rijksuniversiteit groningen

Bachelor Thesis

Comparative analysis of synthetic strategies for the histone deacetylase inhibitor largazole from *Caldora penicillata*

An integrative literature review

Name: Marc Albert Simon

Student number: S3345602

Supervisors: Asst. Prof. Dr. Sandy Schmidt and Michael Runda, MSc

Department: Chemical and Pharmaceutical Biology

Abstract

According to reports provided by the World Health Organization, cancer is the second leading cause of death, accounting for approximately 9.6 million deaths in 2018. A prognosticated global increase of cancer cases among the population can be attributed to various risk factors mainly associated with the increasing age of the population. To counteract this tendency, there is a great demand for new medications to prevent or to treat cancer. In search of that, natural products are considered as promising candidates, representing potential leads for new drug developments. Cyanobacteria produce a large deviation of secondary metabolites due to their ability to survive in different biotopes Among them, largazole, a cyclic depsipeptide produced by *Caldora penicillata*, gained increased attention due to its observed bioactivity as class I histone deacetylase (HDAC), inhibiting the process of cell division. The bio-synthesis of largazole in *Caldora penicillata* is proposed to function via a hybrid NRPS/PKS pathway. Herein, a comparative analysis of available and possible synthetic strategies to synthesize the natural product largazole under economic and environmental aspects is reported. Therefore, recent literature on largazole found on searching tools as Pubmed, Scifinder and Smartcat were used. The results show that the highest yield could be obtained via the total synthesis pathways. This will be compared to a semi-approach based on the four pre-cursor building blocks of largazole found by the Williams group and the biological synthesis via culturing and extraction or *in vitro* synthesis via an enzymatic pathway. The aging of mankind and its additional health risks will cause cancer to be a more common disease in the future. Largazole could be a potent class I HDAC inhibitor therefore, different strategies are proposed for the synthesis of largazole. After comparison of the available and possible synthesis strategies for largazole with economical and sustainable feasibility in mind the best method for the synthesis of largazole to date is total synthesis.

List of abbreviations

Acetyl-CoA	acetyl coenzyme A
ACP	acyl carrier protein
AT	acyltransferase
HDAC	histone deacetylase
KS	ketosynthase
MeOH	methanol
NRPS	non-ribosomal peptide synthase
PCP	peptidyl carrier protein
PKS	polyketide synthase
SULT	sulfotransferases
TE	thioesterase
WHO	World Health Organization
IPP	isopentenyl diphosphate
AACT	Acetoacetyl-CoA thiolase
HMGS	HMG-CoA synthase
MK	Mevalonate kinase
PMK	Mevalonate phosphate kinase
MDC	Mevalonate decarboxylase
IPI	Isopentyl pyrophosphate isomerase
FPS	Farnesyl diphosphate synthase
ADS	Amorpha-4,11-diene synthase
SQS	Squalene synthase
AAR	Artemisinic aldehyde
ALDH1	Aldehyde dehydrogenase
A	Adenylation domain
CMT	C-methyl transferase
GNAT	GCN_5 N-acetyl transferase-like domain
KR	ketoreductase
DH	dehydratase
ER	enoylreductase
R	Reductase domain
C	Condensation domain

Index

1 Introduction

According to the World Health Organization (WHO), the second leading cause of human death is cancer. This disease accounted for 9.6 million deaths in 2018, with a rising tendency (1). This increase can be attributed to an aging human population for which cancer is considered an increasingly severe health risk in the ongoing future. The problem of obsolescence is linked to exposure to physical carcinogens such as radiation, chemical carcinogens, e.g. tobacco or food contaminants, and biological carcinogens such as the human papillomavirus, potentially causing crevicular cancer. Thus, there is a great demand for new medications to treat cancer and hence is the focus of ongoing drug development research. Considering the fact that 60 % of approved anti-cancer drugs are derived from natural sources, extending the scope for bioactive natural products could play a significant role in the ongoing search for new medications to treat cancer.

1.1 Natural products in drug discovery

Over decades, natural products and their broad range of intrinsic bioactivities have gained considerable interest from the scientific community. Inspired by the exciting antimicrobial activities of penicillin from *Penicillium chrysogenum* discovered by Fleming in 1928, much effort has been made to extend its applicability. Therefore, rational semi-, as well as total-synthetic approaches are reported in the literature aiming at the derivatization of penicillin (2). Approaches utilizing semi-synthetic and total synthetic strategies led to revolutionizing changes also for other natural products found in different organisms. One of these natural products is artemisinin, initially isolated from *Artemisia annua* in 1987 (Figure 1).

A) **B)**

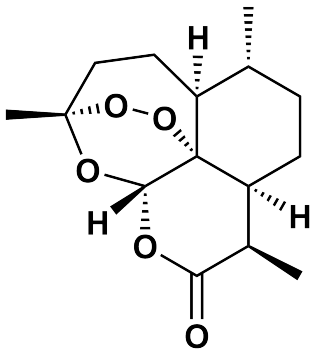


Figure 1. Artemisinin from Artemisia annua. A) Branch of Artemisia annua; B) Molecular structure of artemisinin. (Source: www.conifers.org/ta/Taxus_brevifolia.php)

This terpene metabolite shows beneficial bioactivities against malaria (2). However, the product yield of artemisinin obtained from *Artemisia annua* is relatively low, leading to a

global shortage of supply of this compound. To overcome this problem, alternative strategies to produce artemisinin via semi-synthetic or total synthetic strategies have been developed. The semi-synthesis of artemisinin can be achieved by producing artemisinic acid as a precursor via fermentation of engineered *Saccharomyces* cerevisiae, followed by chemical conversion to artemisinin (3). In 2018 Tang and coworkers revealed another concise semi-synthesis strategy for artemisinin via four consecutive chemical steps. The author proposed the synthetic chemical approach as a low-cost and efficient method to produce the anti-malarial drug artemisinin (4). Another example of a natural product is paclitaxel (Taxol®), a plant-derived chemotherapeutic isolated from the bark of the Pacific Yew (*Taxus brevifolia*) is used to treat different types of cancers, but primarily used to treat breast cancer (Figure 2) (5).

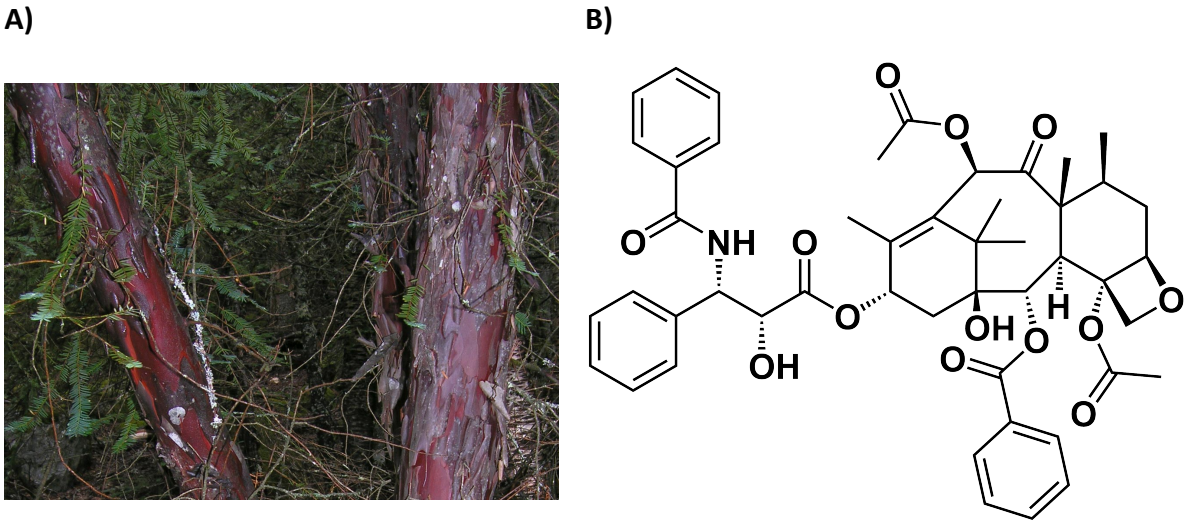


Figure 2. Anti-cancer drug paclitaxel (Taxol®) from Pacific Yew. *A) Bark of Pacific* Yew *(Taxus brevifolia); B) Molecular structure of paclitaxel (Taxol®).(Source: www.conifers.org/ta/Taxus_brevifolia.php)*

Like Taxol®, there are other natural products that show promising bioactivities, potentially effective against cancers. The cyclic depsipeptide largazole from the cyanobacterial strain *Symploca. sp.* shows *in vivo* anti-cancer activity due to its capability to function as histone deacetylase (HDAC) class I inhibitor. Revealing growth inhibitory activity and the ability to target transformed cells over non-transformed cells, largazole is considered as an important drug candidate in a more specific treatment of cancer (6).

1.2 Biosynthesis of natural products

Most of the pharmacologically active natural products are secondary metabolites produced by an organism. The pathways that are used in general to produce natural products are the acetate, shikimic acid, mevalonate and the methylerythritol phosphate pathways. What these pathways have in common are that they start with acetyl-coenzyme A(acetyl-CoA) ,except for the shikimic acid pathway(7). Via the acetate pathway the organism can produce secondary metabolites assigned to the classes of fatty acids and polyketides. The shikimic acid pathway is involved in biosynthesis of aromatic compounds such as the aromatic amino acids tryptophane or tyrosine and phenylalanine which serve as molecular precursors for various alkaloids. The mevalonate pathway and the methylerythritol phosphate pathway are responsible for the production of various terpenoid and steroid building blocks (7). For example, *Artemisia annua* uses both the mevalonate pathway (Figure 3) as well as the methylerythritol phosphate pathway to produce isopentenyl diphosphate (IPP), a precursor for the biosyntheis for artemisinin (8). For largazole another biosynthesis mechanism is proposed, this secondary metabolite is produced via a hybrid non-ribosomal peptide synthetase (NRPS)/ polyketide synthase (PKS) pathway.

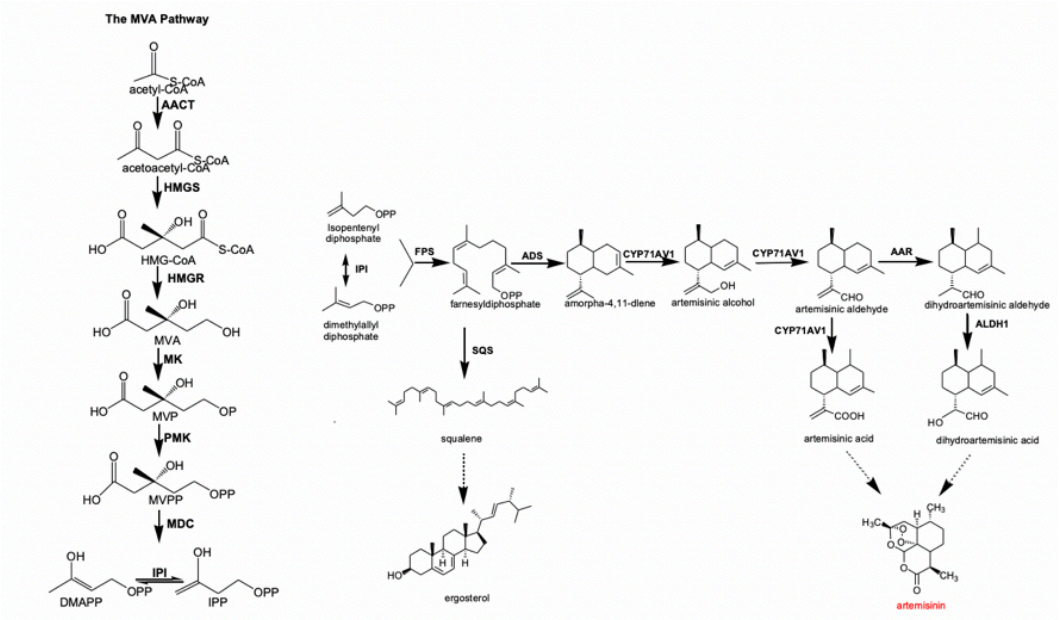


Figure 3. Biosynthetic pathways from Artemisia annua involved in the production of artemisinin (8).

1.3 Non-ribosomal peptide synthetase/ polyketide synthase

Unlike protein synthesis via ribosomes facilitating the translation machinery, biosynthesis of non-ribosomal peptides in bacteria and fungi is driven by the NRPS pathway, the PKS pathway,

or a hybrid mechanism of both. It is assumed that the biosynthesis of the cyclic depsipeptide largazole from *Symploca sp.* is driven by an NRPS/PKS pathway. Natural products produced via PKS pathway can be assigned to three main types of enzymatic reactions defined as type I PKS, type II PKS and type III PKS. Type I PKSs are heteromultimeric protein complexes composed of several enzyme subunits. These subunits are organized in various modules, which consist of various domains. Typically, an elongation module of the PKS I pathway consists of three domains, including an acyltransferase (AT) for loading of the chain extender unit, an acyl carrier protein (ACP) for tethering and a ketosynthase (KS) domain to catalyze a condensation reaction between the ACP and the KS active site. To terminate chain elongation of the PKS pathway, termination modules are added. These consist of a thioesterase (TE) or a reductive domain. The TE domain is used for hydrolysis, forming a linear polyketide. This domain can also cause cyclization to form a macrocyclic polyketide or cause sulfonate transfer and decarboxylation with an extra sulfotransferase (SULT) domain to form a terminal olefin. The R domain is used to form aldehydes (9) (Figure 4). In contrast to most peptides, these peptides are synthesized via an enzyme-controlled process in which the amino acid is added via the enzyme's specificity. The NRPS pathway has many similarities with the modular arrangement of PKS type I. NRPS can be divided into two types, defined as NRPS(I) and NRPS(II). The NRPS pathway starts with the activation of respective amino acids via adenylation with ATP to corresponding adenosine monophosphate esters. These activated intermediates are then accessible for the NRPS pathway by forming a thioester bond with a thiolation domain considered as the initial step of the chain elongation process. Besides the thiolation domain, a typical elongation module of the NRPS pathway consists of an adenylation(A) domain, a peptidyl carrier protein (PCP) and a condensation(C) domain, often arranged in the order of C-A-PCP. The A domain is used to load the chain, the PCP domain is used for tethering and the C domain is used for condensation to catalyze the peptide bond formation. Termination of the chain elongation is induced by a TE domain, resulting in a linear or a macrocyclic peptide dependent on the formation of ester or amide linkages between the amino acids during biosynthesis (7).

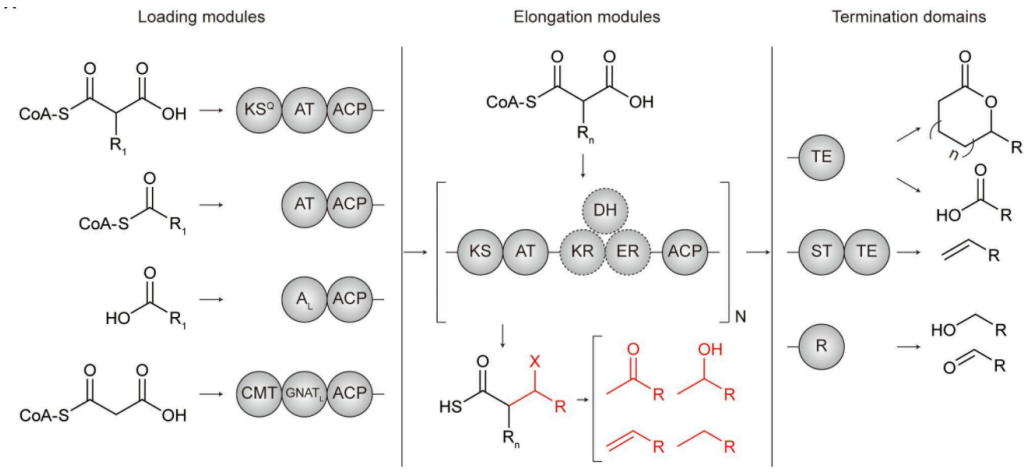


Figure 4. A schematic representation of the available modules for PKS natural product synthesis(9)

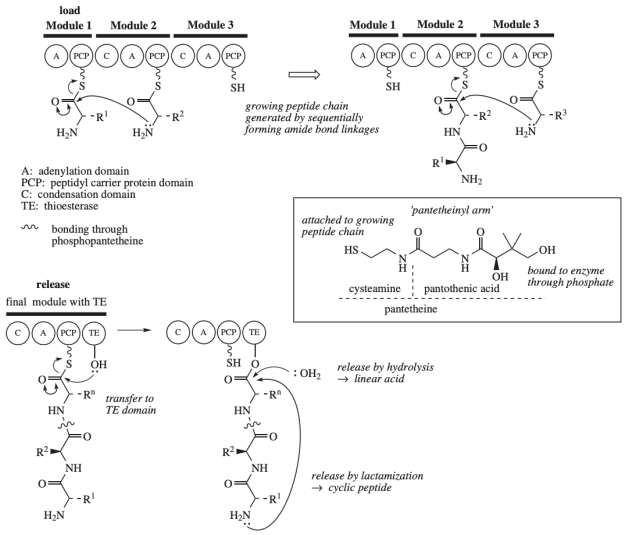


Figure 5. A schematic representation of the available modules for NRPS natural product synthesis(7).

1.4 Cultivation and extraction

A straightforward approach to obtain natural products is to extract the secondary metabolites directly from its natural producers. Taking advantage of the highly specific nature of biosynthetic pathways within organisms, the use of additional chemicals or enzymes could be circumvented. Thus, this approach is considered less harmful to the environment due to the lack of potential toxic catalysts, the avoidance of harsh reaction conditions, and low waste material formation. Besides described beneficial environmental aspects, the extraction of natural products from the natural producers constitutes an economically feasible way to obtain desired products. However, one disadvantage of this method is that the synthesis of a specific bioactive compound is dependent on the growth and production of the organism(10). A low production/biomass ratio was the primary driving force in searching for alternative

strategies to obtain the anti-malaria drug artemisinin from *Artemisia annua* in reasonable quantities. Therefore, scientists developed new methods such as total and semi-synthetic strategies to improve the supply of artemisinin. Metabolic engineering of natural producer organisms is a powerful strategy to enhance desired product yields *in vivo*. However, elucidation of the respective gene cluster involved in the biosynthesis of the desired compounds and knowledge of competitive metabolic pathways is crucial. Another possibility is to search for an alternative host organism that can be engineered to produce precursor molecules for the desired natural product, as applied for the semi-synthesis of artemisinin(11). By the use of alternative engineered organisms, low product yields in the natural producer organism could be circumvented (11).

1.5 Total synthesis of natural products

The first chemist that achieved to synthesize a natural product was Friedrich Wöhler in 1828, who reported the successful total synthesis of urea, without the need for biological processes. By that, it was demonstrated that it is possible to synthesize naturally occurring molecules by chemical means in the laboratory (12). Nowadays, the total synthesis of natural products is not limited to simple compounds such as urea, but it is possible to synthesize the most complex molecules found in nature. Artemisinin is a well-known example of a natural product on which organic chemists were challenged to find a strategy to synthesize this potent anti-malarial drug chemically to overcome relatively low product yields obtained by extracting artemisinin from the tree *Artemisia annua*. Therefore, an independent strategy of synthesizing the drug artemisinin in the lab could solve this global shortage. Another benefit is that when a strategy is developed, modifications to the natural product scaffold can be conducted in order to overcome clinical limitations such as poor solubility, which has been done for paclitaxel (5). The challenges of proposing a strategy for the total synthesis of natural products are to keep the synthesis steps minimalized and to perform it economically beneficial. This means that organic chemists need to think about ways to specify their reactions to minimalize the waste of unwanted intermediate products and lower costs by using readily available chemical compounds (12).

1.6 Semi-synthesis of natural products

Another type of chemical synthesis is semi-synthesis. Unlike the complete chemical conversion of commercially available precursors into the desired product in total synthetic approaches, the concept of semi-synthesis aims to combine chemical conversion with the inherent reactivity of enzymes or microorganisms, functioning as biocatalysts. Due to a global shortage of artemisinin, an efficient semi-synthetic strategy to produce this drug in reasonable quantities gained high interest. As reported, engineered yeast strains can be used as biocatalysts to produce artemisinic acid, a late-stage precursor which can be converted to artemisinin via a two-step chemical reaction. Developing semi-synthetic methods could therefore decrease the costs and secure the supply of artemisinin in a more specific manner than total synthesis. Challenges to propose a semi-synthesis strategy for natural products are that the gene-cluster of the natural product has to be discovered and that a suitable host organism has to be found which can be modified to produce the selected precursor molecule (3).

1.7 Natural products from cyanobacteria

Cyanobacteria, commonly termed as blue-green algae, are a group of ancient Gram-negative photoautotrophic bacteria(13,17,27). They live in a wide range of biotopes like in the ocean, freshwater, thermal springs, or terrestrial environments. Cyanobacteria possess c-phycocyanin used for photosynthesis, there are two processes which could occur a light-driven process in the presence of water and visible (sun) light and the Calvin cycle in which abundant atmospheric carbon dioxide (CO_2) is assimilated and subsequently converted into energy-rich organic molecules. Besides photosynthesis under aerobic conditions via the interaction of photosystem I and photosystem II, these organisms can also perform photosynthesis under aerobic conditions via photosystem I. By that, cyanobacteria can produce a broad range of primary and secondary metabolites referred to as natural products. In particular, beneficial bioactivities observed for compounds assigned to the latter group of natural products have gained increased attention over the past decades, considered potential leads in drug development research (13). An example of a natural product originating from a secondary metabolite found in cyanobacteria is cryptophycin 1 obtained from cyanobacterial strain *Nostoc. sp.* used to treat a wide range of solid tumors (6, 14). Largazole is another bioactive secondary metabolite found in cyanobacteria. Bioactivity assays revealed promising

HDAC class I inhibitory effects. Furthermore, its chemical scaffold shows high similarity to the known HDAC inhibitor FK228, a macrocyclic depsipeptide isolated from the bacterial strain *Chromobacterium violaceum*.

1.8 Chemical diversity of natural products from cyanobacteria

Cyanobacteria produce a wide range of natural bioactive products. Thus, they are able to settle in diverse habitats. Earlier studies identified ten distinct classes of secondary metabolites. Besides diverse compounds assigned to alkaloids, cyanobacteria produce various depsipeptides, lipopeptides, macrolides, peptides, terpenes, polysaccharides, lipids, polyketides and others. Peptides represent the largest group of metabolites due to the non-ribosomal and polyketide synthase present in the cyanobacterial strains (15). Due to the described broad scope of natural products, cyanobacteria are considered promising model organisms for identifying potential leads for novel drug development.

1.9 Largazole

The compound largazole is produced by the cyanobacterial strain *Caldora penicillata.* This secondary metabolite caught researchers' attention due to its potent HDAC class I inhibitory activities and its potent anti-cancer synergy with symplostatin 4, another natural product coproduced by the same *Caldora* strain. The compound largazole heritages its name from the lake Key Largo Florida where the cyanobacterial strain was extracted and the two azole groups (Figure 6). The compound largazole contains a functional thioester group which is important for its mechanism of action. Due to this mechanism largazole possesses growth inhibitory activity, with specific targeting of transformed cells over non-transformed cells. Which is tested on fibroblastic cancer cells such as colorectal carcinoma and breast cancer and healthy cells(16). Furthermore, largazole consist of an (*S*)-valine , a thiazole-thiazoline unit, a unique 3-hydroxy-7-mercapto-hept-4-enoic acid (16).

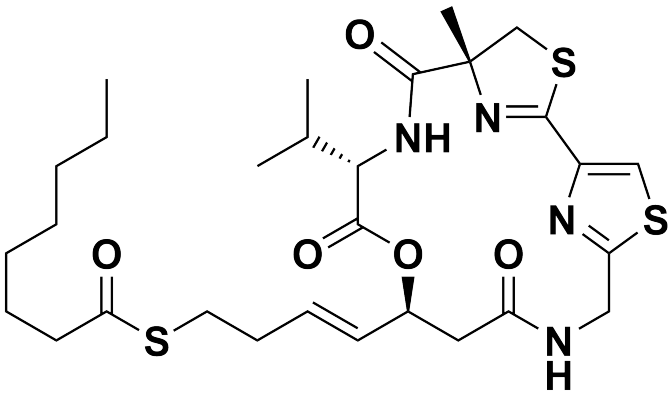


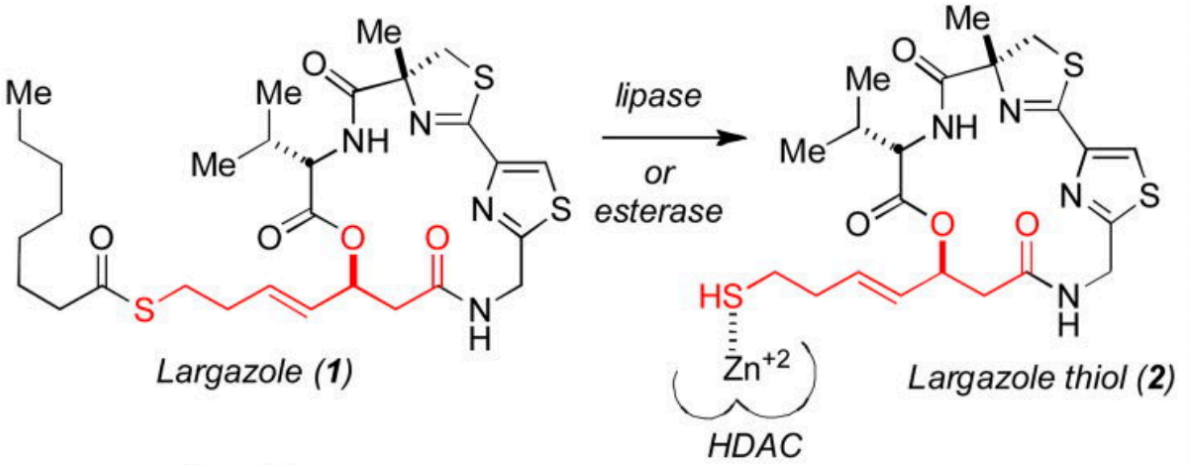
Figure 6. The structure of the natural product largazole

1.9.1 Taxonomy of the producing strains

Initially, the producing strain for largazole was defined as *Symploca. sp*, which was retrospectively renamed to *Caldora penicillata*(16). The morphology is described as amorphic feathery or puffball clumps, with deviating colors ranging from red to orange and green. The filaments of this strain are made of entangled and thin smaller unbranched filaments. The cells are predominantly barrel-or cylindrical-shaped. This species can be found in many coral reef habitats where associated members are attached to hard bottoms. Among 17 identified specimens of the *Caldora* family, only ten are known to produce the secondary metabolite largazole, which most of them were found in the aqueous environment of Key Largo, Florida. Described spatial dependency suggests that environmental factors in the lake may contribute to the capability of cyanobacteria to produce largazole (17).

1.9.2 Bioactivity of largazole

The core of the molecule largazole consists of a macrocyclic depsipeptide containing a (*S*)-valine, thiazole-thiazoline fragment and an unsaturated thioester. The latter is susceptible to hydrolysis, causing the prodrug to become the active ligand for HDAC inhibition (Scheme 1)(18). The active compound largazole thiol (2) can complex Zn^{2+} in the active site of HDACs (Figure 6). To understand why largazole could be an important drug candidate to battle against certain types of cancer, it is essential to understand the principle of HDAC inhibitors.



Scheme 1. The bioactivation of largazole (18).

1.9.3 HDAC inhibitors

HDACs are enzymes that catalyze the deacetylations of the lysine residues of histones, causing the negatively charged DNA to interact with these proteins, resulting in a chromatin arrangement and thereby regulate transcription. Nowadays, 18 classes of HDACs are known and divided in Zn^{2+}-dependent (I, II, IV) and Zn^{2+}-independent, NAD-dependent enzyme classes (III). Overexpression of HDACs of class I, which are highly expressed in prostate, lung, colon and breast tissues, cause the formation of tumors (19). By inhibiting HDACs activity, cell cycle arrest and apoptosis are induced while angiogenesis is reduced. Therefore, HDAC inhibitors are potent drug candidates to treat these types of cancers. When an HDAC inhibitor is applied, cell cycle gene expression of cyclin-dependent kinase inhibitor p21 is increased. This increase of p21 will lead to a decrease in dimer formation of the cyclin-dependent kinases and therefore induce cell cycle arrest. Apoptosis is induced by upregulating the pro-apoptotic and downregulating the anti-apoptotic genes. HDAC inhibitors will therefore initiate both the intrinsic and the extrinsic apoptosis pathway, influencing TNF related apoptosis inducing ligand (TRAIL), death receptor 5 (DR5) and the Fas ligand via the extrinsic pathway. The pro-apoptotic genes which are upregulated for the intrinsic pathway are BAX,BAK and APAF and the anti-apoptotic gene Bcl-2 is down regulated. Finally, HDAC inhibitors perform antiangiogenic effects by down regulating the angiogenic genes vascular endothelial growth factor(VEGF) and endothelial nitric oxide synthase (eNOS)(20). The function of HDAC inhibitors is schematically shown in (Figure 7)

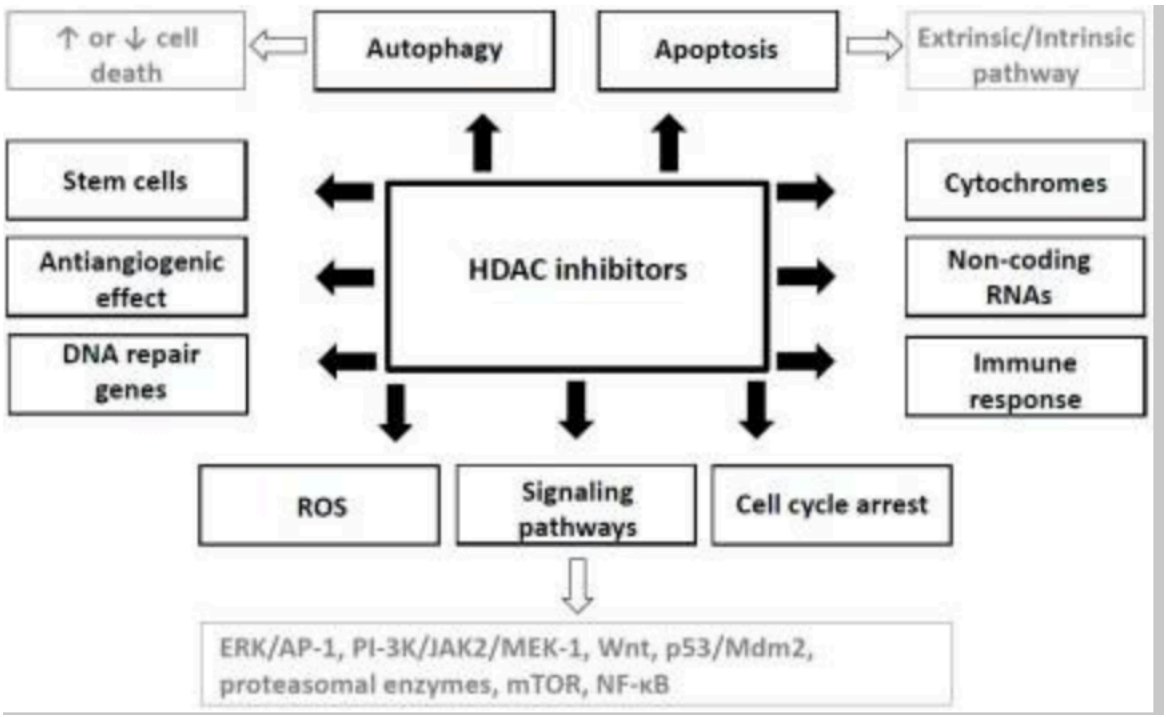


Figure 7. schematic overview on the working mechanism of HDAC inhibitors

2 Thesis aims

As reported for the natural product artemisinin, different synthetic approaches could be considered to obtain natural products independent of the natural producer organism's metabolic condition. However, synthetic strategies all have their advantages and disadvantages. The aim of this thesis is a comparative analysis of available as well as possible approaches to obtain the natural product largazole in reasonable quantities under the consideration of economical and environmentally sustainable feasibility. The potential of enzymatic, semi-synthetic and total synthetic approaches compared to the direct isolation of largazole from the naturally producing strain *Caldora penicillata* are reviewed and discussed in detail.

3 Materials and methods

To write this review on the different strategies to obtain largazole, different searching machines have been used. The searching machines included: SmartCat (library catalog of University Groningen), SciFinder and PubMed. For this review, all articles were included from the discovery of largazole until today, focused on recent literature.

For this thesis the literature is used to provide a comprehensive overview on this topic. The searching terms for the articles that were used are: 'cancer', 'natural products', 'gene cluster', 'largazole', 'HDAC' or 'histone deacetylase inhibitors', 'symploca sp' 'caldora penicillata', 'FK228', 'artemisinin', 'Taxol', 'NRPS' or 'non-ribosomal peptide synthetase', 'PKS' or 'polyketide synthase,' 'semi-synthesis', 'total synthesis'.

4 Results

The need for new cancer therapeutics is of high interest in the ongoing drug development research. Bioactivity assays revealed that largazole, a secondary metabolite found in cyanobacteria, can inhibit HDAC. Thus, largazole is considered as a promising chemical lead in the search of new cancer therapeutics. A comparison between available approaches as well as possible approaches to obtain largazole are represented, considering the obtained yield and the synthesis steps that are needed to obtain largazole.

4.2 Total synthesis

Since the promising bioactivity of largazole towards cancer cells was revealed, different research groups have been busy discovering a strategy to produce this compound via chemical synthesis. By today, eleven protocols dealing with the total synthesis of largazole have been described in the literature. Herein, three selected approaches, chosen based on the product yields and synthetic steps required, are reviewed and discussed in more detail.

4.2.1 Williams total synthesis

One of the strategies aiming at the chemical synthesis of largazole is described by Williams et al. Respective approach is based on the four precursor molecules α-methylcysteine, (*S*)-valine, thiazole and (*S*)-3-hydroxy-7-mercaptohept-4-enoic acid (Figure 8)**.** In the corresponding

protocol, eight linear reaction steps are required, resulting in an overall product yield of 37%, which corresponds to roughly 12-19 mg of largazole (Scheme 3). Besides the advantage of using readily available precursors for the synthesis of largazole, the macrocyclization strategy constitute a major challenge in this approach (18).

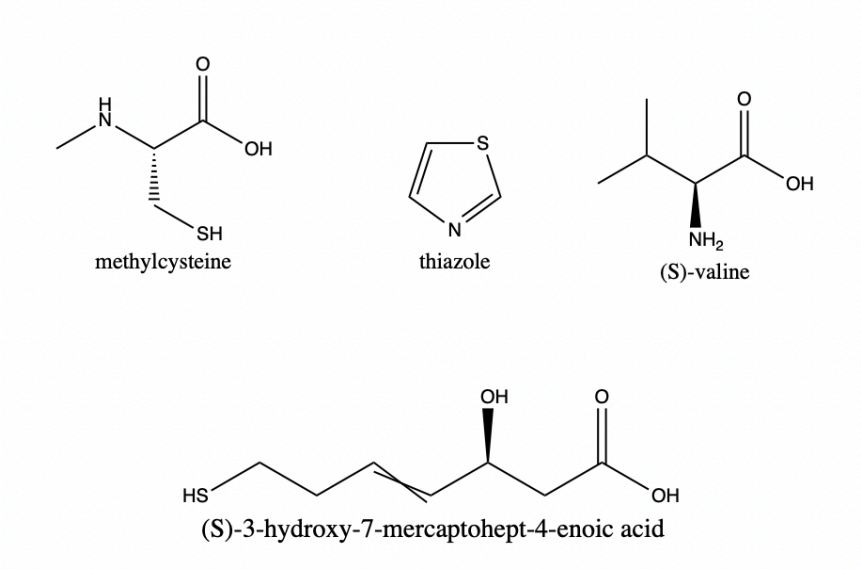
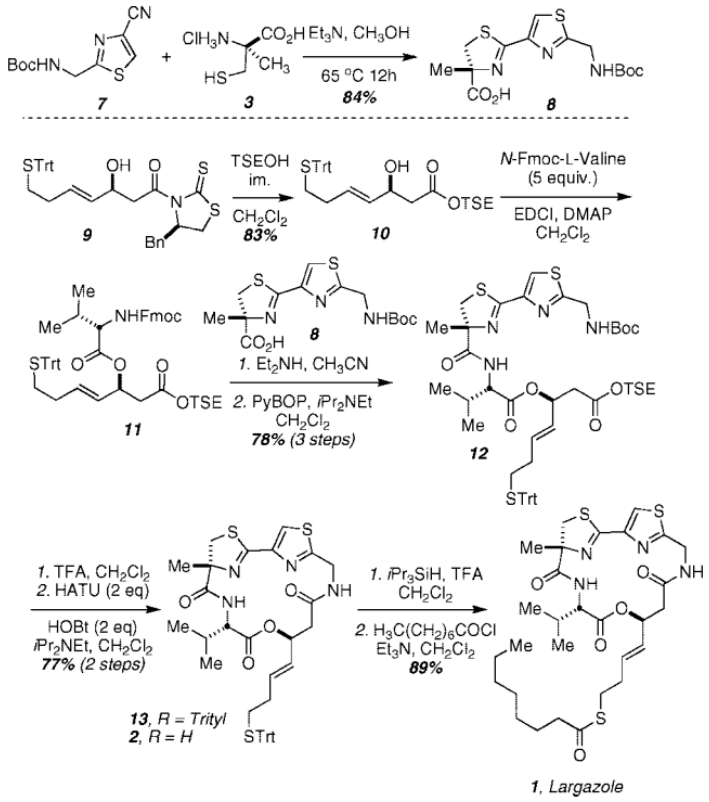


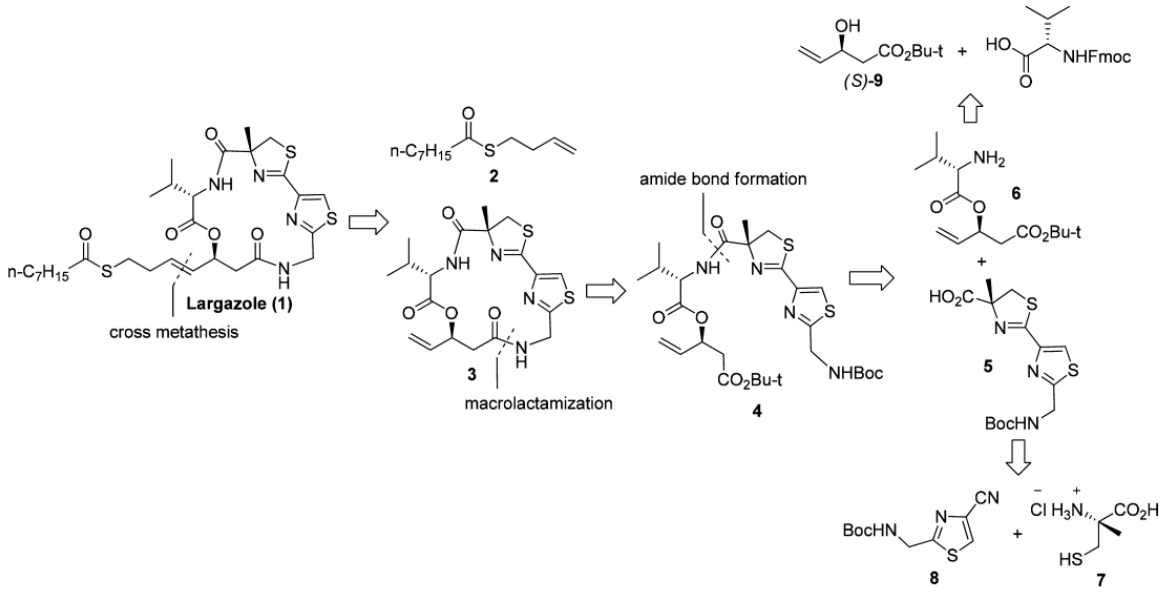
Figure 8. The four precursor building blocks found by retro synthetic analysis of the Williams group(18)



Scheme 3. The proposed synthesis strategy according to the group of Williams (18).

4.2.2 Luesch total synthesis

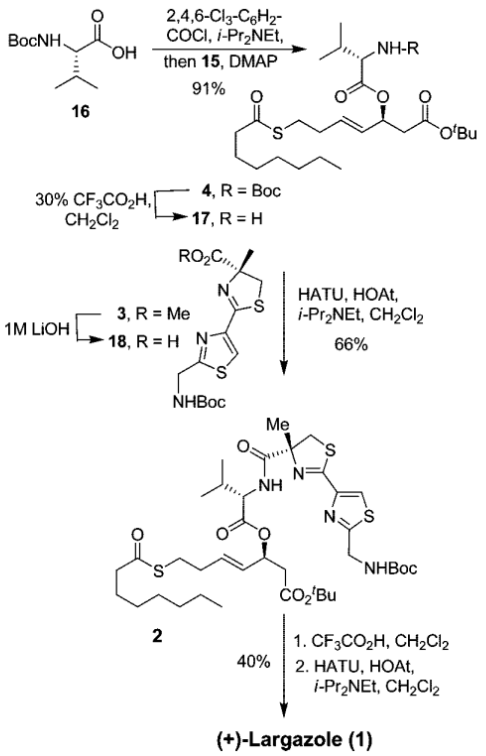
Recently, the research group of Luesch proposed a protocol for the scale-up total synthesis of largazole. This research group achieved to install the thioester moiety of largazole via an olefin cross-metathesis. Unlike Williams et al., Luesch and coworkers used three chemical precursor molecules as available starting compounds, including an acetyl analog indicated with number 5, a thiol analog indicated with number 8 and a hydroxyl analog indicated with number 9 used to form a linear precursor as an intermediate for the macrocyclization reaction are shown in scheme 4. The final product largazole could be obtained in eight linear steps with an overall yield of 21 % corresponding to roughly 7.5-22 mg of largazole (Scheme 4) (21, 22).



Scheme 4. The proposed strategy for the chemical synthesis of precursor molecules for largazole proposed by the group of Luesch (22)

4.2.3 Ghosh total synthesis

The group of Ghosh investigated a strategy for the enantioselective total synthesis of largazole. This group also uses a cross-metathesis reaction like Luesch et al. For the chemical synthesis proposed by Ghosh, four precursor molecules were used: (*R*)-2-methyl cysteine, a thiazole acid, a thioester and an optically active allylic alcohol. With these starting materials via cross-metathesis, the linear precursor of largazole could be obtained. After forming the 16 membered cycloamide, a two-step reaction resulted in (+)-largazole with an isolated yield of 40 %, which corresponds to 6.3 mg of final product (**Fout! Verwijzingsbron niet gevonden.**10) (23).



Scheme 4. The total synthesis proposed by the group of Ghosh (23).:

4.3 Semi-synthetic approach

As shown in the example of artemisinin, a semi-synthetic approach to produce largazole is conceivable. Although no publications have been done on this topic yet, it could be advantageous to produce one of the precursor molecules of largazole through another organism. Based on previous literature, α-methyl cysteine, thiazole, (*S*)-valine and (*S*)-3-hydroxy-7-mercaptohept-4-enoic acid are considered as key building blocks in the total synthesis of largazole. Directed metabolic engineering of other organisms than *Caldora penicillata can* reduce the number of chemical steps needed for the total synthesis and to aid a more specific way to synthesize largazole.

4.3.1 (*S*)-3-hydroxy-7-mercaptohept-4-enoic acid

The bacterial strain *Chromobacterium violaceum* produces the HDAC inhibitor FK228, via a hybrid NRPS/PKS route, conserved in an identified gene cluster. FK228 is a potent HDAC class I inhibitor, which belongs to the depsipeptide macrocycle molecular family revealing a high structural homology to largazole. Thus, the NRPS/PKS route can ultimately be used for a semi-synthetic approach to produce a specific precursor for the synthesis of largazole. The molecular structures of largazole and FK228 exhibit a (*S*)-3-hydroxy-7-mercaptohept-4-enoic

acid moiety, which is a common motif for natural products which are known to inhibit HDACs. If the known gene cluster for biosynthesis of FK228 is closely discussed, it is striking that the DEPd gene module 4 could be an important module to synthesize the precursor needed to produce largazole. This is shown in figure 9 which shows the modules used to synthesize FK228(24). The DEPd module a version of a linear sequence which looks like the (*S*)-3-hydroxy-7-mercaptohept-4-enoic acid of largazole.(18, 24).

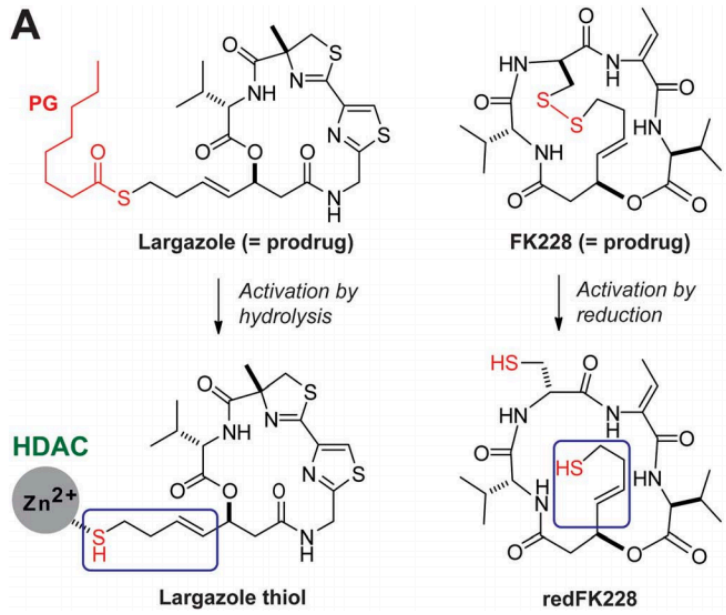


Figure 8. Comparison of the potent HDAC inhibitors FK228(romidepsin) and largazole (24). PG: protective group

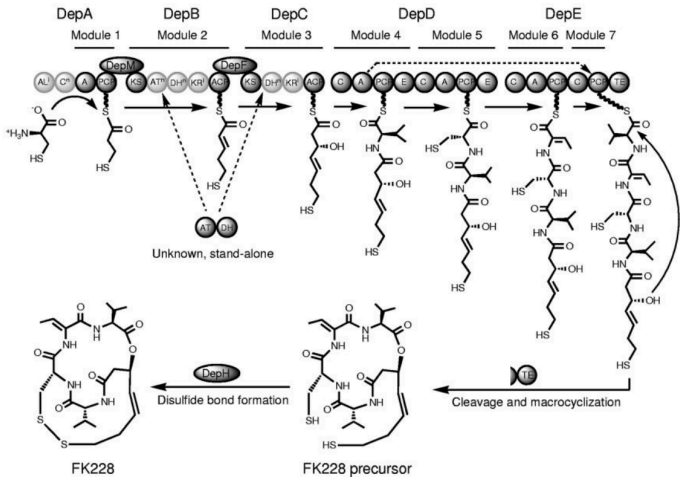


Figure 9. hybrid NRPS/PKS synthesis pathway of FK228(24).

4.3.2 (S)-valine and α-methylcysteine

(*S*)-valine and α-methylcysteine are two compounds used in the total synthesis of largazole(18). These two compounds can be inter alia found in *Carissa edulis* (bush plum). Its fruit shows a relative content of 1.28 mg/g cysteine and 1.22 mg/g valine (25). Based on that,

extracting the amino acids directly from the fruits could be considered an initial step in a possible semi-synthetic route for the production of largazole.

4.3.3 Thiazole

Thiazole is a heterocyclic compound often found as a functional group in natural products. As indicated in previous literature, thiazole containing natural products commonly exhibit biological activities against several microbial organisms such as fungi. For example, *Streptomyces atroolivaceus* produces leinamycin, a natural thiazole-containing natural product, synthesized via an elucidated NRPS/PKS pathway. Respective gene clusters involved in the biosynthesis of the thiazole are indicated as CY-FP and OX-RP. Exploiting this biosynthetic pathway can be considered a possible strategy to use this host organism as a potential biocatalyst to produce the needed thiazole precursor for the semi-synthesis of largazole (26).

4.4 Culturing ,Extraction and isolation

The strain *Symploca sp.* was discovered in 1979 by Pearson et al. Later on, this strain was renamed to *Caldora penicillata* by the research group of Luesch(16). The strain of *Caldora penicillata,* which produces largazole, can be found in Key Largo, Florida and is generally categorized as a type III aerobic, nitrogen-fixing and non-heterocystous organism that occurs in tropical environments. To culture this strain under optimum conditions, it is crucial to meet the ideal requirements, which could be achieved by mimicking the environmental characteristics of its natural habitat. For cultivating *Caldora penicillate*, optimum growth conditions can be established via an environment with high light attenuation, limited diffusion causing anoxia at night, high oxygen concentration, and a low pH, with a water temperature that fits the mean temperature of the lake (17, 27).

The research group of Luesch described a protocol for the extraction of largazole from *Caldora penicillate* strain collected from its natural habitat Key Largo, Florida(21). After freeze-drying of the cells, liquid-liquid extraction with methanol (MeOH)/ethyl acetate (ration 1:1) resulted in a lipophilic extract with a mass of 0.29 g. This lipophilic extract was portioned in hexane and an aqueous solution of MeOH (80 %). The more polar MeOH was then concentrated and fractionated with dichloromethane (CH_2Cl_2) for reversed-phase HPLC, in which, as indicated,

largazole eluted at a retention time of 61.5 minutes with the use of an YMC pack ODS-AQ column and a flow rate of 2.0ml/min at a wavelength range from 220-254nm. A final amount of 1.2 mg largazole could be isolated from 290 mg of Caldora penicillate sample as reported. The obtained amount of largazole corresponds to a total product yield of 0.41 % (28).

4.5 In vitro biosynthesis of largazole

The gene cluster for the biosynthesis of largazole from *Caldora penicillata* has yet to be published. Hence, just speculation about a possible NRPS/PKS biosynthetic pathway can be done. Another potential approach to synthesize largazole could be the enzyme-catalyzed *in vitro* synthesis by utilizing enzymes encoded from the organism's metabolic gene cluster. Therefore, desired enzymes could be first overexpressed in a suitable host organism and then used as biocatalysts for the cell-free enzyme-catalyzed synthesis of largazole. The use of complement enzyme engineering techniques to improve enzyme activity and, consequently, the overall product formation must be considered in establishing an economically feasible process (29).

5 Discussion

The present discussion section compares different synthetic strategies aiming to produce largazole at reasonable quantities based on expected product yields and synthetic steps required. The comparison between different strategies is important to identify the best strategy to produce largazole. First, the total synthesis strategies followed by the semi-synthetic approach and finally, the extraction of largazole from the natural producer organisms will be compared and discussed.

5.1 Total synthesis

The chemical synthesis of the natural product largazole attracted considerable interest of scientists in the organic chemistry community. Nowadays, eleven strategies are proposed for the chemical synthesis of largazole. Among these protocols, three showed promising product yields and displayed low numbers of linear steps required for synthesis. While the Williams total synthesis includes eight reaction steps resulting in a product yield of 37 % (12 – 19 mg)(18), the Luesch synthesis aims to produce largazole via eight reaction steps, and

an expected product yield of 19 % (7.5 mg - 22 mg)(21). On the other hand, the Ghosh synthesis describes a chemical synthesis strategy for largazole with a reported yield of 40 % (6.3 mg) in two linear steps(23). Only one synthesis strategy proposed a protocol for the upscale process for this chemical synthesis which was the one of the Luesch group. Based on the production yield and synthesis steps between these three strategies the highest yield 40% and lowest synthesis steps could be found in the enantioselective method described by Ghosh, although this is not the most common used method to synthesize largazole which are the Williams synthesis or the Luesch synthesis. Hence, the total synthesis of Luesch shows to be the most promising strategy for the total synthesis of largazole. Despite the Williams synthesis has a higher product yield 37% and an equal amount of synthesis steps needed. The reason for this is a briefly described upscaling process of the Luesch group(22). In the future, these strategies can be used to modify the chemical scaffold of largazole. Whereas the group of Luesch already experimented with improving the Zinc binding affinity for the future development of largazole analogs(30). This binding site is very important for the use of largazole as a drug because this causes largazole to form a complex in the active site of HDAC causing the drug to be more potent. Another possibility of modifying the natural product largazole is in the depsipeptide ring. Almaliti et al. (2016) reported that modifying the depsipeptide ring of largazole could positively affect its HDAC inhibitor selectivity (31). Corresponding study also shows that replacing the thiazoline ring and removing the thiazole-thiazoline group do not higher inhibitory effects but represent exciting compounds that could be targets of future optimization studies (31). Although these methods are used to modify and control the supply of largazole, not one of the synthesis strategies showed to be a sustainable method for the environment because the use of the chemical catalyst needed for the total synthesis of largazole examples of these reagents are the lawessons reagent which is used for the enantioselective synthesis of largazole(26) or the grubbs catalyst used in the Luesch total synthesis(21,22).

5.2 Semi-synthesis

A more sustainable and environmentally benign approach to obtain largazole in reasonable quantities could be a semi-synthetic strategy, which, however, has not been published yet. As emphasized in the result section, approaches to biologically synthesize some of the precursor molecules used for synthesizing largazole by using different microorganisms are conceivable.

First, the use of the bacterial strain *Chromobacterium violaceum* for the production of precursor (*S*)-3-hydroxy-7-mercaptohept-4-enoic acid group(24). This organism utilizes this precursor to synthesize the HDAC inhibitor FK228 produced via a hybrid NRPS/PKS pathway. The genes involved in the biosynthesis of this precursor are conserved in the elucidated *dep* gene cluster. This module could be an essential lead for the semi-synthesis of largazole by using engineered *Chromobacterium violaceum* or another microorganism such as *E. coli* harboring the respective gene cluster as biocatalysts.

The precursor thiazole could be produced by *Streptomyces atroolivaceus.* This organism produces leinamycin via a hybrid NRPS/PKS pathway where the thiazole fragment is encoded via CY-FP and OX-RP(26). Besides engineering the natural producer strain *Streptomyces atroolivaceus*, respective genes could be cloned into another host organism to synthesize the thiazole precursor for the semi-synthesis of largazole. (*S*)-Valine and α-methylcysteine are two amino acids that can readily be extracted from the fruit of *Carissa edulis*. A problem by performing the extraction method is that this strategy is dependent on the concentration in the fruit, whereas buying these amino acids on the market is relatively cheap . A problem with these four building blocks is that these compounds are proposed by the retrosynthetic analysis of the Williams group and therefore only applicable for the corresponding protocol. However, the (*S*)-3-hydroxy-7-mercaptohept-4-enoic acid fragment is also reported in the total synthesis of Luesch.

5.3 Culturing, extraction and isolation

The simplest and straight forward way to obtain largazole is to isolate and extract the compound directly from the organism *Caldora penicallata.* The result section describes a protocol from the group of Luesch where they extract largazole from the obtained sample from Key Largo, Florida. This method showed that 290mg of cyanobacterial sample resulted in a yield of 0.41 % (1.2 mg) largazole. This result shows that the supply of largazole from the cyanobacterial strain seems to be very low. Although not described by the authors, the choice of growth medium such as BG-11 could have an impact on the production of largazole. Janson et al. (1998) showed that the *Symploca antlantica* Pcc 8002 showed growth in the artificial sea medium ASN(III). Since this organism belongs to the same family, this medium could also be a potential candidate for culturing *Caldora penicillata* to produce largazole(32). Another problem with the protocol for extraction of the Luesch group is that they did not describe

what parameters they used to contain and grow the sample of *Caldora penicillata.* Since the metabolic gene cluster of largazole has not been elucidated yet, no *in vitro* approach for largazole is published. The enzyme-based synthesis could be a promising strategy for the sustainable production of largazole in cell-free systems (29). Future studies should identify the metabolic gene cluster of largazole to propose a semi-synthetic pathway or propose an enzyme-based synthesis to aid in a more environmentally sustainable method for the production of largazole. A reason for this is that enzymes are efficient catalyst, which can act under mild conditions and can catalyze a lot of reactions with high selectivity (33). Disadvantages of using enzyme catalyst are that these enzymes could be unstable when used for commercial applications like thermostability. Another challenge for the usages of a multi-enzyme synthesis is the large amount of enzyme which is needed and the low concentration of intermediates (34).

6 Conclusion

Cancer will be a more common disease due to the obsolescence of mankind. Therefore, the need for new potential drug candidates is needed. Natural products found in cyanobacteria show to have a considerable potential for new drug candidates such as largazole. This molecule could be assigned to the molecular class of depsipeptides, revealing beneficial bioactivities against many different sorts of cancers. Largazole shows exciting traits such as targeting transformed cells (fibroblastic cancer cells) over non-transformed cells (fibroblastic cells) based on its potential HDAC class I inhibition properties. This review demonstrates a lot on this topic which is currently not known or has to be optimized according to the synthesis of the natural product largazole. Therefore, it can be concluded that to this date, the total synthesis pathway is the most optimal pathway for the synthesis of largazole from *Caldora penicillata*. The most promising strategy for the large-scale production of largazole is the strategy proposed by Luesch et al. Thereby, a yield of 19 %(7.5-22 mg) in 8 linear steps with a protocol for the up-scale process of largazole is indicated. While no publications regarding a semisynthetic approach for the synthesis of largazole could be found, combining chemical with enzymatic synthesis seems highly promising to aid in the total synthesis by synthesizing more precursor molecules in a more selective way. For the culturing and extraction of

largazole from *Caldora penicillata* it showed the problem of a very low yield 0.41 %(1.2 mg) from 290 mg cyanobacteria. Finally, the strategy for *in vitro* synthesis of largazole utilizing the enzymes encoded in the biosynthetic gene cluster also shows a possibility, but the gene cluster responsible for the production of largazole has yet to be published. What is the best strategy to synthesize the natural product largazole derived from cyanobacterial strain *Caldora penicillata* in the laboratory? According to the published literature, this would be the total synthesis of largazole mainly due to the experience and the optimized protocols for the total synthesis.

7 References

1. World Health Organization. Cancer 2020 [cited 2020 23-09]. Available from: https://www.who.int/health-topics/cancer#tab=tab_1.
2. Daniel A. Dias SU, Ute Roessner,. A historical overview of natural products in drug discovery. Metabolites. 2012;2:303-36.
3. Stephanie H.Kung SL, Abhishek Murarka, Derer Mcphee and Chris J.Paddon. Approaches and recent developments for the commercial production of semisynthetic artemisinin. Frontiers in plant science. 2018.
4. Xiaoping Tang MD, Thomas Wirth, Rudolf K.Allemann,. Concise synthesis of artemisinin from a farnesyl diphospate analogue. Bioorganic and medicinal chemistry. 2018;26(7):1314-9.
5. Cragg G.M Pezzuto J.M. Natural Products as a Vital Source for the Discovery of Cancer Chemotherapeutic and Chemopreventive Agents. Medicinal principles and practice 2016;25:41-59.
6. Subramaniyan Vijayakumar MM. Pharmaceutical applicatiokns of cyanobacteria. Journal of acute medicine. 2015;5:15-23.
7. Dewick PM. Medicinal Natural Products a biosynthetic approach. 3 ed: Wiley; 2009. 509 p.
8. Lixia Zeng MC, Xiaozhong Lan and Zhihua Liao,. Metabolic regulation of the artemisinin biosynthetic pathway in Artemisia annua L. Journal of medicininal Plants research. 2012;6:7-
9. Soonkyu Hwang NL, Suhyung Cho, Bernhard Palsson and Byung-Kwan Cho,. Repurposing Modular Polyketide Synthases and Non-ribosomal Peptide Synthetases for Novel Chemical Biosynthesis. frontiers in molecular biosciences. 2020.
10. L.Harvey A. Natural products in drug discovery. drug discovery today. 2008;13(19-20):894-901.
11. Inge Kjærbølling UHM, Tammi Vesth, Mikael R.Andersen,. Strategies to establish the link between biosynthetic gene clusters and secondary metabolites. Fungal Genetics and Biology. 2019;130:107-21.
12. K.C. Nicolaou SR, and Ruocheng Yu,. Total Synthesis Endeavors and Their Contributions to Science and Society: A Personal Account. Chinese chemical society. 2019.
13. Rahul Kunwar Singh SPT, Ashwani K Rai and Tribhuban M Mohapatra. Cyanobacteria: an emerging source for drug discovery. journal of antibiotics. 2011;64:401-12.
14. Vivek K. Bajpai SS, Sung-Min Kang, Seung Kyu Hwang,Xinjie Song, Yun Suk Huh and Young-Kyu Han,. Developments of Cyanobacteria for Nano-Marine Drugs: Relevance of Nanoformulations in Cancer Therapies. Marine drugs. 2018;16(6).
15. Justine Demay CB, Anita Reinhardt and Benjamin Marie,. Natural Products from Cyanobacteria: Focus on BeneficialActivities. Marine drugs. 2019;17(6).
16. Jiyong Hong and Hendrik Luesch. Largazole: From Discovery to Broad-Spectrum Therapy. Natural product reports. 2012;29(4):449-56.
17. Niclas Engene AT, Lilibeth A. Salvador-Reyes, Hendrik Luesch, Valerie J. Paul,. Caldora penicillata gen. nov., comb. nov. (Cyanobacteria), a pantropical marine species with biomedical relevance. journal of phycology. 2015;51:670-81.
18. Albert Bowers NW, Jack Taunton,| Stuart L. Schreiber, James E. Bradner, and Robert M. Williams,. Journal of american chemical society. Total Synthesis and Biological Mode of Action of Largazole: A Potent Class I Histone Deacetylase Inhibitor. 2008;130:11219-22.

19. Omar Khan NBLT. HDAC inhibitors in cancer biology: emerging mechanisms and clinical applications. immunology and cell biology. 2011;90(1).
20. Tomas Eckschlager JP, Marie Stiborova and Jan Hrabeta,. Histone Deacetylase Inhibitors as Anticancer Drugs. international journal of molecular sciences. 2017;18(7).
21. Yongcheng Ying KT, Hyoungsu Kim, Jiyong Hong, and, Hendrik Luesch. Total Synthesis and Molecular Target of Largazole, a Histone Deacetylase Inhibitor. Journal of american chemical society. 2008;130:8455-9.
22. Qi-Yin Chen PRC, and Hendrik Luesch,. Process Development and Scale-up Total Synthesis of Largazole, a Potent Class I Histone Deacetylase Inhibitor. Organic process research and development. 2018;22:190-9.
23. Arun K. Ghosh and Sarang Kulkarni. Enantioselective Total Synthesis of (+)-Largazole, a Potent Inhibitor of Histone Deacetylase. Organic letters. 2008;10(17):3907-9.
24. Yi-Qiang Cheng MY, and Andrea M. Matter,. Characterization of a Gene Cluster Responsible for the Biosynthesis of Anticancer Agent FK228 in Chromobacterium violaceum No. 968. Applied and Environmental Microbiology. 2007;73(11):3460-9.
25. Vinod Kumar AS, Ravdeep Kaur, Ashwani Kumar Thukral, Renu Bhardwaj, Parvaiz Ahmad. Differential distribution of amino acids in plants. Amino Acids. 2017;49:821-69.
26. Yi-Qiang Cheng G-LT, and Ben Shen,. Identification and Localization of the Gene Cluster Encoding Biosynthesis of the Antitumor Macrolactam Leinamycin in Streptomyces atroolivaceus S-140. journal of Bacteriology. 2002;184(24):7013-24.
27. Brian A. Whitton. Ecology of cyanobacteria II. A.Whiton B, editor. Dordrecht, Heidelberg, New York, London: Springer; 2012. 753 p.
28. Kanchan Taori YLD, Valerie J. Paul Dr. , Hendrik Luesch Prof. Dr.,. Combinatorial Strategies by Marine Cyanobacteria: Symplostatin 4, an Antimitotic Natural Dolastatin 10/15 Hybrid that Synergizes with the Coproduced HDAC Inhibitor Largazole. Chemistry and Biochemistry. 2009;10(10):1634-9.
29. Aaron Cravens JP, and Christina D. Smolke,. Synthetic biology strategies for microbial biosynthesis of plant natural products. nature communications. 2019;10(2142).
30. Bumki Kim RR, Hyunji Lee,Guqin Shi, Sabrina L. Zeller, Chenglong Li, Hendrik Luesch, and Jiyong Hong ,. Synthesis and Biological Evaluation of Largazole Zinc-Binding Group Analog. Bioorganic and medicinal chemistry. 2017;25(12):3077-86.
31. Jehad Almaliti AAA-H, Ahmed T Negmeldin , Christin L Hanigan , Lalith Perera , Mary Kay H Pflum , Robert A Casero Jr , L M Viranga Tillekeratne,. Largazole Analogues Embodying Radical Changes in the Depsipeptide Ring: Development of a More Selective and Highly Potent Analogue. Medicinal chemistry. 2016;8(23):10642-60.
32. Sven Janson a AM, Birgitta Bergman,. The presence and expression of hetR in the non-heterocystous cyanobacterium Symploca PCC 8002. Federation of European Microbiolical Societies. 1998;168:173-9.
33. Faber K. Biotransformations in organic chemistry. 7th ed: Springer; 2017.
34. Jung-Min Choi S-S, Hak-Sung Kim. Industrial applications of enzyme biocatalysis: Current status and future aspects. biotechnology advances 2015;33(7):1453-4.