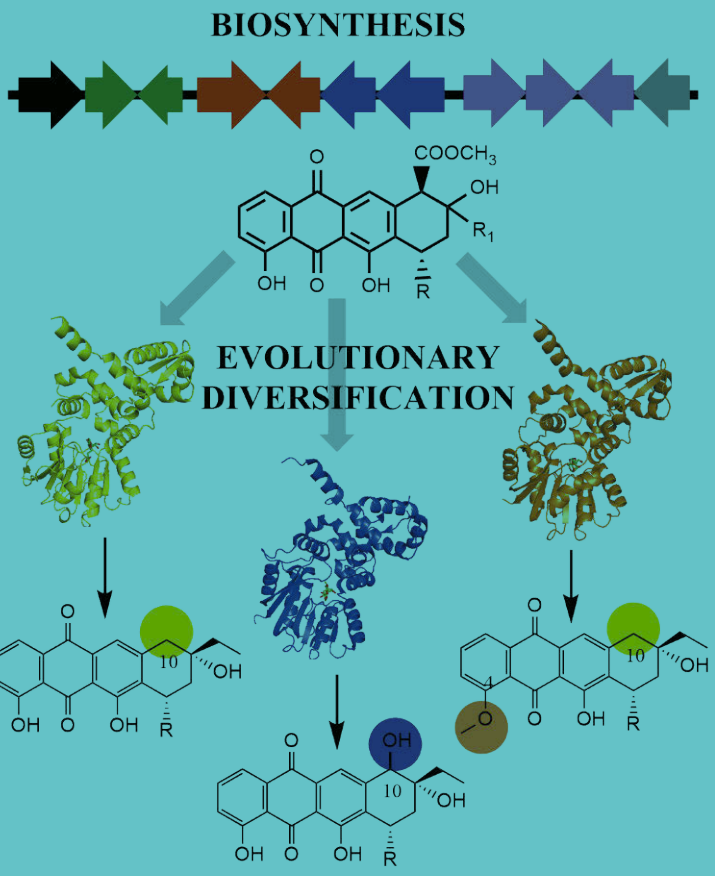
TURUN
YLIOPISTO
UNIVERSITY
OF TURKU



BIOSYNTHESIS AND EVOLUTION OF ANTHRACYCLINES

Benjamin Nji Wandi

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 649 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2021

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

BIOSYNTHESIS AND EVOLUTION OF ANTHRACYCLINES

Benjamin Nji Wandi

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 649 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2021

University of Turku

Faculty of Technology
Department of Life Technologies
Biochemistry
Doctoral Program in Molecular Life Science

Supervised by

Professor Mikko Metsä-Ketelä, Ph.D.
Department of Life Technologies
University of Turku
Finland

Reviewed by

Rajaram Venkatesan, Ph.D., Docent
Faculty of Biochemistry and Molecular Medicine
University of Oulu
Finland

Silvan Scheller, Assistant Professor, Biochemistry
Department of Bioproducts and Biosystems
Aalto University
Finland

Opponent

Steven Van Lanen, Professor
Department of Pharmaceutical Sciences
University of Kentucky
United States of America

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Cover image: Benjamin Nji Wandi

ISBN 978-951-29-8538-8 (PRINT)
ISBN 978-951-29-8539-5 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)
Painosalama, Turku, Finland 2021

UNIVERSITY OF TURKU
Faculty of Technology
Department of Life Technologies
Biochemistry
BENJAMIN NJI WANDI: Biosynthesis and Evolution of Anthracyclines
Doctoral Dissertation, 173 pp.
Doctoral Program in Molecular Life Science
August 2021

ABSTRACT

Natural products are a rich source of medication leads, accounting for about two-thirds of all antibiotics and one-third of all anticancer drugs currently in use. Anthracyclines, which belong to the type II family of aromatic polyketides, are a medically essential class of natural products. They have one or more deoxysugar moieties attached at C–7 via *O*–glycosylation and are made up of the tetracyclic 7,8,9,10-tetrahydro-5,12-naphtacenoquinone aglycone chromophore. Anthracyclines are produced by soil-dwelling Gram-positive Actinobacteria. They are cytotoxic, but also cardiotoxic, which limits their clinical use. Minor changes in anthracyclines can reduce or eliminate their unwanted side-effects, while preserving their cytotoxicity.

The literature review covers the biosynthesis of anthracyclines from the starting unit until the modification of the common aglycone scaffold. Furthermore, it depicts evolutionary events that have occurred during diversification of anthracycline biosynthetic pathways with special focus on the so-called tailoring steps. Finally, it explains the use of chimeragenesis as a protein engineering technique to trace the evolutionary events and to generate novel catalysts.

This thesis focused on the biosynthesis of nogalamycin and elucidated the individual steps in the biosynthesis of L-rhodosamine using seven different enzymes. Furthermore, L-rhodosamine was attached to an anthracycline aglycone acceptor with a dual linkage system using the glycosyltransferase SnogD and the α-ketoglutarate–dependent monooxygenase SnoK. My work uncovered the 2"–hydroxylation function of the Rieske enzyme SnoT, which is a critical post-modification essential for nogalamycin bioactivity. This study also established the order of late-stage modification reactions with 2"–hydroxylation as the first step, followed by the 2–5" carbocyclization and 4"–epimerization by SnoT, SnoK and SnoN, respectively.

My thesis exploited chimeragenesis to uncover the evolutionary paths of anthracycline methyltransferases and α-ketoglutarate-dependent monooxygenases. The work elucidated how the 4" epimerase SnoN has evolved via gene duplication and functional differentiation of the 2–5" carbocyclase SnoK. In contrast, the 10–decarboxylase TamK and the 10–hydroxylase RdmB have evolved without a gene duplication event from the 4–*O*–methyltransferase DnrK. This study explains how the increase of anthracycline diversity may have occurred through subtle changes in the sequences of biosynthetic enzymes.

TURUN YLIOPISTO
Teknillinen tiedekunta
Bioteknologian laitos
Biokemia
BENJAMIN NJI WANDI: Biosynthesis and Evolution of Anthracyclines
Väitöskirja, 173 s.
Molekulaaristen biotieteiden tohtoriohjelma (DPMLS)
Elokuu 2021

TIIVISTELMÄ

Luonnonyhdisteet ovat lääkeaineaihioiden rikas lähde, joiden avulla on kehitetty noin kaksi kolmasosaa kliinisessä käytössä olevista antibiooteista ja kolmasosa syöpälääkkeistä. Antrasykliinit ovat lääkekehityksen kannalta tärkeä luonnonyhdisteiden luokka, mitkä kuuluvat tyypin II aromaattisiin polyketideihin. Antrasykliinejä tuottavat maaperässä elävät Gram-positiiviset Aktinobakteerit. Yhdisteet ovat sytotoksisia, mutta myös sydäntoksisia, mikä rajoittaa niiden kliinistä käyttöä. Pienet muutokset antrasykliinien rakenteissä voivat kuitenkin vähentää tai poistaa ei-toivottuja haittavaikutuksia, samalla kuitenkin säilyttäen sytotoksisuuden.

Kirjallisuuskatsaus kuvaa antrasykliinien biosynteesin yhteisen antrasykliini-hiilirungon muodostumiseen saakka. Tämän lisäksi tutkielma kuvaa antrasykliinien evoluutiota ja monimuotoisuuden syntymistä keskittyen erityisesti niin sanottuihin muokkausreaktioihin. Lopuksi tutkielmassa käydään läpi kuinka proteiinien muokkaukseen käytettävää kimerageneesi-menetelmää voidaan käyttää proteiini-evoluutiossa tapahtuvien muutosten tutkimiseen ja uusien proteiinikatalyyttien kehittämiseen.

Tutkielmani keskittyy nogalamysiinin biosynteesiin, missä selvitin kuinka L-rhodosamiini sokeri muodostuun seitsemän entsyymin toimesta. Tämän lisäksi liitimme L-rhodosamiinin antrasykliiniin kaksinkertaisella sidoksella käyttäen glykosyylitransferaasia SnogD ja α-ketoglutaraattista riippuvaista oksygenaasia SnoK. Työni paljasti, että nogalamysiinin bioaktiivisuuden kannalta kriittisen 2" hydroksylaatioryhmän asettaa paikalle Rieske entsyymi SnoT. Tutkielmani selvitti myös myöhäisen vaiheen muokkausreaktion järjestyksen, missä ensin tapahtuu SnoT proteiinin katalysoima 2" hydroksylaatio, minkä jälkeen SnoK entsyymi katalysoi 2–5" karbosyklisaatiota ja SnoN proteiini 4" epimerisaatiota.

Käytin kimerageneesi-menetelmää selvittääkseni eri antrasykliinien biosynteesiin liittyvien metyylitransferaasien ja α-ketoglutaraattista riippuvaisten oksygenaasien evoluutiota. Työni selvitti kuinka 4" epimeraasi SnoN on muodostunut muinaisen 2–5" karbosyklaasin SnoK tuotosta vastaavan geenin kahdentumisen ja funktionaalisen erilaistumisen seurauksena. Tämä on erilainen kuin mitä on tapahtunut 10–dekarboksylaasin TamK ja 10–hydroksylaasin RdmB kohdalla, missä proteiinien evoluution on tapahtunut 4–*O*–metyylitransferaasi DnrK proteiinista ilman geenien kahdentumista. Työni kuvaa kuinka hyvin pienet muutokset biosynteesistä vastuussa olevissa proteiineissa voivat johtaa antrasykliinien monimuotoisuuden kasvuun.

Table of Contents

Abbreviations of Amino Acid Residues

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Abbreviations

1/2/3D	One/two/three-dimensional
AAME	Aklanonic acid methyl ester
ACP	Acyl carrier protein
ARO	Aromatase
ATP	Adenosine triphosphate
CLF	chain length factor
CO_2	Carbon dioxide
CoA	Coenzyme A
COSY	Correlation spectroscopy
CYC	Cyclase
dTDP	deoxythymidine diphosphate
dUDP	2´-deoxyuridine-5´-diphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESI-MS	Electrospray ionization mass spectrometry
FAS	Fatty acid synthase
Fe^{2+}/ Fe(II)	Iron (2+)
GT	Glycosyltransferase
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single-quantum correlation spectroscopy
HSQCDE	Heteronuclear single-quantum and distortionless enhancement correlation
KR	Ketoreductases
KS	Ketosynthase
L-Glu	L-glutamate
MCAT	malonyltransferase
minPKS	Minimal polyketide synthase
NAME	Nogalonic acid methyl ester
NAD^+	Nicotinamide adenine dinucleotide, oxidized
NADH	Nicotinamide adenine dinucleotide, reduced
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced

NDP	Nucleoside diphosphate
NMP	Nucleoside monophosphate
NTP	Nucleoside triphosphate
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
OD_{600}	Optical density at 600 nm
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PKS	Polyketide synthase
PLP	Pyridoxal 5´–phosphate
R.M.S.D	Root mean square deviation
SAH	*S*–adenosyl-L-homocysteine
SAM	*S*–adenosyl methionine
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TDP	Thymidine diphosphate
TLC	Thin layer chromatography
TOCSY	Total correlation spectroscopy
UV/Vis	Ultraviolet/visible light
WT	Wild type
α-KG	Alpha-ketoglutarate/alpha-oxoglutarate

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Siitonen, V., **Nji Wandi, B.**, Törmänen, A.-P., and Metsä-Ketelä, M. (**2018**) Enzymatic Synthesis of the C-Glycosidic Moiety of Nogalamycin R. ACS Chemical Biology*,* **13**, pg 2433–2437.

II **Benjamin Nji Wandi.**, Vilja Siitonen, Kaisa Palmu and Mikko Metsä-Ketelä. (**2020**) The Rieske Oxygenase SnoT Catalyzes 2"–Hydroxylation of l–Rhodosamine in Nogalamycin Biosynthesis. *ChemBioChem*, **21**, pg 3062–3066.

III **Nji Wandi, B.**, Siitonen, V., Dinis, P., Vukic, V., Salminen, T.A., and Metsä-Ketelä, M. (**2019**) Evolution–Guided Engineering of Non–Heme Iron Enzymes Involved in Nogalamycin Biosynthesis. *The FEBS Journal,* **287**, pg 2998–3011.

IV Pedro Dinis,[1] Heli Tirkkonen,[1] **Benjamin Nji Wandi**, Vilja Siitonen, Jarmo Niemi, Thadée Grocholski and Mikko Metsä-Ketelä. (**2021**) Evolution–Inspired Engineering of Anthracycline Methyltransferases. *Nature Chemical Biology* (Submitted).
[1] – equal contribution

1 Introduction

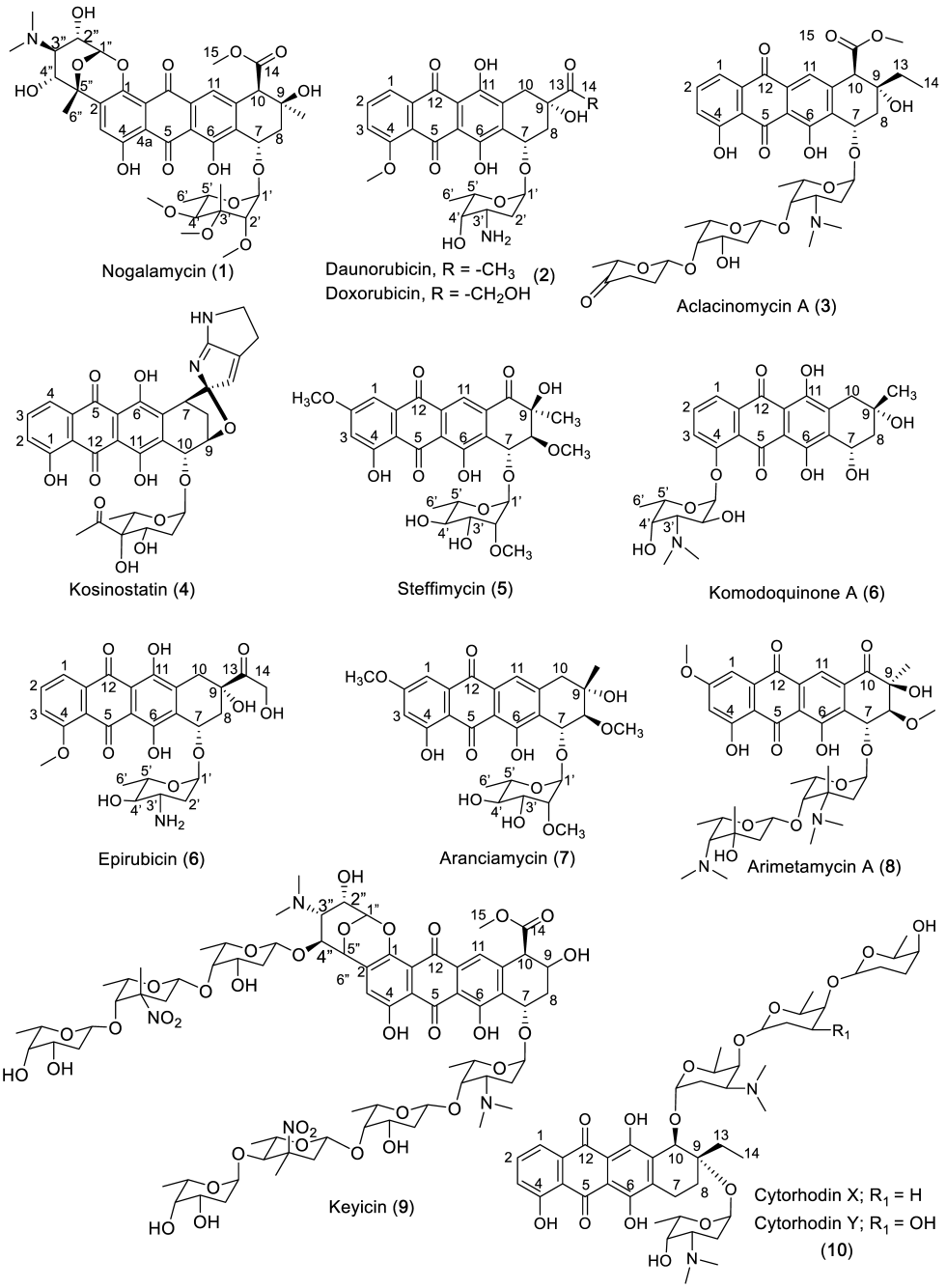
Microbial natural products and semi-synthetic derivatives are a rich source of medication leads, accounting for about two-thirds of antibiotics (Newman & Cragg, 2012) and one-third of anticancer agents (Newman & Cragg, 2016) in clinical use over the last 35 years. Anthracyclines, which belong to the type II family of aromatic polyketides, are a medically essential class of natural products (Metsä-Ketelä *et al*., 2008; Hertweck *et al*., 2007; Hutchinson, 1997). Soil-dwelling Gram-positive Actinobacteria produce these metabolites, and at least 408 bacterial-derived anthracyclines have been identified to date (Elshahawi *et al*., 2015). Anthracyclines are among the most popular anticancer drugs currently available, with compounds like doxorubicin being studied in more than 500 clinical trials around the world (Pang *et al*., 2013). Anthracyclines and their derivatives are known to exert their cytotoxic effect through DNA intercalation and poisoning of topoisomerases II (Nitiss, 2009). This class of compounds have carbohydrate units typically positioned at C-7 (scheme 1), which are known to improve the bioactivities of anthracyclines by facilitating their binding to DNA (Frederick *et al*., 1990; Temperini *et al*., 2005).

Minor structural changes to these anthracyclines may have a major effect on their bioactivity. This encouraged natural product chemists early on in the 1970s to generate semi-synthetic derivatives with improved bioactivity while maintaining low levels of toxicity (Woodward R.B. *et al*., 1981; Chu D.T.W, 1995). The elucidation of the biosynthetic logic of anthracyclines in the 2000s has opened up possibilities for metabolic engineering to generate biosynthetic derivatives for the production of novel anthracyclines (Metsä-Ketelä *et al*., 2008). An attractive approach has been to introduce genes from related pathways in anthracycline producing strains to generate hybrid metabolites that share features of the naturally occurring compounds. The strategy depends on the ability of the foreign genetic material to interact with the existing metabolic pathways, and has been most successfully utilized in the so-called tailoring and glycosylation steps of anthracyclines. However, many promising gene combinations have been nonfunctional (Kulowski *et al*., 1999; Wohlert, 2001), which has hindered the expansion of such a trial-and-error approach. For these reasons considerable effort has been made in the determination of 3D-structures of

the proteins involved in anthracycline biosynthesis in the past decade to obtain a more thorough understanding of the factors that govern substrate recognition and catalysis (Schneider, 2005; Wong & Khosla, 2012). This has enabled scientists to comprehend the possibilities and limitations of the biosynthetic combinatorial approaches. It has also paved the way for other promising applications, such as site-directed and directed evolution mutagenesis, which have aided in the diversification of natural products.

Anthracyclines and fatty acids appear to share a common evolutionary origin. Previous study have suggested that the polyketide synthase (PKS), which produces secondary metabolites (anthracyclines) evolved from the fatty acid synthase (FAS) that is known to produce primary metabolites (fatty acids). Many similarities exist with the enzymes involved in the biosynthesis of both anthracyclines and saturated fatty acids (Smith S. & Tsai S.C., 2007). For example, both use common precursors, similar chemistry and have similar architectural design. In spite of their similarities, FAS is involved in primary metabolism solely and committed in producing entirely saturated fatty acids whereas the PKS remarkably possess the flexibility that allows adaptation to evolutionary pressure to produce a broad range of structurally complex natural products (Smith S. & Tsai S.C., 2007). Over time, evolutionary pressure on *Streptomyces* bacteria has led to the appearance of new biosynthetic pathways, which have contributed greatly to the diversity of the compounds produced by these bacteria (Fischbach *et al*., 2008). These diverse compounds provide significant advantages to these bacteria over other organisms in their competitive environment. This evolutionary pressure has led to the appearance of closely structurally related proteins in anthracycline biosynthetic pathways or among related pathways that may have completely distinct catalytic properties from each other. For instance, methyl transferases acting as mono-oxygenases and mono-oxygenases acting as cyclases (Siitonen *et* al., 2016; Siitonen *et al*., 2012a; Jansson, 2003).

The enzymatic biosynthesis of some common anthracyclines (aclacinomycin, doxorubicin, daunorubicin, nogalamycin, elloramycin, keyicin and tetracenomycin) is described in detail below with emphasis on nogalamycin and aclacinomycin. In addition, the evolutionary event of gene duplication that have contributed greatly to the chemical diversity of secondary metabolites is addressed later.



Scheme 1: Chemical structures of selected anthracyclines.

1.1 Biosynthesis of anthracyclines

1.1.1 Biosynthesis of anthracyclinone carbon scaffolds aklavinone and nogalavinone

Anthracyclines have a tetracyclic 7,8,9,10-tetrahydro-5,12-naphtacenoquinone aglycone chromophore with one or more deoxysugar moieties attached at C-7 through *O*-glycosylation (Brockmann & Brockmann, 1963) (scheme 1). The enzymatic synthesis of the anthracycline core is quite similar for most of the anthracyclines and the structural variation observed on different pathways is generated during later biosynthetic stages. The biosynthesis begins with a sequence of Claisen condensation reactions catalyzed by the minimal polyketide synthase (minPKS) (Dickens *et al*., 1997; Hutchinson, 1997; Kantola *et* al., 2000; *Alexeev et* al., 2007; Dreier & Khosla, 2000; Katz & Donadio, 1993; Khosla *et* al., 1999; McDaniel *et al*., 1993). This minPKS is made up of four different subunits; an acyl carrier protein (ACP), which carries the growing polyketide chain; a malonyl-CoA:ACP malonyltransferase (MCAT), which malonylates ACP; a heterodimeric core of the complex of homologous ketosynthase (KS) and chain length factor (CLF) subunits. The KS (also referred to as the KS–alpha subunit) catalyzes the chain initiation and elongation reactions while the CLF (also known as the KS–beta subunit) determines the length of the polyketide chain (*McDaniel et* al., 1993; *Tang et* al., 2003).

1.1.2 Biosynthesis of aklanonic and nogalonic acids

The biosynthesis of the ketoacyl chain starts with the malonylation of ACP by MCAT, which attaches the malonate moiety to ACP through a phosphopantetheine arm (Keatinge-*Clay et* al., 2003; Florova *et* al., 2002). When the malonyl-ACP encounters the KS-CLF heterodimer, the biosynthesis of the polyketide chain is initiated (Figure 1). Like all previously studied MCAT enzymes, the *act* MCAT (malonyl-CoA:ACP malonyltransferase from the actinorhodin pathway) contains a catalytic serine nucleophile (Ser97) in its active site, which have been shown to be responsible for the transfer of malonate to ACP (Szafranska *et al*., 2001; Joshi *et* al., 1998; Gokhale *et al*., 1998; Kremer *et* al., 2001). The malonyl group undergoes decarboxylation and the resulting acetyl moiety is transferred to the active site of the KS that contains the catalytically active cysteine residue needed for the Claisen-type condensation reaction. The crystal structure of the priming ketosynthase (KS) has highlighted the importance of this conserved cysteine in catalysis, whose nucleophilic attack is enhanced by a proximal histidine residue. The *act* KS (ketosynthases from actinorhodin pathway) contains four amino acid residues that

are conserved in all known ketosynthases. These residues; Cys169, His309, Lys341 and His346 play an important role in the decarboxylative priming of KS and polyketide chain extension (Figure 2A and 2B). Before the decarboxylation, the Cys169 and His346 form a catalytic dyad, where the acyl chain is attached, whereas His309 plays a role in positioning the malonyl-ACP and facilitate the formation of carbanion by interacting with the thioester carbonyl. Lys341 functions in enhancing the rate of decarboxylation of malonyl-ACP through electrostatic interaction. It has been proposed that after every C–C bond formation, the KS dissociate from the ACP and prior to this dissociation, the newly extended acyl chain is transferred back from the ACP pantetheine to the Cys169 of the KS (Dreier & Khosla, 2000). In addition, structural studies have shown that the architecture of the acyl group binding pocket determines the substrate specificity of this enzyme (*Davies et* al., 2000; Pan *et al*., 2002). The polyketide chain is repeatedly elongated by a series of extension cycles with malonyl units until the poly-β-ketoacyl chain reaches the desired length (Figure 1, step 1). The biosynthesis of aclacinomycin, doxorubicin and daunorubicin carrying an ethyl side chain at C-9 (Scheme 1) begins from a longer propionate starter unit (Figure 1).

The ability of the ACP to interact with unrelated enzymes in this pathway is one of the most fascinating aspect about this enzyme. First, the phosphopantetheine prosthetic group is attached to apo-ACP by a holo-ACP synthase, which then activates the protein (Parris *et* al., 2000; Lambalot *et al*., 1996). Secondly, MCAT transfers the malonyl group from malonyl-CoA to ACP and, thirdly, the ACP encounters and interacts with the KS-CLF heterodimer. Crystallography studies on the *act apo*-ACP reveals a hydrophobic core and buried amino acids with hydrophilic groups, mainly Arg72 and Asn79, which are proposed to play a role in catalysis (Crump *et al*., 1997). The side chains of arginine are usually known to be involved in protein-ligand interactions and in hydrogen bonding to separate acceptors in a bidentate arrangement (Mrabet *et al*., 1992; Borders *et al*., 1994). These two amino acids are conserved in all the PKS ACPs and may perhaps be accountable for the stabilization of the growing polyketide chain.

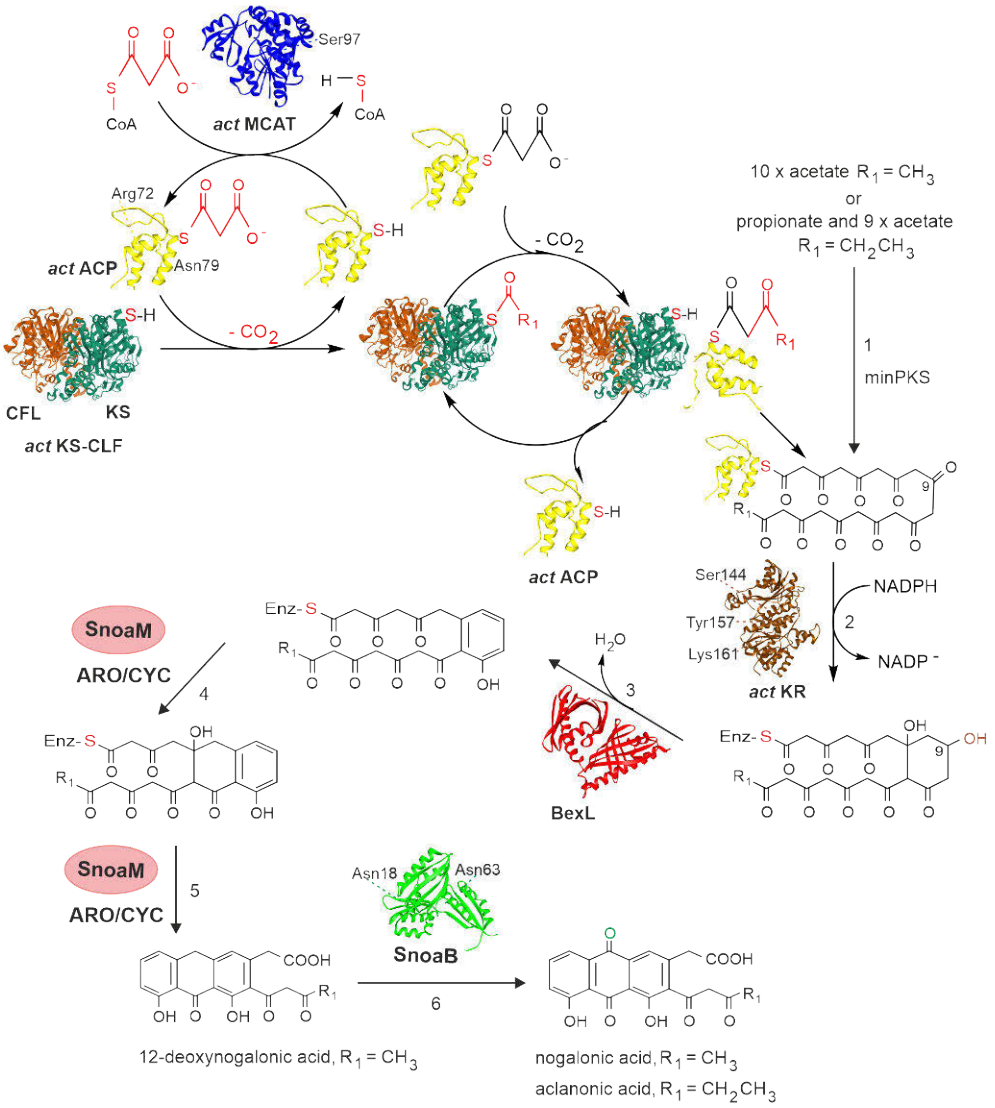


Figure 1: Illustration of the initial steps involved in the biosynthetic pathway of type II polyketides. The PDB code of the enzymes shown are: *act* KS-CLF (orange/dark green), 1TQY; *act* ACP (yellow), 2AF8; *act* MCAT (violet), 1NM2; *act* KR (brown), 1X7H; BexL (red), 4XRW; SnoaB (green), 3KG0.

The minPKS complex produces a labile poly-β-keto intermediate as its final product. Unless the biosynthetic enzymes required for the next step are present, this unstable intermediate is known to undergo several cyclization events generating a range of differently folded polyketide shunt products. The presence of other specialized enzymes in this pathway such as Ketoreductases (KR), cyclases (CYC) and

aromatases (ARO) guide the unstable intermediate to fold in the correct pattern to form a stable polyaromatic intermediate (Hertweck *et al*., 2007; Hopwood *et al*., 1997). Studying the *in vitro* structure/function relationship of these enzymes have been quite challenging due to the lack of stable intermediates. However, more specific knowledge has been gained from the study of compounds formed by blocked mutants and molecular genetic research. After the complete biosynthesis of the polyketide chain and cyclization of the first ring between C–7 and C–12 (numbered from the enzyme-bound carboxyl termini), the keto group at C–9 position is reduced by a ketoreductase (Figure 1, step 2). This ketoreductase belongs to the short–chain alcohol dehydrogenase superfamily of enzymes that contain the characteristic catalytic Ser–Tyr–Lys triad and require NAD(P)H as a cofactor for the reduction of the keto group (Bartel *et al*., 1990; Korman *et al*., 2004; Hadfield *et al*., 2004). The minPKS and the KR appear to be responsible for the correct regiochemistry of the cyclization of this first ring. Structural studies of *act* KS–CLF (from the actinorhodin pathway, which is not an anthracycline but belongs to the same family of aromatic type II polyketides) show that the only way a full length octaketide can fit into the substrate binding tunnel is by positioning C–7 adjacent to C–12, hence favoring the C7–C12 cyclization event (Keatinge–Clay *et al*., 2004). In the absence of the downstream KR, misfolded polyketides have been detected, which indicates that the KR plays a role in the regiospecificity of the first ring cyclization (Kantola *et al*., 1997; Kunnari *et al*., 1999). This enzyme–catalyzed ketoreduction occur in almost all anthracycline biosynthetic pathways as the majority of them belong to the class of reduced polyketides. The active site of these ketoreductases contain the typical catalytic tetrad (Asn-Ser-Tyr-Lys) as revealed by the crystal structure of actinorhodin ketoreductase (*act* KR) (Korman *et al*., 2004; Hadfield *et al*., 2004). The reaction mechanism involves this tetrad, NADPH and four water molecules, wherein both Ser144 and Tyr157 form hydrogen bonds with the ketone substrate. Tyr157 together with Lys161 are shown to form additional hydrogen bonds with the ribose and nicotinamide ring of NADPH. The crystalline water molecules that are involved in the proton-relay network are positioned via hydrogen bond by an Asn114 residue (Figure 2C). This network is similar to the proton-relay water network observed in *E. coli* FabG-$NADP^+$ (Price *et al*., 2004). The next step is a hydride transfer from NADPH to the substrate resulting in the formation of an alkoxide. Ser144 and Tyr157 stabilizes the alkoxide, while the tyrosyl proton is being transferred to it. The proton-relay water network replenishes the extracted proton. This proton-relay water network sequentially includes the 2–OH of NADPH ribose and the N–H of Lys161, followed by the four water molecules (Figure 2C).

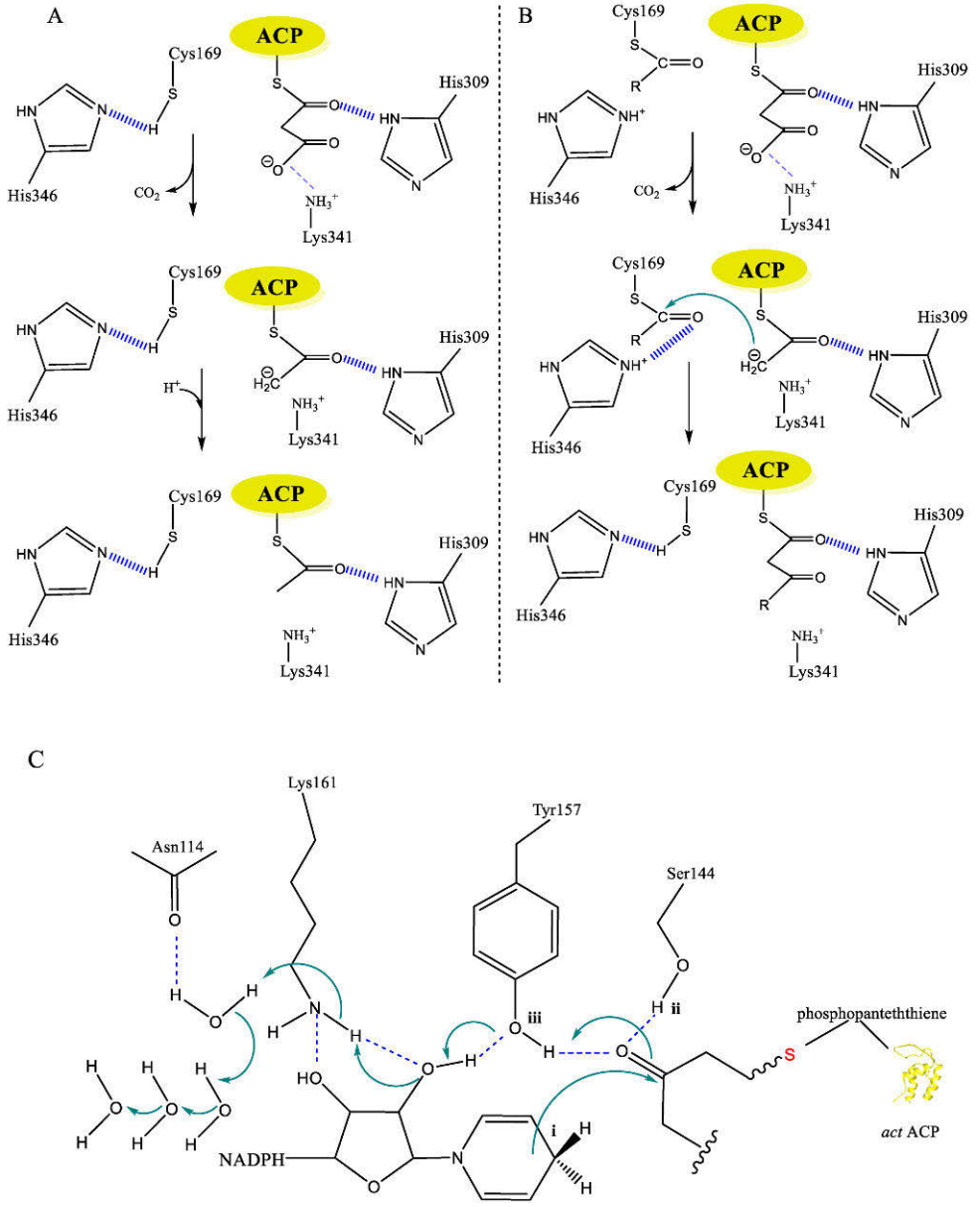
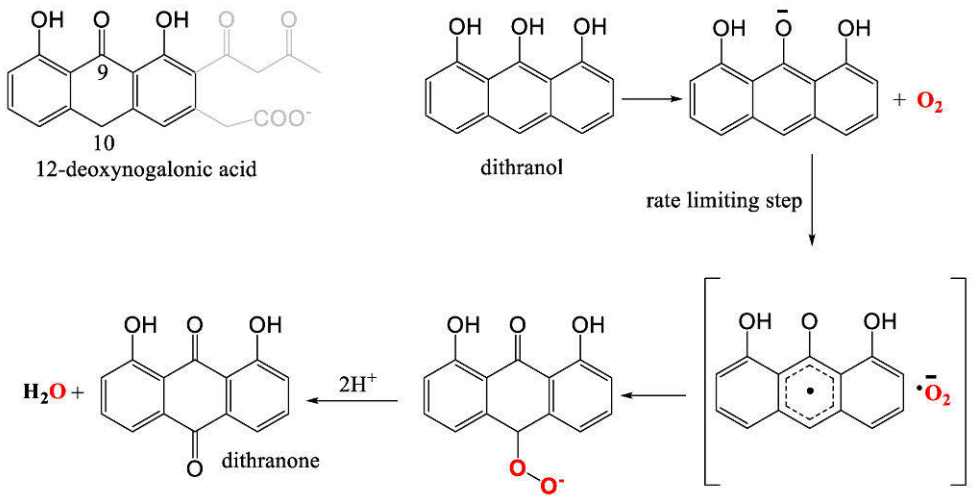


Figure 2: The *act* KS and *act* KR reaction mechanism. A) *act* KS decarboxylative priming and B) polyketide chain extension via carbon-carbon bond formation as suggested by Dreier and Khosla (Dreier & Khosla, 2000). C) Presentation of the C-9 ketoreduction by *act* KR via the catalytic tetrad Asn114–Ser144–Tyr157–Lys161, NADPH and four water molecules. The docking points (**i, ii** and **iii**) are labeled in the order of the proton-relay event (Korman *et al*., 2004).

After the ketoreduction, gene products of *bexL*, *dpsF*, *snoaE* and *aknE1* carry out aromatization of the first ring (Caldara-Festin *et al*., 2015; Grimm *et al*., 1994; Ylihonko *et al*., 1996; Räty *et al*., 2002). Bioconversional studies with BexL have illustrated that the hydroxyl group is eliminated as a water molecule upon aromatization (Figure 1, step 3). The next step is the cyclization of the second and third ring (Figure 1, step 4 and 5) by the gene products of *snoaM, dpsY* and *aknW* (Yliohonko *et al*., 1996; Lomovskaya *et al*., 1998; Räty *et al*., 2000). Up to date, there have been no intermediates identified with only two cyclized rings, indicating that these cyclases are responsible for cyclization of both the second and third rings in anthracycline biosynthesis. (Hautala *et al*., 2003).

The formation of the first stable intermediates (nogalonic and aklanonic acids) is catalyzed by monooxygenases, which oxidize the C-12 position (Figure 1, step 6). SnoaB and DnrG (also referred to as DauG) from the nogalamycin and doxorubicin pathways, respectively, have been identified to be responsible for this oxidation reaction (Ylihonko *et al*., 1996; Grimm *et al*., 1994; Ye *et al*., 1994). Homologs of these monooxygenases from the aclacinomycin pathway (AknX) (Chung *et al*., 2002) and the *act* pathway (ActVA-orf6) (Kendrew *et al*., 1997) have been shown *in vitro* to catalyze this oxidation reaction without the assistance of any cofactors, metals or prosthetic groups commonly needed for the activation of molecular oxygen. They belong to the mechanistically and structurally diverse family of cofactorless oxygenases (Fetzner *et al*., 2002). Structural studies of the first cofactorless monooxygenase ActVA-orf6 revealed a ferrodoxin-like fold (Sciara *et al*., 2003) and the activation of molecular oxygen (a key feature of the reaction) is carried out by the substrate itself. The mechanism is similar to the reaction of FAD/FMN with oxygen in many flavoenzymes. The protein environment of these cofactor-independent monooxygenases facilitates direct electron transfer, while the enzymes accelerate rate-limiting electron transfer by strongly lowering the reaction barrier (Melodie *et al*., 2019; Emerald *et al*., 2019) (Scheme 2). Since the substrate has a very weak C–H bond that can be activated by a weak oxidant, nogalamycin monooxygenase does not require a metal cofactor for oxidizing its substrate (Cantú *et al*., 2018).



Scheme 2: Cofactor-independent monooxygenation as proposed by Melodie M.M. *et al.*, 2019 (Figure 1, step 6). Dithranol is a 12–deoxynogalonic acid substrate analog that transfers an electron to oxygen to produce a dithranyl radical and superoxide pair. The radicals recombine to form a peroxy adduct, which breaks down into dithranone and water.

1.1.3 Biosynthesis of aklavinone and nogalavinone

The aglycone is the key intermediate of all anthracycline pathways on which enzymatic modifications are carried out. Aklavinone and nogalavinone are the aglycones of the aclacinomycin and nogalamycin pathways, respectively (Figure 3A, red box). After the formation of the aklanonic and nogalonic acids, the biosynthesis proceed with methylation of the carboxyl group by *S*-adenosyl-L-methione (SAM)-dependent methyl transferase, forming an aklanonic acid methyl ester (methyl aklanonate) (Figure 3, step 7). This bioconversion has been studied using DnrC from the daunorubicin pathway of *Streptomyces peucetius* (Madduri *et al*., 1995; Dickens *et al*., 1995).

The stereospecific cyclization of the fourth ring follows. This cyclization event is catalyzed by ester cyclases like SnoaL and AknH (Madduri & Hutchinson, 1995; Kallio *et al*., 2006; Sultana *et al*., 2004; Torkkell *et al*., 2000; Kendrew *et al*., 1999) that determine the 9*S* and 9*R* stereochemistry observed in nogalamycin and aclacinomycin, respectively. Structural studies of SnoaL and AknH show that these cyclases carry out an aldol condensation that does not employ Schiff-base formation as in class I aldolases or metal ions as in class II aldolases (Sultana *et al*., 2004; Kallio *et al*., 2006) (Figure 3, step 8). The key step in the reaction mechanism of these cyclases is the abstraction of a C–H proton, forming an enzyme stabilized enolate. In class I aldolases, this enolate is stabilized through covalent catalysis

involving a Schiff base (Gefflaut *et al*., 1995), whereas in class II aldolases the enolate is stabilized by a metal ion (Morse & Horecker, 1968).

Unlike the class I and II aldolases, the anthracycline cyclases have been shown to use a different mechanistic approach for the stabilization of the enolate (Sutana *et al*., 2004; Kendrew *et al*., 1999). They employ an acid-base chemistry to catalyze the intramolecular aldol condensation. In nogalamycin biosynthetic pathway, the cyclase SnoaL contains Asp121 in the active site that abstracts the C–H proton of C-10 of the substrate (Figure 3B, step 1). The enolate formed is stabilized partly by hydrogen bond formed between the carbonyl oxygen at C-14 and the Gln105, and perhaps more significantly by the delocalization of the electron pair over the extended π system of the nearby ring systems. At this point, the cyclization event proceeds with the nucleophilic attack of the carbonyl carbon at C-9 by the stabilized enolate intermediate (Figure 3B, step 2) producing a negative charge at the carbonyl oxygen. The Asp121 in its protonated state stabilizes the negatively charged carbonyl oxygen (Figure 2B, step 3) and the transfer of the proton results in the formation of nogalaviketone.

At this stage of the biosynthetic pathway, an NAD(P)H-dependent ketoreductase (KR) performs an important ketoreduction at C-7. All anthracycline gene clusters have been shown to encode the C-7 KR and this ketoreduction (Figure 3A, step 9) is required prior to the attachment of the deoxysugar at C-7 (Dickens *et al*., 1996; Kantola *et al*., 2000).

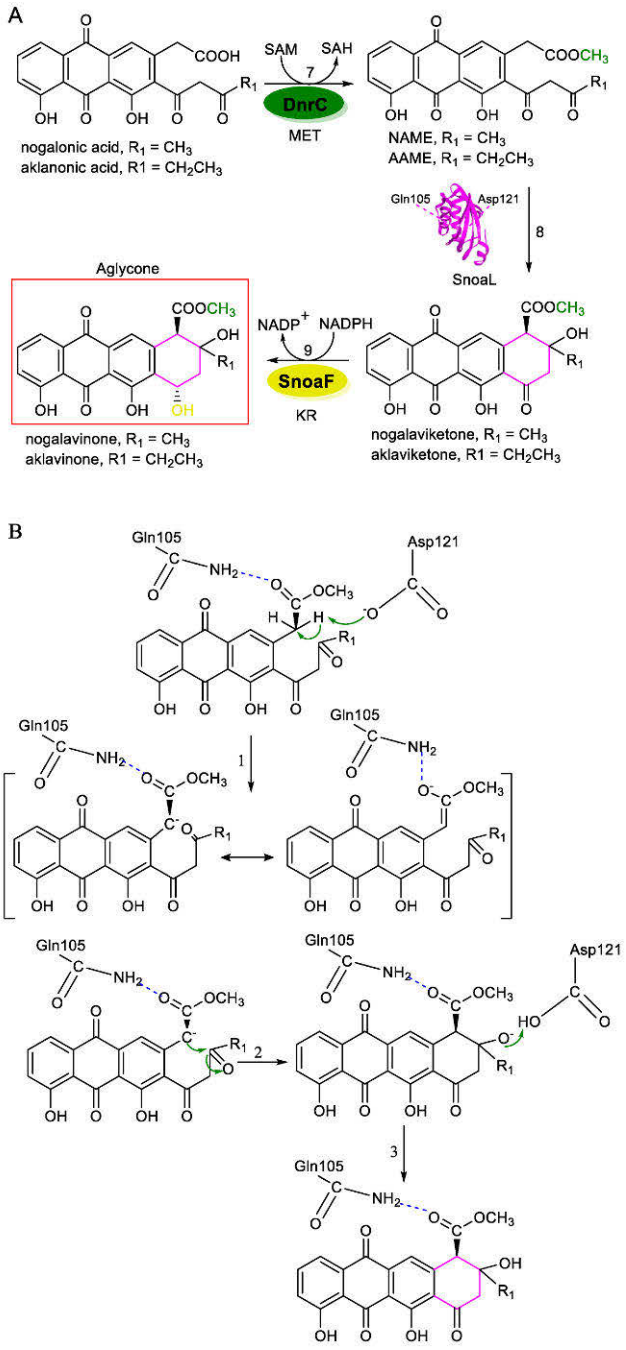


Figure 3: Formation of the aglycone (aklavinone and nogalavinone). A) Simplified model for the biosynthesis of the key intermediate of anthracycline pathway, sequentially involving a methyltransferase, a cyclase and a ketoreductase. The PDB code of SnoaL (purple) is 1SJW. B) Mechanistic illustration of the fourth ring cyclization by SnoaL as proposed by Sultana A., et al. (Sultana *et al*., 2004).

1.1.4 Diversification of anthracyclines

The majority of anthracyclines have a similar biosynthetic logic until the step where the common aglycone is been modified by tailoring enzymes (scheme 1). The biosynthetic gene clusters encode conserved enzymes for the synthesis of the aglycone scaffold (Figure 1 & 3A), but also have diverse sets of accessory tailoring enzymes that are not generally conserved (Fischbach M.A *et al*., 2008; Walsh C.T. *et al*., 2001; Metsä-Ketelä M. *et al*., 2008; Jensen P.R. *et al*., 2016). This results in complex secondary metabolic pathways, involving proteins that can catalyze complex sequential chemical reactions that are stereospecific and regioselective (Malik V.S., 1980; Robbins T. *et al*., 2016; Metsä-Ketelä M. *et al*., 2008; Rabe P. *et al*., 2018; Herr C.Q. & Hausinger R.P., 2018). The chemical diversity of anthracyclines as shown in scheme 1 is greatly promoted via the use of unusual dedicated precursor substrates by the biosynthetic enzymes and by the tailoring reactions catalyzed.

The aglycone moieties, aklavinone and nogalamycinone are the key intermediates of aclacinomycin and nogalamycin biosynthetic pathways respectively, from which the end products are obtained following post-polyketide modifications. The great diversity among these anthracyclines is generated by chemical modifications such as hydroxylations, methylations, decarboxylations and glycosylations, that modifies the typical 7,8,9,10–tetrahydro–5,12–naphthacenoquinone aglycone chromophore (Nji Wandi *et al*., 2020; Siitonen *et al*., 2018; Metsä-Ketelä *et al*., 2008; Hertweck *et al*., 2007; Hutchinson *et al*., 1997).

In the keyicin (**9**) and nogalamycin (**1**) pathways, the aglycone moiety is not only hydroxylated at the C-7 position (like in most anthracyclines), but also at the C-1 position. Investigations of the nogalamycin pathway revealed that a two-component monooxygenase SnoaW/SnoaL2 is accountable for the installation of the C-1 hydroxyl group (Siitonen *et al*., 2012a). The attachment of L-nogalose at C-7 through a canonical *O*–glycosidic linkage by the gene product of *snogE* precede this unique C-1 hydroxylation (Figure 4). The C-1 hydroxylation is an critical step in the biosynthesis of nogalamycin because the C-1 hydroxyl group is needed for the attachment of the amino sugar L-rhodosamine by the glycosyltransferase SnogD (Siitonen *et al*., 2012b; Siitonen *et al*., 2018) (Figure 4). One distinctive feature of nogalamycin is the dual attachment of nogalamine to the aglycone via an *O*–glycosidic bond and a C–C linkage by SnogD and the non-heme iron α-ketoglutarate (α-KG)–dependent carbocyclase SnoK respectively (Siitonen *et al*., 2018). Previous studies have confirmed that the substrate and catalytic promiscuity displayed by these biosynthetic enzymes are the main driving forces behind the diversification of anthracyclines (Metsä-Ketelä M., 2017).

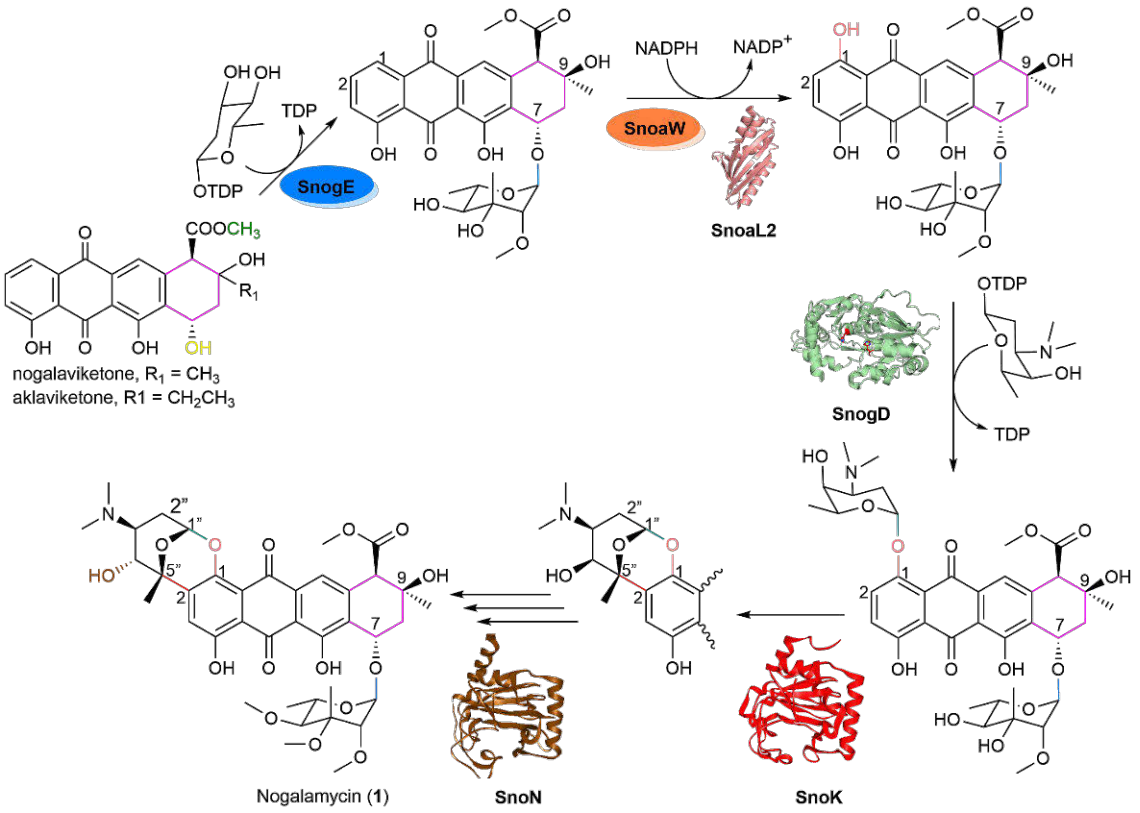


Figure 4: Post-polyketide modifications of the aglycone in anthracycline biosynthesis. The enzyme structures shown and the respective PDB codes are: SnoaL2 (violet), 2GEX; SnogD (green), 4AMB; SnoK (red), 5EPA; SnoN (brown), 5EP9.

1.2 Evolution of anthracyclines

Evolutionary events are mostly driven by DNA mutations, with selective pressure contributing to the emergence of distinct new characteristics (Futuyma, D. J.; Kirkpatrick, M. *Evolution*. (Sinauer) 2017). However, the knowledge of how novel enzymatic functions appear in the first place after these evolutionary events is limiting. Several evolutionary experiments have emphasized the relevance of functional promiscuity and suggested that the novel functions evolve as hidden, secondary catalytic activities, during a time when the old activities are still preserved (Conant & Wolfe, 2008; Des Marais & Rausher, 2008; Sikosek *et al*., 2012). When the selective pressure responsible for the gene duplication event of these bifunctional catalytic entities is relieved, each daughter gene undergo specialization and mature to possess new enzyme activity. Nature appears to recycle secondary structure components on a continual basis in order to incorporate functional promiscuity

(Khan & Ghosh, 2015) and the build-up of additional mutations leads to refinement of activity towards a completely new enzyme function (Höcker *et al*., 2013). Enzyme structures are more conserved than the sequences they are made up of as a consequence of this reuse of structural elements (Siltberg-Liberles *et al*., 2011; Harms & Thornton, 2013), leading to the majority of nature's complex enzyme functions being accomplished by a minimal number of protein superfamilies using a small number of structural types. (Murzin *et al*., 1995; Finn *et al*., 2014).

Many enzymes in secondary metabolic pathways have gradually evolved to harbour novel catalytic functions that differ completely from the ancestral enzymes. In primary metabolism, a gene duplication event is essential for new activities to emerge, since the original functions of the genes typically need to be preserved (Conant & Wolfe, 2008). This is in contrast to secondary metabolism, where alterations in gene functions have been observed directly with concomitant loss of the original function (Grocholski T *et al*., 2019). Such modifications without a gene duplication event are possible due to the non-essential nature of genes involved in secondary metabolism. For *Streptomyces* soil bacteria, this mechanism allows for rapid acclimation to changing environmental conditions and to further increase the chemical diversity of bioactive substances that the bacteria are able to produce (Fewer & Metsä-Ketelä, 2019). Gene duplication events do occur readily on secondary metabolism pathways, but their predominant role is to extend the biosynthetic pathways to allow the bacteria to produce even more complex natural products (Fewer & Metsä-Ketelä, 2019).

1.2.1 Gene duplication

The duplication of genes can be consider as a process through which new genetic materials are generated during molecular evolution. Duplication of genes in natural product biosynthetic gene clusters has provided a means for the bacteria to synthesize more chemically complex metabolites from ancestral pathways. Bacteria with gene clusters that encode two or more copies of homologous genes (Fischback *et al*., 2008) are presumed to have the benefit of producing new metabolites with enhanced or altered biological activity that provide a competitive advantage in their living environment. The difference in the chemistry catalyzed by these structurally similar duplicated gene products is likely due to both substrate and catalytic promiscuity of these enzymes (Matsuda *et al*., 2018; Siitonen *et al*., 2016; Beinker *et al*., 2006).

A typical example of gene duplication event is found in the nogalamycin and aclacinomycin (Scheme 1, **1** and **3** respectively) biosynthetic gene clusters. As described earlier, nogalamycin contain an amino sugar moiety that is dually attached to the aglycone via an O–C and a C–C bond. (Siitonen *et al*., 2012b; Siitonen *et al*., 2016). Prior to the attachment of the amino sugar, an essential C1–hydroxylation reaction is performed by a two-component mono-oxygenase system composed of an

SDR-enzyme SnoaW and a hydroxylase SnoaL2 (Figure 4) (Siitonen *et al*., 2012a; Zhang *et al*., 2017). The SnoaL2 and AclR (Figure 5A) are structurally similar (Beinker *et al*., 2006) to the canonical polyketides cyclases such as SnoaL (from the nogalamycin pathway) and AknH (from the aclacinomycin pathway) (Figure 5A), known to catalyzed the closure of the forth ring of anthracycline (Figure 5B) (Sultana *et al*., 2004 & Kallio *et al*., 2006). Previous studies have revealed that *snoaL* is the ancestral gene that was duplicated into *snoaL2*, because SnoaL2 was shown to perform minor ancestral cyclization reaction (Beinker *et al*., 2006).

Another gene duplication event has occurred on the nogalamycin biosynthetic pathways. After the attachment of the L-rhodosamine amino sugar by SnogD via a O–C linkage, the carbocyclase SnoK establishes the second C2–C5" linkage (Figure 4) (Siitonen *et al*., 2016). A gene duplication in this pathway has led to the appearance of SnoN from SnoK. SnoN is structurally identical to SnoK (Figure 5A), but harbors C4" epimerization activity (Figure 4) (Siitonen *et al*., 2016).

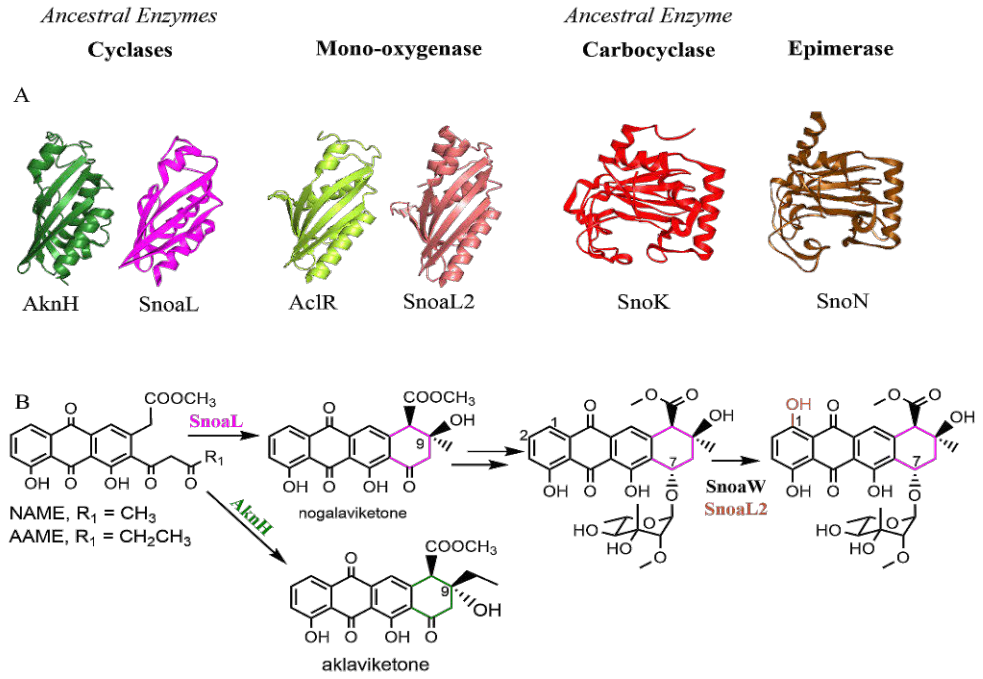


Figure 5: Evolution of secondary metabolic enzymes in anthracycline biosynthetic pathways. A) The fourth ring cyclases AknH (green) and SnoaL (pink) are from the aclacinomycin and nogalamycin biosynthetic pathways, respectively. An unusual two-component monooxygenase mechanism has emerged as a result of gene duplication involving the cyclase-like AclR (limon) and SnoaL2 (salmon) to function together with the quinone reductases AclQ and SnoaW, respectively. B) An additional gene duplication event during the tailoring steps of nogalamycin biosynthesis. Duplication of the C2–C5" nonheme iron– and 2αKG–dependent carbocyclase SnoK (red) into the C4" epimerase SnoN (brown) (Figure 4).

1.2.2 Evolution of enzyme function without gene duplication

Several examples of direct evolution of genes without gene duplication events have been observed on secondary metabolic pathways. Anthracycline biosynthetic gene clusters commonly harbor SAM-dependent methyltransferases that have co-evolved *in situ* with the core genes of the pathways (Grocholski T *et al*., 2019). However, DnrK from the daunomycin pathway and RdmB from the rhodomycin pathway are 52 percent identical in sequence and use the same substrate, but they catalyze different reactions; Dnrk catalyzes a 4–*O*–methyltransfer reaction, whereas RdmB is a 10–hydroxylase (Grocholski T *et al*., 2015; Jansson A *et al*., 2003). In addition, EamK from the komodoquinone pathway has been observed to catalyze solely 10-decarboxylation (Grocholski T *et al*., 2019). Using structure-based information, the region responsible for this functional differentiation was determined and narrowed down to an insertion of a serine residue in DnrK, which was sufficient to introduce the 10–hydroxylation activity. (Figure 6A) (Grocholski *et al*., 2015).

A second example can be found from the biosynthetic pathways of angucycline antibiotics. On the gaudimycin pathway, the short chain alcohol dehydrogenase/reductase (SDR) CabV catalyzes the 6–ketoreduction of an angucycline substrate. The enzyme shares 68% sequence identity with LanV from the related landomycin pathway. LanV uses an earlier biosynthetic intermediate as a substrate and catalyzes a similar ketoreduction, but with opposite stereoselectivity. Exchanging distinct loops between LanV and CabV separately, and in diverse combinations, established a gradual shift in activities of these two protein types, with quadruple chimeric enzymes capable of converting 80%–90% of the substrate (Figure 6B) (P. Patrikainen *et al*., 2014).

A third example involved the conversion of an *O*–glycosyltrasnsferase to a *C*–glycosyltransferase. The pharmaceutical significance is that *C*–glycosides are more stable toward chemical hydrolysis whereas *O*–glycosidic bonds are acid-sensitive. In this protein engineering study, the *O*–glycosidic bond forming LanGT2 (Figure 6C, orange) from landomycin biosynthetic pathway was utilized as a template. The donor sequences (Figure 6C, pink) were obtained from the natural C–C glycosyltransferase UrdGT2 of the urdamycin biosynthetic pathway. This chimeragenesis approach identify a switch of enzymatic activity from *O*–glycosylation to *C*–glycosylation through an insertion of a 10–amino acid segment (Figure 6C)(Härle *et al*., 2011; Tam *et al*., 2015).

1.3 Protein engineering by chimeragenesis

Chimeragenesis is a specialized type of protein engineering in which naturally occurring homologous protein sequences are shuffled individually and in different combinations to mix the parental proteins' subfunctionalities. This has been used in

a number of experiments to create novel enzyme catalysts that are precisely suited for the reaction of interest (Bornscheuer, U. & Pohl, M 2001, *et al* 2002 & Woodley, J. M.2013). For example, combining structural features from two enzyme active sites may result in new enzyme functions in terms of substrate specificity, catalytic rates, or even chemistry.

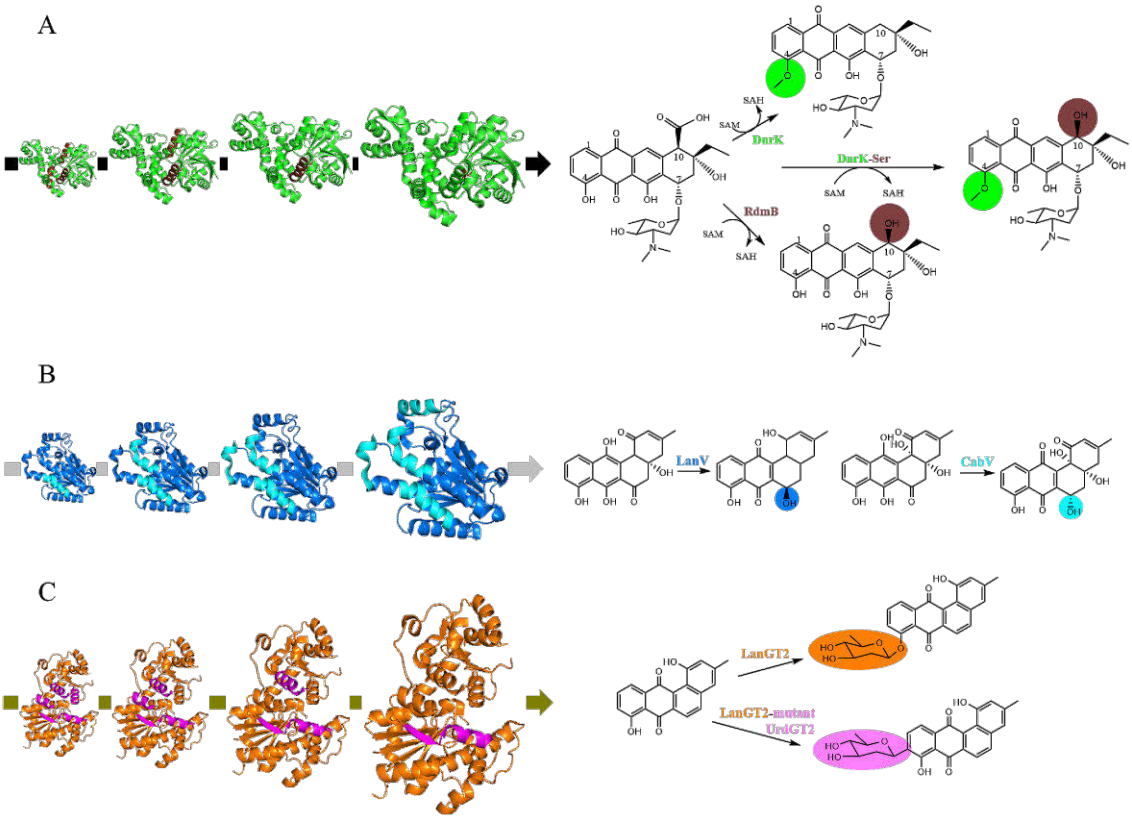


Figure 6: Different protein engineering approaches via chimeragenesis. A) The SAM–dependent enzymes DnrK (green) and RdmB (brown regions) use the same substrate, but perform distinct reactions. DnrK catalyzes 4–O–methylation (green circle), while RdmB is responsible for 10–hydroxylation (brown circle). Exchanging regions between these enzymes and narrowing down, led to the identification of the structural determinant responsible for the difference in the reaction catalyzed. DnrK activity was changed by inserting a single serine residue in α16 of the protein. B) LanV (blue) and CabV (cyan) are NADPH-dependent 6–ketoreductases that have distinct substrate specificities and catalyze reduction with opposite stereochemistry (blue and cyan circle, respectively). Exchanging more regions gradually improved the switch in activity up to 80%–90% when four regions were exchanged. C) LanGT2 (orange) and UrdGT2 (pink regions) are *O*– and *C*–glycosyltransferases, respectively. The use of chimeragenesis to engineer proteins led to the discovery of a 10-amino-acid region that was sufficient for a change in regiospecificity, turning LanGT2 into a C–glycosyltransferase.

Various chimeragenesis approaches have been used to develop new and/or modified enzymatic functions, as well as to investigate the structure/function relationships between enzymes (Balabanova *et al*., 2015). This thesis exploited the most common example of chimeragenesis in which DNA sequences from homologous proteins are exchanged individually and in different combinations to form chimeric proteins (Figure 6). There are many different types of complex engineering approaches, but they can be categorized into three major methods: rational design, random chimeragenesis or semi-rational design (Eriksen *et al*., 2015).

In rational design, information of the structure/function of homologous proteins is used to create modifications that are more likely to produce the desired effect. The lack of structural knowledge for the targeted proteins is problematic and limits the method's availability, but it also reduces the size of the chimeragenesis library, making it possible to test the function of each chimeric protein individually (Balabanova *et al*., 2015; Eriksen *et al*., 2015). Random chimeragenesis involves randomly combining homologous sequences to produce a wide library of sequences from which the target chimeric enzyme can be identified. Random chimeragenesis, unlike rational design, does not require structural knowledge about the target proteins, but it does rely heavily on a high-throughput screening method. Semi-rational design falls in between these techniques and exploits structural information to reduce the sequence space of the library, but still relies on screening methods to identify the desired enzymatic function (Balabanova *et al*., 2015; Eriksen *et al*., 2015).

2 Aims of Study

Generally, the objective of this thesis was to gain insight into anthracycline biosynthesis and to understand how these pathways have been formed during evolution. This thesis aims further to provide explanations for how minor changes among homologous biosynthetic enzymes may lead to the appearance of new catalysts.

Specifically, the aims were to:

a) to understand the complete biosynthesis, modification and attachment of the unique nogalamine moiety of nogalamycin (**1**) by ten enzymes. (Articles I and II).

b) to uncover how a gene duplication event has led to functional differentiation of the nonheme iron and 2-oxoglutarate-dependent mono-oxygenases SnoK and SnoN in nogalamycin biosynthetic pathway (Article III).

c) to gain insight how the functions of S-adenosyl methionine dependent methyltransferases have diversified in anthracycline biosynthetic pathways (Article IV).

3 Materials and Methods

An overview of the materials and methods used are presented here. The experimental details can be found in the original publications (I - IV).

3.1 Molecular Biology

The native gene and chimeric gene constructs were amplified using standard Molecular Biology techniques such as polymerase chain reaction (PCR). High fidelity Phusion DNA polymerase (New England Biolabs) was used in the PCR reactions. The oligonucleotide primers used in this thesis work are presented in their respective articles (I -IV). Sticky ends restriction enzymes, T4 ligase (Thermo Scientific) and the pBAD vector system were used for molecular cloning. *Escherichia coli* TOP10 (Invitrogen, ThermoFisher Scientific, Waltham, MA, USA) was used as the cloning and production host. The purifications and isolations of DNA was performed using GeneJET plasmid miniprep kit and GeneJET gel extraction kit (ThermoFisher Scientific). Isolation of DNA from *Streptomyces* species and protoplast transformations were performed using conventional techniques (see article II for details).

3.2 Recombinant protein expression, purification and analysis

The wild type enzymes and the chimeras containing an additional AHHHHHHHRSAD sequences were produced in *E. coli*. The cells were grown either in Erlenmeyer flasks or a Fermenter (Bioengineering) at 30–37 °C until the OD600 reached 0.5–0.7 after which the cells were induced with L-arabinose. Protein expression was allowed for 15–18 h at room temperature (RT). After the production of the enzymes, the cells were pelleted and resuspended in the desired buffer and stored at -20 °C. Before purification, the *E.coli* cells were thawed on ice and lysed by sonication. The lysed cells were spun by centrifugation to remove the debris, and the crude lysate was mixed with TALON Superflow (GE Healthcare) to bind the recombinant enzymes. After binding, the impurities were washed away from the column and the desired enzymes were eluted from the

column. The PD-10 column (GE Healthcare) was used for desalting, and the proteins were stored in the desired buffer in 40% (v/v) glycerol at -20 °C. The purity of all the purified wild type and chimeric enzymes was evaluated by SDS-PAGE. The concentrations of the purified enzymes were estimated photometrically at 280 nm (NanoDrop2000; Thermo Scientific, Vantaa, Finland) and stored at -20 °C for further use.

3.3 Production and purification of substrates and standards

The substrates and standards used in this thesis were produced from *S.albus* and purified in multiple steps. The *S.albus* strains were cultivated for 5–7 days and the cells were harvested by centrifugation and discarded as the compounds are secreted into the growth media. We used XAD7-resin (20 g·L-1; Rohm and Haas) to extract the compounds from the supernatant. After binding the compounds to the resin, we washed the resin several times with water to remove the culture media. The bound metabolites were now extracted from the resin with methanol. The majority of the impurities in the crude extract were removed by running the crude extract through a gel filtration LH20 column (GE Healthcare) in methanol. The fractions containing the desired metabolites were determined by SCL-10Avp HPLC with a SPD-M10Avp diode array detector (Shimadzu) using a Kinetex column (2.6 µm C18 100 Å, LC Column 100 x 4.6 mm, Ea, Phenomenex). The fractions were concentrated with -rotavapor RII (BUCHI) and subjected to preparative HPLC (LC-20AP, model; CBM-20A, SHIMADZU) with a reverse phase column (SunFire Prep C18, 5 mm, 10 × 250 mm; Waters). For all metabolites, a mobile phase gradient from 10% acetonitrile to 70% acetonitrile including 18 mM ammonium acetate, pH 3.6, was used.

For the structural elucidation of novel reaction products, several large scale (15–45 mL reaction volume) *in vitro* reactions with purified enzymes were carried out. The reaction products were extracted with chloroform and purified by size-exclusion chromatography [Sephadex LH-20 (GE Healthcare)] and preparative HPLC.

3.4 Enzymatic reactions

The enzymatic assays for end-point measurements were usually carried out with a reaction volume of 200 µl and in cases where novel reaction products were to be characterized and structurally elucidated, reaction volumes from 15–45 mL were used. The concentration of the enzymes, substrates, cofactors, cosubstrates, reaction conditions and reaction buffers used vary among the articles (I - IV). At the end of

every enzymatic reaction, the content of the reactions was extracted with chloroform and the chloroform was dried either with concentrator plus (Eppendorf, Hørsholm, Denmark) (for 200 µl reactions) or with Rotavapor RII (Büchi) (large scale reactions for structural elucidation of new products) and dissolved with methanol. The reaction products were analyzed by a SCL-10Avp HPLC with a SPD-M10Avp diode array detector (Shimadzu) using a Kinetex column (2.6 µm C18 100 Å, LC Column 100 x 4.6 mm, Ea, Phenomenex).

3.5 Nuclear magnetic resonance (NMR)

The purified and dried compounds obtained from preparative HPLC were further dried under nitrogen flow in the dark after which the compounds were dissolved in deuterated methanol and analyzed by NMR (600 MHz Bruker with a AVANCE-III NMR-system) equipped with a liquid nitrogen cooled Prodigy TCI (inverted CryoProbe) at 298 K. The experiments conducted include 1D (^{1}H) analyses (^{13}C, ^{1}H, 1D-TOCSY) and 2D measurements; Correlation spectroscopy (COSY), heteronuclear multiple bond correlation (HMBC), heteronuclear single quantum coherence-HSQC, heteronuclear single-quantum and distortionless enhancement correlation (HSQCDE), nuclear overhauser effect spectroscopy (NOESY). Spectral analysis was done with Topspin (Bruker Biospin). The high resolution mass (MicroTOF-Q, Bruker Daltonics) was performed with direct injection of purified compounds dissolved in methanol.

3.6 Crystallography

TamK and TamK RRRT were produced in *E.coli* TOP10 as N-terminal histidine tagged recombinant proteins. The proteins were purified with TALON Superflow and eluted in buffer B pH7.5 [50 mM Tris, 100 mM NaCl, 300 mM imidazole, 10% (v/v) glycerol]. The enzymes were desalted with a PD-10 column and concentrated to 20-40 mg/ml in C-buffer pH 7.5 [25 mM Tris, 200 mM NaCl, 5% (v/v) glycerol]. The protein was analyzed on SDS PAGE and co-crystallized with either compound **20** (or **21**) and the cosubstrate SAM by hanging drop vapour diffusion in buffer D (0.04M Potassium phosphate monobasic, 16% w/v PEG8000, 20% glycerol). Droplets of the protein solutions were mixed in different volume ratios (1:1, 2:1 and 1:2) with the reservoir solution (buffer D). Crystals of both enzymes appeared after four days and the crystals were frozen in liquid nitrogen prior to data collection. Crystallographic data of wild type TamK and TamK RRRT were collected at the beamline ESRF ID23 of the European Synchrotron Radiation Facility (Grenoble, France), to a resolution of 1.5 Å and 1.9 Å, respectively (Table 3). The structures of both complexes were determined with molecular replacement using Phenix

Molecular Replacement program (Phaser)(McCoy AJ *et al*., 2007), with DnrK WT (PDBID: 1TW2) as the initial model. WinCOOT (Emsley P. et al., 2010) and PHENIX were used to build and refine the models. The C2 model of TamK and TamK RRRT crystallized as a monomer containing SAH.

4 Results and Discussion

This thesis took advantage of years of structural studies and functional diversity of the enzymes residing on related anthracycline pathways. The biosynthesis of nogalamycin has been studied for more than 25 years and the last remaining unknown steps are reported in this work. These are related to the aminoglycosylation of nogalamycin and the formation of the atypical epoxyoxocin ring system. The systematic synthesis of the L-rhodosamine carbohydrate moiety, which enhances the biological activity of nogalamycin, clarifies that the modification of the sugar to L-nogalamine occurs after glycosylation. We identified that the Rieske enzyme SnoT is responsible for the installment of the 2″–hydroxyl group and elucidated the reaction order of the post-glycosylation events.

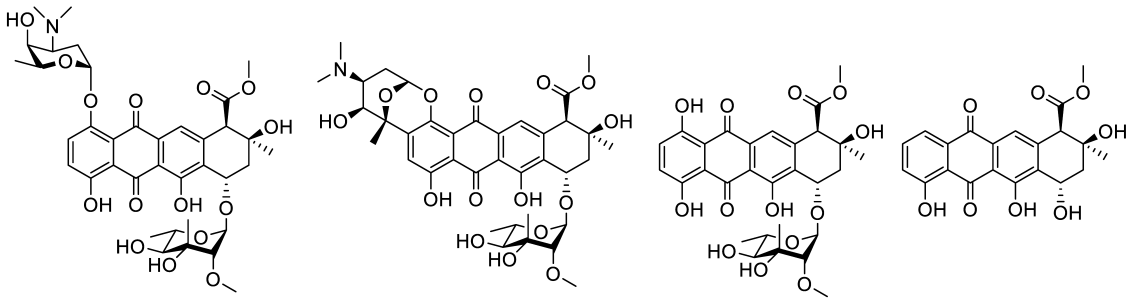
Furthermore, we performed extensive structure/function studies in order to understand how homologous enzymes have evolved to catalyze distinct chemistry on anthracycline biosynthesis pathways. The anthracycline biosynthetic enzymes were engineered by exchanging short segments (that may be responsible for their functional differentiation) of homologous proteins in different combinations to generate novel catalysts and products. After the identification of the key segments, the individual amino acid(s) responsible for the emergence of the novel enzyme function were pinpointed more precisely by narrowing down the originally selected segment.

4.1 Biosynthesis of the amino sugar moiety of nogalamycin R (Article I)

Carbohydrate moieties are important for the bioactivity of anthracycline anticancer agents such as **1** and **9**, which contains L-nogalose and L-nogalamine units. The L-nogalose is attached through a canonical *O*–glycosidic linkage, while the L-nogalamine is connected via an unusual dual linkage composed of C−C and *O*–glycosidic bonds. L-rhodosamine is a common sugar unit found in anthracyclines such as doxorubicin, aclacinomycins, komodoquinone, keyicin and cytorhodins (**2**, **3**, **6**, **9**, **10**, Scheme 1, respectively).

The dual attachment of the amino sugar is quite unique to **1** and **9**. This uniqueness motivated us to investigate its biosynthesis and attachment. We carried

out the systematic enzymatic synthesis of the amino sugar unit of nogalamycin and attached it to the acceptor compound via an O–C and a C–C bond in two steps.



nogalamycin K (**11**) nogalamycin R (**12**) **13** nogalamycinone (**14**)

Scheme 3: Selected compounds relevant to this section.

4.1.1 Biosynthesis of L-rhodosamine from TDP-α-D-glucose

We performed the enzymatic synthesis of the amino sugar of nogalamycin in a single-step using seven different enzymes; RmlB (a homolog of the 4,6-dehydratase SnogK from *E.coli* K12 (Stevenson *et al*., 1994)), a 2,3-dehydratase SnogH, a pyridoxal 5′–phosphate (PLP)-dependent SnogI, an epimerase SnogF, a ketoreductase SnogG and the methyl transferases, SnogX and SnogA. Due to the instability of some of the recombinant enzymes, the enzymes were immobilized and the starting substrate and cofactors were loaded at once. The first reaction was the 4,6-dehydration of the starting substrate (**15**) by RmlB (Figure 7), followed by another 2,3-dehydration by SnogH leading to the formation of TDP-3,4-diketo-2,6-dideoxy-α-D-glucose. The next step was the transamination of the resulting intermediate by the pyridoxal 5′–phosphate (PLP)–dependent SnogI to generate TDP-3-amino-4-keto-2,3,6-trideoxy-α-D-glucose. The synthesis then proceed through 5–epimerization by SnogF and 4–ketoreduction by SnogG to form TDP-L-daunosamine, which was then methylated twice by SnogX and SnogA using SAM to generate TDP-L-rhodosamine (**16**, Figure 7).

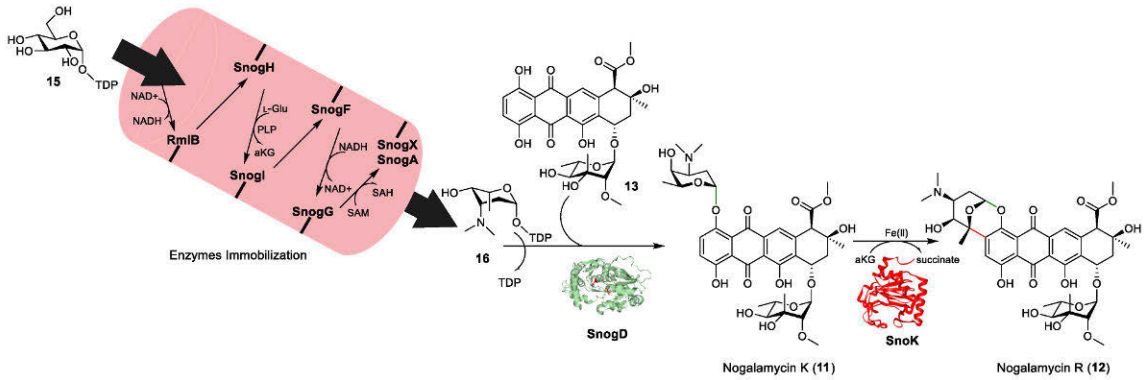


Figure 7: One-step enzymatic synthesis of TDP-L-rhodosamine and its dual attachment to the aglycone. Seven enzymes responsible for the biosynthesis of TDP-L-rhodosamine were immobilized and the substrate (**15**), together with all cofactors and cosubstrates were added and incubated. The resulting product (**16**) was attached at two positions on the acceptor compound (**13**); firstly by the glycosyltransferase SnogD and secondly by the carbocyclase SnoK to form nogalamycin K (**11**, green bond) and nogalamycin R (**12**, red bond), respectively.

4.1.2 Methylation of L-rhodosamine

The nogalamycin gene cluster contains two homologous genes that code for methyl transferases, SnogX and SnogA (54% sequence identity) that were both included in the reaction mixture for the biosynthesis of L-rhodosamine. The participation of both methyltransferases in this reaction was unclear to us and prompted to investigate the role of SnogX and SnogA in the biosynthesis of this amino sugar. Unexpectedly, the presence of both methyltransferases was not critical for the reaction and **16** could be detected with either SnogX or SnogA only. Our reaction with SnogX after 2 h of incubation produced solely **16** (Figure 8A), while we could observed a monomethylated intermediate in the reaction with SnogA (Figure 8B). Nevertheless, this monomethylated intermediate was converted to **16** after prolonging the reaction time to 5 h (Figure 8B). This result implies that SnogA is a poor converter of the monomethylated intermediates to **16**, which shows that the biological role of SnogA in nogalamycin biosynthesis might be to perform the first methylation reaction while SnogX is responsible for the second methylation reaction, although both enzymes are capable of performing dimethylations.

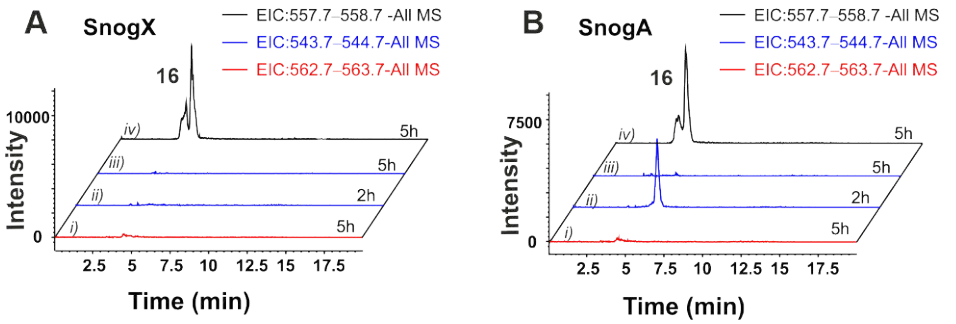


Figure 8: LC-MS analysis and comparison of the *in vitro* products of SnogX and SnogA. (A) Analysis of TDP-L-rhodosamine biosynthesis using SnogX as the methyltransferase. (i) Extracted ion chromatogram (EIC) for the consumed substrate **15** after 5 h incubation. (ii) EIC for the monomethylated product after 2 h incubation. (iii) EIC for the monomethylated product after 5 h incubation. (iv) EIC for the dimethylated product **16** after 5 h incubation. (B) Analysis of TDP-L-rhodosamine biosynthesis using SnogA as the methyltransferase. (i) EIC for the consumed substrate **15** after 5 h incubation. (ii) EIC for the monomethylated product after 2 h incubation. (iii) EIC for the monomethylated product after5 h incubation. (iv) EIC for the dimethylated product **16** after 5 h incubation.

4.1.3 The dual attachment of L-rhodosamine to the acceptor aglycone

In nogalamycin (**1**), the points of attachment of the amino sugar to the aglycone is unique (scheme 1, **1**). The amino sugar is dually attached via an O–C and a C–C linkage. After synthesis of L-rhodosamine (**16**), the next step was the attachment of **16** to the aglycone via the O–C and the C–C linkage using the glycosyltransferase SnogD and the carbocyclase SnoK, respectively (Figure 7). This dual attachment experiment was performed in order to confirm the identity of the synthesized **16**. We observed the expected product **12** after incubation of the two substrates in the presence of the glycosyltransferase SnogD (Siitonen V. *et al*., 2012a), the carbocyclase SnoK (Siitonen *et al*., 2016), α-ketoglutarate, Fe(II) and ascorbate (Figure 7).

Many glycosyltransferases are known to require the assistance of P450–like enzymes in order to catalyze their reaction (Moncrieffe *et al*., 2012). For example, AknS/AknT in aclacinomycin biosynthesis pathway catalyzed the transfer of L-rhodosamine sugar moeity to the C-7 position of the aglycone (Lu *et al*., 2005). Nogalamycin gene cluster contains such P450-like enzyme, SnogN. We were unsure whether SnogN plays an assisting role in the transfer of the aminosugar moiety (**16**) to the aglycone. To be sure whether the corresponding gene *snogN* can play the assisting role in the transfer of **16**, we knocked out the gene from the cosmid pSnogaori and transformed *S. albus* with the engineered cosmid pSnoΔgN. We cultured this strain and the analysis of culture extract showed that the main metabolite produced was nogalamycinone (**14**, scheme 3), which implies that SnogN

may be working in partnership with the TDP-L-nogalose transferase SnogE (Figure 4). Therefore, the auxiliary SnogN has more to do with the chemical structure of the aglycone acceptor molecule instead of the donor TDP-aminosugar in nogalamycin biosynthesis. To confirm this result, we complemented the mutation to *snogN* with an intact copy of the gene, which reestablished the production of the double glycosylated metabolite nogalamycin R (**12**, scheme 3).

In this study, using seven enzymes, we synthesized L-rhodosamine-thymidine diphosphate (TDP) from α-D-glucose-L-TDP. We completed the dual linkage system in a second phase by employing the glycosyl transferase SnogD and the α-ketoglutarate dependent oxygenase SnoK to attach the aminosugar to an anthracycline aglycone acceptor. Furthermore, our findings show that the glycosylation-facilitating auxiliary P450-type protein SnogN is interestingly linked to the attachment of the neutral sugar L-nogalose rather than the aminosugar L-nogalamine in nogalamycin biosynthesis.

4.2 C2″–hydroxylation of nogalamycin (Article II)

The carbohydrate moieties of secondary metabolites are known to play an essential role in the bioactivity of these compounds (Weymouth-Wilson, 1997). Late stage biochemical modifications such as hydroxylation and methylation of these sugar moieties greatly improves the bioactivity of these compounds. Anthracyclines (scheme 1) exert their biological effect mainly through DNA binding and poisoning of topoisomerases (Nitiss, 2009). Its two carbohydrate units, L-nogalose and L-nogalamine that interact with the minor and major grooves of DNA, respectively, facilitate the intercalation of nogalamycin anti-cancer agent (**1**, scheme 1) into the DNA double helix (Nitiss, 2009). Analysis of the nogalamycin biosynthetic gene cluster brought our attention to a putative Rieske oxygenase *snoT*, as the possible candidate for 2″–hydroxylation of nogalamycin. The hydroxyl group at position C2″ of nogalamycin is essential for its binding to the DNA (Smith *et al*., 1995). This essential late stage hydroxylation in the biosynthesis of nogalamycin motivated us to study the function of *snoT*, which appeared to be the potential candidate for this modification.

4.2.1 The function and natural substrate of SnoT

We investigated the function of SnoT by expressing *snoT* in a heterologous host (*Streptomyces albus*/pSnogaori), which harbors a cosmid encoding the majority of the genes in nogalamycin pathway responsible for biosynthesis of **12** (scheme 3). We analyzed the culture extracts of *S. albus*/pSnogaori+T and found a new compound with a +16 Da increase in mass.

Rieske oxygenases usually need electron transfer components (e.g., ferredoxin and ferredoxin reductase) for their catalysis. Our analysis of the nogalamycin gene cluster does not reveal any gene products for transfer of reducing equivalents from NADPH to SnoT. To complement for this, we used *E. coli* cell-free lysate extracts for all *in vitro* reactions of SnoT. The approach allowed us to observe the appearance of trace quantities of a compound with a +16 Da increase in mass to **11** by LC–MS. This result showed that the unnatural electron transfer components from *E. coli* were compatible with SnoT. We then tested SnoT with compound **11**, **12** and **19** and found out that **11** is the natural substrate of SnoT, since **12** and **19** were not modified (Figure 9). LC–MS analysis of the reaction reveals that SnoT was able to convert substrate **11** to product **17**, which was confirmed by NMR.

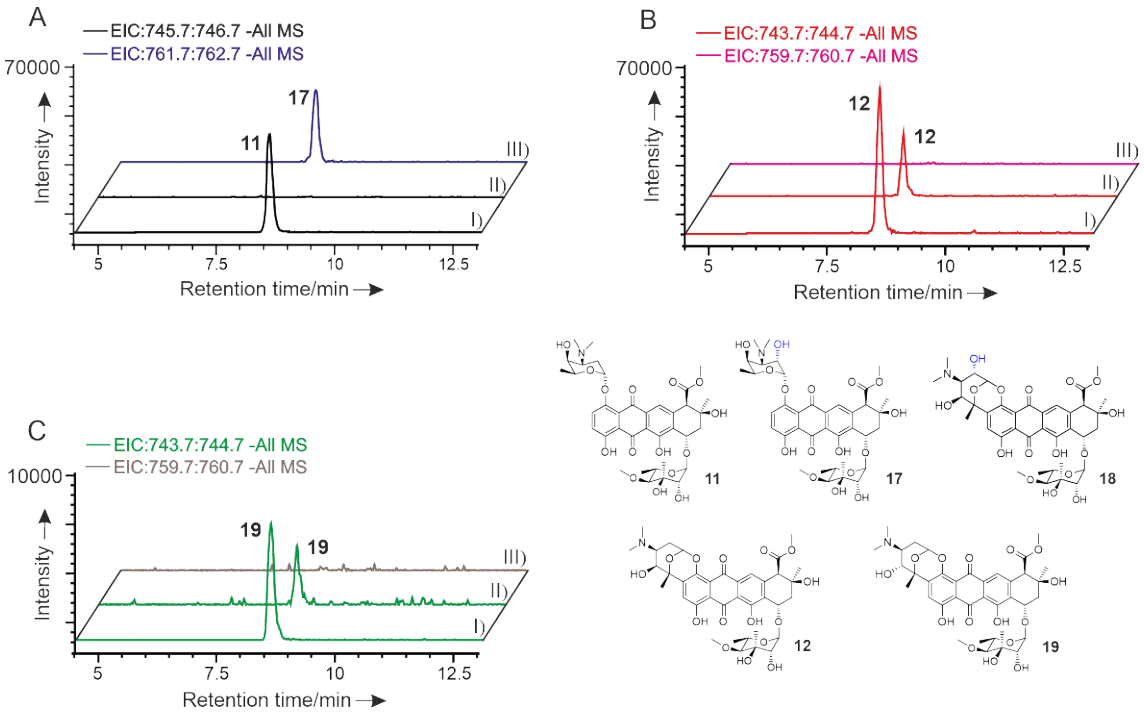


Figure 9: Verification of the natural substrate of SnoT. A) Enzymatic conversion of nogalamycin K (**11**) to nogalamycin T (**17**) by SnoT. I) Purified compound **11** with the extracted ion chromatogram (EIC) for the mass of **11**. II) EIC of the reaction product for the mass of **11**. III) EIC of the reaction product for the mass of **17**. B) Enzymatic conversion of nogalamycin R (**12**). The enzyme did not consume **12**. I) EIC for the purified compound **12**. II) EIC of the reaction product for the mass of **12**. III) EIC of the reaction product for the mass of nogalamycin TR (**18**, see Fig.11 below). C) Enzymatic conversion of nogalamycin RE (**19**). SnoT did not consume **19**. I) EIC for the purified compound **19**. II) EIC of the reaction product for the mass of **19**. III) EIC of the reaction product for the mass of nogalamycin TRE.

4.2.2 Structural elucidation of SnoT reaction product

Our results from the LC–MS analysis of the reaction product of SnoT suggested that it is the expected compound **17**. In order to identify the product, we performed large–scale production and purification of substrate **11** from *S. albus*/pSnoΔK that harbors a cosmid from which the carbocyclase gene *snoK* has been deleted (Siitonen *et al.*, 2016). We conducted several reactions with the purified **11** and SnoT cell-free lysates at room temperature to allow the bioconversion to occur. We quenched the reaction and extracted its contents with chloroform and then purified it with LH–20 and preparative HPLC. Our structural interpretation by ^{1}H NMR and COSY, HSQCDE, HMBC and HR-MS (ESI *m/z* $[M–H]^-$ obs. 760.2833, calc. 760.2822) established that the reaction product was **17** (Figure 10).

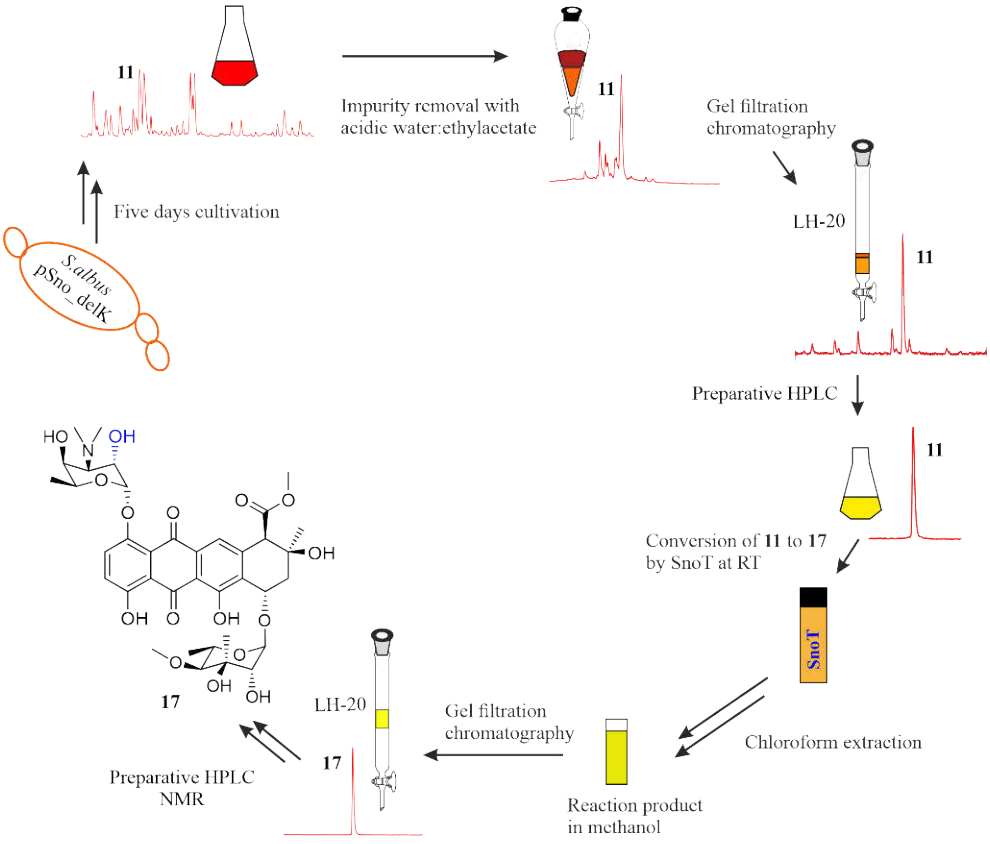


Figure 10: The stepwise large-scale production and purification of nogalamycin K (**11**) and its enzymatic conversion to nogalamycin T (**17**) by SnoT. The chemical structure of **17** was confirmed by NMR.

4.2.3 Reaction order during tailoring steps in nogalamycin biosynthesis

Enzyme substrate promiscuity, the ability of an enzyme to use a wide range of substrates to catalyze the same or similar chemical reactions, and enzyme catalytic promiscuity, which describes a situation where an enzyme catalyses secondary reactions that differ from the canonical primary reaction (Copley S.D., 2015), have posed a great challenge in determining the order of the tailing reactions in anthracycline biosynthesis. This has made deciphering the reaction order of the final stages of nogalamycin biosynthesis challenging. Our result for the determination of the natural substrate of SnoT revealed that the enzyme could modify only compound **11** that lacks the additional C2–C5″ bond as a substrate, and that SnoT did not hydroxylate compound **12** and **19** (Figure 9). We previously demonstrated that SnoK can convert substrate **11** to compound **12**, which was further epimerized to **19** *in vivo* by SnoN. However, *in vitro* studies of the SnoN reaction has been quite challenging, because its reactions with compounds **11** and **12** yielded degradation products (Siitonen *et al*., 2016). We performed a series of complicated reactions in order to determine the order of reaction among the tailoring enzymes; SnoT, SnoK and SnoN. We found out that SnoT was only capable to catalyze 2″–hydroxylation of compounds that lack the C2–C5″ bond such as **11** to generate **17**. In addition, the carbocyclase SnoK have a broad substrate promiscuity, but the biological function is probably the conversion of **17** to **18** (Figure 11). Unlike the carbocyclase SnoK, the epimerase SnoN only accept intermediates containing the C2–C5″ bond such as nogalamycin R (**12**) and nogalamycin TR (**18**), with only **12** residing on the pathway towards the end product nogalamycin (**1**). We established that the reaction order begins with the 2″–hydroxylation by SnoT and continues with 2–5″ carbocyclization and 4″–epimerization by the non-heme iron and 2–oxoglutarate-dependent enzymes SnoK and SnoN, respectively (Figure 11).

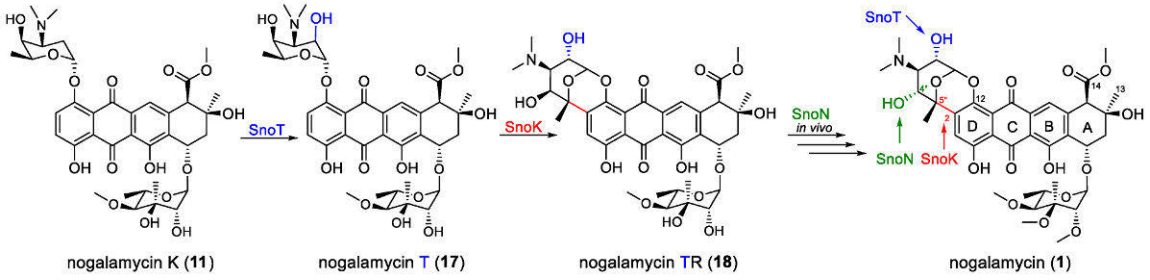


Figure 11: The order of reaction for late-stage tailoring steps in the biosynthesis of nogalamycin. This experimental work proves that the initial step is hydroxylation of the C2″–position of L-nogalamine by SnoT, which is followed by the C2″–C5″ carbocyclization and C4″ epimerization catalyzed by SnoK and SnoN, respectively.

In this work, these set of complicated experiments elucidate the last remaining unknown steps in the biosynthesis of nogalamycin (**1**) and clarify how the 2″–hydroxy and 4″–hydroxy groups of L–nogalamine, which are important for DNA interaction, are assembled.

4.3 Evolution by gene duplication in nogalamycin biosynthesis pathway (Article III)

Streptomyces are known to generate thousands of chemically complex natural products with potent biological activities. Rapid evolution of secondary metabolism has been the main driving force of the chemical diversity found in these microbes. The biosynthetic pathway of the anthracycline nogalamycin contains two non-heme iron and 2–oxoglutarate-dependent monooxygenases, SnoK and SnoN (Figure 12). These monooxygenases are structurally similar but catalyze distinct chemical reactions; SnoK is a C2–C5″ carbocyclase, whereas SnoN catalyzes stereoinversion at the adjacent C4″ position (Figure 11).

4.3.1 Tracing the evolutionary path by chimeragenesis

Chimeragenesis has been a useful tool to uncover the evolutionary aspects of biosynthetic enzymes involved in the secondary metabolism in *Streptomyces* and in other natural product producing bacteria. This tool provided us with knowledge about the evolutionary events that had occurred in these bacteria and provided insight into the great chemical diversity of anthracyclines. Many secondary metabolic enzymes have been shown to harbor high sequence and structure similarity, but catalyze entirely unrelated reactions from each other (see Figure 5). Here we took advantage of the structural information of homologous enzymes to identify segments that could be responsible for the functional differentiation. Once the segments are shown to be critical for the catalysis, we examined them more precisely using structure-based information to pinpoint the amino acid(s) responsible for the catalysis.

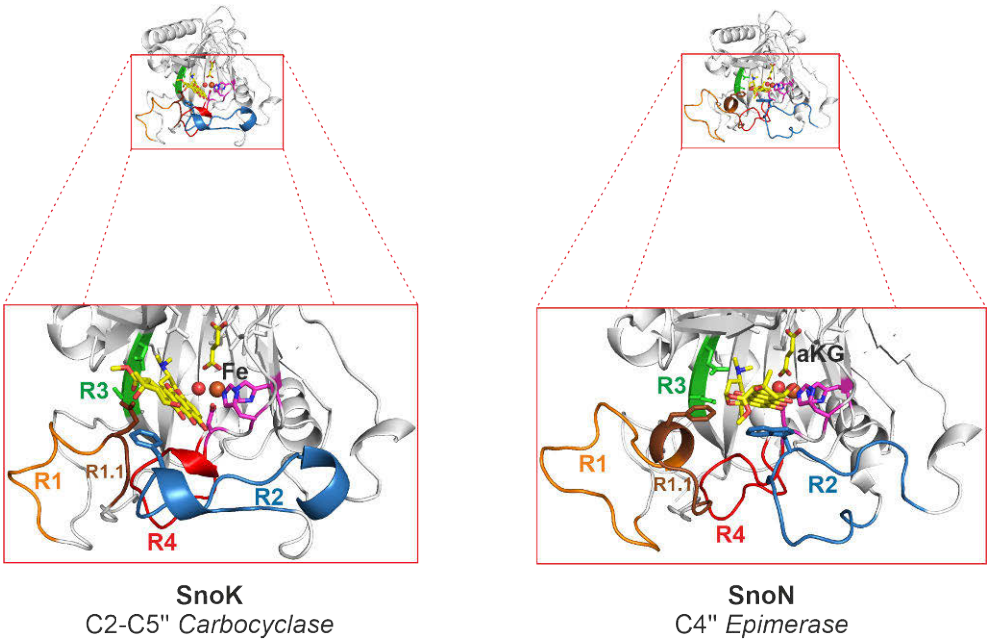


Figure 12: An overview of the crystal structure of SnoK and SnoN in complex with the substrate (**11**). The regions exchanged in this study are labelled from R1–R4. SnoK is a C2–C5" carbocyclase whereas SnoN is a C4" epimerase that catalyzed the later steps in the biosynthesis of nogalamycin. The difference in the reaction catalyzed by these enzymes lies on the orientation and positioning of the substrate in front of the Fe^{3+}–superoxide species.

4.3.2 Selection of regions for chimeragenesis

We were excited to trace back the evolutionary event that is responsible for the difference in the reaction catalyzed by SnoK and SnoN. We superimposed the crystal structures of these enzymes and identified four structural regions that could be involved in the functional differentiation (Figure 12; R1–R4 and Table 1). To examine the significance of the selected regions for catalysis in SnoN and SnoK, we exchanged the regions individually and in various combination and generated 30 chimeric constructs. We produced the histidine-tagged chimeric proteins in *E. coli* TOP10 and purified them by affinity chromatography. We named the chimeric enzymes based on the protein scaffold, followed by identification of the exchanged region (for instance, SnoN R3 consists of the SnoN protein, where the R3 region has been replaced with the sequence from SnoK). We utilized compound **11** (Figure 14) as the substrate to investigate the relative activities of the chimeric enzymes, since this substrate can be consumed by both wild type enzymes.

Table 1. Amino acid sequence of the four distinct regions (R1-R4) between SnoK and SnoN.

Region (R)	SnoK	SnoN
R1.1	GVQ (238-240)	TG I A F L (241-246)
R1	GVQDGHLSRLS (238-248)	TGIAFLDDLPGTGADPLREGA (241-261)
R2	EDRQEHMTSFAEFRDLADVW (164-183)	PDPDTGDEPWAGAFTR (173-186)
R3	YSSD (103-106)	FKLE (109-112)
R4	DDETGLPLNELSAT (122-135)	HDAFAFPFSTAGTA (131-144)

4.3.3 Relative activity of chimeras

The results from the enzymatic assays established that all of the regions we selected for chimeragenesis were essential for catalysis (Table 2). One SnoK-based chimera containing region R1 from SnoN, SnoK R1 displayed significant epimerization activity (32%). The regions we exchanged in SnoK R2 and SnoK R4 influenced catalysis in the same way, although not as significant as the SnoK R1 chimera. SnoK R2 and SnoK R4 gained 7% and 17% epimerization activity, respectively (Table 2 and Figure 4 in publication III). Both of these SnoK chimeras had lost considerable amounts of their activity, since most of the substrate **11** was left unconsumed after the reactions were quenched (Table 2). We were surprised to find out that the SnoK R3 chimera was completely inactive and did not consumed the substrate (Table 2). Our results from the multiple chimeric enzymes confirmed that SnoK R12 displayed minor carbocyclization and epimerization activities of 10% and 12%, respectively. The chimera SnoK R14 still harbored 4% of native carbocyclization activity, while all other chimeric enzymes had completely lost their activities (Table 2).

Using the wild type SnoN as a template to generate chimeric proteins containing sequences from SnoK was unfavorable for the enzymatic activity. None of the SnoN-based chimeras gained activity toward the SnoK carbocyclization reaction. Our analysis of the single chimeric enzymes showed that SnoN R1, SnoN R3 and SnoN R4 preserved only 12%, 30% and 14% of relative activity, respectively (Table 2). The SnoN R2 consumed 81% of substrate **11** in comparison with the native SnoN. This result implied that the exchanged R2 sequence from SnoK was not detrimental to the enzymatic activity of SnoN R2 chimera (Table 2). For all the SnoN chimeras containing multiple regions from SnoK, just four of them; SnoN R12, SnoN R13, SnoN R23 and SnoN R24 preserved 6%, 29%, 7% and 7% of relative activity, respectively.

Table 2. Overall analysis of the enzymatic reactions of all chimeric enzymes.

Enzyme	% Substrate	% SnoK product	% SnoN product
Neg. control (no enzyme)	100	0	0
SnoK (pos. control)	0	100	0
SnoN (pos. control)	0	0	100
SnoK R1	0	68±0	32±0
SnoK R2	85±1	8±1	7±1
SnoK R3	100	0	0
SnoK R4	48±4	34±3	17±1
SnoK R12	78±2	10±1	12±1
SnoK R13	100	0	0
SnoK R14	96±0	4±0	0
SnoK R23	100	0	0
SnoK R24	100	0	0
SnoK R34	100	0	0
SnoK R23	100	0	0
SnoK R124	100	0	0
SnoK R134	100	0	0
SnoK R234	100	0	0
SnoK R1234	100	0	0
SnoN R1	88±0	0	12±0
SnoN R2	0	0	100
SnoN R3	70±3	0	30±3
SnoN R4	86±2	0	14±2
SnoN R12	94±1	0	6±1
SnoN R13	71±5	0	29±5
SnoN R14	100	0	0
SnoN R23	93±0	0	7±0
SnoN R24	93±1	0	7±1
SnoN R34	100	0	0
SnoN R123	100	0	0
SnoN R124	100	0	0
SnoN R134	100	0	0
SnoN R234	100	0	0
SnoN R1234	100	0	0
SnoK R1.1	0	58±2	42±5

To confirm that the loss of enzymatic activity by the SnoN chimeras was due to our engineering work but not unfolded chimeric enzymes, we performed CD spectroscopy. Our results showed that the all the SnoN chimeras without activity were folded (Figure 13) and that the inactivity resulted from the engineering efforts.

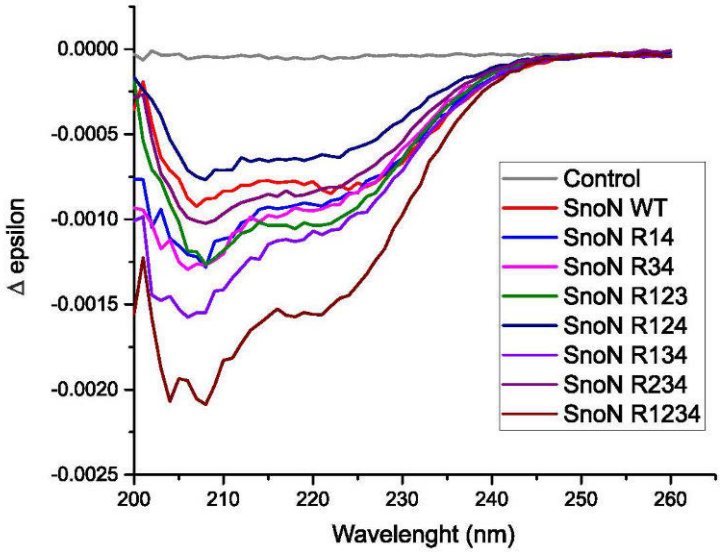


Figure 13: CD spectra of the inactive SnoN chimeras (SnoN R14, SnoN R34, SnoN R123, SnoN R124, SnoN R134, SnoN R234, and SnoN R1234) compared to buffer control and SnoN WT. The inactive chimeric proteins are folded and the inactivation is likely due to the engineering efforts.

4.3.4 Evolution of SnoK to SnoN

4.3.4.1 Role of region R1 in the evolution of epimerization activity

We used molecular modeling and docking calculations to understand the structural changes in our chimeric proteins and further investigated how the modifications in SnoK R1 resulted to the appearance of SnoN epimerization activity. Our modeling suggested that region R1 has a significant impact in catalysis and that, this region provides several critical interactions with the ligand. The results also revealed that the main changes in the active site of SnoK R1 in comparison with the native enzyme were the V239SnoK/F245SnoN and Q240SnoK/L246SnoN replacements (Figure 14A and 14B). Interestingly, the R1 loop in SnoK R1 was shown to be stabilized in a similar way as in the native SnoN and the position of F245SnoN corresponds to that of W180 in SnoN. As a result of this, the substrate (**11**) was stacked between F245SnoN and W57 in the SnoK R1 complexes (Figure 14C). This result indicates that these stacking interactions might have been the main evolutionary event to reposition **11** for the SnoN epimerization reaction.

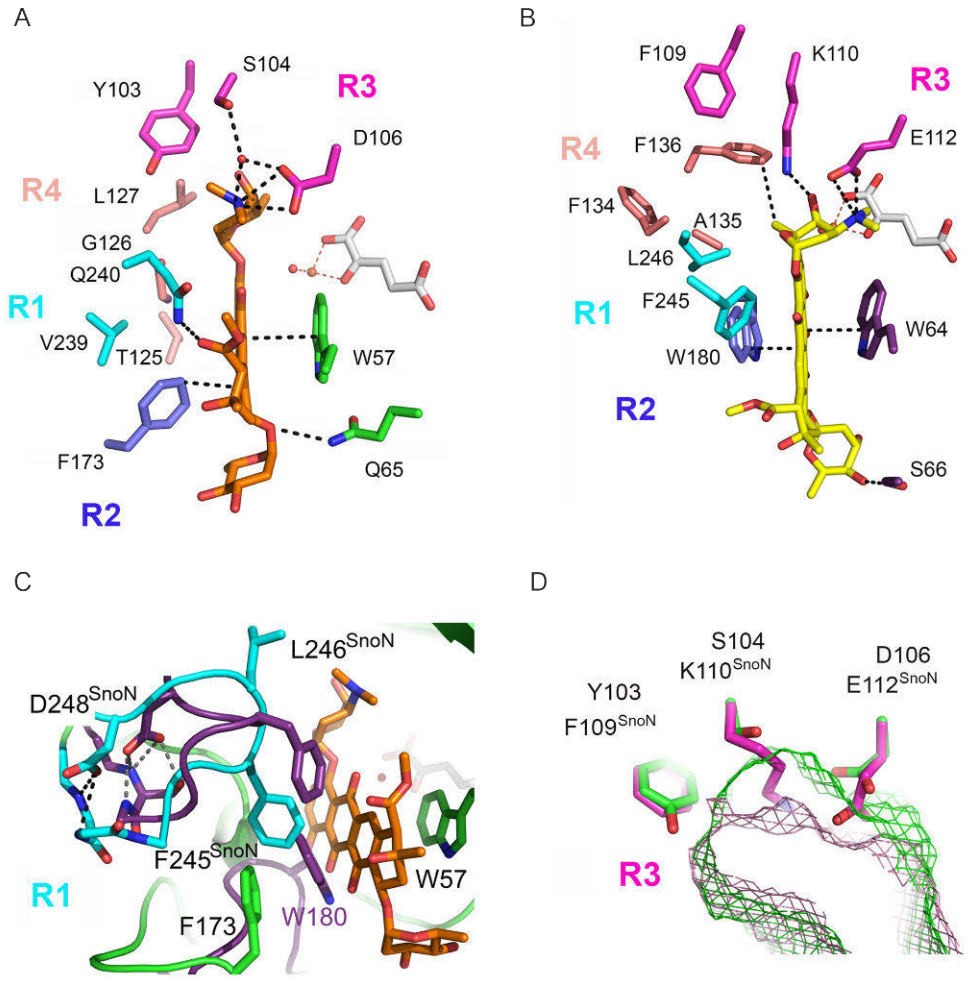


Figure 14: Comparison of SnoK and SnoN ligand binding properties and structural analysis of SnoK chimeras R1 and R3 with docked ligands by molecular modeling. A) Active site of SnoK docked with compound 11. B) Active site of SnoN with compound 12. The cosubstrate α-KG is shown in gray, and the iron (Fe) and water molecules as orange and red spheres, respectively. The conserved tyrosine (Y68 in SnoK/Y74 in SnoN) describing the binding site near the acidic residue (D106 in SnoK/E112 in SnoN) is not shown for clarity. The amino acid residues interacting with compound 12 are shown as sticks and the key hydrophobic (< 5 Å distance) and polar interactions as dashed lines. C) Comparison of the R1 loop (cyan) of SnoK R1 (green) and wild-type SnoN (purple). D) Comparison of the active site cavities of SnoK R3 chimera (magenta) and native SnoK (green). The reduced size of SnoK R3 active site cavity in comparison with the wild-type SnoK resulted in the loss of carbocyclization activity.

We further analyzed the R1 region and found out that the loop containing F245SnoN and L246SnoN was stabilized by a structurally conserved hydrogen bond between the main chain and a conserved aspartate (D248SnoN in Figure 14C). In order to validate the significance of this key area, we narrowed down the original R1 region

from TGIAFLDDLPGTGADPLREGA (241-261) to TGIAFL (241-246), which we referred to as R1.1 (Table 1, Figure 12, dark brown). We used this region to replace R1.1 of SnoK [GVQ (238-240), Table 1] and generated SnoK R1.1 chimeric enzyme. Our analysis of the enzymatic activity of SnoK R1.1 with **11** revealed an increase in the epimerization activity when compared with SnoK R1 (Table 2 and Figure 4 of in publication III). The results obtained here established that the extension of the tip of the R1 loop and formation of the aromatic stacking interactions might have been the original evolutionary events that is responsible for the appearance of epimerization activity (Figure 15).

4.3.4.2 Importance of region R3 in carbocyclization activity

From our structural data, we deduced that region R3 have a critical role of fine-tuning the positioning of the sugar unit of the substrate. We predicted that the complete loss of enzymatic activity in the SnoK R3 chimera (Table 2 and Figure 4 in publication III) might have resulted from the reduced size of its active site cavity in comparison with the native SnoK. The S104SnoK/K110SnoN and D106SnoK/ E112SnoN replacements have incorporated longer side chains in the R3 region (Figure 14D). Consequently, the substrate (**11**) could not move further enough into the active site for the SnoK carbocyclization reaction to occur in any R3 chimeras of SnoK. In addition to the longer side-chain residues introduced in all R3 chimeras of SnoK, the absence of the hydroxyl group due to the Y103SnoK/F103SnoN replacement might have contributed to the loss of the carbocyclization activity. Our results established that the amino acid residues in region R3 are highly effective in eliminating the ancestral SnoK carbocyclization activity, but they are not critical for the SnoN epimerization reaction, as revealed by the activity of the SnoN R3 chimera (Table 2 and Figure 4 in publication III).

We concluded from our results that the carbocyclase SnoK is the ancestral form of the enzyme from which SnoN has evolved to catalyze stereoinversion at the neighboring carbon (Figure 15).

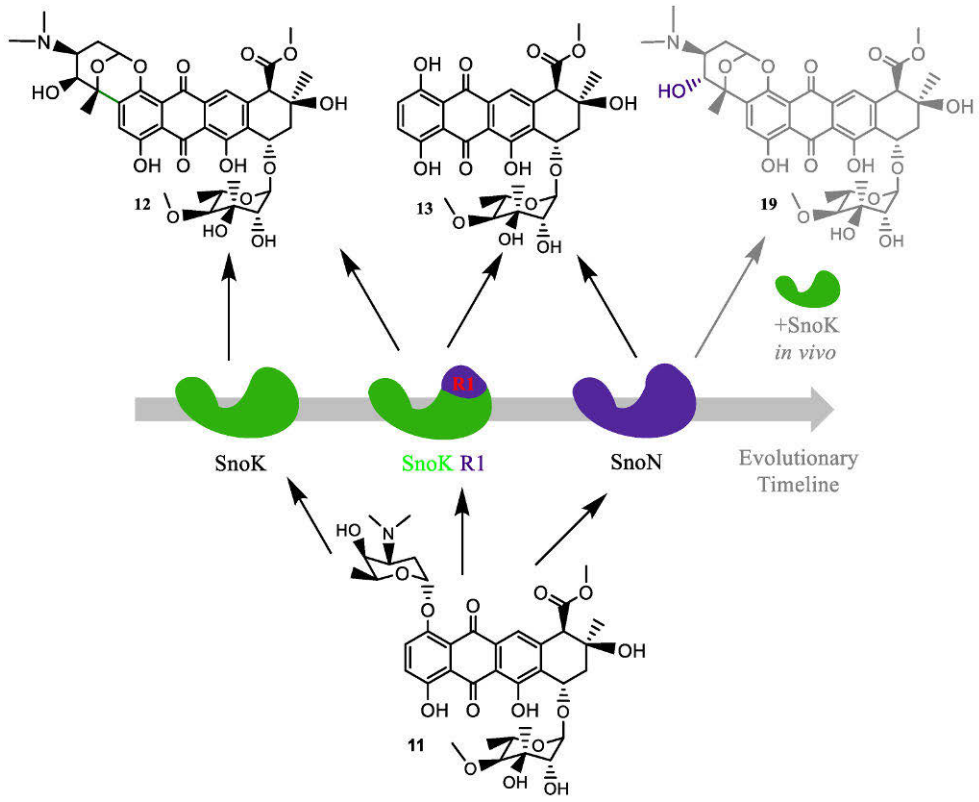


Figure 15: Illustration of the evolution event of gene duplication of SnoK to SnoN in nogalamycin biosynthetic pathway. The carbocyclase SnoK is the ancestral form of the enzyme from which the epimerase SnoN has evolved.

According to the findings of this study, the carbocyclase SnoK is the ancestral form of the enzyme from which SnoN evolved to catalyze stereoinversion at the adjacent carbon. The insertion of three residues at the C-terminus, which allow the substrate to be repositioned in front of the iron center, is likely the most important step in the emergence of epimerization activity. Changes in four amino acids around the iron center prevented the substrate from aligning correctly for the establishment of the C2–C5″ bond, resulting in the loss of the original carbocyclization activity. These results have provided detailed insights into the evolutionary events that have allowed *Streptomyces* soil bacteria to become the major source of antibiotics and antiproliferative agents.

4.4 Evolution of anthracycline methyltransferases (Article IV)

In anthracycline biosynthesis, late stage modifications such 4–*O*–methylation, 10–hydroxylation and 10–decarboxylation of the common aglycone scaffold are a source of structural diversity (Grocholski T *et al*., 2019; Grocholski T *et al*., 2015; Jansson A *et al*., 2005, Madduri *et al*., 1993). Remarkably, these diverse chemical transformations are catalyzed by related SAM methyltransferase-like proteins; DnrK, TamK, RdmB and CalMB. DnrK is a canonical *S*–adenosyl-L-methionine (SAM)–dependent methyltransferase that catalyzes 4–*O*–methylation in daunorubicin biosynthesis in *S. peucetius* (Madduri *et al*., 1993). Phylogenetic analysis of these methyltransferase-like proteins showed that DnrK preserved the native methylation activity, although the enzyme has evolved to catalyze an additional 10–decarboxylation (Figure 16). RdmB (53.0% sequence identity to DnrK) from the L-rhodomycin pathway in *S. purpurascens* and CalMB (56.7% sequence identity to RdmB) from *Streptomyces* sp. CcalMP8WS are 10–hydroxylases requiring SAM, molecular oxygen and a thiol reducing agent for activity (Jansson A *et al*., 2003 & 2005) (Figure 16). RdmB has evolved to lost its native methylation activity and in return, it has gained the ability to hydroxylate the C10 position. CalMB on the other hand, which is expected to be a 10–decarboxylase, has evolved to catalyze 10–hydroxylation like RdmB, while TamK from *S. Tsukubensis* catalyzes solely 10–decarboxylation (Figure 16).

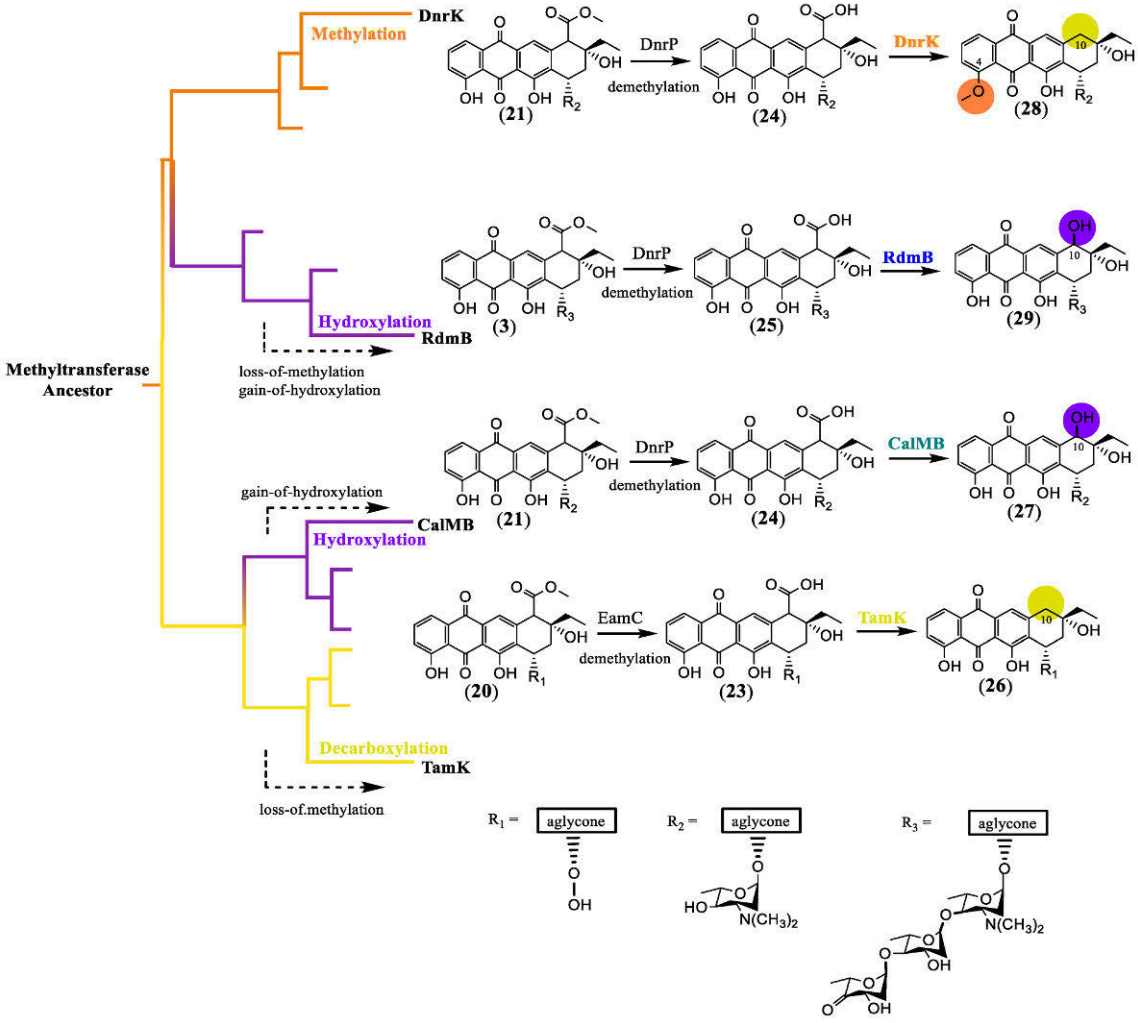


Figure 16: Evolution of the SAM-dependent methyltransferase family studied. The branch containing the 4–*O*–methyltransferase DnrK (orange-yellow) maintained the ancestral methylation activity, as well as moonlighting decarboxylation activity, the latter being responsible for the diversification observed in this family. At the opposite end of the family tree lie the 10–decarboxylases, such as TamK (yellow), that evolved to decarboxylate aglycone compounds as well as aclacinomycin T, losing their ancestral methylase activity. Convergent evolution has led to both these branches to gain 10–hydroxylation mechanisms, exemplified in RdmB (blue) and CalMB (green), together with broadening of substrate specificity.

4.4.1 Structure of wild type TamK

The structures of the 4–*O*–methyltransferase DnrK and the 10–hydroxylase RdmB have been solved previously, but to date no representatives of 10–decarboxylases have been available. We solved the structure of TamK at a resolution of 1.5 Å, which revealed that the overall structure consists of twenty-one helices and nine β–sheets (Figure 17A). The crystal structure of TamK contains the cosubstrate SAH but lacks the ligand, which made it difficult to understand the mechanism of the 10–decarboxylation. However, we could deduce the mechanism from the structure of DnrK TDDD, which is a DnrK chimera containing region R1 from TamK (R1 modulates enzymatic activity of these enzymes) (Figure 17B, yellow segment).

We crystallized this chimera in complex with the compound **21** and proposed a mechanism for 10-decarboxylation, where the formation of a 10–carbanion intermediate is the key step (Figure 17B). In this mechanism, a fully conserved arginine residue (R303 in DnrK and R306 in DnrK TDDD) (Figure 1 in publication IV), which is within hydrogen bonding distance from the 10–carboxyl group, is responsible for the initiation of the 10–decarboxylation reaction (Grocholski T *et al*., 2015; Jansson A *et al*., 2005). Structural analysis of DnrK TDDD showed that Y299 forms a polar contact with the carboxylate unit for the correct position of the leaving group while L303 and P344 (DnrK TDDD numbering) create a hydrophobic environment that destabilizes the ground state of the substrate, favouring the release of neutral carbon dioxide molecule (Figure 17B).

TamK, DnrK and RdmB are from the same protein family (Figure 16) and their overall fold is very similar (Figure 18A), although there are several hotspots of divergence (Figure 18B, red zones). Structural comparison between TamK and DnrK (64.7% sequence identity) reveals an r.m.s.d. difference of 3.108 Å for 331 Cα atoms. TamK and RdmB share 51.6% sequence identity and an r.m.s.d. difference of 3.322 Å for 331 Cα atoms (Figure 18B).

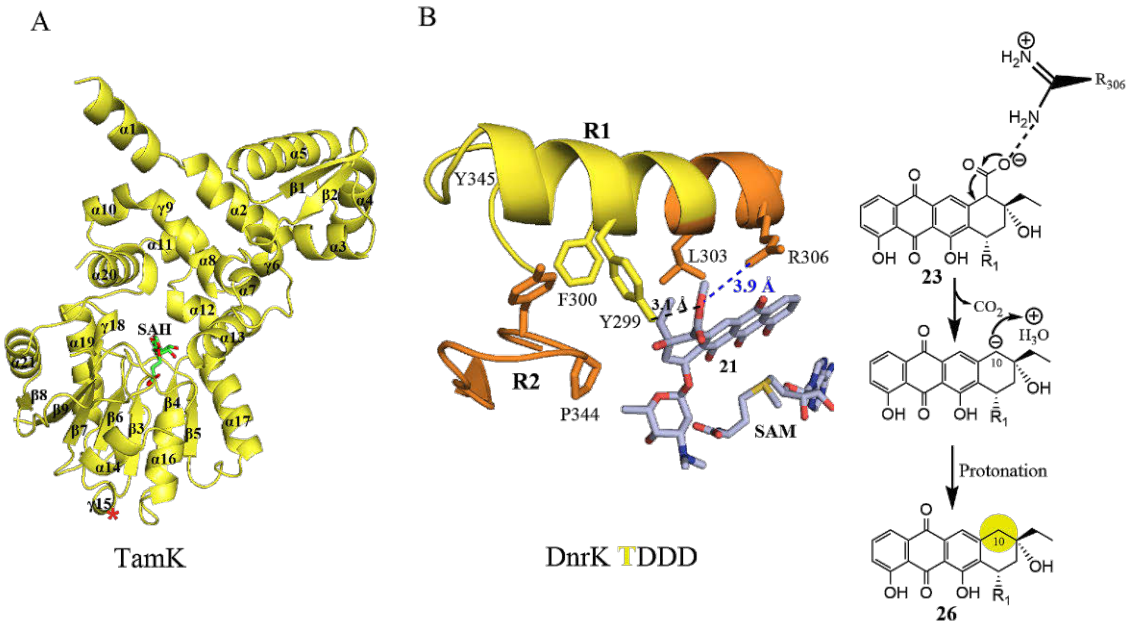


Figure 17: Schematic view of the monomer unit of TamK–SAH and mechanism of 10–decarboxylation. A) Crystal structure of wild type TamK. The bound cofactor (SAH) is shown in *green ball-and-stick* models and labeled. Secondary structure elements in the monomer are labeled. Native TamK contains twenty-one helices and nine β–sheets. The additional γ15 helix of native TamK is starred. B) Proposed mechanism of 10–decarboxylation. The crystal structure of DnrK TDDD showing the conserved R306 that initiates the formation of a 10–carbanion intermediate while L303 and P344 create a hydrophobic environment, which destabilizes the ground state of the substrate and favours the release of carbon dioxide.

TamK is composed of an N–terminal domain mainly constituted of helical structures except for two β–strands, a middle all-helical domain, and the C–terminal Rossman-like fold. The N–terminal appears to be the hybridization domain like in RdmB and DnrK. The C–terminal Rossman-like fold comprises of a central parallel β–sheet (β3–β9) surrounded by ten helices (α12–α21) (Figure 17A). The difference in the secondary structure elements among these proteins is that both TamK and DnrK have an additional helix (twenty-one helices and nine β–sheets) while RdmB has twenty helices and nine β–sheets. Like in RdmB and DnrK, the C–terminal domain of TamK also contains the binding site for the cofactor SAM with the conserved DLGGG*X*G fingerprint (Figure 17A and Figure 19). The substrate is expected to be positioned between the middle and C–terminal domains as it is shown in RdmB (Figure 19), and the residues from both of these domains are involved in binding the substrate.

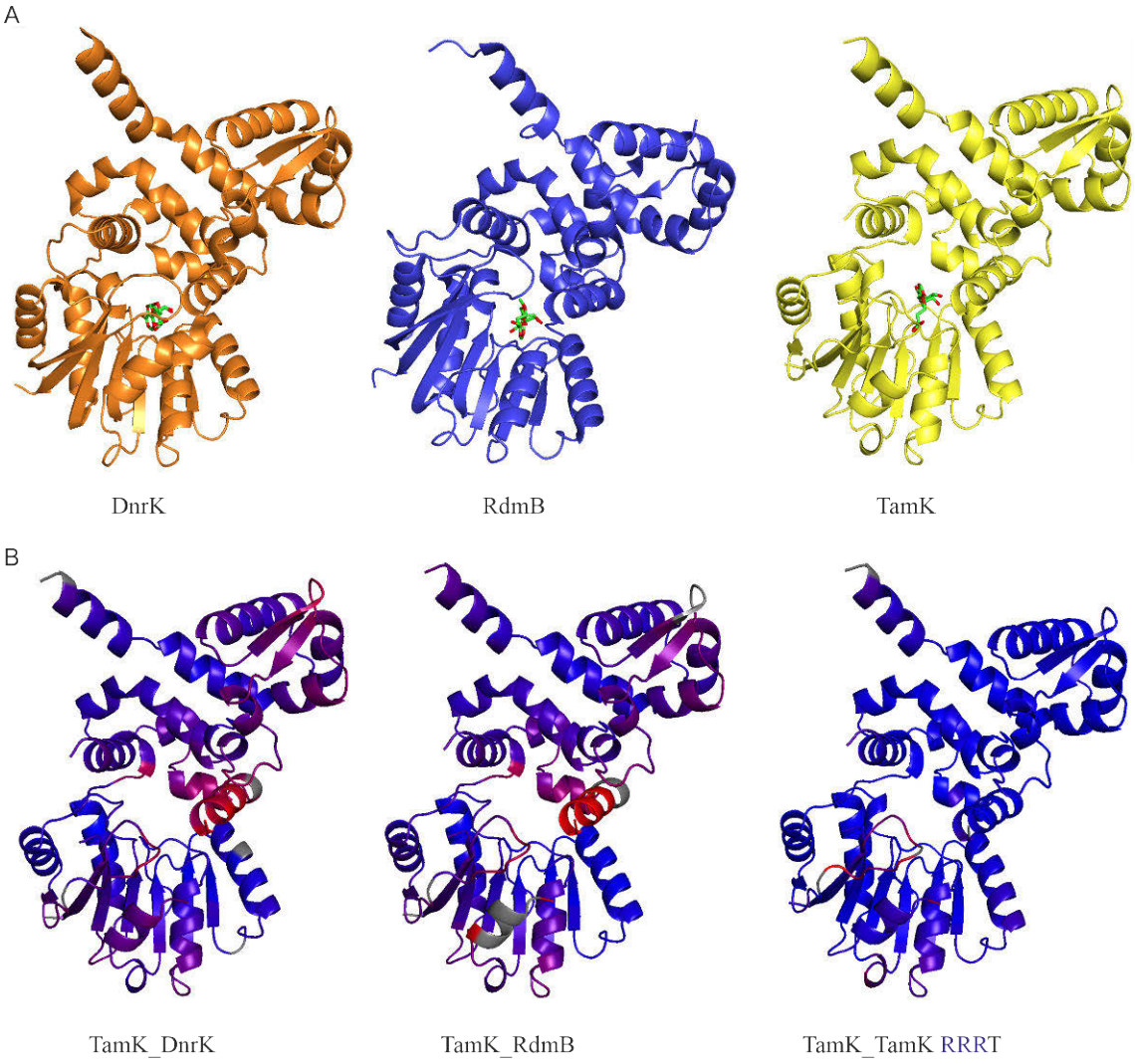


Figure 18: Native structures of the DnrK, RdmB and TamK. A) Structural similarity among DnrK, RdmB and TamK. Their overall fold is very similar but the catalyzed distinct reactions; DnrK is a 4–*O*–methylase, RdmB is a 10–hydroxylase while TamK in a 10–decarboxylase. B) Illustrates the differences among these proteins with Root Mean Square Deviation (r.m.s.d) of 3.108, 3.322 and 0.284 after superimposing TamK with DnrK, RdmB and TamK RRRT, respectively. The blue color specifies the minimum pairwise r.m.s.d and the red indicates the maximum. All unaligned residues are colored gray.

4.4.2 Chimeragenesis of anthracycline methyltransferases

We utilized extensive chimeragenesis (Dinis P *et al*., 2019) to probe factors that govern catalysis and substrate recognition among the SAM methyltransferase-like enzymes. We generated 24 chimeric enzymes using DnrK and TamK scaffolds and

sequences from the four proteins (DnrK, TamK, RdmB and CalMB). We selected four regions (R1–R4) from these enzymes and exchanged them in different combination using two protein scaffolds in order generate the chimeric enzymes. The influence of each region (R1–R4) (Figure 19) on catalysis and substrate specificity have previously been investigated. The first region (R1) is the major determinant for the type reaction catalyzed by the enzyme (Grocholski T *et al*., 2015). Regions R2 and R3 determine the substrate specificity of the enzyme. R2 appears to be mobile and allows the entry of the unique substrate into the active site of the enzyme while R3 wraps around the carbohydrate unit of the substrate and contribute to the substrate specificity of the enzyme. Region R4 is the binding site of the cosubstrate SAM and because SAM is essential for the methylation activity of this enzyme family, the orientation of the bound SAM is critical for catalysis. This SAM binding site is highly conserved in the enzyme family under study, with the exception of the 10–decarboxylase TamK that contain two mutations G188A/K189P (Figure 1C in publication IV). Hence, R4 plays a significant role in determining the methylation activity in these enzymes.

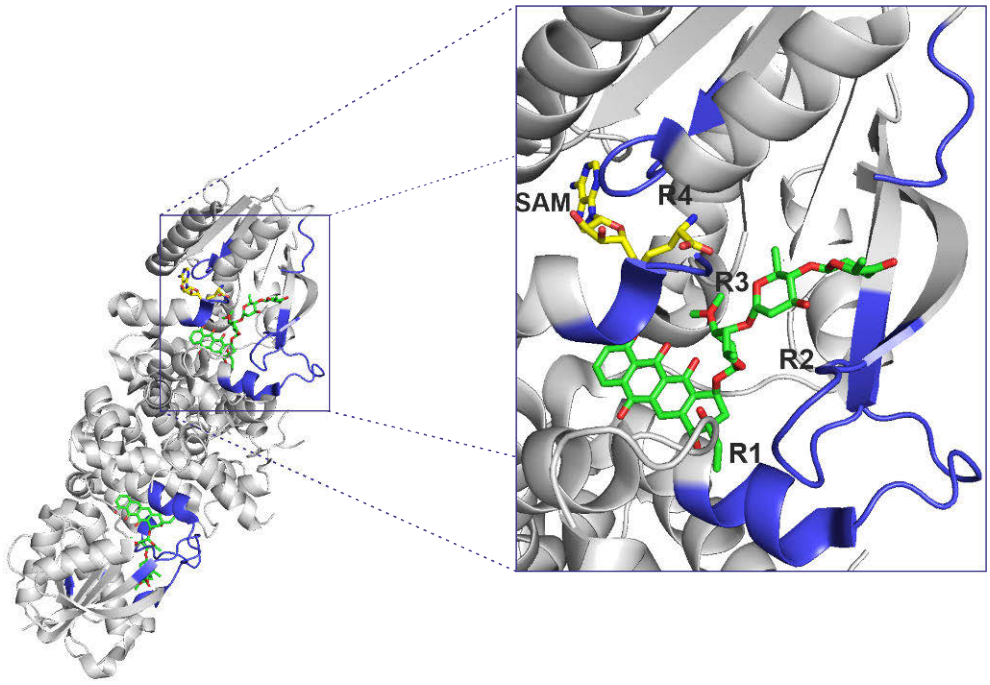


Figure 19: Overall fold of 10–hydroxylase RdmB, with the critical regions (R1–R4, purple) exchanged in the chimeragenesis studies. Region (R1) is the major determinant for the type reaction catalyzed by the enzyme. Regions R2 and R3 determine the substrate specificity of the enzyme. R2 is mobile and allows the entry of the substrate into the active site while R3 wraps around the carbohydrate unit of the substrate. Region R4 is the binding site of the cosubstrate SAM.

Due to the complex nature of the constructs, the chimeras are named in a way that the scaffold enzyme is firstly mentioned, followed by a sequence of four letters, corresponding to the origin of the regions R1–R4 (D for DnrK, R for RdmB, T for TamK and C for CalMB). To exemplify, wild type TamK would be termed TamK TTTT as all four regions are derived from TamK, while TamK RRRT is a chimera with a TamK scaffold, regions R1, R2 and R3 from RdmB and DnrK RTTD is a chimera with region R1 from RdmB and regions R2 and R3 from TamK.

4.4.3 Enzymatic activities of the chimeric proteins

We performed enzymatic assays of all the chimeras with three different substrates; the aklavinone (**20**); the monoglycosidic aclacinomycin T (**21**); and the triglycosidic aclacinomycin A (**3**) (Figure 2 in publication IV). An initial 15–methylesterase reaction either by DnrP (glycosylated substrates) (Dickens *et al*., 1997) or EamC (aklavinone substrate) (Grocholski T *et al*., 2019) was required in order to allow 10–decarboxylation and 10–hydroxylation to occur, which generated the following intermediates; **23**, **24** and **25** (Figure 16). We extracted these intermediates with the 10–carboxyl groups in the dark and subsequently used them as substrates for the chimeras together with the reducing agent dithiotreitol (DTT) and SAM.

Our results showed that 16 chimeras harbored diverse enzymatic activities and the majority of reaction products could be identified by comparison to standards obtained from previous studies (Figure 2 of in publication IV). The DnrK-based chimera DnrK RTTD and three other chimeras appeared to expand the substrates specificity for 4–O–methylation and 10–hydroxylation towards the aglycone (**20**) to produce a novel product (**30**) (Figure 20A). We confirmed the identity of this new compound (**30**) by comparing it with an acid-hydrolyzed known standard (**32**) using analytical HPLC (Figure 20B). In addition, we also noted that ten chimeras produced a novel product (**31**), which appeared to be an elimination product based on HR-MS measurement ($[M-H]^-$, ESI- obs. 335.0923, calc. 335.0925), when **23** was used as a substrate (Figure 2 in publication IV). The highest yields were detected with TamK RRRT (Figure 20A and Figure 2C in publication IV), which was crystallized during this study (Figure 21A). We performed this elimination reaction in large-scale to obtain sufficient material for structure elucidation of **31**. NMR-experiments (^1H, ^{13}C, HSQCDE, COSY and HMBC) revealed that an additional 9,10–elimination reaction had occurred in **31** (Figure 20A). The novel compound **31** differs from other aclacinomycins in the partially aromatic nature of the A-ring, where the chemical shifts at C9 (δ 148.6), C10 (δ 119.4) and H10 (δ 7.61) correspond to aromaticity, whereas those at C7 (δ 60.0), H7 (δ 5.38), OH7 (δ 2.27), C8 (δ 35.8) and H8 (δ 2.64, 2.71) do not.

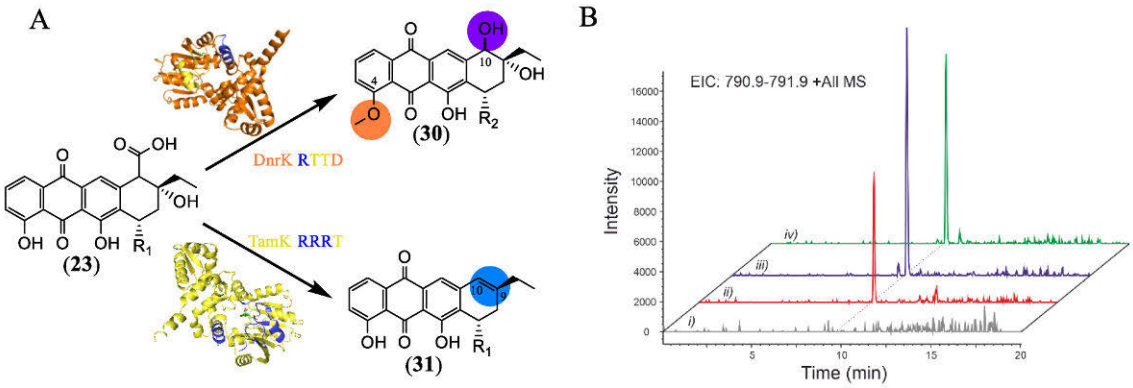


Figure 20: Enzymatic synthesis of novel compounds. A) Enzymatic synthesis of two novel compounds. DnrK RTTD gained the ability to act as a 4–*O*–methylase and at the same time a 10–hydroxylase towards the demethylated aglycone (**23**) to produce the novel product (**30**). The 10–decarboxylase TamK based chimera, TamK RRRT appeared to be an eliminase catalyzing a 9,10–elimination when **23** is used as the substrate. B) Analysis of the DnrK RTTD product (**30**) by LC–MS. i) EIC for the aklavinone substrate (**20**), ii) EIC for acid-hydrolyzed standard, iii) EIC for DnrK RTTD reaction product (**30**), iv) EIC for DnrK RTTD product (**30**) mixed with the acid-hydrolyzed standard.

4.4.4 Structure of the 9,10–elimimase TamK RRRT

In order to understand the 9,10–elimination reaction, we solved the crystal structure of TamK RRRT to a resolution of 1.9 Å (Figure 21A). The native 10–decarboxylase TamK and the novel chimeric 9,10–eliminase TamK RRRT share 94.1% sequence identity and the structures displayed an r.m.s.d. difference of 0.284 Å for 331 Cα atoms (Figure 18B). The differences between TamK WT and TamK RRRT arise from the exchange of three critical regions (termed R1, R2 and R3) from RdmB (Figure 19). These regions, firstly discussed in Grocholski, 2015, are critical for binding of the substrate in the active site. Region R1 is a helix/loop that closes the active site near the carboxyl subunit of the substrate, and was identified as responsible for modulating enzymatic activity; R2 is a loop that interacts with both R1 and R3, and can be seen in both closed and open conformations (Jansson, 2004); R3, together with R2 interacts with the sugar-binding domain (Figure 19). Despite these changes, the structures of TamK and TamK RRRT are very similar although the native enzyme have twenty-one helices while the chimera TamK RRRT have twenty helices. This chimera lacks the γ15 helix shown in the wild type TamK (red star, Figure 17A) because of the regions exchanged. It is possible more changes that are significant could be observed in the exchanged regions, but due to the lack of substrate in both TamK WT and TamK RRRT, both region R1 and R3 are disordered and difficult to build, and no conclusions can be drawn on the effects of chimeragenesis.

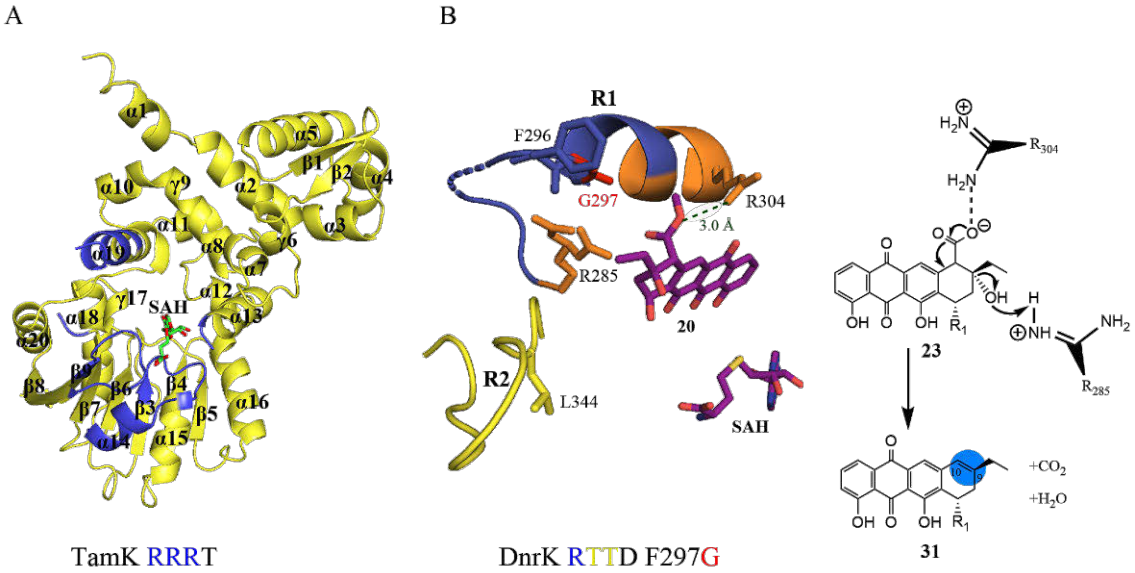


Figure 21: Schematic view of the monomer unit of TamK RRRT–SAH and mechanism of 9,10–elimination. A) The crystal structure of the 9,10–eliminase TamK RRRT bound with SAH (*green ball-and-stick* models and labeled). Secondary structure elements in the monomer are labeled. TamK RRRT has twenty helices and nine β–sheets. The purple regions in TamK RRRT are the exchanged regions (R1–R3) from RdmB (see Figure 19). B) Structure elucidation and proposed mechanism of the 9,10–elimination reaction. The crystal structure of DnrK RTTD F296G shows R285 in two distinct conformations, one of them bringing it closer to the substrate. This indicates a higher degree of freedom, which will likely allow R285 to interact with the substrate.

The mechanism of 9,10–elimination could be deduced from the structure of DnrK RTTD F297G, which was crystallized in complex with the anthracycline ligand (Figure 21B and Figure 4C in publication IV). The data demonstrated that the changes caused in region R1 by the single point mutation (F297G) affected the neighbouring residue R285, which is seen adopting two conformations, one of them that brings it closer to the substrate (Figure 21B). Our structural data suggested that the initial step of the 9,10–elimination reaction is the 10–decarboxylation by R304 (DnrK RTTD numbering). The next step involves the loss of the hydroxyl group at C9, where the interaction with R285 might have led to protonation and generation of a better leaving group (Figure 21B).

In this engineering work, we were able to show factors that govern gain-of-hydroxylation, loss-of-methylation, and substrate selection using a combination of structural and biochemical studies. The engineering effort extended the catalytic repertoire by generating a novel 9,10-eliminase, as well as 4–*O*–methylation and 10–decarboxylation of non-natural substrates. The study sheds light on how minor

alterations in biosynthetic enzymes can lead to an increase in the diversity of microbial natural products.

Table 3. Statistics of data collection and structure refinement.

	TamK	**TamK RRRT**
Data collection		
Beam line	ESRF ID23	ESRF ID23
Wavelength (Å)	0.87313	0.87313
Space group	C121	C121
Cell axes (Å)	122.835, 39.692, 97.977	122.774 40.163 99.214
Cell angles (degrees)	90.000, 114.076, 90.000	90.000, 115.059, 90.000
Resolution (Å)	56.07-1.51	36.76-2.11
Rmerge (%)	3.199 (60.1)	3.205 (28.39)
I/σI	15.69 (2.24)	10.78 (2.33)
Completeness (%)	96.51 (85.60)	98.92 (99.56)
Multiplicity	1.9 (1.8)	1.9 (2.0)
Number of reflections	125763 (10636)	49321 (4938)
Number of unique reflections	65678 (5773)	25374 (2498)
Wilson B-factor (Å^2)	24.12	40.62
Refinement		
Resolution (Å)	28.04-1.51 (1.57-1.5)	33.16-2.11 (2.18-2.11)
R_{work}/R_{free}	0.19/0.21	0.21/0.25
Number of non-hydrogen atoms/ mean B-factor (Å^2)	2784/30.64	2530/46.11
Protein	2511	2438
Water	251	92
R.m.s.d. bond lengths (Å)	0.006	0.004
R.m.s.d. bond angles (degrees)	0.97	0.78
Ramachandran plot		
Residues in favoured regions (%)	99.69	98.43
Residues in allowed regions (%)	0.31	1.57

Statistics for the highest-resolution shell are shown in parentheses.

5 Conclusion and Future Perspective

Anthracyclines have been the cornerstone in cancer treatment for several decades and are currently included worldwide in 500 clinical trials to explore better combinations. One limiting factor in their clinical use has been their severe side effects. Minor modifications in the structures of these compounds may have a significant improvement on the bioactivity while maintaining low levels of toxicity.

In my Ph.D. research, the complete biosynthesis and attachment of the unique nogalamine moiety of the anticancer agent nogalamycin (**1**) was established using nine enzymes. An essential modification for the bioactivity of nogalamycin, the C2″ hydroxylation, was also reported. In addition, the order of reaction of the final stages in nogalamycin biosynthesis was pointed out. We now know which proteins are responsible for the complete biosynthesis and attachment of nogalamine, and understand how this unusual amino sugar of nogalamycin is biosynthesized and atypically connected through both canonical *O*–glycosidic and C2–C5″ bond to the aglycone. Furthermore, we are now aware of the function of SnoT as a 2″–hydroxylase and that this modification precedes the C2–C5″ carbocyclization and C4″ epimerization by SnoK and SnoN, respectively.

In addition, this thesis gained insight into how anthracycline pathways may have been formed during evolution and provided explanations on how minor changes among homologous biosynthetic enzymes have led to the appearance of new catalysts. We uncovered the gene duplication event that led to the functional differentiation of the SnoK and SnoN in nogalamycin biosynthetic pathway. Through protein engineering by chimeragenesis, my work revealed how the functions of *S*-adenosyl methionine dependent methyltransferases have diversified in anthracycline biosynthetic pathways and generated novel compounds whose bioactivity have not yet been tested.

The advantage of chimeragenesis is that there is little or no solubility issues with the newly generated chimeras. This is mainly because the templates used for the generation of the chimeras are from pairs of naturally evolved enzymes that are soluble, and partly because of the way the chimeragenesis regions were selected. During selection, the suitable segment boundaries are determined by super-positioning the homologous proteins and extending the initial interchangeable region

in both N– and C–terminal directions until the sites where both the sequences and spatial positions are conserved.

The efforts to understand the biosynthesis and evolution of anthracyclines in a broader sense has been progressing in the past years, as was elaborated in the Literature Review section of my thesis. In addition to the biosynthesis of some common anthracyclines, the evolutionary events with or without gene duplication, which have contributed greatly to the chemical diversity of secondary metabolites have been discussed, and the progress is quite significant. The methods presented herein on how to generate novel compounds were well exploited in my thesis to generate new compounds. Therefore, my Ph.D. research has opened the gateway for achieving the long-awaited breakthrough of generating novel drugs with improved clinical efficacy and no severe side effects.

Despite the accomplishments of anthracyclines in cancer chemotherapy, improved agents are still urgently needed, because their use has been limited by severe side effects. With the complete knowledge of the biosynthetic steps and modifications of anthracyclines gained in my thesis, another fruitful and feasible future prospect is the possibility to engineer the biosynthetic enzymes themselves in order to produce novel compounds or modify the existing ones to improve their bioactivity while eliminating their toxicity. Up to date, only a few examples of such engineering exist in the field of natural product research. Moreover, knowing the functions of all enzymes involved in nogalamycin and related anthracycline biosynthesis have paved the path to generate novel anticancer agents by metabolic engineering.

The use of chimeragenesis as an engineering tool to study enzyme evolution and to generate novel catalysts as presented by my work, has provided great ideas and motivations in the long-term for the generation of novel drugs. It will be great to see the enzymes and compounds that we have studied as components in a well-designed machineries or engineered strains when new anticancer agents are created through metabolic engineering.

Acknowledgemets

My work was performed solely at the Department of Biochemistry, University of Turku. The Academy of Finland, Turku University Foundation and the Finnish Academy of Science and Letters financially supported my study. I wish to thank the Turku University Foundation, Computer Life Science – Åbo Akademi University and UTUGS for travel grants. I am thankful for all financial support that I received during my study.

I want to thank Professor Mikko Metsä-Ketelä for his excellent supervision skills. I also wish to appreciate the support from the members of my advisory committee; Prof. Lari Lehtiö, Dr. Jarmo Käpylä and Assistant Prof. Pauli Kallio. Thank you all for the yearly guidance and advice on how to successfully complete my study within the prescribed period.

I would like to thank the Dean and the Head of the Department for providing the work environment and equipment.

Great thanks to my opponent Prof. Steven Van Lanen and reviewers; Dr. Rajaram Venkatesan and Prof. Silvan Scheller for their time and efforts invested in my thesis.

Special thanks to Vilja and Kaisa for guiding me through the purification and analysis of metabolites at the beginning of my thesis. Thank you Dinis for your brillant ideas and great support with crystallography. I also wish to thank all my co-authors from the University of Turku, Åbo Akademi University, Turku-Finland and from the University of Novi Sad, Serbia for the signification role played in our publications.

Thanks to Anu Hirvensalo, Heli Kalevo, Hannele Heinonen and Satu Jasu for the technical support. Thank you Petja for assisting with CD measurements and LC–MS at the Chemistry Department. I gratefully acknowledge the access to the synchrotron radiation at the European Synchrotron Radiation Facility and the MAX IV X–ray Facility Lund, Sweden.

To all my current and former lab mates, I thank you for making everyday life at the university easy and good.

List of References

Alexeev I., Sultana A., Mäntsälä P., Niemi J. & Schneider G. (**2007**) Aclacinomycin oxidoreductase (AknOx) from the biosynthetic pathway of the antibiotic aclacinomycin is an unusual flavoenzyme with a dual active site. *Proc Natl Acad Sci USA. 104*, 6170–6175.

Balabanova L., Golotin V., Podvolotskaya A. & Rasskazov V. (**2015**) Genetically modified proteins: Functional improvement and chimeragenesis. *Bioengineered 6***,** 262–274.

Bartel P.L., Zhu C.B., Lampel J.S., Dosch D.C., Connors N.C., Strohl W.R., Beale J.M Jr., Floss H.G. (**1990**) Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions. *J Bacteriol. 172*, 4816–4826.

Beinker P., Lohkamp B., Peltonen T., Niemi J., Mäntsälä P. & Schneider G. (**2006**) Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics. *J Mol Biol. 359*, 728–740.

Borders C.L. Jr., Broadwater J.A., Bekeny PA., Salmon J.E., Lee A.S., Eldridge A.M. & Pett V.B. (**1994**) A structural role for arginine in proteins: multiple hydrogen bonds to backbone carbonyl oxygens. *Protein Sci. 3*, 541–548.

Bornscheuer U.T. & Pohl M. (**2001**) Improved biocatalysys by directed evolution and rational protein design. *Biocatal. Biotransformation 5*, 137–143.

Brockmann H. & Brockmann H. Jr. (**1963**) Rhodomycine, VIII; Antibiotica aus Actinomyceten, L. δ-Rhodomycinon. *Chem Ber 96*, 1771.

Burton S.G., Cowan D.A. & Woodley J.M. (**2002**) The search for the ideal biocatalyst. *Nat. Biotechnol. 20*, 37–45.

Caldara-Festin G., Jackson D.R., Barajas J.F., Valentic T.R., Patel A.B., Aguilar S., Nguyen M., Vo M., Khanna A., Sasaki E., Liu H.W. & Tsai S.C. (**2015**) Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases. *Proc Natl Acad Sci USA. 112*, 6844–6851.

Cantú Reinhard F.G., DuBois J.L. & de Visser S.P. (**2018**) Catalytic Mechanism of Nogalamycin Monoxygenase: How Does Nature Synthesize Antibiotics without a Metal Cofactor? *J Phys Chem B. 122*, 10841–10854.

Chu D.T.W (**1995**) Section Review Anti-infectives: Recent developments in 14– and 15–membered macrolides. *Exp Opin Invest Drugs 4*, 65–94.

Chung J.Y., Fujii I., Harada S., Sankawa U. & Ebizuka Y. (**2002**) Expression, purification, and characterization of AknX anthrone oxygenase, which is involved in aklavinone biosynthesis in *Streptomyces galilaeus*. *J Bacteriol. 184*, 6115–6122.

Claesson M., Siitonen V., Dobritzsch D., Metsä-Ketelä M. & Schneider G. (**2012**) Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in *Streptomyces nogalater*. *FEBS J 279*, 3251–3263.

Conant G.C. & Wolfe K.H. (**2008**) Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet. 9*, 938–950.

Copley S.D. (**2015**) An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci. 40*, 72–78.

Crump M.P., Crosby J., Dempsey C.E., Parkinson J.A., Murray M., Hopwood D.A. & Simpson T.J. (**1997**) Solution structure of the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2). *Biochemistry 36*, 6000–6008.

Daughtry K.D., Xiao Y., Stoner-Ma D., Cho E., Orville A.M., Liu P. & Allen K.N. (**2012**) Quaternary ammonium oxidative demethylation: X-ray crystallographic, resonance Raman, and UV-visible spectroscopic analysis of a Rieske-type demethylase. *J Am Chem Soc. 134*, 2823−2834.

Davies C., Heath R.J., White S.W. & Rock C.O. (**2000**) The 1.8 A crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from escherichia coli. *Structure. 8*, 185–195.

Des Marais D.L. & Rausher M.D. (**2008**) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature 454*, 762–765.

Dickens M.L., Priestley N.D. & Strohl W.R. (**1997**) In vivo and in vitro bioconversion of epsilon-rhodomycinone glycoside to doxorubicin: functions of DauP, DauK, and DoxA. *J Bacteriol. 179*, 2641–2650.

Dickens M.L., Ye J. & Strohl W.R. (**1995**) Analysis of clustered genes encoding both early and late steps in daunomycin biosynthesis by *Streptomyces sp*. strain C5. *J Bacteriol. 177*, 536–543.

Dickens M.L., Ye J. & Strohl W.R. (**1996**) Cloning, sequencing, and analysis of aklaviketone reductase from *Streptomyces sp*. strain C5. *J Bacteriol. 178*, 3384–3388.

Dinis P., Wandi B.N., Grocholski T. & Metsä-Ketelä M (**2019**) Chimeragenesis for biocatalysis. in Biomass, Biofuels, Biochemicals: Advances in Enzyme Technology, *Elsevier*. *1*, 389–418. doi:10.1016/B978-0-444-64114-4.00014-5

Dreier J. & Khosla C. (**2000**) Mechanistic analysis of a type II polyketide synthase. Role of conserved residues in the beta-ketoacyl synthase-chain length factor heterodimer. *Biochemistry. 39*, 2088–2095.

Elshahawi S.I., Shaaban K.A., Kharel M.K. & Thorson J.S. (**2015**) A comprehensive review of glycosylated bacterial natural products. *Chem Soc Rev. 44*, 7591–697.

Emerald E.S., MacHale L.T., Szilagyi R.K. & DuBois J.L. (**2019**) How Chemical Environment Activates Anthralin and Molecular Oxygen for Direct Reaction. *J Org Chem. 85*, 1315–1321.

Emsley P., Lohkamp B., Scott W.G. & Cowtan K. (**2010**) Features and development of Coot. *Acta Crystallogr D. 66,* 486–501.

Eriksen D.T., Lian J. & Zhao H. (**2014**) Protein design for pathway engineering. *J. Struct. Biol. 185*, 234–242.

Fetzner S. (**2002**) Oxygenases without requirement for cofactors or metal ions. *Appl Microbiol Biotechnol. 60*, 243–257.

Fewer D.P. & Metsä-Ketelä M. (**2019**) A pharmaceutical model for the molecular evolution of microbial natural products. *FEBS J. 287*, 1429–1449.

Finn R.D., Bateman A., Clements J., Coggill P., Eberhardt R.Y., Eddy S.R., Heger A., Hetherington K., Holm L., Mistry J., Sonnhammer E.L, Tate J. & Punta M. (**2014**) Pfam: the protein families database. *Nucleic Acids Res. 42*. 222–230.

Fischbach M.A., Walsh C.T. & Clardy J. (**2008**) The evolution of gene collectives: How natural selection drives chemical innovation. *Proc Natl Acad Sci USA 105*, 4601–4608.

Florova G., Kazanina G. & Reynolds K.A. (**2002**) Enzymes involved in fatty acid and polyketide biosynthesis in *Streptomyces glaucescens*: role of FabH and FabD and their acyl carrier protein specificity. *Biochemistry. 41*, 10462–10471.

Frederick C.A., Williams L.D., Ughetto G., van der Marel G.A., van Boom J.H., Rich A. & Wang A.H. (**1990**) Structural comparison of anticancer drug-DNA complexes: adriamycin and daunomycin. *Biochemistry. 29*, 2538–2549.

Futuyma D.J. & Kirkpatrick M. Evolution, Sinauer Associates Incorporated, 2017.

Gefflaut T., Blonski C., Perie J. & Willson M. (**1995**) Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects. *Prog Biophys Mol Biol. 63*, 301–340.

Gokhale R.S., Lau J., Cane D.E. & Khosla C. (**1998**) Functional orientation of the acyltransferase domain in a module of the erythromycin polyketide synthase. *Biochemistry. 37*, 2524–2528.

Grimm A., Madduri K., Ali A. & Hutchinson C.R. (**1994**) Characterization of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene. 151*, 1–10.

Grocholski T., Dinis P., Niiranen L., Niemi J. & Metsä-Ketelä M. (**2015**) Divergent evolution of an atypical *S*-adenosyl-l-methionine-dependent monooxygenase involved in anthracycline biosynthesis. *Proc Natl Acad Sci USA 112*, 9866–9871.

Grocholski T., Yamada K., Sinkkonen J., Tirkkonen H., Niemi J. & Metsä-Ketelä M (**2019**) Evolutionary trajectories for the functional diversification of anthracycline methyltransferases. *ACS Chem Biol 14*, 850–856.

Hadfield A.T., Limpkin C., Teartasin W., Simpson T.J., Crosby J. & Crump M.P. (**2004**) The crystal structure of the actIII actinorhodin polyketide reductase: proposed mechanism for ACP and polyketide binding. *Structure. 12*, 1865–1875.

Härle J., Günther S., Lauinger B., Weber M., Kammerer B., Zechel D.L., Luzhetskyy A. & Bechthold A. (**2011**) Rational design of an aryl-C-glycoside catalyst from a natural product O-glycosyltransferase. *Chem Biol. 18*, 520–530.

Harms M. J. & Thornton J.W. (**2013**) Evolutionary biochemistry: Revealing the historical and physical causes of protein properties. *Nat. Rev. Genet. 14*, 559–571.

Hautala A., Torkkell S., Räty K., Kunnari T., Kantola J., Mantsälä P., Hakala J. & Ylihonko K. (**2003**) Studies on a second and third ring cyclization in anthracycline biosynthesis. *J Antibiot (Tokyo). 56*, 143–153.

Herr C.Q. & Hausinger R.P. (**2018**) Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases. *Trends Biochem Sci*. *43*, 517–532.

Hertweck C., Luzhetskyy A., Rebets Y. & Bechthold A. (**2007**) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep. 24*, 162–190.

Höcker B. (**2013**) Engineering chimaeric proteins from fold fragments: 'hopeful monsters' in protein design. *Biochem. Soc. Trans. 41*, 1137–1140.

Hopwood D.A. (**1997**) Genetic Contributions to Understanding Polyketide Synthases. *Chem Rev. 97*, 2465–2498.

Hutchinson C.R. (**1997**) Biosynthetic Studies of Daunorubicin and Tetracenomycin C. *Chem Rev. 97*, 2525-2536.

Jansson A., Niemi J., Lindqvist Y., Mäntsälä P. & Schneider G. (**2003**) Crystal structure of aclacinomycin-10-hydroxylase, a *S*-adenosyl-L-methionine-dependent methyltransferase homolog involved in anthracycline biosynthesis in *Streptomyces purpurascens*. *J Mol Biol. 334*, 269–280.

Jansson A., Koskiniemi H., Erola A., Wang J., Mäntsälä P., Schneider G. & Niemi J. (**2005**) Aclacinomycin 10-hydroxylase is a novel substrate-assisted hydroxylase requiring S-adenosyl-L-methionine as cofactor. *J Biol Chem*. *280*, 3636–3644.

Jensen P.R. (**2016**) Natural products and the gene cluster revolution. *Trends Microbiol*. *24*, 968–977.

Joshi A.K., Witkowski A. & Smith S. (**1998**) The malonyl/acetyltransferase and beta-ketoacyl synthase domains of the animal fatty acid synthase can cooperate with the acyl carrier protein domain of either subunit. *Biochemistry. 37*, 2515–2523.

Kallio P., Sultana A., Niemi J., Mäntsälä P. & Schneider G. (**2006**) Crystal structure of the polyketide cyclase AknH with bound substrate and product analogue: implications for catalytic mechanism and product stereoselectivity. *J Mol Biol. 357*, 210–220.

Kantola J., Blanco G., Hautala A., Kunnari T., Hakala J., Mendez C., Ylihonko K., Mäntsälä P. & Salas J. (**1997**) Folding of the polyketide chain is not dictated by minimal polyketide synthase in the biosynthesis of mithramycin and anthracycline. *Chem Biol. 4*, 751–755.

Kantola J., Kunnari T., Hautala A., Hakala J., Ylihonko K. & Mäntsälä P. (**2000**) Elucidation of anthracyclinone biosynthesis by stepwise cloning of genes for anthracyclines from three different *Streptomyces spp*. *Microbiology (Reading). 146*, 155–163.

Katz L. & Donadio S. (**1993**) Polyketide synthesis: prospects for hybrid antibiotics. *Annu Rev Microbiol. 47*, 875–912.

Keatinge-Clay A.T., Maltby D.A., Medzihradszky K.F., Khosla C. & Stroud R.M. (**2004**) An antibiotic factory caught in action. *Nat Struct Mol Biol. 11*, 888–893.

Keatinge-Clay A.T., Shelat A.A., Savage D.F., Tsai S.C., Miercke L.J., O'Connell J.D. 3rd, Khosla C. & Stroud R.M. (**2003**) Catalysis, specificity, and ACP docking site of *Streptomyces coelicolor* malonyl-CoA:ACP transacylase. *Structure. 11*, 147–154.

Kendrew S.G., Hopwood D.A. & Marsh E.N. (**1997**) Identification of a monooxygenase from *Streptomyces coelicolor* A3(2) involved in biosynthesis of actinorhodin: purification and characterization of the recombinant enzyme. *J Bacteriol. 179*, 4305–4310.

Kendrew S.G., Katayama K., Deutsch E., Madduri K. & Hutchinson C.R. (**1999**) DnrD cyclase involved in the biosynthesis of doxorubicin: purification and characterization of the recombinant enzyme. *Biochemistry. 38*, 4794–4799.

Khan T. & Ghosh I. (**2015**) Modularity in protein structures: Study on all-alpha proteins. *J. Biomol. Struct. Dyn. 33*, 2667–2681.

Khosla C., Gokhale R.S., Jacobsen J.R. & Cane D.E. (**1999**) Tolerance and specificity of polyketide synthases. *Annu Rev Biochem. 68*, 219–253.

Korman T.P., Hill J.A., Vu T.N. & Tsai S.C. (**2004**) Structural analysis of actinorhodin polyketide ketoreductase: cofactor binding and substrate specificity. *Biochemistry. 43*, 14529–14538.

Kremer L., Nampoothiri K.M., Lesjean S., Dover L.G., Graham S., Betts J., Brennan P.J., Minnikin D.E., Locht C. & Besra G.S. (**2001**) Biochemical characterization of acyl carrier protein (AcpM) and malonyl-CoA:AcpM transacylase (mtFabD), two major components of Mycobacterium tuberculosis fatty acid synthase II. *J Biol Chem. 276*, 27967–27974.

Kulowski K., Wendt-Pienkowski E., Han L., Yang K., Vining L.C. & Hutchinson C.R. (**1999**) Functional characterization of the jadl gene as a cyclase forming angucyclinones. *J Am Chem soc 121*, 1786.

Kunnari T., Ylihonko K., Hautala A., Klika K.D., Mäntsälä P. & Hakala J. (**1999**) Incorrectly folded aromatic polyketides from polyketide reductase deficient mutants. *Bioorg Med Chem Lett. 9*, 2639–2642.

Lambalot R.H., Gehring A.M., Flugel R.S., Zuber P., LaCelle M., Marahiel M.A., Reid R., Khosla C. & Walsh C.T. (**1996**) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol. 3*, 923–936.

Lomovskaya N., Doi-Katayama Y., Filippini S., Nastro C., Fonstein L., Gallo M., Colombo A.L. & Hutchinson C.R. (**1998**) The *Streptomyces peucetius* dpsY and dnrX genes govern early and late steps of daunorubicin and doxorubicin biosynthesis. *J Bacteriol. 180*, 2379–2386.

Lu W., Leimkuhler C., Gatt G. J., Kruger R.G., Oberthü M., Kahne, D. & Walsh C.T. (**2005**) AknT is an activating protein for the glycosyltransferase AknS in L-aminodeoxysugar transfer to the aglycone of aclacinomycin A. *Chem. Biol. 12*, 527−534.

Madduri K., Torti F., Colombo A.L. & Hutchinson C.R. (**1993**) Cloning and sequencing of a gene encoding carminomycin 4-O-methyltransferase from Streptomyces peucetius and its expression in Escherichia coli. *J Bacteriol. 175*, 3900–3904.

Madduri K. & Hutchinson C.R. (**1995**) Functional characterization and transcriptional analysis of a gene cluster governing early and late steps in daunorubicin biosynthesis in *Streptomyces peucetius*. *J Bacteriol. 177*, 3879–3884.

Malik V.S. (**1980**) Microbial secondary metabolism. *Trend Biochem Sci. 5*, 68–72.

Matsuda Y., Bai T., Phippen C.B.W., Nødvig C.S., Kjærbølling I., Vesth T.C., Andersen M.R., Mortensen U.H., Gotfredsen C.H., Abe I. & Larsen T.O. (**2018**) Novofumigatonin biosynthesis involves a non-heme iron-dependent endoperoxide isomerase for orthoester formation. *Nat Commun. 9*, 2587.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C. & Read R.J. (**2007**) Phaser crystallographic software. *J Appl Crystallogr 40,* 658–674.

McDaniel R., Ebert-Khosla S., Hopwood D.A. & Khosla C. (**1993**) Engineered biosynthesis of novel polyketides. *Science. 262*, 1546–1550.

Melodie M.M., Ellis E.S., Carney T.J., Brushett F.R., DuBois J.L. (**2019**) How a cofactor-free protein environment lowers the barrier to O2 reactivity. *J Biol Chem 294*, 3661–3669.

Metsä-Ketelä M. (**2017**) Evolution inspired engineering of antibiotic biosynthesis enzymes. *Org Biomol Chem 15*, 4036–4041.

Metsä-Ketelä M., Niemi J., Mäntsälä P. & Schneider G. (**2008**) Anthracycline Biosynthesis: Genes, Enzymes and Mechanisms In Anthracycline Chemistry and Biology I: Biological Occurrence and Biosynthesis, Synthesis and Chemistry (Krohn K, ed), pp. 101–140. Springer-Verlag, Berlin/Heidelberg.

Moncrieffe M.C., Fernandez M.-J., Spiteller D., Matsumura H., Gay N.J., Luisi B.F. & Leadlay P.F. (**2012**) Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. *J. Mol. Biol. 415*, 92−101.

Morse D.E. & Horecker B.L. (**1968**) The mechanism of action of aldolases. *Adv Enzymol Relat Areas Mol Biol. 31*, 125–181.

Mrabet N.T., Van den Broeck A., Van den brande I., Stanssens P., Laroche Y., Lambeir A.M., Matthijssens G., Jenkins J., Chiadmi M., van Tilbeurgh H., *et al.* (**1992**) Arginine residues as stabilizing elements in proteins. *Biochemistry. 31*, 2239–2253.

Murzin A.G., Brenner S.E., Hubbard T. & Chothia C. (**1995**) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol. 247*, 536–540.

Newman D.J. & Cragg G.M. (**2012**) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod 75*, 311–335.

Newman D.J. & Cragg G.M. (**2016**) Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod 79*, 629–661.

Nitiss J.L. (**2009**) Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer. 9*, 338–350.

Pan H., Tsai S.C., Meadows E.S., Miercke L.J., Keatinge-Clay A.T., O'Connell J., Khosla C. & Stroud R.M. (**2002**) Crystal structure of the priming beta-ketosynthase from the R1128 polyketide biosynthetic pathway. *Structure. 10*, 1559–1568.

Pang B., Qiao X., Janssen L., *et al*. (**2013**) Drug-induced histone eviction from open chromatin contributes to the chemotherapeutic effects of doxorubicin. *Nat Commun. 4*, 1908.

Parris K.D., Lin L., Tam A., Mathew R., Hixon J., Stahl M., Fritz C.C., Seehra J. & Somers W.S. (**2000**) Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure. 8*, 883–895.

Patrikainen P., Niiranen L., Thapa K., Paananen P., Tähtinen P., Mäntsälä P., Niemi J. & Metsä-Ketelä M. (**2014**) Structure-based engineering of angucyclinone 6-ketoreductases. *Chem Biol. 21*, 1381–1391.

Price A.C., Zhang Y.M., Rock C.O. & White S.W. (**2004**) Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG. *Structure. 12*, 417–428.

Rabe P., Kamps J.J.A.G., Schofield C.J. & Lohans C.T. (**2018**) Roles of 2-oxoglutarate oxygenases and isopenicillin N synthase in b-lactam biosynthesis. *Nat Prod Rep*. *35*, 735–756.

Robbins T., Liu Y.C., Cane D.E. & Khosla C. (**2016**) Structure and mechanism of assembly line polyketide synthases. *Curr Opin Struct Biol*. *41*, 10–18.

Räty K., Kantola J., Hautala A., Hakala J., Ylihonko K. & Mäntsälä P. (**2002**) Cloning and characterization of *Streptomyces galilaeus* aclacinomycins polyketide synthase (PKS) cluster. *Gene. 293*, 115–122.

Räty K., Kunnari T., Hakala J., Mäntsälä P. & Ylihonko K. (**2000**) A gene cluster from *Streptomyces galilaeus* involved in glycosylation of aclarubicin. *Mol Gen Genet. 264*, 164–172.

Schneider G. (**2005**) Enzymes in the biosynthesis of aromatic polyketide antibiotics. *Curr Opin Struct Biol. 15*, 629–636.

Sciara G., Kendrew S.G., Miele A.E., Marsh N.G., Federici L., Malatesta F., Schimperna G., Savino C. & Vallone B. (**2003**) The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *EMBO J. 22*, 205–215.

Siitonen V., Blauenburg B., Kallio P., Mäntsälä P. & Metsä-Ketelä M. (**2012a**) Discovery of a two-component monooxygenase SnoaW/SnoaL2 involved in nogalamycin biosynthesis. *Chem Biol 19*, 638–646.

Siitonen V., Claesson M., Patrikainen P., Aromaa M., Mäntsälä P., Schneider G. & Metsä-Ketelä M. (**2012b**) Identification of late-stage glycosylation steps in the biosynthetic pathway of the anthracycline nogalamycin. *ChemBioChem 13*, 120–128.

Siitonen V., Nji Wandi B., Törmänen A.P. & Metsä-Ketelä M. (**2018**) Enzymatic Synthesis of the C-Glycosidic Moiety of Nogalamycin R. *ACS Chem Biol. 13*, 2433–2437.

Siitonen V., Selvaraj B., Niiranen L., Lindqvist Y., Schneider G. & Metsä-Ketelä M. (**2016**) Divergent non-heme iron enzymes in the nogalamycin biosynthetic pathway. *Proc Natl Acad Sci USA 113*, 5251–5256.

Sikosek T., Chan H. S. & Bornberg-Bauer E. (**2012**) Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness. *Proc. Natl. Acad. Sci. 109*, 14888–14893.

Siltberg-Liberles J. (**2011**) Evolution of structurally disordered proteins promotes neostructuralization. *Mol. Biol. Evol. 28*, 59–62.

Smith C.K, Davies G.J, Dodson E.J & Moore M.H. (**1995**) DNA-nogalamycin interactions: the crystal structure of d(TGATCA) complexed with nogalamycin. *Biochemistry*. *34*, 415–25.

Smith S. and Tsai S.C. (**2007**) The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep. 24*, 1041-1072.

Stevenson G., Neal B., Liu D.N., Hobbs M., Packer N.H., Batley M., Redmond J.W., Lindquist L. & Reeves, P. (**1994**) Structure of the 0 Antigen of *Escherichia coli* K-12 and the Sequence of Its *r*lb Gene Cluster. *J. Bacteriol. 176*, 4144−4156.

Sultana A., Kallio P., Jansson A., Wang J.S., Niemi J., Mäntsälä P. & Schneider G. (**2004**) Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. *EMBO J. 23*, 1911–1921.

Szafranska A.E., Hitchman T.S., Cox R.J., Crosby J. & Simpson T.J. (**2002**) Kinetic and mechanistic analysis of the malonyl CoA:ACP transacylase from *Streptomyces coelicolor* indicates a single catalytically competent serine nucleophile at the active site. *Biochemistry. 41*, 1421–1427.

Tam H.K., Härle J., Gerhardt S., Rohr J., Wang G., Thorson J.S., Bigot A., Lutterbeck M., Seiche W., Breit B., Bechthold A. & Einsle O. (**2015**) Structural characterization of O- and C-glycosylating variants of the landomycin glycosyltransferase LanGT2. *Angew Chem Int Ed Engl. 54*, 2811–2815.

Tang Y., Tsai S.C. & Khosla C. (**2003**) Polyketide chain length control by chain length factor. *J Am Chem Soc. 125*, 12708–12709.

Temperini C., Cirilli M., Aschi M. & Ughetto G. (**2005**) Role of the amino sugar in the DNA binding of disaccharide anthracyclines: crystal structure of the complex MAR70/d(CGATCG). *Bioorg Med Chem. 13*, 1673–9.

Torkkell S., Kunnari T., Palmu K., Hakala J., Mäntsälä P. & Ylihonko K. (**2000**) Identification of a cyclase gene dictating the C-9 stereochemistry of anthracyclines from *Streptomyces nogalater*. *Antimicrob Agents Chemother. 44*, 396–399.

Walsh C.T., Chen H., Keating T.A., Hubbard B.K., Losey H.C., Luo L., Marshall C.G., Miller D.A. & Patel H.M. (**2001**) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol*. *5*, 525–534.

Weymouth-Wilson, A. C. (**1997**) The role of carbohydrates in biologically active natural products. *Nat. Prod. Rep*. *14*, 99−110.

Wohlert S.E., Wendt-Pienkowski E., Bao W. & Hutchinson C.R. (**2001**) Production of aromatic minimal polyketides by the daunorubicin polyketide synthase genes reveals the incompatibility of the heterologous DpsY and JadI cyclases. *J Nat Prod. 64*, 1077–1080.

Wong F.T. & Khosla C. (**2012**) Combinatorial biosynthesis of polyketides--a perspective. *Curr Opin Chem Biol. 16*, 117–123.

Woodley J.M. (**2013**) Protein engineering of enzymes for process applications. *Curr. Opin. Chem. Biol. 17*, 310–316.

Woodward R B, Logusch E, Nambiar K P, Sakan K, Ward D E, Au-Yeung B W, Balaram P, Browne L J, Card P J, Chen C H, et al. (**1981**) Asymmetric total synthesis of erythromycin. 3. Total synthesis of erythromycin. *J Am Chem Soc 103*, 3215–3217.

Ye J., Dickens M.L., Plater R., Li Y., Lawrence J. & Strohl W.R. (**1994**) Isolation and sequence analysis of polyketide synthase genes from the daunomycin-producing *Streptomyces sp*. strain C5. *J Bacteriol. 176*, 6270–6280.

Ylihonko K., Tuikkanen J., Jussila S., Cong L. & Mäntsälä P. (**1996**) A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and complementation of early-block mutations in the anthracycline pathway. *Mol Gen Genet. 251*, 113–120.

Zhang Z., Gong Y-K., Zhou Q., Hu Y., Ma H-M., Chen Y-S., Igarashi Y., Pan L. & Tang G-L. (**2017**) Hydroxyl regioisomerization of anthracycline catalyzed by a four-enzyme cascade. *Proc Natl Acad Sci USA 114*, 1554–1559.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2021

ISBN 978-951-29-8538-8 (PRINT)
ISBN 978-951-29-8539-5 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)