TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. A I OSA - TOM. 386

ASTRONOMICA - CHEMICA - PHYSICA - MATHEMATICA

TYPE II AROMATIC POLYKETIDE BIOSYNTHETIC TAILORING ENZYMES: DIVERSITY AND ADAPTATION IN *STREPTOMYCES* SECONDARY METABOLISM

by

Pauli Kallio

TURUN YLIOPISTO
Turku 2008

From the Department of Biochemistry and Food Chemistry
University of Turku
Turku, Finland

Supervised by

Docent Jarmo Niemi, Ph.D.
Department of Biochemistry and Food Chemistry
University of Turku
Turku, Finland

and

Professor Emeritus Pekka Mäntsälä, Ph.D.
Department of Biochemistry and Food Chemistry
University of Turku
Turku, Finland

Reviewed by

Professor Pirkko Heikinheimo, Ph.D.
Institute of Biotechnology
University of Helsinki
Helsinki, Finland

and

Professor Lloyd Ruddock, Ph.D.
Department of Biochemistry
University of Oulu
Oulu, Finland

Opponent

Professor Gregory L. Challis, Ph.D.
Department of Chemistry
University of Warwick
Coventry, United Kingdom

ISBN 978-951-29-3753-0 (PRINT)
ISBN 978-951-29-3754-7 (PDF)
ISSN 0082-7002
Painosalama Oy – Turku, Finland 2008

To my Tribe

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, referred to as I-VI in the text, as well as on additional unpublished data.

I **Sultana, A., Kallio, P., Jansson, A., Wang, J-S., Niemi, J., Mäntsälä, P. and Schneider, G. (2004)** Crystallization and preliminary crystallographic data of SnoaL, a polyketide cyclase in nogalamycin biosynthesis. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1118-1120.

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

III **Kallio, P., Sultana, A., Niemi, J., Mäntsälä, P. and 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.

IV **Koskiniemi, H., Metsä-Ketelä, M., Dobritzsch, D., Kallio, P., Korhonen, H., Mäntsälä, P., Schneider, G., Niemi, J. (2007)** Crystal structures of two aromatic hydroxylases involved in the early tailoring Ssteps of angucycline biosynthesis. *J. Mol. Biol.* **372**, 633-648.

V **Kallio, P., Liu, Z., Mäntsälä, P., Niemi, J., Metsä-Ketelä, M. (2008)** A nested gene in *Streptomyces* bacteria encodes a protein involved in quaternary complex formation. *J. Mol. Biol.* **375**, 1212-1221.

VI **Kallio, P., Liu, Z., Mäntsälä, P., Niemi, J. and Metsä-Ketelä, M. (2008)** Sequential action of two flavoenzymes PgaE and PgaM in angucycline biosynthesis: Chemoenzymatic synthesis of gaudimycin C. *Chemistry & Biology* **15**, 157-166.

ABSTRACT

Members of the bacterial genus *Streptomyces* are well known for their ability to produce an exceptionally wide selection of diverse secondary metabolites. These include natural bioactive chemical compounds which have potential applications in medicine, agriculture and other fields of commerce. The outstanding biosynthetic capacity derives from the characteristic genetic flexibility of *Streptomyces* secondary metabolism pathways: i) Clustering of the biosynthetic genes in chromosome regions redundant for vital primary functions, and ii) the presence of numerous genetic elements within these regions which facilitate DNA rearrangement and transfer between non-progeny species. Decades of intensive genetic research on the organization and function of the biosynthetic routes has led to a variety of molecular biology applications, which can be used to expand the diversity of compounds synthesized. These include techniques which, for example, allow modification and artificial construction of novel pathways, and enable gene-level detection of silent secondary metabolite clusters. Over the years the research has expanded to cover molecular-level analysis of the enzymes responsible for the individual catalytic reactions. *In vitro* studies of the enzymes provide a detailed insight into their catalytic functions, mechanisms, substrate specificities, interactions and stereochemical determinants. These are factors that are essential for the thorough understanding and rational design of novel biosynthetic routes.

The current study is a part of a more extensive research project (Antibiotic Biosynthetic Enzymes; www.sci.utu.fi/projects/biokemia/abe), which focuses on the post-PKS tailoring enzymes involved in various type II aromatic polyketide biosynthetic pathways in *Streptomyces* bacteria. The initiative here was to investigate specific catalytic steps in anthracycline and angucycline biosynthesis through *in vitro* biochemical enzyme characterization and structural enzymology. The objectives were to elucidate detailed mechanisms and enzyme-level interactions which cannot be resolved by *in vivo* genetic studies alone. The first part of the experimental work concerns the homologous polyketide cyclases SnoaL and AknH. These catalyze the closure of the last carbon ring of the tetracyclic carbon frame common to all anthracycline-type compounds. The second part of the study primarily deals with tailoring enzymes PgaE (and its homolog CabE) and PgaM, which are responsible for a cascade of sequential modification reactions in angucycline biosynthesis.

The results complemented earlier *in vivo* findings and confirmed the enzyme functions *in vitro.* Importantly, we were able to identify the amino acid -level determinants that influence AknH and SnoaL stereoselectivity and to determine the complex biosynthetic steps of the angucycline oxygenation cascade of PgaE and PgaM. In addition, the findings revealed interesting cases of enzyme-level adaptation, as some of the catalytic mechanisms did not coincide with those described for characterised homologs or enzymes of known function. Specifically, SnoaL and AknH were shown to employ a novel acid-base mechanism for aldol condenzation, whereas the hydroxylation reaction catalysed by PgaM involved unexpected oxygen chemistry. Owing to a gene-level fusion of two ancestral reading frames, PgaM was also shown to adopt an unusual quaternary sturucture, a non-covalent fusion complex of two alternative forms of the protein. Furthermore, the work highlighted some common themes encountered in polyketide biosynthetic pathways such as enzyme substrate specificity and intermediate reactivity. These are discussed in the final chapters of the work.

ABBREVIATIONS

AAME	aklanonic acid methyl ester
AKV	aklavinone
amu	atomic mass unit
CDA	calcium dependent antibiotic
CLF	chain length factor
DEAE	diethylamino ethanol
DMSO	dimethyl sulfoxide
ES	electro spray
FAD	flavin adenine dinucleotide
GST	glutathione S-transferase
HGT	horizontal gene transfer
HPLC	high performance liquid chromatography
KS	ketosynthase
LC-MS	liquid chromatography - mass spectrometry
Mb	megabase
MDR	multidrug resistance
MHBH	3-hydroxybenzoate hydroxylase
minPKS	minimal polyketide synthase complex
NADPH	nicotinamide adenine dinucleotide phosphate
NAME	nogalonic acid methyl ester
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
ORF	open reading frame
PDB	Protein Data Bank
pHBH	para-hydroxybenzoate hydroxylase
PHHY	phenol hydroxylase
PKS	polyketide synthase complex
QM	quinone methide
SDR	short-chain alcohol reductase/dehydrogenase
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sp.	species
TIR	terminal inverted repeat
TOF	time of flight
UV-Vis	ultraviolet-visible

1. INTRODUCTION

1.1 *Streptomyces* diversity and secondary metabolism

1.1.1 Background to antibiotic research

The beginning of the antibiotic research era was marked by the discovery of penicillin, a secondary metabolite of *Penicillium notatum* mold, by Alexander Fleming in 1928. Although it was not until over 10 years later the first successful human antibiotic treatments were carried out, realization of the potential of microbes to produce useful bioactive compounds and advances in mass-producing these products soon launched world-wide interest in natural product research. Massive international soil screening projects were carried out in Europe, North America and Japan through 1950's to 70's aiming to find producers of new effective drug molecules. One of the most prominent groups of microbial secondary metabolite producers proved to be common gram-positive soil bacteria of the genus *Streptomyces*. These bacteria are abundant inhabitants in soil, and display exceptional metabolic diversity. Despite their leading role over the "golden years" of antibiotic research, however, only a limited number of the thousands of isolated bioactive compounds proved pharmacologically effective and not too toxic for human use. Since then the focus has polarized towards other producers such as fungi, lichens and so-called rare actinomycete species. Yet even today *Streptomyces*-derived metabolites continue to uphold a central role in the clinical treatment of bacterial infections and various types of cancer. For decades *Streptomyces* have also been subjected to intense genetic study of the secondary metabolism pathways. Gradually the research has expanded towards protein-level study to reveal molecular mechanisms behind the biosynthetic diversity - to understand the limitations and possibilities of different technologies in the design of novel compounds with improved clinical effects and toxicity profiles.

1.1.2 Diversity of *Streptomyces* species

One gram of soil may contain over 1.5 x 10^{10} microbial cells (Torsvik *et al.*, 1990) emphasizing the concepts of competition and competitive advantage in this natural habitat. Based on mathematical models and genetic diversity, the number of different strains within such populations has been estimated to range from nearly 40 000 (Curtis *et al.*, 2002) to up to several millions species (Gans *et al.*, 2005) in pristine uncontaminated soil. Many of the most important non-fungal contributors to this diversity are members of a heterogeneous group of gram-positive bacteria actinomycetes (order *Actinomycetales*), which have a vital role in carbon turnover in nature as decomposers of organic matter such as cellulose, lignin and chitin. The largest and perhaps the most significant taxonomic sub-group of this order is the genus *Streptomyces* (family *Streptomycetaceae*), which is known and extensively studied for having a complex life cycle and an exceptional capacity to produce bioactive metabolites of potential medical, clinical and agricultural interest. Historically,

taxonomic classification of *Streptomyces* has been difficult due to extensive morphological, physiological and biochemical heterogeneity. Even with today's chemotaxonomic and molecular techniques combined with numerical taxonomic methods, distinction of taxa within the genus is not clear-cut, and the results vary depending on the approach and identification matrices used (Anderson & Wellington, 2001; Manfio *et al.*, 1995). In 1999 the number of validly assigned *Streptomyces* species has been stated to be over 450 (al-Tai *et al.*, 1999). The corresponding number is presently 542 and 38 subspecies as cited in the *Approved Lists of Bacterial Names* (Skerman *et al.*, 1980) or validly published in the *International Journal of Systematic Bacteriology* (IJSB) or in the *International Journal of Systematic and Evolutionary Microbiology* (IJSEM). Underlining the lack of systematic tools for specification and comparison, the number of taxa listed under the genus on the NCBI (National Center for Biotechnology Information; www.ncbi.nlm.nih.gov) taxonomy server is presently in several thousands.

1.1.3 *Streptomyces* life cycle and adaptation

Streptomyces are exceptionally well adapted to the competition and the quickly changing conditions of the soil. This is reflected in the complexity of their life cycle by bacterial standards. During the vegetative phase of their growth when nutrients are abundant the cells elongate and invade the soil by forming a network of branched filaments called mycelium. Environmental stress such as lack of nutrients triggers the formation of aerial hyphae (sporophores) which extend above the colony or into the air pockets in soil and eventually septate into separate spores (conidia). The molecular mechanisms regulating the morphological and chemical differentiation are still poorly understood. At the onset of periods of metabolic stress *Streptomyces* species synthesize and accumulate triacylglycerols as high-energy reserve compounds (Arabolaza *et al.*, 2008; Olukoshi & Packter, 1994), a capability very uncommon to prokaryotes outside the *Actinomycetales* order (Alvarez & Steinbuchel, 2002). Non-optimal growth conditions also set off a transition to secondary metabolism functions, including the induction of the production of various bioactive compounds which are excreted as a chemical defense against microbial rivals. These include a wealth of growth inhibitors and microbicidal compounds. In addition, metabolites which participate in active transport of scarce nutrients to provide competitional advantage are also synthesized. The number and functional range of the distinct compounds naturally produced by *Streptomyces* bacteria reflects their genetic potential and flexibility, and their substantial interest to humans.

1.1.4 *Streptomyces* as a source of bioactive metabolites

It has been estimated that of the excess one million reported natural products, over 50000 are of microbial origin (Berdy, 2005). About half of these have been associated with a certain cellular or molecular-level biological function and, consequently, are commonly referred to as "bioactive compounds". These include over 17000 so-called "classical antibiotics" with antimicrobial, anti-tumor and/or antiviral activities, as well as compounds such as growth promoters, immunosuppressants, herbicides and anti-

paracitic agents. Remarkably, about a third of all microbial bioactive compounds reported by the year 2002 (~7600 out of ~22000 metabolites) derive from *Streptomyces* species (Berdy, 2005). According to mathematical predictions, this is yet a minor fraction of the total number of different compounds potentially produced by the genus (Watve *et al.*, 2001).

Unequivocal classification of the metabolites found in *Streptomyces* is complicated by the vast structural and functional diversity, but most fall under one of the three major compound classes; peptide-type compounds, macrocyclic lactones and various quinone derivatives. If the metabolites are grouped based on their biosynthetic origin, one of the most significant entities are the polyketide compounds, which comprise a major share of natural products used in human medicine. These include compounds such as tetracycline (Chopra & Roberts, 2001; Zakeri & Wright, 2008) and erythromycin (McGuire *et al.*, 1952; Shah, 1998) effective against bacterial infections, the antifungal agents candicidin (Waksman *et al.*, 1965) and amphotericin (Abadi, 1995; Dutcher, 1968), anthracycline derivatives such as doxorubicin (Skovsgaard & Nissen, 1975) used in cancer treatment, and the immunosuppressant rapamycin (Saunders *et al.*, 2001; Sehgal, 2003; Vezina *et al.*, 1975) used to prevent organ transplant rejection. Polyketides can be subdivided into categories (type I, II and III) on the basis of their biosynthesis (see section 1.3 for a description of type II polyketide biosynthesis). Type II aromatic polyketides are further classified into 7 structurally distinct groups, anthracyclines, angucyclines, aureolic acids, tetracycline-type compounds, tetracenomycins, pradimicins and benzoisochromanequinones (BIQ) (Figure 1). Anthracyclines and angucyclines, the subjects of this study, are described in more detail in later sections 4.2.1 and 4.3.1.

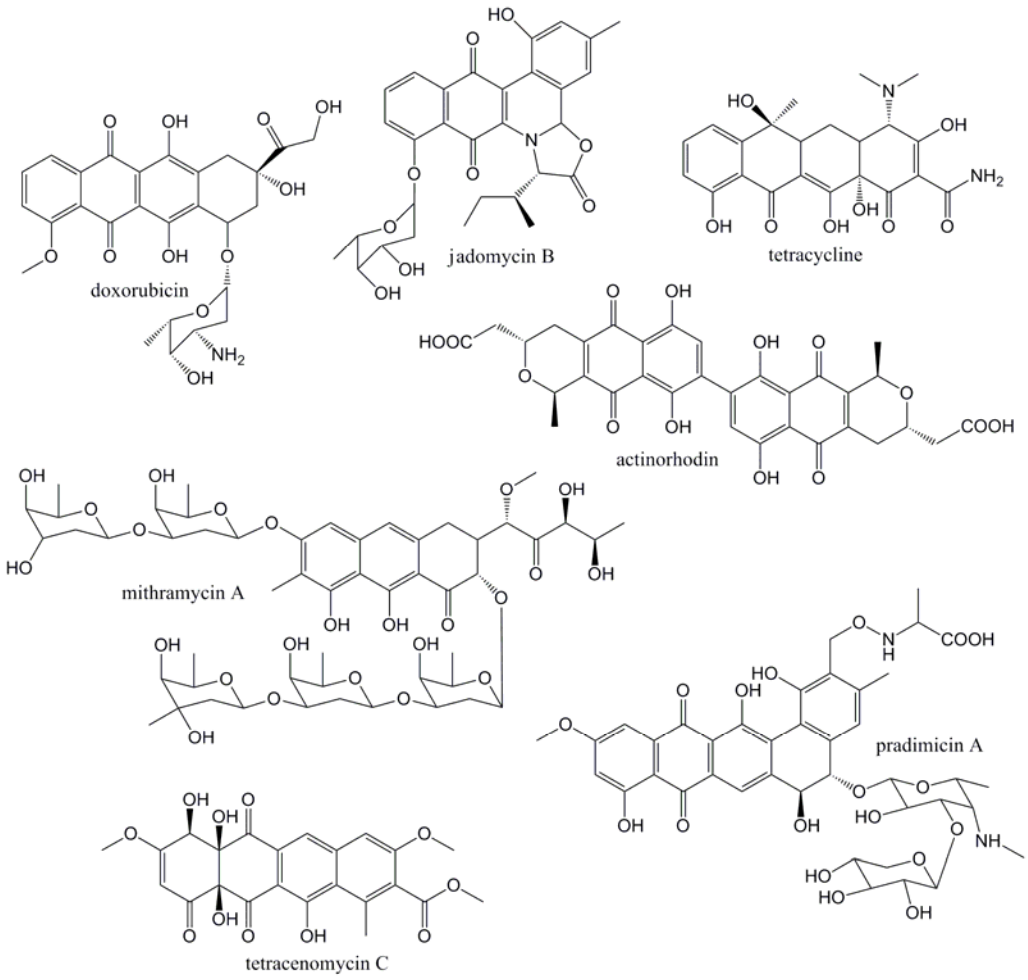

Figure 1. Typical examples of different groups of type II aromatic polyketide compounds: Doxorubicin (anthracycline), jadomycin B (angucycline), tetracycline, actinorhodin (benzoisochromanequinone), mithramycin A (aureolic acid), pradimicin A and tetracenomycin C.

1.1.5 Need for new natural products and use by humans

Despite the vast selection of reported compounds, there is an ever-increasing need for new pharmacologically active molecules and potential drug leads. One of the standing goals is to find alternatives to overcome adverse clinical effects which presently compromise the use of many drugs in human therapeutics. This is classically exemplified by anthracycline-induced cumulative cardiotoxicity (Chen *et al.*, 2007; Gianni *et al.*, 2008; Keefe, 2001) which together with rapidly emerging multidrug resistance (MDR) (Nobili *et al.*, 2006) typically restrains prolonged cancer treatment. As demonstrated by the immunosuppressor cyclosporin A (CSA) used for reversal of MDR (Aouali *et al.*, 2005) and antioxidants applied to overcome anthracycline cardiotoxicity (Hideg & Kalai, 2007), the screening of new compound derivatives is closely accompanied by the study of auxiliary effectors for improved clinical results.

Another obvious motive for the research is to provide alternative treatment against infectious diseases caused by resistant pathogenic micro-organisms, bacteria, viruses, parasites or fungi, which continue as the leading cause of death and a drain of resources world-wide (World Health Report 2007 / WHO; www.who.int/whr/en). Whatever the specific use may be, better understanding of the molecular mechanics behind various clinical conditions has in part rationalized the field of drug design, while the development of high throughput screens for different molecular targets requires a constant input of new ligands.

Only a small fraction of the bioactive compounds found in nature have been applied in any commercial purposes in medicine, pharmacology or agriculture. However, as illustrated by the comprehensive review by Newman and Cragg (Newman & Cragg, 2007), natural products continue to uphold a central role in the development of new drugs. Of the 1184 new chemical entities (NCEs) approved worldwide for human use between the years 1981 and 2006 over 50% were natural products or their derivatives (Newman & Cragg, 2007). The corresponding proportion of naturally-derived and/or semisynthetically modified antibacterial drugs was over 70% and for anticancer drugs just under 50%. This highlights the continuing role of multidisciplinary research in the discovery of new compounds from natural sources, including the yet largely unexplored genetic biosynthetic potential of *Streptomyces*.

1.2 Organization of the *Streptomyces* genome: Evolved to evolve and diverge

1.2.1 *Streptomyces* genome background

The research related to *Streptomyces* genetics date back to mid 60's when the first genetic recombination linkage map studies of *S. coelicolor* A3(2) were carried out by David Hopwood (Hopwood, 1965). Since then, inspired by the evident genetic potential in the synthesis of bioactive molecules, profound gene-level studies have continued. These have been reviewed by various authors (Chen *et al.*, 2002; Hopwood, 1999; Hopwood, 2006; Metsä-Ketelä *et al.*, 2007). So far the genomes of four *Streptomyces* species have been fully sequenced: *S. coelicolor* 8 667 507 bp/7825 ORFs (Bentley *et al.*, 2002) *S. avermitilis* 9 025 608 bp/7574 ORFs (Ikeda *et al.*, 2003), *S. griseus* 8 545 929bp/7138 ORFs (Ohnishi *et al.*, 2008) and *S. scabies* 10 148 695 bp (Sanger Institute Project, unpublished 11/2008; www.sanger.ac.uk/Projects/S_scabies). *Streptomyces* genome is unusually large with a high GC content (~70%), and organized in a linear chromosome with primary and secondary gene functions clearly distributed at distinct conserved and more variable regions. The chromosome replicates bidirectionally from a centered origin of replication in a process involving terminal protein (TP) mediated circularization. The secondary metabolic functions are mainly located at the distal ends of the chromosome arms. Importantly, this chromosomal arrangement is closely affiliated with the evolution and divergence of the secondary metabolite routes found in *Streptomyces* (Fischer *et al.*, 1997b). As will be seen, this is also reflected at the level of individual enzymes catalyzing the biosynthetic reactions along these pathways.

1.2.2 Genome size reflects a diversity of functions

The first obvious characteristic of *Streptomyces* genome is its exceptional size. Despite rather extensive variation between the genome size of different species, ranging from about 6.5Mb in *S. ambofaciens* (Leblond *et al.*, 1990) to over 10Mb reported for *S. scabies*, they are amongst the largest encountered in bacteria. Typically *Streptomyces* harbor more open reading frames than the single celled eukaryote *S. cerevisiae* (Goffeau *et al.*, 1996), and in comparison to the sporulating gram-positive soil bacterium *Bacillus subtilis*, the genome size and the number of potential open reading frames are roughly double (Kunst *et al.*, 1997). The enormous genome size reflects the complex life cycle of *Streptomyces* bacteria and adaptability to different stress factors encountered in the quickly changing conditions of the soil. A large fraction of the total genes have been suggested to be involved with environmental adaptation, response to external stimuli, and regulation between the primary and secondary metabolism functions. The focal point of genetic research has centered on these "excess" gene functions, primarily the production of bioactive secondary metabolites, genes responsible for the biosynthetic events and their regulation.

1.2.3 Linear genome and distribution of functions

Distinctively, although not exclusively in the bacterial kingdom, the *Streptomyces* chromosome is linear (Lin *et al.*, 1993; Volff & Altenbuchner, 2000). Different gene elements are distributed unevenly between the center and the distal regions of the chromosome, and are expressed during different phases of the life cycle (Karoonuthaisiri *et al.*, 2005). The central core region of roughly 6.5 Mb harbors a majority of essential primary metabolism genes such as those responsible for cell division, DNA replication, transcription and translation. Between different *Streptomyces* species, this region appears highly conserved in genetic content as well as gene order (Ikeda *et al.*, 2003). In contrast, the telomeric chromosomal arm regions show substantial variability and polymorphism from one species to another and contain non-essential genes largely redundant for primary metabolism (Bentley *et al.*, 2002; Ikeda *et al.*, 2003). These regions carry secondary metabolism gene elements related to adaptation to physical stress conditions, and most biosynthetic clusters involved in polyketide antibiotic production.

1.2.4 Genome susceptible to alteration

The high frequency of genetic alteration in the *Streptomyces* chromosomal arm regions was characterized as "instability" or "plasticity" years before molecular mechanisms behind the phenomenon were known. The variability is not related to linearity as such (Fischer *et al.*, 1997a; Wenner *et al.*, 2003), but to specific gene elements in the chromosomal arm regions, which facilitate various processes of extensive DNA rearrangement, deletion and duplication. These regions contain long terminal inverted repeat (TIR) sequences which vary in size up to hundreds of kilobases, and participate in homologous recombination events resulting in intra-chromosomal arm replacement and rearrangement (Fischer *et al.*, 1998; Widenbrant *et al.*, 2007). This is expectedly

enhanced by the geometry of the chromosome, as a single crossing-over event is sufficient for the exchange process in a linear DNA (Volff & Altenbuchner, 1998). TIRs are also found in linear plasmids which can integrate into the host genome (Gravius *et al.*, 1994; Kinashi *et al.*, 1992; Pandza *et al.*, 1998) facilitating plasmid-mediated interspecies conjugation (Kinashi *et al.*, 1992; Wang *et al.*, 1999). In addition, an exceptionally large number of transposase genes, phage-plasmid integrase genes, and corresponding transposable element sequences have been concentrated in the chromosomal arm regions (Bentley *et al.*, 2002; Ikeda *et al.*, 2003). Altogether, these sequences actively participate in the rearrangement of the genetic material within and between *Streptomyces* species, providing a genetic basis for the diversity of bioactive compounds observed in nature.

Phylogenetic sequence analysis has provided evidence that genes involved in *Streptomyces* secondary metabolism are subjected to horizontal gene transfer (HGT) (Egan *et al.*, 2001; Metsä-Ketelä *et al.*, 2002), a central phenomenon in molecular evolution between non-progeny bacterial species (Jain *et al.*, 2002; Jain *et al.*, 2003). In the work of Metsä-Ketelä and co-workers, extensive sequence comparison of over 100 *Streptomyces* species revealed disparate clustering of specific gene elements central to aromatic polyketide biosynthesis and the corresponding 16S rRNA sequences, demonstrating that exchange of genetic material from one species to another must have contributed to the natural diversification of the biosynthetic pathways. Recently, direct inter-species HGT has also been reported between the more distantly related taxa *Streptomyces padanus* and *Rhodococcus fascians* (Kurosawa *et al.*, 2008). This study illustrated that transfer of *Streptomyces* DNA in a competitive co-culture was accompanied by the appearance of capacity to produce novel aminoglycoside antibiotics by the recipient.

1.2.5 Secondary metabolism biosynthetic clusters

As opposed to primary metabolism genes which are spread out in the genome, the genes involved in secondary metabolism pathways are characteristically arranged together in defined continuous genetic regions or "clusters". Importantly, this organization allows the detection and isolation of complete functional biosynthetic pathways for biochemical study and manipulation. The clusters typically compose of up to dozens of genes required for the synthesis of a specific metabolite including carbon chain assembly, modification and regulation (see Figure 2).

The true distribution and divergence of secondary metabolite clusters in nature cannot be estimated on the basis of culture screens, as only a small minority of potential producers are culturable in laboratory conditions and only a fraction of these can be induced to produce the target metabolites. However, the genetic sequence data available at present has enabled different genome-mining approaches to detect and characterize biosynthetic pathways expected to produce formerly unknown products (Challis, 2008a; Corre & Challis, 2007; Gross *et al.*, 2007). Based on genome sequencing projects, *S. coelicolor* harbors 20 distinct secondary metabolite clusters

(Bentley *et al.*, 2002), whereas 30 putative clusters are found in *S. avermitilis* (Ikeda *et al.*, 2003). These include biosynthetic pathways for different siderophores, hopanoids, terpenes and pigments (Bentley *et al.*, 2002; Ikeda *et al.*, 2003; Omura *et al.*, 2001)as well as for various antibiotic compounds. *S. coelicolor* produces three known antibiotics; the type II polyketide actinorhodin, oligopyrrole prodiginine antibiotic complex and a non-ribosomal peptide antibiotic CDA (Bentley *et al.*, 2002). *S. avermitilis* has biosynthetic clusters for a number of type I polyketide compounds including avermectin, two putative type II aromatic polyketides and non-ribosomal peptide antibiotics. The most recently published genome sequence of *S. griseus* (Ohnishi *et al.*, 2008) reveals 34 secondary metabolite clusters, including clusters for the aminoglycoside antibiotic streptomycin, two putative lantibiotic peptide antibiotics and several non-ribosomal peptide antibiotics. Although complete genomic sequence data is not available aside from these examples, numerous individual antibiotic clusters from other *Steptomyces* species have been detected and characterized to varying degrees.

The NCBI CoreNucleotide database provides a means to approximate the variety of presently identified and sequenced clusters related to *Streptomyces* secondary metabolism. Search for "*Streptomyces* cluster" expected to cover most of the deposited target sequences, yields presently over 1200 distinct hits, most of which are overlapping partial biosynthetic clusters or refer to bacterial taxa other than *Streptomyces*. Inspection and manual processing of the search results one by one reveals about two hundred partial or complete clusters assigned to identified *Streptomyces* bioactive metabolite (or metabolite group) biosynthesis. These originate from altogether about a hundred different *Streptomyces* taxa, species, subspecies or strains. Many species were found to contain several different clusters, as in the case of *Streptomyces antibioticus* which harbours identified pathways for oleandomycin, oviedomycin and simocyclinone biosynthesis. In parallel, several pathways have been discovered in different species, such as the validamycin cluster in *S. hygroscopicus* and *S. tsusimaensis*, or the landomycin E cluster in *S. cyanogenus* and *S. globisporus*. Using a more strictly limited search "*Streptomyces* cluster complete" for parallel comparison and validation, about 60 of the gene clusters were denoted as *complete sequences*, covering all the known groups of bioactive metabolites produced by different *Streptomyces* species. The only complete cluster sequenced from two sources was the aminocyclitol antibiotic spectinomycin route cloned from *S. spectabilis* and *S. flavopersicus*. Apparently, the results are not comprehensive and only provide a generalized estimation. Many fully sequenced clusters, including those encoding for CDA and avermectin, are left out because, presumably, they have not been deposited in the data bank as separate complete entities.

So far the biosynthetic gene clusters of eight different anthracycline-type aromatic polyketide compounds have been cloned and sequenced entirely or in part (Metsä-Ketelä *et al.*, 2007). As related to this work (I-III), these include the nogalamycin cluster *sno* from *S. nogalater* strain ATCC 21451 (Ylihonko *et al.*, 1996) and the partially sequenced aclacinomycin biosynthetic clusters from two *S. galilaeus* strains

ATCC31615 (Räty *et al.*, 2000; Räty *et al.*, 2002b) and 3AR-33 (Chung *et al.*, 2002) (Figure 2). The available genetic sequence data on related angucycline-type compounds is more limited (Decker & Haag, 1995; Han *et al.*, 1994; Palmu *et al.*, 2007; Westrich *et al.*, 1999), largely because most of the existing biosynthetic pathways are silent in laboratory conditions (Metsä-Ketelä *et al.*, 2002) and undetected in conventional production screens. The two partially sequenced angucycline clusters subjected to this study (IV-VI), *pga* from *Streptomyces* sp. PGA64 and *cab* from *Streptomyces* sp. H021, are examples of pathways detected by DNA fingerprinting from non-producing phenotypes (Palmu *et al.*, 2007) (Figure 2).

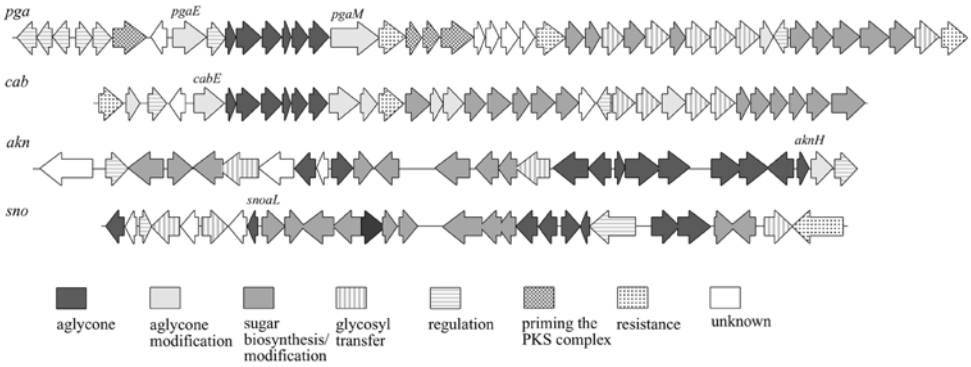

Figure 2. Graphic representation of the genes in sequenced biosynthetic clusters *pga*, *cab, akn* and *sno*. The genes encoding for the target enzymes of this study are shown as labelled arrows.

1.3 Biosynthesis of type II aromatic polyketides and genetic engineering

Type II aromatic polyketide biosynthetic routes have been extensively studied to elucidate the nature and order of the sequential enzymatic reactions. One typical gene-level approach employs extrachromosomal expression plasmids harboring complete or partial biosynthetic clusters, which can be induced in heterologous host cells to produce polyketide intermediates for analysis (Challis, 2008b). Together with enzyme-level *in vitro* functional and structural studies the details of many biosynthetic steps are now known and have been reviewed by various authors (Hertweck *et al.*, 2007; Metsä-Ketelä *et al.*, 2007; Niemi *et al.*, 2002).

Type II aromatic polyketides are synthesised in a cascade of sequential enzyme catalysed reactions analogous to fatty acid biosynthesis. First, a linear polyketide chain of typically 16-26 carbons (Ridley *et al.*, 2008) is assembled from a starter unit (acetate or less commonly propionate or longer) and two carbon acetate extender units in repeated Claisen condensation reactions. This is carried out by the so-called minimal polyketide synthase enzyme complex (minPKS) which consists of ketoacylsynthase/ chain length factor heterodimer (KS_α/KS_β or KS/CLF) and acyl carrier protein (ACP). The highly reactive carbon chain is then reduced, cyclized and aromatized to produce the first carbon ring. Aside from the carbon chain length, the biosynthesis of most type

II aromatic polyketides follows the same route up to this point (Figure 3). Anthracycline-type compounds are then typically converted into a common tetracyclic precursor aklavinone (AKV) (Arora, 1985). The reaction sequence involves the formation of the next two carbon rings, oxygenation (7,12 -quinone formation) and C10 substituent methylation, which in the aclacinomycin pathway yields aklanonic acid methyl ester (AAME). This is followed by the cyclization of the fourth ring, and ketoreduction at C7. The nogalamycin pathway is one of the rare exceptions to this route, and proceeds *via* nogalonic acid methyl ester (NAME) to nogalavinone, a product with the opposite stereochemistry at C9. In the biosynthesis of angucycline-type compounds the corresponding steps are catalyzed by a specific cyclase, which alone completes the tetracyclic carbon ring backbone and produces UWM6, a common intermediate in various angucycline pathways. In the following stages different pathways diverge into distinct routes by the action of post-PKS tailoring enzymes responsible for specific glycosylation, oxygenation, ketoreduction, and dehydration reactions, steps which account for the vast diversity of the secondary metabolites encountered in nature. In the context of the present work the term "tailoring" (reaction/enzyme) is used loosely and expected also to cover the 4th ring cyclization step of anthracyclines.

Scientific interest in the biosynthetic routes derives largely from their inherent flexibility, which allows recombination of genes or partial clusters from different origins to generate entirely new pathways. Ever since the development of specific molecular biology techniques, combinatorial biosynthesis and pathway engineering approaches have been explored for their potential in designing novel metabolites (Hutchinson & McDaniel, 2001; Kantola *et al.*, 2003; McDaniel *et al.*, 1995; McDaniel *et al.*, 1999; Weber *et al.*, 2003). The first reported demonstration of possibilities of genetic engineering was the production of mederrhodin, a hybrid between medermycin and actinorhodin by David Hopwood and coworkers (Hopwood *et al.*, 1985). Since then the methods have been applied, for example, to combine genes from different pathways to produce novel compounds (Kunnari *et al.*, 1999; McDaniel *et al.*, 1993; Miyamoto *et al.*, 2000; Remsing *et al.*, 2003), to generate more efficient production mutant strains (Lomovskaya *et al.*, 1999) and to replace chemical synthesis methods (Madduri *et al.*, 1998). However, systematic methods to efficiently harness the naturally existing genetic potential still remain to be elucidated. To supplement gene-level research in understanding the biosynthetic pathways, research has expanded to different fields of enzyme study, *in vitro* protein chemistry, high-resolution structural biology and bioinformatics.

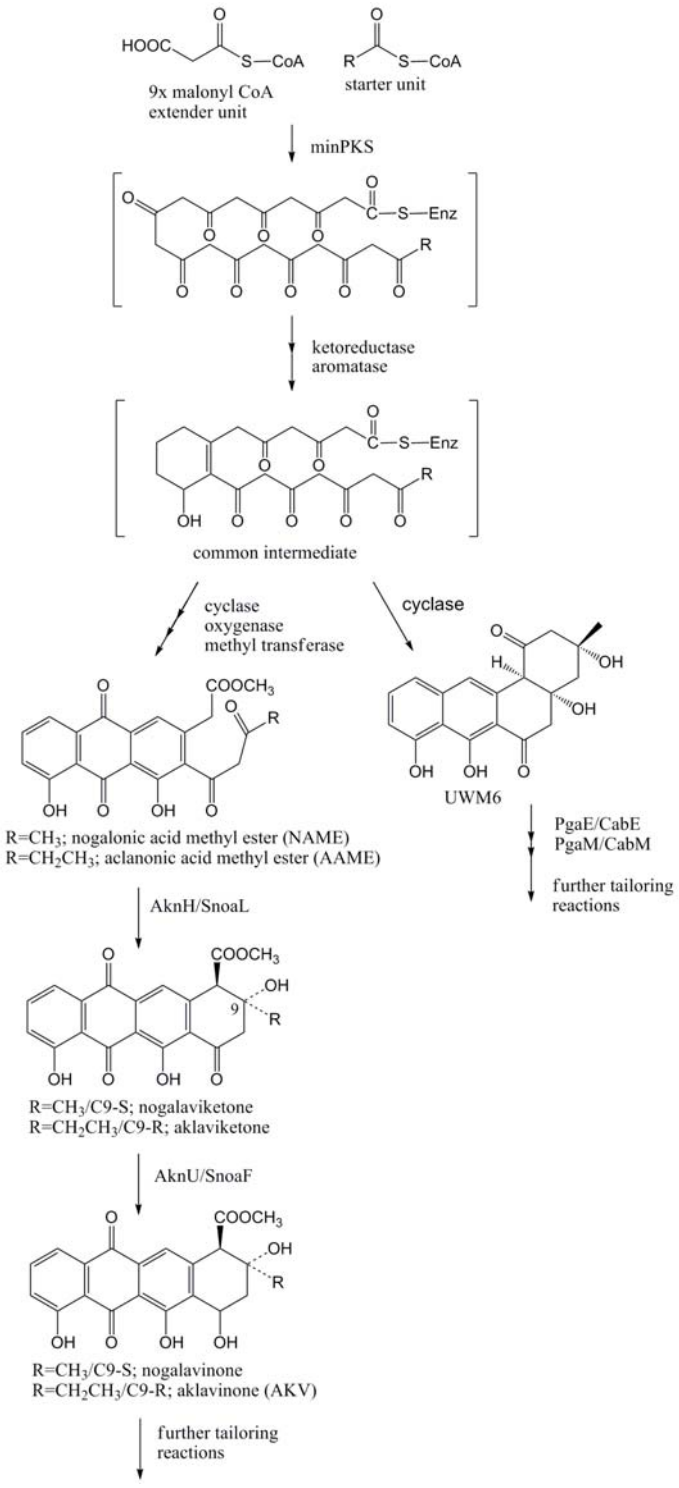

Figure 3. Simplified representation of some common steps in anthracycline and angucycline biosynthesis. R=CH_3 in nogalamycin pathway and typical angucycline biosynthetic routes. R=CH_2CH_3 in aclacinomycin pathway. The intermediates NAME and UWM6 are the substrates used in this study.

1.4 Tailoring enzyme *in vitro* structure/function analysis

1.4.1 Goals of enzyme research

One of the primary aims of molecular-level enzyme research is to provide a detailed understanding of the mechanisms and interactions behind observed biological functions. The *in vitro* approach enables the isolation of individual biosynthetic events, and investigation of enzyme biochemical properties and catalytic steps in a defined environment. Together with high-resolution structural information and mutational analysis, it is a powerful tool to study enzyme mechanisms and interactions at atomic level. One of the standing goals is to combine the knowledge of different disciplines to construct models for prediction and design, as in this case, to rationalize pathway modification and combinatorial biosynthesis. Structure/function information has already been applied to predict enzyme substrate specificity (Yadav *et al.*, 2003) and to pre-screen functional ligands from a pool of chemically diverse molecules (Kim *et al.*, 2007). It can also be used to simulate the native catalytic reaction whenever the biological substrate is not available for biochemical studies (Korman *et al.*, 2008) or to evaluate the roles of specific residues in the active site against a putative function (Ames *et al.*, 2008).

1.4.2 Enzyme research, past and present

The enzymological era of polyketide biosynthesis research began in the 1990's when the first crude cell extract fractionations and *in vitro* characterizations were carried out. Amongst the first post-PKS polyketide tailoring enzymes to be isolated and studied was carminomycin 4-O-methyltransferase DnrK (then called CMT or COMT). It was partially purified from the wild type producer *Streptomyces sp*. strain C5 by ammonium sulphate precipitation and DEAE-Sepharose fractionation (Connors & Strohl, 1993). The efficiency of protein recovery procedures became more effective by heterologous over-expression systems in different *Streptomyces* species and *E. coli,* which were adapted to purify DnrK further (Madduri *et al.*, 1993) and in production of enzymes such as cytochrome P-450 DoxA from the doxorubicin biosynthetic pathway (Walczak *et al.*, 1999). The purification procedures were further enhanced by the use of specific affinity tags in protein extraction, which provided a routine means of purifying protein preparates to near homogeneity. This paved the way for more elaborate *in vitro* characterization methods and protein crystal structure determination. The first crystal structure of an aromatic polyketide tailoring enzyme, oxygenase Act VA-Orf6 (*S. coelicolor*, actinorhodin biosynthesis) was solved in 2003 (Sciara *et al.*, 2003). It was soon followed by the high-resolution structures of aclacinomycin methyl esterase RdmC (Jansson *et al.*, 2003b) and aclacinomycin-10-hydroxylase RdmB (Jansson *et al.*, 2003a) (*S. purpurascens* rhodomycin biosynthesis) later the same year. Despite the extensive research around the aromatic polyketide biosynthesis, only around a dozen tailoring enzyme structures had been deposited in Protein Data Bank (PDB; www.rcsb.org) by the beginning of 2008, and only about half with a bound ligand in the active site. The enzymes discussed in this work are listed in Table 1. Numerous publications have focused on the enzymology of the earlier biosynthetic PKS steps (most recently reviewed by Hertweck *et al.* (2007) and Metsä-Ketelä *et al.* (2007)) but they are not discussed in the scope of this work.

Table 1. Type II aromatic polyketide tailoring enzymes (including fourth-ring cyclases) with resolved crystal structures.

Enzyme name (PDB accession code)	**Function**	**Source organism; biosynthesis route**	**Primary reference**
AclR (2gey)	aklavinone-1-hydroxylase	*S. galilaeus;* cinerubin	Beinker *et al.*, 2006
ActVA-Orf6 (1n5s;1n5t;1n5v;1lq9;1n5q)	6-deoxydihydro-kalafungin monooxygenase	*S. coelicolor;* actinorhodin	Sciara *et al.*, 2003
AknH (2f98;2f99)	aklanonic acid methyl ester cyclase	*S. galilaeus;* aclacinomycin	III
AknOx (2ipi)	aclacinomycin (rhodinose/cinerolose A) oxidoreductase	*S. galilaeus;* aclacinomycin	Alexeev *et al.*, 2007
CabE (2qa2)	UWM6-12-hydroxylase	*Streptomyces sp*. H021; unknown angucycline	IV
DnrK (1tw2)	carminomycin-4-O-methyltransferase	*S. peucetius;* daunorubicin	Jansson *et al.*, 2004
PgaE (2qa1)	UWM6-12-hydroxylase	*Streptomyces sp*. PGA64; unknown angucycline	IV
RdmB (1xdu;ixds;1r00;1qzz)	aclacinomycin-10-hydroxylase	*S. purpurascens;* β-rhodomycin	Jansson *et al.*, 2003a
RdmC (1q0z;1q0r)	aclacinomycin methylesterase	*S. purpurascens;* β-rhodomycin	Jansson *et al.*, 2003b
SnoaL (1sjw)	nogalonic acid methyl ester cyclase	*S. nogalater;* nogalamycin	II
SnoaL2 (2gex)	aklavinone-1-hydroxylase	*S. nogalater;* nogalamycin	Beinker *et al.*, 2006
TcmI (1tuw)	tetracenomycin F2 cyclase	*S. glaucescens*; tetracenomycin	Thompson *et al*., 2004
urdGT2 (2p6p)	dTDP-D-Olivose-transferase	*S. fradiae*; urdamycin	Mittler *et al*., 2007

2. AIMS OF THE PRESENT STUDY

This work focuses on the enzymology of *Streptomyces* secondary metabolism, specifically on selected post-PKS tailoring enzymes of type II aromatic polyketides. The target enzymes are the homologous polyketide cyclases SnoaL and AknH involved in anthracycline biosynthesis (I-III), and PgaE/CabE and PgaM which catalyze consecutive reaction steps in angucycline biosynthesis (IV-VI). The specific aims of the research were i) to set up preparative and analytical systems to enable the enzyme *in vitro* studies, ii) to verify the catalytic functions of the enzymes, and iii) to investigate their molecular-level reaction mechanisms and interactions.

In a broader sense the aim was to evaluate the functional and mechanistic features of these biosynthetic enzymes in general; to see if the enzymes share any common characteristics, and to see if and how the genetic plasticity of *Streptomyces* genome is reflected in the function of individual biosynthetic enzymes.

3. MATERIALS AND METHODS

General: Here is a brief outline of the most common techniques, materials and methods I have used in this study. More comprehensive descriptions are presented in the corresponding original publications I-VI.

DNA amplification: The high-fidelity Phusion DNA polymerase (Finnzymes, Finland) was used for all PCR amplifications. The reactions were carried out with 3% DMSO to facilitate DNA strand separation as instructed by the manufacturer. PCR primers were ordered from MWG (Germany) or TAGC Copenhagen (Denmark). All final constructs were verified by sequencing.

Plasmid constructs: The plasmids used in this study were constructed by conventional molecular biology techniques (Sambrook *et al.*, 1989) in combination with commercial DNA isolation kits (Qiagen GmbH, Germany)**.** The *E. coli* plasmids were modified from a commercial vector pBAD/HisB (Invitrogen, USA) and the final *Streptomyces* plasmids were derivatives of pIJE486 (Ylihonko *et al.*, 1994). The constructs used for the production of substrates NAME and UWM6 were pSY15 (Kunnari *et al.*, 1999) and pMC6BD (Metsä-Ketelä *et al.*, 2003) respectively.

Targeted mutagenesis: Due to the high GC content in *Streptomyces* DNA, many commercial mutagenesis kits have been found to be inefficient or unreliable. For this reason, all targeted mutagenesis constructs were made using a conventional 2-step 4-primer PCR method (Higuchi *et al.*, 1988). The method employs specific mismatch primers to introduce desired nucleotide substitutions in the amplified genes in two successive PCR reaction steps.

Transformation methods: *E. coli* transformations were carried out either by electroporation or heat-shock transformation of calcium chloride competent cells (Ausubel *et al.*, 1994). *Streptomyces* protoplasts were prepared and transformed as instructed in Hopwood *et al.,* 1985.

Cell lines: The *E. coli* strain used for the studies was the commercial TOP10 acquired from Invitrogen (USA). The *Streptomyces* host strain used for protein over expression and substrate production was *S. lividans* TK2 (Hopwood *et al.*, 1985).

Culture conditions: Depending on the specific application, cells were cultured at 37ºC, 30ºC or at room temperature in volumes ranging from 250ml flask cultures to 5L fermentor batches. The *E. coli* cultures were grown in rich 2xYT (yeast extract / tryptone) medium, whereas TSB (tryptic soy broth) medium (Kieser *et al.*, 2000) was typically used for the protein over expression in *Streptomyces*. Enzyme selenomethionine derivatives were produced in minimal drop-out medium (Sultana *et al.*, 2007). The medium used for substrate production was E1, with or without XAD resin as an adsorbent (Ylihonko *et al.*, 1994).

Protein purification and analysis: The target enzymes were typically overproduced as N-terminal histidine-tag or GST fusion proteins. The enzymes were purified to homogeneity in different combinations of specific affinity chromatography, gel filtration and anion exchange, using pre-packed commercial Amersham (Amersham Biosciences, UK) columns and ÄKTA FPLC (Amersham Biosciences, UK) chromatography system. Purity of protein samples was verified by SDS electrophoresis. Occasionally, purified proteins/variants were analyzed by protein mass spectrometry (TOF MS ES+) to confirm the exact mass.

Enzyme assays and reaction product analysis: *In vitro* enzyme reactions were carried out in 1ml volume in buffer conditions and concentrations optimized for each specific case. The reactions were monitored either kinetically using a UV-Vis spectrophotometer, or by qualitative/quantitative HPLC analysis of reaction products isolated by organic solvent extraction (chloroform or ethyl acetate). For some applications the reactions were coupled to secondary enzymatic conversion steps, and/or to analysis by LC-MS or NMR.

4. RESULTS AND DISCUSSION

4.1 Introduction

This section summarizes the most important findings of the experimental work presented in the original publications I-VI. Selected central themes are then discussed in a broader context in relation to literature.

The experimental work was carried out in close co-operation between our research team (Antibiotic Biosynthetic Enzymes, Department of Biochemistry and Food Chemistry, University of Turku, Finland) and our collaborators (Molecular Structural Biology, Department of Medical Biology and Biophysics, Karolinska Institutet, Stockholm Sweden). DNA constructs*, heterologous protein expression and purification, substrate isolation, activity assays and enzyme functional characterization were done in our laboratory. Protein structural studies including crystallization, X-ray analysis and structure determination (I-IV) were carried out by our collaborators.

*with the exception of the SnoaL variant constructs, which were produced by Jishu Wang at Karolinska Institutet.

4.2 Polyketide cyclases SnoaL and AknH in anthracycline biosynthesis (original publications I-III)

4.2.1 Introduction to SnoaL and AknH

Anthracyclines are amongst the most extensively studied group of type II aromatic polyketides. They are derivatives of the tetracyclic 7,8,9,10-tetrahydro-5,12-naphtacenoquinone structure, and the biologically active species are typically (poly)glycosylated at position C7 by deoxysugars and aminodeoxysugars. Anthracyclines are best known for their clinical applications in cancer treatment, as exemplified by daunorubicin, doxorubicin and epirubicin, which have been used for decades in the clinical treatment of various types of solid and hematopoietic tumors. Over the years in the search for more effective drugs with less severe side effects, thousands of anthracycline derivatives have been found and characterized, and the biosynthesis of several model compounds have been elucidated in detail. One of these is aclacinomycin A (aclarubicin) (Figure 4) which has been used in medicine, and thoroughly studied both for its clinical activities (Karanes *et al.*, 1990; Nitiss *et al.*, 1997; Wakabayashi *et al.*, 1980; Wang *et al.*, 2002) and its biosynthesis (Alexeev *et al.*, 2007; Jansson *et al.*, 2003b; Jansson *et al.*, 2005; Leimkuhler *et al.*, 2007; Lu *et al.*, 2004; Räty *et al.*, 2002a; Räty *et al.*, 2002b). Nogalamycin (Figure 4) is another anthracycline antitumor antibiotic, which has been under focus in relation to its mode of action (Box, 2007; Williams *et al.*, 1990), genetics and biosynthesis (Torkkell *et al.*, 1997; Torkkell *et al.*, 2001; Ylihonko *et al.*, 1996). In comparison to typical anthracyclines, nogalamycin is characterized by an unusual glycosylation (nogalamine attached to C1 and C2) (Wiley *et al.*, 1977) and an opposite C9 configuration (Arora, 1983).

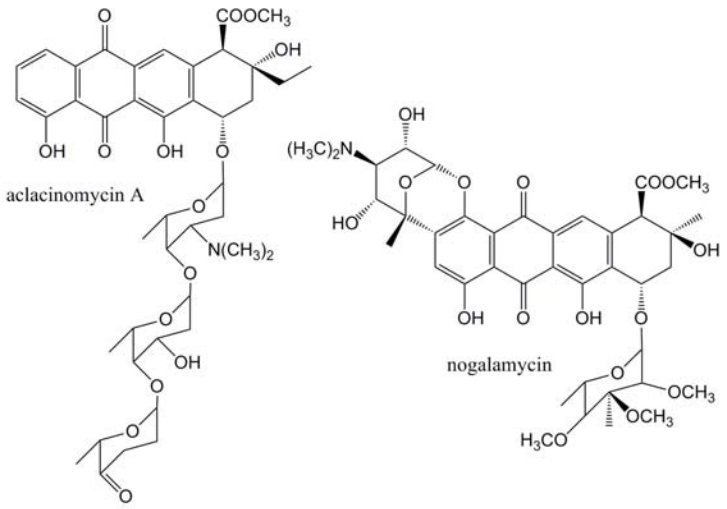

Figure 4. Aclacinomycin A and nogalamycin, the products of *akn* (*S. galilaeus*) and *sno* (*S. nogalater*) pathways, respectively.

SnoaL and AknH are representatives of a group of small aromatic polyketide cyclases, which are responsible for closing the last aglycone carbon ring in various anthracycline pathways; completing the tetracyclic core for successive pathway-specific tailoring reactions (Figure 3). AknH catalyses the conversion on aklanonic acid methyl ester (AAME) to aklaviketone in aclacinomycin A biosynthesis in *S. galilaeus*. SnoaL catalyses the corresponding reaction in nogalamycin biosynthesis in *S. nogalater*, converting nogalonic acid methyl ester (NAME) to nogalaviketone *in vivo* (Torkkell *et al.*, 2000). The enzymes are both 144 amino acids in length and 67% identical at the amino acid level. They also share substantial sequence similarity with other members of the cyclase class such as RdmA (rhodomycin biosynthesis; *S. purpurascens*) (Niemi & Mäntsälä, 1995), DnrD (doxorubicin biosynthesis; *S. peucetius*) (Kendrew *et al.*, 1999), and DauD (daunomycin biosynthesis; *Streptomyces sp* strain C5) (Dickens *et al.*, 1995).

The reactions catalyzed by SnoaL and AknH are of specific interest because they determine the stereochemical outcome of the final biosynthesis products. SnoaL has been shown to be solely responsible for the unusual C9-S configuration observed in ring A of nogalamycin (Torkkell *et al.*, 2000). AknH in turn directs the biosynthesis towards AKV the most common precursor of anthracycline-type polyketides with the opposite C9-R stereochemistry. As the primary initiative to our enzyme-level study, we set out to investigate *in vitro* the molecular basis behind this enzyme stereoselectivity.

4.2.2 Proteins and substrate for characterization

As the initial step of the experimental study, the entire genes encoding for SnoaL and AknH were cloned into several modified pBAD/HisB -based plasmid expression vectors and verified by sequencing. These vectors were later used as templates for active site variant constructs listed in tables 2 and 3. The enzymes were heterologously expressed in *E. coli* as N-terminal histidine-tag fusion proteins (I-III). As an exception, due to instability resulting from active site amino acid substitutions, the AknH variants were produced as GST (glutathione S-transferase) fusion proteins (III). The corresponding non-mutated AknH/GST was prepared in parallel for comparison. The optimized soluble recovery yields were typically in the milligram-range per liter culture medium. The proteins were purified using different combinations of consecutive preparative chromatography steps including affinity chromatography, gel filtration and anion exchange. For each preparate, purity and protein size were verified by SDS-PAGE electrophoresis, followed by biochemical and/or structural characterization.

The biological substrate for SnoaL, NAME, was provided by Dr. Tero Kunnari for the initial experiments, and later produced and purified using a hybrid production strain *S. argillaceus* /pSY15 (Ylihonko *et al.*, 1996). NAME was also used for all AknH activity measurements because the natural substrate AAME was unavailable.

4.2.3 Cyclase reactions *in vitro*

Both enzymes were active and shown to convert NAME *in vitro* (Figure 5). The reactions were monitored using a HPLC-based end-point method, either as a decrease of the substrate (II) or as an increase of the cyclized product (III). The reactions were typically carried out near the saturation limit of the substrate, well above the physiological substrate concentrations and above the expected K_M value range of the enzymes. In these conditions the quantitation of end product formation proved more accurate than the comparison of modest changes in integrated substrate peak areas. The end product approach also enabled the stereochemical assays described in section 4.2.4.

The conversion of NAME could also be detected spectrophotometrically as a slight increase in absorbance (Δ_{max} at 420nm), but due to small differences in signal intensities and resulting poor reproducibility, the method was not useful for quantitation (results unpublished). In an attempt to set up a more sensitive kinetic assay, we decided to couple the reactions to a secondary enzyme-catalyzed step with a potentially higher signal. The idea was to use the consecutive biosynthetic enzymes, the 7-ketoreductases SnoaF and AknU (see Figure 3) from the nogalamycin and aclacinomycin pathways, respectively. The genes were cloned into expression constructs, followed by enzyme production and purification, but the soluble recovery yields remained too low for practical use (results unpublished).

Altogether, kinetic characterization of the cyclases was hindered by the lack of a reliable continuous assay and the means of determining the absolute substrate and

product concentrations as the molar absorption coefficients were not known. The estimated k_{cat} values for the enzymes in the reaction conditions were around $1s^{-1}$, with roughly approximated K_M values in the low micro-molar range.

NAME and the cyclization products are pure bright yellow in methanol and in the aqueous buffers used in the assays. Upon extended incubation the reaction products were shown to slowly undergo further conversion into a reddish compound, expectedly as a result of spontaneous loss of the C9 hydroxyl group and aromatization (Figure 5). The same product was observed in the SnoaL and AknH crystal structure complexes (chapter 4.2.5).

4.2.4 C9 stereochemical assay

The stereochemically distinct C9-S and C9-R cyclization products of SnoaL and AknH could not be differentiated based on their UV-Vis spectra or separated in conventional reverse phase HPLC. However, we demonstrated that the two products could be distinguished from one another in a coupled enzymatic assay with a stereospecific hydroxylase RdmE (aklavinone-11-hydroxylase) (Niemi *et al.*, 1999) *in vitro*. RdmE was shown to exclusively use the C9-R stereoisomer as a substrate, leaving the C9-S product unreacted (Figure 5). The hydroxylated product analog could then be separated by HPLC from the initial cyclization products and quantitated to monitor the relative amount of each form present (Figure 2 in III). As anticipated, AknH produced exclusively the C9-R stereoisomer (auraviketone) whereas SnoaL only produced the C9-S form (nogalaviketone).

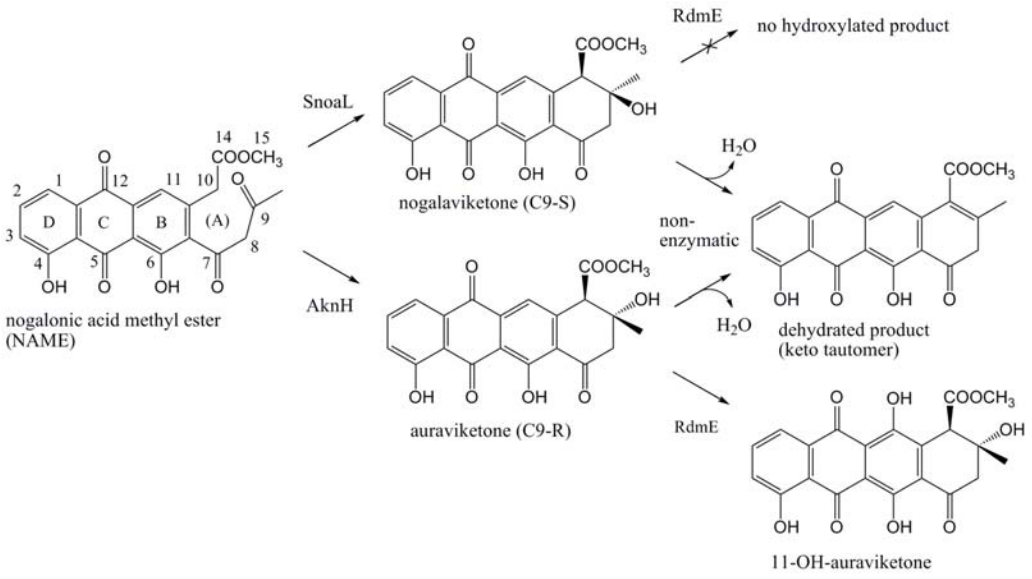

Figure 5. The *in vitro* cyclase reactions of SnoaL and AknH with NAME, coupled to RdmE-catalyzed hydroxylation. Both cyclase products also may undergo slow spontaneous dehydration leading to the loss of C9 stereochemistry.

4.2.5 Structures of SnoaL and AknH

As the first representative of this family of homologous polyketide cyclases, the 3D X-ray crystal structure of SnoaL was resolved at 1.35Å resolution (PDB accession code 1sjw) (I-II). The structure was acquired using single isomorphous replacement (SIR) with uranium derivative anomalous scattering. SnoaL was crystallized in the presence of the NAME substrate, and the electron density clearly revealed a bound ligand in the active site. The X-ray crystal structure of AknH (PDB accession code 2f98) was resolved by molecular replacement using the SnoaL structure as a template (III). The AknH structure was determined in complex with two different ligands, derivatives of NAME and AKV at resolution 2.1Å and 1.9Å respectively.

Structural analysis revealed that SnoaL (Figure 6) and AknH belong to the superfamily of α+ß barrels, and are representatives of the scytalone dehydratase fold (Lundqvist *et al.*, 1994). The two enzymes are very similar in their overall tertiary and quaternary structures. Both polypeptides fold into a single domain which consists of five-stranded mixed ß-sheet, two α-helices and two 3_{10}-helices (Figure 3 in II; Figure 3a in III). These regular secondary structure elements form a distorted α+ß barrel which encloses a large hydrophobic active site cavity. Analysis suggests that the native quaternary structure of the proteins is a tetramer composed of a pair of homodimeric subunits (Figure 5 in II; Figure 3b in III). The quaternary structures are stabilized by a network of hydrogen bonds and expected to undergo a conformational change upon substrate binding and product release. The main structural differences between the SnoaL and AknH reside in the loop region between helices H3 and H4 (residues 40-43) and at the C-terminal end (residues 135-141).

In both cyclases amino acid residues from nearly all the regular secondary structural elements line the substrate binding site, with most of the hydrophobic residues arranged in the vicinity of the ring A atoms which participate in the reaction. Although the observed ligands in the structure complexes had undergone enzymatic and/or spontaneous chemical changes, the orientations of the SnoaL/NAME and AknH/AKV derivatives were expected to resemble those in the native productive Michaelis' complexes (Figure 7 in II; Figure 6 in III). In both cases the main enzyme-ligand interactions were van der Waals forces and hydrophobic contacts. In the SnoaL complex the only direct hydrogen bond was observed between the amide group of Gln105 and C14 carbonyl oxygen of the ligand. Sequence comparison revealed that this glutamine residue is conserved in the known homologs of the cyclase family. Also in AknH the Gln105 was shown to participate in a network of hydrogen bonds which interacts with the bound ligand. Another invariant residue found in the active site cavities in the proximity of the reacting C9 of both enzymes was Asp121, which was expected to have a role in the catalysis. Based on the structural information several active site variants were prepared and characterized (Table 2) to clarify the reaction mechanism employed by the cyclases.

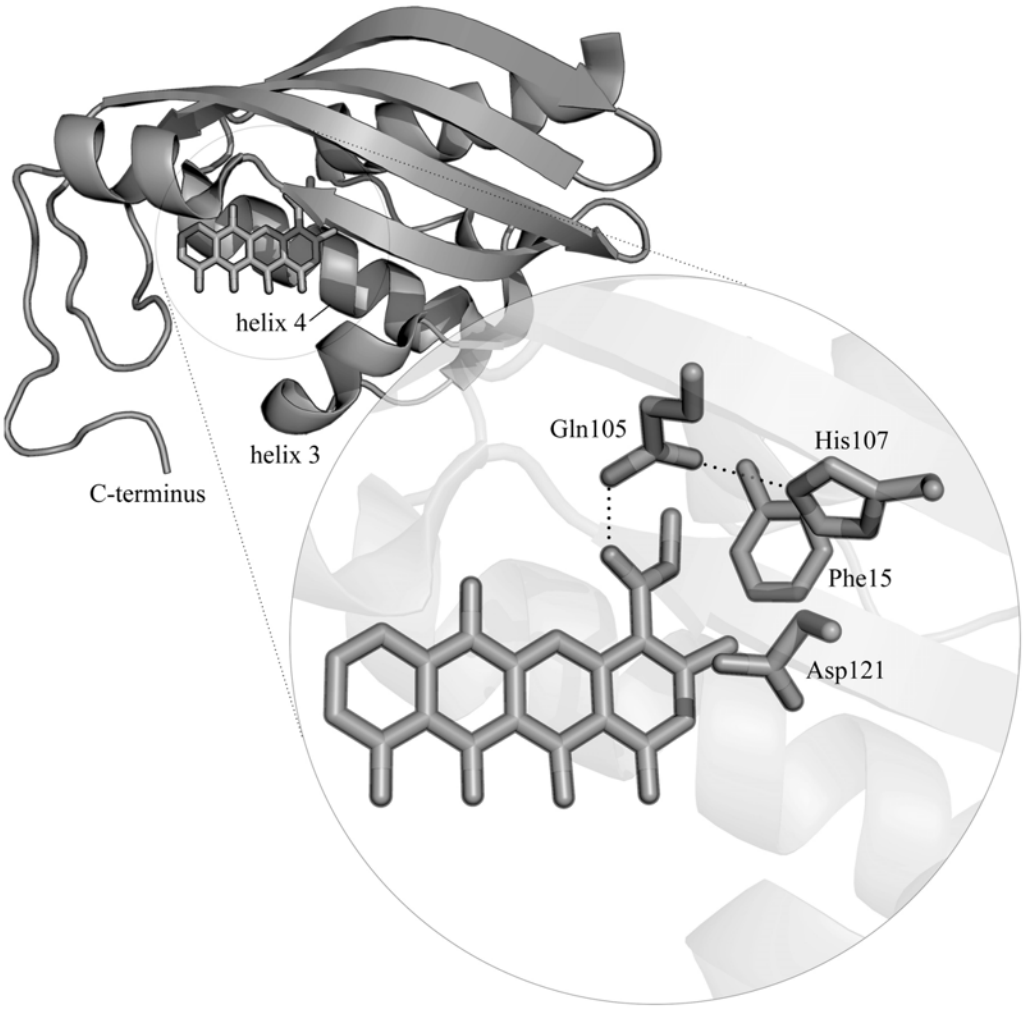

Figure 6. Representation of the SnoaL monomer X-ray crystal structure with a bound ligand. The key active site residues under study are shown in the close-up. The ligand is a cyclized product derivative, which has lost the C9 hydroxyl substituent in unspecific dehydration, resulting in ring A aromatization and loss of C9 stereochemistry.

4.2.6 Cyclase reaction mechanism

In chemical terms the reaction catalysed by SnoaL and AknH is an intramolecular aldol condenzation. This reaction is initiated by proton abstraction at the reacting carbon, resulting in a carbanion/enol(ate) intermediate. Typically this step is facilitated by stabilizing the intermediate either through covalent Schiff base formation (class I aldolases) or by using metal ion stabilization (class II aldolases). However, the cyclases SnoaL and AknH do not have either an appropriately positioned lysine residue or a metal ion cofactor required for these mechanisms. Instead, the catalysis may proceed through an unusual aspartate-driven acid-base mechanism. According to the model, the

reaction is initiated by the removal of the C10 proton by the invariant residue Asp121 (Figure 7; step 1). In addition to the favorable orientation of this residue, its role is emphasized by the fact that the point mutation Asp121Ala either completely inactivates the enzyme (SnoaL) or seriously impaires the catalytic activity (0.2% residual activity in AknH) (Table 2). In contrast to the classical aldolases, the essential intermediate stabilization is proposedly brought about by π electron delocalization over the extended aglycone ring system (discussed in chapter 4.5.3). Additional yet less significant stabilization is provided by hydrogen bond formation between C14 carbonyl oxygen and the invariant residue Gln105. In accordance, the point mutation Gln105Ala moderately decreased the relative enzyme activities of SnoaL (15%) and AknH (6%). A comparable effect in AknH resulted from the substitution of the conserved residue His107, which also participates in the stabilizing hydrogen bond network. The subsequent step in the catalysis is an intermolecular nucleophilic attack of the carbanion intermediate on the C9, which leads to ring A formation (Figure 7; step 2). To complete the reaction cycle, the remaining oxyanion/alcoholate at C9 is protonated, proposedly by Asp121 in SnoaL, or by a catalytic water observed in the crystal structure in AknH (Figure 7; step 3).

Table 2. Relative activities of SnoaL and AknH active site variants in comparison to the corresponding non-mutated species.

Cyclase	Point mutation	Variant relative activity (%) vs. wild type
SnoaL	Asp121Ala	no detectable activity
SnoaL	Gln105Ala	15
AknH	Asp121Ala	0.2
AknH	Gln105Ala	6
AknH	His107Ala	2

4.2.7 Cyclase stereochemical determinants

SnoaL and AknH are highly similar in primary structure, 3D fold and catalytic mechanism, yet when provided with the same substrate, the reaction products differ in the stereochemistries of C9 of the ring A. The opposite conformations must be determined by structural differences between the active sites, and apparently in part, by a hydrophobic area surrounding one side of the reacting atoms. The topology of the SnoaL-ligand complex implicates that the C9-S product is formed as a result of nucleophilic attack on the *re* face of the C9 aldehyde (Figure 7; step 2). This reaction is energetically favored because it orients the forming alcoholate towards the protonated Asp121, and away from the substantially hydrophobic environment of the opposite face. Intriguingly, the distribution of hydrophobicity around the forming ring A is different in cyclases which produce the C9-R stereoisomer, and as in AknH, the nucleophilic attack takes place on the *si* face (Figure 7; step 2). Sequence analysis shows that these enzymes have a conserved residue Tyr15 instead of a highly nonpolar Phe15 located within a van der Waals distance from C9 of the product in SnoaL. The

substitution Phe15Tyr in SnoaL, however, did not alter the reaction stereoselectivity. In contrast, the corresponding reverse substitution Tyr15Phe in AknH resulted in a partial loss of specificity and formation of the unnatural C9-S product (20% of the total yield) (Figure 2c in III). Further decrease of polarity in the hydrophobic area near the Asp121 in AknH was achieved by a second substitution Asn51Leu. This additional mutation Asn51Leu with Tyr15Phe completely abolished AknH stereoselectivity, resulting in the formation of a racemic mixture of C9-S and C9-R stereoisomers (50% each) (Figure 2d in III). The remaining structural factors which determine the absolute stereochemistries were left unindentified, but may involve the loop residues between helices H3 and H4 or the C-terminal amino region (see Figure 6) where the most significant structural differences were observed. Despite the fact that complete reversal of stereochemistries was not achieved, this work demonstrated that molecular-level information on enzyme structure and function may enable enzyme modification to specifically alter enzyme enantioselectivity by targeted mutagenesis.

Table 3. Stereoselectivity and relative activities of SnoaL and AknH active site variants in comparison to the corresponding wild type species.

Cyclase	**Point mutation**	**Relative activity (%) vs. wildtype**	**Reaction product C9 stereochemistry**
SnoaL	Phe15Tyr	5	C9-S / no change in stereochemistry
AknH	Tyr15Phe	45	~20% change from C9-R to C9-S
AknH	Tyr15Phe & Asn51Leu	20	~50% change from C9-R to C9-S

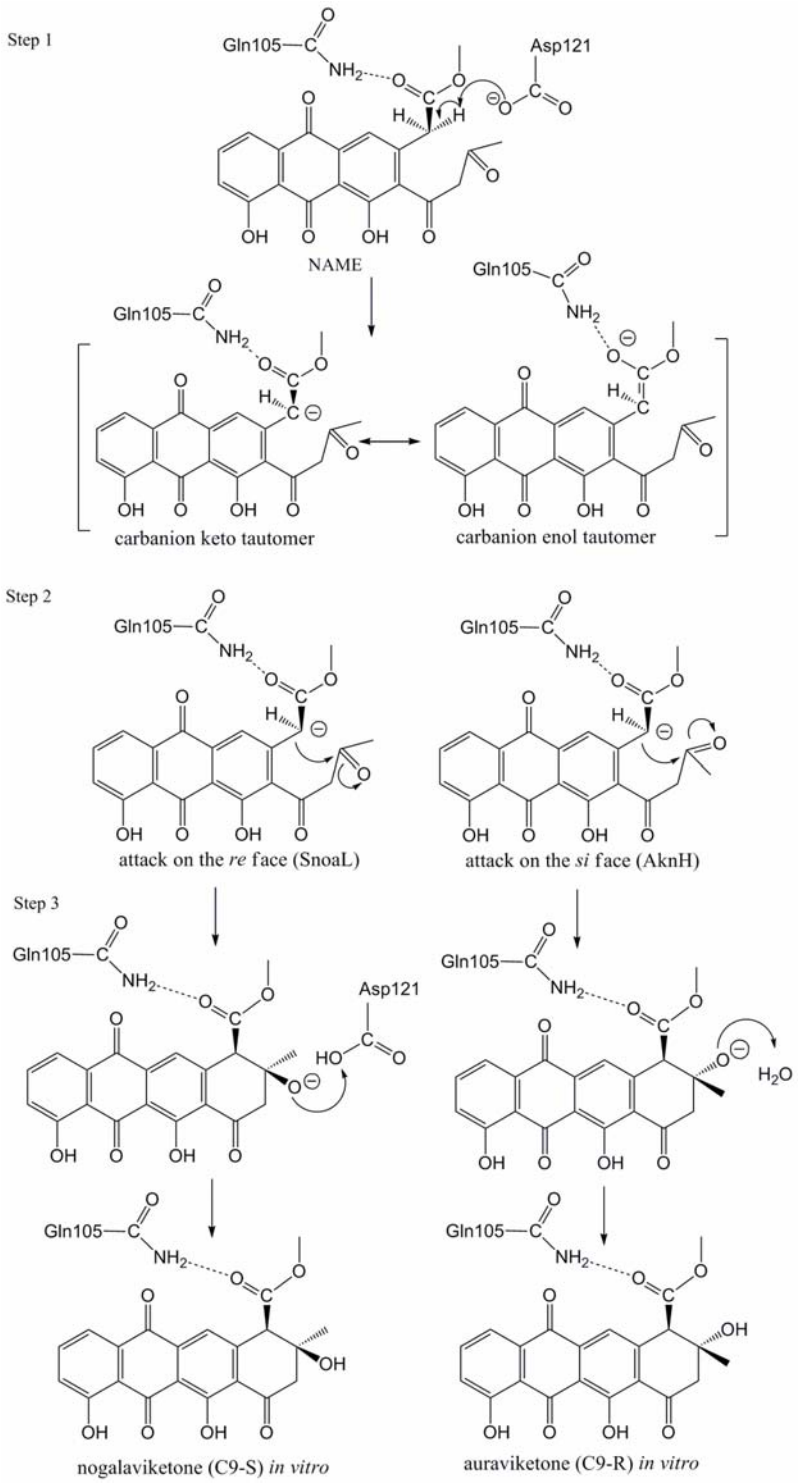

Figure 7. The proposed cyclase catalytic mechanism and determination of stereoselectivity in SnoaL and AknH (intermediate aromatic resonance stabilization not shown).

4.3 Oxygenation reactions in angucycline biosynthesis (original publications IV-VI)

4.3.1 Background to angucycline research

Angucyclines are a distinct group of natural aromatic polyketides produced by various *Streptomyces* species (Rohr & Thiericke, 1992). They are closely related to anthracyclines, and characterized by the angular shape of the tetracyclic benz[a]anthracene ring frame. The first angucycline structures were described already in the mid 60's (Dann *et al.*, 1965), yet today only a little over one hundred different structures are known. DNA fingerprinting studies have indicated that angucycline clusters may be much more common in nature than this number of discovered compounds implies (Metsä-Ketelä *et al.*, 2002). Angucyclines have been shown to possess a variety of bioactive properties (Abdelfattah *et al.*, 2003; Antal *et al.*, 2005; Bringmann *et al.*, 2005; Korynevska *et al.*, 2007), but so far they have not been used for any direct medical applications.

The most extensively characterized representatives of the angucycline family include landomycin (Henkel *et al.*, 1990), urdamycin (Drautz *et al.*, 1986) (Figure 8) and jadomycin (Ayer *et al.*, 1991), for which many of the biosynthetic steps have been described in detail. The biosynthesis follows the typical steps of type II polyketide assembly, diverging on the angucycline pathway as a result of specific cyclase which orients the forming 4th ring in an angled position (Metsä-Ketelä *et al.*, 2003) (Figure 3). There has been certain ambiguity, however, in regard to some of the subsequent post-PKS tailoring steps. Specifically, the successive oxygenation steps which follow the aglycone 4th ring cyclization are not yet fully understood, and due to difficulties in determining the catalytic events involved *in vivo*, the reaction cascades have sometimes been described as "biosynthetic black boxes" (Rix *et al.*, 2005). The primary initiative to this experimental work was to study these biosynthetic steps by isolating the individual enzyme-catalyzed reactions in selected model pathways *in vitro*. The aim was to confirm the catalytic functions of specific enzymes (PgaE, PgaM and CabE), and to study the mechanisms involved.

The target enzymes PgaE and PgaM derive from partial cryptic angucycline-type gene cluster *pga* from *Streptomyces* sp. PGA61 (Metsä-Ketelä *et al.*, 2004), whereas CabE originates from the related *cab* cluster from *Streptomyces* sp. HO21 (Palmu & Kunnari, 2002) (Figure 2). The clusters were initially detected in a genetic screen specific for type II aromatic polyketide PKS regions (Metsä-Ketelä *et al.*, 1999), cloned and reconstructed for *in vivo* gene function analysis (Palmu *et al.*, 2007). PgaE and CabE are homologous to flavin-dependent hydroxylases and mono-oxygenases, and, based on *in vivo* studies, they were expected to participate in UWM6 C12 hydroxylation. Sequence analysis implicates that PgaM is a two-domain protein which consists of a N-terminal region homologous to flavoprotein mono-oxygenases fused to C-terminal domain homologous to short-chain alcohol reductases/dehydrogenases (SDR). These two putative functional domains have been suggested to participate in the angucycline

core C12b hydroxylation and 2,3-dehydration, respectively (Palmu *et al.*, 2007). There has been, however, uncertainty about the reaction sequence and the chemical nature of the true reaction product.

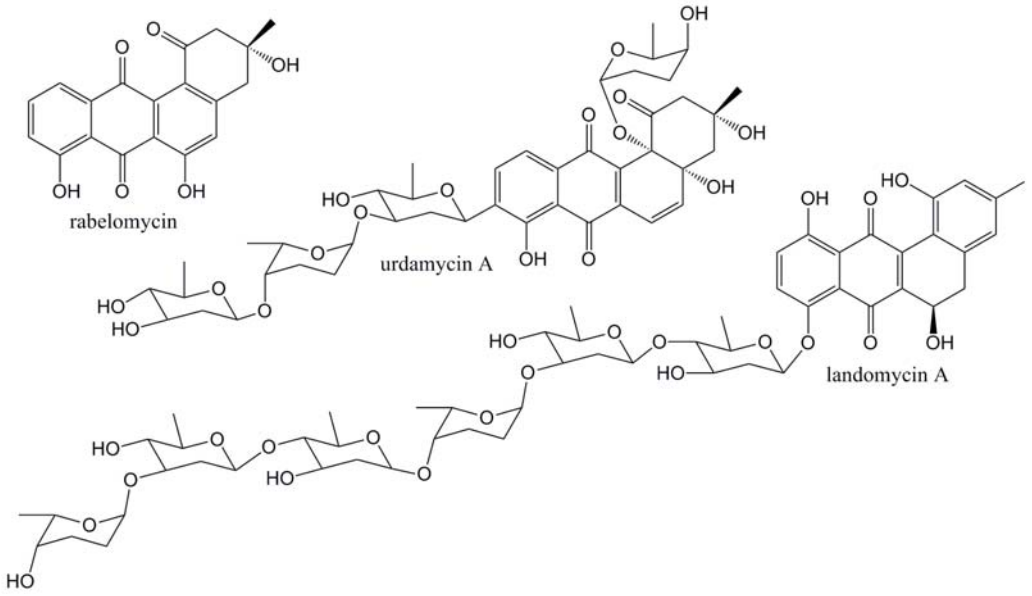

Figure 8. Rabelomycin, a common intermediate and/or shunt product observed in various angucycline pathways. Urdamycin and landomycin are products of *urd* and *lan* pathways, repectively. The final products of the *pga* and *cab* pathways are unknown.

4.3.2 Enzymes and substrate for the studies

The entire open reading frames encoding for PgaE, CabE and PgaM were cloned into various plasmid expression constructs and verified by sequencing. The PgaE construct was later used as a template for several active site variants listed in table 4. PgaE and CabE were expressed as N-terminal polyhistidine tag fusion proteins in *E. coli* TOP10. Due to low recovery yields in *E. coli*, PgaM was typically produced in a corresponding construct in *S. lividans* TK24. The enzymes were purified using combinations of chromatographic techniques including nickel affinity chromatrography and preparative gel filtration, followed by analytical SDS-PAGE chromatography to confirm purity. The purified enzymes remained active in storage and were used for structural and biochemical characterization.

The substrate UWM6 was produced and purified from *Streptomyces* TK24/pMC6BD (Metsä-Ketelä *et al.*, 2003) a construct which harbors all the biosynthetic genes of the pathway prior to the gene encoding for PgaE. Purified UWM6 was apparently stable and did not show signs of degradation or spontaneous conversion when stored solubilized in methanol.

4.3.3 Sequence of reactions catalyzed by PgaE and PgaM; formation of gaudimycin C

PgaE and PgaM were shown to catalyse a series of consecutive reactions *in vitro* (VI). The first step, conversion of the biosynthetic intermediate UWM6, was catalysed by PgaE. This reaction could be directly monitored as consumption of UWM6, either spectrophotometrically at the substrate absorption maximum or by using quantitative HPLC-based end-point system optimized for the purpose (Figure 2 in VI). The PgaE/UWM6 conversion product, however, could not be isolated or detected. In contrast to the corresponding reaction *in vivo* which had been previously shown to yield rabelomycin (Figure 8) (Palmu *et al.*, 2007), the reaction *in vitro* invariably resulted in product degradation. This was observed in HPLC as the formation of multiple unstable minor products (Figure 3 in VI). The degradation could be avoided by coupling the PgaE reaction to the next biosynthetic steps catalysed by PgaM. Incubation of PgaM together with PgaE and UWM6 resulted in the formation of a new stable product which could be extracted for detailed study. The coupled reaction was shown to be highly dependent on simultaneous presence of the two enzymes, and sensitive to variations in the relative enzyme concentrations, observed as intermediate product breakdown in unoptimal reaction conditions (Figure 4 in VI). Although the results implicated that the two enzymes may require direct contact with eachother to carry out the sequential reactions, physical interaction or complex formation could not be observed.

LC-MS analysis revealed that the product of the coupled PgaE/PgaM reaction with UWM6 had a mass (356m/z), 14amu higher than that of UWM6 (342m/z). To determine the chemical changes involved, we set off to investigate the product structure by NMR spectroscopy. Successive preparative-scale *in vitro* enzyme reactions followed by consecutive purification steps were carried out to produce sufficient material for the analysis. Eventually, a complete NMR structure of the compound was resolved (Table 2 in VI), confirming the compound as a novel angucycline intermediate. The structure was different from the corresponding *in vivo* product of the homologous *cab* pathway, gaudimycin B, only in the stereochemical orientation of the C6-OH substituent. Interestingly, it was also different from the equivalent *in vivo pga* product, gaudimycin A, in which the C6 hydroxyl group was completely missing. Consequently, the *in vitro* product was named gaudimycin C (Figure 9). The structure of gaudimycin C allowed us to deduce the chemical steps of the PgaE/PgaM reaction cascade. Significantly, it included hydroxylation at positions C12 and C12b, C7,12- quinone formation, C2,3-dehydration and stereospecific ketoreduction at C6.

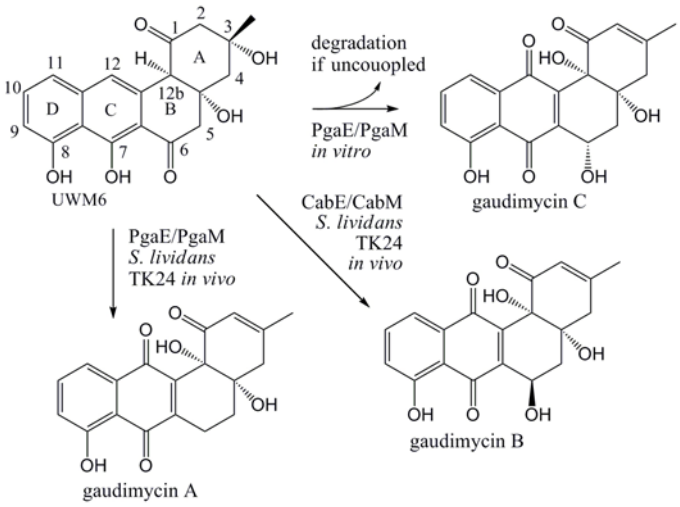

Figure 9. *In vitro* reaction of PgaE and PgaM with UWM6 leads to the formation of gaudimycin C. The corresponding *in vivo* reaction in heterologous host *S. lividans* TL24 yields gaudimycin A (6-dehydroxy gaudimycin C), whereas the reactions catalyzed by homologous CabE and CabM produce gaudimycin B (C6 stereoisomer of gaudimycin C).

4.3.4 PgaE functions as UWM6 C12 hydroxylase *in vitro*

The first direct evidence of PgaE function and mechanism was the requirement for oxygen; the conversion of UWM6 by PgaE was inhibited in anaerobic conditions *in vitro*. In parallel, the reaction was dependent on equimolar input of NADPH as expected based on homology with flavoprotein hydroxylases and mono-oxygeneses. Direct confirmation of hydroxyl group addition by intermediate structure analysis was not possible because of the instability of the PgaE/UWM6 reaction product. This was, however, verified indirectly by LC-MS $^{18}O_2$ incorporation analysis of the coupled PgaE/PgaM reaction. It revealed that one of the two newly-attached oxygen atoms in gaudimycin C originated from O_2 (Table 1 in VI), in agreement with the expected hydroxylase function and the observed dependence for molecular oxygen.

Based on the structures of gaudimycin C and the corresponding *in vivo* (shunt) product rabelomycin, the target position for the hydroxylation was deduced to be C12. Apparently, the resulting UWM6 7,12-dihydroquinone is highly reactive, and spontaneously undergoes further reactions leading to breakdown *in vitro*, unless coupled to the subsequent enzyme-stabilized steps. The oxidation events which lead to the formation of the observed 7,12-quinone are not yet clearly understood, but may take place non-enzymatically if the rest of the molecule is correctly stabilized. Mechanistic features of the PgaE reaction were investigated in a structure based enzyme study (IV) presented in the following chapters.

4.3.5 PgaE/CabE crystal structure overview

The 3D structures PgaE and its homolog CabE were determined and studied in relation to characterized proteins similar in fold and function to understand the molecular basis of the catalysis (IV). The X-ray crystal structure of PgaE (PDB accession code 2qa1) was determined at 1.8Å resolution with multiple anomalous diffraction using selenomethionine derivative for phasing. The structure of CabE (PDB accession code 2qa2) was solved by molecular replacement at 2.7Å resolution using the PgaE structure as a template. Ligand-bound complexes were not obtained despite attempts to incorporate several different compounds into the crystals by co-crystallization and soaking.

The 3D structure analysis revealed PgaE and CabE to be highly similar members of para-hydroxybenzoate hydroxylase (pHBH) fold family (Figure 3 in IV). Three previously characterized members are flavin-dependent aromatic hydroxylases pHBH (Weijer *et al.*, 1983; Wierenga *et al.*, 1979), MHBH (Hiromoto *et al.*, 2006) and PHHY (Enroth *et al.*, 1998) which resemble the target enzymes in the overall fold, although primary sequence similarity is relatively low. The PgaE (Figure 10) and CabE subunits consist of three distinct domains. A large FAD binding domain, with the highest similarity with the structure homologs, is intervened in the primary structure by a smaller middle domain. The putative enzyme active site cavity is located at the interface of these two domains. The active site is hydrophobic in character and significantly large in comparison to the other members of the fold family, as expected by the relatively large size of the aromatic polyketide substrate. Most residues lining the active site cavity are conserved between the two proteins. The third, C-terminal, domain is of yet unknown function. It has a thioredoxin fold yet the putative catalytic residues have not been conserved. The native quaternary structure is a dimer formed *via* interactions at the FAD binding domain interface. The quaternary structures resemble pHBH but differ significantly from the other two members.

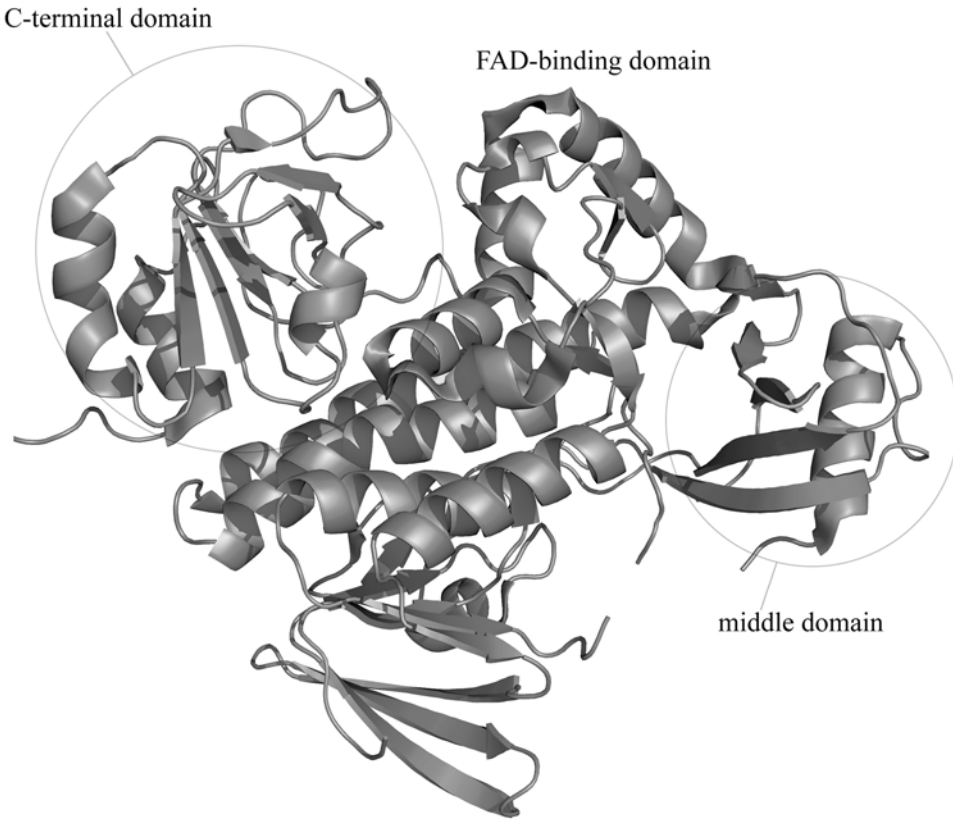

Figure 10. X-ray crystal structure of PgaE monomer showing the relative orientation of the three domains.

4.3.6 PgaE/CabE mutagenesis studies and mechanistic implications

The members of the fold family employ a complex catalytic mechanism in which the NADPH/FAD/O_2 redox cycle is coupled with conformational change to bring about substrate hydroxylation. This involves a transition between the so-called in-conformation in which the FAD isoalloxazine ring is tilted towards the substrate in the active site, and the out-conformation which allows contact between FAD and the co-substrate NADPH. The cycle is initiated when substrate binding triggers the transition towards the out-conformation, followed by the reduction of FAD by NADPH (reductive half-reaction). The resulting $FADH^-$ anion is attracted by the positive electrostatic field of the active site, which induces the switch back to the in-conformation. In the subsequent step the reduced flavin reacts with molecular oxygen (oxidative half-reaction), which eventually leads to flavin hydroperoxide formation and substrate hydroxylation. Structural analysis suggests that this model also applies for PgaE and CabE. The bound FAD is well defined in both crystal structures in the in-conformation. In one of the putative productive substrate binding modes found by molecular docking with PgaE, the FAD C4a which participates in oxygen activation is positioned in the proximity of substrate C12, the target for hydroxylation. Based on the structural data and comparison with the characterized homologs, several different PgaE

active site variants (His73Ala, Lys92Gln, Pro282Ser and Asn289Asp) were designed and prepared for biochemical analysis to study the catalytic mechanisms involved.

A key step in the pHBH redox cycle is an enzyme-catalyzed substrate deprotonation, which contributes to the transition from the in- to the out-conformation. In the modeled CabE and PgaE enzyme-substrate complexes there are only two suitably positioned residues, His73 and Lys92, which could potentially act as such catalytic bases. In targeted mutagenesis studies, however, the substitutions His73Ala and Lys92Gln were shown not to reduce the enzyme activity (Table 4). Another residue affiliated with the switch to the out-conformation is the strictly conserved residue corresponding to Pro282 in PgaE and CabE, which has been proposed to trigger the FAD movement as a response to substrate binding and deprotonation in pHBH. Substitution of the residue by serine has been reported to seriously impair the activity of pHBH (Palfey *et al.,* 2002) and PHHY (Xu *et al.,* 2002), but for PgaE the mutation (Pro282Ser) did not have an equivalent effect (Table 4). Taken together, the results imply that the enzyme i) may not require specific residues for substrate deprotonation, and ii) employs some alternative mechanism independent of interactions between Pro282 and the substrate to facilitate the FAD in/out conformational transition. Interestingly, two of the variants (His73Ala and Pro282Ser) were shown to have a somewhat increased catalytic activity in comparison to the wild type enzyme. As the measurements were carried out in substrate concentrations which are substantially higher than expected for physiological conditions, the changes are assumed to be due to k_{cat} rather than K_M. This suggests that these substitutions somehow decrease the energy barrier required to overcome to induce the conformational change, perhaps *via* favorable alterations in the electrostatic environment of the FAD binding site. Another residue conserved in the fold family is an asparagine (Asn289 in PgaE and CabE) located in the proximity of FAD isoalloxazine ring in the out-conformation. It is expected to have a role in aligning FAD in a favorable orientation for the reduction by NADPH. Substitution of the residue by a negatively charged aspartate (Asn289Asp) resulted in the drop of relative enzyme activity down to 1% in PgaE (Table 4), an effect equivalent to that reported for pHBH (Palfey *et al.,* 1994). The substitution apparently prevents the formation of a hydrogen bond between the residue 289 and the carbonyl oxygen of the cofactor, and at the same time destabilizes the nascent negative charge formed on FAD upon reduction.

Table 4. The specific activities of PgaE active site variants in relation to the wild type enzyme.

PgaE variant	Relative activity vs. wild type (%)
His73Ala	250
Lys92Gln	100
Pro282Ser	250
Asn289Asp	1

4.3.7 PgaM is a multifunctional oxygenase/reductase

Functional characterization of PgaM was complicated by the fact that stable substrate was not available, and the measurements had to be carried out in tandem with the PgaE reaction (VI). In addition, the number of sequential chemical changes in one reaction cycle had been found to be unusually high as seen in the structure of gaudimycin C, yet none of the intermediate products were sufficiently stable for isolation and analysis. Multiplicity of the reactions could be explained at least in part by the dipartite structure of PgaM, which contains two distinct putative functional domains (V; described in chapter 4.3.8). Although the sequential reactions could not be separated from each other, the results were consistent with earlier reports, supporting the role of the C-terminal SDR domain as a 2,3-dehydratase, and the N-terminal domain as a C12b hydroxylase (VI).

Perhaps as the most intriguing mechanistic finding, LC-MS analysis of reactions carried out in $H_2{}^{18}O$ unambiguously demonstrated that the C12b oxygen attached to gaudimycin C derives from water, not air (Table I in VI). This is in disagreement with the reactions in homologous pathways (Udvarnoki *et al.*, 1992), and inconsistent with any previously proposed reaction mechanism for the C12b hydroxylation (Udvarnoki *et al.*, 1992; Rix *et al.*, 2003). The experiments also confirmed several things about the remaining reaction step left unaccounted for, the C6 ketoreduction. First, the ketoreduction was shown to be stereospecific, and to yield a product which differs from the corresponding *cab* product in the orientation of the C6 hydroxyl substituent. Second, this hydroxyl group has been lost in the equivalent *in vivo* product gaudimycin A most likely as a result of endogenous dehydration by the host organism *S. lividans* TK24. In addition, all C12b-hydroxylated 12,7-quinone angucycline structures we found in on-line database structure searches had undergone ketoreduction at C6, suggesting that these two steps may somehow be mechanistically linked.

The observations led us to propose a reaction mechanism to explain the role of PgaM oxygenase domain in C12b hydroxylation and C6 ketoreduction in gaudimycin C biosynthesis (Figure 11). The mechanism combines the central concepts of reactive quinone methide (QM) intermediate formation observed for related anthracycline structures (Freccero, 2004), and a mechanism earlier described for an unrelated group of FAD-dependent hydroxylases involved in toxic phenol/benzo compound degradation (Cunane *et al.*, 2000; Fraaije & van Berkel, 1997; Hopper & Cottrell, 2003). The proposed reaction involves the generation of an angucycline QM intermediate, which leads to enzyme-stabilized incorporation of oxygen at C12b from a molecule of H_2O, accompanied by a mechanistically-coupled stereospecific ketoreduction at C6. Due to the presence of numerous simultaneous reactions, verification of the model by kinetic means has not been possible. However, enzyme structural information, together with detailed understanding of the FAD redox cycle in the future may allow us to examine and adjust the model further.

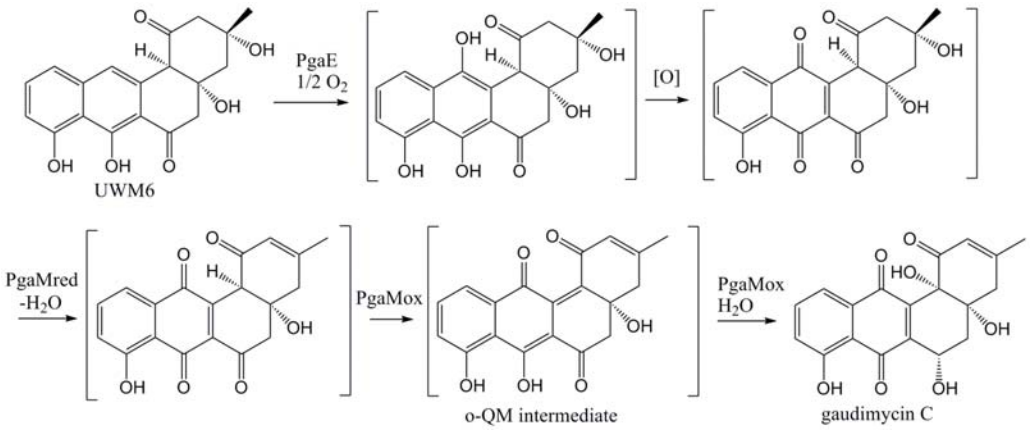

Figure 11. Presented model for PgaE and PgaM reactions (VI)

4.3.8 PgaM; genetic organization and quaternary complex

Besides the unexpected catalytic features associated with PgaM, the enzyme also has a distinct quaternary structure. The gene encoding for PgaM was found to exhibit an unusual DNA-level organization, which leads to the formation of two different translational products and their association in a non-covalent complex (V).

The gene *PgaM* has evolved as a consequence of a fusion between two consecutive ancestral reading frames. The stop codon at the end of the first gene (encoding for a FAD-dependent flavoprotein oxygenase; denoted as oxygenase) has been inactivated by a point mutation, while the start codon and Shine Dalgarno sequence in front of the downstream gene (encoding for a short-chain alcohol dehydrogenase/reductase; denoted as reductase) have remained intact. As a result, PgaM is expressed as a full-length oxygenase/reductase fusion, and as a separate reductase identical to the C-terminal domain *in vitro* (see Figure 2 in V). The two co-expressed forms associate into a native PgaM heterocomplex, which appears to have a structural role in the fusion enzyme stability. If isolated from one another, the solubility of each subunit is significantly decreased. The full-length form of PgaM alone was shown to retain the native enzyme activity (Figure 5 in V), suggesting that the separate reductase does not have a direct catalytic role. It however, seemed to have an effect on the PgaM FAD oxidation state (Figure 6 in V), yet the biological significance of this remains unclear. Based on the observations we proposed a quaternary complex model for PgaM, which explains the stability-related behavior of the protein *in vitro* (Figure 4 in V). According to the model, the separate reductase fragment is required to seal off the C-terminal end of the full-length protein to prevent PgaM oligomerization.

Sequence comparison between different biosynthetic routes reveals three alternative native forms of homologous systems (Table 1 in V), which represent successive stages of mechanistic adaptation in parallel angucycline pathways. In the ancestral form of the pathway the two genes are translated separately as independent functional units, as in simocyclinone (Trefzer *et al.*, 2002) pathway. PgaM represents the following genetic

variation after the loss of the intervening stop codon, an organization also found in landomycin (Zhu *et al.*, 2005) and urdamycin (Rix *et al.*, 2003) pathways. Presumably the most recently developed system is found in the auricin pathway (Novakova *et al.*, 2002) where the reductase start codon and ribosomal binding site have been lost, and apparently the corresponding protein is expressed exclusively in the fusion form.

4.4 Enzyme stereoselectivity and stereospecificity

Stereochemistry is an important concept in natural product research, which tightly connects the disciplines of medical research and enzymology. Enzyme-catalyzed reactions specify the overall stereochemical structure of biosynthetic products, which in turn determines their biological properties, interactions with target receptor molecules, binding sites and enzymes (Kulig *et al.*, 2004; Lu, 2007). In relation to the present work, this is exemplified by the effect of C9 stereochemistry on the pharmacological properties of aclarubicin and nogalamycin (Arora, 1983), the respective products of the *akn* and *sno* pathways, or the difference in C4' stereochemistry between epirubicin and doxorubicin (Jain *et al.*, 1985; Weiss *et al.*, 1986).

The stereochemistry of aromatic polyketide compounds is defined in various stereoselective enzymatic cyclization, oxidation and ketoreduction steps, which give rise to pathway-specific chiral centers on the aglycone ring frame. Opposite stereochemical enzyme systems are found in parallel biosynthetic routes, and represent an important structural element in natural product diversity. In addition to SnoaL and AknH which define the above-mentioned C9 stereochemistry in nogalamycin and aclacinomycin biosynthesis (III; Torkkell *et al.*, 2000), PgaM and its homolog CabM are responsible for the opposite C6 configurations in angucycline biosynthesis (VI; Palmu *et al.*, 2007). In analogy, the orientation of C3 substituent in dihydrogranaticin and medermycin biosynthesis is determined by gra-ORF5/6 (Taguchi *et al.*, 2001) and med-ORF12 (Li *et al.*, 2005), homologs of actVI-ORF1 ketoreductase in actinorhodin biosynthesis, respectively. The stereoselective properties appear to be, at least in some cases, specified by a relatively small number of residues. As a result, detailed understanding of the interactions may enable targeted enzyme modification towards a favoured stereochemical outcome (III; Nakajima *et al.*, 1998; Nakajima *et al.*, 1999).

In parallel to the end product stereoselectivity, many enzymes are stereospecific towards a certain chiral form of the substrate (*cf.* substrate specificity; chapter 4.5.2). Consequently, the stereochemical structure may have a profound effect on the compatibility between consecutive biosynthetic steps, but at present, the influence of compound structure on possible steric or chemical constraints cannot be easily determined without experimental case-studies. RdmE, for example, was shown to hydroxylate only the C9-R cyclization product of NAME *in vitro*, whereas the opposite C9-S isoform was left completely unreacted (III). In analogy, the C6 hydroxyl substituent of the PgaM reaction product was lost as a result of endogenous dehydroxylation *in vivo*, yet on the opposite stereoisomer of the *cab* pathway retained the corresponding OH group in equivalent conditions (VI; Palmu *et al.*, 2007). In other

cases, chirality may have no evident impact on enzyme function, as exemplified by late-stage tailoring enzymes in doxorubicin biosynthesis, which have been reported to be unaffected by the change in substrate C4' stereochemistry (Madduri *et al.*, 1998). Streospecificity is fundamentally determined by the same physicochemical parameters as substrate specificity in general, discussed in chapter 4.5.2.

4.5 Common enzymatic themes and repeated phenomena

4.5.1 Introduction: Enzyme-level flexibility

Secondary metabolism pathways are relatively susceptible to changes, because i) mutations are not expected to cause acute impairment of essential cellular functions as in primary metabolism which stringently regulates growth and development, and ii) alteration increases the potential for chemical diversity which may (in simplified terms) lead to selective advantage. This has steered the evolution of the secondary metabolism pathways, and can be seen as certain "developmental freedom" at different organizational levels. Enzyme-level studies of type II aromatic polyketide pathways have provided evidence of various common characteristics which support this concept, although due to insufficient reference material, generalizations are difficult to make.

In context with this work, the following chapters summarize some related themes associated with the enzymes. These include, specifically, determinants of substrate specificity, and enzyme mechanistic evolution.

4.5.2 Relaxed enzyme substrate specificity

Enzymes involved in different secondary metabolism pathways have been suggested to share a common characteristic of relaxed (or broad) substrate specificity, an ability to convert a range of different precursors (Firn & Jones, 2003). This property is one of the central axioms in *Screening Hypothesis*, a simplified model to explain the multiplicity of naturally occurring secondary metabolites (Firn & Jones, 2000; Firn & Jones, 2003). According to the model, broad enzyme substrate specificity is favored as a way to efficiently increase chemical diversity at low metabolic cost (Firn & Jones, 2003) (as opposed to the strictly defined reactions in primary metabolism where such redundancy could have drastic effects and impair viability). In essence, it allows the generation of new functional gene combinations in nature, and specifically, explains how the genetic flexibility characteristic to *Streptomyces* biosynthetic clusters (discussed in chapter 1.2.4) is manifested at enzyme level. To provide further product diversity, it also enables enzymes to convert biosynthetic intermediates of parallel co-existing pathways, a phenomenon frequently observed as unexpected endogenous reactions *in vivo*. In parallel, broad substrate specificity is a basis for various research applications that employ genetic engineering techniques to combine and reorganize genetic material from different biosynthetic origins.

As postulated, it has become a generally accepted fact that enzymes involved in type II aromatic polyketide biosynthesis display relatively relaxed substrate specificity.

Provided that the atoms which directly encage in the reaction are preserved, many of the tailoring enzymes tolerate rather extensive variation in substrate chemical structure, size (Kim *et al.*, 2007; Pacholec *et al.*, 2005), aglycone ring frame topology (Decker & Haag, 1995), substitution (Dickens *et al.*, 1997; Jansson *et al.*, 2004) and/or glycosylation patterns (Jansson *et al.*, 2003b). In the present work the property was demonstrated by AknH *in vitro* (III). It was shown to readily convert NAME, which in comparison to the biological substrate AAME, has a methyl instead of an ethyl substituent attached on the reacting carbon C9.

The underlying determinants of relaxed substrate specificity have not been well described, yet recently, enzyme structure/function characterization has revealed some general factors concerning substrate binding, active site topology and ligand geometry. The enzyme binding pockets are typically composed of batches of hydrophobic residues, which pack against the non-polar aromatic aglycone ring system and the sugar residues. Active site amino acid conservation between orthologous/homologous enzymes is rather low, yet most mutations are conservative, and retain the hydrophobic nature of the pocket (Beinker *et al.*, 2006; Jansson *et al.*, 2003b; Jansson *et al.*, 2004). In agreement with the expected relaxed substrate specificity, substrate recognition and binding are often primarily defined by relatively unspecific interactions (II; Alexeev *et al.*, 2007; Jansson *et al.*, 2003b; Jansson *et al.*, 2004; Jansson *et al.*, 2005; Sciara *et al.*, 2003;). The main physical forces involved are hydrophobic and van der Waals interactions (Bolam *et al.*, 2007), and instead of a network of strictly defined hydrogen bonds, recognition appears to be determined by the overall topology, polarity and shape of the active site cavity. Typically only one or few of many possible hydrogen bonds are observed between the ligand and the active site residues (II; Jansson *et al.*, 2003b; Jansson *et al.*, 2004; Jansson *et al.*, 2005), although additional indirect hydrogen bonds may be formed *via* water molecules located in the active site.

The research has in some cases provided direct molecular-level information on why some potential ligands do not function as substrates for specific enzymes. For example RdmC, which catalyses C15 demethoxylation of various anthracycline precursors, only converts compounds which are glycosylated. Although the sugar residues point away from the active site cavity and do not directly participate in the reaction, they have an important orientational role in substrate alignment through a hydrogen bond formed between the enzyme and the first carbohydrate moiety (Jansson *et al.*, 2003b). In comparison, unlike RdmC which does not discriminate between mono- and triglycosylated ligands, DnrK exclusively accepts substrates with a single sugar substituent. Structural analysis has suggested that the alignment of larger glycosyl groups is prevented by enzyme conformational change induced upon binding (Jansson *et al.*, 2004). Detailed understanding on such interactions could generate new possibilities for rational pathway design, and tools to replace approaches which simply rely on trial and error. This has already been demonstrated in practice, as structural information on substrate recognition and catalysis has been used to screen for functional ligands for DnrK *in silico* (Kim *et al.*, 2007).

4.5.3 Unusual enzyme mechanisms and substrate-induced stabilization

Characterization of enzymes involved in aromatic polyketide biosynthesis has revealed a number of unusual catalytic mechanisms, which differ from those described for typical model enzymes or homologs of known function. Based on detailed structure/ function analysis some of these enzymes lack cofactors usually required for catalysis. Besides SnoaL (II) and AknH (III), examples include oxygenase Act VA-Orf6 (Sciara *et al.*, 2003) and aclacinomycin 10-hydroxylase RdmB (Jansson *et al.*, 2005). In several other cases, specific catalytic residues (indispensable for catalysis) have not been detected in combined structure comparison and targeted mutagenesis analysis. This has been demonstrated by PgaE (IV) tetracenomycin F2 cyclase (Thompson *et al.*, 2004) and O-methyltransferase DnrK in daunorubicin biosynthesis (Jansson *et al.*, 2004) in which specific activities were only moderately affected upon substitution of putative catalytic bases essential for substrate deprotonation in related proteins. In addition, there are examples of enzymes which differ mechanistically from their homologs in some other aspects. These include PgaM described in this study (V-VI) and RdmB, which closely resembles typical methyltransferases in structure (Jansson *et al.*, 2003a), but exclusively catalyses hydroxylation (Jansson *et al.*, 2005). These unusual enzymes do not share a common fold or function, and intriguingly, the only obvious unifying theme is that they all catalyse reactions of related highly conjugated polycyclic substrates.

It appears that the observed mechanistic properties are at least in part linked with the chemical nature of the substrate ring scaffold. It has been postulated that the substrate may adopt a direct catalytic role in reaction initiation and/or transition state stabilization (II; Fetzner, 2007; Jansson *et al.*, 2005; Sciara *et al.*, 2003), in analogy to the function of transition metal ions, flavin or pterin cofactors used by many enzymes. As per the organic cofactors, aromatic polyketides are characterized by polycyclic structures with multiple conjugated double bonds. For example AKV contains eight conjugated double bonds distributed over a four-ringed antracene quinone frame. Such structures allow efficient delocalization of the aglycone π electrons over an extended aromatic or aromatic-like system, which promotes charge distribution and stability of deprotonated reaction intermediates. This in turn reduces the free energy of transition state which must be overcome to drive the reaction forward. The phenomenon was first observed with the cofactorless oxygenase Act VA-Orf6 from *S. coelicolor*, which employs the substrate ring system to overcome the energy barrier of oxygen activation (Sciara *et al.*, 2003). An equivalent mode of action has been reported for aclacinomycin 10-hydroxylase (RdmB) (Jansson *et al.*, 2005), which uses delocalization of the π electrons over the substrate aromatic ring system for carbanion intermediate stabilization and oxygen activation. This model of substrate-induced stabilization has been mostly discussed in context with enzymes involved with oxygen chemistry (Fetzner, 2007) but as exemplified by SnoaL (II), the principle is not confined to this group of enzymes alone.

Additional reactivity is provided by the presence of multiple hydroxyl and keto groups, typical substituents of polyketide pathway intermediates. They contribute to charge distribution in the ring system by enabling keto-enol tautomerization, and in parallel, may facilitate substrate deprotonation. Specifically, α-hydrogen atoms adjacent to

carbonyl oxygen groups are expected to be removed with relative ease in the hydrophobic environment of the active site pocket. This deprotonation would produce a reactive carbanion species able to readily undergo further reactions. Such "substrate-facilitated proton abstraction" may, in part, explain the redundancy of typical catalytic residues observed in some enzymes, and has been suggested as the initial reaction step in the cyclization reaction catalyzed by SnoaL (II) and hydroxylation reaction catalysed by PgaM (VI).

4.5.4 Multifunctional enzymes

Another intriguing property associated with some of the enzymes is multifunctionality; participation of a single gene product in several sequential reactions. Such enzymes have been found in various type II polyketide biosynthetic pathways (VI; Alexeev *et al.*, 2007; Chen *et al.*, 2005; Dickens *et al.*, 1997; Mayer *et al.*, 2005; Rafanan *et al.*, 2000), and take part in different oxygenation, reduction and/or dehydration reactions. The experimental findings presented in this study showed that the reaction sequence catalyzed by PgaE and PgaM involve five distinct reactions, two hydroxylation steps, hydroxyquinone oxidation, ketoreduction and dehydration (VI). The possible mechanisms of this multistep reaction have been covered in chapters 4.3.4 and 4.3.7. Other examples include homologous enzymes TcmG and ElmG, which both catalyse three stepwise hydroxylations in tetracenomycin biosynthesis (Rafanan *et al.*, 2000). The reactions follow an intricate monoxygenase-dioxygenase mechanism, involving oxirane ring opening by a molecule of H_2O to introduce two oxygen atoms from O_2 and a single water-derived oxygen. Also cytochrome P-450 enzyme DoxA in doxorubicin biosynthesis is responsible for three oxygenation events. It catalyses hydroxylation and subsequent keto group formation at C13, together with the spontaneous dehydration of a dihydroxy intermediate, and hydroxylation of C14 (Dickens *et al.*, 1997; Walczak *et al.*, 1999). The phenomenon is not restricted to the aromatic substrate systems alone, but is also observed for example in FAD-dependent oxygenase EncM (*S. maritimus*) which catalyses *in vivo* multiple reactions on non-aromatic, yet extremely reactive substrate in enterocin biosynthesis (Xiang *et al.*, 2004).

Although in some multifunctional enzymes the specific residues responsible for different reactions have been identified (Alexeev *et al.*, 2007), the number of putative catalytic sites does not always correlate with the number of observed changes. This suggests that the reactions may involve mechanistically interlinked "spontaneous" non-enzymatic chemical steps induced by the inherent instability of the intermediate products. Such unspecific (and in this case unwanted) reactivity is characteristic for the PKS steps, but also observed in association with the post-PKS tailoring reactions as, for example, dehydration (II-III), more extensive degradation (VI), or other changes (Jansson *et al.*, 2003b) *in vitro*. It could be argued that the instability is one of the key factors which has contributed to the evolvement of multifunctional enzymes in aromatic polyketide biosynthesis; in combination with relaxed substrate specificity, spontaneous reactions would be expected to be an efficient way to increase natural secondary metabolite diversity.

5. CONCLUDING REMARKS

The experimental work presented in this thesis provides detailed information on the enzymology of specific reaction steps, mechanisms and interactions in selected type II polyketide biosynthesis routes. The results emphasize the importance of individual enzyme case-studies in the field of antibiotic research, because they are required to build up a complete, molecular-level blueprint of the reaction pathways. The high-resolution structure/function analysis in the first part of the study (AknH and SnoaL) allowed us to identify some of the amino acid determinants of cyclase stereoselectivity in anthracycline biosynthesis. The findings implicated that the reaction product stereochemistry, a central chemical feature which largely determines the biological activity of the pathway products, may in some cases be specified by only a small number of residues. Such information could potentially be used for targeted enzyme-level modification to steer a biosynthetic reaction towards a more desired stereochemical outcome. The enzyme *in vitro* characterization in the second part of the study (PgaE and PgaM) provided insight on the complex post-PKS oxygenation cascade common to many angucycline pathways. The work underlined the concepts of timing and coupling in successive reactions, and demonstrated that the biosynthetic steps are not always separable from one another, and may involve complex chemistry difficult to resolve *in vivo*.

The results presented here are not merely isolated examples of specific phenomena, but may provide direct evidence of how these biosynthetic pathways have evolved. While the morphological, biochemical and genetic diversity of *Streptomyces* have been widely documented, reports on enzyme-level adaptation and diversification are still limited. This study described several cases of enzyme mechanistic adaptation: In comparison to characterized model enzymes, the polyketide cyclases SnoaL and AknH are devoid of specific functional elements (cofactor/amino-acid) usually required for intermediate stabilization, whereas PgaE and CabE appear to use an alternative mechanism for the initial catalytic step of substrate deprotonation. PgaM employs an unusual flavoprotein redox cycle and oxygen chemistry to bring about hydroxylation, and at the same time presents an interesting case of structural evolution - translational fusion of two ancestral reading frames. These divergences illustrate that the enzymes in type II aromatic polyketides are subjected to change, perhaps as a result of the need to adapt to rapidly changing biosynthetic contexts in response to gene-level flexibility. Two central factors which may contribute to the process of adaptation are high intermediate reactivity and broad enzyme substrate specificity characteristic to the biosynthetic steps. However, more extensive enzyme-level research is needed to understand the process in order to make valid generalizations.

ACKNOWLEDGEMENTS

The experimental work presented in this thesis was conducted over the years 2002-2007 at the Department of Biochemistry and Food Chemistry, University of Turku, Finland. The work was carried out under the supervision of Ph.D. Jarmo Niemi and Professor Emeritus Pekka Mäntsälä in the research project Antibiotic Biosynthetic Enzymes (www.sci.utu.fi/projects/biokemia/abe) in close cooperation (I-IV) with the Molecular Structural Biology research laboratory (Department of Medical Biochemistry and Biophysics, Karolinska Institute, Sweden; http://phillips.mbb.ki.se). The work was funded by The National Graduate School in Informational and Structural Biology (ISB), Academy of Finland, and Turku University Foundation.

I want to thank my supervisor Jarmo Niemi, and the group leaders Pekka Mäntsälä and Gunter Schneider for the opportunity to work in such an inspiring project! Thank you for your support and encouragement in the experimental work, as well as for your contribution to the writing process of the manuscripts.

I want to express my gratitude and respect to all the hardworking people who participated in the work here and in Sweden. Special commendation for the pleasant cooperation and friendship to Azmiri Sultana, Hanna Koskiniemi, Anna Jansson and the rest of the structural biology crew. Also a big handshake to my comrade Liu Zhanliang for the mutual scientific engagement. A particular superthanks to Ph.D. Mikko Metsä-Ketelä for inspiration, expertise, critical views, and a shoulder to cry against.

I also want to acknowledge the rest of the people in our research team and the department for providing an excellent working environment for scientific research. This includes all my fellow researchers, the professors Jyrki Heino and Reijo Lahti, the M.Sc. students I have been privileged to work with. Not forgetting the administrative and the technical staff. And of course, an additional thank you to Anssi Malinen for the weekly therapy sessions and help in enlightning new potential paths for my future scientific studies.

I want to thank the reviewers of my thesis, Professor Lloyd Ruddock and Professor Pirkko Heikinheimo for all the effort, constructive criticism and valuable comments.

Also cheers to my ISB committee supervisor Ph.D. Tiina Salminen, and all the rest of the friends I have made in ISB over the years. ISB rules, no question!

For most of my achievements I owe to people outside the scientific field, in particular my family and friends who have provided invaluable support along the journey! Thank you for everything! My deepest gratitude and respect belong to Liinu and the kids; for standing by me, understanding me and standing up with me!

REFERENCES

Abadi, J. (1995) Amphotericin B. *Pediatr. Rev.* **16**, 363-364.

Abdelfattah, M., Maskey, R.P., Asolkar, R.N., Grun-Wollny, I. & Laatsch, H. (2003) Seitomycin: Isolation, structure elucidation and biological activity of a new angucycline antibiotic from a terrestrial streptomycete. *J. Antibiot. (Tokyo)* **56**, 539-542.

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. U. S. A.* **104**, 6170-6175.

al-Tai, A., Kim, B., Kim, S.B., Manfio, G.P. & Goodfellow, M. (1999) *Streptomyces malaysiensis* sp. nov., a new streptomycete species with rugose, ornamented spores. *Int. J. Syst. Bacteriol.* **49**, 1395-1402.

Alvarez, H.M. & Steinbuchel, A. (2002) Triacylglycerols in prokaryotic microorganisms. *Appl. Microbiol. Biotechnol.* **60**, 367-376.

Ames, B.D., Korman, T.P., Zhang, W., Smith, P., Vu, T., Tang, Y. & Tsai, S.C. (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5349-5354.

Anderson, A.S. & Wellington, E.M. (2001) The taxonomy of *Streptomyces* and related genera. *Int. J. Syst. Evol. Microbiol.* **51**, 797-814.

Antal, N., Fiedler, H.P., Stackebrandt, E., Beil, W., Stroch, K. & Zeeck, A. (2005) Retymicin, galtamycin B, saquayamycin Z and ribofuranosyllumichrome, novel secondary metabolites from *Micromonospora* sp. Tu 6368. I. Taxonomy, fermentation, isolation and biological activities. *J. Antibiot. (Tokyo)* **58**, 95-102.

Aouali, N., Eddabra, L., Macadre, J. & Morjani, H. (2005) Immunosuppressors and reversion of multidrug-resistance. *Crit. Rev. Oncol. Hematol.* **56**, 61-70.

Arabolaza, A., Rodriguez, E., Altabe, S., Alvarez, H. & Gramajo, H. (2008) Multiple pathways for triacylglycerol biosynthesis in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* **74**, 2573-2582.

Arora, S.K. (1983) Molecular structure, absolute stereochemistry, and interactions of nogalamycin, a DNA-binding anthracycline antitumor antibiotic. *J. Am. Chem. Soc.* **105**, 1328-1332.

Arora, S.K. (1985) Structure of aklavinone, a DNA binding anthracycline antibiotic. *J. Antibiot. (Tokyo)* **38**, 1788-1791.

Ausubel, F.M., Brent, R., Kinston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. (1994) *Current protocols in molecular biology* John Wiley & Sons Inc. / USA.

Ayer, S.W., McInnes, A.G., Thibault, P., Walter, J.A., Doull, J.L., Parnell, T. & Vining, L.C. (1991) Jadomycin, a novel 8H-benz[b]oxazolo[3,2-f]phenanthridine antibiotic from from *Streptomyces venezuelae* ISP5230. *Tetrahedron Lett.* **32**, 6301.

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.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. & Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147.

Berdy, J. (2005) Bioactive microbial metabolites. *J. Antibiot. (Tokyo)* **58**, 1-26.

Bolam, D.N., Roberts, S., Proctor, M.R., Turkenburg, J.P., Dodson, E.J., Martinez-Fleites, C., Yang, M., Davis, B.G., Davies, G.J. & Gilbert, H.J. (2007) The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5336-5341.

Box, V.G. (2007) The intercalation of DNA double helices with doxorubicin and nogalamycin. *J. Mol. Graph. Model.* **26**, 14-19.

Bringmann, G., Lang, G., Maksimenka, K., Hamm, A., Gulder, T.A., Dieter, A., Bull, A.T., Stach, J.E., Kocher, N., Muller, W.E. & Fiedler, H.P. (2005) Gephyromycin, the first bridged angucyclinone, from *Streptomyces griseus* strain NTK 14. *Phytochemistry* **66**, 1366-1373.

Challis, G.L. (2008a) Genome mining for novel natural product discovery. *J. Med. Chem.* **51**, 2618-2628.

Challis, G.L. (2008b) Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* **154**, 1555-1569.

Chen, B., Peng, X., Pentassuglia, L., Lim, C.C. & Sawyer, D.B. (2007) Molecular and cellular mechanisms of anthracycline cardiotoxicity. *Cardiovasc. Toxicol.* **7**, 114-121.

Chen, C.W., Huang, C.H., Lee, H.H., Tsai, H.H. & Kirby, R. (2002) Once the circle has been broken: Dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet.* **18**, 522-529.

Chen, Y.H., Wang, C.C., Greenwell, L., Rix, U., Hoffmeister, D., Vining, L.C., Rohr, J. & Yang, K.Q. (2005) Functional analyses of oxygenases in jadomycin biosynthesis and identification of JadH as a bifunctional oxygenase/dehydrase. *J. Biol. Chem.* **280**, 22508-22514.

Chopra, I. & Roberts, M. (2001) Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232-260.

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.

Connors, N.C. & Strohl, W.R. (1993) Partial purification and properties of carminomycin 4-O-methyltransferase from *Streptomyces* sp. strain C5. *J. Gen. Microbiol.* **139**, 1353-1362.

Corre, C. & Challis, G.L. (2007) Heavy tools for genome mining. *Chem. Biol.* **14**, 7-9.

Cunane, L.M., Chen, Z.W., Shamala, N., Mathews, F.S., Cronin, C.N. & McIntire, W.S. (2000) Structures of the flavocytochrome p-cresol methylhydroxylase and its enzyme-substrate complex: Gated substrate entry and proton relays support the proposed catalytic mechanism. *J. Mol. Biol.* **295**, 357-374.

Curtis, T.P., Sloan, W.T. & Scannell, J.W. (2002) Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10494-10499.

Dann, M., Lefemine, D.V., Barbatschi, F., Shu, P., Kunstmann, M.P., Mitscher, L.A. & Bohonos, N. (1965) Tetrangomycin, a new quinone antibiotic. *Antimicrob. Agents Chemother. (Bethesda)* **5**, 832-835.

Decker, H. & Haag, S. (1995) Cloning and characterization of a polyketide synthase gene from *Streptomyces fradiae* Tu2717, which carries the genes for biosynthesis of the angucycline antibiotic urdamycin A and a gene probably involved in its oxygenation. *J. Bacteriol.* **177**, 6126-6136.

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.

Drautz, H., Zahner, H., Rohr, J. & Zeeck, A. (1986) Metabolic products of microorganisms. 234. Urdamycins, new angucycline antibiotics from *Streptomyces fradiae*. I. Isolation, characterization and biological properties. *J. Antibiot. (Tokyo)* **39**, 1657-1669.

Dutcher, J.D. (1968) The discovery and development of amphotericin B. *Dis. Chest* **54**, 296-298.

Egan, S., Wiener, P., Kallifidas, D. & Wellington, E.M. (2001) Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. *Antonie Van Leeuwenhoek* **79**, 127-133.

Enroth, C., Neujahr, H., Schneider, G. & Lindqvist, Y. (1998) The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. *Structure* **6**, 605-617.

Fetzner, S. (2007) Cofactor-independent oxygenases go it alone. *Nat. Chem. Biol.* **3**, 374-375.

Firn, R.D. & Jones, C.G. (2003) Natural products--a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20**, 382-391.

Firn, R.D. & Jones, C.G. (2000) The evolution of secondary metabolism - a unifying model. *Mol. Microbiol.* **37**, 989-994.

Fischer, G., Decaris, B. & Leblond, P. (1997a) Occurrence of deletions, associated with genetic instability in *Streptomyces ambofaciens*, is independent of the linearity of the chromosomal DNA. *J. Bacteriol.* **179**, 4553-4558.

Fischer, G., Kyriacou, A., Decaris, B. & Leblond, P. (1997b) Genetic instability and its possible evolutionary implications on the chromosomal structure of *Streptomyces*. *Biochimie* **79**, 555-558.

Fischer, G., Wenner, T., Decaris, B. & Leblond, P. (1998) Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14296-14301.

Fraaije, M.W. & van Berkel, W.J. (1997) Catalytic mechanism of the oxidative demethylation of 4-(methoxymethyl)phenol by vanillyl-alcohol oxidase. Evidence for formation of a p-quinone methide intermediate. *J. Biol. Chem.* **272**, 18111-18116.

Freccero, M. (2004) Quinone methides as alkylating and cross-linking agents. *Mini-Reviews in Organic Chemistry* **1**, 403-415.

Gans, J., Wolinsky, M. & Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**, 1387-1390.

Gianni, L., Herman, E.H., Lipshultz, S.E., Minotti, G., Sarvazyan, N. & Sawyer, D.B. (2008) Anthracycline cardiotoxicity: From bench to bedside. *J. Clin. Oncol.* **26**, 3777-3784.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S.G. (1996) Life with 6000 genes. *Science* **274**, 546, 563-567.

Gravius, B., Glocker, D., Pigac, J., Pandza, K., Hranueli, D. & Cullum, J. (1994) The 387 kb linear plasmid pPZG101 of *Streptomyces rimosus* and its interactions with the chromosome. *Microbiology* **140**, 2271-2277.

Gross, H., Stockwell, V.O., Henkels, M.D., Nowak-Thompson, B., Loper, J.E. & Gerwick, W.H. (2007) The genomisotopic approach: A systematic method to isolate products of orphan biosynthetic gene clusters. *Chem. Biol.* **14**, 53-63.

Han, L., Yang, K., Ramalingam, E., Mosher, R.H. & Vining, L.C. (1994) Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in *Streptomyces venezuelae* ISP5230. *Microbiology* **140**, 3379-3389.

Henkel, T., Rohr, J., Beale, J.M. & Schwenen, L. (1990) Landomycins, new angucycline antibiotics from *Streptomyces* sp. I. Structural studies on landomycins A-D. *J. Antibiot. (Tokyo)* **43**, 492-503.

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.

Hideg, K. & Kalai, T. (2007) Novel antioxidants in anthracycline cardiotoxicity. *Cardiovasc. Toxicol.* **7**, 160-164.

Higuchi, R., Krummel, B. & Saiki, R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: Study of protein and DNA interactions. *Nucleic Acids Res.* **16**, 7351-7367.

Hiromoto, T., Fujiwara, S., Hosokawa, K. & Yamaguchi, H. (2006) Crystal structure of 3-hydroxybenzoate hydroxylase from *Comamonas testosteroni* has a large tunnel for substrate and oxygen access to the active site. *J. Mol. Biol.* **364**, 878-896.

Hopper, D.J. & Cottrell, L. (2003) Alkylphenol biotransformations catalyzed by 4-ethylphenol methylenehydroxylase. *Appl. Environ. Microbiol.* **69**, 3650-3652.

Hopwood, D.A. (1965) A circular linkage map in the actinomycete *Streptomyces coelicolor*. *J. Mol. Biol.* **12**, 514-516.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Burton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. & Schrempf, H. (1985) *Genetic manipulation of Streptomyces: A laboratory manual*. John Innes Foundation / Norwich, UK.

Hopwood, D.A. (2006) Soil to genomics: The *Streptomyces* chromosome. *Annu. Rev. Genet.* **40**, 1-23.

Hopwood, D.A. (1999) Forty years of genetics with *Streptomyces*: From *in vivo* through *in vitro* to *in silico*. *Microbiology* **145**, 2183-2202.

Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A., Floss, H.G. & Omura, S. (1985) Production of 'hybrid' antibiotics by genetic engineering. *Nature* **314**, 642-644.

Hutchinson, C.R. & McDaniel, R. (2001) Combinatorial biosynthesis in microorganisms as a route to new antimicrobial, antitumor and neuroregenerative drugs. *Curr. Opin. Investig Drugs* **2**, 1681-1690.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Omura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526-531.

Jain, K.K., Casper, E.S., Geller, N.L., Hakes, T.B., Kaufman, R.J., Currie, V., Schwartz, W., Cassidy, C., Petroni, G.R. & Young, C.W. (1985) A prospective randomized comparison of epirubicin and doxorubicin in patients with advanced breast cancer. *J. Clin. Oncol.* **3**, 818-826.

Jain, R., Rivera, M.C., Moore, J.E. & Lake, J.A. (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**, 1598-1602.

Jain, R., Rivera, M.C., Moore, J.E. & Lake, J.A. (2002) Horizontal gene transfer in microbial genome evolution. *Theor. Popul. Biol.* **61**, 489-495.

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.

Jansson, A., Koskiniemi, H., Mäntsälä, P., Niemi, J. & Schneider, G. (2004) Crystal structure of a ternary complex of DnrK, a methyltransferase in daunorubicin biosynthesis, with bound products. *J. Biol. Chem.* **279**, 41149-41156.

Jansson, A., Niemi, J., Lindqvist, Y., Mäntsälä, P. & Schneider, G. (2003a) 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., Niemi, J., Mäntsälä, P. & Schneider, G. (2003b) Crystal structure of aclacinomycin methylesterase with bound product analogues: Implications for anthracycline recognition and mechanism. *J. Biol. Chem.* **278**, 39006-39013.

Kantola, J., Kunnari, T., Mäntsälä, P. & Ylihonko, K. (2003) Expanding the scope of aromatic polyketides by combinatorial biosynthesis. *Comb. Chem. High Throughput Screen.* **6**, 501-512.

Karanes, C., Crowley, J., Sawkar, L., Tranum, B.L., Hicks, W.J., Bonnet, J.D., Dabich, L., Dana, B.W. & Salmon, S. (1990) Aclacinomycin A in the treatment of multiple myeloma: A Southwest Oncology Group study. *Invest. New Drugs* **8**, 101-104.

Karoonuthaisiri, N., Weaver, D., Huang, J., Cohen, S.N. & Kao, C.M. (2005) Regional organization of gene expression in *Streptomyces coelicolor*. *Gene* **353**, 53-66.

Keefe, D.L. (2001) Anthracycline-induced cardiomyopathy. *Semin. Oncol.* **28**, 2-7.

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.

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. & Hopwood, D.A. (2000) *Practical Streptomyces genetics*. The John Innes Foundation / Norwich, England.

Kim, N.Y., Kim, J.H., Lee, Y.H., Lee, E.J., Kim, J., Lim, Y., Chong, Y. & Ahn, J.H. (2007) O-methylation of flavonoids using DnrK based on molecular docking. *J. Microbiol. Biotechnol.* **17**, 1991-1995.

Kinashi, H., Shimaji-Murayama, M. & Hanafusa, T. (1992) Integration of SCP1, a giant linear plasmid, into the *Streptomyces coelicolor* chromosome. *Gene* **115**, 35-41.

Korman, T.P., Tan, Y.H., Wong, J., Luo, R. & Tsai, S.C. (2008) Inhibition kinetics and emodin cocrystal structure of a type II polyketide ketoreductase. *Biochemistry* **47**, 1837-1847.

Korynevska, A., Heffeter, P., Matselyukh, B., Elbling, L., Micksche, M., Stoika, R. & Berger, W. (2007) Mechanisms underlying the anticancer activities of the angucycline landomycin E. *Biochem. Pharmacol.* **74**, 1713-1726.

Kulig, K., Nowicki, P. & Malawska, B. (2004) Influence of the absolute configuration on pharmacological activity of antihypertensive and antiarrhythmic drugs. *Pol. J. Pharmacol.* **56**, 499-508.

Kunnari, T., Kantola, J., Ylihonko, K., Klika, K.D., Mäntsälä, P. & Hakala, J. (1999) Hybrid compounds derived from the combination of anthracyclin and actinorhodin biosynthetic pathways. *J. Chem. Soc. Perkin Trans* **2**, 1649-1652.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F. & Danchin, A. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249-256.

Kurosawa, K., Ghiviriga, I., Sambandan, T.G., Lessard, P.A., Barbara, J.E., Rha, C. & Sinskey, A.J. (2008) Rhodostreptomycins, antibiotics biosynthesized following horizontal gene transfer from *Streptomyces padanus* to *Rhodococcus fascians*. *J. Am. Chem. Soc.* **130**, 1126-1127.

Leblond, P., Francou, F.X., Simonet, J.M. & Decaris, B. (1990) Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofaciens* strains. *FEMS Microbiol. Lett.* **60**, 79-88.

Leimkuhler, C., Fridman, M., Lupoli, T., Walker, S., Walsh, C.T. & Kahne, D. (2007) Characterization of rhodosaminyl transfer by the AknS/AknT glycosylation complex and its use in reconstituting the biosynthetic pathway of aclacinomycin A. *J. Am. Chem. Soc.* **129**, 10546-10550.

Li, A., Itoh, T., Taguchi, T., Xiang, T., Ebizuka, Y. & Ichinose, K. (2005) Functional studies on a ketoreductase gene from *Streptomyces* sp. AM-7161 to control the stereochemistry in medermycin biosynthesis. *Bioorg. Med. Chem.* **13**, 6856-6863.

Lin, Y.S., Kieser, H.M., Hopwood, D.A. & Chen, C.W. (1993) The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol. Microbiol.* **10**, 923-933.

Lomovskaya, N., Otten, S.L., Doi-Katayama, Y., Fonstein, L., Liu, X.C., Takatsu, T., Inventi-Solari, A., Filippini, S., Torti, F., Colombo, A.L. & Hutchinson, C.R. (1999) Doxorubicin overproduction in *Streptomyces peucetius*: Cloning and characterization of the *dnrU* ketoreductase and *dnrV* genes and the *doxA* cytochrome P-450 hydroxylase gene. *J. Bacteriol.* **181**, 305-318.

Lu, H. (2007) Stereoselectivity in drug metabolism. *Expert Opin. Drug Metab. Toxicol.* **3**, 149-158.

Lu, W., Leimkuhler, C., Oberthur, M., Kahne, D. & Walsh, C.T. (2004) AknK is an L-2-deoxyfucosyltransferase in the biosynthesis of the anthracycline aclacinomycin A. *Biochemistry* **43**, 4548-4558.

Lundqvist, T., Rice, J., Hodge, C.N., Basarab, G.S., Pierce, J. & Lindqvist, Y. (1994) Crystal structure of scytalone dehydratase--a disease determinant of the rice pathogen, *Magnaporthe grisea*. *Structure* **2**, 937-944.

Madduri, K., Kennedy, J., Rivola, G., Inventi-Solari, A., Filippini, S., Zanuso, G., Colombo, A.L., Gewain, K.M., Occi, J.L., MacNeil, D.J. & Hutchinson, C.R. (1998) Production of the antitumor drug epirubicin (4'-epidoxorubicin) and its precursor by a genetically engineered strain of *Streptomyces peucetius*. *Nat. Biotechnol.* **16**, 69-74.

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.

Manfio, G.P., Zakrzewska-Czerwinska, A.,E. & Goodfellow, M. (1995) Towards minimal standards for the description of *Streptomyces* species. *Biotekhnologia* **7-8**, 242-253.

Mayer, A., Taguchi, T., Linnenbrink, A., Hofmann, C., Luzhetskyy, A. & Bechthold, A. (2005) LanV, a bifunctional enzyme: Aromatase and ketoreductase during landomycin A biosynthesis. *Chembiochem* **6**, 2312-2315.

McDaniel, R., Ebert-Khosla, S., Hopwood, D.A. & Khosla, C. (1995) Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature* **375**, 549-554.

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

McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M. & Ashley, G. (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1846-1851.

McGuire, J.M., Bunch, R.L., Anderson, R.C., Boaz, H.E., Flynn, E.H., Powell, H.M. & Smith, J.W. (1952) Ilotycin, a new antibiotic. *Schweiz. Med. Wochenschr.* **82**, 1064-1065.

Metsä-Ketelä, M., Niemi, J., Mäntsälä, P. & Schneider, G. (2007) Anthracycline Biosynthesis: Genes, Enzymes and Mechanisms in *Topics in Current Chemistry.* Springer Berlin / Heidelberg.

Metsä-Ketelä, M., Halo, L., Munukka, E., Hakala, J., Mäntsälä, P. & Ylihonko, K. (2002) Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. *Appl. Environ. Microbiol.* **68**, 4472-4479.

Metsä-Ketelä, M., Palmu, K., Kunnari, T., Ylihonko, K. & Mäntsälä, P. (2003) Engineering anthracycline biosynthesis toward angucyclines. *Antimicrob. Agents Chemother.* **47**, 1291-1296.

Metsä-Ketelä, M., Salo, V., Halo, L., Hautala, A., Hakala, J., Mäntsälä, P. & Ylihonko, K. (1999) An efficient approach for screening minimal PKS genes from *Streptomyces*. *FEMS Microbiol. Lett.* **180**, 1-6.

Metsä-Ketelä, M., Ylihonko, K. & Mäntsälä, P. (2004) Partial activation of a silent angucycline-type gene cluster from a rubromycin beta producing *Streptomyces* sp. PGA64. *J. Antibiot. (Tokyo)* **57**, 502-510.

Mittler, M., Bechthold, A., Schulz, G.E. (2007) Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin *J. Mol. Biol.* **372**, 67-76.

Miyamoto, Y., Ohta, S., Johdo, O., Nagamatsu, Y. & Yoshimoto, A. (2000) Production of a new hybrid anthracycline 4-O-methylepelmycin by heterologous expression of *dnrK* in epelmycin-producing *Streptomyces violaceus*. *J. Antibiot. (Tokyo)* **53**, 828-836.

Nakajima, K., Kato, H., Oda, J., Yamada, Y. & Hashimoto, T. (1999) Site-directed mutagenesis of putative substrate-binding residues reveals a mechanism controlling the different stereospecificities of two tropinone reductases. *J. Biol. Chem.* **274**, 16563-16568.

Nakajima, K., Yamashita, A., Akama, H., Nakatsu, T., Kato, H., Hashimoto, T., Oda, J. & Yamada, Y. (1998) Crystal structures of two tropinone reductases: Different reaction stereospecificities in the same protein fold. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 4876-4881.

Newman, D.J. & Cragg, G.M. (2007) Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **70**, 461-477.

Niemi, J., Kantola, J., Ylihonko, K. & Mäntsälä, P. (2002) Anthracycline biosynthesis: Steps, enzymes and genes in *Microbial Secondary Metabolites: Biosynthesis, Genetics and Regulation* (F. Fierro & J.F. Martin, eds.) pp. 122-139, Research Signpost / Kerala, India.

Niemi, J. & Mäntsälä, P. (1995) Nucleotide sequences and expression of genes from *Streptomyces purpurascens* that cause the production of new anthracyclines in *Streptomyces galilaeus*. *J. Bacteriol.* **177**, 2942-2945.

Niemi, J., Wang, Y., Airas, K., Ylihonko, K., Hakala, J. & Mäntsälä, P. (1999) Characterization of aklavinone-11-hydroxylase from *Streptomyces purpurascens*. *Biochim. Biophys. Acta* **1430**, 57-64.

Nitiss, J.L., Pourquier, P. & Pommier, Y. (1997) Aclacinomycin A stabilizes topoisomerase I covalent complexes. *Cancer Res.* **57**, 4564-4569.

Nobili, S., Landini, I., Giglioni, B. & Mini, E. (2006) Pharmacological strategies for overcoming multidrug resistance. *Curr. Drug Targets* **7**, 861-879.

Novakova, R., Bistakova, J., Homerova, D., Rezuchova, B. & Kormanec, J. (2002) Cloning and characterization of a polyketide synthase gene cluster involved in biosynthesis of a proposed angucycline-like polyketide auricin in *Streptomyces aureofaciens* CCM 3239. *Gene* **297**, 197-208.

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M. & Horinouchi, S. (2008) The genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **190**, 4050-4060.

Olukoshi, E.R. & Packter, N.M. (1994) Importance of stored triacylglycerols in *Streptomyces*: Possible carbon source for antibiotics. *Microbiology* **140**, 931-943.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. & Hattori, M. (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12215-12220.

Pacholec, M., Tao, J. & Walsh, C.T. (2005) CouO and NovO: C-methyltransferases for tailoring the aminocoumarin scaffold in coumermycin and novobiocin antibiotic biosynthesis. *Biochemistry* **44**, 14969-14976.

Palfey, B.A., Basu, R., Frederick, K.K., Entsch, B. & Ballou, D.P. (2002) Role of protein flexibility in the catalytic cycle of p-hydroxybenzoate hydroxylase elucidated by the Pro293Ser mutant. *Biochemistry* **41**, 8438-8446.

Palfey. B.A., Entsch, B., Ballou, D.P. & Massey. V. (1994) Changes in the catalytic properties of p-hydroxybenzoate hydroxylase caused by the mutation Asn300Asp. *Biochemistry* **33**, 1545-1554.

Palmu, K. & Kunnari, T. (2002) *Gene cluster for rabelomycin biosynthesis and its use to generate compounds for drug screening.* **PCT/FI2002/000214**.

Palmu, K., Ishida, K., Mäntsälä, P., Hertweck, C. & Metsä-Ketelä, M. (2007) Artificial reconstruction of two cryptic angucycline antibiotic biosynthetic pathways. *Chembiochem* **8**, 1577-1584.

Pandza, S., Biukovic, G., Paravic, A., Dadbin, A., Cullum, J. & Hranueli, D. (1998) Recombination between the linear plasmid pPZG101 and the linear chromosome of *Streptomyces rimosus* can lead to exchange of ends. *Mol. Microbiol.* **28**, 1165-1176.

Rafanan, E.R.,Jr, Hutchinson, C.R. & Shen, B. (2000) Triple hydroxylation of tetracenomycin A2 to tetracenomycin C involving two molecules of O_2 and one molecule of H_2O. *Org. Lett.* **2**, 3225-3227.

Räty, K., Hautala, A., Torkkell, S., Kantola, J., Mäntsälä, P., Hakala, J. & Ylihonko, K. (2002a) Characterization of mutations in aclacinomycin A-non-producing *Streptomyces galilaeus* strains with altered glycosylation patterns. *Microbiology* **148**, 3375-3384.

Räty, K., Kantola, J., Hautala, A., Hakala, J., Ylihonko, K. & Mäntsälä, P. (2002b) 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.

Remsing, L.L., Gonzalez, A.M., Nur-e-Alam, M., Fernandez-Lozano, M.J., Brana, A.F., Rix, U., Oliveira, M.A., Mendez, C., Salas, J.A. & Rohr, J. (2003) Mithramycin SK, a novel antitumor drug with improved therapeutic index, mithramycin SA, and demycarosyl-mithramycin SK: Three new products generated in the mithramycin producer *Streptomyces argillaceus* through combinatorial biosynthesis. *J. Am. Chem. Soc.* **125**, 5745-5753.

Ridley, C.P., Lee, H.Y. & Khosla, C. (2008) Evolution of polyketide synthases in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4595-4600.

Rix, U., Remsing, L.L., Hoffmeister, D., Bechthold, A. & Rohr, J. (2003) Urdamycin L: A novel metabolic shunt product that provides evidence for the role of the urdM gene in the urdamycin A biosynthetic pathway of *Streptomyces fradiae* TU 2717. *Chembiochem* **4**, 109-111.

Rix, U., Wang, C., Chen, Y., Lipata, F.M., Remsing Rix, L.L., Greenwell, L.M., Vining, L.C., Yang, K. & Rohr, J. (2005) The oxidative ring cleavage in jadomycin biosynthesis: A multistep oxygenation cascade in a biosynthetic black box. *Chembiochem* **6**, 838-845.

Rohr, J. & Thiericke, R. (1992) Angucycline group antibiotics. *Nat. Prod. Rep.* **9**, 103-137.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular cloning: A laboratory manual (2nd edition).* Cold Spring Harbor Laboratory Press / Cold Spring Harbor, New York.

Saunders, R.N., Metcalfe, M.S. & Nicholson, M.L. (2001) Rapamycin in transplantation: A review of the evidence. *Kidney Int.* **59**, 3-16.

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.

Sehgal, S.N. (2003) Sirolimus: Its discovery, biological properties, and mechanism of action. *Transplant. Proc.* **35**, 7S-14S.

Shah, F. (1998) Erythromycin. *Pediatr. Rev.* **19**, 140-141.

Skerman, V.B.D., McGowan, V. & Sneath, P.H.A. (1980) Approved lists of bacterial names. *Int. J. Syst. Bacteriol.* **30**, 225-420.

Skovsgaard, T. & Nissen, N.I. (1975) Adriamycin, an antitumour antibiotic: A review with special reference to daunomycin. *Dan. Med. Bull.* **22**, 62-73.

Sultana, A., Alexeev, I., Kursula, I., Mäntsälä, P., Niemi, J. & Schneider, G. (2007) Structure determination by multiwavelength anomalous diffraction of aclacinomycin oxidoreductase: Indications of multidomain pseudomerohedral twinning. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 149-159.

Taguchi, T., Ebizuka, Y., Hopwood, D.A. & Ichinose, K. (2001) A new mode of stereochemical control revealed by analysis of the biosynthesis of dihydrogranaticin in *Streptomyces violaceoruber* Tu22. *J. Am. Chem. Soc.* **123**, 11376-11380.

Thompson, T.B., Katayama, K., Watanabe, K., Hutchinson, C.R. & Rayment, I. (2004) Structural and functional analysis of tetracenomycin F2 cyclase from *Streptomyces glaucescens*. A type II polyketide cyclase. *J. Biol. Chem.* **279**, 37956-37963.

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.

Torkkell, S., Kunnari, T., Palmu, K., Mäntsälä, P., Hakala, J. & Ylihonko, K. (2001) The entire nogalamycin biosynthetic gene cluster of *Streptomyces nogalater*: Characterization of a 20-kb DNA region and generation of hybrid structures. *Mol. Genet. Genomics* **266**, 276-288.

Torkkell, S., Ylihonko, K., Hakala, J., Skurnik, M. & Mäntsälä, P. (1997) Characterization of *Streptomyces nogalater* genes encoding enzymes involved in glycosylation steps in nogalamycin biosynthesis. *Mol. Gen. Genet.* **256**, 203-209.

Torsvik, V., Goksoyr, J. & Daae, F.L. (1990) High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**, 782-787.

Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C., Fiedler, H.P., Welzel, K., Vente, A. & Bechthold, A. (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob. Agents Chemother.* **46**, 1174-1182.

Udvarnoki, G., Henkel, T., Machinek, R. & Rohr, J. (1992) Biosynthetic origin of the oxygen atoms in aquayamycin: Aspects for the biosynthesis of the urdamycin family and for aquayamycin-containing angucycline antibiotics in general. *J. Org. Chem.* **57**, 1274-1276.

Vezina, C., Kudelski, A. & Sehgal, S.N. (1975) Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *J. Antibiot. (Tokyo)* **28**, 721-726.

Volff, J.N. & Altenbuchner, J. (2000) A new beginning with new ends: Linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol. Lett.* **186**, 143-150.

Volff, J.N. & Altenbuchner, J. (1998) Genetic instability of the *Streptomyces* chromosome. *Mol. Microbiol.* **27**, 239-246.

Wakabayashi, T., Oki, T., Tone, H., Hirano, S. & Omori, K. (1980) A comparative electron microscopic study of aclacinomycin and adriamycin cardiotoxicities in rabbits and hamsters. *J. Electron. Microsc. (Tokyo)* **29**, 106-118.

Waksman, S.A., Lechevalier, H.A. & Schaffner, C.P. (1965) Candicidin and other polyenic antifungal antibiotics. *Bull. World Health Organ.* **33**, 219-226.

Walczak, R.J., Dickens, M.L., Priestley, N.D. & Strohl, W.R. (1999) Purification, properties, and characterization of recombinant *Streptomyces* sp. strain C5 DoxA, a cytochrome P-450 catalyzing multiple steps in doxorubicin biosynthesis. *J. Bacteriol.* **181**, 298-304.

Wang, J., Maitani, Y. & Takayama, K. (2002) Antitumor effects and pharmacokinetics of aclacinomycin A carried by injectable emulsions composed of vitamin E, cholesterol, and PEG-lipid. *J. Pharm. Sci.* **91**, 1128-1134.

Wang, S.J., Chang, H.M., Lin, Y.S., Huang, C.H. & Chen, C.W. (1999) *Streptomyces* genomes: Circular genetic maps from the linear chromosomes. *Microbiology* **145**, 2209-2220.

Watve, M.G., Tickoo, R., Jog, M.M. & Bhole, B.D. (2001) How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176**, 386-390.

Weber, T., Welzel, K., Pelzer, S., Vente, A. & Wohlleben, W. (2003) Exploiting the genetic potential of polyketide producing streptomycetes. *J. Biotechnol.* **106**, 221-232.

Weijer, W.J., Hofsteenge, J., Beintema, J.J., Wierenga, R.K. & Drenth, J. (1983) p-Hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. 2. Fitting of the amino-acid sequence to the tertiary structure. *Eur. J. Biochem.* **133**, 109-118.

Weiss, R.B., Sarosy, G., Clagett-Carr, K., Russo, M. & Leyland-Jones, B. (1986) Anthracycline analogs: The past, present, and future. *Cancer Chemother. Pharmacol.* **18**, 185-197.

Wenner, T., Roth, V., Fischer, G., Fourrier, C., Aigle, B., Decaris, B. & Leblond, P. (2003) End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens*. *Mol. Microbiol.* **50**, 411-425.

Westrich, L., Domann, S., Faust, B., Bedford, D., Hopwood, D.A. & Bechthold, A. (1999) Cloning and characterization of a gene cluster from *Streptomyces cyanogenus* S136 probably involved in landomycin biosynthesis. *FEMS Microbiol. Lett.* **170**, 381-387.

Widenbrant, E.M., Tsai, H.H., Chen, C.W. & Kao, C.M. (2007) *Streptomyces coelicolor* undergoes spontaneous chromosomal end replacement. *J. Bacteriol.* **189**, 9117-9121.

Wierenga, R.K., de Jong, R.J., Kalk, K.H., Hol, W.G. & Drenth, J. (1979) Crystal structure of p-hydroxybenzoate hydroxylase. *J. Mol. Biol.* **131**, 55-73.

Wiley, P.F., Kelly, R.B., Caron, E.L., Wiley, V.H., Johnson, J.H., MacKellar, F.A. & Mizsak, S.A. (1977) Structure of nogalamycin. *J. Am. Chem. Soc.* **99**, 542-549.

Williams, L.D., Egli, M., Qi, G., Bash, P., van der Marel, G.A., van Boom, J.H., Rich, A. & Frederick, C.A. (1990) Structure of nogalamycin bound to a DNA hexamer. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2225-2229.

Xiang, L., Kalaitzis, J.A. & Moore, B.S. (2004) EncM, a versatile enterocin biosynthetic enzyme involved in *Favorskii* oxidative rearrangement, aldol condensation, and heterocycle-forming reactions. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15609-15614.

Xu, D., Enroth, C., Lindqvist, Y., Ballou, D.P. & Massey, V. (2002) Studies of the mechanism of phenol hydroxylase: effect of mutation of proline 364 to serine. *Biochemistry* **41**, 13627-13636.

Yadav, G., Gokhale, R.S. & Mohanty, D. (2003) Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**, 335-363.

Ylihonko, K., Hakala, J., Kunnari, T. & Mäntsälä, P. (1996) Production of hybrid anthracycline antibiotics by heterologous expression of *Streptomyces nogalater* nogalamycin biosynthesis genes. *Microbiology* **142**, 1965-1972.

Ylihonko, K., Hakala, J., Niemi, J., Lundell, J. & Mäntsälä, P. (1994) Isolation and characterization of aclacinomycin A-non-producing *Streptomyces galilaeus* (ATCC 31615) mutants. *Microbiology* **140**, 1359-1365.

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.

Zakeri, B. & Wright, G.D. (2008) Chemical biology of tetracycline antibiotics. *Biochem. Cell Biol.* **86**, 124-136.

Zhu, L., Ostash, B., Rix, U., Nur-E-Alam, M., Mayers, A., Luzhetskyy, A., Mendez, C., Salas, J.A., Bechthold, A., Fedorenko, V. & Rohr, J. (2005) Identification of the function of gene *lndM2* encoding a bifunctional oxygenase-reductase involved in the biosynthesis of the antitumor antibiotic landomycin E by *Streptomyces globisporus* 1912 supports the originally assigned structure for landomycinone. *J. Org. Chem.* 70, 631-638.