University of Jyväskylä

Faculty of Mathematics and Science

Department of Biological and Environmental Science

Division of Biotechnology

Generation and Utilisation of Natural Product Library for Bioactivity Screening and Drug Discovery

Master's thesis, 2006

Author: **Pirkko Lepola**

Supervisor: **Docent Päivi Tammela**

Plants may;

Kill you fast

Kill you slowly

Dress you

Feed you

Cure you

Robert Verpoorte, 2005

Preface

This Master's thesis was carried out at the Drug Discovery and Development Technology Center (DDTC). The DDTC is an interdisciplinary research project belonging to the Faculty of Pharmacy at the University of Helsinki; it focuses on research work in drug discovery, development of key technologies, discovery of new drug candidates, and education of high level researchers.

I wish to express my gratitude to Professor Pia Vuorela for offering me the possibility to carry out this work at DDTC. I also want to express my deepest gratitude to my supervisor, Docent Päivi Tammela, Ph.D. who introduced me to the fascinating field of pharmacognosy and the plant based drug discovery. Her positive attitude and tireless efforts to guide my work have been the most important support during this study. I also want to express my warm thanks to research scientist Päivi Oinonen, M.Sc., whose technical expertise was remarkably essential in this work. I am also grateful to Jaakko Salonen for his valuable assistance during the process. I am grateful for the friendly and supportive guidance and practical help of all the researchers and personnel of the Drug Discovery and Development Technology Center during this study. In addition, I would like to thank my friend, Kirsi Hanslin, who offered her time and expertise in the English language to proofread this text.

I wish to thank Professor Christian Oker-Blom for having the opportunity to do all my studies under his excellent guidance. I appreciate and respect his educational expertise applied into practice in the field of biotechnology. And finally, I wish to thank him for examining and marking this paper.

Most of all, I want to express my gratitude to my beloved husband Vesa and my children, Lauri, Liisa and Lasse. Without their support, patience, encouragement and never ending love, I could not have been able to complete my studies and write this thesis. In addition to my family, I dedicate this work to my mother Raili, who has always supported me in my decisions to study further.

Abstract

University of Jyväskylä **Abstract of Master's thesis**
Faculty of Mathematics and Science

Author:	Pirkko Lepola
Title of thesis:	Generation and Utilisation of Natural Product Library for Bioactivity Screening and Drug Discovery
Finnish title:	Luonnonainekirjaston perustaminen bioaktiivisuusseulontaa ja lääkekehitystutkimusta varten
Date:	25.08.2006
Pages:	77+4
Department:	Department of Biological and Environmental Science
Division:	Biotechnology
Supervisor:	Docent Päivi Tammela[1]

New drugs are needed continuously. Natural products offer almost unlimited source of bioactive compounds and novel chemical structures for new drugs. By creating natural product libraries suitable for bioactivity screening, drug discovery from natural products can be intensified with shortened timelines. In this work a natural product (NP) library system was generated for the Drug Discovery and Development Technology Center (DDTC) which is an interdisciplinary research unit belonging to the Faculty of Pharmacy at the University of Helsinki.

The pilot NP library, including sample material and data base, was generated of 40 dried, randomly selected Finnish plants collected during the summer 2000. The plant samples were originally either whole plants or plants divided into two, four or six parts. Methods to produce the library material of plant parts were extraction (100% methanol), lyophilization and microfractionation by using the high-performance liquid chromatography (HPLC). The primary profiling was done by HPLC, using the reverse-phase column and gradient elution (methanol-water, B% 5-95) coupled with diode-array detection with two wavelengths (A: λ 230 nm, B: λ 280 nm). The integration of the NP library into further bioactivity screening was enhanced by creating a library data management system by using relational database, Microsoft Access XP (2002) software. The coding system was designed on the basis of the library classification and the nature of the samples.

As a result, the 23 735-membered NP library consists of four classes (I-IV): (I) Crude natural product material, (II) Lyophilized crude natural product extracts, (III) Crude natural product extracts in DMSO, and (IV) Fractionated natural product extracts. Total pivotal quantity of sub-samples (the library material) in Class I was 145, in Class II 145, in Class III 145, and in Class IV 23 300 microfractions in 290 96-well microplates. The NP-library material samples are stored either in dry conditions or in a freezer in the facilities of DDTC, and the library data is in forms of created NP-library data base (Microsoft Access), and as a relevant data in separate DDTC-NP-Library binders. The NP-library database has a reservation for two additional classes (V: Natural compounds and VI: Synthetic compounds) which may be added later on. The controlled and easy-to-operate NP-library enables DDTC, and possible co-operative parties, to perform further screening processes for discovering bioactive compounds or lead structures of potential new drug candidates from natural products more rapidly by shortening the total process time. The Microsoft Access software offers practical platform for both entering, searching and modifying the data. The library can be expanded and several types of additional complex data can be added to the library data base whenever needed.

Keywords: Natural product, library, drug discovery, HPLC

[1]Drug Discovery and Development Technology Center (DDTC), Division of Pharmaceutical Biology, Faculty of Pharmacy, University of Helsinki, Helsinki.

Tiivistelmä

Jyväskylän yliopisto **Pro gradu –tutkielman tiivistelmä**
Matemaattis-luonnontieteellinen tiedekunta

Tekijä:	Pirkko Lepola
Tutkielman nimi:	Luonnonainekirjaston perustaminen bioaktiivisuusseulontaa ja lääkekehitystutkimusta varten
English title:	Generation and Utilisation of Natural Product Library for Bioactivity Screening and Drug Discovery
Päivämäärä:	25.08.2006
Sivumäärä:	77+4
Laitos:	Bio- ja ympäristötieteiden laitos
Oppiaine:	Biotekniikka
Tutkielman ohjaaja:	Dos. Päivi Tammela[1]

Uusia lääkkeitä tarvitaan jatkuvasti. Koska luonnonaineet sisältävät bioaktiivisia yhdisteitä ja uusia kemiallisia rakenteita, ne tarjoavat lähes rajattoman uusien lääkeaineiden lähteen. Bioaktiivisuusseulontaan soveltuvan luonnonainekirjaston luomisen avulla voidaan luonnonaineisiin perustuvaa lääkekehitystutkimusta tehostaa lyhentäen prosessin normaalisti vaatimaa aikaa. Tässä työssä perustettiin luonnonaine (LA) kirjastojärjestelmä Lääkkeen keksintä- ja kehitysteknologian keskukselle (DDTC), joka on monitieteinen tutkimusyksikkö Helsingin yliopiston farmasian tiedekunnassa.

Kokeellinen LA-kirjasto, joka sisältää sekä näytemateriaalia että tietokannan, perustettiin kesällä 2000 satunnaisen lajivalinnan perusteella kerätyistä 40:stä suomalaisesta kasvista. Jokainen kasvi oli alun perin joko kokonaisena tai jaoteltuna kahteen, neljään tai kuuteen kasvin osaan. Kasvin osista tuotetun kirjaston menetelminä olivat uutto (100% metanoli), kylmäkuivaus ja mikrofraktiointi käyttäen korkean erotuskyvyn nestekromatografiaa (HPLC; high-performance liquid chromatography). Näytteiden esiprofilointi tehtiin HPLC:lla, käyttäen käänteisfaasikolonnia ja gradienttieluutiota (metanoli-vesi, B% 5-95), yhdistettynä diodirividetektoriin ja mittaamalla absorbanssi kahdella aallonpituudella (A: λ 230 nm, B: λ 280 nm). Integrointia jatkossa tehtäviin bioaktiivisuusseulontoihin vahvistettiin luomalla tiedonhallintajärjestelmä Microsoft Access XP (2002) relaatiotietokantaan.

Tuloksena oli 23 735 näytettä sisältävä 4-luokkainen LA-kirjasto (I-IV): (I) Kuivarohdokset, (II) Kuivauutteet, (III) Raakauutteet dimetyylisulfoksidiin (DMSO) liuotettuina, sekä (IV) Fraktioidut uutteet. Keskeisimpien näytteiden kokonaislukumäärät kirjastossa ovat luokassa I 145 kpl, luokassa II 145 kpl, luokassa III 145 kpl ja luokassa IV 23 300 mikrofraktiota 290:llä 96-kuoppaisella näytelevyllä. LA-kirjaston näytemateriaali säilytetään joko kuiva-olosuhteissa tai pakastettuna DDTC:n tiloissa, ja kirjaston tiedot ovat sekä tietokantatiedostoina (Microsoft Access) että soveltuvin osin tulostettuina erillisissä DDTC LA-kirjasto –kansioissa. Tietokantaan on tehty varaukset mahdollisille myöhemmin lisättäville luokille, jotka ovat (V) Luonnon puhdasaineet ja (VI) Synteettiset yhdisteet. Kontrolloitu ja helppokäyttöinen LA-kirjasto edistää sekä DDTC:n että mahdollisten yhteistyöosapuolten seulontatyötä, tarkoituksena löytää luonnonaineista bioaktiivisia yhdisteitä tai mahdollisten uusien lääkeaineiden alkurakenteita nopeammin säästämällä kokonaisprosessin aikaa. Microsoft Access tarjoaa käytännöllisen tietokannan sekä tiedon syöttämiseen, etsintään että muokkaamiseen. Kirjastoa voi myös laajentaa ja sinne voi lisätä monen tyyppistä lisätietoa tarpeen mukaan.

Avainsanat: Luonnonaine, kirjasto, lääkekehitystutkimus, HPLC

[1] Lääkkeen keksintä- ja kehitysteknologian keskus (DDTC), Farmaseuttisen biologian osasto, Farmasian tiedekunta, Helsingin yliopisto, Helsinki.

Table of Contents

Abbreviations

ADME	absorption, distribution, metabolism, excretion
AU	absorbance
BAC	bacterial artificial chromosome
B (%B)	the (% amount) strong solvent in a binary solvent mobile phase (mixture)
BAS	bioactivity screening
BH	library code for Class I samples (BAS Herb)
BE	library code for Class II samples (BAS Extract)
BED	library code for Class III samples (BAS Extract in DMSO)
BEF	library code for Class IV samples (BAS Extract in fraction)
DDTC	Drug Discovery and Development Technology Center
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DOS	diversity-oriented synthesis
DYMONS	diversity-modified natural scaffolds
FS	flower stem
HPLC	high-performance liquid chromatography
HTS	high-throughput screening
L	leaf
LC	liquid chromatography
LS	leaf stem
MEMO	text type that stores up to 64,000 characters in Microsoft Access
MS (Access)	Microsoft Access
MS	mass spectrometry
NMR	nuclear magnetic resonance
NP	natural product
NRPS	non-ribosomal peptide synthetase

OLE	An OLE (Object Linking and Embedding) object is a sound, picture, or other object such as a Word document or Excel spreadsheet that is created in another program. This data type is used to embed an OLE object or link to the object in the Microsoft Access database.
R	root
S	stem
SAR	structure-activity relationship
TLC	thin layer chromatography

1. Introduction

1.1 Drug discovery is demanding, risky and expensive

New drugs are needed continuously and the pharmaceutical industry is increasingly looking for more effective ways of screening the active compounds with lower costs and shortened timelines. The industry is moving away from the linear process of compound optimization towards a parallel strategy. This means shaping the profile of chemical entities in a multidimensional manner which allows the properties of a molecule to be appropriately balanced in a rapid, iterative fashion (Bleicher *et al.*, 2003). The natural product chemistry and organic synthesis, as well as new technologies together with existing libraries of previously discovered natural compounds are of tremendous help for screening and discovery of novel compounds from natural products, and for investigation of their biological activities. (Abel *et al*., 2002; Eldridge *et al*., 2002; Shen, 2004; FitzGerald, 2005; Shang and Tan, 2005; Tan, 2005; Van Lanen and Shen, 2006)

The research work of discovering and developing new medicines has been regarded as an increasingly complex and fast-paced area in both industry and academia, where it is essential to use the expertise of many scientific disciplines and technologies to resolve multi-parametric parallel processes as fast as possible (Eldridge *et al.*, 2002; Bleicher *et al*., 2003). Pharmaceutical drug discovery is a massive process measured by both time and money (Fig.1). Starting from target selection, identification and validation, it includes intensive research work with several types of knowledge, such as disease etiology and pathology, biological pathways and disease models. By using the libraries together with high-throughput-screening (HTS) it is possible to identify from a larger compound collection the most attractive and active compounds, and their ability to bind or otherwise inhibit specific macromolecular targets. Several possibilities, “hits” continue the process.

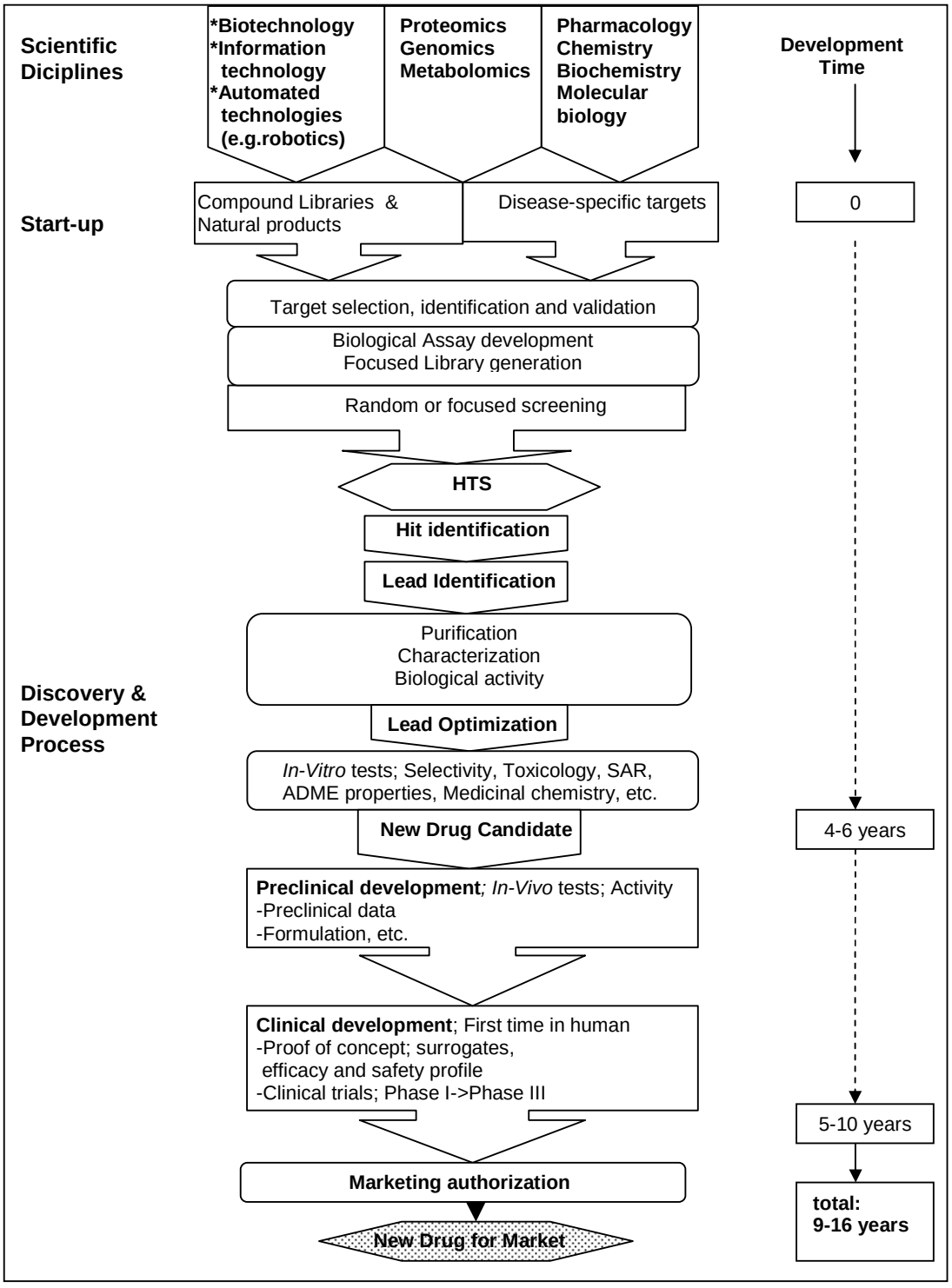

Figure 1. The general view of the drug discovery and development process.

The optimization of the lead compound needs a comprehensive assessment of chemical integrity, synthetic accessibility, functional behaviour, structure-activity-relationship (SAR) as well as bio-physicochemical and ADME (absorption, distribution, metabolism and excretion) properties. At least two series with best development potential are selected for further testing and profiling, synthesis and formulation, in case of unexpected failures due to unpredictable factors, for example toxicological findings in animal studies (*in vivo*) in preclinical phase (Bleicher *et al.*, 2003). The clinical development phases (I-III), e.g. studies with humans, are the most expensive parts of the development process. These studies are highly controlled by regulatory authorities and demand efforts from quite a number of people from different working sectors to finally proof the safety and efficacy of the new drug.

All in all, the whole discovery and development process for one lead compound can take approximately 10 to 15 years of development work before the marketing authorization can be applied for the new drug. By that time the average development costs have risen up to €1-billion ($800-million) and the figure is still lacking the additional marketing costs (Uehling, 2004; Davies, 2006).

1.2 Natural products in drug discovery

1.2.1 Definition of natural products

Natural product is the commonly used term of all biologically active living organisms, or part of organism (e.g. a leaf or a flower of a plant or an isolated organ of an animal). The definition includes plants, bacterial and fungal species, animals, insects and marine invertebrates, which have not been subjected to any treatment except perhaps to a simple process of preservation such as drying, and which can be used as a source of food, remedies, preventives and drugs. An extract of an organism or exudates can also be regarded as a natural product as well as a pure compound isolated from a plant or an animal. The term *crude drug* is used in pharmacognosy (i.e. the study of pharmacologically active natural products; *pharmakon*=drug and *gnosis*=knowledge) for those natural

products which are not pure compounds, such as plants or extracts (Samuelsson, 1992). The term of *natureceuticals* is used by some researchers when referring to the pharmacologically active natural products (McCurdy and Scully, 2005).

Natural products have played a major role in medicinal history. The traditional medicinal treatment system was created around natural products (Samuelsson, 1992). Natural products are a huge and mainly unrevealed source of chemical and functional diversity, and therefore an ideal source and starting point for screening pharmacologically active small molecules. The reason why natural products have become effective drugs in a wide variety of therapeutic indications, is the privileged structures selected by the evolutionary pressures to interact with a wide variety of proteins and other biological targets for specific purposes. Natural products provide a wide range of bioactive compounds which can be used as a source of novel drug leads, or as plant extracts in conventional way or as complementary therapy (Newman *et al.*, 2003; Vuorela *et al.*, 2004; Koehn and Carter, 2005).

1.2.2 Secondary metabolism and secondary products

In nature, survival and propagation are the basic functions of plant cells. These functions are based on the *primary,* or basic metabolism. This refers to all biochemical processes necessary for the normal anabolic and catabolic pathways which result in assimilation, respiration, transport and differentiation of a cell including general modification and synthesis of carbohydrates, proteins, fats and nucleic acids. These pathways exist in cells of all living organisms and are found to be essentially the same, apart from minor variations. But, plants cells produce far more chemical compounds than are necessary for their basic functions, and which have a much more limited distribution in nature. These compounds are so called secondary products, or *byproducts,* generated by *secondary* metabolism, which consists of a large number of diverse processes of the certain differentiated plant cell types found in only specific organisms, or groups of organisms, and are an expression of the individuality of species. These secondary products are not necessary for the cells themselves but may be useful for the plant as whole, e.g. give plants their colour, flavour and smell (Hiltunen and Holm, 2000).

Many of the natural products used as medicines, pesticides and spices belong to secondary metabolites. These metabolites are derived from the primary products, such as amino acids or nucleotides, by modification (methylation, hydroxylation, glycosylation etc.) and are therefore more complex. Plant pigments, alkaloids, isoprenoids, terpenes and waxes are secondary products. A number of these products are bacteriocidal, repellent by bad taste, or even poisonous. For example, the large majority of alkaloids are known as “mind altering drugs” such as nicotine, cocaine, ephedrine, dopamine and morphine. These are synthesized from amino acids or derived from purines or pyrimidines. (Samuelsson, 1992; Hiltunen and Holm, 2000). Secondary metabolites are found also in other bioactive organisms than plants. Today over 100 000 secondary metabolites have been recognized, but it still represents only a small quantity of the natural products existing in the nature (Vuorela *et al.*, 2004). During the evolutionary processes, animals developed a variety of dependencies to phytochemicals and used these products as precursors for the synthesis of vital or beneficial molecules in animal body. Secondary plant products have been used for thousands of years and the secondary metabolites represent a tremendous resource for scientific and clinical research. Compared to the situation in the preceding decades, such biologically active compounds of natural origin have nowadays a more promising role as potential new medicines. This progress is due to developed technologies in bioactivity screening and in generation of various types of screening libraries, as well as to the integration of combinatorial chemistry with natural product drug discovery.

1.2.3 Natural products as the source of new drug candidates

Natural products offer an almost unlimited resource for drug discovery. The estimated quantity of plant species existing in the nature is approximately 420 000. This includes at least 250 000 living species of the flowering plants which have been classified, but still, only 10% of the higher terrestrial plants have been systematically investigated. Additionally, the potential of the marine environment, bacterial and fungal species has not been fully exploited; less than 1% of bacterial and 5% of fungal species are currently known. More over, 61% of the 877 small molecule new chemical entities introduced as drugs worldwide during the years 1981-2002 can be traced to or were inspired by natural products. During the past 25 years about 60-75% of drugs approved to treat infectious

disease and cancer have been of natural origin, and approximately 25% of western medicines are derived from plant compounds. Despite the efforts and work with combinatorial chemistry during the past years in the pharmaceutical industry, the natural products field is still producing about 50 % of all small molecules (Houghton, 2000; Newman *et al.,* 2003; Vuorela *et al.,* 2004; Verpoorte, 2005).

Natural products are still being widely used in the form of medicinal plants, herbal extracts and finished products, the so-called phytopharmaceuticals. Today approximately 80% of the world population relies on traditional plant-based medicines for primary health care (Verpoorte, 2005). Natural products from micro-organisms have been a source of cancer drugs for many decades. Again, antitumour activity has also been found in compounds originating from other sources, such as trees (paclitaxel) or marine invertebrates like a sponge (discodermolide) (Koehn and Carter, 2005). Antibiotics have been discovered from bacteria and fungus (e.g. penicillin) and anti-inflammatory drugs from plants (e.g. salisylic acid). In a addition to these, natural products have offered treatment for pain (analgesic substances) and increasingly for many other indications in prophylactic or therapeutic forms, such as antioxidatives or T-cell supressors (Rauha *et al.,* 2000; Galvez *et al.,* 2005; Koehn and Carter, 2005; Lee *et al.,* 2005; McAlpine *et al.,* 2005; McCurdy and Scully, 2005; Singh and Barret, 2006).

As a natural product sample can contain 10-100s of different components, the sample has to be processed to have a single biologically active compound. The first step is to manipulate the sample in order to get an extract or a concentrate for further experiments. Sample extract can be fractionated by using chromatography techniques (e.g. reverse phase-HPLC) in order to separate and identify the active fractions. The complexity of extracts causes problems, and the acceptable purity and yield requires several steps of cleanup and separation before preparative production can be done. Active fractions are purified at the microgram level for rescreening and confirmation of activity. Structures of active compounds are elucidated by spectroscopic techniques (e.g. NMR spectroscopy and mass spectrometry). Additional scale-up purification is needed, because tens of milligrams may be required to elucidate the structure of the active compound. The purification process

is again scaled up to gram-level after preliminary SAR-studies for subsequent medicinal chemistry, e.g. synthetic work and for animal studies, with the novel compound (Fig. 2) (Eldridge *et al.*, 2002; Vuorela *et .al.*, 2004; Koehn and Carter, 2005).

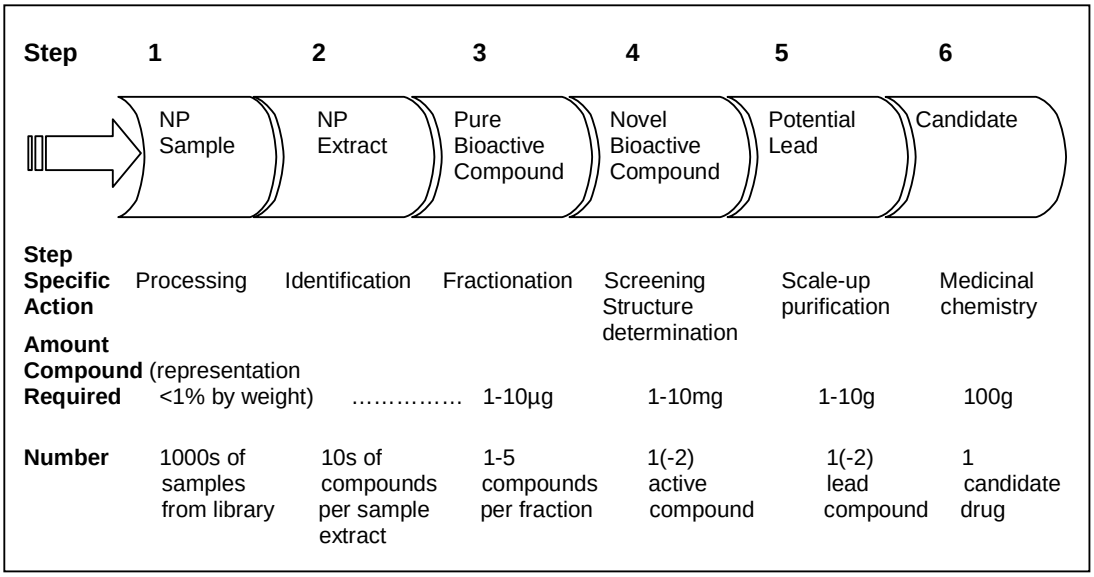

Figure 2. Chemical process for natural product (drug) discovery (modified from Koehn and Carter, 2005).

1.2.4 Natural products in drug development

The major problem with using natural products for drug development occurs in the isolation and purification of the active principles from a complex matrix. Natural products may also be limited in supply owing to sourcing limitations or impracticality of synthesis. As cells of natural products or organisms produce a mixture of different chemicals each present at a very low concentration level, it would take a huge amount of original product to produce the wanted compound on industrial scale. Quite commonly the target compounds are present in less than 1% of the weight of the crude extract. For example, to get 60 grams of discodermolide (mentioned previously), an anticancer compound produced by the rare Caribbean sponge (*Discodermia dissoluta*), would require 3000 kg of dried sponge. That is more sponge than exists in the world (Koehn and Carter, 2005).

Another example of the practical problems due to the supply is the anticancer drug paclitaxel, which was isolated from the bark of Pacific yew tree (*Taxus brevifolia).* The structure was elucidated by 1971 but led to human clinical trials as late as 1983. The development process of paclitaxel was difficult, for example, the low solubility made formulating into a stable product problematic, and its low natural abundance required large-scale extraction from its native source. The yield of active principle was in range 0.007-0.014 %. This meant the need of huge quantities of bark to produce required amount of paclitaxel for drug development; one hundred-years-old Pacific yew tree has 3 kg of bark which leads to 300 mg of Taxol (trade name for paclitaxel). Almost 30 000 kg of bark were extracted in 1989 for large-scale clinical trials of Taxol. Now, the major (late) intermediate in the biosynthesis of Taxol (10-decacetylbaccatin, 10-DAB) can be obtained from the leaves (needles) of many species of yew, and at concentration in excess of 0.1%. New chemical methods allow the synthesis of Taxol from10-DAB. Additionally, the semi-synthesis of Taxol also facilitates generation of Taxol analogues, which may have important value in the future (Hiltunen and Holm, 2000; Koehn and Carter, 2005).

Moreover than limitations, the natural products seem to have many advantages in drug discovery, such as the biological activity, complex novel structures and different structural features than synthetic compounds (Boldi, 2004; Vuorela *et al.*, 2004; Verpoorte, 2005).

1.2.5 Production of secondary metabolites

Fermentation is a useful and an old method to be used especially in case of micro-organisms like bacteria, which have an excellent expression rate of secondary metabolites. It also provides an economical way to generate natural product libraries. One alternative method to produce secondary metabolites is plant cell cultures, where it is possible to grow cell lines that produce biologically active, very specific and expensive compounds which are too complex or impossible to synthesize, or whose marketing prize (value) is high enough to make it worthwhile or profitable to produce it (Samuelsson, 1992; Hiltunen and Holm, 2000).

Each plant cell, containing all the genetic information relating to the whole plant, enables to start the culture with a single cell which is allowed to multiply by division to form a tissue of loosely attached cells called *callus.* In practice though, the callus culture is started with a small piece of plant tissue. This callus culture can be kept growing in a suitable medium, but it is an unhomogenous and a fairly slowly growing culture. As suspended in suspension culture, the callus culture cells exist separately in the medium or form small aggregates. The cell lines grow in three phases; lag-phase, exponential-phase and stationary-phase. A maximum production of secondary metabolites is detected during the stationary phase. If the cells are alive, the secondary metabolites are not excreted from the plant cells as they do from the microbial cells. The extraction of secondary metabolites can be done by mechanical or enzymatic degradation of cell walls, after collecting the cells at a suitable stage, or pumping the medium through material (e.g. ion exchanger) which absorbs the metabolites and from which they can subsequently be recovered. The cell cultures are faster (in weeks) in producing the secondary metabolites than growing the whole plants (in years) or other natural products, e.g. trees (decades). However it is questionable whether this method is cost-effective enough to harvest these byproducts by using callus or suspension cultures. Some products are produced in industrial scale in massive bioreactors for economical reasons (Samuelsson, 1992; Hiltunen and Holm, 2000).

There are certain methods to increase this often unstable and low-yield process of producing the secondary metabolites. One way is the use of *elicitors*. These elicitors are certain compounds which are released from the plant cell walls or the cell walls of a micro-organism when it attacks a plant. These (elicitor) compounds have been isolated from several bacteria and fungi. Additionally, some physical and chemical stress factors; such as various radiations, heat- and cold shocks, ethylene gas, fungisides, antibiotics, salts of heavy metals or high salt content can function as elicitors. The elicitors activate those genes in the plant which code for the enzymes needed for phytoalexins. Phytoalexins are plants´ own antibiotically active secondary metabolites which protect the plant against many microbes and infections but are not normally present. Phytoalexins are harmful to the infecting organism and are part of the defence system. These substanses are found to be in many plant species. Using elicitation process, by adding the elicitor into the medium in the

end of exponential phase, it is possible to increase the excretion of the desired metabolite as much as many hundred folds within a few hours (Samuelsson, 1992).

1.2.6 Genomics and biotechnology for secondary metabolites

Today, by analyzing the gene sequences and by using combinatorial biosynthesis and genome mining, that is, the analysis of gene sequences and gene clusters involved in encoding secondary metabolites in biosynthesis, it is possible to discover novel secondary metabolites (Shen, 2004; Lautru *et al.*, 2005; McAlpine *et al.*, 2005). For example, in microbial genomes, the natural-product biosynthetic genes are present in clusters and thereby it is possible to estimate the biosynthetic potential for a given organism by mining the whole-genome sequence. After the whole-genome sequencing, it has been discovered that the biosynthetic potential for natural products in micro-organisms has been greatly under-explored by traditional methods, due to many new revealed biosynthetic gene clusters other than currently known metabolites for given organism. Additionally, the variation of a few common biosynthetic machineries can account for a vast structural diversity observed in natural products (Van Lanen and Shen, 2006). As examples, the utility of genome mining has been demonstrated with *Streptomyces aizunensis* showing its potential to produce a novel antifungal compound (McAlpine *et al.*, 2005), and with *Streptomyces coelicolor*, which revealed several gene clusters encoding new non-ribosomal peptide synthetase (NRPS) systems not associated with known metabolites (Lautru *et al.*, 2005). The analysis of genome and specifically the genes encoding enzymes in the biosynthetic pathways has lead to the understanding of the biosynthesis of complex secondary metabolites. By combining the genomics and information technology it is possible to analyze the genome and predict the structure, completely or partially, that is capable of producing active components after isolation of an organism.

Secondary metabolites are very important class of natural products for industry and biomedical applications. Secondary metabolite pathway engineering is an approach to produce previously inaccessible compounds in microbial cells. The biosynthesis of novel products and directed synthesis of desired products at higher production levels are obtained by metabolic engineering of both native and heterologous secondary metabolite producing

organisms, utilizing the knowledge of cellular metabolism and extensive screening. By combining the genes responsible for individual metabolic pathway steps from different source organisms it is possible to generate novel branches in metabolic pathways and to biosynthesize previously inaccessible products. This method is known as combinatorial engineering and it is used in addition to metabolic engineering in producing secondary metabolites such as isoprenoids, polyketides and biopolymers, in micro-organisms (Mijts and Schmidt-Dannert, 2003).

The secondary metabolites of marine invertebrates show exceptional promise as potential pharmaceuticals in many therapeutic areas, and are therefore one of the exiting new sources in the area of natural products. However, bioactive compounds found in marine invertebrates have turned out to be extremely difficult to synthesize, and therefore cannot be produced effectively in industrial scale. The exploitation of these metabolites can be hampered by economical aspects related to sustainable supply. One promising option is to clone the genes encoding the biosynthetic expression of a lead metabolite into a surrogate host suitable for industrial scale fermentation (Jaspars *et al.*, 2005; Dunlap *et al.*, 2006,). An example of this kind of recombinant expression of a marine bioactive metabolites has been demonstrated with the heterologous expression of patellamides. This sustainable production method of secondary metabolite from marine invertebrate, the seasquirt *Lissoclinum patella*, includes: the cleaving of the lead metabolite genes, insertion into a vector (BAC), introducing the BAC into the host (*E.coli*), producing the clones containing the vectors with DNA inserts, culturing clones (fermentation), extracting culture media, and screening and identification of the active compound (Jaspars *et al.*, 2005).

1.2.7 Modern methods in exploiting natural compound structures

To optimize productivity, pharmaceutical industry has adopted various methods to produce lead compounds more rapidly and cost-effectively for the use of drug discovery and development. Different modification and synthetisation methods have also been applied to increase the potential of natural products in generating new drugs. If the product supplies of the source are limited, synthetisation can be used in developing natural product scaffolds from a derivative which has been discovered to have promising screening results. Total

synthesis offers a possibility to manipulate a molecule and make variations of its original structure. By this way the limited but crucially important features of the molecule can be improved to increase the potency and a favourable therapeutic window leading to new drugs. This is particularly important in antibiotic design as the bacteria are constantly evolving resistance to treatments (Koehn and Carter, 2005; Singh and Barret, 2006). Furthermore, in cases like discodermolide (mentioned before), where the natural supply is very limited (rare) and the isolation yield really low (~14 mg/kg) it is necessary to find other ways to produce the active compound. With discodermolide the only reasonable way to produce these rare compounds in larger scale was synthetisation, but it took almost two years and 39 different steps to produce only 60 mg of synthetic active compound (Koehn and Carter, 2005).

Sometimes the complexity of a natural product compound may generate problematic side effects when used as a drug for humans. Sometimes the candidate compound may have specific potency for particular target but still not be practical for use. It may be limited in supply, too expensive, or it may have certain unsuitable pharmaceutical properties or metabolic liabilities. One way of limiting these unwanted effects and of overcoming the practical problems, is to create a model for synthetic mimetics by understanding the binding interactions of the natural product. Another option is to construct the molecule by using technology to alter the molecule so that side effects are reduced and efficacy improved. The crucial structural elements required for biological activity are defined, and the potent and selective products can be derived with fewer synthetic steps and at reasonable costs. For example with small peptides, the systematic variation of the individual amino-acid residues (side chains) allows to find the structural features essential for biological activity. The mimic version of the natural product agent still has those chemical and biological features necessary to maintain the biological activity as a drug molecule (Koehn and Carter, 2005). These natural-product analogs or synthetic mimetics, and the libraries of these, can be used further for drug discovery of new bioactive compounds. Furthermore, in case of highly toxic natural products, for example cytotoxic agent for the treatment of cancer, it is possible to overcome the lack of selectivity by conjugate formation as delivery system, where the compound is linked to monoclonal antibodies and is cleaved under intracellular conditions (Koehn and Carter, 2005).

Natural products can also be used as building blocks to produce privileged structural motives. The core structures (skeletons or scaffolds), specific substructures found across a class of natural products, or general structural characteristics of natural products can be used by diversity-oriented synthesis (DOS) of novel small natural product-like molecules. The DOS approach is used to create a collection of compounds that are maximally diverse and to identify which of these new compounds are biologically active (Shang and Tan, 2005; Tan, 2005).

Combinatorial biosynthesis is complementary to chemical synthesis and microbial fermentation covering the gene manipulation of the secondary pathways. It offers an alternative for preparation of complex natural products and their analogs biosynthetically (Shen, 2004). Combinatorial chemistry, also exploitable in natural product drug discovery context, is a widely used method to generate a large collection (e.g. libraries) of compounds by synthesizing combinations of a set of smaller chemical structures simultaneously in a time usually taken to prepare only a few by conventional methods.

The current drug discovery, generally and therefore also from natural products, has moved towards a genomic approach due to advances in genomics, molecular biology and particularly in biotechnology (Koehn and Carter, 2005). The genomic-driven drug discovery has lead to the use of focused libraries of small drug-like molecules for the identification and validation of novel drug targets. This approach, known as chemical genomics or chemogenomics, is greatly aided by computational technologies (Bleicher *et al.*, 2003). One example of this chemogenomics is a ~ 50-membered natural-product inspired synthesized library of protein-reactive chemical genomics probes, which was used to identify a compound inhibiting breast cancer cell proliferation in cell-based screening (Evans *et al.*, 2005). Though combinatorial chemistry will have very important role in modern drug discovery, the future trend appears to be toward the synthesis of complex natural product-like libraries and the combination of natural product synthesis, combinatorial chemistry and combinatorial engineering (Mijts and Schmidt-Dannert, 2003; Newman *et al.*, 2003; Rose and Stevens, 2003; Shang and Tan, 2005).

1.3 Natural product- and natural product-derived libraries

1.3.1 Requirements for a good library

Historically, natural product drug discovery has been very time-consuming and laborious process (Eldridge *et al.*, 2002; Vuorela *et al.*, 2004; Koehn and Carter, 2005). For this reason, the drug discovery processes have been done with the aid of randomly synthesized screening libraries, (i.e. compound collections) not favouring natural products during a long period of time. Nowadays the advanced new technologies in HTS with the combination of combinatorial chemistry have increased the possibilities to use natural products and natural product-derivates in drug discovery, by shortening the cycles in screening process, and by providing large natural product libraries suitable for HTS (Abel *et al.*, 2002; Eldridge *et al.*, 2002; Bleicher *et al.*, 2003).

Generation of a natural product library that fits the need of HTS programmes is the main task in order to speed up and help the screening procedures in natural product- related drug research (Abel *et al.*, 2002; Eldridge *et al.*, 2002 Bleicher *et al.*, 2003; Vuorela *et al.*, 2004; Koehn and Carter, 2005). The library should be suitable for effective and practical use, and the quality and quantity of the library samples play a pivotal role in the success of HTS programmes. Additionally, the samples of natural origin have to be competitive with synthetic compound libraries. Therefore, the library should contain pure high-quality compounds or compound mixtures, stored on microplates that fit HTS programmess. In addition, the library should contain adequate information to allow rapid identification and localisation of the sample or compound of interest (Abel *et al.*, 2002; Vuorela *et al.*, 2004). It is important that the information of the library is in a practical form, a clever data management system and suitable storage conditions are needed in case of further studies. Furthermore, applicable methods allowing further processing such as detection, separation and identification of potential active compounds are usually inevitably required.

There has been a general shift in the pharmaceutical industry away from screening natural product extracts towards screening pre-fractionated extracts or even pure natural product compounds. In addition to pharmaceutical industry, today even many commercial

laboratories (e.g. LGC and Albany Molecular Research, Inc.) as well as research institutes have generated their own natural product libraries (Abel *et al.*, 2002; FitzGerald, 2005, Koehn and Carter, 2005). The aim is pre-screened and pre-developed highly functional and chemically diverse natural compounds, with the knowledge about the relationship between chemistry and biology. And specifically, the economic generation of focused collections of chemically and functionally diverse compound libraries, which can be easily used in different screening procedures. This will increase the chance of finding compounds among natural products which have the ability to interact with human proteins (Abel *et al.*, 2002; Koehn and Carter, 2005). In addition to pre-fractionated and pure natural-product libraries, parallel synthesis gives access to synthetic, semi-synthetic and natural-product-like libraries (Abel *et al.*, 2002; Bleicher *et al.*, 2003; Boldi, 2004; Shang and Tan 2005).

1.3.2 Library material

Library material of original natural products, such as preserved material, cultures of micro-organisms, plant or herbal extracts etc., can be produced by several types of methods described later (chapter 1.4). In addition to the informative data received from the compound analysis, i.e. structure, activity, pharmacological and physiochemical parameters etc., the starting material itself has to be saved as a collection or library in the same way. Natural product library can be composed of samples, such as extracts (mixture), semi-pure mixtures or single purified natural compounds. The difference between these materials is the complexity or heterogeneity of the sample. Extracts are the most complex ones and can include 10-100s of components, whereas semi-pure mixtures only 5-10 compounds and pure products represent one compound. The complexity is the main issue that challenges the HTS programmes (Koehn and Carter, 2005). This material library creates the fundamental foundation for further studies (e.g. bioactivity screening) of the natural product and it has to be saved or stored by using material specific methods relating to the further planned activities (e.g. in cultured form, lyophilized form, frozen form etc.) in different platforms (e.g. different culture plates, microplates, test tubes, solution etc.) (Eldridge *et al.*, 2002; Lee *et al.*, 2005).

1.3.3 Library strategies and approaches

Basically, the pharmaceutical industry and drug research units have had three types of libraries, having the size up to approximately one million entities in large pharmaceutical companies. These libraries represent a) historical collections, which are intermediates and precursors from earlier research programmes, b) natural products, and c) combinatorial libraries (Bleicher *et.al.*, 2003). Nowadays different sized libraries are available for drug discovery process, represented and provided by various commercial laboratories and companies (e.g. LGC and Albany Molecular Research, Inc.). These libraries can be generated and focused by applying several design approaches depending on the starting material or the demand of the research activities (Fig.3).

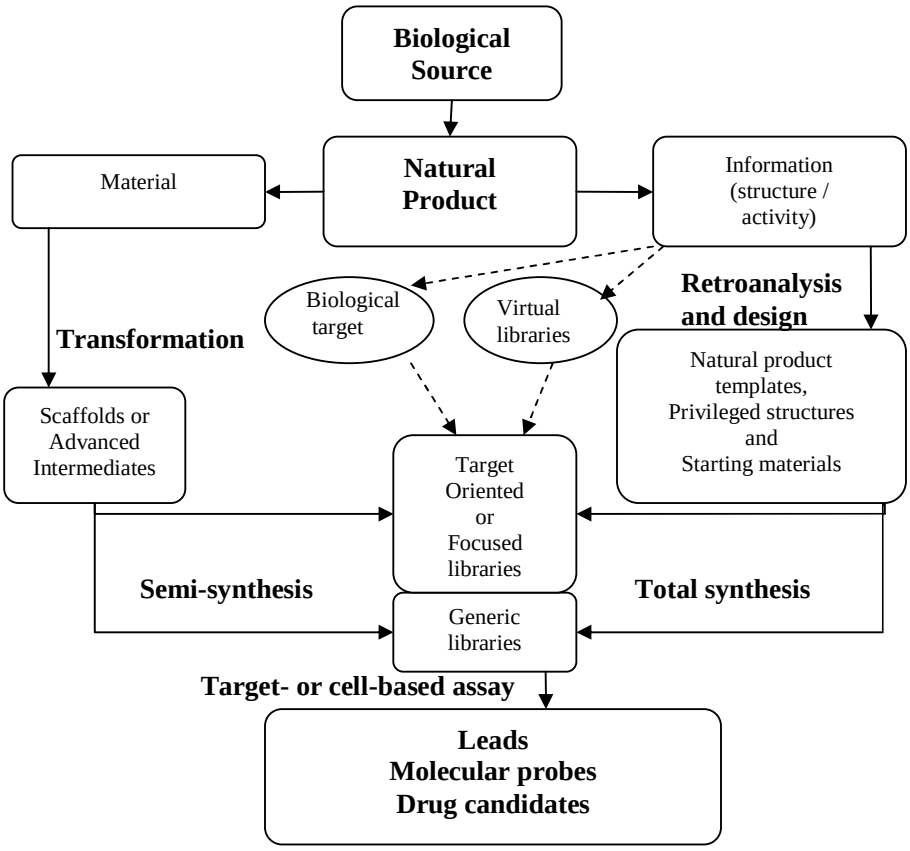

Figure 3. Approaches towards natural-product-derived libraries depending on the accessibility of starting materials and library focus (modified from Abel *et al.*, 2002).

Natural products can be used for libraries by serving start material or by serving structure information leading to different kind of libraries. There are several methods with diverse concepts and construction to produce or generate natural product libraries, or natural product-like libraries. The library elements can include pure natural compounds, semi-pure natural compounds, natural-product-like compounds, or synthetic -, or semi-synthetic compounds derived from natural products, alone or as combinations (Abel *et al.*, 2002; Rose and Stevens, 2003; Boldi, 2004; Shang and Tan, 2005; Tan, 2005; Mang *et al.*, 2006).

1.3.4 Pure -, semi-pure – and prefractionated libraries

A pure compound library includes highly purified and characterized single biologically active compounds from natural sources or derivates with a purity of 90% or more. One example of pure compound libraries, Natural Product Pool, with more than 6000 compounds derived from different European laboratories has been established in Germany. This pool is funded by several industrial partners, and the Pool delivers approximately 1000 pure substances in quantities of 1 mg per compound per year to these industrial partners. The access to subsequent deliveries of a substance or options and licences on substances, as well as possible patent applications, are secured by contract regulations and different guidelines (Abel *et.al.*, 2002) As a demonstration of high-throughput methods applied to the production, analysis and characterization of natural product compound library, a library of *Taxus brevifolia* consisting of 36 000 fractions of detectable compounds was made. From this library, smaller more focused libraries were drawn for screening. A total of 160 preparative HPLC fractions were collected and screened. As a result, a total of 147 bioactive Taxane compounds were detected, identified, analyzed and purified from this smaller library (Eldridge *et al.*, 2002).

The bioLeads (Germany) has processed a combined pure compound library from natural sources such as plants, fungi, terrestrial and marine actinomycetes, and synthesis with detailed structure information. The compound purity is 90-98% and the structures and strereochemistry are confirmed by various physiochemical analytical methods (Abel *et al.*, 2002).

Semi-pure compound library includes semi-characterized entities with purity of greater than 80%. According to the evaluation of 4000-membered non-redundant substances from plant, bacteria and fungi, made by AnalytiCon and Aventis Pharma, the major problem with these compounds was the high number of redundant and ubiquitous (i.e. extra) substances. The comparison between these semi-characterized pure compounds and a combinatorial library (Rhone-Poulenc Rorer) was found to be ideal for modern drug discovery, because this type of combined-library includes derivates from synthesis and natural sources, the compounds are complementary to one another; the distribution of pharmacophores is divided so that natural products contain more oxygen and synthetic substances more nitrogen groups. But, it is expensive and takes much time to generate such a library (Abel *et al.*, 2002).

One practical approach is to generate libraries of prefractionated extracts of natural origin. These libraries can be based on the subfractions of the pre-purified natural product extracts which can be derived from cultures of filtered micro-organisms, such as fungi and bacteria, or from plant material extracts. It is possible to generate thousands of subfractions per month by using semi-automatic processes, and the aliquots of subfractions can be delivered to customers for their screening systems. The repository of these subfractions (e.g. the original fractionated material) can be micro-fractionated by the function of time and re-tested, which allows perform dereplication, i.e., to identify and eliminate known compounds that have been studied in the past from the screening process. By that way it is possible to decrease the number of structures that will need to be fully elucidated, because the natural product mixtures contain compounds that have been previously characterized by the structure and the biological activity (Abel *et al.*, 2002; Singh and Barret, 2006). The natural product extract is identified (chemical fingerprinting) before preparative fractionation to minimize the risk of compound duplicates. Prefractionation offers advantages by shortened timelines and low costs because structures of active fractions are elucidated and only non-replicable substances have to be re-processed further (Abel *et al.*, 2002; Tammela *et al.*, 2004; Wennberg *et al.*, 2004). One library concept of this kind has been generated by a company called bioLeads (Germany) using purified (average 85%) natural products without structure information. The subfractions are further fractionated by

peak-based semi-preparative HPLC, and, within one month, it is possible to get more than 80 000 substances (Abel *et al.*, 2002).

1.3.5 Combinatorial libraries derived from natural products

At the beginning of the 1990s when HTS and combinatorial chemistry first emerged, the drug discovery and search for lead compounds had been done mainly by assaying vast, diversity-driven compound collections created by combinatorial chemistry. These combinatorial libraries were developed by different techniques enabling generation of large numbers of novel synthetic chemicals (Bleicher *et al.*, 2003; Rose and Stevens, 2003) Many pharmaceutical companies maintain basic libraries of several hundred thousand natural and synthetic compounds, but by combinatorial chemistry the library can easily contain millions of compounds. The latest computational technology offers the possibility for rapid virtual screening of active components from these libraries; one library can be screened in few hours. These libraries can be generated relatively easily and the library is reusable and can be re-screened if needed.

Natural products can also be exploited in combinatorial chemistry by looking at scaffold architecture. Natural products may contain novel and potential structure with wanted properties. For example, the common topological pharmacophore patterns between trade drugs and natural products can be systematically explored by creating selected combinatorial libraries based on a combination of natural product-derived and synthetic molecular building blocks. After identification of novel scaffolds a virtual combinatorial library can be generated (Lee and Schneider, 2001). Combinatorial chemistry offers also a possibility to create as many variations as possible around a core chemical scaffold to find out which variants have desirable properties. It is an especially effective method in case of antibiotics. By taking a natural molecular scaffold and modifying it has given several next-generation synthetic antibiotics (Koehn and Carter, 2005). Also, a natural product-based library can be created by using one known compound which is easily available, and has favourable arrangement of functional groups from combinatorial chemistry´s point of view. Such synthesized libraries, like 360-membered andrographolide library (from *Andrographis paniculata*), can be used for checking physiochemical parameters and for

searching the pharmacological relevance of the library members, i.e. natural product derivates (Mang *et al.*, 2006). Additionally, a selection of diversity-modified natural scaffolds (DYMONS) offers one type of biased natural product derived library, in which diverse substituents are introduced to lead-like natural product cores (Abel *et al.*, 2002; Boldi, 2004).

Generic libraries consist of highly diverse compounds covering larger areas in diversity space and are a contrast to focused or biased libraries. These libraries can be used for cell-based and chemical-genetic assays leading to previously unidentified targets and biochemical pathways, and eventually to new drug leads (Abel *et al.*, 2002). Diversity-oriented (DOS) semi-synthesis or diversity-oriented total synthesis are methods to generate natural product-like libraries that explore untapped or underrepresented regions of chemical structure space. These library members created by DOS are skeletally diverse, structurally complex, stereochemically rich and densely functionalized offering easy re-synthesis and access to analogs. These libraries have thereby increasing probability that wide range of biological targets (different proteins) will be binding and interacting with different compounds in the library (Shang and Tan, 2005; Tan, 2005).

1.3.6 Focused libraries derived from natural products

The collaboration of computational scientists and chemists creates library proposals that fit the target structure requirements and that are simultaneously amenable to parallel synthetic assembly. The mechanism of action of a biological target is an important aid in these mechanism-based libraries. Another targeted compound collection type is ligand motif-based library, which is relevant for targets for which very limited or no biostructural information is available. The advantages of natural products being “start-up material” for focused libraries, are the privileged structures with high content of information (3D – structure), biosynthetically derived and encoded in conservative gene clusters, capable to interact with a broad variety of proteins, and not being randomly synthesized. (Bleicher *et al.*, 2003) The selected natural scaffolds are ideal starting points for focused libraries for further screening and synthesis, hopefully giving a dramatic impact on the generation of new lead compounds (Boldi, 2004).

Target-oriented semi-synthesis can be started from isolated natural products or semi-synthetic advanced intermediates, and the strategy of target-oriented total synthesis, is applied if the starting material can not be obtained in sufficient quantities from natural sources, or if substituents need to be varied on positions not accessible by semi-synthesis (Abel *et al.*, 2002; Boldi, 2004; Shang and Tan, 2005; Tan, 2005). The size and design of focused or biased library depends on the knowledge about the addressed biological target. If the target is known for a valid lead, the library of derivates to be synthesised to explore the neighbouring diversity space, is relatively small (Abel *et al.*, 2002).

1.3.7 Virtual libraries in natural product drug discovery

Moreover, sophisticated informatics tools derived from computational chemistry technologies provide a rapid and parallel mining with initial screening processes of the libraries. Virtual library investigations can be also used as a supplementary aid in natural product – or natural-product-derived compound discovery, in re-engineering the process of generating chemistry ideas, and also in screening toxicity, bioavailability and patent issues. The virtual compound collection serves as the foundation for the entire discovery program; library design, hit-validation and the exploration of lead compounds are all functions using the virtual collections. After finding a “hit” compound, the research work can continue with searching technology and return to original virtual collection to generate focused libraries for validating “hits” or explore initial SARs. All the members of the new focused library can now be synthesized and screened (Hecht, 2002).

1.3.8 Future expectations for libraries

From a mathematical point of view, the virtual generation of chemical diversity seemed to be unlimited and by the aid of information technology the possibility to create infinite number of synthetically tractable compounds in combinatorial libraries has fascinated chemists since then (Abel *et al.*, 2002; Bleicher *et al.*, 2003). But, the philosophy behind combinatorial library design has changed, because the early libraries showed disappointing results after biological testing and did not result in the increase of drug candidates (Bleicher *et al.*, 2003; Rose and Stevens, 2003). Today, most of the previous libraries have

been largely eliminated and although the discussion of library size is still going on, efforts have been set towards smaller, high-purity and information rich focused libraries (Bleicher *et al.*, 2003; Rose and Stevens, 2003).

It is expected that the application of HTS technologies will move from pure random testing of large compound pools to iterative rapid feedback screening of smaller, but more focused libraries (Bleicher *et al.*, 2003). Additionally, the design strategies used to generate libraries are more frequently biologically focused, virtually screened for a variety of ADME- and toxicity properties, intending to hit a single biological target or family of related targets (Rose and Stevens, 2003). These focused libraries with a significant medicinal chemistry design component can be generated by support from modern integrated computational and synthetic methods and is now more commonly adapted for finding "hits" and lead compounds (Bleicher *et al.*, 2003; Rose and Stevens, 2003). Additionally, recent efforts have been focused also to smaller natural product-like libraries, created by diversity-oriented synthesis (DOS) for providing more flexible synthesis of the compounds (Shang and Tan, 2005).

1.4 Applied technology in library design and generation

1.4.1 Automation and information technologies in library design

Efficient screening operation includes the management of vast amount of data generated from designing the libraries and screening them trough many different assays. Effective and accurate data management systems are essentially needed when time and effort are expended in the analysis and recording of a vast amount of generated data. Information technology has been of tremendous help in library design programmes (Koehler and Villar, 2000; Lee and Schnider, 2000; Hecht, 2002; Rose and Stevens, 2003; Mang *et al.*, 2005) as well as automation technology. For example, it easies the exploitation of natural products in screening campaigns by microfractionating extracts of biogenic origin into 96-well plates by using high-performance liquid chromatography (HPLC) combined with

automated fraction collector (Tammela *et al.*, 2004; Vuorela *et al.*, 2004; Wennberg *et al.*, 2004).

Using today's enormous virtual libraries to search through chemical space for topologically similar entities compared to known actives (structures), requires fast and effective searching strategies and screening tools, algorithms, for the identification of desirable compounds. Algorithms, like the LASSOO-algorithm, can identify the most different from those already present in a screening set and in the reference set of undesirable compounds, while being simultaneously most similar to set of compounds with desirable characteristics (Koehler and Villar, 2000). Another type of algorithm is based on the topomer-shape similarity, which identifies similar compounds by comparing steric interactions between given query and molecules in a virtual library. This algorithm (made by Tripos^{TM}) can rapidly scan vast libraries at the rate of 2 trillion three-dimensional structures per hour (Hecht, 2000). All in all, there is a huge selection of different kind of algorithms available today for drug discovery programmes to aid high-throughput technology, and these are applied for the validation of compound collections to be purchased for their drug-likeness, chemical diversity and similarity to the existing ones. Despite the wide variety of algorithms, i.e. informatics available, they can be regarded as prediction tools only (Bleicher *et al.*, 2003). High-throughput technology is also not sufficient enough if used alone, but the combination of these two can give competitive advance to improve drug discovery programmes (Hecht, 2000).

1.4.2 Relational databases [MicroSoft Access XP (2002)] in library design

Nowadays, when drug discovery programs are dealing with increasingly complex datasets, a demand for more efficient data capture in practical format has emerged to increase the productivity in discovery projects. Relational database has the ability to represent complex data using normalized data model, and it allows the handling of scientific data of pharmaceutical or biotechnological origin to create reports and search for information by any criteria. Relational databases have a non-hierarchical model in which all data tables are equal and relationships can be created between tables. Therefore, many pharmaceutical and biotechnological companies are currently using relational databases as a current standard

for storing and retrieving data generated from drug discovery programs (Hewitt *et al.*, 2005).

MS Access (Fig. 4) is a relational database management system from Microsoft, packaged with Microsoft Office Professional. Microsoft has released several versions of this program during the last few years, latest being MS Office Access 2003. MS Access is widely used in different forms depending on the usage. It can be applied to all kinds of small business, dealing with data which needs to be stored and classified in sophisticated manner, allowing users to handle and search the data with little effort. Certain features of Access are powerful design tools: rapid report creation, selective queries and easy-to-view -forms. (Töyli, 1998; Florida Gulf Coast University 2000; Hautamäki, 2005; Access 2003 Product Information; Microsoft Office Access –Wikipedia, 2006).

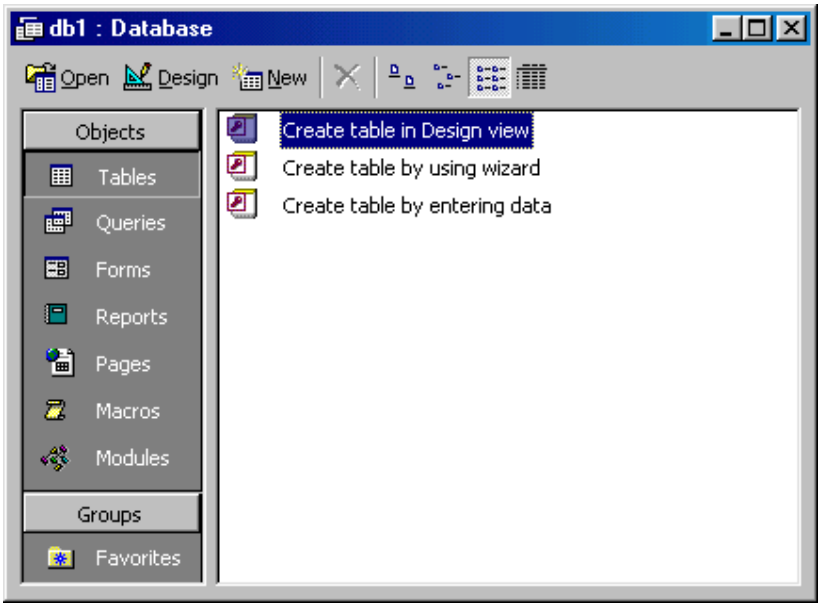

Figure 4. The Database Window (Florida Gulf Coast University, 2000).

The basic element of MS Access is the table which is relatively easy to create. Design View allows defining the fields in the table before adding any data to the datasheet. The window is divided into two parts: a top pane for entering the field name, data type, and an option description of the field, and a bottom pane for specifying field properties (Fig. 5).

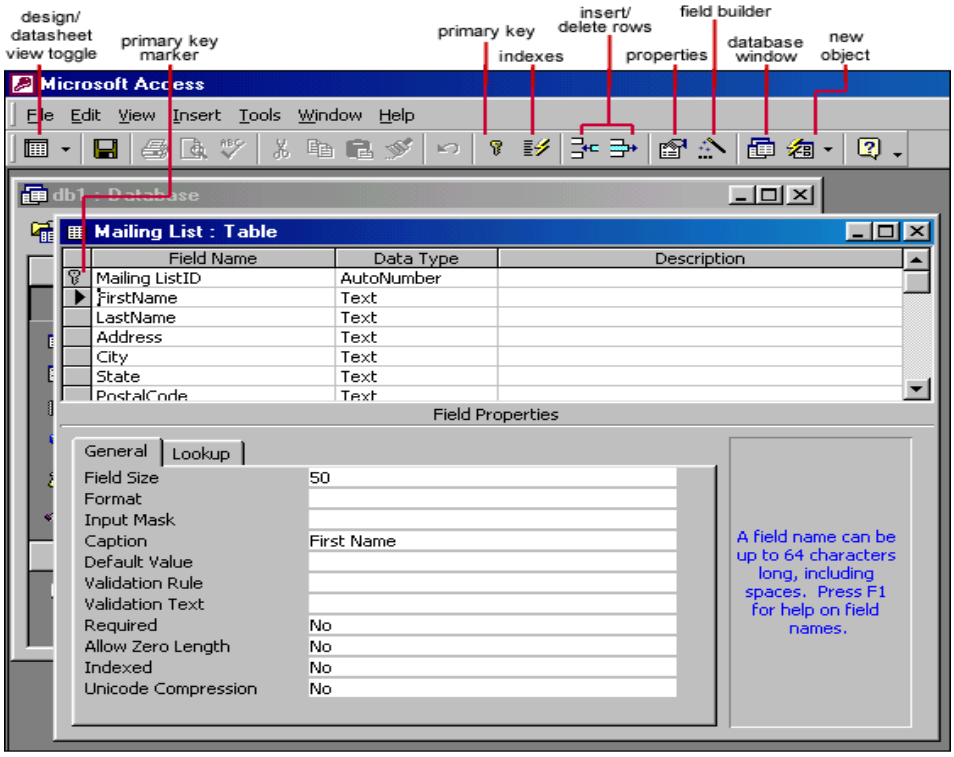

Figure 5. MS Access XP (2002) database design view and icons in the menu bar (Florida Gulf Coast University, 2000).

Data Type is the type of value that will be entered into the fields, for example: text, memo, number, date/time and OLE object, or a hyperlink based information. Also currency, autonumber and yes/no –options can be added (Florida Gulf Coast University, 2000).

OLE Object - An OLE (Object Linking and Embedding) object is a sound, picture, or other object such as a Word document or Excel spreadsheet that is created in another program. Use this data type to embed an OLE object or link to the object in the database.

Text - The default type, text type allows any combination of letters and numbers up to a maximum of 255 characters per field record.

Memo - A text type that stores up to 64,000 characters.

Number - Any number can be stored.

Date/Time - A date, time, or combination of both.

A hyperlink - A hyperlink will link to an Internet or Intranet site, or another location in the database. The data consists of up to four parts each separated by the pound sign (#): DisplayText#Address#SubAddress#ScreenTip. The Address is the only required part of the string. Examples:

Internet hyperlink example: FGCU Home Page#http://www.fgcu.edu#

Database link example: #c:\My Documents\database.mdb#MyTable

Tables are grids similar to Excel worksheet where the information is stored. Because the MS Access is a relational database all the data is saved to the tables. Access provides three ways to create a table; to design it, by using wizard (automatic function) and by entering data. It is also possible to create relationship between two tables including different data and manipulating the data in to the same query. Tables represent object definitions and each row in a table corresponds to one (real world) object. Each object is identified by its unique *primary key*. Every record in a table must have a primary key that differentiates it from every other record in the table. The relationships are represented by parent-foreign key pairs (i.e. parent table and child table(s)). To establish the relationship, the primary key value of the parent object is included as a foreign key in each row of its child table, and it enables to create "one-to-many", "one-to-one" and "many-to-many" relationships between tables (Hewitt *et al.*, 2005).

MS Access Form is actually a "view" in the screen. Forms, queries and reports can be generated from tables. In forms the information (data) from tables are presented in more conveniently and by that way it is easier to read. Sub-forms, including some extra information related to original forms, can be added and viewed easily at the same time with the original form. Queries can be done from several tables at the same time and search for certain information with given criteria. Reports are usable when the data has to be presented in printed form on paper. The report can be created by selecting only the information relevant to the report at the time. Additionally, the software is capable of importing and exporting data, allowing integration with Windows and other platform applications (Töyli, 1998; Florida Gulf Coast University 2000; Hautamäki, 2005; Access 2003 Product Information; Microsoft Office Access –Wikipedia, 2006).

1.5. Generation of a plant-based natural product library

1.5.1 Selection and collection strategies for natural products

There are six commonly used criteria for selecting natural products for the collection (Vuorela *et al.*, 2004):

1. random collection of vegetation in areas supporting a diversity of plant growth, certain surroundings or several places
2. targeted collection using ethnobotanical approach (relationships between plants and people) and knowledge of the traditional use in healing
3. targeting collection specifically to plants more likely to contain bioactive molecules, or by chemotaxonomy (e.g. if a plant seems to be immune from predation it may well be producing toxic substances, or a plant contains same type of chemical entities that have been found active and collected, respectively)
4. collection of species phylogenetically (based on evolution and genetics) closely related to previously known and collected active products
5. selection of species by exploiting genomics (e.g. using structural information about genomes) and metabolomics (analysis of cellular metabolites at the same time; metabolic profiling)
6. by epidemiologic studies (e.g. the relationships between the disease occurrence and the factors causing diseases)

After the selection criteria is defined, the collection procedure of the natural products can be done by several different methods depending on the nature of the targeted product (e.g plants collected by hand, insects or invertebrates by catching equipment, etc.). The amount of plant needed for initial screening purposes is relatively small (about 0.5-2 kg). The collected plant material is identified and catalogued. The material is then dried and stored with proper identification system. The collection catalogue should include all the basic knowledge about the finding or collection place, season, point of time and the specific name of the plant species.

1.5.2 Preservation methods of plant material

The library can be generated from all kinds of plant material. Before library generation, the plant material has to be preserved so that the active compounds will remain unchanged

during storage. In order to keep plant material and the secondary metabolites (effective compounds) stable as long as possible, it is essential to preserve them in a proper way. If the preservation is not done correctly, the cell barriers are quickly broken down and the enzymes get the opportunity to promote various chemical changes in the other cell constituents, e.g. by oxidation or hydrolysis. These processes can be limited by preservation. (Samuelsson, 1992)

The most common method for preserving plant material is drying. Living plant material has a high water content (wood ~40-50% and leaves ~60-90%) and drying decreases the risk of getting e.g. moulds, and stops the enzymic processes. Drying can be done naturally (outside or in big rooms) or artificially (with drying machine). Freeze-drying (lyophilization) is a very mild and commonly used method. Drying delays enzymic reactions because the enzymes are left intact (in inactive form) and will start functioning again as soon as sufficient water is present. Fermentation of the plant material is used to release the active compounds when they are in form of glycosides. The plant material is placed in thick layers, covered and often exposed to raised temperatures (30-40°C) and humidity. Fermentation removes often bitter or unpleasant-tasting substances and promotes the formation of aromatic compounds with a pleasant smell or taste (e.g. vanilla, tea and cacao) (Samuelsson, 1992).

1.5.3 Extraction

In extraction process a small portion of the dry raw material is crushed or powdered depending on the purpose of the study. This preliminary treated material is then soaked and extracted with various solvents. Most plant-derived bioactive molecules are low-molecular-mass substances, soluble in organic solvents of varying polarity (e.g. methanol). After the solvent is added, the extracted material is treated by several different methods depending on the product. Sonication, homogenization, vortex-mixing, percolation and Soxhlet extraction and/or centrifugation are often used methods in the extraction process. Extraction can be done for example three times. Afterwards the extract is often lyophilized (freeze-dryed) in order to remove the solvent directly from a solution while in the frozen state. This relatively gentle method of removing the solvent yields a powdered product for

further treatments (Rauha *et al.*, 2000; Eldridge *et al.*, 2002; Tammela *et al.*, 2004; Vuorela *et al.*, 2004; Wennberg *et al.*, 2004).

1.5.4 Storage systems

The storage system may vary a lot depending on the natural product material which needs to be preserved. Dried crude plant material has to be stored in closed containers or lockers in dry conditions excluded from light. The extracted plant material can be stored after extraction as concentrated extracts or in lyophilized condition in a freezer. The extract can be subjected to HPLC fractionation and profiling, i.e. the material can be microfractionated to 96-well microplates before storing. In that way the extract to be studied is already in more convenient form.

1.6. Bioactivity screening of natural products

1.6.1 Principles of screening

Natural products are sources of chemically diverse compounds of great importance. But, on the other hand this diversity presents significant technical difficulties in screening processes. Typically a natural product is extracted, concentrated, fractionated and purified yielding one ore more biologically active compounds. Previously separation, detection and identification of the biologically active compounds, or biologically active chemical structures of the complex molecules found in natural products, have been very laborious and time-consuming work. The main reason to create and generate different types of compound libraries is their usage in drug discovery screening programmes. Random screening entails the biological screening of vast libraries of chemical compounds and crude extracts. The aim of the screening is to identify a possible bioactive compound or a lead structure (e.g. molecule) of a potential new drug candidate. In the past, the process took many years with no guarantee of finding a lead compound. Today, the modern screening technology enables the screening of vast libraries in large numbers of *in vitro*

tests generating lead structures in a shorter time (Eldridge *et al.*, 2002; Vuorela *et al.*, 2004; Koehn and Carter, 2005; Lee *et al.*, 2005).

1.6.2 Bioassays for bioactivity screening

The main purpose of any drug substance is the activity (effectiveness) and suitability for human use. When the chemistry is adapted to the biology in case of drug discovery, it is possible for example to measure specific macromolecule-ligand interactions by means of various physiochemical methods or integrate immunochemical and enzymatic methods with chemo-analytical systems. In addition to the chemical (physical) properties of the molecule, it is obvious that the lead compound has to also go through appropriate biological test systems to demonstrate pharmacological activity before becoming a drug candidate.

Tests used to detect the biological activity (e.g. cytotoxicity) of an extract (or pure substance isolated from extract) can be defined as bioassays. By using small scale *in vitro* biological test systems (i.e. assay systems) which can be based on culture cells, isolated enzymes or cloned receptors, it is possible to test many samples in a short period of time (Houghton, 2000). The recent developments in biotechnology have had a strong impact on the efficiency of screening methodology in drug discovery, by increasing the range of different cell types offering genetically modified cells, expressing systems not normally found in cells. One application is in the provision of sufficient quantities of cloned human receptors as drug discovery targets. Another application is the development of screening strategies for the identification of novel peptide ligands, in order to identify novel monoclonal antibodies, randomly synthesised peptide fragments, and therapeutically valuable oligonucleotides (Houghton, 2000; Rees, 2001).

Most bioassays are performed using small amounts of test material, usually in microplate format, which can be used in plate readers to measure absorbance or fluorescence of wells to measure the assay end point. The HPLC micro-fractionation is a useful tool for preparing natural products for bioactivity screening and can be applied also to fermentation

broths and synthetic mixtures in addition to plant extracts. The automated micro-fractionation to 96-well microplates can be integrated to a diversity of bioassays such as microbial- or mammalian cell-based assays, and chemical assays, such as enzyme inhibition or receptor-ligand binding assays. Test results can be replicated enabling sound statistical analysis of the results (Houghton, 2000; Rees, 2001; Tammela *et al.*, 2004; Vuorela *et al.*, 2004). Different types of systems modelling metabolic processes in the gut and liver can be used in order to discover the metabolic reactions. Tests at organismic level may be closer to the complex biological reality than single-target assays, but not amenable to HTS. A wide range of bioassay systems are used to test natural product extracts and isolate the bioactive compounds, but the most important component of screening is the use of therapeutically relevant and mechanistically specific bioassays (Houghton, 2000; Rees, 2001).

1.6.3 High-throughput screening (HTS)

High-throughput screening is an “industrialised”, fully automated process for rapid identification of novel active compounds (“hits”) in early phase drug discovery which hopefully result in lead compounds. It is a fast, efficient and reproducible system, which provides a means of evaluating the intrinsic activity of a novel chemical entity in a specified biological system. The technique can measure *in vitro* biological activities, such as receptor binding, or enzyme inhibition/activation by using live cell or transcription assays for thousands of compounds in a day. The starting point can be as large as 1 million compounds, and by running the activity tests the amount of active compounds decreases in every test of tracing the active target; from 1 million to 1000, and finally to a few actives which are tested and selected by certain assays to get one, or few lead compound(s) at the end of the process. A special feature of HTS assay is the using of microplates, such as 96-well (higher density formats are 384-well or 1536-well) plates, amenable to robotic handling. The automated workstations can have several functions, such as liquid handling, shaking, incubation, plate washing, sealing and reading the plates. Data export and –tracking, as well as on-line analysis and scheduling are also the elements of HTS automation (Rees, 2001). The lead compounds found by HTS are not drugs, but more like chemical entities, which specifically inhibit or activate the activity of the chosen drug

target with defined novel structure, and possess a structure-activity relationship (SAR). The leads are less complex than drugs, but have some “drug like”-properties, such as low molecular weight. This kind of lead is also amenable to chemical modification, e.g. is chemically tractable (Rees, 2001).

1.7 High-performance liquid chromatography (HPLC) in drug discovery

1.7.1 Reverse-phase HPLC and gradient elution

One of the main analytical tools in primary separation of natural products is high-performance liquid chromatography (HPLC). There are several characteristic features which make the HPLC suitable for the purpose. These include the excellent fractionation speeds (sometimes a few minutes per sample), superior peak resolution, high degree of automation (including data analysis). Also various sophisticated systems are commercially available. HPLC is a combination of several different chromatographic methods (i.e. reverse-phase, size-exclusion and ion-exchange –based HPLC) and the technological apparatus with pumps (to produce the high pressure to the chromatographic columns), columns, auto-sampler, detector lamps and data capture software within a same system.

Reverse-phase HPLC is among the most commonly used methods in pharmaceutical industry and many natural compounds permit the use of reverse phase HPLC in their analysis by being soluble in polar solvents. This chromatographic method is a very powerful technique because it is capable of separating very similar molecules on the basis of only minor differences of hydrophobic features in their surface. The stationary phase in the column used in reverse phase HPLC is usually a silica support with e.g. C_{18} (Octadecyl) bonded phase, which is less polar than water-organic mobile phase (e.g. water-methanol). Therefore, the sample molecules partition between the polar mobile phase and non-polar stationary phase causes the differential retention of samples depending on their hydrophobicity. When hydrophilic (polar) compounds are less strongly held, and elute first from the column, the more hydrophobic (non-polar) compounds are retained more strongly and elute last. Additionally, the more bonded the carbon (C) in the stationary phase, the

more effective (long lasting) is the retention in the column. In gradient elution the sample migrates through the column with increasing strength of strong solvent (%B), for example, methanol in the mobile phase. At the beginning of the separation when %B is low there is little or no migration. After some time, when %B increases, the sample starts to move through the column, and as time progresses and %B increases further, the migration becomes faster (Snyder, 1997).

Even if methanol is a good organic solvent for extraction, it has certain limitations when used as an organic component in the mobile phase in HPLC gradient elution. The mobile phase must transmit sufficiently at the wavelength used for detection. Due to the limits of increasing baseline noise, which reduces detection sensitivity, the mobile-phase absorbance should usually be less than 0.5 at the wavelength used for detection. For methanol the specified absorbance (AU) of reverse-phase mobile-phase at given wavelengths is as follows:

200 nm	205 nm	210 nm	215 nm	220 nm	230 nm	240 nm	250 nm	260 nm
2.06	1.00	0.53	0.37	0.24	0.11	0.05	0.02	<0.01

One of the typical mixtures with low background absorbance (<0.5 AU) in mobile-phase is methanol-water, and therefore the UV detection at 200 nm or higher can be used in reverse-phase HPLC (Snyder, 1997).

1.7.2 Analytical HPLC separation and micro-fractionation

For primary bioactivity screening the sample can first be separated and micro-fractionated by HPLC by using, for example, the reverse phase column and gradient elution. For the primary screening the injected sample volume can be rather small (e.g. 20 µL per sample). When processing natural products for primary screening by micro-fractionation, the reasonable solution is to collect fractions as a function of time, because the complex matrix, as plant extracts, causes variation to the size of the peaks and it is difficult to collect both smallest and largest peaks. It is also possible (or almost certain) that the sample has compounds that can not be detected as peaks at the wavelength used, and these

would be excluded if only peaks were collected. Micro-fractions are collected on 96-well microplates, usually with a maximum sample volume of 330 µL/well. Using the mentioned maximal volume the possibility to reach the highest obtainable concentration of the sample in the well is increased, still resulting in easily detected peaks as narrow as possible which can be identified to one well (Tammela *et al.,* 2004; Wennberg *et al.,* 2004).

Usually the bioactivity assays can be carried out after the lyophilization of the micro-fraction plates. The micro-fractionated samples are lyophilized on the 96-well microplates at least 12 hours before applied to the bioassay. With the combination of the micro-fractionation method and a bioassay, it is possible to demonstrate which components of the sample show activity. When the sample is fractionated, the inactive components of the sample do not interfere with the active constituents, and it is possible to achieve stronger signals. Fractions on a microplate offer an excellent platform for developing bioactivity assays based on the addition of a chemical reagent or cell suspension to the wells, and by using the microplates the valuable time of research work can be saved. Moreover, the lyophilized microplate can be used as a library platform which can be stored in a freezer (Eldridge *et al.*, 2002; Tammela *et al.*, 2004; Vuorela *et al.,* 2004; Wennberg *et al.*, 2004).

1.7.3 Activity-guided fractionation by HPLC

When the primary screening and separation process is completed, the usual procedure to search for compounds from complex matrices, such as extracts, is activity-guided fractionation. The number of false results can be reduced by fractionating the extract and then testing the activity component by component. The automated HPLC collector fractionates the samples as a function of time or as function of peak signals. The peaks of the chromatogram are linked to detected bioactivity (compounds) in microplate wells. Fractions of a sample showing activity are screened on microplates and are taken through further cycles of separation and testing until pure compounds can be detected from active fractions. Afterwards the active fractions undergo further isolation and purification until pure active principles are obtained. Very often the purification step in the process is rate-limiting and progression is depended upon how many “cycles” of fractionation and bioassay are required. Only if the active principles are deemed of sufficient interest, a

targeted preparative isolation is carried out. (Eldridge *et al.*, 2002; Tammela *et al.*, 2004; Vuorela *et al.*, 2004; Wennberg *et al.*, 2004; Koehn and Carter, 2005).

1.7.4 Identification of the active compound

In natural product drug discovery, the most optimal way in HTS process, is to test and identify only the active components of the extract because identification is costly and time-consuming. The active compounds are determined by using advanced spectroscopic techniques such as high resolution mass spectrometry and multinuclear magnetic resonance spectroscopy. By using hyphenated techniques, such as liquid chromatography with mass spectrometry (LC-MS), it is possible to perform the micro-fractionation simultaneously with HPLC, as well as identification of compounds in the extract. The HPLC fingerprint of the extract can be matched with its activity profile and correlated with spectroscopic data. Structural information obtained can be used for natural products database searches and tentative structural assignments (Kovács *et al.*, 2004; Tammela *et al.*, 2004 Vuorela *et al.*, 2004).

The identification of natural compounds, especially in plant extracts, can be done by a combination of different methods, such as high-performance liquid cromatographic (HPLC), nuclear magnetic resonance (NMR), thin layer cromatography (TLC), liquid chromatography (LC), and e.g. on-line and *in situ* off-line UV-Vis (Ultraviolet/visible) spectroscopy as the detection method. Integration of mass spectrometry (MS) with chromatography provides structural and mass selective detection about the molecular weight and molecular variants. The advances in NMR, MS and HPLC technology have produced hyphenated techniques, such as HPLC-MS, LC-NMR and LC-NMR-MS, which are practical and effective methods for analysis and chemical characterization of complex natural products. The NMR is mostly used for complex mixture analysis. Especially the combination of LC, NMR and MS (LC-NMR-MS) gives comprehensive structural data applicable to the identification of novel drug candidate. After detecting the active fractions, the separation can be transferred directly to preparative scale (Vuorela *et al.*, 2004; Koehn and Carter, 2005).

2. Aim of the Study

Natural products have been, and continue to be, a tremendous source of new pharmacologically effective compounds. It is therefore appropriate to have a ready-to-use natural product library system in-house for the practical reason of speeding up the process - especially if the screening is done on a regularly basis and for ongoing programmes.

The main objective of this master`s thesis was to generate a suitable and applicable natural product library system for the Drug Discovery and Development Technology Center (DDTC) in order to use it in further bioactivity screening processes in drug discovery.

The specific aims of the study were:

1) to create four-class library of plant material of reasonable quantity as an easily available resource for further use
2) to generate applicable database for the data generated from the library creation
3) to demonstrate in practice the suitability of the methods used, including laboratory work procedures, data management system generation and library data creation for the applied implementation

3. Materials and Methods

3.1 Plant material

The library material was obtained from the herb collection of the Division of Pharmaceutical Biology (Faculty of Pharmacy, University of Helsinki). The plants were collected in summer 2000 mainly from Helsinki and North Carelia region, and one sample plant from Mikkeli region. The 44 plants (Table 1) were originally chosen to this herb collection on the basis of random selection. The plants were identified, catalogued and numbered by Pia Fyhrquist, M.Sc., (Department of Biological and Environmental Sciences, University of Helsinki). The catalogue of all the identified plant samples includes the exact information of the specific place and the area of collection, as well as time point of collection. The plant material was dried and added to the herb collection in processing order, and not by a specific species or alphabetic order. Voucher specimens of the species are deposited at the Division of Pharmaceutical Biology, Faculty of Pharmacy, for future verification purposes.

Before the plants were chosen for the library, the whole plant collection was inspected in order to confirm the existence and quantity of each plant sample and voucher specimen, and also for creating comprehensive order for the samples. As a result of this inspection there were 2 plants (Wild chervil and Willow herb) for which no voucher specimen could be located and could thus not be reliably identified any more. Therefore the plants were excluded from the library samples. Additionally, the plant White clover was rejected for the same reason. On the other hand, the specific species of the plant Alsike clover was in some extent unclear, but included, because the voucher specimen was available for species verification. The family of the plant Alsike clover is *Fabaceae* and the species belongs to group of *Trifolium*, but it was not yet confirmed whether it is *T.hybridum* or *T.repens.* The species identification can be done later on. The quantity of the plant Large flowered hempnettle was so small, that it could not be processed further, and therefore also excluded from the library. Latin and English names as well of the herb collection plants were verified from two databases; PLANTS-database and The International Plant Names Index. All samples were measured in quantity approximately enough for extraction and re-packed

into new small library paper bags, coded and listed. The total amount of plant samples included to the library was 40.

Table 1. The herb collection list; plants used for the NP-library. The excluded samples are grey shaded.

Finnish name	Scientific name	English name
ahosuolaheinä	*Rumex acetocella*	Sheep sorrel
alsikeapila	*Trifolium hybridum*	Alsike clover
alsikeapila 2	*Trifolium hybridum*	Alsike clover
harmio	*Berteroa incana*	Hoary false madwort
hiirenvirna	*Vicia cracca*	Bird vetch
kanerva	*Calluna vulgaris*	Heather
kangasmaitikka	*Melampyrum pratense*	Common cow-wheat
karheapillike	*Galeopsis tetrahit*	Brittlestem hempnettle
keltakannusruoho	*Linaria vulgaris*	Butter and eggs
keltamaksaruoho	*Sedum acre*	Goldmoss stonecrop
kirjopillike	*Galeopsis speciosa*	Large flowered hempnettle
kissankello	*Campanula rotundifolia*	Harebell
koiranputki	*Anthriscus sylvestris*	Wild chervil
kurttulehtiruusu	*Rosa rugosa*	Rugosa rose
lehtovirmajuuri	*Valeriana sambusifolia*	Valerian *sambucifolia* J. C.Mikan
luhtalemmikki	*Myosotis scorpioides*	Forget-me-not
maitohorsma	*Epilobium angustifolium*	Willow herb
mesiangervo	*Filipendula ulmaria*	Meadowsweet
metsäapila	*Trifolium medium*	Zigzag clover
niittyleinikki	*Ranunculus acris*	Tall buttercup
niittynätkelmä	*Lathyrus pratensis*	Meadow pea
nukkahorsma	*Epilobium parviflorum*	Smallflower hairy willowherb
ojakärsämö	*Achillea ptarmica*	Sneezeweed
pelto-ohdake	*Cirsium arvense*	Canada thistle
pietaryrtti	*Tanacetum vulgare*	Tansy
pujo	*Artemisia vulgaris*	Common wormwood
puna-ailakki	*Silene dioica*	Red catchfly
ranta-alpi	*Lysimachia vulgaris*	Yellow loosestrife
rantanenätti	*Rorippa palustris*	Bog yellowcress
rantatädyke	*Veronica longifolia*	Longleaf speedwell
rohtoraunioyrtti	*Symphytum officinale var.Bohemicum*	Common comfrey
sananjalka	*Pteridium aquilinum*	Western brackenfern
sarjakeltano	*Hieracium umbellatum*	Narrowleaf hawkweed
särmäkuisma	*Hypericum maculatum*	Imperforate St. John's-wort
saunakukka	*Tripleurospermum inodorum*	Scentless Mayweed
seittitakiainen	*Arctium tomentosum*	Woolly burrdock
siankärsämö	*Achillea millefolium*	Yarrow
syysmaitiainen	*Leontodon autumnalis*	Fall dandelion
tahmavillakko	*Senecio viscosus*	Sticky ragwort
tarharaunioyrtti 1	*Symphytum asperum*	Prickly comfrey
tarharaunioyrtti 2	*Symphytum asperum*	Prickly comfrey
vahamaksaruoho	*Sedum telephium*	Witch's moneybags
valkoapila	*Trifolium repens ?*	White clover
valkopeippi	*Lamium album*	White deadnettle

The major part of the specimens in the collection, were originally divided into two, four or six samples, depending on the part of the plant. The most common parting system was based on four parts of the plant; 1.flower (F), 2.stem (S), 3.leave (L), and 4.root (R) One plant had extra parts, such as flower stem (FS) and leave stem (LS). One plant was collected in one piece, including the upper part from the ground without roots.

3.2 Chemicals and water

The extraction (for library Class II) was done with 100% methanol (HPLC Grade, Rathburn, Walkerburn, Scotland). DMSO (Dimethyl sulfoxide, Merck KGaA, Germany) was used as a solvent for preparing the stock solutions (for Class III). The HPLC analysis was performed by using water, which was purified via a Millipore MilliQ filtration machine, and methanol (HPLC grade, Rathburn, Walkerburn, Scotland) was used as the organic solvent in the mobile phase.

3.3 Instrumentation

Primary grinding of the plant samples was done first in mortars and then by electric Moulinex-mixer (rough materials). The secondary grinding (smooth material), powder, was done by using electric IKA Labortechnik A 10 –mill (IKA LABORTECHNIK STAUFEN, Germany). Sonication was carried out with the frequency of 50/60 Hz and centrifugation at 20ºC with 1500 rpm. Samples were lyophilized (condenser -55.5ºC, pressure < 1 hPa) by using HETO LyoPro 3000 freeze-dryer (Heto-Holten A/S, Denmark).

Analytical HPLC (high-performance liquid chromatography) separations were performed on Discovery HSC18 column (25 cm x 4.6 mm i.d., 5µm, Supelco, Bellefonte, USA), supplemented with a guard column (SecurityGuard^{TM}, HPLC Guard Cartridge System, Phenomenex, USA). The HPLC system consisted of a PerkinElmer series 200 LC pump and Autosampler (Norwalk, CT., USA) with 200 µL loop, PerkinElmer LC 235 C Diode Array Detector (Norwalk, CT., USA) and PE Nelson 600 series link (Norwalk, CT., USA). Automated fraction collector, Gilson FC 204 (Middleton, WI., USA) connected to the HPLC system, was used for the micro-fractionation (Fig. 6).

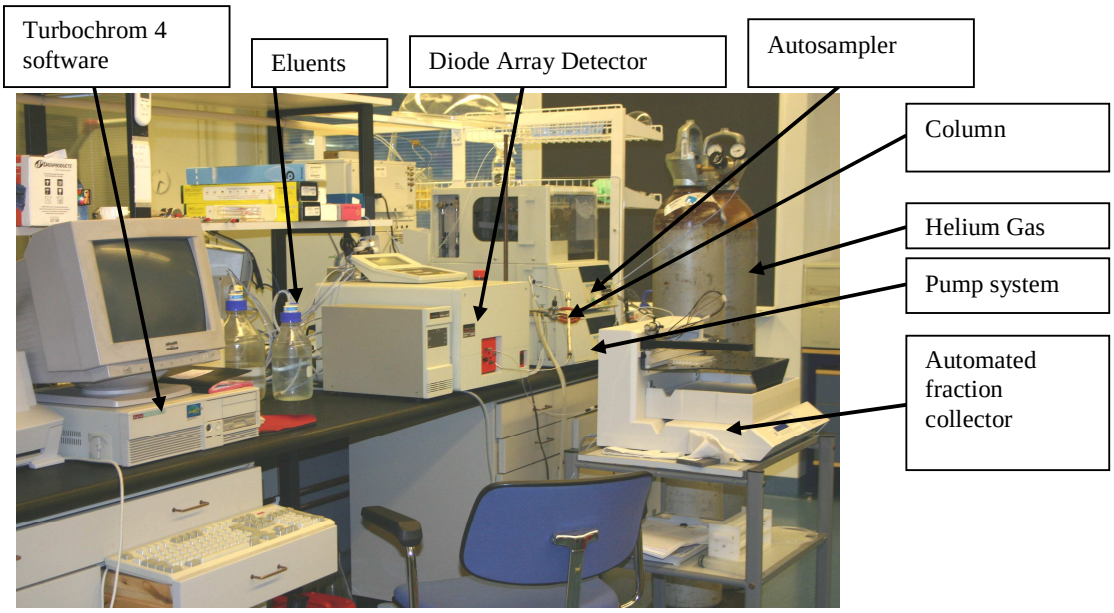

Figure 6. The HPLC (high-performance liquid chromatography) apparatus at the Division of Pharmaceutical Biology, Faculty of Pharmacy.

3.4 Computer programs

The software used in HPLC was PerkinElmer Nelson Turbochrom 4 software, version 4.1. (Norwalk, CT., USA) and Microsoft Access XP 2002 (Microsoft, USA) was applied to generate the database for the library.

3.5 Grinding and powdering the plant material (Class I: BH -samples)

Air-dried plant samples were originally divided into 4 or 6 sub-samples depending on the plant part. Approximately a handful of each plant sample was ground and powdered, in order to get homogenous start-up material for the library. Primary grinding of the rough plant samples, such as roots and some stems, were first crushed by hand in mortars and then ground by using simple electric Moulinex-mixer. The secondary grinding was done by using electric IKA-mill. The smoother samples were ground only with the IKA-mill. The powdered material was collected, packed in paper bags, coded and stored at dry and dark conditions. These samples formed the Class I library material. Total number of samples was 145.

3.6 Extraction of the plant material

Air-dried and powdered plant material (1g per sample) was weighed into a test tube, 5 ml methanol (100%) was added and suspension was stirred. The test tube was sonicated for 10 minutes and centrifugated for 15 minutes at 20ºC (1500 rpm). This extraction was performed three times on each sample, and the supernatant was collected after every extraction. The combined supernatants were collected to weighed vial (20mL) and vial cap was closed. Samples were frozen (-20ºC) before lyophilization.

3.7 Lyophilization and weighing the yield (Class II: BE -samples)

The extracts were freeze-dried (lyophilized) for 2 x 24 hours, and the dryness was checked after the first 24 h by scraping the sample loose from the bottom of the vial. In case of “bubbling” in the lyophilizator, the extracts were divided into two weighed vials. After lyophilization the samples were weighed and removed from vials to 1.8 mL cryotubes (Nunc CryoTubeVials, Nunc A/S, Denmark). Samples were frozen (-70ºC) in format of two 81 Cryotube –racks (Nalgene CryoBox). A map of the racks was designed for the library (Appendix 1). These samples formed the Class II library material and the total number of samples was 145.

3.8 Preparation of stock solutions in DMSO (Class III: BED -samples)

For preparing stock solutions for storage and for HPLC analysis and microfractionation, a portion of the dry extract was dissolved in DMSO (Dimethyl sulfoxide) generating the Class III material. Samples were categorized on the basis of the sample amount in lyophilized form. If the sample amount was under 60 mg, the sub-sample size for the DMSO solution was about 10 mg, and if above 60 mg, the sub-sample size was about 20mg. The sub-samples were weighed into 1.4 mL sample tubes (ScreenMates Alphanumeric Tubes, MATRIX, USA). The sub-samples were dissolved into DMSO in a concentration of 40 mg/mL. After solvent addition the colour of the samples was recorded and classified by using the DDTC library sample colour classification (Appendix 2). The total number of samples in Class III was 145. Samples were stored in the freezer (-20ºC) in

format of two 96- ScreenMates Tubes-racks (MATRIX, USA). A map of the racks was designed for the library (Appendix 2).

3.9 Analytical HPLC separation and fractionation

For the HPLC analysis 50µL of each Class III sample was pipetted into limited volume inserts of 0.3 mL (BROWN Chromatography Supplies Inc., Germany) inside the HPLC sample vial. The analytical HPLC separation was performed by using reverse phase method and linear gradient elution. Before the analytical column, the samples were filtered through a quard column which was changed every time in case of pump pressure exceeding the regular function level of 3500PSI. The mobile phase consisted of methanol (B) and water (A).

The injection volume for all samples was 20µL and the flow rate was 1mL/ min^{-1}. The gradient profile 5-95 (B%) for 48 minutes was applied to all samples. Total run time of the method used in HPLC analysis was 75 minutes, including experiment time of 60 minutes and steady flow time of 15 min. The experiment time consisted of a linear gradient for 48 minutes, steady 95 (B%) level for 10 minutes and linear gradient 95-5 (B%) for 2 minutes. The fraction collection time was 48 minutes, starting from 17 minutes and lasting for 65 minutes. The 2 minutes lag between sample injection and collection start, and the 2 minutes delay after linear gradient were due to the flow time when the sample has eluted through the column. The method (Fig.7) was saved to the Turbochrom 4 software database to the path D:\PIRKKO\METHOD\METDDTC.MTH. The detection was performed by using two different wavelengths: λ_a = 230 nm and λ_b = 280 nm. Separated samples were micro-fractionated into two coded (A+B Plates) 96-well microplates (Nunc, Roskilde, Denmark) with volume of approx. 300 µL/well by using automated fraction collector. Direction of the fractionation was a serpentine pattern. Fractionation was performed as a function of time, with the collection time of 18 s. per well for 48 minutes. All the columns of plate A were filled, but the last 4 columns of plate B were left empty for the possible reference use. The total number of plates produced was 290 and the final number of micro-fractions 23 200.

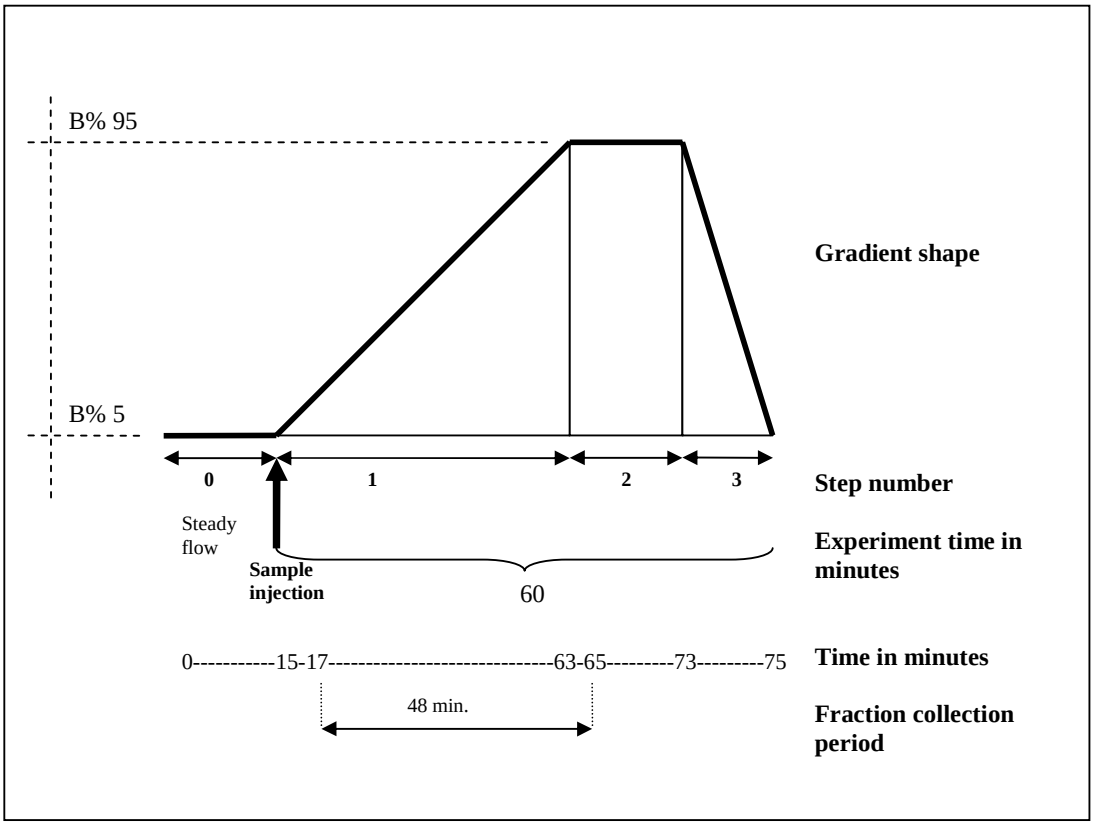

Figure 7. HPLC profile of the method METDDTC.

The time point of the sample migrated through the column to microplate as a sub-fraction, was calculated for each sample and a map of plates A and B was designed for the library (Fig. 8). This map is directly proportional to the chromatograms, so that activity peaks and time points can be identified and located to a certain fraction in a certain well (Fig. 9). The default reports and chromatograms of both wavelengths of all samples were saved to the program files, separate discs and printed for the library binders.

DDTC BEF FRACTIONS ORIGINAL 96-WELL PLATE A

A1 0.0-0.3	A2 4.5-4.8	A3 4.8-5.1	A4 9.3-9.6	A5 9.6-9.9	A6 14.1-14.4	A7 14.4-14.7	A8 18.9-19.2	A9 19.2-19.5	A10 23.7-24.0	A11 24.0-24.3	B12 28.5-28.8
B1 0.3-0.6	B2 4.2-4.5	B3 5.1-5.4	B4 9.0-9.3	B5 9.9-10.2	B6 13.8-14.1	B7 14.7-15.0	B8 18.6-18.9	B9 19.5-19.8	B10 23.4-23.7	B11 24.3-24.6	C12 28.2-28.5
C1 0.6-0.9	C2 3.9-4.2	C3 5.4-5.7	C4 8.7-9.0	C5 10.2-10.5	C6 13.5-13.8	C7 15.0-15.3	C8 18.3-18.6	C9 19.8-20.1	C10 23.1-23.4	C11 24.6-24.9	D12 27.9-28.2
D1 0.9-1.2	D2 3.6-3.9	D3 5.7-6.0	D4 8.4-8.7	D5 10.5-10.8	D6 13.2-13.5	D7 15.3-15.6	D8 18.0-18.3	D9 20.1-20.4	D10 22.8-23.1	D11 24.9-25.2	E12 27.6-27.9
E1 1.2-1.5	E2 3.3-3.6	E3 6.0-6.3	E4 8.1-8.4	E5 10.8-11.1	E6 12.9-13.2	E7 15.6-15.9	E8 17.7-18.0	E9 20.4-20.7	E10 22.5-22.8	E11 25.2-25.5	F12 27.3-27.6
F1 1.5-1.8	F2 3.0-3.3	F3 6.3-6.6	F4 7.8-8.1	F5 11.1-11.4	F6 12.6-12.9	F7 15.9-16.2	F8 17.4-17.7	F9 20.7-21.0	F10 22.2-22.5	F11 25.5-25.8	G12 27.0-27.3
G1 1.8-2.1	G2 2.7-3.0	G3 6.6-6.9	G4 7.5-7.8	G5 11.4-11.7	G6 12.3-12.6	G7 16.2-16.5	G8 17.1-17.4	G9 21.0-21.3	G10 21.9-22.2	G11 25.8-26.1	H12 26.7-27.0
H1 2.1-2.4	H2 2.4-2.7	H3 6.9-7.2	H4 7.2-7.5	H5 11.7-12.0	H6 12.0-12.3	H7 16.5-16.8	H8 16.8-17.1	H9 21.3-21.6	H10 21.6-21.9	H11 26.1-26.4	H12 26.4-26.7

DDTC BEF FRACTIONS ORIGINAL 96-WELL PLATE B

A1 28.8-29.1	A2 33.3-33.6	A3 33.6-33.9	A4 38.1-38.4	A5 38.4-38.7	A6 42.9-43.2	A7 43.2-43.5	A8 47.7-48.0	**A9**	**A10**	**A11**	**A12**
B1 29.1-29.4	B2 33.0-33.3	B3 33.9-34.2	B4 37.8-38.1	B5 38.7-39.0	B6 42.6-42.9	B7 43.5-43.8	B8 47.4-47.7	**B9**	**B10**	**B11**	**B12**
C1 29.4-29.7	C2 32.7-33.0	C3 34.2-34.5	C4 37.5-37.8	C5 39.0-39.3	C6 42.3-42.6	C7 43.8-44.1	C8 47.1-47.4	**C9**	**C10**	**C11**	**C12**
D1 29.7-30.0	D2 32.4-32.7	D3 34.5-34.8	D4 37.2-37.5	D5 39.3-39.6	D6 42.0-42.3	D7 44.1-44.4	D8 46.8-47.1	**D9**	**D10**	**D11**	**D12**
E1 30.0-30.3	E2 32.1-32.4	E3 34.8-35.1	E4 36.9-37.2	E5 39.6-39.9	E6 41.7-42.0	E7 44.4-44.7	E8 46.5-46.8	**E9**	**E10**	**E11**	**E12**
F1 30.3-30.6	F2 31.8-32.1	F3 35.1-35.4	F4 36.6-36.9	F5 39.9-40.2	F6 41.4-41.7	F7 44.7-45.0	F8 46.2-46.5	**F9**	**F10**	**F11**	**F12**
G1 30.6-30.9	G2 31.5-31.8	G3 35.4-35.7	G4 36.3-36.6	G5 40.2-40.5	G6 41.1-41.4	G7 45.0-45.3	G8 45.9-46.2	**G9**	**G10**	**G11**	**G12**
H1 30.9-31.2	H2 31.2-31.5	H3 35.7-36.0	H4 36.0-36.3	H5 40.5-40.8	H6 40.8-41.1	H7 45.3-45.6	H8 45.6-45.9	**H9**	**H10**	**H11**	**H12**

Figure 8. The microfractionation map of microplates A and B.

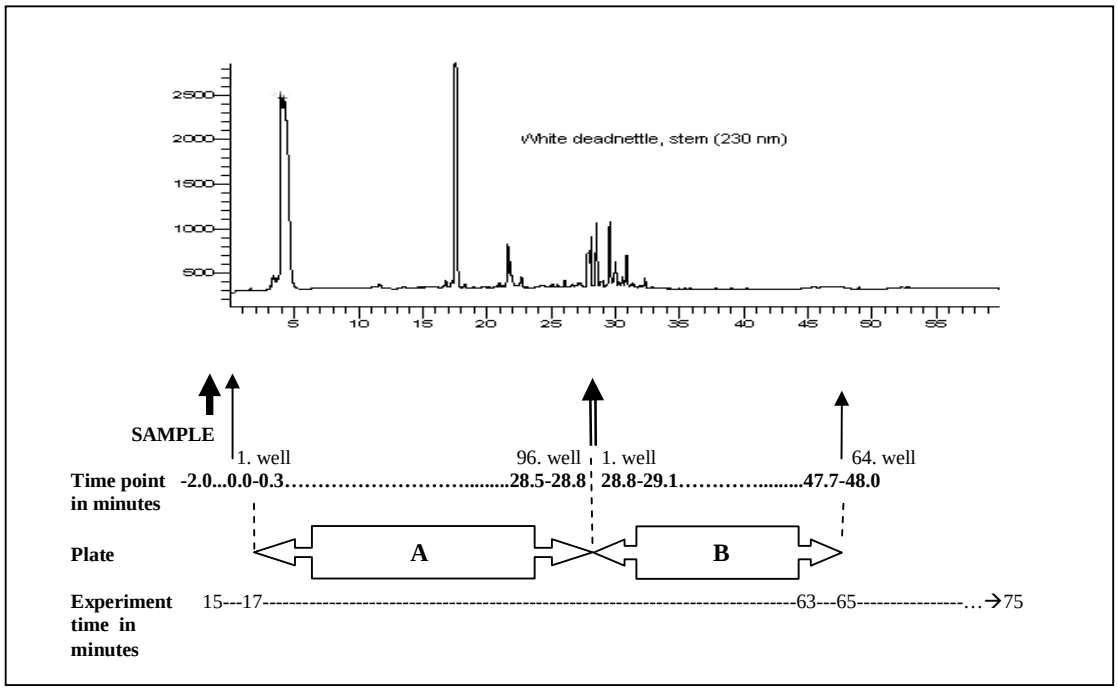

Figure 9. The correlation between microfrationation map and the HPLC chromatogram.

3.10 Lyophilization of microfraction plates (Class IV: BEF -samples)

After fractionation, the plates were covered with plastic lids, frozen and lyophilized for 24 hours. After lyophilization the plastic lids were removed and the plates were covered tightly with non-sterile sealing tapes (Nunc Sealing Tape, Nalge Nunc International, USA) and stored in freezer (-20ºC). These plates formed the Class IV of the DDTC NP Library.

3.11 The library material and colour classification

The creation of the four main classes for the library material was based upon the stages of the process: the start up material (first stage), the extracts (second stage), the extracts in DMSO (third stage), and the final fractionated and lyophilized samples (fourth stage). Every step of the process created its own type of samples which were classified in numerical order according to the stage of the process. Additionally, the colours of all the samples of Class III were recorded after the solvent (DMSO) addition. After all samples were dissolved the colour classification was designed on the basis of the sample colours, and this formed the DDTC NP library sample colour classification (Appendix 3).

3.12 The design of the library data management

The design of the library data base was based on the MS Access XP (2002) software features. Being a relational database, and therefore having relation building capabilities between the separately created tables, it formed the structure of the library. The structure was therefore created to include all the library classes in separate tables which have a relation (e.g. connection) to protocol tables which in turn are related to the main classes. Relationships were created between tables as follows;

DDTC NP LIBRARY Class I ⟷ **Drying protocol for plant material**
DDTC NP LIBRARY Class II ⟷ **Extraction method protocol**
DDTC NP LIBRARY Class III ⟷ **Lyophilization procedure protocol**
DDTC NP LIBRARY Class IV ⟷ **HPLC procedure protocol**

All classes (I-IV) were created into their own tables, with duplicates of each class. The first, the original class groups (Class I-IV) include the various imported objects (OLE-objects) from other Windows programs, e.g. library maps (MS Excel), color classification of the samples (MS Word) and HPLC chromatograms (MS Paint). The duplicates (Class I-IV without OLE and MEMO objects) do not have these imported objects and therefore are usable for various forms of functions, such as making the queries, inside the program. The data input was performed simultaneously with the library material preparation and formation. (Fig. 10)

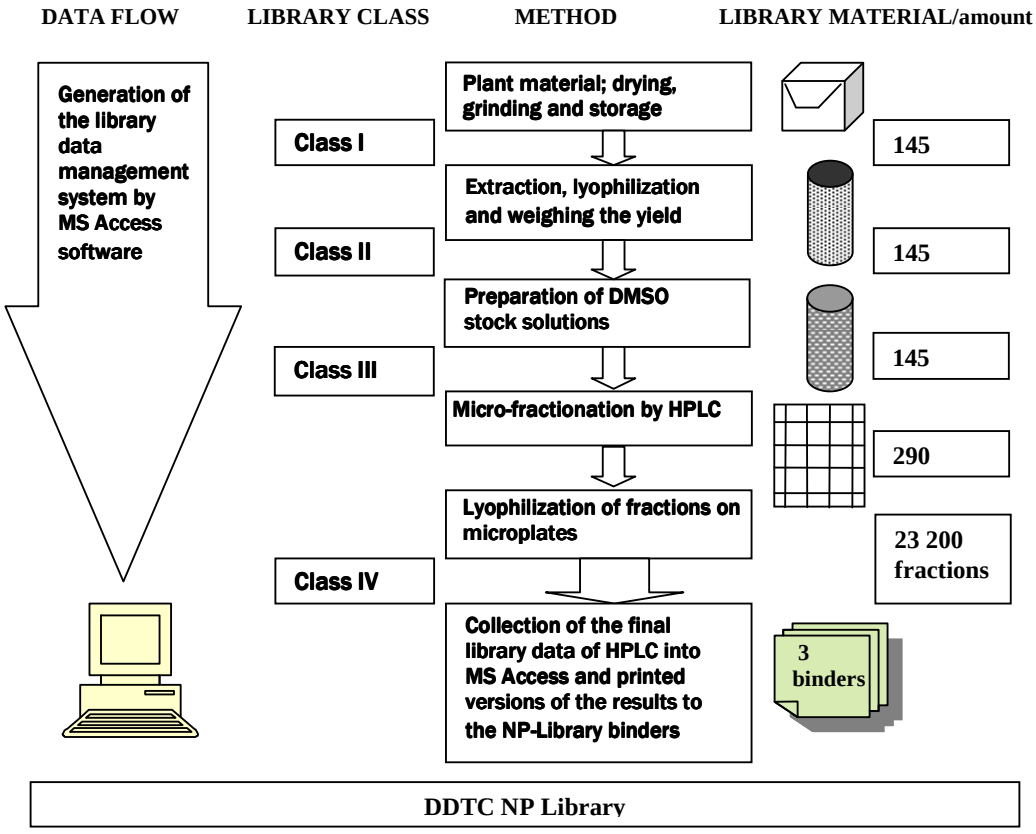

Figure 10. Schematic representation of the natural product library generation.

3.13 Library coding design

The coding system was designed on the basis of the owner's name (DDTC), generation year (mainly during the year 2005), the producer group (Bioactivity Screening Group, BAS) and the nature of the sample. The year was shortened to be -05 and letter B was chosen to represent the Bioactivity Screening Group (BAS). The second and third letters represent the nature, category and the method of treatment of the Classes I-IV, such as herb (H), extract (E), sample in DMSO (D) or sample fraction (F). The library categories were divided to 4 separate classes all representing different kind of start-up material:

Class I. **Crude natural product material**: B01-B09, B10->Bn
DDTC-05-**BH**-01a -> **BH**-44d

Class II. **Lyophilized crude natural product extracts**: B01-B09, B10->Bn
DDTC-05-**BE**-01a-> **BE**-44d

Class III. **DMSO solutions of crude natural product extracts**: B01-B09, B10->Bn
DDTC-05-**BED**-01a -> **BED**-44d

Class IV. **Fractions of natural product extracts**: B01-B09, B10->Bn
DDTC-05-**BEF**-01aA->**BEF**-44dB

Additionally, two classes were designed for the coding system in case of later use and data input to the classes:

Class V. **Natural compounds**: **N**01-**N**09, **N**10->Nn : DDTC-05-NC-01 -> n

Class VI. **Synthetic compounds**: **S**01-**S**09, **S**10->Sn : DDTC-05-SC-01 -> n

In the coding system the abbreviation F, L, S, R, FS and LS indicate the following plant parts; flower, leaf, stem, root, flower stem and leaf stem, respectively. One plant, collected as a whole plant, was coded as F/L/S. Additionally the individual plants were numbered from 1 to 44, according to the original herb collection (see chapter 3.1. Note: plants are not numbered in alphabetical order). In the database (MS Access) the coding was designed to have two separate codes. The first code indicates the program primary key-code and the second the library classification. The primary key-code of the library (obligatory MS Access feature) named as Library number, was selected to be letter B and the number of

the sample was added in numerical order. In MS Access, number 0 had to be added to the first 9 numbers, and the numbering was designed to be 01 until 09, and from number 10 onwards normal two-figure numbers were used. Some samples were given additional sub-numbers -1 and -2 (e.g. B03b-1 and B03b-2) due to division of samples at the lyophilization stage. The MS Access processes only descending or ascending orders for numbers and alphabets. A corresponding alphabetical order had to be created in order to input each sample of the same plant into the same table. A small letter: a, b, c, d, e or f was included to the library number, as well as to the library code and to indicate the plant part similarly to abbreviations F, L, S, R, FS and LS, respectively (Table 2) .

Table 2. Example of the key code (library number) and library code in DDTC NP Library.

Library number	**Library Code**
B01a	DDTC-05-BH-01-a
B01b	DDTC-05-BH-01-b
B01c	DDTC-05-BH-01-c
B01d	DDTC-05-BH-01-d
B02b	DDTC-05-BH-02-b
B02c	DDTC-05-BH-02-c
B03a	DDTC-05-BH-03-a
B03b-1	DDTC-05-BH-03-b-1
B03b-2	DDTC-05-BH-03-b-2
B03c	DDTC-05-BH-03-c
B03d	DDTC-05-BH-03-d

4. Results

4.1 Library volumes and quantities

The natural product library was generated of 40 plants of Finnish origin, by using high performance liquid chromatography (HPLC) for sample processing and by applying MS Access software platform for creating a library database. In Class I, the sample volume in paper bags was not measured. Some of the bags have several grams of the ground plant material, and some only few milligrams, but the original plant material is available in bigger quantities in the herb collection, and therefore easily available if needed. The average yield of the samples in Class II varied from 22.10 mg to 233.00 mg and the average final individual sample quantity was about 133 mg. Two of these samples (B10d and B16c) showed not to have the calculated volume left after generating the Class III samples when estimated visually, and therefore these samples are no more available in extracted form. The library volume of the Class III samples varied between 200µL and 470 µL, and the average volume of the final Class III samples was 440 µL per sample. In Class IV micro-fractionated samples the volume was approx. 300µL per well before lyophilization. The final number of library samples was 725 samples in Classes I-III and 23 300 fractions in Class IV, giving a total number of 23 735 members to the DDTC NP Library (Table 3).

Table 3. Size of the DDTC NP Library.

Library category	**Sample quantity**	**Library size (MS Access)**
Class I	145	-
Class II	145	-
Class III	145	-
Class IV	23 300 (fractions)	-
Total	**23 735**	**176 MB**

4.2 Library data documentation

The generation of the library produced a large amount of data. In order to have the most beneficial and practical system for further use, the library data was stored mainly to the DDTC NP Library database (MS Access).

Additionally, the HPLC default reports and chromatograms at two wavelengths (λ_a = 230 nm and λ_b = 280 nm) were documented for each sample (Fig. 11 and Fig.12) and stored in Turbochrom 4 software database (path:D:\Pirkko\DDTC\Results), on separate discs (8 discs) and in printed paper versions. These were also copied to the DDTC NP Library database, as well as several other documents from Word-, Excel- and Paint programmes. The total number of chromatograms was 290. The discs and paper versions of these chromatograms were organized in three binders named DDTC NP Library binders I-III.

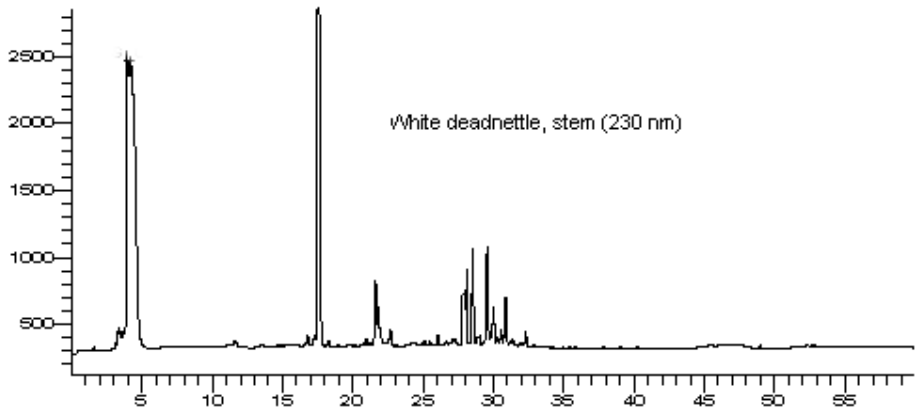

Figure 11. HPLC chromatogram of the *Lamium album* (White deadnettle, stem) extract at the wavelength of 230 nm.

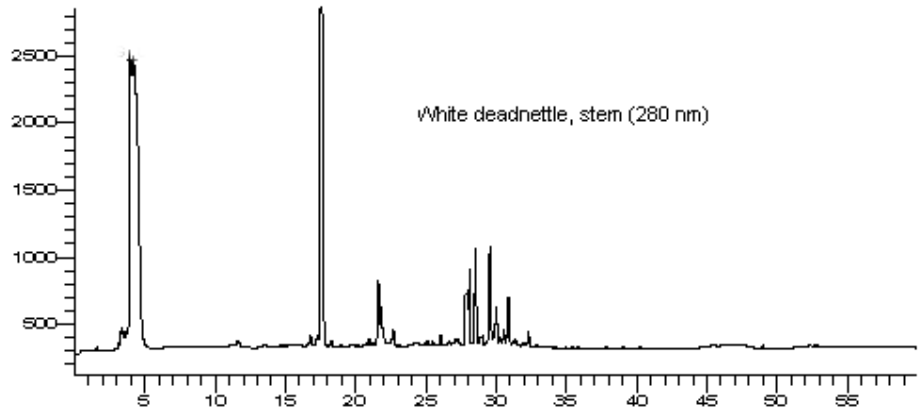

Figure 12. HPLC chromatogram of the *Lamium album* (White deadnettle, stem) at the wavelength of 280 nm.

4.3 Data in the DDTC NP Library database (MS Access)

The MS Access database allows creation of several types of forms, queries and reports from the created tables. The DDTC NP Library database consists of tables, forms and reports (see chapter 1.4.2). The main category is tables. At the time of generating the library the number of tables was 16 (Table 4). The forms of all four library classes and methods or procedures were created from tables and sub-forms were included for the extra information (Table 5). Queries were not created, because there was no practical or actual need for queries at the moment, and those can be done separately from the existing tables and forms when needed. There were two reports; one is an example, the other represents the relationships between the tables (Table 6). The total size of the DDTC NP Library database was 176 MB (see Table 3). The table 1 (Instruction for the library use) is included as Appendix 4.

Table 4. Tables in the DDTC NP Library database.

1.	A: Start page; open and double click word document (incl. Instructions for the library use)
2.	Class I without OLE and MEMO data
3.	Class II without OLE and MEMO data
4.	Class III without OLE and MEMO data
5.	Class IV without OLE and MEMO data
6.	DDTC NP LIBRARY Class I
7.	DDTC NP LIBRARY Class II
8.	DDTC NP LIBRARY Class III
9.	DDTC NP LIBRARY Class IV
10.	General DDTC NP LIBRARY all classes I-IV
11.	General subform
12.	Drying protocol for plant material
13.	Extraction method protocol
14.	Lyophilization procedure protocol
15.	HPLC Procedure protocol
16.	Protocol number

The data in the tables varies according to the class type; To give an example, the table number 2, Class I without OLE and MEMO objects, includes the following information of each sample: Library number, Library Code, Species, Family, English name, Finnish name, Identification, Classification, Date of Collection, Collection place, Sample storage (e.g. conditions) and Storage location. Similarly the information relevant to certain class type was stored into other tables (tables 3-9). Because MS Access specific features cannot to create queries from OLE and MEMO objects, there are two different tables of every class (see chapter 3.12: The design of the library data management). Tables 6-9 include additional data (chromatograms, colour classification and excel-tables, e.g. maps of classes II-IV). Tables 2-5 do not have this data. The protocols include detailed procedure information for producing the class specific samples.

Table 5. Forms in the DDTC NP Library database.

1. DDTC NP LIBRARY Class I
2. DDTC NP LIBRARY Class II
3. DDTC NP LIBRARY Class III
4. DDTC NP LIBRARY Class IV
5. Extraction method protocol
6. General Subform
7. Drying protocol for plant material
8. General DDTC NP LIBRARY all classes I-VI
9. Lyophilization procedure protocol
10. HPLC procedure protocol

Table 6. Reports in the DDTC NP Library database.

1. DDTC NP LIBRARY Class II Report
2. Relationships for DDTC NP LIBRARY

4.4 Library material storage system

Class I samples were packaged into paper bags and stored in the DDTC herb collection room, in dry and dark conditions. Library samples of Class II-IV were stored at DDTC freezer at -20ºC or -70 ºC and maps of these storage systems were created for the sample identification. The sample storage system (Table 7) was chosen according to the nature of the sample and maps were designed for the library classes II-IV (see Fig. 7 and Appendices 1 and 2).

Table 7. The storage system of the DDTC NP Library, based on the nature of the material.

Class I:	In paper bags / DDTC Herb Collection Room
Class II:	In Cryotubes (Cryotube Map 1+2) / DDTC Freezer -70ºC
Class III:	In ScreeMatesTubes (ScreenMatesTube Map 1+2) / DDTC Freezer -20ºC
Class IV:	In 96-well micro-plates (96WellPlate Map A+B) / DDTC Freezer -20ºC

5. Discussion

5.1 Natural product library in four classes

Natural products are still a huge and mainly unrevealed source of chemical and functional diversity. They provide a wide range of bioactive compounds which can be used as a source for novel drug leads, and are therefore an ideal starting point for screening pharmacologically active small molecules (Abel *et al.*, 2002; Bleicher *et al.*, 2003; Newman *et al.*, 2003; Vuorela *et al.*, 2004; Jaspars *et al.*, 2005; Koehn and Carter, 2005; Verpoorte, 2005). New drugs are needed continuously and the discovery programs increasingly need more effective ways of screening the active compounds from natural products, with lower costs and shortened timelines; in the past the research into natural products declined, mainly because of the lack of compatibility of natural product extracts with high-throughput screening programmes (Abel *et al.*, 2002, Bleicher *et al.*, 2003, Koehn and Carter, 2005). Today, the value of natural products has again been realized and they are more widely adapted to HTS programmes through a general shift from screening the extracts to pre-fractionated extracts, or even pure natural product libraries. Nowadays, in addition to the pharmaceutical industry, many commercial laboratories as well as research institutes have generated various kinds of natural product libraries in-house. There is specifically a need for economic generation of focused collections of chemically and functionally diverse compound libraries, which can be easily used in different screening procedures (Abel *et al.*, 2002, FitzGerald, 2005; Koehn and Carter, 2005).

As natural products are very complex compound combinations, the sample has to be processed to have finally a single biologically active compound. The sample can be manipulated with various methods in order to first get an extract and then the fractionated sample for the screening process. As the process is very time-consuming it is obviously practical to have pre-processed, e.g. fractionated samples already available in form of a larger collection. To make the process more economical, it is necessary to shorten the process time by producing larger quantities of samples that represent larger variations of species. At the same time a natural product library available for further experiments is generated (Abel *et al.*, 2002; Eldridge *et al.*, 2002; Bleicher *et al.*, 2003; Vuorela *et al.*,

2004; Koehn and Carter, 2005). Undoubtedly, it is equally meaningful to generate a high-quality library with detailed information of the natural product source. In case of activity found from the library fractions it is important that the original source can be traced easily to enable further studies with the specific sample. For example, knowing the exact place of collection in addition to the name and species, the plant material can be recollected in case of limited amounts of material from the first collection. That was the reason why the database and coding system for this NP library were created.

Despite some species being very rare with limited supply, the modern methods can be used to construct new mimic versions and synthesize effective new drugs derived from natural products. Further more, by using modern methods of natural product chemistry and synthetisation, chemogenomics, proteomics and the possibilities of biotechnology, it is possible to engineer the secondary metabolic pathways as well as to sequence, clone and express the metabolites in suitable host systems; At the same time it is possible to generate large and diverse libraries as well as focused small libraries, and again increase the possibilities to discover new drugs of natural origin. (Mijts and Schmidt-Dannert, 2003; Boldi, 2004; Shen, 2004; Evans *et al.*, 2005; Jaspars *et al.*, 2005; Koehn and Carter, 2005; Lautru, *et al.*, 2005; McAlpine *et al.*, 2005; Shang and Tan, 2005; Dunlap *et al.*, 2006; Mang *et al.*, 2006; Van Lanen and Shen, 2006).

The creation of a material library in form of four classes, where all the types of sample material are available and ready-to-use, is highly beneficial and practical like all libraries in case of further use. In addition to being directly adaptable to screening, it will also save time and resources in the following steps of drug discovery because the original sample type can be processed from certain preliminary stages onward and not from the beginning; This is especially useful in case of plant extracts, for which the processing time is long. All the samples in every class were created using the class-specific protocol. This is a critical feature to ensure uniform quality of the samples for the preparative HPLC analysis.

5.2 Suitability of the methods and procedures

The chosen laboratory procedures were found to be the most appropriate and feasible for this type of work. The samples can be processed in groups including several samples at a time, which saves time in each procedure step. The protocols are easy to copy and apply to all kinds of plant material.

A critical point in the generation process was also the observation how much the quantity of lyophilized extracts differed from each other despite the fact that the original extracted sample material was the same (1g). As a result, two of the Class II samples ran out totally after preparing the Class III samples. This suggests that the extracts should be prepared in larger quantities, for example of 2-3g of original ground and powdered material, or even more, depending on how much of the lyophilized extract is needed to store.

Usually good analytical results will be obtained with careful selection of the wavelength used for detection, but this requires knowledge of the UV spectra (and molar absorptivity values) of individual sample components (Snyder, 1997). In case of crude plant extracts where one sample may contain more than 10s of different components, the selection of the wavelength has to be done by some average level suitable for many compound types, having regard for possible limitations due to the chosen mobile phase. In order to minimize the background absorbance in the mobile phase with methanol-water mixture, the chosen wavelength has to be higher than 200 nm, and there are several compound types detected between the wavelengths of 210-300 (Snyder, 1997; see chapter 1.7.1). In the light of the examples of some previous studies with plant material where the chosen wavelength for UV detection has been 255 nm (Tammela *et al.*, 2004) and 280 nm (Kovács *et al.*, 2004, Wennberg *et al.*, 2004), and in order to cover a wider range of detection of different compounds, the detection was performed by using two different wavelengths; λ_a = 230 nm and λ_b = 280 nm.

Again, the method for the analytical HPLC should be developed and optimized according to the samples to be analysed (Snyder, 1997). In this case, where over a hundred (145)

samples had to be processed efficiently, and it was not possible to optimize nor to develop sample specific methods, it is clear that not all the possible compounds of these samples are detectable in the library documentation. Being aware of the requirements for high-quality analysis, the samples were treated as well as the circumstances allowed: the guard column was used for filtration, and the same analytical HPLC run was performed to all samples. The chromatograms showed roughly uniform quality, and some similarity between the samples of same plant part from different plants could be detected. The results of the chromatograms were also in most cases successfully clear in form of peak detection at both wavelengths. This indicates that the selected wavelengths cover the molar absorptivity of several clearly detectable compound types, and can be used when the mobile phase consists of methanol. The reverse phase HPLC is the method of choice for the separation of various pharmacologically effective compounds from unknown complex matrix of plant extracts. The column proved to produce clear retentions which could be detected as selectable peaks in chromatograms.

Despite the aid of new technology, the preparation and manipulation of natural product samples is still time-consuming work. Based on the fact that HPLC analysis and microfractionation of the extracts took 75 minutes per one sample plus change time, the average time to produce 5 to 7 microfractionated samples was one working day. The generation of the library took several months. From this point of view the production of this kind of a library is quite demanding work and needs a lot of resources and materials to work with.

Another feature which was considered as a delay at the lyophilization stage was the nature of methanol-water mixtures in extract form in vials and also in fractions on microplates. The extracts and the fractions had a tendency to bubble, e.g. rise from the vials or from bottom of wells when exposed to the vacuum of the lyophilizator, especially if the fraction was further pipetted to the plates with cone-shaped wells. For this reason some of the extracts needed to be divided into two vials due to the strong “bubbling” effect. After several tests with microplates, the decision was made to freeze-dry the samples onto the

microfractionation plates which have flat bottom wells. Still, the plates needed to be covered with lids to prevent the “rising” of the fractions up from the wells during the first minutes in the lyophilizator.

One important aspect is also the final concentration of the microfractionated samples in the wells. Compounds in the extracts are present at a very low concentration; it is quite common that the target compounds are present in less than 1% by weight of the crude extract (Koehn and Carter, 2005). Despite the relatively high concentration (40mg/mL) at the beginning, the resulting concentration of individual compounds in the wells, and whether or not it is enough for the detection of the biological activity, can be argued. Additionally, the microfractionated plates are mostly for one-time use, e.g. non-reusable, due to the low concentration and format (wells to be filled with another solvent for screening the possible activity).

Lastly, there are not so many laboratory procedures which can be said to have an esthetic aspect, but in this work the beautiful and bright colours of the methanol extracts (Class I) as well as DMSO solutions (Class III); different variations of red, green and yellow, were visually effective. I was pleased to see these colours as one result of the work, and therefore, from my point of view, also worth of mentioning.

5.3 MS Access as a library database

At the beginning of library design, there were several options of database softwares available with a variety of functions and different features. The Microsoft Access software was chosen for creating this library database for practical and economical reasons. Firstly, this software is highly appropriate for building this kind of data collection system, and is generally known to be suitable for drug discovery programmes (Access 2003 Product Information; Hewitt *et al.*, 2005). Secondly, the MS Access had been used previously in other departments at the University of Helsinki and has been proven to be quite practical in use (according to provided example on material how data was saved for similar purposes).

Thirdly, this program was already available at the university without any extra costs or licences (the other optional databases should have been purchased and the licences of such databases are remarkably expensive), and therefore the further usage of the library is more economical and easier, and it can be spread to all the persons who need the library data in their studies or for adding the information to the database.

Although this program has certain limitations such as structural and chemical analysis by using its own databases, the search can be done outside the program by using the hyperlink and OLE object linking systems; Also the structures and the molecule-specific information can be added to the MS Access, because it allows importing and saving different kinds of objectives into the database. All in all, this program has proved to be practical and applicable to the kind of library system in question. It can be extended, manipulated and new data can be easily added, and tables or forms can be created depending on the nature of the need.

5.4 Library database implementation

The DDTC NP Library is ready for use without any extra training. The basic information concerning the use of MS Access as software is available in two different forms (i.e. a printed version in library binders and at the workstation / on computer screen as files). Besides the actual DDTC NP Library binders it can be easily scanned from the same opening screen window (Hautamäki, 2005; Töyli, 1998). Also, the various free information databases from internet can be used to gain more information about the software (e.g. Access 2003 Product Information, Florida Gulf Coast University, Technology Skills Orientation; Access 2000 tutorial, Microsoft Office Access –Wikipedia, the free encyclopedia 2006).

5.5 Library use in DDTC

The DDTC has planned a restricted use for the library and the database, so that there will be nominated persons to have the rights to amend the library that is, to add or change the data. Otherwise the data can be distributed to other persons depending on the need and purpose for to view only -use. These could be situations where the data would be used for other related studies, or as a reference material.

However, the ultimate goal of this library material is to use it for the screening purposes in drug discovery programmes to find bioactivity from the samples. If activity is detected in the screening process, the original sample can be easily collected from the library samples, and further analytical separations and activity tests can be performed. When a pure natural compound is found, the information of the compound can be added to the library database. After further processing the material samples of these compounds may also be added to the Class V, Natural Compounds, which is already in reserve in the library.

The most beneficial function of this library and database is therefore saving time in screening process, as it is for other similar libraries widely used in commercial private laboratories, research institutes and among the pharmaceutical industry. It is also possible to add both the old and new data to the database from completed or ongoing research programmes of the DDTC, or as an extended use, from the other departments of the university if the results are related to natural products. The information from other databases, i.e. chemical, structural or otherwise related material in different forms (pictures, text, tables) can be imported to the NP Library database, and by that way expand the utility in many ways. Besides the software version of the library, the actual library material is also in ready-to-use format in the freezer. That is why the whole library system is easily implemented in practice and the usage of both library formats can be controlled by the DDTC.

5.6 Library use for educational purposes

One beneficial future prospect could be the use of the library for educational purposes at the Faculty of pharmacy. There could be partial selective processing of the library samples as a demonstration in education, or as starting material for laboratory courses, or for further studies in form of graduation projects. The protocols used in the library generation could be copied and used for other natural product materials, and the idea can be expanded to cover many other natural products than plants.

5.7 Library use internally within the Helsinki University

As natural products cover a wide range of materials, there are many possibilities to have different kinds of natural product material to use in bioactivity screening for drug discovery. Microbiological samples, semi-synthetic or the synthetic compounds derived from the natural products can serve as a sample resource for the expanded library system. The information of the new compounds can be saved, for example, to the Class VI (synthetic compounds), or to the other tables in MS Access database. New classes can be created easily. Moreover, in case of existing data in other databases or platforms, these can be connected or converted to MS Access by the copy/paste functionality, or by using various import and export features.

5.8 Library use externally by contracts

The library material can also be used externally by contracts done with DDTC. In these cases the DDTC can provide the final micro-fractions on plates for e.g. screening purposes, and if activity can be found, the further use of the initial samples can be contracted and agreed separately, as have been done by some commercial libraries (Abel *et al.*, 2002). In these cases the original identification data is protected and the other party of the contract is provided only with the codes of the samples.

6. Conclusion

As many of the already discovered breakthrough compounds have been of natural origin, there is no doubt that there will be many more new drugs in the future which are derived or originate from the natural products; This is because natural products are still a vast and mainly unknown recourse of bioactive metabolites, and the coding genes of these metabolites. The need for new medicinal compounds will increase, as the use of genomic data (i.e. by genomics, proteomics etc.) will increase the knowledge of both novel biological targets and of secondary metabolites, as will also do various engineering methods of secondary metabolite pathways. One very interesting future aspect will be the possible demand of individualized medicines. That may promote drug discovery of natural products, in order to develop more options to choose from to get more specified, individually suitable medication to fulfill the need covering differences in individual ADME properties and genetic metabolic mutations. In addition, natural product derived pharmacophores have more potential of having oral bioavailability compared to biological products (e.g. gene products). More options can be obtained by using combined technology, synthesis and medicinal chemistry, and by that way the possibility to exploit unique scaffolds found from natural product's secondary metabolites can be maximized. At the end of the day, these will create new innovations to pharmaceutical industry and be beneficial to all patients.

The exploitation of high-quality and high-throughput technologies will increase in the future, and the combination of different disciplines into integrated processes will have definitive impact to drug discovery programmes. The HTS technologies will move towards smaller and more focused compound libraries, and therefore different types of focused libraries types are needed increasingly. Natural product libraries are also beneficial tools for the identification and validation of new drug targets, and have a valuable role in drug discovery by serving faster development process.

Moreover, the applications of information technology will inevitably increase also possibilities of computational design and simulation in pharmaceutical industry. These are

already greatly developed but will become more sophisticated, and have a big impact and more important role in drug discovery programmes maximizing successful outcome. The data from natural product libraries can be easily exploited and used also with these information technology applications.

Few suggested improvements would still be ideal in case of expansion and for new library generation. First of all, these include good planning strategies before the actual laboratory procedures to avoid shortage of supplies and delays in actual procedures. Secondly, the average amount of starting material should be big enough to ensure that there will be no material loss during the procedures. And thirdly, any new laboratory procedures should be tested before processing large amounts of samples to avoid unexpected problems which could delay or change the planned procedures, and in case of change – to be prepared to change the protocol without hazarding the result.

As a result of this work, a pilot 23 735 membered natural product library was generated for the Drug Discovery and Development Technology Center (DDTC) to use in bioactivity screening; I believe that the library suitable for high-throughput screening (HTS) reduces the overall time of the discovery process. Moreover, the whole library system – the database and library material – can be used for different purposes depending on the current need and the library supply status. Hopefully, there will be many useful and exiting findings following the screening processes in the future, and most of all I wish, that the library could be exploited, developed and expanded further so that it can serve the purposes of many adopted usage forms.

References

Abel, U., Koch, C., Speitling, M., and Hansske, F.G. 2002. Modern methods to produce natural-product libraries. *Curr.Opin.Chem.Biol.* 6:453-458.

Access 2003 Product Information. Microsoft Office Online. Available from: http://www.microsoft.com/office/access/prodinfo/default.mspx

Albany Molecular Research, Inc. Discovery Services: Natural Product Libraries. Available from: http://www.albmolecular.com/discovery/natprod.html

Bleicher, K.H., Bohm, H.J., Muller, K., and Alanine, A.I. 2003. Hit and lead generation: beyond high-throughput screening. *Nat.Rev.Drug Discov.* 2:369-378.

Boldi, A.M. 2004. Libraries from natural product-like scaffolds. *Curr.Opin.Chem.Biol.* 8:281-286.

Davies K., 2006. The True Cost of Drug Discovery. BIO-IT World, June 06,2006. Available from: http://www.bio-itworld.com

Forthcoming: Dunlap, W. C., Jaspars, M., Hranueli, D., Battershill, C. N., Peric-Concha, N., Zucko, J., Wright, S. and Long, P. F. 2006. New Methods for Medicinal Chemistry – Universal Gene Cloning and Expression Systems for Production of Marine Bioactive Metabolites. *Curr.Med.Chem.*13(6):697-710.

Eldridge, G.R., Vervoort, H.C., Lee, C.M., Cremin, P.A., Williams, C.T., Hart, S.M., Goering, M.G., O'Neil-Johnson, M., and Zeng, L. 2002. High-throughput method for the production and analysis of large natural product libraries for drug discovery. *Anal.Chem.* 74:3963-3971.

Evans, M.J., Saghatelian, A., Sorensen, E.J., and Cravatt, B.F. 2005. Target discovery in small-molecule cell-based screens by in situ proteome reactivity profiling. *Nat.Biotechnol.* 23:1303-1307.

FitzGerald, Garret A. 2005. Anticipating Change in Drug Development: The Emerging Era of Translational Medicine and Therapeutics. *Nature Reviews Drug Discovery.* 4: 815-818.

Florida Gulf Coast University, Technology Skills Orientation; Access 2000 tutorial. Available from: http://www.fgcu.edu/support/office2000/access/tables.html

Galvez, M., Martin-Cordero, C., Houghton, P.J., and Ayuso, M.J. 2005. Antioxidant activity of methanol extracts obtained from Plantago species. *J.Agric.Food Chem.* 53:1927-1933.

Hautamäki,J. 2005. Microsoft Access-tietokantaohjelman perusteet. Merkonomin virtuaalikoulu. Available from: http://www.hai.cop.fi/henkilokunta/Jouko.Hautamaki/access/relaatiotietokanta/relaatiotietokanta.html

Hecht, P. High-troughput screening: Beating the odds with informatics-driven chemistry. [Feature] Current Drug Discovery, January 2002; 21-24. Available from: http://www.currentdrugdiscovery.com.

Hewitt, R., Gobbi, A. and Lee, M.-L. 2005. A Searching and Reporting System for Relational Databases Using a Graph-Based Metadata Representation *J. Chem. Inf. Model.*, **45** (4), 863 -869.

Hiltunen R. and Holm. Y., editors. 2000. Farmakognosia - Farmaseuttinen Biologia. Helsinki University Press, Helsinki.

Houghton, P.J. 2000. Use of Small Scale Bioassays in the Discovery of Novel Drugs from Natural Sources. *Phytotherapy Research.* 14:419-423.

IPNI. The International Plant Names Index. Available from: http://www.ipni.org/index.html

Jaspars, M., Long, P. F., Dunlap, W. C., Battershill, C. N. Sustainable Methods for the Production of Bioactive Marine Metabolites.[Presentation] Natural Products in Drug Discovery &Development, London,June 2005.

Koehler, R.T., Villar, H.O., 2000. Design of screening libraries biased for pharmaceutical discovery. *Journal of Computational Chemistry*. 21(13):1145-1152.

Koehn, F.E., and Carter, G.T. 2005. The evolving role of natural products in drug discovery. *Nat.Rev.Drug Discov*. 4:206-220.

Kovács, G., Kuzovkina, I. N., Szoke É., Kursinszki L. 2004. HPLC Determination of Flavonoids in Hairy-Root Cultures of *Scutellaria baicalensis* Georgi. *Chromatographia.* 60: S81 - S85.

Lautru, S., Deeth, R.J., Bailey, L.M., and Challis, G.L. 2005. Discovery of a new peptide natural product by Streptomyces coelicolor genome mining. *Nat.Chem.Biol.* 1:265-269.

Lee, M-L., and Schneider, G. 2001. Scaffold Architecture and Pharmacophoric Properties of Natural Products and Trade Drugs: Application in the Design of Natural Product-Based Combinatorial Libraries *J. Comb. Chem.*, 3: 284 -289.

Lee, S.I., Hyun, P.M., Kim, S.H., Kim, K.S., Lee, S.K., Kim, B.S., Maeng, P.J., and Lim, J.S. 2005. Suppression of the onset and progression of collagen-induced arthritis by chebulagic acid screened from a natural product library. *Arthritis Rheum.* 52:345-353.

LGC Business Sectors: Pharmaceuticals and Biotechnology. Available from: http://www.lgc.co.uk/service.

Mang, C., Jakupovic, S., Schunk, S., Ambrosi, H.D., Schwarz, O., and Jakupovic, J. 2006. Natural products in combinatorial chemistry: an andrographolide-based library. *J.Comb.Chem.* 8:268-274.

McAlpine, J.B., Bachmann, B.O., Piraee, M., Tremblay, S., Alarco, A.M., Zazopoulos, E., and Farnet, C.M. 2005. Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *J.Nat.Prod.* 68:493-496.

McCurdy, C.R., and Scully, S.S. 2005. Analgesic substances derived from natural products (natureceuticals). *Life Sci.* 78:476-484.

Microsoft Office Access –Wikipedia, the free encyclopedia: 2006. Available from: http://en.wikipedia.org/wiki/Microsof_Office_Access

Mijts, B.N., and Schmidt-Dannert, C. 2003. Engineering of secondary metabolite pathways. *Curr.Opin.Biotechnol.* 14:597-602.

Newman, D.J., Cragg, G.M., and Snader, K.M. 2003. Natural products as sources of new drugs over the period 1981-2002. *J.Nat.Prod.* 66:1022-1037.

PLANTS Database. United States Department of Agriculture, Natural Resources Conservation Service. Available from: http://plants.usda.gov/index.html

Rauha, J.P., Remes, S., Heinonen, M., Hopia, A., Kahkonen, M., Kujala, T., Pihlaja, K., Vuorela, H., and Vuorela, P. 2000. Antimicrobial effects of Finnish plant extracts containing flavonoids and other phenolic compounds. *Int.J.Food Microbiol.* 56:3-12.

Rees, S. 2001. Principles of High Throughput Screening. Lecture, Post Graduate Medical School, University of Surrey, U.K.

Rose, S., and Stevens, A. 2003. Computational design strategies for combinatorial libraries. *Curr.Opin.Chem.Biol.* 7:331-339.

Samuelsson, G. 1992. Drugs of Natural Origin. A textbook of Pharmacognosy. 3rd ed. Swedish Pharmaceutical Press, Stockholm.

Shang, S., and Tan, D.S. 2005. Advancing chemistry and biology through diversity-oriented synthesis of natural product-like libraries. *Curr.Opin.Chem.Biol.* 9:248-258.

Shen, B. 2004. Accessing natural products by combinatorial biosynthesis. *Sci.STKE.* 2004:pe14.

Singh, S.B., and Barrett, J.F. 2006. Empirical antibacterial drug discovery--foundation in natural products. *Biochem.Pharmacol.* 71:1006-1015.

Snyder, L. R., Kirkland, J. J. and Glajch, J. L. 1997. Practical HPLC Method Development. 2nd ed. John Wiley & Sons, Inc., USA.

Tammela, P., Wennberg, T., Vuorela, H., and Vuorela, P. 2004. HPLC micro-fractionation coupled to a cell-based assay for automated on-line primary screening of calcium antagonistic components in plant extracts. *Anal.Bioanal Chem.* 380:614-618.

Tan, D.S. 2005. Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nat.Chem.Biol.* 1:74-84.

Töyli, J. 1998. Tietojärjestelmän toteutus (TTE.135) Microsoft Access -97. Vaasan yliopisto/avoin yliopisto, Tietotekniikan ja tuotantotaloudenlaitos, s. 1-41.

Uehling, M. D. 2004. New Drug Costs Sky-High: $1.7B. BIO-IT World, January 12. Available from: http://www.bio-itworld.com

Van Lanen, S.G., and Shen, B. 2006. Microbial genomics for the improvement of natural product discovery. *Curr.Opin.Microbiol.* 9; 1-9.

Verpoorte, R. 2005. Dispelling the "Single compound/single target" paradigm-Evaluating metabolomics as a new tool for drug discovery. [Presentation] Natural Products in Drug Discovery &Development, London,June 2005.

Vuorela, P., Leinonen, M., Saikkuc, P., Tammela, P., Rauhala, J.P., Wennberg, T., Vuorela, H. 2004. Natural products in the process of finding new drug candidates. *Curr.Med.Chem.* 11:1375-1389.

Wennberg, T., Kreander, K., Lähdevuori, M., Vuorela, H., Vuorela, P. 2004. Primary Screening of Natural Products Using Micro Fractionation Combined with a Bioassay. *Journal of Liquid Chromatography & Related Technologies.* 27(16); 2573 – 2592.

Appendices

1. Cryotube Map 1+2, BE samples in 81 format
2. ScreenMatesTube Map 1+2, BED samples in 96 format
3. DDTC NP library sample colour classification
4. DDTC NP library instructions for users (in Finnish)