unveiling the significance of these genes for AGA and related diseases. Overlaps between the chosen gene network and AGA and hypotrichosis disease networks indicated their relevance. Furthermore, potential connections or shared molecular mechanisms were discovered between AGA and conditions such as atrichia, wooly hair, prostate neoplasms, and diabetes mellitus.

Insights into drug metabolism, specifically for Finasteride, Minoxidil, and Dutasteride, were obtained through Tool Service 16, revealing the relevance of Sulphotransferase 1A1 (SULT1A1) in minoxidil metabolism and suggesting the potential of apigenin in enhancing minoxidil effectiveness.

The application of Tool Service 17 provided valuable insights into the potential binding interactions between the selected targets and drugs/phytochemicals relevant to AGA. Most of the drugs currently indicated for AGA exhibited a high binding interaction with the androgen receptor (AR), validating their known mechanisms of action. Additionally, ten phytochemicals showed promising binding interactions with the AR binding site, indicating their potential as leads for novel AGA treatments.

The integration of Tool Services 18, 19, and 20 further enhanced our understanding of AGA's molecular mechanisms, highlighting the significance of the Wnt signalling pathway and its genes. Pathway visualisation and comparative toxigogenomic data confirmed our results and linked phytochemicals to AGA-related genes, providing additional evidence for their therapeutic potential.

The comprehensive utilisation of these Tool Services and databases has significantly contributed to our understanding of AGA's molecular mechanisms, potential targets, and therapeutic approaches. The identification of key pathways, the associations between phytochemicals and AGA-related genes, and the visualisation of interconnected pathways offers valuable insights for future research and the development of targeted therapies for AGA.

6.6 Results from the application of Tool Services 21-22

6.6.1 Results from the application of Tool Service 21 (QSAR-DB predictions)

This service is taking advantage of QSARDB's API (Ruusmann, Sild and Maran, 2015) to make predictions for properties such as a compounds melting point, Abraham's solvent coefficients and descriptors. The latter is of great importance in the prediction of drug transport through biological barriers. Results of applying this Tool Service to the 34 phytochemicals are reported in Table 6.6.1.

A. Melting Point

Table 6.6.1: Predicted melting point values for the phytochemicals using Tool Service 21

Phytochemical	Predicted values Melting Point (°C)
adenosine	220.6
alpha-caryophyllene	19.3
alpha-curcumene	11.9
alpha-pinene	-5.0
alpha-thujene	-22.6
apigenin	230.6
arachidic acid	73.2

benzaldehyde	-8.6
beta-elemene	44.7
beta-myrcene	-52.5
beta-sitosterol	132.8
beta-terpinene	-61.7
caffeic acid	169.6
chlorogenic acid	192.2
ferulic acid	156.1
gallic acid	214.7
kaempferol	232.4
linoleic acid	38.8
luteolin	231.3
myristic acid	51.9
oleanolic acid	163.7
oleic acid	44.5
palmitic acid	58.0
palmitoleic acid	27.2
p-coumaric acid	154.5
p-hydroxybenzoic acid	199.8
pinocarvone	58.5
protocatechuic acid	215
quercetin	232.7
rutin	213.4
stearic acid	67.9
ursolic acid	168.5
vanillic acid	210.9
vanillin	79.9

B. Abraham's Descriptors

Table 6.6.2: Predicted Abrahams descriptors values for the phytochemicals using Tool Service 21

	Abraham's descriptors				
Phytochemical	V	B	A	S	E
adenosine	1.742	1.822	0.87	2.14	2.024
alpha-caryophyllene	1.92	0.218	0	0.498	0.611
alpha-curcumene	1.897	0.263	0	0.861	0.936
alpha-pinene	1.22	0.212	0	0.264	0.544
alpha-thujene	1.159	0.187	0	0.246	0.539
apigenin	1.861	1.094	0.89	2.183	2.304
arachidic acid	2.427	0.698	0.44	1.003	0.398

benzaldehyde	0.862	0.488	0	1.008	0.872
beta-elemene	1.892	0.245	0	0.718	1.007
beta-myrcene	1.329	0.249	0	0.344	0.463
beta-sitosterol	3.695	1.759	0.18	1.882	2.009
beta-terpinene	1.078	0.234	0	0.317	0.485
caffeic acid	1.243	0.801	0.97	1.483	1.205
chlorogenic acid	2.364	2.379	0.78	2.724	2.201
ferulic acid	1.394	0.816	0.74	1.487	1.154
gallic acid	1.099	0.819	1.16	1.617	1.291
kaempferol	1.918	0.956	0.58	2.02	2.54
linoleic acid	2.412	0.775	0.4	1.006	0.449
luteolin	1.918	0.897	0.57	1.969	2.606
myristic acid	2.029	0.604	0.46	0.79	0.402
oleanolic acid	3.889	2.006	0.39	2.399	1.74
oleic acid	2.492	0.767	0.4	0.99	0.438
palmitic acid	2.349	0.777	0.32	0.822	0.424
palmitoleic acid	2.285	0.744	0.34	0.846	0.433
p-coumaric acid	1.177	0.755	0.81	1.336	1.151
p-hydroxybenzoic acid	0.991	0.718	0.78	1.154	1.134
pinocarvone	1.238	0.633	0	0.894	0.844
protocatechuic acid	1.041	0.759	1.04	1.449	1.162
quercetin	1.986	1.437	0.65	2.391	2.655
rutin	4.104	3.444	0.6	3.87	4.142
stearic acid	2.583	0.832	0.4	1.056	0.426
ursolic acid	3.888	2.131	0.34	2.405	1.789
vanillic acid	1.15	0.805	0.79	1.487	1.155
vanillin	1.113	0.748	0.35	1.325	1.078

Table 6.6.3: Predicted Abrahams solvent coefficient values for the phytochemicals using Tool Service 21

	Abraham's solvent coefficients					
Phytochemical	v	b	a	s	e	C
adenosine	3.753	-4.265	0.1	0.023	0.367	0.048
alpha-caryophyllene	4.492	-4.925	-3.399	-1.124	0.586	0.128
alpha-curcumene	4.491	-4.916	-3.399	-1.431	0.578	0.129
alpha-pinene	4.483	-4.763	-3.358	-1.354	0.571	0.127
alpha-thujene	4.486	-4.764	-3.404	-1.419	0.576	0.128
apigenin	4.024	-4.278	-0.354	0.028	0.339	0.049
arachidic acid	4.15	-4.413	-0.359	-0.43	0.488	0.057
benzaldehyde	4.153	-4.695	-0.756	-0.125	0.306	0.073
beta-elemene	4.439	-4.881	-2.904	-0.949	0.561	0.14
beta-myrcene	4.499	-5.023	-3.433	-1.407	0.547	0.179

beta-sitosterol	4.177	-4.283	-0.33	-1.006	0.525	0.095
beta-terpinene	4.508	-4.869	-3.432	-1.432	0.592	0.121
caffeic acid	4.078	-4.313	-0.386	-0.056	0.298	0.055
chlorogenic acid	3.898	-4.274	-0.416	-0.025	0.353	0.051
ferulic acid	4.029	-4.316	-0.444	-0.077	0.296	0.057
gallic acid	4.038	-4.328	-0.302	0.03	0.279	0.062
kaempferol	4.011	-4.302	-0.361	0.042	0.347	0.051
linoleic acid	4.148	-4.4	-0.383	-0.465	0.47	0.06
luteolin	4.023	-4.331	-0.358	0.015	0.346	0.05
myristic acid	4.15	-4.411	-0.359	-0.43	0.487	0.055
oleanolic acid	4.053	-4.349	-0.274	-0.13	0.455	0.111
oleic acid	4.15	-4.412	-0.374	-0.431	0.486	0.057
palmitic acid	4.15	-4.411	-0.359	-0.43	0.487	0.056
palmitoleic acid	4.15	-4.408	-0.374	-0.431	0.485	0.059
p-coumaric acid	4.033	-4.292	-0.416	-0.065	0.313	0.062
p-hydroxybenzoic acid	3.913	-4.251	-0.341	-0.002	0.283	0.069
pinocarvone	4.123	-4.732	-0.568	-0.137	0.342	0.09
protocatechuic acid	3.947	-4.283	-0.295	0.026	0.266	0.062
quercetin	4.028	-4.336	-0.38	0.027	0.359	0.05
rutin	3.878	-4.287	-0.312	0.017	0.366	0.058
stearic acid	4.15	-4.413	-0.359	-0.43	0.488	0.057
ursolic acid	4.055	-4.344	-0.385	-0.126	0.452	0.113
vanillic acid	3.951	-4.297	-0.429	0.022	0.273	0.061
vanillin	3.956	-4.279	-0.336	0.007	0.27	0.055

By analysing the Abrahams descriptors of different compounds, researchers can make informed decisions regarding their drug-likeness and suitability for further development.

By considering the solvent coefficients, researchers can gain insights into the compounds' behavior in biological systems and make informed decisions about its potential as a drug candidate.

Utilising predicted Abrahams descriptors and solvent coefficients, we can narrow down the focus to compounds with desirable physicochemical properties and drug-like characteristics, thus saving time and resources in the drug discovery process. These computational tools allow for the screening and prioritisation of large compound libraries, helping researchers identify potential lead compounds and optimize their properties early in the discovery phase. Additionally, these descriptors and coefficients can guide medicinal chemists in designing and modifying molecules to improve their drug properties, such as enhancing solubility, permeability, and target interactions.

6.6.2 Results from the application of Tool Service 22 (QSAR classification)

As mentioned above, the androgen receptor (AR) is one of the most important targets implicated in androgenetic alopecia. To demonstrate the ability of the QSAR Tool Service, an example will be given. A dataset reported by Zang et al. (2016) containing 206 active and 1,600 inactive compounds, in terms of AR binding ability was used. The dataset included 1,445 descriptors generated using the PADEL software. Using Tool Service 22, a data frame was created binding descriptors with outcome data (active/inactive) and CAS IDs. As the best way to validate the predictive ability of a model is to perform its external validation, the overall dataset was randomly divided into a training and a test set with a proportion 80% (1,442 compounds) -20% (364 compounds). Random forest and Naive Bayes algorithms were used to model the data.

ROC curves were employed to evaluate the performance of the two models. The area under the ROC curve (AUC) is commonly used to evaluate the performance of binary classification models, because of its simplicity and insensitivity to class skewness. The higher the area under the ROC curve the better the prediction of the specific model.

The results for the area under the ROC curve were:

	Methods	AUC Score
1	**Random Forest**	**0.9836588**
2	**Naive Bayes**	**0.7865459**

and can be visualised in Figure 6.6.1 below.

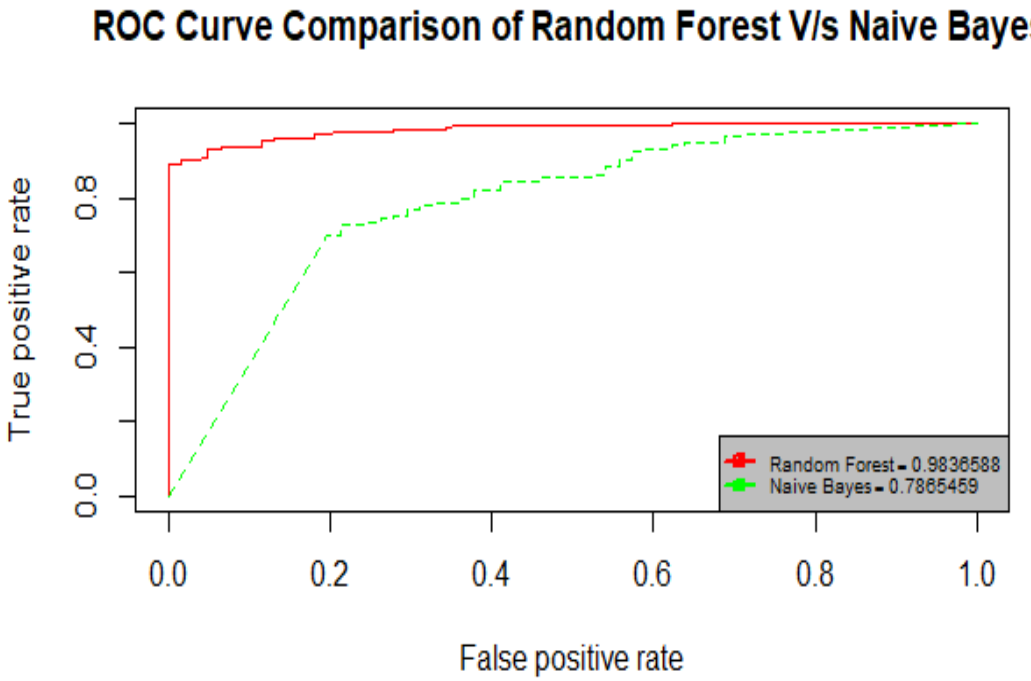


Figure 6.6.1: ROC curves for Random Forest and Naïve Bayes models for the AR dataset

In pattern recognition and classification, precision is the fraction of true positive results amongst all positives and sensitivity is the fraction of true positives amongst all results. These two fractions are considered a measure of relevance for our models. The F1 score takes into account both of these parameters. We could say that its the harmonic mean of accuracy and sensitivity. F1 scores approaching 1 correspond to a perfect test accuracy.

Classification Statistics were calculated based on the formulae below:

Accuracy	(True Positive+True Negative)/Total
Sensitivity	True Positive/(True Positive+False Negative)
Specificity	True Negative/(True Negative+False Positive)
Precision	True Positive/(True Positive+False Positive)
F1 score	2*(Precision*Sensitivity)/(Precision+Sensitivity)

and the results were:

	Methods	Accuracy	Sensitivity	Specificity	AUC	F1Score
1	Random Forest	0.9013333	0.4098361	0.9968153	0.9836588	0.5747126
2	Naive Bayes	0.7386667	0.7540984	0.7356688	0.7865459	0.5747126

In the context of predicting the binding ability of molecules to the AR (and assuming that positive cases represent molecules that bind to the AR), prioritising sensitivity can be crucial for the following reasons:

Importance of detecting true positives: Sensitivity, also known as the true positive rate or recall, measures the model's ability to correctly identify positive cases. In this scenario, identifying molecules that bind to the AR receptor (true positives) is of significant interest. False negatives, which represent molecules that actually bind to the AR but are misclassified as non-binding, could have potential implications in drug discovery or other applications. Prioritising sensitivity ensures that a higher proportion of true positive cases are correctly identified, reducing the chances of missing potential AR-binding molecules.

Cost of false negatives: False negatives occur when the model fails to identify molecules that actually bind to the AR receptor. In drug discovery or other molecular binding prediction tasks, false negatives can have serious consequences. It could result in missing potential lead compounds or inhibiting molecules, leading to delays in research or the development of ineffective drugs. Prioritising sensitivity helps minimize the risk of false negatives by aiming to capture as many true positives as possible.

Trade-off with specificity: While sensitivity focuses on correctly identifying positive cases, specificity measures the model's ability to correctly identify negative cases. In this case, negative cases would represent molecules that do not bind to the AR receptor. Although specificity is still important, prioritising it over sensitivity could result in a higher number of false negatives (misclassifying AR-binding molecules as non-binding), which is generally more undesirable than false positives (misclassifying non-binding molecules as binding).

Considering these reasons, prioritising sensitivity in this scenario of predicting the binding ability of molecules to the AR receptor would be beneficial to minimize the risk of missing potential AR-binding molecules and the associated implications in drug discovery or other

applications. Hence, although the RF model seems to outperform the Naive Bayes model in terms of accuracy, specificity and AUC, its low sensitivity would lead us to the assumption that the Naïve Bayes model will perform better in this case where correctly identifying positive cases is more crucial.

The most relevant descriptors are shown below in Table 6.6.4 and Figure 6.6.2:

Table 6.6.4: The molecular descriptors relevant to the RF model produced

Descriptor	Description	Class
BCUTc-1h	Represents the low highest partial charge weighted BCUTS	2D
piPC6	Conventional bond order ID number of order 6 (ln(1+x)	2D
MDEC-23	Molecular distance edge between all secondary and tertiary carbons	2D
MLFER_L	Solute gas-hexadecane partition coefficient	2D
AATSC1p	Average centered Broto-Moreau autocorrelation - lag 1 / weighted by polarizabilities	2D
SdssC	Sum of atom-type E-State: =C<	2D
maxwHBa	Maximum E-States for weak Hydrogen Bond acceptors	2D
SP-6	Simple path, order 6	2D
GATS1p	Geary autocorrelation - lag 1 / weighted by polarizabilities	2D
VE1_Dzv	Coefficient sum of the last eigenvector from Barysz matrix / weighted by van der Waals volumes	2D
VP-5	Valence path, order 5	2D
SP-7	Simple path, order 7	2D
maxaasC	Maximum atom-type E-State: :C:-	2D
maxaaCH	Maximum atom-type E-State: :CH:	2D
MWC4	Molecular walk count of order 4 (ln(1+x)	2D
BCUTp-1h	Represents the low highest polarizability weighted BCUTS	2D
GATS8c	Geary autocorrelation - lag 8 / weighted by charges	2D
VPC-5	Valence path cluster, order 5	2D
SpMax2_Bhp	Largest absolute eigenvalue of Burden modified matrix - n 2 / weighted by relative polarizabilities	2D
piPC9	Conventional bond order ID number of order 9 (ln(1+x)	2D
TpiPC	Total conventional bond order (up to order 10) (ln(1+x))	2D
piPC10	Conventional bond order ID number of order 10 (ln(1+x)	2D
GATS1c	Geary autocorrelation - lag 1 / weighted by charges	2D
MATS3s	Moran autocorrelation - lag 3 / weighted by I-state	2D
ETA_Alpha	Sum of alpha values of all non-hydrogen vertices of a molecule	2D
AATSC1i	Average centered Broto-Moreau autocorrelation - lag 1 / weighted by first ionisation potential	2D

GATS5e	Geary autocorrelation - lag 5 / weighted by Sanderson electronegativities	2D
SpMax2_Bhv	Largest absolute eigenvalue of Burden modified matrix - n 2 / weighted by relative van der Waals volumes	2D
SpMin2_Bhi	Smallest absolute eigenvalue of Burden modified matrix - n 2 / weighted by relative first ionisation potential	2D

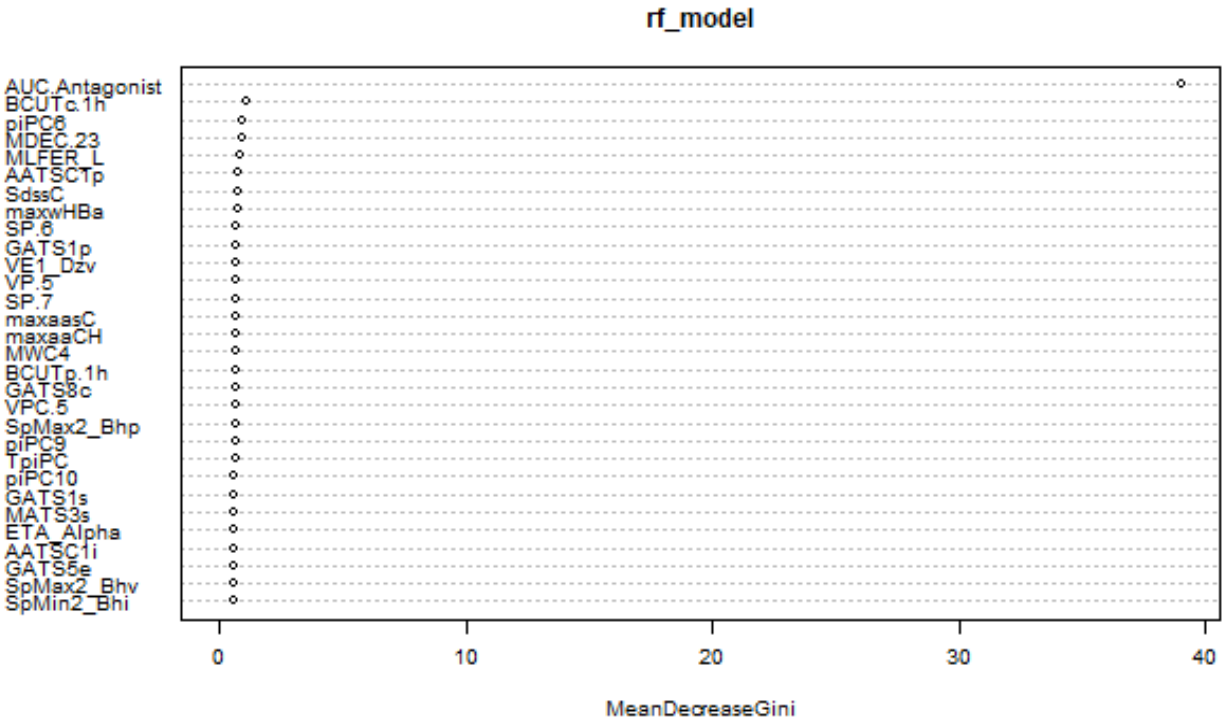


Figure. 6.6.2: A diagram of the most important molecular descriptors in the RF model as was produced by Tool Service 22

The QSAR classification model using Tool Service 22, has several potential impacts in the field of *in silico* drug discovery of ethnopharmacologically used plants. These impacts can be summarized as follows:

Predictive Ability: The model demonstrates a high level of predictive ability in identifying molecules with binding potential to the androgen receptor (AR). The calculated AUC scores indicate strong performance of both the random forest and Naive Bayes models. This suggests that the model could serve as a reliable tool for predicting the AR binding ability of compounds, aiding in the identification of potential lead compounds or drug candidates.

Efficient Screening: By utilising the QSAR model, researchers can effectively screen a large number of compounds to prioritize those with a higher likelihood of binding to the AR. This can significantly reduce the time and cost associated with experimental screening, as compounds can be prioritized for further investigation based on their predicted binding ability.

Identification of Relevant Molecular Descriptors: The model identifies the most relevant molecular descriptors associated with the AR binding ability. These descriptors provide valuable insights into the structural features and properties that contribute to the interaction between compounds and the AR receptor. This information can guide further research and optimisation of compounds to enhance their binding affinity and therapeutic potential.

Optimisation of Drug Discovery Process: The model offers a means to optimize the drug discovery process by integrating computational methods early on. By utilising *in silico* tools like QSAR, researchers can prioritize compounds with higher binding potential, reducing the number of compounds that need to be synthesized and experimentally tested. This can accelerate the drug discovery process and increase the chances of identifying promising candidates for further development.

Risk Mitigation: Prioritising sensitivity in the model helps minimize the risk of false negatives, ensuring that potential AR-binding molecules are not missed. This is particularly important in drug discovery, as false negatives can lead to the exclusion of potentially effective compounds, delaying research progress and hindering the development of novel therapeutics.

Overall, the impact of this model lies in its potential to enhance the efficiency, accuracy, and cost-effectiveness of drug discovery efforts targeting the androgen receptor. By leveraging computational approaches and predictive modeling, researchers can focus their resources on compounds with higher binding potential, leading to more targeted and successful drug development processes**.**

Summary of Tool Services 21 and 22

Tool Service 21 utilized QSAR-DB predictions to determine various properties of the 34 phytochemicals, including melting point and Abraham's descriptors.

In drug discovery processes, the prediction of melting points plays a significant role. Accurate melting point predictions allow researchers to assess the solid-state properties and thermal stability of compounds. This information is vital for formulation development, process optimisation, and determining appropriate storage conditions. By incorporating melting point predictions into the tool, researchers gain an additional tool to evaluate compound properties and make informed decisions in the drug discovery process.

The Abraham's descriptors, which are important in predicting drug transport through biological barriers, were also determined for each phytochemical. By analysing these descriptors, researchers can assess the drug-likeness and suitability of the compounds for further development. Additionally, Tool Service 21 provided information on the Abraham's solvent coefficients. These coefficients can offer insights into the behavior of the compounds in biological systems, helping researchers make informed decisions about their potential as drug candidates.

By utilising the predicted Abrahams descriptors and solvent coefficients, we can narrow down their focus to compounds with desirable physicochemical properties and drug-like characteristics. This approach saves time and resources in the drug discovery process by enabling the screening and prioritisation of large compound libraries. Furthermore, these descriptors and coefficients can guide medicinal chemists in designing and modifying molecules to improve their drug properties, such as enhancing solubility, permeability, and target interactions.

In the second part of this section, Tool Service 22 was applied to a dataset of compounds with known androgen receptor (AR) binding ability. The dataset included both active and inactive compounds, and 1,445 descriptors were generated using the PADEL software. The data was divided into training and test sets, and random forest and Naive Bayes algorithms were used to model the data. The performance of the models was evaluated using ROC curves, which plot the true positive rate against the false positive rate. The area under the ROC curve (AUC) was calculated to assess the prediction performance of the models. The random forest model achieved an AUC score of 0.9836588, , while the Naive Bayes model had an AUC score of 0.7865459.

Classification statistics, including accuracy, sensitivity, specificity, precision, and F1 score, were calculated to further evaluate the models. Accuracy measures the overall correctness of the predictions, sensitivity quantifies the ability to correctly identify positive instances, specificity measures the ability to correctly identify negative instances, precision represents the proportion of true positives among positive predictions, and the F1 score combines precision and sensitivity. These metrics provide a comprehensive evaluation of the models' performance.

CHAPTER 7: Conclusions and future work

7.1 Summary of the main findings

The purpose of this thesis was to develop computational tools that will improve various drug discovery and development procedures enabling us to leverage the wealth of traditionally used natural products in ethnopharmacology, and suggest novel and safe lead compounds for the treatment of androgenetic alopecia.

More than a third of the 69 plants studied belong to four taxonomic families that share many similarities in terms of phylogeny and phytochemical content.

Structural similarity studies revealed fifteen pairs with significant (Tanimoto > 0.85) resemblances between the phytochemicals. No significant structural similarities of phytochemicals and currently used drugs in AGA were observed. Interestingly many of the phytochemicals exhibited similarities in structure with UV protectant, vascular protectant and anti-inflammatory drugs. Indeed, UV exposure, inflammation, and dermal microcirculation have all been implicated in different theories of AGA's pathogenesis.

Analyses of alopecia as a side effect and as an adverse drug reaction confirmed literature results on suspected drug groups. These unwanted 'off-target' drug effects can be attributed to genes that may play a role in the apoptosis of hair and steroid metabolism.

Gene-disease association networks revealed a high correlation of the chosen genes with hypotrichoses (diseases that share a similar phenotype to AGA) and to a lesser extent, with prostatic cancer. Gene-Pathways network studies revealed significant links of the chosen genes with three signalling pathways (Wnt, Notch and Hedgehog) which are important in hair development and growth.

A QSAR model was developed and validated to predict the binding abilities of phytochemicals to the androgen receptor (AR). The Random Forest model demonstrated a high area under the ROC curve score of 0.9836, indicating strong overall performance. However, it exhibited limitations in accurately identifying phytochemicals that bind to the AR. In contrast, the Naïve Bayes algorithm-based model showed a reasonable balance between sensitivity and

specificity, suggesting its potential for effectively predicting AR binding of phytochemicals. Results were also confirmed through molecular docking studies. Ten of the thirty-four phytochemicals showed similar binding interaction for the binding site of the androgen receptor to drugs currently used in androgenetic alopecia and can potentially serve as leads for the design of novel AGA treatments.

7.2 Overall Conclusions

This thesis has made significant advancements in both *in silico* tools for drug discovery and treatments for androgenetic alopecia. The developed computational tool, which automates processes for ethnopharmacological data retrieval, enhances the identification and taxonomy of plants through the use of DNA barcodes, and effectively addresses nomenclature complexities, surpasses previous approaches (Padhy et al., 2023). By integrating various methodologies and functionalities for identification of leads and targets and testing, this tool stands out from existing tools in this field (Shaker et al., 2021).

Moreover, the research findings on potential phytochemical leads for androgenetic alopecia, including their structural resemblances to UV screens, vascular protectants, and anti-inflammatory drugs, provide novel insights into the mechanisms of AGA pathogenesis. These findings can greatly contribute to our current understanding of the disease (Kinoshita-Ise, Fukuyama, and Ohyama, 2023).

The contributions of this thesis exemplify how it has significantly advanced research in the fields of *in silico* drug discovery of natural products and the treatment of androgenetic alopecia. By introducing an innovative and comprehensive computational tool with enhanced functionalities and providing valuable insights into the disease mechanisms, this thesis paves the way for further advancements in the field.

7.3 Future Work

The results of this thesis offer significant potential to advance to both *in vitro* and *in vivo* experimentation for at least ten phytochemical leads. The developed tool enables the identification of new lead compounds from natural products, providing opportunities for further investigation in laboratory settings.

Other steps include the expansion of current work to other diseases and the addition of more capabilities in the computational tool itself. The plan is to focus on *in silico* drug discovery for other dermatological conditions while also integrating other tools which will allow us to:

- study the synergistic effects of phytochemicals in plant extracts
- target 'undruggable' proteins via their miRNA gene regulators and propose new treatments that currently seem impossible to cure
- accurately estimate the thermodynamics and kinetics associated with drug-target recognition and binding by using Molecular Dynamics simulations
- introduce deep learning techniques to design novel compounds with desired activities.

7.4 Overall impact of this thesis

In conclusion, this thesis contributes to the field of *in silico* drug discovery, harnessing the power of ethnopharmacological knowledge to address the pressing issue of androgenetic alopecia. Through the development of an innovative tool utilising the R language, users may successfully unlock a wealth of new potential lead compounds from natural products deeply rooted in traditional medicinal practices. Furthermore, the profound research findings provide invaluable insights into unravelling the underlying causes of androgenetic alopecia. Moreover, the thesis sets a visionary course for the future, envisioning the extension of the tool's capabilities to combat a plethora of other diseases and incorporating advanced methodologies like deep learning and molecular dynamics simulations. As such, this thesis attempts to pave the pathway for enhanced *in silico* drug discovery and transcend the boundaries of traditional medicine.

Bibliography

Acharya, S., Yao, J., Li, P., Zhang, C., Lowery, F.J., Zhang, Q., Guo, H., Qu, J., Yang, F., Wistuba, I.I. and Piwnica-Worms, H., 2019. Sphingosine kinase 1 signaling promotes metastasis of triple-negative breast cancer. *Cancer research*, *79*(16), pp.4211-4226.

Ahmad, W., ul Haque, M.F., Brancolini, V., Tsou, H.C., Ul Haque, S., Lam, H., Aita, V.M., Owen, J., Deblaquiere, M., Frank, J. and Cserhalmi-Friedman, P.B., 1998. Alopecia universalis associated with a mutation in the human hairless gene. *Science*, *279*(5351), pp.720-724.

Ahn, Y., Sims, C., Logue, J.M., Weatherbee, S.D. and Krumlauf, R., 2013. Lrp4 and Wise interplay controls the formation and patterning of mammary and other skin appendage placodes by modulating Wnt signalling. *Development*, *140*(3), pp.583-593.

Ahouansou, S., Le Toumelin, P., Crickx, B. and Descamps, V., 2007. Association of androgenetic alopecia and hypertension. *European Journal of Dermatology*, *17*(3), pp.220-222.

Ali, S. and Singh, G., 2010. Radiation-induced alopecia. *International Journal of Trichology*, *2*(2), p.118.

Alonso, L., Okada, H., Pasolli, H.A., Wakeham, A., You-Ten, A.I., Mak, T.W. and Fuchs, E., 2005. Sgk3 links growth factor signalling to maintenance of progenitor cells in the hair follicle. *The Journal of Cell Biology*, *170*(4), pp.559-570.

Alonso, L. and Fuchs, E., 2006. The hair cycle. *Journal of Cell Science*, *119*(3), pp.391-393.

Amory, J.K., Wang, C., Swerdloff, R.S., Anawalt, B.D., Matsumoto, A.M., Bremner, W.J., Walker, S.E., Haberer, L.J. and Clark, R.V., 2007. The effect of 5α-reductase inhibition with dutasteride and finasteride on semen parameters and serum hormones in healthy men. *The Journal of Clinical Endocrinology & Metabolism*, *92*(5), pp.1659-1665.

Amunategui, M 2015, Github, accessed 1 February 2016, <https://github.com/amunategui/pubmed-query/blob/master/pubmed-query.R>

Anastassakis, K., 2022. *Androgenetic Alopecia from A to Z: Vol. 2 Drugs, Herbs, Nutrition and Supplements*. Springer Nature, p.43.

Andl, T., Reddy, S.T., Gaddapara, T. and Millar, S.E., 2002. WNT signals are required for the initiation of hair follicle development. *Developmental Cell*, *2*(5), pp.643-653.

Anitua, E., Pino, A., Martinez, N., Orive, G. and Berridi, D., 2017. The effect of plasma rich in growth factors on pattern hair loss: a pilot study. *Dermatologic Surgery*, *43*(5), pp.658-670.

Ardic, U.A. and Ercan, E.S., 2017. Resolution of methylphenidate osmotic release oral system-induced hair loss in two siblings after dose escalation. *Pediatrics International*, *59*(11), pp.1217-1218.

Arias-Santiago, S., Arrabal-Polo, M.A., Buendía-Eisman, A., Arrabal-Martín, M., Gutiérrez-Salmerón, M.T., Girón-Prieto, M.S., Jimenez-Pacheco, A., Calonje, J.E., Naranjo-Sintes, R., Zuluaga-Gomez, A. and Ortega, S.S., 2012. Androgenetic alopecia as an early marker of benign prostatic hyperplasia. *Journal of the American Academy of Dermatology*, *66*(3), pp.401-408.

Badri T, Kumar D D. Minoxidil. [Updated 2019 Oct 9]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2019 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482378/

Bajorath, J., 2015. Computer-aided drug discovery. *F1000Research*, *4*.

Baker, C 2019. patentsview: An R Client to the 'PatentsView' API. R package version 0.2.2, ac cessed 5 June 2019 <https://CRAN.R-project.org/package=patentsview>
Banbury, B.L. and O'Meara, B.C., 2014. Reol: R interface to the Encyclopaedia of Life. *Ecology and evolution*, *4*(12), pp.2577-2583.

Baran, R., 1989. Explosive eruption of pyogenic granuloma on the scalp due to topical combi nation therapy of minoxidil and retinoic acid. *Dermatology*, *179*(2), pp.76-78.

Barman, J.M., 1964. Method, technique and computations in the study of the trophic state of the human hair. *Journal of Investigative Dermatology*, *42*, pp.421-425.

Barter, D.A., 1989. Hair loss in a child associated with naproxen. *BMJ: British Medical Journal*, *298*(6669), p.325.

Bartosová, L., Jorda, V. and Stáva, Z., 1984. *Diseases of the hair and the scalp*. Karger Publishers.

Bassino, E., Antoniotti, S., Gasparri, F. and Munaron, L., 2016. Effects of flavonoid derivatives on human microvascular endothelial cells. *Natural product research*, *30*(24), pp.2831-2834.

Bassino, E., Gasparri, F. and Munaron, L., 2020. Protective Role of Nutritional Plants Containing Flavonoids in Hair Follicle Disruption: A Review. *International Journal of Molecular Sciences*, *21*(2), p.523.

Batrinos, M.L., 2014. The endocrinology of baldness. *Hormones*, *13*(2), pp.197-212.

Bazzano, G.S., Terezakis, N. and Galen, W., 1986. Topical tretinoin for hair growth promotion. *Journal of the American Academy of Dermatology*, *15*(4), pp.880-893.

Benigni, R. and Bossa, C., 2008. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. *Mutation Research/Reviews in Mutation Research*, *659*(3), pp.248-261.

Bergfeld, W., 2009. Diffuse hair loss: its triggers and management. *Cleveland Clinic journal of medicine*, *76*(6), p.361.

Bergfeld, W.F., 1988. Etiology and diagnosis of androgenetic alopecia. *Clinics in dermatology*, *6*(4), pp.102-107.

Bergfeld, W.F., 1995. Androgenetic alopecia: an autosomal dominant disorder. *The American journal of medicine*, *98*(1), pp. S95-S98.

Biggs, L.C. and Mikkola, M.L., 2014, January. Early inductive events in ectodermal appendage morphogenesis. In *Seminars in cell & developmental biology* (Vol. 25, pp. 11-21). Academic Press.

Bilgin, Y.Y., Yagci, I., Tunc, S. and Atagun, M.I., 2015. Hair loss due to aripiprazole use: a case report. *The European Research Journal*, *1*(2), pp.85-87.

Birch, M.P., Messenger, J.F. and Messenger, A.G., 2001. Hair density, hair diameter and the prevalence of female pattern hair loss. *British Journal of Dermatology*, *144*(2), pp.297-304.

Blain, R. and Kazantzis, G., 2015. Thallium. In *Handbook on the Toxicology of Metals* (pp. 1229-1240). Academic Press.

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L. and Fuchs, E., 2004. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell*, *118*(5), pp.635-648.

Blumeyer, A., Tosti, A., Messenger, A., Reygagne, P., Del Marmol, V., Spuls, P.I., Trakatelli, M., Finner, A., Kiesewetter, F., Trüeb, R. and Rzany, B., 2011. Evidence-based (S3) guideline for the treatment of androgenetic alopecia in women and in men. *JDDG: Journal der Deutschen Dermatologischen Gesellschaft*, *9*, pp. S1-S57.

Booth, C. and Potten, C.S., 2000. Keratinocyte growth factor increases hair follicle survival following cytotoxic insult. *Journal of investigative dermatology*, *114*(4), pp.667-673.

Borum, M.L. and Cannava, M., 1997. Diffuse alopecia associated with omeprazole. *The American journal of gastroenterology*, *92*(9), pp.1576-1576.

Bosticardo, M., Yamazaki, Y., Cowan, J., Giardino, G., Corsino, C., Scalia, G., Prencipe, R., Ruffner, M., Hill, D.A., Sakovich, I. and Yemialyanava, I., 2019. Heterozygous FOXN1 variants cause low TRECs and severe T cell lymphopenia, revealing a crucial role of FOXN1 in supporting early thymopoiesis. *The American Journal of Human Genetics*, *105*(3), pp.549-561.

Botchkarev, V.A, 2003. Stress and the hair follicle. Exploring the connections. *The American journal of Pathology.* 162(3), p.709-712

Botchkarev, V.A., Botchkareva, N.V., Roth, W., Nakamura, M., Chen, L.H., Herzog, W., Lindner, G., McMahon, J.A., Peters, C., Lauster, R. and McMahon, A.P., 1999. Noggin is a mesenchymally derived stimulator of hair-follicle induction. *Nature cell biology*, *1*(3), p.158.

Botchkareva, N.V., Botchkarev, V.A. and Gilchrest, B.A., 2003, June. Fate of melanocytes during development of the hair follicle pigmentary unit. In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 8, No. 1, pp. 76-79). Elsevier.

Bourgeois, J.A., 1996. Two cases of hair loss after sertraline use. *Journal of clinical psychopharmacology*, *16*(1), pp.91-92.

Bradley, J.C., Abraham, M.H., Acree, W.E. and Lang, A.S., 2015. Predicting Abraham model solvent coefficients. *Chemistry Central Journal*, *9*(1), pp.1-10.

Braun, L., Tiralongo E., Wilkinson J.M, Spitzer O., et al. (2010). Perceptions, use and attitudes of pharmacy customers on complementary therapy and pharmacy practice. *BMC Complementary and Alternative Medicine, 10,38.*

Brockschmidt, F.F., Heilmann, S., Ellis, J.A., Eigelshoven, S., Hanneken, S., Herold, C., Moebus, S., Alblas, M.A., Lippke, B., Kluck, N. and Priebe, L., 2011. Susceptibility variants on chromosome 7p21. 1 suggest HDAC9 as a new candidate gene for male-pattern baldness. *British Journal of Dermatology*, *165*(6), pp.1293-1302.

Brownstein, M.H., 1979. Trichilemmal horn: cutaneous horn showing trichilemmal keratinization. *British Journal of Dermatology*, *100*(3), pp.303-309.

Buhl, A.E., Waldon, D.J., Baker, C.A. and Johnson, G.A., 1990. Minoxidil sulfate is the active metabolite that stimulates hair follicles. *Journal of Investigative Dermatology*, *95*(5), pp.553-557.

Burg, D., Yamamoto, M., Namekata, M., Haklani, J., Koike, K. and Halasz, M., 2017. Promotion of anagen, increased hair density and reduction of hair fall in a clinical setting following identification of FGF5-inhibiting compounds via a novel 2-stage process. *Clinical, cosmetic and investigational dermatology*, *10*, p.71.

Callahan, C.A., Ofstad, T., Horng, L., Wang, J.K., Zhen, H.H., Coulombe, P.A. and Oro, A.E., 2004. MIM/BEG4, a Sonic hedgehog-responsive gene that potentiates Gli-dependent transcription. *Genes & development*, *18*(22), pp.2724-2729.

Cao, Y., Charisi, A., Cheng, L.C., Jiang, T. and Girke, T., 2008. ChemmineR: a compound mining framework for R. *Bioinformatics*, *24*(15), pp.1733-1734.

Carlson, M., 2016. UniProt. ws: A package for retrieving data from the UniProt web service. *R Package version 2.18*, pp.1-3.

Carruthers, R., 1968. Contraceptives and Alopecia. *British medical journal*, *2*(5596), p.52.

Cayuela, L Stein A, and Oksanen J 2017. Taxonstand: Taxonomic Standardization of Plant Spe cies Names. R package version 2.1., accessed 22 March 2018 <https://CRAN.R-project.org/package=Taxonstand>

Cecchin, E., De Mattia, E., Mazzon, G., Cauci, S., Trombetta, C. and Toffoli, G., 2014. A pharmacogenetic survey of androgen receptor (CAG) n and (GGN) n polymorphisms in patients experiencing long term side effects after finasteride discontinuation. *The International journal of biological markers*, *29*(4), pp.310-316.

Chamberlain, S., 2017a. bold: interface to Bold Systems' API'. R package version 0.4. 0.

Chamberlain, S., 2017b. ritis: Integrated Taxonomic Information System. Available from: <https://cran.r-project.org/web/packages/ritis/ritis.pdf>

Chamberlain, S., Szoecs, E., Foster, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J. and Oksanen, J., 2017. Package 'taxize'. *Taxonomic Information from Around th e Web*.

Chen, B., Choi, H., Hirsch, L.J., Moeller, J., Javed, A., Kato, K., Legge, A., Buchsbaum, R. and Detyniecki, K., 2015. Cosmetic side effects of antiepileptic drugs in adults with epilepsy. *Epilepsy & Behavior*, *42*, pp.129-137.

Chuang, Y.C., Chang, W.N., Chen, I.L., Yang, J.Y., Ho, J.C. and Kuo, H.W., 2002. Topiramate-Induced Hair Loss: Case Report. *Dermatology and Psychosomatics/Dermatologie und Psychosomatik*, *3*(4), pp.183-184.

Chumlea, W.C., Rhodes, T., Girman, C.J., Johnson-Levonas, A., Lilly, F.R., Wu, R. and Guo, S.S., 2004. Family history and risk of hair loss. *Dermatology*, *209*(1), pp.33-39.

Chuong, C.M. ed., 1998. *Molecular basis of epithelial appendage morphogenesis* (Vol. 1). Landes Bioscience.

Concha, J.S.S. and Werth, V.P., 2018. Alopecias in lupus erythematosus. *Lupus science & medicine*, *5*(1), p.e000291.

Cotsarelis, G., 2006. Epithelial stem cells: a folliculocentric view. *Journal of Investigative Dermatology*, *126*(7), pp.1459-1468.

Cranwell, W. and Sinclair, R., 2016. Male androgenetic alopecia. In *Endotext [Internet]*. MDText. com, Inc.

Crum Brown, A. and Fraser, T.R., 1868. On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia. *Journal of anatomy and physiology*, *2*(2), p.224.

D'Amico, A.V. and Roehrborn, C.G., 2007. Effect of 1 mg/day finasteride on concentrations of serum prostate-specific antigen in men with androgenic alopecia: a randomised controlled trial. *The Lancet Oncology*, *8*(1), pp.21-25.

Danilenko, D.M., Ring, B.D. and Pierce, G.F., 1996. Growth factors and cytokines in hair follicle development and cycling: recent insights from animal models and the potentials for clinical therapy. *Molecular Medicine Today*, *2*(11), pp.460-467.

D'Arcy, P.F. and Gerber, D., 1987. Adverse Reactions of Piroxicam. *Drug Intelligence & Clinical Pharmacy*, *21*(9), pp.707-710.

Das Gupta, R. and Fuchs, E., 1999. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development*, *126*(20), pp.4557-4568.

Datta, K., Singh, A.T., Mukherjee, A., Bhat, B., Ramesh, B. and Burman, A.C., 2009. Eclipta alba extract with potential for hair growth promoting activity. *Journal of ethnopharmacology*, *124*(3), pp.450-456.

Dawber, R.P.R (ed.), 1997. *Diseases of the Hair and Scalp* (3rd ed.).Oxford, U.K: Blackwell science.

Dawber, R.P., 1988. The embryology and development of human scalp hair. *Clinics in dermatology*, *6*(4), pp.1-6.

De Villez, R.L., 1986. *The growth and loss of hair*. The Upjohn Co. Kalamazoo, Michigan.

Dhingra, N. and Bhagwat, D., 2011. Benign prostatic hyperplasia: An overview of existing treatment. *Indian journal of pharmacology*, *43*(1), p.6.

Diamanti-Kandarakis, E., 1999. Current aspects of antiandrogen therapy in women. *Current pharmaceutical design*, *5*, pp.707-724.

Dinh, Q.Q. and Sinclair, R., 2007. Female pattern hair loss: current treatment concepts. *Clinical interventions in aging*, *2*(2), p.189.

Directive, O.J., 2010. 63/EU of the European Parliament and of the Council on the protection of animals used for scientific purposes. *Official Journal of the European Union*, *276*, p.56.

Doucet, J.P. and Weber, J., 1996. *Computer-aided molecular design: theory and applications*. Elsevier.

Du, X., Tabeta, K., Hoebe, K., Liu, H., Mann, N., Mudd, S., Crozat, K., Sovath, S., Gong, X. and Beutler, B., 2004. Velvet, a dominant Egfr mutation that causes wavy hair and defective eyelid development in mice. *Genetics*, *166*(1), pp.331-340.

Ebling, F.J., 1987. Biology of hair follicles. *Dermatology in general medicine*, pp.213-219.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A. and McMahon, A.P., 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell*, *75*(7), pp.1417-1430.

Edenberg, H.J. and Liu, Y., 2009. Laboratory methods for high-throughput genotyping. *Cold Spring Harbor Protocols*, *2009*(11), pp.pdb-top62.

Ehrlich, H.C. and Rarey, M., 2012. Systematic benchmark of substructure search in molecular graphs-From Ullmann to VF2. *Journal of cheminformatics*, *4*(1), p.13.

Ekor, M., 2014. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. *Frontiers in Pharmacology*, *4*, p.177.

Elliott, K., Messenger, A.G. and Stephenson, T.J., 1999. Differences in hair follicle dermal papilla volume are due to extracellular matrix volume and cell number: implications for the control of hair follicle size and androgen responses. *Journal of investigative dermatology*, *113*(6), pp.873-877.

Ellis, J.A., Sinclair, R. and Harrap, S.B., 2002. Androgenetic alopecia: pathogenesis and potential for therapy. *Expert reviews in molecular medicine*, *4*(22), pp.1-11.

Enoch, S.J., Madden, J.C. and Cronin, M.T.D., 2008. Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR and QSAR in Environmental Research*, *19*(5-6), pp.555-578.

Erdogan, B., 2017. Anatomy and Physiology of Hair. In *Hair and Scalp Disorders*. IntechOpen.

Evans, L.M., Cowey, S.L., Siegal, G.P. and Hardy, R.W., 2009. Stearate preferentially induces apoptosis in human breast cancer cells. *Nutrition and cancer*, *61*(5), pp.746-753.

Fabbrocini, G., Cantelli, M., Masarà, A., Annunziata, M.C., Marasca, C. and Cacciapuoti, S., 2018. Female pattern hair loss: a clinical, pathophysiologic, and therapeutic review. *International journal of women's dermatology*, *4*(4), pp.203-211.

Farah, M.H., Edwards, R., Lindquist, M., Leon, C. and Shaw, D., 2000. International monitoring of adverse health effects associated with herbal medicines. *Pharmacoepidemiology and Drug Safety*, *9*(2), pp.105-112.

Favier, B., Fliniaux, I., Thélu, J., Viallet, J.P., Demarchez, M., Jahoda, C.A. and Dhouailly, D., 2000. Localisation of members of the notch system and the differentiation of vibrissa hair follicles: receptors, ligands, and fringe modulators. *Developmental dynamics: an official publication of the American Association of Anatomists*, *218*(3), pp.426-437.

Fawzi, M.M., Mahmoud, S.B., Ahmed, S.F. and Shaker, O.G., 2016. Assessment of vitamin D receptors in alopecia areata and androgenetic alopecia. *Journal of cosmetic dermatology*, *15*(4), pp.318-323.

FDA Adverse Reporting System (FAERS), accessed 15 March 2016 <https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-latest-quarterly-data-files>

Feinerer, I and Hornik, K 2018, tm: Text Mining Package. R package version 0.7-6. accessed 1 8 February 2018 <https://CRAN.R-project.org/package=tm>

Ferry, J.J., Forbes, K.K., VanderLugt, J.T. and Szpunar, G.J., 1990. Influence of tretinoin on the percutaneous absorption of minoxidil from an aqueous topical solution. *Clinical Pharmacology & Therapeutics*, *47*(4), pp.439-446.

Gai, W.T., Yu, D.P., Wang, X.S. and Wang, P.T., 2016. Anti-cancer effect of ursolic acid activates apoptosis through ROCK/PTEN mediated mitochondrial translocation of cofilin-1 in prostate cancer. *Oncology Letters*, *12*(4), pp.2880-2885.

Gan, D.C. and Sinclair, R.D., 2005, December. Prevalence of male and female pattern hair loss in Maryborough. In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 10, No. 3, pp. 184-189). Elsevier.

Garcia AM, McLaren Jr CE, Meyskens FL. Melanoma: is hair the root of the problem? Pigment cell & melanoma research. 2011 Feb;24(1):110.

Garza, L.A., Liu, Y., Yang, Z., Alagesan, B., Lawson, J.A., Norberg, S.M., Loy, D.E., Zhao, T., Blatt, H.B., Stanton, D.C. and Carrasco, L., 2012. Prostaglandin D2 inhibits hair growth and is

elevated in bald scalp of men with androgenetic alopecia. *Science translational medicine*, *4*(126), pp.126ra34-126ra34.

Gasteiger, J. and Engel, T. eds., 2006. *Chemoinformatics: a textbook*. John Wiley & Sons.

Gat, U., DasGupta, R., Degenstein, L. and Fuchs, E., 1998. De novo hair follicle morphogenesis and hair tumors in mice expressing a truncated β-catenin in skin. *Cell*, *95*(5), pp.605-614.

Godden, J.W., Stahura, F.L. and Bajorath, J., 2000. Variability of molecular descriptors in compound databases revealed by Shannon entropy calculations. *Journal of chemical information and computer sciences*, *40*(3), pp.796-800.

Grant, B.J., Rodrigues, A.P., ElSawy, K.M., McCammon, J.A. and Caves, L.S., 2006. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, *22*(21), pp.2695-2696.

Greco, V., Chen, T., Rendl, M., Schober, M., Pasolli, H.A., Stokes, N., dela Cruz-Racelis, J. and Fuchs, E., 2009. A two-step mechanism for stem cell activation during hair regeneration. *Cell stem cell*, *4*(2), pp.155-169.

Grisoni, F., Consonni, V. and Todeschini, R., 2018. Impact of molecular descriptors on computational models. In *Computational Chemogenomics* (pp. 171-209). Humana Press, New York, NY.

Groppe, J., Greenwald, J., Wiater, E., Rodriguez-Leon, J., Economides, A.N., Kwiatkowski, W., Affolter, M., Vale, W.W., Belmonte, J.C.I. and Choe, S., 2002. Structural basis of BMP signalling inhibition by the cystine knot protein Noggin. *Nature*, *420*(6916), p.636.

Grover, C. and Khurana, A., 2013. Telogen effluvium. *Indian journal of dermatology, venereology and leprology*, *79*(5), p.591.

Guha, R 2018, rpubchem: An Interface to the PubChem Collection, accessed 17 March 2018 < https://github.com/rajarshi/cdk>

Guha, R. and Willighagen, E., 2012. A survey of quantitative descriptions of molecular structure. *Current topics in medicinal chemistry*, *12*(18), pp.1946-1956.

Guha, R., 2018. Fingerprint: functions to operate on binary fingerprint data.

Guha, R., Charlop-Powers, Z., Schymanski, E. and Charlop-Powers, M.Z., 2017. Package 'rcdk'.

Guin, D., Rani, J., Singh, P., Grover, S., Bora, S., Talwar, P., Karthikeyan, M., Satyamoorthy, K., Adithan, C., Ramachandran, S. and Saso, L., 2019. Global text mining and development of pharmacogenomic knowledge resource for precision medicine. *Frontiers in Pharmacology*, *10*.

Guo, X. and Wang, X.F., 2009. Signaling cross-talk between TGF-β/BMP and other pathways. *Cell research*, *19*(1), p.71.

Gupta, A.K. and Carviel, J., 2016. A mechanistic model of platelet-rich plasma treatment for androgenetic alopecia. *Dermatologic Surgery*, *42*(12), pp.1335-1339.

Gupta, S. and Major, L.F., 1991. Hair loss associated with fluoxetine. *The British Journal of Psychiatry*, *159*(5), pp.737-738.

Gupta, S. and Masand, P.S., 2000. Citalopram and Hair Loss. *Primary care companion to the Journal of clinical psychiatry*, *2*(2), pp.61-62.

Gustavsen, J.A., Pai, S., Isserlin, R., Demchak, B. and Pico, A.R., 2019. RCy3: Network biology using Cytoscape from within R. *F1000Research*, *8*.

Hajheydari, Z. and Makhlogh, A., 2008. Cutaneous and Mucosal Manifestations in Patients on Maintenance Hemodialysis A Study of 101 Patients in Sari, Iran.:86-90

Hagemann, T., Schlütter-Böhmer, B., Allam, J.P., Bieber, T. and Novak, N., 2005. Positive lymphocyte transformation test in a patient with allergic contact dermatitis of the scalp after short-term use of topical minoxidil solution. *Contact Dermatitis*, *53*(1), pp.53-55.

Hagenaars, S.P., Hill, W.D., Harris, S.E., Ritchie, S.J., Davies, G., Liewald, D.C., Gale, C.R., Porteous, D.J., Deary, I.J. and Marioni, R.E., 2017. Genetic prediction of male pattern baldness. *PLoS genetics*, *13*(2), p. e1006594.

Hamilton, J.B., 1942. Male hormone stimulation is prerequisite and an incitant in common baldness. *American Journal of Anatomy*, *71*(3), pp.451-480.

Hamilton, J.B., 1951. Patterned loss of hair in man: types and incidence. *Annals of the New York Academy of Sciences*, *53*(3), pp.708-728.

Hammett, L.P., 1937. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *Journal of the American Chemical Society*, *59*(1), pp.96-103.

Hann, M., Hudson, B., Lewell, X., Lifely, R., Miller, L. and Ramsden, N., 1999. Strategic pooling of compounds for high-throughput screening. *Journal of chemical information and computer sciences*, *39*(5), pp.897-902.

Hansch, C., Leo, A. and Hoekman, D.H., 1995. *Exploring QSAR: fundamentals and applications in chemistry and biology* (Vol. 557). Washington, DC: American Chemical Society.

HapMap Consortium, T.I., 2005. A haplotype map of the human genome. *Nature*, *437*(7063), pp.1299-320.

Hardy, M.H., 1992. The secret life of the hair follicle. *Trends in Genetics*, *8*(2), pp.55-61.

Harel, S., Higgins, C.A., Cerise, J.E., Dai, Z., Chen, J.C., Clynes, R. and Christiano, A.M., 2015. Pharmacologic inhibition of JAK-STAT signaling promotes hair growth. *Science Advances*, *1*(9), p.e1500973.

Harkey, M.R., 1993. Anatomy and physiology of hair. *Forensic science international*, *63*(1-3), pp.9-18.

Harmon, C.S. and Nevins, T.D., 1993. IL-1 alpha inhibits human hair follicle growth and hair fiber production in whole-organ cultures. *Lymphokine and cytokine research*, *12*(4), pp.197-203.

Harvey, A.L. and Waterman, P.G., 1998. The continuing contribution of biodiversity to drug discovery. *Current Opinion in Drug Discovery & Development*, *1*(1), pp.71-76.

Hattori, M., Tanaka, N., Kanehisa, M. and Goto, S., 2010. SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic acids research*, *38*(suppl_2), pp.W652-W656.

He, H., Xie, B. and Xie, L., 2018. Male pattern baldness and incidence of prostate cancer: A systematic review and meta-analysis. *Medicine*, *97*(28).

Heilmann-Heimbach, S., Herold, C., Hochfeld, L.M., Hillmer, A.M., Nyholt, D.R., Hecker, J., Javed, A., Chew, E.G., Pechlivanis, S., Drichel, D. and Heng, X.T., 2017. Meta-analysis identifies novel risk loci and yields systematic insights into the biology of male-pattern baldness. *Nature communications*, *8*, p.14694.

Heilmann-Heimbach, S., Herold, C., Hochfeld, L.M., Hillmer, A.M., Nyholt, D.R., Hecker, J., Javed, A., Chew, E.G., Pechlivanis, S., Drichel, D. and Heng, X.T., 2017. Meta-analysis identifies novel risk loci and yields systematic insights into the biology of male-pattern baldness. *Nature communications*, *8*, p.14694.

Helfer, E.L., Miller, J.L. and Rose, L.I., 1988. Side-effects of spironolactone therapy in the hirsute woman. *The Journal of Clinical Endocrinology & Metabolism*, *66*(1), pp.208-211.

Herman, A. and Herman, A.P., 2016. Mechanism of action of herbs and their active constituents used in hair loss treatment. *Fitoterapia*, *114*, pp.18-25.

Hevener, K.E., Ball, D.M., Buolamwini, J.K. and Lee, R.E., 2008. Quantitative structure–activity relationship studies on nitrofuranyl anti-tubercular agents. *Bioorganic & medicinal chemistry*, *16*(17), pp.8042-8053.

Higgins, C.A., Westgate, G.E. and Jahoda, C.A., 2009. From telogen to exogen: mechanisms underlying formation and subsequent loss of the hair club fiber. *Journal of Investigative Dermatology*, *129*(9), pp.2100-2108.

Hillmer, A.M., Brockschmidt, F.F., Hanneken, S., Eigelshoven, S., Steffens, M., Flaquer, A., Herms, S., Becker, T., Kortüm, A.K., Nyholt, D.R. and Zhao, Z.Z., 2008. Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nature genetics*, *40*(11), p.1279.

Hindorff, L.A., Junkins, H.A., Mehta, J.P. and Manolio, T., 2012. A catalog of published genome-wide association studies (https://www.ebi.ac.uk/gwas/).

Hirschhorn, J.N. and Daly, M.J., 2005. Genome-wide association studies for common diseases and complex traits. *Nature reviews genetics*, *6*(2), p.95

Hirsso, P., Laakso, M., Matilainen, V., Hiltunen, L., Rajala, U., Jokelainen, J. and Keinänen-Kiukaanniemi, S., 2006. Association of insulin resistance linked diseases and hair loss in elderly men. Finnish population-based study. *Central European journal of public health*, *14*(2).

Hoffmann, R. and Happle, R., 2000. Current understanding of androgenetic alopecia. Part II: clinical aspects and treatment. *European Journal of Dermatology*, *10*(5), pp.410-7.

Hollis, D.E., Chapman, R.E., Panaretto, B.A. and Moore, G.P.M., 1983. Morphological changes in the skin and wool fibres of Merino sheep infused with mouse epidermal growth factor. *Australian journal of biological sciences*, *36*(4), pp.419-434.

Hu, R., Xu, F., Sheng, Y., Qi, S., Han, Y., Miao, Y., Rui, W. and Yang, Q., 2015. Combined treatment with oral finasteride and topical minoxidil in male androgenetic alopecia: a randomized and comparative study in Chinese patients. *Dermatologic therapy*, *28*(5), pp.303-308.

Hunter, L. and Cohen, K.B., 2006. Biomedical language processing: what's beyond PubMed? *Molecular cell*, *21*(5), pp.589-594.

Hussein, S.M., Duff, E.K. and Sirard, C., 2003. Smad4 and β-catenin co-activators functionally interact with lymphoid-enhancing factor to regulate graded expression of Msx2. *Journal of Biological Chemistry*, *278*(49), pp.48805-48814.

Inaba, M. and Inaba, Y., 1996. Androgenetic alopecia. In *Androgenetic Alopecia* (pp. 161-168). Springer, Tokyo.

Inamadar, A.C., Palit, A. and Ragunatha, S., 2014. *Textbook of Pediatric dermatology*. JP Medical Ltd.

Iorizzo, M., Vincenzi, C., Voudouris, S., Piraccini, B.M. and Tosti, A., 2006. Finasteride treatment of female pattern hair loss. *Archives of dermatology*, *142*(3), pp.298-302.

Ito, M. and Kizawa, K., 2001. Expression of calcium-binding S100 proteins A4 and A6 in regions of the epithelial sac associated with the onset of hair follicle regeneration. *Journal of Investigative Dermatology*, *116*(6), pp.956-963.

Ito, M., 1990. The morphology and cell biology of the hair apparatus: recent advances.

Ito, M., Kizawa, K., Toyoda, M. and Morohashi, M., 2002. Label-retaining cells in the bulge region are directed to cell death after plucking, followed by healing from the surviving hair germ. *Journal of Investigative Dermatology*, *119*(6), pp.1310-1316.

Jacobs, J.P., Szpunar, C.A. and Warner, M.L., 1993. Use of topical minoxidil therapy for androgenetic alopecia in women. *International journal of dermatology*, *32*(10), pp.758-762.

Jain, P.K., Joshi, H. and Dass, D.J., 2012. Drug that causes hair loss and promotes hair growth-A review. *Int J Res Pharm Biomed Sci*, *3*(4), pp.1476-82.

Jaworsky, C., Kligman, A.M. and Murphy, G.F., 1992. Characterization of inflammatory infiltrates in male pattern alopecia: implications for pathogenesis. *British Journal of Dermatology*, *127*(3), pp.239-246

Johnson, E., 1977. Cycles and patterns of hair growth. *The physiology and pathophysiology of the skin*, *4*, pp.1237-1254.

Johnson, E., 1977. Environmental effects on the hair follicle. *Physiology and Pathophysiology of the Skin. Academic Press, London*, pp.1237-1249.

Jönsson, E.H., Bendas, J., Weidner, K., Wessberg, J., Olausson, H., Wasling, H.B. and Croy, I., 2017. The relation between human hair follicle density and touch perception. *Scientific reports*, *7*(1), p.2499.

Jung, H.J. and Chang, B.S., 2016. Ultrastructural characteristics of neonate scalp hair. *Indian Journal of Science and Technology*, *9*(26), pp.1-7.

Kaliyadan, F., Nambiar, A. and Vijayaraghavan, S., 2013. Androgenetic alopecia: an update. *Indian Journal of Dermatology, Venereology, and Leprology*, *79*(5), p.613.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M., (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research*, *38*(suppl_1), pp.D355-D360.

Kanwar, A. and Narang, T., 2013. Anagen effluvium. *Indian journal of dermatology, venereology and leprology*, *79*(5), p.604.

Kar, S. and Roy, K., 2012. QSAR of phytochemicals for the design of better drugs. *Expert opinion on drug discovery*, *7*(10), pp.877-902.

Karabulut, S. and Cakir, S., 2016. A case with alopecia related to risperidone use/Risperidon kullanimina bagli alopesi gelisen bir olgu. *Anadolu Psikiyatri Dergisi*, *17*, p.55.

Karashima, T., Tsuruta, D., Hamada, T., Ono, F., Ishii, N., Abe, T., Ohyama, B., Nakama, T., Dainichi, T. and Hashimoto, T., 2012. Oral zinc therapy for zinc deficiency-related telogen effluvium. *Dermatologic therapy*, *25*(2), pp.210-213.

Kataria, V., Wang, H., Wald, J.W. and Phan, Y.L., 2017. Lisinopril-induced alopecia: a case report. *Journal of pharmacy practice*, *30*(5), pp.562-566.

Katsimbri, P., Bamias, A. and Pavlidis, N., 2000. Prevention of chemotherapy-induced alopecia using an effective scalp cooling system. *European journal of cancer*, *36*(6), pp.766-771.

Kazius, J., McGuire, R. and Bursi, R., 2005. Derivation and validation of toxicophores for mutagenicity prediction. *Journal of medicinal chemistry*, *48*(1), pp.312-320.

Kere, J., Srivastava, A.K., Montonen, O., Zonana, J., Thomas, N., Ferguson, B., Munoz, F., Morgan, D., Clarke, A., Baybayan, P. and Chen, E.Y., 1996. X–linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. *Nature genetics*, *13*(4), p.409.

Kil, M.S., Kim, C.W. and Kim, S.S., 2013. Analysis of serum zinc and copper concentrations in hair loss. *Annals of Dermatology*, *25*(4), pp.405-409.

Kimelman, D. and Xu, W., 2006. β-Catenin destruction complex: insights and questions from a structural perspective. *Oncogene*, *25*(57), p.7482.

Kinoshita-Ise, M., Fukuyama, M. and Ohyama, M., 2023. Recent Advances in Understanding of the Etiopathogenesis, Diagnosis, and Management of Hair Loss Diseases. *Journal of Clinical Medicine*, *12*(9), p.3259.

Komen, M.M., Smorenburg, C.H., van den Hurk, C.J. and Nortier, J.W., 2013. Factors influenci ng the effectiveness of scalp cooling in the prevention of chemotherapy-induced alopecia. *Th e oncologist*, *18*(7), p.885.

Kovalchik, S 2017, RISmed: Download Content from NCBI Databases. R package version 2.1.7. accessed 18 February 2018, accessed 13 February 2016, < https://CRAN.R-project.org/packa ge=RISmed>

Krallinger, M. and Valencia, A., 2005. Text-mining and information-retrieval services for molecular biology. *Genome biology*, *6*(7), p.224.

Kulessa, H., Turk, G. and Hogan, B.L., 2000. Inhibition of Bmp signalling affects growth and differentiation in the anagen hair follicle. *The EMBO journal*, *19*(24), pp.6664-6674.

Kumar, P., Singh, S., Handa, V. and Kathuria, H., 2018. Oleic Acid Nanovesicles of Minoxidil for Enhanced Follicular Delivery. *Medicines*, *5*(3), p.103.

Kwon, O.S., Pyo, H.K., Oh, Y.J., Han, J.H., Lee, S.R., Chung, J.H., Eun, H.C. and Kim, K.H., 2007. Promotive effect of minoxidil combined with all-trans retinoic acid (tretinoin) on human hair growth in vitro. *Journal of Korean medical science*, *22*(2), pp.283-289.

Lachgar, S., Charveron, M., Gall, Y. and Bonafe, J.L., 1998. Minoxidil upregulates the expression of vascular endothelial growth factor in human hair dermal papilla cells. *The British journal of dermatology*, *138*(3), pp.407-411.

Lai, J.J., Chang, P., Lai, K.P., Chen, L. and Chang, C., 2012. The role of androgen and androgen receptor in skin-related disorders. *Archives of dermatological research*, *304*(7), pp.499-510.

Lampl, Y., Gilad, R., Sarova-Pinchas, I. and Barak, Y., 1996. Hair loss-an adverse reaction to treatment with vigabatrin. *Acta therapeutica*, *22*(1), pp.51-55.

Laurent, M.R., Hammond, G.L., Blokland, M., Jardí, F., Antonio, L., Dubois, V., Khalil, R., Sterk, S.S., Gielen, E., Decallonne, B. and Carmeliet, G., 2016. Sex hormone-binding globulin regulation of androgen bioactivity in vivo: validation of the free hormone hypothesis. *Scientific reports*, *6*, p.35539.

Lee, D.J. and Ewer, K., 2006. Implanon and alopecia. *Journal of Family Planning and Reproductive Health Care*, *32*(4), p.255.

Lee, W.S., Lee, H.J., Choi, G.S., Cheong, W.K., Chow, S.K., Gabriel, M.T., Hau, K.L., Kang, H., Mallari, M.R., Tsai, R.Y. and Zhang, J., 2013. Guidelines for management of androgenetic alopecia based on BASP classification–the Asian consensus committee guideline. *Journal of the European Academy of Dermatology and Venereology*, *27*(8), pp.1026-1034.

Leiros, G.J., Attorresi, A.I. and Balaña, M.E., 2012. Hair follicle stem cell differentiation is inhibited through cross-talk between Wnt/β-catenin and androgen signalling in dermal papilla cells from patients with androgenetic alopecia. *British Journal of Dermatology*, *166*(5), pp.1035-1042.

Levy, L.L. and Emer, J.J., 2013. Female pattern alopecia: current perspectives. *International journal of women's health*, *5*, p.541.

Li, N., Liu, S., Zhang, H.S., Deng, Z.L., Zhao, H.S., Zhao, Q., Lei, X.H., Ning, L.N., Cao, Y.J., Wang, H.B. and Liu, S., 2016. Exogenous R-spondin1 induces precocious telogen-to-anagen transition in mouse hair follicles. *International journal of molecular sciences*, *17*(4), p.582.

Li, R., Brockschmidt, F.F., Kiefer, A.K., Stefansson, H., Nyholt, D.R., Song, K., Vermeulen, S.H., Kanoni, S., Glass, D., Medland, S.E. and Dimitriou, M., 2012. Six novel susceptibility Loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS genetics*, *8*(5), p. e1002746.

Lin, W.H., Xiang, L.J., Shi, H.X., Zhang, J., Jiang, L.P., Cai, P.T., Lin, Z.L., Lin, B.B., Huang, Y., Zhang, H.L. and Fu, X.B., 2015. Fibroblast growth factors stimulate hair growth through β-catenin and Shh expression in C57BL/6 mice. *BioMed research international*, *2015*.

Lindner, G., Botchkarev, V.A., Botchkareva, N.V., Ling, G., Van Der Veen, C. and Paus, R., 1997. Analysis of apoptosis during hair follicle regression (catagen). *The American journal of pathology*, *151*(6), p.1601.

Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, *23*(1-3), pp.3-25.

Liu, R., Yu, X. and Wallqvist, A., 2015. Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries. *Journal of cheminformatics*, *7*(1), p.4.

Llau, M.E., Viraben, R. and Montastruc, J.L., 1995. Drug-induced alopecia: review of the literature. *Therapie*, *50*(2), pp.145-150.

Lotufo, P.A., Chae, C.U., Ajani, U.A., Hennekens, C.H. and Manson, J.E., 2000. Male pattern baldness and coronary heart disease: The Physicians' Health Study. *Archives of internal medicine*, *160*(2), pp.165-171.

Ludwig, E., 1977. Classification of the types of androgenetic alopecia (common baldness) occurring in the female sex. *British Journal of Dermatology*, *97*(3), pp.247-254.

Lowry, W.E., Blanpain, C., Nowak, J.A., Guasch, G., Lewis, L. and Fuchs, E., 2005. Defining the impact of β-catenin/Tcf transactivation on epithelial stem cells. *Genes & development*, *19*(13), pp.1596-1611.

Lyle, S., Christofidou-Solomidou, M., Liu, Y., Elder, D.E., Albelda, S. and Cotsarelis, G., 1998. The C8/144B monoclonal antibody recognizes cytokeratin 15 and defines the location of human hair follicle stem cells. *Journal of cell science*, *111*(21), pp.3179-3188.

Lynfield, Y.L. and Macwilliams, P., 1970. Shaving and hair growth. *Journal of Investigative Dermatology*, *55*(3), pp.170-172.

Madden, J., Nelms, M., Cronin, M. and Enoch, S., 2014. Identification of structural alerts for mitochondrial toxicity using chemotyper. *Toxicol. Lett*, *229*, p.S162.

Mahé, Y.F., Michelet, J.F., Billoni, N., Jarrousse, F., BTS, Buan, B., BTS, Commo, S., BTS, Saint-Léger, D. and Bernard, B.A., 2000. Androgenetic alopecia and microinflammation. *International journal of dermatology*, *39*(8), pp.576-584.

Malkud, S., 2015. Telogen effluvium: a review. *Journal of clinical and diagnostic research: JCDR*, *9*(9), p. WE01.

Mansouri, P., Mortazavi, M., Eslami, M. and Mazinani, M., 2005. Androgenetic alopecia and coronary artery disease in women. *Dermatology online journal*, *11*(3).

Massart, D.L., 1988. *Chemometrics: A Textbook (Data Handling in Science and Technology; V. 2)*. John Wiley and Sons Incorporated.

Matilainen, V.A., Mäkinen, P.K. and Keinänen-Kiukaanniemi, S.M., 2001. Early onset of androgenetic alopecia associated with early severe coronary heart disease: a population-based, case-control study. *Journal of cardiovascular risk*, *8*(3), pp.147-151.

Mavromoustakos, T., Durdagi, S., Koukoulitsa, C., Simcic, M., G Papadopoulos, M., Hodoscek, M. and Golic Grdadolnik, S., 2011. Strategies in the rational drug design. *Current medicinal chemistry*, *18*(17), pp.2517-2530.

McLean, R.M. and Harrison-Woolrych, M., 2007. Alopecia associated with quetiapine. *International clinical psychopharmacology*, *22*(2), pp.117-119.

Meier, N., Dear, T.N. and Boehm, T., 1999. Whn and mHa3 are components of the genetic hierarchy controlling hair follicle differentiation. *Mechanisms of development*, *89*(1-2), pp.215-221.

Melkote, S., Dhurat, R.S., Palav, A. and Jerajani, H.R., 2009. Alopecia in congenital hidrotic ectodermal dysplasia responding to treatment with a combination of topical minoxidil and tretinoin. *International journal of dermatology*, *48*(2), pp.184-185.

Messenger, A.G., 2000. Thyroid hormone and hair growth. *The British journal of dermatology*, *142*(4), pp.633-634.

Meyer, H., 1899. Welche Eigenschaft der Anaesthetica bedingt ihre narkotische Wirkung? *Naunyn-Schmiedeberg's Arch Exp Pathol Pharmakol*, *42*, pp.109-118.

Meyer, H.C., 1979. Alopecia associated with ibuprofen. *JAMA*, *242*(2), pp.142-142.

MHRA Yellow Card Scheme, accessed 10 March 2016 <https://yellowcard.mhra.gov.uk>

Michel, L., Reygagne, P., Benech, P., Jean-Louis, F., Scalvino, S., Ly Ka So, S., Hamidou, Z., Bianovici, S., Pouch, J., Ducos, B. and Bonnet, M., 2017. Study of gene expression alteration

in male androgenetic alopecia: evidence of predominant molecular signalling pathways. *British Journal of Dermatology*, *177*(5), pp.1322-1336.

Millar, S.E., 2002. Molecular mechanisms regulating hair follicle development. *Journal of Investigative Dermatology*, *118*(2), pp.216-225.

Millar, S.E., Willert, K., Salinas, P.C., Roelink, H., Nusse, R., Sussman, D.J. and Barsh, G.S., 1999. WNT signaling in the control of hair growth and structure. *Developmental biology*, *207*(1), pp.133-149.

Miranda, B.H., Tobin, D.J., Sharpe, D.T. and Randall, V.A., 2010. Intermediate hair follicles: a new more clinically relevant model for hair growth investigations. *British Journal of Dermatology*, *163*(2), pp.287-295.

Mo, S. and Cui, Z., 2012. Regulation of canonical Wnt signaling during development and diseases. In *Embryogenesis*. IntechOpen.

Montagna, W. and Carlisle, K., 1990. Structural changes in ageing skin. *British Journal of Dermatology*, *122*, pp.61-70.

Muller, M., Jasmin, J.R., Monteil, R.A. and Loubiere, R., 1991. Embryology of the hair follicle. *Early human development*, *26*(3), pp.159-166.

Müller-Röver, S., Foitzik, K., Paus, R., Handjiski, B., van der Veen, C., Eichmüller, S., McKay, I.A. and Stenn, K.S., 2001. A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. *Journal of Investigative Dermatology*, *117*(1), pp.3-15.

Mullin, R., 2014. Cost to develop new pharmaceutical drug now exceeds $2.5 B. *Scientific American*, *24*.

Muniz, C.E., Salem, R.B. and Director, K.L., 1982. Hair loss in a patient receiving lithium. *Psychosomatics*, *23*(3), pp.312-313.

Muro, Y., Sugiura, K. and Akiyama, M., 2016. Cutaneous manifestations in dermatomyositis: key clinical and serological features—a comprehensive review. *Clinical reviews in allergy & immunology*, *51*(3), pp.293-302.

Myung, P. and Ito, M., 2012. Dissecting the bulge in hair regeneration. *The Journal of clinical investigation*, *122*(2), pp.448-454.

Nakamura, M., Schneider, M.R., Schmidt-Ullrich, R. and Paus, R., 2013. Mutant laboratory mice with abnormalities in hair follicle morphogenesis, cycling, and/or structure: an update. *Journal of dermatological science*, *69*(1), pp.6-29.

Narisawa, Y. and Kohda, H., 1996. Two-and three-dimensional demonstrations of morphological alterations of early anagen hair follicle with special reference to the bulge area. *Archives of dermatological research*, *288*(2), pp.98-102.

Nendza, M., 2012. *Structure—Activity Relationships in Environmental Sciences* (Vol. 6). Springer Science & Business Media.

Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A. and Myatt, G., 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships: The report and recommendations of ecvam workshop 52. *Alternatives to Laboratory Animals*, *33*(2), pp.155-173.

Nieves, A. and Garza, L.A., 2014. Does prostaglandin D2 hold the cure to male pattern baldness? *Experimental dermatology*, *23*(4), pp.224-227.

Nishimura, E.K., Jordan, S.A., Oshima, H., Yoshida, H., Osawa, M., Moriyama, M., Jackson, I.J., Barrandon, Y., Miyachi, Y. and Nishikawa, S.I., 2002. Dominant role of the niche in melanocyte stem-cell fate determination. *Nature*, *416*(6883), p.854.

Noveen, A., Jiang, T.X., Ting-Berreth, S.A. and Choung, C.M., 1995. Homeobox genes Msx-1 and Msx-2 are associated with induction and growth of skin appendages. *Journal of Investigative Dermatology*, *104*(5), pp.711-719.

Nusse, R. and Varmus, H.E., 1982. Many tumors induced by the mouse mammary tumour virus contain a provirus integrated in the same region of the host genome. *Cell*, *31*(1), pp.99-109.

Oak, M.H., Bedoui, J.E., Madeira, S.F., Chalupsky, K. and Schini-Kerth, V.B., 2006. Delphinidin and cyanidin inhibit PDGFAB-induced VEGF release in vascular smooth muscle cells by preventing activation of p38 MAPK and JNK. *British journal of pharmacology*, *149*(3), pp.283-290.

Oh, B.R., Kim, S.J., Moon, J.D., Kim, H.N., Kwon, D.D., Won, Y.H., Ryu, S.B. and Park, Y.I., 1998. Association of benign prostatic hyperplasia with male pattern baldness. *Urology*, *51*(5), pp.744-748.

Oh, J.W., Kloepper, J., Langan, E.A., Kim, Y., Yeo, J., Kim, M.J., Hsi, T.C., Rose, C., Yoon, G.S., Lee, S.J. and Seykora, J., 2016. A guide to studying human hair follicle cycling in vivo. *Journal of Investigative Dermatology*, *136*(1), pp.34-44.

Olsen, E.A., 2001. Female pattern hair loss. *Journal of the American Academy of Dermatology*, *45*(3), pp. S70-S80.

Olsen, E.A., Hordinsky, M., Whiting, D., Stough, D., Hobbs, S., Ellis, M.L., Wilson, T., Rittmaster, R.S. and Dutasteride Alopecia Research Team, 2006. The importance of dual 5α-reductase inhibition in the treatment of male pattern hair loss: results of a randomized placebo-controlled study of dutasteride versus finasteride. *Journal of the American Academy of Dermatology*, *55*(6), pp.1014-1023.

Olsen, E.A., Whiting, D.A., Savin, R., Rodgers, A., Johnson-Levonas, A.O., Round, E., Rotonda, J., Kaufman, K.D. and Male Pattern Hair Loss Study Group, 2012. Global photographic assessment of men aged 18 to 60 years with male pattern hair loss receiving finasteride 1 mg or placebo. *Journal of the American Academy of Dermatology*, *67*(3), pp.379-386.

Ornitz, D.M. and Itoh, N., 2015. The fibroblast growth factor signaling pathway. *Wiley Interdisciplinary Reviews: Developmental Biology*, *4*(3), pp.215-266.

Orwin, A., 1983. Hair loss following lithium therapy. *British Journal of Dermatology*, *108*(4), pp.503-504.

Oshima, H., Rochat, A., Karzai, C., Kobayashi, K. and Barrandon, Y., 2001. Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell*, *104*(2), pp.233-245.

Osman, M. and McCauley, M.D., 2016. Psychiatric shades of grey: mirtazapine-induced hair discoloration and hair loss. *Irish journal of psychological medicine*, *33*(3), pp.175-178.

Otberg, N., Finner, A.M. and Shapiro, J., 2007. Androgenetic alopecia. *Endocrinology and metabolism clinics of North America*, *36*(2), pp.379-398.

Overton, E., 1896. Ueber die osmotischen Eigenschaften der Zelle in ihrer Bedentung für die Toxicologie und Pharmakologie. *Vierteljahrsschr. Narturforsch. Ges. Zuerich*, *41*, p.383.

Padhy, I., Mahapatra, A., Banerjee, B. and Sharma, T., 2023. Computational approaches in drug discovery from phytochemicals. In *Phytochemistry, Computational Tools and Databases in Drug Discovery* (pp. 57-88). Elsevier.

Pagès, H., Aboyoun, P., Gentleman, R. and DebRoy, S., 2018. Biostrings: efficient manipulation of biological strings. R package version 2480.

Paik, J.H., Yoon, J.B., Sim, W.Y., Kim, B.S. and Kim, N.I., 2001. The prevalence and types of androgenetic alopecia in Korean men and women. *British Journal of Dermatology*, *145*(1), pp.95-99.

Palombo, V., Milanesi, M., Sferra, G., Capomaccio, S., Sgorlon, S. and D'Andrea, M., 2020. PANEV: an R package for a pathway-based network visualization. *BMC bioinformatics*, *21*(1), pp.1-7.

Panteleyev, A.A., Jahoda, C.A. and Christiano, A.M., 2001. Hair follicle predetermination. *Journal of Cell Science*, *114*(19), pp.3419-3431.

Papadatos, G., Gaulton, A., Hersey, A. and Overington, J.P., 2015. Activity, assay and target data curation and quality in the ChEMBL database. *Journal of computer-aided molecular design*, *29*(9), pp.885-896.

Paradisi, R., Porcu, E., Fabbri, R., Seracchioli, R., Battaglia, C. and Venturoli, S., 2011. Prospective cohort study on the effects and tolerability of flutamide in patients with female pattern hair loss. *Annals of Pharmacotherapy*, *45*(4), pp.469-475.

Paterson, H., Clifton, J., Miller, D., Ashton, J. and Harrison-Woolrych, M., 2007. Hair loss with use of the levonorgestrel intrauterine device. *Contraception*, *76*(4), pp.306-309.

Pathomvanich, D., Pongratananukul, S., Thienthaworn, P. and Manoshai, S., 2002. A random study of Asian male androgenetic alopecia in Bangkok, Thailand. *Dermatologic surgery*, *28*(9), pp.804-807.

Patrizi, A., Savoia, F., Negosanti, F., Posar, A., Santucci, M. and Neri, I., 2005. Telogen effluvium caused by magnesium valproate and lamotrigine. *Acta dermato-venereologica*, *85*(1), pp.77-78.

Paus, R. and Cotsarelis, G., 1999. The biology of hair follicles. *New England journal of medicine*, *341*(7), pp.491-497.

Pecoraro, V., Astore, I., Barman, J. and Araujo, C.I., 1964. The normal trichogram in the child before the age of puberty. *Journal of Investigative Dermatology*, *42*(6), pp.427-430.

Peters, E.M., Müller, Y., Snaga, W., Fliege, H., Reißhauer, A., Schmidt-Rose, T., Max, H., Schweiger, D., Rose, M. and Kruse, J., 2017. Hair and stress: A pilot study of hair and cytokine balance alteration in healthy young women under major exam stress. *PloS one*, *12*(4), p. e0175904.

Petiot, A., Conti, F.J., Grose, R., Revest, J.M., Hodivala-Dilke, K.M. and Dickson, C., 2003. A crucial role for Fgfr2-IIIb signalling in epidermal development and hair follicle patterning.

Picard, C., Jonville-Bera, A.P., Billard, C. and Autret, E., 1997. Alopecia associated with gabapentin: First case. *Annals of Pharmacotherapy*, *31*(10), pp.1260-1260.

Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y. and Hinds, D.A., 2016. Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics*, *48*(7), p.709.

Piérard, G.E., Pierard-Franchimont, C., Nikkels-Tassoudji, N., Nikkels, A.F. and Léger, D.S., 1996. Improvement in the inflammatory aspect of androgenetic alopecia. A pilot study with an antimicrobial lotion. *Journal of dermatological treatment*, *7*(3), pp.153-157.

Piérard-Franchimont, C., Quatresooz, P. and Piérard, G.E., 2010. Effect of UV radiation on scalp and hair growth. In *Aging hair* (pp. 113-121). Springer, Berlin, Heidelberg.

Piérard-Franchimont, C., Uhoda, I., Saint-Léger, D. and Piérard, G.E., 2002. Androgenic alopecia and stress-induced premature senescence by cumulative ultraviolet light exposure. *Exogenous Dermatology*, *1*(4), pp.203-206.

Pierre, S.A.S., Vercellotti, G.M., Donovan, J.C. and Hordinsky, M.K., 2010. Iron deficiency and diffuse nonscarring scalp alopecia in women: more pieces to the puzzle. *Journal of the American Academy of Dermatology*, *63*(6), pp.1070-1076.

Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F. and Furlong, L.I., 2020. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids research*, *48*(D1), pp.D845-D855.

Pinkus, H., Iwasaki, T. and Mishima, Y., 1981. Outer root sheath keratinization in anagen and catagen of the mammalian hair follicle. A seventh distinct type of keratinization in the hair follicle: trichilemmal keratinization. *Journal of anatomy*, *133*(Pt 1), p.19.

Pirastu, N., Joshi, P.K., De Vries, P.S., Cornelis, M.C., McKeigue, P.M., Keum, N., Franceschini, N., Colombo, M., Giovannucci, E.L., Spiliopoulou, A. and Franke, L., 2017. GWAS for male-

pattern baldness identifies 71 susceptibility loci explaining 38% of the risk. *Nature communications*, *8*(1), p.1584.

Pitchot, W. and Ansseau, M., 2001. Venlafaxine-induced hair loss. *American Journal of Psychiatry*, *158*(7), pp.1159-a.

Plikus, M., Wang, W.P., Liu, J., Wang, X., Jiang, T.X. and Chuong, C.M., 2004. Morpho-regulation of ectodermal organs: integument pathology and phenotypic variations in K14-Noggin engineered mice through modulation of bone morphogenic protein pathway. *The American journal of pathology*, *164*(3), pp.1099-1114.

Plikus, M.V., Mayer, J.A., de La Cruz, D., Baker, R.E., Maini, P.K., Maxson, R. and Chuong, C.M., 2008. Cyclic dermal BMP signalling regulates stem cell activation during hair regeneration. *Nature*, *451*(7176), p.340.

Porter, R.M., Lunny, D.P., Ogden, P.H., Morley, S.M., McLean, W.I., Evans, A., Harrison, D.L., Rugg, E.L. and Lane, E.B., 2000. K15 expression implies lateral differentiation within stratified epithelial basal cells. *Laboratory investigation*, *80*(11), p.1701.

Potter, C.S., Peterson, R.L., Barth, J.L., Pruett, N.D., Jacobs, D.F., Kern, M.J., Argraves, W.S., Sundberg, J.P. and Awgulewitsch, A., 2006. Evidence that the satin hair mutant gene Foxq1 is among multiple and functionally diverse regulatory targets for Hoxc13 during hair follicle differentiation. *Journal of Biological Chemistry*, *281*(39), pp.29245-29255.

Powell, B.C., Passmore, E.A., Nesci, A. and Dunn, S.M., 1998. The Notch signalling pathway in hair growth. *Mechanisms of development*, *78*(1-2), pp.189-192.

Pummila, M., Fliniaux, I., Jaatinen, R., James, M.J., Laurikkala, J., Schneider, P., Thesleff, I. and Mikkola, M.L., 2007. Ectodysplasin has a dual role in ectodermal organogenesis: inhibition of Bmp activity and induction of Shh expression. *Development*, *134*(1), pp.117-125.

Randall, V.A., 2008. Androgens and hair growth. *Dermatologic therapy*, *21*(5), pp.314-328.

Randić, M., 1991. Generalized molecular descriptors. *Journal of Mathematical Chemistry*, *7*(1), pp.155-168.

Randić, M., 1996. Molecular bonding profiles. *Journal of Mathematical Chemistry*, *19*(3), pp.375-392.

Rani, J. and Ramachandran, S., 2015. pubmed. mineR: An R package with text-mining algorit hms to analyse PubMed abstracts. *Journal of biosciences*, *40*(4), pp.671-682.

Rathnayake, D. and Sinclair, R., 2010. Innovative use of spironolactone as an antiandrogen in the treatment of female pattern hair loss. *Dermatologic clinics*, *28*(3), pp.611-618.

Rebora, A., 1997. Telogen effluvium. *Dermatology*, *195*(3), pp.209-212.

Rebora, A. and Guarrera, M., 2004. Teloptosis and kenogen: two new concepts in human trichology. *Archives of dermatology*, *140*(5), pp.619-620.

Rebora, A., 2004. Pathogenesis of androgenetic alopecia. *Journal of the American Academy of Dermatology*, *50*(5), pp.777-779.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R. and Lander, E.S., 2001. Linkage disequilibrium in the human genome. *Nature*, *411*(6834), p.199.

Rendl, M., Lewis, L. and Fuchs, E., 2005. Molecular dissection of mesenchymal–epithelial interactions in the hair follicle. *PLoS biology*, *3*(11), p.331.

Reynolds, E.L., 1951. The appearance of adult patterns of body hair in man. *Annals of the New York Academy of Sciences*, *53*(3), pp.576-584.

Rhodes, T., Girman, C.J., Savin, R.C., Kaufman, K.D., Guo, S., Lilly, F.R., Siervogel, R.M. and Chumlea, W.C., 1998. Prevalence of male pattern hair loss in 18–49 year-old men. *Dermatologic surgery*, *24*(12), pp.1330-1332.

Richards, J.B., Yuan, X., Geller, F., Waterworth, D., Bataille, V., Glass, D., Song, K., Waeber, G., Vollenweider, P., Aben, K.K. and Kiemeney, L.A., 2008. Male-pattern baldness susceptibility locus at 20p11. *Nature genetics*, *40*(11), p.1282.

Richards, K.A., Fukai, K., Oiso, N. and Paller, A.S., 2001. A novel KIT mutation results in piebaldism with progressive depigmentation. *Journal of the American Academy of Dermatology*, *44*(2), pp.288-292.

Richardson, G.D., Bazzi, H., Fantauzzo, K.A., Waters, J.M., Crawford, H., Hynd, P., Christiano, A.M. and Jahoda, C.A., 2009. KGF and EGF signalling block hair follicle induction and promote interfollicular epidermal fate in developing mouse skin.

Richet, O., 1893. Chloralose, a derivative of chloral, and its physiological properties. *American Journal of Pharmacy (1835-1907)*, p.394.

Rishikaysh, P., Dev, K., Diaz, D., Qureshi, W., Filip, S. and Mokry, J., 2014. Signaling involved in hair follicle morphogenesis and development. *International journal of molecular sciences*, *15*(1), pp.1647-1670.

Rogers, M.A., Langbein, L., Praetzel-Wunder, S., Winter, H. and Schweizer, J., 2006. Human hair keratin-associated proteins (KAPs). *International review of cytology*, *251*, pp.209-263.

Rogers, N.E. and Avram, M.R., 2008. Medical treatments for male and female pattern hair loss. *Journal of the American Academy of Dermatology*, *59*(4), pp.547-566.

Rook, A., 1965. Endocrine influences on hair growth. *British medical journal*, *1*(5435), p.609.

Roth, W., Deussing, J.A.N., Botchkarev, V.A., Pauly-Evers, M., Saftig, P., Hafner, A., Schmidt, P., Schmahl, W., Scherer, J., Anton-Lamprecht, I. and Von Figura, K., 2000. Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and perturbation of hair follicle cycling. *The FASEB Journal*, *14*(13), pp.2075-2086.

Roy, K., Kar, S., Das, R.N., Roy, K., Kar, S. and Das, R.N., 2015. Chemical information and descriptors. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, pp.47-80.

Roy, S., Qiao, T., Wolff, C. and Ingham, P.W., 2001. Hedgehog signaling pathway is essential for pancreas specification in the zebrafish embryo. *Current Biology*, *11*(17), pp.1358-1363.

Rubio-Gonzalez, B., Juhász, M., Fortman, J. and Mesinkovska, N.A., 2018. Pathogenesis and treatment options for chemotherapy-induced alopecia: a systematic review. *international journal of dermatology*, *57*(12), pp.1417-1424.

Ruusmann, V., Sild, S. and Maran, U., 2015. QSAR DataBank repository: open and linked qualitative and quantitative structure–activity relationship models. *Journal of cheminformatics*, *7*(1), p.32.

Saitoh, M. and Uzuka, M., 1969. Rate of Hair Growth in Advances in Biology of Skin. *Hair Growth*, *9*, pp.183-201.

Samplaski, M.K., Lo, K., Grober, E. and Jarvi, K., 2013. Finasteride use in the male infertility population: effects on semen and hormone parameters. *Fertility and sterility*, *100*(6), pp.1542-1546.

Samudra, A.P. and Sahinidis, N.V., 2013. Optimization-based framework for computer-aided molecular design. *AIChE Journal*, *59*(10), pp.3686-3701.

Samuelov, L., Sprecher, E. and Paus, R., 2015. The role of P-cadherin in skin biology and skin pathology: lessons from the hair follicle. *Cell and tissue research*, *360*(3), pp.761-771.

Sayin, A. and Turkoglu, S., 2016. Methylphenidate-related severe and diffuse hair loss: A case report. *Klin Psychopharmacol B*, *26*, p.327.

Schlake, T., 2007, April. Determination of hair structure and shape. In *Seminars in cell & developmental biology* (Vol. 18, No. 2, pp. 267-273). Academic Press.

Schmidt, F., Matter, H., Hessler, G. and Czich, A., 2014. Predictive *in silico* off-target profiling in drug discovery. *Future Medicinal Chemistry*, *6*(3), pp.295-317.

Schmidt-Ullrich, R. and Paus, R., 2005. Molecular principles of hair follicle induction and morphogenesis. *Bioessays*, *27*(3), pp.247-261.

Sclafani, A.P., 2014. Platelet-rich fibrin matrix (PRFM) for androgenetic alopecia. *Facial Plastic Surgery*, *30*(02), pp.219-224.

Scott, G., Ewing, J., Ryan, D. and Abboud, C., 1994. Stem cell factor regulates human melanocyte-matrix interactions. *Pigment cell research*, *7*(1), pp.44-51.

Seal, A 2014, Github, accessed July 2018, <https://github.com/abhik1368/dsdht/blob/master/Cheminformatics/QSAR.R>

Seal, A 2015, Github, accessed July 2018 <https://github.com/abhik1368/CheminfoR>

Seco, C.Z., de Castro, L.S., Van Nierop, J.W., Morín, M., Jhangiani, S., Verver, E.J., Schraders, M., Maiwald, N., Wesdorp, M., Venselaar, H. and Spruijt, L., 2015. Allelic mutations of KITLG, encoding KIT ligand, cause asymmetric and unilateral hearing loss and Waardenburg syndrome type 2. *The American Journal of Human Genetics*, *97*(5), pp.647-660.

Segal, A.S., 2002. Alopecia associated with atorvastatin. *The American journal of medicine*, *113*(2), p.171.

Sekiya, T., Adachi, S., Kohu, K., Yamada, T., Higuchi, O., Furukawa, Y., Nakamura, Y., Nakamura, T., Tashiro, K., Kuhara, S. and Ohwada, S., 2004. Identification of BMP and activin membrane-bound inhibitor (BAMBI), an inhibitor of transforming growth factor-β signaling, as a target of the β-catenin pathway in colorectal tumour cells. *Journal of Biological Chemistry*, *279*(8), pp.6840-6846.

Sennett, R. and Rendl, M., 2012, October. Mesenchymal–epithelial interactions during hair follicle morphogenesis and cycling. In *Seminars in cell & developmental biology* (Vol. 23, No. 8, pp. 917-927). Academic Press.

Severi, G., Sinclair, R., Hopper, J.L., English, D.R., McCredie, M.R.E., Boyle, P. and Giles, G.G., 2003. Androgenetic alopecia in men aged 40–69 years: prevalence and risk factors. *British Journal of Dermatology*, *149*(6), pp.1207-1213.

Shaker, B., Ahmad, S., Lee, J., Jung, C. and Na, D., 2021. *In silico* methods and tools for drug discovery. *Computers in biology and medicine*, *137*, p.104851.

Shanshanwal, S.J. and Dhurat, R.S., 2017. Superiority of dutasteride over finasteride in hair regrowth and reversal of miniaturization in men with androgenetic alopecia: A randomized controlled open-label, evaluator-blinded study. *Indian Journal of Dermatology, Venereology, and Leprology*, *83*(1), p.47.

Shapiro, J. and Kaufman, K.D., 2003, June. Use of finasteride in the treatment of men with androgenetic alopecia (male pattern hair loss). In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 8, No. 1, pp. 20-23). Elsevier.

Sharma, A., Goren, A., Dhurat, R., Agrawal, S., Sinclair, R., Trüeb, R.M., Vañó-Galván, S., Chen, G., Tan, Y., Kovacevic, M. and Situm, M., 2019. Tretinoin enhances minoxidil response in androgenetic alopecia patients by upregulating follicular sulfotransferase enzymes. *Dermatologic Therapy*, *32*(3), p.e 12915.

Sharma, R.P. and Chopra, V.L., 1976. Effect of the Wingless (wg1) mutation on wing and haltered development in Drosophila melanogaster. *Developmental biology*, *48*(2), pp.461-465.

Sheehan, T.M.T., 2006. Hair in Toxicology: An Important Bio-Monitor Edited by Desmond J. Tobin.

Shin, H.S., Won, C.H., Lee, S.H., Kwon, O.S., Kim, K.H. and Eun, H.C., 2007. Efficacy of 5% minoxidil versus combined 5% minoxidil and 0.01% tretinoin for male pattern hair loss. *American journal of clinical dermatology*, *8*(5), pp.285-290.

Shorter, K., Farjo, N.P., Picksley, S.M. and Randall, V.A., 2008. Human hair follicles contain two forms of ATP-sensitive potassium channels, only one of which is sensitive to minoxidil. *The FASEB Journal*, *22*(6), pp.1725-1736.

Shum, K.W., Cullen, D.R. and Messenger, A.G., 2002. Hair loss in women with hyperandrogenism: four cases responding to finasteride. *Journal of the American Academy of Dermatology*, *47*(5), pp.733-739.

Shuper, A., Stahl, B., Weitz, R., Labrada, L.M. and Depot, M., 1985. Carbamazepine-induced hair loss. *Drug intelligence & clinical pharmacy*, *19*(12), pp.924-925.

SIDER DB, accessed 18 April 2016,<http://sideeffects.embl.de>

Sinclair, R., Patel, M., Dawson Jr, T.L., Yazdabadi, A., Yip, L., Perez, A. and Rufaut, N.W., 2011. Hair loss in women: medical and cosmetic approaches to increase scalp hair fullness. *British Journal of Dermatology*, *165*, pp.12-18.

Sinclair, R., Wewerinke, M. and Jolley, D., 2005. Treatment of female pattern hair loss with oral antiandrogens. *British Journal of Dermatology*, *152*(3), pp.466-473.

Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S.L., Digles, D. and Ehrhart, F., 2018. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research*, *46*(D1), pp.D661-D667.

Slominski, A., Paus, R., Plonka, P., Chakraborty, A., Maurer, M., Pruski, D. and Lukiewicz, S., 1994. Melanogenesis during the anagen-catagen-telogen transformation of the murine hair cycle. *Journal of investigative dermatology*, *102*(6), pp.862-869.

Soo, C. and Lorenz, H.P., 2004. Anatomy and Embryology of the Skin. In *Handbook of Plastic Surgery* (pp. 77-80). CRC Press.

Spearman, R.I.C., 1977. Hair follicle development, cyclical changes and hair form. *The physiology and pathophysiology of the skin*, *4*, pp.1255-1292.

Sperling, L.C., 1991. Hair anatomy for the clinician. *Journal of the American Academy of Dermatology*, *25*(1), pp.1-17.

Sperling, L.C., 1999. Hair density in African Americans. *Archives of dermatology*, *135*(6), pp.656-658.

Srettabunjong, S., Patompakdeesakul, P. and Limawongpranee, S., 2016. Relative studies between hair index, hair area, and medullary index with age and sex in Thai scalp hair. *Forensic science international*, *267*, pp.196-203.

Stein, K.M., Odom, R.B., Justice, G.R. and Martin, G.C., 1973. Toxic alopecia from ingestion of boric acid. *Archives of dermatology*, *108*(1), pp.95-97.

Stelnicki, E.J., Kömüves, L.G., Kwong, A.O., Holmes, D., Klein, P., Rozenfeld, S., Lawrence, H.J., Adzick, N.S., Harrison, M. and Largman, C., 1998. HOX homeobox genes exhibit spatial and temporal changes in expression during human skin development. *Journal of Investigative Dermatology*, *110*(2), pp.110-115.

Stenn, K., 2005. Exogen is an active, separately controlled phase of the hair growth cycle. *Journal of the American Academy of Dermatology*, *52*(2), pp.374-375.

Stenn, K.S. and Paus, R., 2001. Controls of hair follicle cycling. *Physiological reviews*, *81*(1), pp.449-494.

Stenn, K.S., 2001. Insights from the asebia mouse: a molecular sebaceous gland defect leading to cicatricial alopecia. *Journal of cutaneous pathology*, *28*(9), pp.445-447.

Su, L.H. and Chen, T.H.H., 2007. Association of androgenetic alopecia with smoking and its prevalence among Asian men: a community-based survey. *Archives of dermatology*, *143*(11), pp.1401-1406.

Sundberg, J.P., Rourk, M.H., Boggess, D., Hogan, M.E., Sundberg, B.A. and Bertolino, A.P., 1997. Angora mouse mutation: altered hair cycle, follicular dystrophy, phenotypic maintenance of skin grafts, and changes in keratin expression. *Veterinary pathology*, *34*(3), pp.171-179.

Takashima, I., Iju, M. and Sudo, M., 1981. Alopecia androgenetica—its incidence in Japanese and associated conditions. In *Hair research* (pp. 287-293). Springer, Berlin, Heidelberg.

Tang, P.H., Chia, H.P., Cheong, L.L. and Koh, D., 2000. A community study of male androgenetic alopecia in Bishan, Singapore. *Singapore Med J*, *41*(5), pp.202-205.

Tansey, E.A. and Johnson, C.D., 2015. Recent advances in thermoregulation. *Advances in physiology education*, *39*(3), pp.139-148.

Tenenbaum, D., (2019). KEGGREST: Client-side REST access to KEGG. R package version 1.24. 0. 2019.

Testa, B. and Kier, L.B., 1991. The concept of molecular structure in structure–activity relationship studies and drug design. *Medicinal research reviews*, *11*(1), pp.35-48.

Thompson, I.M., Goodman, P.J., Tangen, C.M., Lucia, M.S., Miller, G.J., Ford, L.G., Lieber, M.M., Cespedes, R.D., Atkins, J.N., Lippman, S.M. and Carlin, S.M., 2003. The influence of finasteride on the development of prostate cancer. *New England journal of medicine*, *349*(3), pp.215-224.

Todeschini, R. and Consonni, V., 2008. *Handbook of molecular descriptors* (Vol. 11). John Wiley & Sons.

Todeschini, R. and Consonni, V., 2009. *Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references* (Vol. 41). John Wiley & Sons.

Tosti, A. and Pazzaglia, M., 2007. Drug reactions affecting hair: diagnosis. *Dermatologic clinics*, *25*(2), pp.223-231.

Tosti, A., Misciali, C., Piraccini, B.M., Peluso, A.M. and Bardazzi, F., 1994. Drug-induced hair loss and hair growth. *Drug safety*, *10*(4), pp.310-317.

Trüeb, R.M., 2002. Molecular mechanisms of androgenetic alopecia. *Experimental gerontology*, *37*(8-9), pp.981-990.

Trüeb, R.M., 2004. Finasteride treatment of patterned hair loss in normoandrogenic postmenopausal women. *Dermatology*, *209*(3), pp.202-207.

Trüeb, R.M., 2008. Diffuse hair loss. In *Hair Growth and Disorders* (pp. 259-272). Springer, Berlin, Heidelberg.

Trüeb, R.M., 2010. Chemotherapy-induced alopecia. *Current opinion in supportive and palliative care*, *4*(4), pp.281-284.

Twigg, J. and Majima, S., 2014. Consumption and the constitution of age: Expenditure patterns on clothing, hair and cosmetics among post-war 'baby boomers'. *Journal of Aging Studies*, *30*, pp.23-32.

Van den Bemt, P.M., Brodie-Meijer, C.C., Krijnen, R.M. and Nieboer, C., 1999. Drug induced alopecia. *Nederlands tijdschrift voor geneeskunde*, *143*(19), p.990.

Vandiviere, H.M., Dale, T.A., Driess, R.B. and Watson, K.A., 1971. Hair-shaft diameter as an index of protein-calorie malnutrition. *Archives of Environmental Health: An International Journal*, *23*(1), pp.61-66.

Veniaminova, N.A., Vagnozzi, A.N., Kopinke, D., Do, T.T., Murtaugh, L.C., Maillard, I., Dlugosz, A.A., Reiter, J.F. and Wong, S.Y., 2013. Keratin 79 identifies a novel population of migratory epithelial cells that initiates hair canal morphogenesis and regeneration. *Development*, *140*(24), pp.4870-4880.

Verma, J., Khedkar, V.M. and Coutinho, E.C., 2010. 3D-QSAR in drug design-a review. *Current topics in medicinal chemistry*, *10*(1), pp.95-115.

Vincent, M. and Yogiraj, K., 2013. A descriptive study of alopecia patterns and their relation to thyroid dysfunction. *International journal of trichology*, *5*(1), p.57.

Walters, W.P. and Murcko, M.A., 2002. Prediction of 'drug-likeness'. *Advanced drug delivery reviews*, *54*(3), pp.255-271.

Wang, Y., Backman, T.W., Horan, K. and Girke, T., 2013. fmcsR: mismatch tolerant maximum common substructure searching in R. *Bioinformatics*, *29*(21), pp.2792-2794.

Wang, Y.Y., Yang, Y.X., Zhao, R., Pan, S.T., Zhe, H., He, Z.X., Duan, W., Zhang, X., Yang, T., Qiu, J.X. and Zhou, S.F., 2015. Bardoxolone methyl induces apoptosis and autophagy and inhibits epithelial-to-mesenchymal transition and stemness in esophageal squamous cancer cells. *Drug design, development and therapy*, *9*, p.993.

Whiting, D.A., 1996. Chronic telogen effluvium. *Dermatologic clinics*, *14*(4), pp.723-731.

Warnock, J.K., 1991. Psychotropic medication and drug-related alopecia. *Psychosomatics: Journal of Consultation and Liaison Psychiatry*.

Warnock, J.K., Sieg, K., Willsie, D., Stevenson, E.K. and Kestenbaum, T., 1991. Drug-related alopecia in patients treated with tricyclic antidepressants. *Journal of Nervous and Mental Disease*.

Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, *28*(1), pp.31-36.

Weininger, D., 1990. SMILES. 3. DEPICT. Graphical depiction of chemical structures. *Journal of chemical information and computer sciences*, *30*(3), pp.237-243.

Weininger, D., Weininger, A. and Weininger, J.L., 1989. SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of chemical information and computer sciences*, *29*(2), pp.97-101.

Winter, D.J., 2017. *rentrez: An R package for the NCBI eUtils API* (No. e3179v2). PeerJ Preprin t.

Wright, E.S., 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R. *R Journal*, *8*(1).

Xu, F., Sheng, Y.Y., Mu, Z.L., Lou, W., Zhou, J., Ren, Y.T., Qi, S.S., Wang, X.S., Fu, Z.W. and Yang, Q.P., 2009. Prevalence and types of androgenetic alopecia in Shanghai, China: a community-based study. *British Journal of Dermatology*, *160*(3), pp.629-632.

Yamada, T., Hara, K., Umematsu, H. and Kadowaki, T., 2013. Male pattern baldness and its association with coronary heart disease: a meta-analysis. *BMJ open*, *3*(4), p. e002537.

Yamaguchi, K., Shirakabe, K., Shibuya, H., Irie, K., Oishi, I., Ueno, N., Taniguchi, T., Nishida, E. and Matsumoto, K., 1995. Identification of a member of the MAPKKK family as a potential mediator of TGF-β signal transduction. *Science*, *270*(5244), pp.2008-2011.

Yamazaki, M., Tsuboi, R., Lee, Y.R., Ogawa, H., Ishidoh, K. and Mitsui, S., 1999, December. Hair cycle-dependent expression of hepatocyte growth factor (HGF) activator, other proteinases, and proteinase inhibitors correlates with the expression of HGF in rat hair follicles. In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 4, No. 3, pp. 312-315). Elsevier.

Yap, C.X., Sidorenko, J., Wu, Y., Kemper, K.E., Yang, J., Wray, N.R., Robinson, M.R. and Visscher, P.M., 2018. Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nature communications*, *9*(1), p.5407.

Yassa, R. and Ananth, J., 1983. Hair loss in the course of lithium treatment: a report of two cases. *The Canadian Journal of Psychiatry*, *28*(2), pp.132-133.

Yazdabadi, A. and Sinclair, R., 2011. Treatment of female pattern hair loss with the androgen receptor antagonist flutamide. *Australasian Journal of Dermatology*, *52*(2), pp.132-134.

Yazici, K.U. and Percinel, I., 2015. Quetiapine induced hair loss in an adolescent case/Bir ergen olguda ketiyapin kullanimina bagli saç dökülmesi. *Anadolu Psikiyatri Dergisi*, *16*(2), p.143.

Yeon, J.H., Jung, J.Y., Choi, J.W., Kim, B.J., Youn, S.W., Park, K.C. and Huh, C.H., 2011. 5 mg/day finasteride treatment for normoandrogenic Asian women with female pattern hair loss. *Journal of the European Academy of Dermatology and Venereology*, *25*(2), pp.211-214.

Yilmaz, Y., Tasdemir, H.A. and Paksu, M.S., 2009. The influence of valproic acid treatment on hair and serum zinc levels and serum biotinidase activity. *european journal of paediatric neurology*, *13*(5), pp.439-443.

Yip, L., Rufaut, N. and Sinclair, R., 2011. Role of genetics and sex steroid hormones in male androgenetic alopecia and female pattern hair loss: an update of what we now know. *Australasian Journal of Dermatology*, *52*(2), pp.81-88.

Yoo, H.G., Chang, I.Y., Pyo, H.K., Kang, Y.J., Lee, S.H., Kwon, O.S., Cho, K.H., Eun, H.C. and Kim, K.H., 2007. The additive effects of minoxidil and retinol on human hair growth in vitro. *Biological and Pharmaceutical Bulletin*, *30*(1), pp.21-26.

Yoo, H.G., Kim, J.S., Lee, S.R., Pyo, H.K., Moon, H.I., Lee, J.H., Kwon, O.S., Chung, J.H., Kim, K.H., Eun, H.C. and Cho, K.H., 2006. Perifollicular fibrosis: pathogenetic role in androgenetic alopecia. *Biological and Pharmaceutical Bulletin*, *29*(6), pp.1246-1250.

Yu, V., Juhász, M., Chiang, A. and Mesinkovska, N.A., 2018. Alopecia and associated toxic agents: a systematic review. *Skin appendage disorders*, *4*(4), pp.245-260.

Zalsman, G., Sever, J. and Munitz, H., 1999. Hair loss associated with paroxetine treatment: a case report. *Clinical neuropharmacology*, *22*(4), pp.246-247.

Zang, Q., Kleinstreuer, N., Allen, D., Casey, W. and Judson, R.,2016. QSAR Modeling for the Predictions of Androgen Receptor Pathway Activity, accessed 10 June 2017, <https://github.com/zang1123/AR-QSAR-Modeling>

Zhang, S., 2011. Computer-aided drug discovery and development. In *Drug Design and Discovery* (pp. 23-38). Humana Press.

Zhang, Y., Tomann, P., Andl, T., Gallant, N.M., Huelsken, J., Jerchow, B., Birchmeier, W., Paus, R., Piccolo, S., Mikkola, M.L. and Morrisey, E.E., 2009. Reciprocal requirements for EDA/EDAR/NF-κB and Wnt/β-catenin signaling pathways in hair follicle induction. *Developmental cell*, *17*(1), pp.49-61.

Zhou, P., Byrne, C., Jacobs, J. and Fuchs, E., 1995. Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate. *Genes & development*, *9*(6), pp.700-713

Zhou, Z., Song, S., Gao, Z., Wu, J., Ma, J. and Cui, Y., 2019. The efficacy and safety of dutasteride compared with finasteride in treating men with androgenetic alopecia: a systematic review and meta-analysis. *Clinical interventions in aging*, *14*, p.399.

Zhu, H.L., Gao, Y.H., Yang, J.Q., Li, J.B. and Gao, J., 2018. Serenoa repens extracts promote hair regeneration and repair of hair loss mouse models by activating TGF-β and mitochondrial signaling pathway. *European Review of Medical and Pharmacological Science*, *22*(12), pp.4000-8.

Zhu, Y.Y., Huang, H.Y. and Wu, Y.L., 2015. Anticancer and apoptotic activities of oleanolic acid are mediated through cell cycle arrest and disruption of mitochondrial membrane potential in HepG2 human hepatocellular carcinoma cells Retraction in/10.3892/mmr. 2021.12079. *Molecular medicine reports*, *12*(4), pp.5012-5018.

Zimmerman, L.B., De Jesús-Escobar, J.M. and Harland, R.M., 1996. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell*, *86*(4), pp.599-606.

Zou, X., Hong, Z. and Zhou, D., 2014. Hair loss with levetiracetam in five patients with epilepsy. *Seizure*, *23*(2), pp.158-160.