Studying Protein Interactions
of a Bacterial Type II Polyketide Synthase

A thesis submitted to the University of London for the degree of Doctor of Philosophy

by

Gaetano Castaldo

Microbiology Group
Department of Pharmaceutics
School of Pharmacy
University of London
29/39 Brunswick Square
London WC1N 1AX
United Kingdom

September 2007
DECLARATION

This thesis describes the research conducted at the School of Pharmacy, University of London between September 2003 and June 2007 under the supervision of Dr. Paul Long. I certify that the research described is original and that any parts of the work that have been conducted by collaboration are clearly indicated. I also certify that I have written all the text herein and have clearly indicated by suitable citation any part of this dissertation that has already appeared in publication.

Signature

Date 7/03/2008
'Niente potrà mai allontanare Nino dal Sole' 

'Nothing will ever take Nino far away from the Sun'

Valerio Gigliotti, 1996
ACKNOWLEDGEMENTS

I would like first to thank my supervisor Dr. Paul F. Long for his invaluable support, enthusiasm and for making me understand the real meaning of the expression ‘no pain, no gain’. I would also like to thank my co-supervisor Dr. John Crosby of University of Bristol for his suggestions and for mass spectrometry analysis. Also, thanks to Mr. Pakorn Wattana-Amorn of University of Bristol for the kind gift of TEV protease. I would like to thank Mr. Jurica Zucko and Professor Daslav Hranueli from University of Zagreb (Croatia) for the support about homology modelling. Thanks to Dr. Smith and Professor Anne Stephenson for useful advices for the yeast two-hybrid system. An huge thanks goes to Sibylle Heidelberger for the invaluable help for the very final mass spectrometry analysis. A special thank to Dr. Duska Vujaklija for showing me the art, if the not the science, of preparing *Streptomyces* protoplasts. Thanks also to the School of Pharmacy for the economic support.

Thank to all the nice people that have meet over 4 years in the Microbiology Group. In particular, I would like to thank Prof. Peter Taylor and for his useful advices Dr Paul Stapleton. A special thanks goes to Tsveta and Natasa for nurturing me and for endless amusement. Also, I would like to thank everyone at the School of Pharmacy for making these 4 years such an unforgettable time. Thanks to everyone for the good time in London in particular to Mr CSI Fiata for the incredibly funny nights out all over London and suburbs (mm mm).

I want to reserve a special thank to my flatmates Sonja and David for being absolutely great during all this time we have lived together. I will always be indebted to you guys as well as to Linda and John for their constant support and help in every occasion. Sorry for the pain I have caused since the very first months in London (e.g. fridge, freezer)! Hope Linda is going to get her third PhD soon! I consider all 4 (soon more…) of you like a branch of my family. I must admit without you my time in London would have been more and more difficult!

A special thank goes to all my good friends in Italy. In particular, I want to thank Daniela ‘bionda’ for our invaluable and unique friendship. Also, thank to Luca ‘il pazzone’ for the support and enthusiasm. Thanks to all the people of Naples for welcoming and leaving me always with an everlasting smile.
Finally, my biggest thank goes to my dad Giancarlo and my mum Maria for putting up with an absent son (one day I’ll come back…) and to my brother Dario (D8 meditronic dubwise) for making me feel like I have never left home. Also, thank to all my relatives for their enthusiasm and infinite support. In particular, I want to dedicate this thesis to my grandfather Antonio still waiting for his grandson to bring him a medicine that could ‘take away from him 20-30 years’! Without all of you I couldn’t have done it and I would never be what I am.


ABSTRACT

Daunorubicin (DNR) and its C-14 hydroxylated derivative doxorubicin (DXR) are the most widely used anthracyclines as anti-tumour agents. DNR and DXR are produced by the soil bacteria *Streptomyces peucetius* through a biosynthetic pathway that employs a type II polyketide synthase (PKS). Type II PKSs consist of several discrete, monofunctional proteins that form a dissociable complex. Studies on enzyme complex formation and substrate channelling are essential for a better understanding of metabolism and could lead to the generation of novel compounds by ‘combinatorial’ biosynthesis. The 21-carbon atom aromatic polyketide, aklanonic acid, is the first enzyme-free intermediate in the DNR/DXR biosynthetic pathway. The DNR/DXR PKS is composed of eight proteins: DpsABCDG form the ‘minimal’ PKS, responsible for the condensation reactions between the propionate starter unit and nine malonate extender units, whereas DpsEFY catalyse successive modifications of the carbon backbone such as reduction, aromatisation and cyclisation. This type II PKS is intriguing since it contains a type III ketosynthase (DpsC) that selects the starter unit and a putative malonyl/acyltransferase (DpsD) whose role seems obscure.

The network of protein interactions within this complex have been investigated using a yeast two-hybrid system (GAL4), affinity chromatography (Tandem Affinity Purification) and computer-aided protein docking simulations (Hex 4.5 software). The results have led to the proposal of a head-to-tail arrangement for the ‘minimal’ PKS suggested by the interactions established by DpsA, DpsB and DpsD. A putative role for DpsD is suggested as physical inhibitor of the incorporation of acetate in the priming reaction allowing the choice of propionate. Also, a structural role for the cyclase DpsY is proposed perhaps to maintain the overall structural integrity of the complex. This represents the first study attempting to analyse *in vivo* protein interactions forming a type II PKS. The purification method has allowed isolation of DpsA and DpsB and *in silico* docking simulations have produced results consistent with the proposed arrangements of proteins based on yeast two-hybrid assays.
# TABLE OF CONTENTS

DECLARATION .................................................................................................................. 2
ACKNOWLEDGEMENTS ................................................................................................. 4
PUBLICATIONS .............................................................................................................. 6
ABSTRACT ..................................................................................................................... 7
TABLE OF CONTENTS ................................................................................................ 8
INDEX OF FIGURES ..................................................................................................... 12
INDEX OF TABLES ....................................................................................................... 14
ABBREVIATIONS ......................................................................................................... 15

## CHAPTER 1
Introduction .................................................................................................................. 17

### 1.1 Secondary metabolism and natural products ....................................................... 17
#### 1.1.1 Streptomyces and antibiotic production ......................................................... 17

### 1.2 Fatty acid biosynthesis ....................................................................................... 20

### 1.3 Polyketide biosynthesis ..................................................................................... 22
#### 1.3.1 Type I polyketide synthases ......................................................................... 23
##### 1.3.1.1 Erythromycin polyketide synthase ...................................................... 24
##### 1.3.1.2 Other Type I PKSs ............................................................................ 25
#### 1.3.2 Type II Polyketide synthases ...................................................................... 28
##### 1.3.2.1 ‘Minimal’ PKS ................................................................................. 29
##### 1.3.2.2 Ketoreductases ................................................................................. 33
##### 1.3.2.3 Cyclases .......................................................................................... 34
##### 1.3.2.4 Aromatases ...................................................................................... 36
#### 1.3.3 Type III Polyketide synthases ................................................................... 38

### 1.4 Combinatorial biosynthesis .............................................................................. 38
#### 1.4.1 Engineering Type I PKSs .......................................................................... 39
##### 1.4.1.1 DEBS 1-TE ..................................................................................... 39
##### 1.4.1.2 Alteration of AT domain ..................................................................... 41
##### 1.4.1.3 Swapping Loading modules ................................................................ 41
##### 1.4.1.4 Incorporation of alternative starter units ......................................... 42
##### 1.4.1.5 Modification of post-PKS tailoring steps ........................................... 43
#### 1.4.2 Combinatorial potential of Type I PKS ....................................................... 43
##### 1.4.2.1 Structure of Type I PKS multienzyme .............................................. 45
##### 1.4.2.2 Protein interactions on DEBS ............................................................ 47
#### 1.4.3 Engineering Type II PKSs ........................................................................ 48

### 1.5 DNR/DXR Polyketide synthase .......................................................................... 52
#### 1.5.1 Clinical use of DNR and DXR ................................................................. 56
#### 1.5.2 DNR and DXR analogues ......................................................................... 58

### 1.6 Aim of the project ............................................................................................. 59

## CHAPTER 2
Materials and Methods ............................................................................................... 61

### 2.1 Materials ............................................................................................................ 61
#### 2.1.1 Chemicals and reagents .............................................................................. 61
##### 2.1.1.2 Oligonucleotides ............................................................................ 61
#### 2.1.2 Buffers and solutions .................................................................................. 61
#### 2.1.3 Culture media ............................................................................................ 62
#### 2.1.4 Antibiotic concentrations .......................................................................... 64
#### 2.1.5 Source of enzymes ..................................................................................... 64

### 2.2 Growth and manipulation of microorganisms .................................................... 64
#### 2.2.1 Strains .......................................................................................................... 65
#### 2.2.2 Plasmids ....................................................................................................... 66
2.2.3 Manipulation of *Escherichia coli* cultures

2.2.3.1 Preparation and transformation of chemically competent *E. coli* cells

2.2.4 Manipulation of *Saccharomyces cerevisiae* cultures

2.2.4.1 Preparation and transformation of *S. cerevisiae* competent cells

2.2.5 Manipulation of *Streptomyces coelicolor* cultures

2.2.5.1 Growth of *S. coelicolor* cultures

2.2.5.2 Preparation and transformation of *S. coelicolor* A3(2) protoplasts

2.3 DNA manipulation techniques

2.3.1 Plasmid DNA preparation

2.3.2 DNA restriction digestion

2.3.3 Agarose gel electrophoresis and recovery of DNA

2.3.4 Ligation of DNA

2.3.5 DNA sequencing

2.3.6 Polymerase Chain Reaction (PCR)

2.4 Yeast two-hybrid system methods

2.4.1 Extraction of proteins from *S. cerevisiae

2.4.2 *LacZ* assay

2.5 Protein chemistry

2.5.1 Protein quantification

2.5.2 SDS-Polyacrylamide Gel Electrophoresis

2.5.3 Western Blot analysis

2.5.4 Source of antibodies

2.6 RNA methods

2.6.1 Extraction of RNA from *S. coelicolor

2.6.2 Electrophoresis of RNA through Formaldehyde gels

2.6.3 DNAse treatment of RNA

2.6.4 Reverse transcriptase (RT-PCR)

2.7 Tap tag method

2.7.1 Growth of *S. coelicolor* cells

2.7.2 Preparation of the protein extracts from *S. coelicolor* cells

2.7.3 Purification of Tap tagged protein

2.7.3.1 Binding to IgG Sepharose

2.7.3.2 Cleavage with Tobacco etch virus (TEV) protease

2.7.3.3 Binding to Calmodulin Sepharose

2.7.3.4 Elution from Calmodulin Sepharose

2.8 Protein identification by mass spectrometry

2.8.1 Excision and destaining of protein bands following SDS-PAGE

2.8.2 In-gel trypsin digestion

2.8.2.1 Reduction and alkylation of proteins

2.8.2.2 Protein digestion and extraction of digested peptides

2.9 Computer Software

CHAPTER 3

Investigation of protein interactions of the 'minimal' DNR/DXR PKS using yeast two-hybrid system

3.1 Introduction to protein interactions

3.2 Introduction to the yeast two-hybrid system

3.2.1 Yeast transcription factors

3.2.2 Yeast two-hybrid system

3.2.3 Reporter genes and yeast strain
3.2.4 MATCHMAKER™ GAL4 yeast two-hybrid system 3 (Clontech) ...... 90
   3.2.4.1 Yeast strain AH109 ................................................................. 90
   3.2.4.2 GAL4 expression vectors ........................................................... 91
3.2.5 Sensitivity and limitations of the yeast two-hybrid system .......... 95
3.2.6 Type II PKS and DNR/DXR PKS .................................................... 97

3.3 Results ............................................................................................................. 98
   3.3.1 PCR amplification of dps genes .............................................................. 98
   3.3.2 Cloning strategies for the dps genes ....................................................... 98
   3.3.3 DNA sequencing .................................................................................... 102
   3.3.4 Expression of recombinant proteins ..................................................... 102
   3.3.5 Yeast two-hybrid system assay .............................................................. 106
       3.3.5.1 Nutritional selection .................................................................... 109
       3.3.5.2 β-galactosidase activity ............................................................... 109

3.4 Discussion ....................................................................................................... 112

CHAPTER 4 .............................................................................................................. 124
Investigation of protein interactions of the entire DNR/DXR PKS using yeast
two-hybrid system ................................................................................................. 124
4.1 Introduction ...................................................................................................... 124
4.2 Results ............................................................................................................. 124
   4.2.1 PCR amplification and cloning of the dps genes ..................................... 124
   4.2.2 Expression of recombinant proteins ................................................... 125
   4.2.3 Yeast two-hybrid system assay ............................................................ 130
       4.2.3.1 Nutritional selection ................................................................... 130
       4.2.3.2 β-galactosidase activity ............................................................... 132

4.3 Discussion ....................................................................................................... 136

CHAPTER 5 .............................................................................................................. 141
Tandem Affinity Purification of the ‘minimal’ DNR/DXR PKS .................... 141
5.1 Purification of protein complexes .............................................................. 141
   5.1.2 Tandem Affinity Purification ................................................................. 141
       5.1.2.1 Sensitivity and limitations of the TAP method .................................. 144
       5.1.2.2 TAP method for ‘minimal’ DNR/DXR PKS ................................... 145
5.2 Results ............................................................................................................. 145
   5.2.1 Cloning strategy for TAP tagging of DpsA and expression of the
       ‘minimal’ DNR/DXR PKS ....................................................................... 145
   5.2.2 Considerations upon tcmJ and design of oligonucleotide primers .......... 150
   5.2.3 Expression of the ‘minimal’ DNR/DXR PKS ......................................... 155
   5.2.4 TAP method .......................................................................................... 157
5.3 Discussion ....................................................................................................... 162

CHAPTER 6 .............................................................................................................. 168
Investigation of Dps proteins interactions by computational methods .......... 168
6.1 Introduction ...................................................................................................... 168
   6.1.2 Protein databases .................................................................................... 169
   6.1.3 Docking methods .................................................................................... 170
       6.1.3.1 The Hex Program ......................................................................... 172
   6.1.4 First approach mode and DeepView modelling ..................................... 173
   6.1.5 Use of computer-aided tools for the investigation of protein interaction
       between Dps proteins ............................................................................... 175
6.2 Results ............................................................................................................. 175
   6.2.1 Prediction of tertiary structures of Dps proteins .................................... 175
       6.2.1.1 DpsA and DpsB ........................................................................... 175
       6.2.1.2 DpsD .......................................................................................... 178
       6.2.1.3 DpsE .......................................................................................... 178
INDEX OF FIGURES

Figure 1.1 ............................................................................................................................19
Figure 1.2 ............................................................................................................................21
Figure 1.3 ............................................................................................................................26
Figure 1.4 ............................................................................................................................31
Figure 1.5 ............................................................................................................................37
Figure 1.6 ............................................................................................................................40
Figure 1.7 ............................................................................................................................46
Figure 1.8 ............................................................................................................................49
Figure 1.9 ............................................................................................................................50
Figure 1.10 ...........................................................................................................................53
Figure 1.11 ...........................................................................................................................57
Figure 3.1 ............................................................................................................................86
Figure 3.2 ............................................................................................................................89
Figure 3.3 ............................................................................................................................92
Figure 3.4 ............................................................................................................................93
Figure 3.5 ............................................................................................................................94
Figure 3.6 ............................................................................................................................96
Figure 3.7 ...........................................................................................................................100
Figure 3.8 ..........................................................................................................................101
Figure 3.9 ..........................................................................................................................103
Figure 3.10 ..........................................................................................................................104
Figure 3.11 ..........................................................................................................................107
Figure 3.12 ..........................................................................................................................108
Figure 3.13 ..........................................................................................................................110
Figure 3.14 ..........................................................................................................................111
Figure 3.15 ..........................................................................................................................113
Figure 3.16 ..........................................................................................................................115
Figure 3.17 ..........................................................................................................................117
Figure 3.18 ..........................................................................................................................118
Figure 3.19 ..........................................................................................................................119
Figure 3.20 ..........................................................................................................................123
Figure 4.1 ............................................................................................................................128
Figure 4.2 ............................................................................................................................129
Figure 4.3 ............................................................................................................................131
Figure 4.4 ............................................................................................................................133
Figure 4.5 ............................................................................................................................134
Figure 4.6 ............................................................................................................................135
Figure 4.7 ............................................................................................................................137
Figure 5.1 ............................................................................................................................142
Figure 5.2 ............................................................................................................................147
Figure 5.3 ............................................................................................................................148
Figure 5.4 ............................................................................................................................149
Figure 5.5 ............................................................................................................................151
Figure 5.6 ............................................................................................................................152
Figure 5.7 ............................................................................................................................153
Figure 5.8 ............................................................................................................................154
Figure 5.9 ............................................................................................................................156
Figure 5.10 ..........................................................................................................................158
Figure 5.11 ..........................................................................................................................159
INDEX OF TABLES

Table 3.1 .............................................................................................................................. 99
Table 3.2 .............................................................................................................................. 105
Table 4.1 .............................................................................................................................. 126
Table 4.2 .............................................................................................................................. 127
ABBREVIATIONS

In addition to those already listed in *Biochemical Journal*, the following abbreviations are used in this thesis:

3D three-dimensional
6-dEB 6-deoxyerythronolide B
6MSA 6-methylsalicylic acid
AA aklanonic acid
ACP acyl carrier protein
ACN acetonitrile
act actinorhodin
ARO aromatase
ARO/CYC aromatase/cyclase
AT acyltransferase
CBP calmodulin-binding peptide
CHS chalcone synthase
DBD DNA-binding domain
DEBS 6-deoxyerythronolide B synthase
DH dehydratase
DNR daunorubicin
DO drop-out
dps DNR/DXR biosynthetic cluster
DXR doxorubicin
ER enoylreductase
FAS fatty acid synthase
fren frenolicin
FRET fluorescence resonance energy transfer
GFP green-fluorescent protein
gra granaticin
Ha Hemagglutinin
KR ketoreductase
KS ketosynthase
KSα α-subunit ketosynthase
KSβ β-subunit ketosynthase
KSIII type III ketosynthase
MAT malonyl-CoA:ACP acyltransferase
MCS multiple cloning site
MSAS 6-MSA producing PKS
PKS polyketide synthase
PICS pikromycin producing PKS
PPTase 4'-phosphopantetheine transferase
ProtA protein A of *Staphylococcus aureus*
RBS ribosome binding site
RAPS rapamycin producing PKS
RT-PCR reverse transcriptase PCR
SD synthetic defined
TAD transcriptional activator domain
TAP tandem affinity purification
TCA trichloracetic acid
tcm tetracenomycin
TE thioesterase
TEV tobacco etch virus
UAS upstream activation sequence
CHAPTER 1

Introduction

1.1 Secondary metabolism and natural products

Secondary metabolites have always been a rich source of structurally diverse molecules that have proven beneficial to human health. Historically the therapeutic use of these natural products can be traced as far back as to ancient Egypt and Romans times when extract of willow was used as antipyretic and analgesic (Vainio and Morgan, 1997). In 1899, the introduction of acetylsalicylic acid (Aspirin) marked the birth of the first semi-synthetic drug obtained by chemical modification of a natural product (Dreser, 1899). Many other compounds derived from natural sources have been discovered and now have widespread use. Cancer and infectious diseases are mostly treated using natural product-derived medicines (Cragg et al., 1997) and also agriculture has taken advantage of the use of secondary metabolites, such as bialaphos, to protect the growth of agronomically high-value crop plants (Saxena and Pandey, 2001). As a group of compounds, secondary metabolites have a tremendous economic importance with some compounds, such as erythromycins, having a market of over 1 billion dollars per year (Demain, 2000). Typically, these compounds are synthesised through additional metabolic pathways and are not essential for normal growth, survival, development and reproduction of the producing organism (Williams et al., 1989).

1.1.1 Streptomyces and antibiotic production

More than 5,000 known antibiotic compounds are produced by bacteria or fungi and of these over two-thirds are produced by members of the actinomycete genera (Demain, 2000). Actinomycetes have also been found to be prolific producers of other useful secondary metabolites and common products including immunomodulators, enzyme inhibitors, insecticides and herbicides (Omura, 1992). Actinomycetes are soil dwelling bacteria initially regarded as fungi, as they share many morphological similarities, however, electron microscopy revealed that members of this genus lack a nuclear membrane (Hopwood, 1967) and were reclassified as prokaryotes. Actinomycetes are filamentous Gram-positive bacteria and have a high guanine and cytosine (G+C) nucleic acid content (>60 mol%) compared to other bacteria. Their life-cycle is unusual for bacteria as it involves complex multi-cellular growth which, under conditions of nutrient limitation, leads to the development of aerial hyphae bearing reproductive
spores (Kieser et al., 2000). Associated with the transition period between vegetative growth and morphological differentiation, there is the production of secondary metabolites such as polyketides and other antibiotic compounds.

Members of the genus *Streptomyces* are responsible for the production of 70-80% of antibiotics produced by actinomycetes, with smaller contributions from genera such as *Saccharopolyspora*, *Amycolatopsis*, *Micromonospora* and *Actinoplanes* (Kieser et al., 2000). *Streptomyces coelicolor* A3 (2) is the best genetically known actinomycete used as model for the biology of the entire genus. The whole genome of *Streptomyces coelicolor* has recently been sequenced showing the presence of a 8.7 megabases (Mb) linear chromosome that can be divided in a central core of 4.9 Mb and a right and left arms of 1.5 Mb and 2.3 Mb respectively (Bentley et al., 2002). Virtually all genes coding for essential functions such as DNA replication, transcription and translation are concentrated in the core whereas gene clusters coding for non essential function, such as secondary metabolites production, lie in the arms. The terminals regions of the chromosome are unstable, with extensive deletions and highly conserved inverted repeats. These regions, prone to insertions and transpositions of non-essential genes may allow the organism to sequester beneficial genetic material from its environment (Bentley et al., 2001). Interestingly *S. coelicolor* A3 (2) and the industrial relevant microorganism *Streptomyces avermilitis* dedicate respectively 5% and 6% of their genome, corresponding to 23 and 30 gene clusters, to the production of known or predicted secondary metabolites (Bentley et al., 2002; Ikeda et al., 2003). Recently, completion of the genome sequence of the erythromycin producer microorganism *Saccharopolyspora erythraea* has identified at least 25 gene clusters coding for secondary metabolites (Oliynyk et al., 2007).

Polyketides are secondary metabolites exhibiting a vast range of structural and functional diversity, and possess a wealth of medicinally important properties including antibiotic, anticancer, antiparasitic, antifungal and immunosuppressive activities. The diverse polyketide structures can be sub-divided into broad chemical groups typically aromatic, macrolide, polyether and polyene polyketides (Fig. 1.1). The term ‘polyketide’ derives historically from an intermediate in the formation of aromatic plant metabolites from the multiple condensations of small chain carbon precursors, which were called ketene groups (CH2=C=O) (Collie, 1907). Later, isotope tracer experiments also showed that saturated fatty acid biosynthesis proceeds by similar repeated
Figure 1.1 Examples of structure of different polyketides
condensations and decarboxylations of small chain carbon units such as malonyl-CoA (Gibson et al., 1958; Wakil, 1961). The aromatic polyketide 6-methylsalicylic acid (6MSA) produced by Penicillium patulum was similarly shown to be derived from C-2 units (Birch et al., 1955) revealing a close analogy between the biosynthetic pathways to polyketides and to saturated fatty acids (Birch, 1967). Since the study of fatty acid synthases (FASs) has started much earlier, our knowledge about polyketide biosynthesis largely relies on these previous studies.

1.2 Fatty acid biosynthesis

Fatty acids are essential components of cells involved in the structural formation of all biological membranes, as for example phospholipids, and also represent an important form of energy storage as triacylglycerols (Smith, 1994). Fatty acid biosynthesis is catalysed by fatty acid synthases (FASs) through sequential condensations between two carbon atom units derived from malonyl-CoA until a chain of the required length is reached. The overall reaction proceeds via Claisen condensations coupling the formation of C-C bond with the energetically favorable decarboxylation of malonyl-CoA. The chemical steps of chain extension are shown in Figure 1.2.

For the biosynthesis of a typical saturated fatty acid such as palmitic acid (16 carbon atom), a starter acetyl unit is loaded onto the cysteine thiol of the ketosynthase (KS) by the initiation enzyme malonyl-CoA:ACP acyltransferase (MAT). Similarly, a malonyl-CoA unit is transferred by the MAT to the terminal thiol of the 4'-phosphopantetheneine arm of the acyl carrier protein (ACP). The 4'-phosphopantetheneine prosthetic group is a post-translational modification and functions as a long flexible arm that carries the growing chain and delivers it to the various enzymes responsible for chain processing. The KS catalyses the decarboxylation of the malonate and then couples the condensation with the bond acetyl group resulting in the formation of a 3-ketoacyl-ACP thioester. The β-keto group undergoes reduction reactions via three sequential steps catalysed by a ketoreductase (KR), a dehydratase (DH) and an enoylreductase (ER) to yield a fully saturated chain longer than the original by two methylene units (Fig. 1.2). This sequence of reactions completes the first round of chain extension. The cycle is then repeated after transfer of the saturated chain from the ACP, which subsequently is loaded with a new malonate unit, to the KS. Successive cycles lead eventually to a chain of the required length (usually 14, 16 or 18 carbons) with the release from ACP as a free acid by the action of a thioesterase (TE) (Wakil, 1989).
Figure 1.2 Fatty acid biosynthesis
R represents a methyl group for the first cycle and the extending polyketide chain in subsequent cycles. ACP - acyl carrier protein; MAT - malonyl-CoA:ACP acyltransferase; KS - ketosynthase; KR - ketoreductase; DH - dehydratase; ER - enoylreductase; TE - thioesterase.
All FASs have the same set of catalytic centres (MAT, KS, ACP, KR, DH, ER and TE), but the structural organization of these activities depends on the host organism. In most bacteria and plants they reside in individual fully dissociable polypeptides that may form a multiprotein complex. This organization is called type II FAS (Magnuson et al., 1993). In yeast and mammals the catalytic activities are grouped to form extremely large multifunctional polypeptides (Smith, 1994), in two nonidentical polypeptides in yeast with αβ organization (Schweizer, 1980) and two identical polypeptides in animals with an organization (Wakil et al., 1983). This structural organization is designated as type I. Fatty acids display considerable structural diversity including altered chain length, due to tissue specific expression of alternative thioesterases, odd numbered chain lengths, by incorporation of propionyl-CoA as starter unit and branched-chains by utilization of methylmalonyl-CoA as extender unit (Fernandes et al., 1997). Increased diversity amongst fatty acids is due to different levels of insaturation; unsaturated fatty acid can be generated de novo (Wallis et al., 2002) or by desaturation of already synthesised saturated precursors (Behrouzian et al., 2003).

1.3 Polyketide biosynthesis

Polyketide synthases (PKSs) utilize similar chemistry and involve analogous enzymatic activities to FASs which also share much amino acid sequence homology with PKS, especially around the active sites (Hutchinson and Fujii, 1995). However, important features distinguish the mechanism of fatty acid and polyketide biosynthesis. Firstly, while fatty acid biosynthesis is rarely primed with precursors other than acetyl-CoA, PKSs utilize a wider repertoire of both starter and extender CoA esters. For example, acetyl-, propionyl-, isobutyryl- and n-butyryl-CoA esters can all be used in vitro as starter units for erythromycin biosynthesis (Wiesmann et al., 1995). Secondly, the β-keto group processing after the condensation reactions is usually incomplete in polyketide biosynthesis giving rise to keto-, hydroxyl, enoyl or alkyl functional groups. These factors allow cyclisation of the final product and introduction of a wide diversity of chemical side chains, usually sugars, and various chiral centres.

Polyketide synthases can be divided into distinct functional types in a fashion similar to the FASs. Type I PKSs or modular PKSs, are large multifunctional enzymes that have discrete functional units with individual active sites, called domains. The domains responsible for the condensation and modification of an acyl unit into the polyketide chain are grouped together to form a module. Each module, therefore, consists of a
ketosynthase domain, an acyltransferase (AT) domain and an acyl carrier protein, plus can contain one, two or all of the reductive domains namely, ketoreductase, dehydratase and enoyl reductase. The organisation of modules, with their domain activities, therefore represents the combinatorial programming required to synthesise polyketides.

Type II PKSs or aromatic PKSs, consist of several small monofunctional proteins loosely associate in a complex. These PKSs contain a set of proteins called 'minimal' PKS that selects a starter unit (most often acetate) and catalyses the condensation reactions with malonate extender units. The 'minimal' PKS is composed of two KS subunits (KSα and KSβ), a malonyl-CoA:ACP acyltransferase that loads the acyl carrier protein with the acyl units. Type II PKSs also contain proteins with different enzymatic activities such as ketoreductases (KRs), aromatases (AROs) and cyclases (CYCs) to generate the natural product (McDaniel et al., 1995). In addition to type I and type II, other class of PKSs exists both in plants and bacteria that interestingly does not involve ACP and utilize acyl-CoA esters as substrate directly. Type III PKSs, also called chalcone synthases, are small homodimeric proteins that choose the starter unit, carry out the condensing reactions and release the final product after cyclisation (Funa et al., 1999).

The product of the PKSs is very often only an intermediate on the way to becoming a bioactive compound. Once it is released from the PKS, the polyketide is modified by other 'post-PKS' enzymes which may add functional groups (methyl, hydroxyl, and carbonyl groups) and typically glycosilate (Gaisser and Leadlay, 1998; Borisova et al., 1999; Tang and McDaniel, 2001; Blanco et al., 2001). In the case of a group of polyketides called polyethers, such activities carry out a series of oxidations and ring-forming reactions, which totally alter the molecular shape (Oliynyk et al., 2003; Hughes-Thomas et al., 2003).

1.3.1 Type I polyketide synthases
Type I PKSs are massive multidomain proteins responsible for the biosynthesis of complex compounds that display the greatest structural and functional diversity of all the PKSs classes. Pioneering studies conducted in 1976 provided purification of the fungal PKS responsible for the biosynthesis of the polyketide 6-methylsalicylic acid, naturally produced by P. patulum (Vogel and Lynen, 1975; Dimroth et al., 1976). The 6-MSA polyketide synthase (MSAS) was the first fungal PKS gene to be characterized.

and showed that polyketide and fatty acid biosynthesis have different metabolic pathways in the same host producer (Vogel and Lynen, 1975). Also, structural analysis of the MSAS provided a paradigm for the structural organization of the type I PKS.

### 1.3.1.1 Erythromycin polyketide synthase

The complex multienzyme responsible for the production in *Saccharopolyspora erythraea* of the macrolide polyketide antibiotic erythromycin A, was the first modular PKS gene cluster to be sequenced. The genetic sequence showed what was known as a type I arrangement and revealed that the genes encoding for the PKS were clustered together in consecutive open reading frames. The assumption that antibiotic biosynthetic genes are clustered together with the gene that confers self-resistance *ermE* (Bibb *et al.*, 1985) and the observation that mutations in a region of the genome near to *ermE* disabled the production of erythromycin A (Weber *et al.*, 1989), led to the identification of three open reading frames *eryAI*, *eryAII* and *eryAIII* by chromosome walking around the *ermE* genetic locus (Cortes *et al.*, 1990). These three ORFs encode for three giant multienzyme polypeptides each approximately 350 kDa, DEBS 1, 2 and 3; (Cortes *et al.*, 1990; Donadio *et al.*, 1991). These enormous proteins constitute the 6-deoxyerythronolide B synthase (DEBS) and are involved in the production of the first free intermediate 6-deoxyerythronolide B (6-dEB) that forms the aglycone core of erythromycin A (Martin *et al.*, 1967).

The biosynthesis of 6-dEB is carried out by decarboxylative condensation reactions between the starter unit, propionyl-CoA and six methylmalonyl-CoA extender units. The 6-dEB lactone product is further modified by hydroxylation of the aglycone core at C-6, glycosylated with attachment of the sugars L-mycarose and D-deosamine (which is then O-methylated to give cladinose) at C-3 and C-5 respectively, and finally hydroxylated at C-12 to yield active erythromycin A (Staunton, 1998) (Fig. 1.3). Sequence comparisons with type I FAS revealed a linear arrangement of separate catalytic sites for each biosynthetic step, and that these were grouped into six modules consisting of the active sites (KS, AT, ACP) and where appropriate a variable set of domains (KR, DH, ER and TE) required for one cycle of chain extension and chain release. In addition, DEBS1 is fronted by an AT-ACP didomain loading module that is responsible for the choice of propionate as starter unit derived from propionyl-CoA (Marsden *et al.*, 1994). The TE domain is located at C-terminal of DEBS3 and catalyses
the cyclisation of the heptaketide chain to macrolactone releasing the polyketide from the enzyme (Aggarwal et al., 1995) (Fig. 1.3).

The biosynthesis carried out by DEBS is performed in an assembly line fashion; once a module has acted on a nascent acyl chain, the chain is passed to the next downstream module. Evidence of this processive nature has been elucidated by deletion of large part of the gene encoding for the KR domain in module 5 that resulted in a strain producing two erythromycin analogues with an unreduced β-keto group in place of hydroxyl (Donadio et al., 1991). Also, amino acid substitutions that disabled the NADPH-binding site of the ER in module 4, led to the accumulation of an erythromycin analogue presenting a double bond in the expected position (Donadio et al., 1993). The one-to-one correspondence between domains and biosynthetic transformations explains, with a concept called 'colinearity', how genetic programming is achieved (Bevitt et al., 1992).

1.3.1.2 Other Type I PKSs
Since the first report of the DEBS sequence, more than 30 other type I PKSs have been sequenced and published including those responsible for the biosynthesis of the immunosuppressor rapamycin (Schwecke et al., 1995), the antibiotics pikromycin (Xue et al., 1998), tylosin (Gandecha et al., 1997) and monensin (Leadlay et al., 2001). Comparison of the sequences of all the different PKSs has revealed that modular PKS can contain different numbers of extension modules in a single PKS polypeptide and that sometimes domains are present in some modules, although analysing the structure of the final product these would be expected to be unnecessary. Some of these domains are evidently inactive by sequence analysis while others appear to be potentially active but not required for the processing of the polyketide chain. The PKS responsible for the biosynthesis of the immunosuppressor rapamycin (RAPS) (Schwecke et al., 1995) presents this discrepancy between the sequence of domains and the required set of transformation necessary for the production of the polyketide product. RAPS is composed of three extraordinary large multienzymes RAPS 1 (4 modules)(900 kDa), 2 (6 modules)(1.07 MDa) and 3 (4 modules)(660 kDa). Analysis of the primary structure of the polypeptides shows that modules 3 and 6 present potentially active KR and DH activities not required for the final structure of the polyketide; module 3 also contains a potentially active ER domain. The characterization of other type I PKSs and the elucidation of the mechanisms of biosynthesis have highlighted examples of
Figure 1.3 Erythromycin A biosynthesis
The assembly line fashion of the erythromycin producing PKS (DEBS). DEBS1, 2 and 3 and their reaction products are shown in red, blue and green respectively.
diversion from the general rule of colinearity for modular PKS and that some type I PKSs that do not completely fit in the general classification system (Moss et al., 2004).

A phenomenon termed 'skipping' has been observed in some modular PKSs where a round of extension is omitted resulting in a final product smaller than expected. Attempts to generate new polyketides by inserting either module 2 or 5 from RAPS in between the module 1 and 2 of DEBS1, led to a mutated strain of *S. erythraea* that produced the predicted new tetraketide resulting from the incorporation of malonate as extender unit which is dictated by the RAPS AT (Rowe et al., 2001). However, the major products of the strain were the two triketide lactones normally produced by *S. erythraea* eryA* mutants harbouring only DEBS1 (Rowe et al., 1998). Direct ACP-ACP transfer of the acyl chain without any loading on the KS domain and subsequently with no round of extension, has been proposed as molecular mechanism behind the module skipping (Thomas et al., 2002). Some modular PKS appear to use a 'programmed skipping' mechanism for the biosynthesis of their normal products. The pikromycin (PICS) PKS in *Streptomyces venezuelae* ATCC 15439 contains six extension modules and is capable of production of five different antibiotics; three are hexaketides and the other two are heptaketides (Xue et al., 1998). The biosynthesis of these compounds by one PKS has been demonstrated to occur by a mechanism of skipping of the last extension module resulting in the accumulation of also hexaketides in addition to the expected heptaketides (Beck et al., 2002).

In addition to the 'skipping', another aberrant process termed 'stuttering' has been observed occurring in some modular PKSs. Some synthases can use a module iteratively, catalysing an additional round of chain extension to give a product larger than predicted from the sequence of the PKS (Wilkinson et al., 2000). A mutated strain of *S. erythraea* was found to produce minor metabolites closely related to 6-dEB. Structural analysis of these new compounds revealed that they were synthesised by incorporation of either acetyl- or propionyl-CoA; ability to use different starter had been already described for DEBS (Wiesmann et al., 1995). Surprisingly these metabolites incorporated an extender unit more than expected for erythromycin analogues. In particular, the processing of the β-keto group of the additional unit appeared to be complete. The only module in DEBS that contains all the enzymatic activities for a full reduction of the keto group is module 4 which is located at the C-terminal of DEBS2. The octaketides should have been synthesised by DEBS using module 4 twice in an
iterative fashion generating one extender unit longer polyketide. The molecular details
of this aberrant process are unclear as normally native modular PKSs exert a rigorous
control over chain length. It could be speculated that, as module 4 is present at the
C-terminal of DEBS2 and the successive step is the transfer of the nascent acyl chain
onto the KS of module 5 on the adjacent polypeptide DEBS3, there is a competition
between inter-protein transfer of the acyl chain and iterative use of module 4 with this
latter process to be faster (Moss et al., 2004). The epothilone producing modular PKS in
Streptomyces cellulosum has been described to undergo both process of ‘skipping’ and
‘stuttering’ as revealed by analysis of the fermentation metabolites (Hardt et al., 2001).
The repeated use of certain modules appears to be finely programmed in the
biosynthesis of a number of polyketides including borrelidin (Olano et al., 2003),
aureothin (He et al., 2003), stigmatellin (Gaitatzis et al., 2002) and lankacidin
(Mochizuki et al., 2003).

The comparison of the PKS responsible for the production in P. notatum of 6MSA and
the animal FAS provided the paradigm for the classification of type I PKS for modular
synthases. 6MSAS contains in order KS, AT, DH, KR and ACP domain to catalyse the
biosynthesis of 6-methylsalicylic acid from one molecule of acetate and three of
malonate. This PKS uses iteratively these activities to catalyse three round of chain
extension with different level of reduction of the β-keto group at each stage
(Shoolingin-Jordan et al., 1999). Another fungal PKS that is responsible for the
biosynthesis of the cholesterol-lowering agent lovastatin in Aspergillus terreus
repeatedly uses different set of activities during each cycle of extension. The initial
observation that these modular PKSs that use in iterative fashion some domains were
from fungi, generated their classification as fungal type I PKSs. Successively the
elucidation of the mechanism for the biosynthesis of the polyketide avilamycin in
Streptomyces viridochromogenes Tu57 and calicheamicin in Micromonospora
echinospora ssp. calichensis (Ahlert et al., 2002; Liu et al., 2002) has revealed that
iterative type I PKS are widespread also amongst bacteria. These finding determined the
modern distinction type I PKSs that can be classified in iterative and modular PKSs
(Shen, 2003).

1.3.2 Type II Polyketide synthases
Type II polyketide synthases are a large family of polyketide synthase mechanistically
related to the type II FAS typically found in bacteria and plants, that catalyse the
biosynthesis of aromatic natural product such as actinorhodin, frenolicin, tetracenomycin and daunorubicin (Hopwood, 1997). The immediate product of a type II PKS is a poly-β-keto chain which then undergoes specific downstream modifications such as cyclisation and aromatisation to stabilise the highly reactive product. Typically, a reduction occurs to a specific carbonyl group, C9 for actinorhodin, which strongly influences the subsequent pattern of folding of the final product. In 1984, Malpartida and Hopwood reported the isolation of a 30 kbp chromosomal DNA fragment containing the gene cluster responsible for the production of aromatic polyketide actinorhodin \((act)\) in \textit{Streptomyces coelicolor} (Malpartida and Hopwood, 1984). The first clue of the architecture of the actinorhodin producing type II PKS and gene function came from complementation analysis of different mutants of \textit{S. coelicolor} blocked in different stages of the biosynthesis of the polyketide and from the sequence of part of the gene clusters for granaticin \((gra)\) and tetracenomycin \((tcm)\) (Sherman \textit{et al}., 1989; Bibb \textit{et al}., 1989). The complementation of the lacking activities of the \textit{act} PKS with the homologues from the \textit{gra} and \textit{tcm} PKSs suggested that these PKSs had a type II structure composed of monofunctional proteins for individual reactions of chain assembly and modification similar to those found in \textit{E. coli} FAS (type II).

1.3.2.1 ‘Minimal’ PKS

Expression of the \textit{act} KS genes in the granaticin producer \textit{Streptomyces galilaeus} revealed the effective cooperation of \textit{act} proteins with heterologous subunits demonstrated by the accumulation of novel polyketides and the ability to complement mutations in homologues \textit{gra} subunits (Bartel \textit{et al}., 1990). These experiments marked the first example of functional active hybrid PKS system. Similar experiments were performed on a systematic basis by Khosla, Hopwood and colleagues using the strain of \textit{S. coelicolor} CH999 which is deleted in the bulk of the \textit{act} locus (McDaniel \textit{et al}., 1993) transformed with a plasmid carrying individual \textit{act} PKS gene that could be deleted, mutated or replaced by a heterologous gene (Hopwood, 1997). These experiments were useful to defined the so called ‘minimal’ PKS which they formulated to consist of the KS, and ACP and another KS-like component, chain length factor (CLF), that may have a specific role in determining the length of nascent chain (McDaniel \textit{et al}., 1994).
The CLF subunit is essential for polyketide production and shows remarkable sequence homology to the KS, except for a highly conserved glutamine (or occasionally glutamate) residue in place of the active site cysteine (McDaniel et al., 1993; Sherman et al., 1992). The CLF was originally believed to control the overall length of the final product (McDaniel et al., 1994; Hopwood, 1997); however, later experiments showed that downstream enzymes also have a role in determining chain length which is not only dictated by the ‘minimal’ PKS (Shen et al., 1999; Kramer et al., 1997). On the base of these results, some authors have renamed the condensing enzyme carrying the active cysteine, KSα, and the CLF KSβ (Seow et al., 1997; Meurer et al., 1997). ‘Minimal’ PKSs were created mutating either the active cysteine of the KSα or the glutamine, present KSβ in place of the active cystein, in alanine. Decarboxylation of malonyl-CoA was observed only when the substitution involved the KSα but not the KSβ. These results suggested that the glutamine residue in KSβ have a role in the decarboxylation of the malonate extender unit during the biosynthesis suggesting the property of the protein to act as chain initiation factor that generates acetyl-CoA in situ on the ‘minimal’ PKS (Bisang et al., 1999). Similar results were observed in the related type II FAS whereas the condensing enzyme could be transformed in a decarboxylase by replacing the active cysteine with glutamine (Witkowski et al., 1999). Interestingly, some modular PKSs present a similar component in the loading module designated KSQ, with a glutamine residue in place of the active cysteine, which has decarboxylase activity towards CoA-esters of dicarboxylic acids (Bisang et al., 1999).

Recently Stroud and co-workers have solved the crystal structure of the act KSα/KSβ complex at 2.0 Å resolution (Keatinge-Clay et al., 2004) which formation during aromatic polyketide biosynthesis had already been suggested earlier (Carreras and Khosla, 1998). The crystal structure has revealed an internal cavity at the interface of the two subunits where the polyketide chain is assembled. During the biosynthesis, the growing poly β-ketoacyl chain is highly reactive and must be protected from aberrant intramolecular cyclisation reactions (Harris and Harris, 1986). The amphipathic environment of the tunnel could keep the polyketide backbone extended stabilizing the keto groups which are kept separated. A model for aromatic polyketide biosynthesis involves interaction between the KSα/KSβ heterodimer and the ACP in order to carry out the condensation reactions necessary for the assembly of the polyketide chain (Fig. 1.4).
Figure 1.4 Model for the priming (A), elongation (B) and termination (C) for aromatic PKSs. The model involves dissociation of the ACP at each round of condensation (Dreier and Khosla, 2000).
The interaction with the ACP was suggested by kinetic studies that demonstrated that the apo-form of the protein is an inhibitor of the polyketide biosynthesis (Dreier et al., 1999).

The acyl carrier protein is an essential protein shuttling all the pathway intermediates between the different enzymes. The ACPs are a group of highly related, small acidic protein with molecular weight of ~9 kDa (White et al., 2005). The structure of the act apo-ACP has revealed that the protein is rod-shaped composed exclusively of helical elements; three major helices (1, 2 and 4) a shorter helix (3) and a large loop region connecting helix 1 and 2 (Crump et al., 1997). Conserved acidic residues at N-terminal of α2 helix are postulated to represent a ‘recognition helix’ essential for the interaction of the ACP with its partner proteins (Zhang et al., 2003a). Also, the ACP interacts with the bound acyl intermediate by using the hydrophobic cavity that lies in a cleft flanked by α2 and α3 helices (Roujeinikova et al., 2002). The prosthetic group, 4’-phosphopantetheine, which binds the acyl groups, is covalently attached to the ACP via phosphodiester linkage to the Ser42 in the α2 helix.

The conversion from the inactive apo-form to the active holo-form of the ACP is catalysed by 4’-phosphopantetheine transferase (PPTases) that transfers the 4’phosphopantetheine moiety from the coenzyme A onto Ser42 (for PKS ACPs, Ser36 for FAS ACPs) (Lambalot et al., 1996). Thus, the activity of the ACPs is dependent on its conversion to the active holo-form catalyzed by specific PPtase. Heterologous expression in Escherichia coli of the ACPs from different type II PKSs resulted in only partial or in some cases no modifications at all with the essential 4’-phosphopantetheine cofactor (Crosby et al., 1995). In addition, PPTases have been demonstrated to possess an intrinsic specificity for their own substrates (Lambalot et al., 1996). It has been reported that co-expression in Saccharomyces cerevisiae of the surfactin P-pant transferase (Sfp) from Bacillus subtilis along with the fungal type I MSAs resulted in an increased production of the polyketide (Kealey et al., 1998). act apo-ACP shares a common overall structural organization with E. coli FAS though some structural differences exist. In particular, a hydrophobic core and a number of buried hydrophilic groups, principally Arg72 and Asn79, both of which are 100% conserved amongst PKS ACPs but not in FAS ACPs, may play a role in stabilizing the growing polyketide chain (Crump et al., 1997).
There is some debate about the role of MAT in the aromatic polyketide biosynthesis. Earlier experiments conducted by Hopwood and coworkers (McDaniel et al., 1993) indicated as components of ‘minimal’ type II PKS, the two KS subunits and the ACP. The act PKS gene cluster lacks of an ORF coding for a protein with acyltransferase activity and therefore a MAT was not included as an essential component of the ‘minimal’ PKS. Later studies have indeed demonstrated the existence of a gene (fabD) encoding for a malonyl-CoA:ACP acyltransferase, that lies 2.8 Mb away from the rest of the act cluster, involved in fatty acid biosynthesis (type II) (Revill et al., 1995). The product of fabD gene is able to charge both act and type II FAS ACPs with malonate suggesting that this activity may be shared between FAS and PKS (Summers et al., 1995; Florova et al., 2002). Different studies have provided evidence that the presence of MAT is essential for polyketide synthesis to proceed in reconstituted act and tcm PKSs (Bao et al., 1998; Dreier et al., 1999; Carreras and Khosla, 1998) although, Arthur et al. indicated that self-malonylation is an inherent activity in vitro of chemically synthesised act ACP (Arthur et al., 2005). More recently, the act ACP has been shown to catalyse transfer of malonate to FAS and other PKSs ACPs in vitro (Arthur et al., 2006). However, a study has reported that the $K_M$ values of act ACP malonylation in the presence and in absence of MAT was 79 $\mu$M and 216 $\mu$M respectively (Dreier et al., 1999). These results suggest that at physiologically relevant concentration of individual proteins, the MAT-dependent pathway for malonyl-ACP is dominant (Dreier et al., 1999; Dreier and Khosla, 2000).

1.3.2.2 Ketoreductases

In S. coelicolor, following chain assembly the actinorhodin polyketone backbone is further processed by tailoring enzymes such as ketoreductase, cyclase and aromatase that define the folding pattern of the nascent poly-$\beta$-keto intermediate (Hopwood, 1997). Recently, the crystal structure of the act KR has been determined at 2.5 Å resolution revealing a tetrameric arrangement as dimer of dimers. The binding of the cofactor NADPH stabilizes the complex and exposes the positively charged residues recognised by the ACP for interaction delivering the substrate to the catalytic site (Hadfield et al., 2004). Ketoreductases catalyse the addition onto a keto group of a hydrogen from NAD(P)H, resulting in the formation of a secondary alcohol and in most cases, the reduction occurs at the C9 on the growing poly-$\beta$-keto carbon chain independently from the chain length and nature of starter unit. Since this kind of modification leads the carbon chain to be more prone to aldol condensation for the
closure of the first aromatic ring, ketoreductases have an important role in dictating the structure of the final product (Rawlings, 1999). In addition, type II PKSs lacking of a functional KR may result in the production of incorrectly folded polyketides (Fu et al., 1994). However, some PKSs, such as tcm, do not contain any ketoreductase and have a different folding pattern with a C9-C14 ring closure instead of the C7-C12 amongst the more common reduced systems (Rawlings, 1999).

Most ‘minimal’ PKSs are able to synthesise polyketides in absence of a KR and therefore, it was hypothesised that the reduction of the keto group occurs when the assembly of the nascent chain is complete (Bartel et al., 1990). The observation that the enterocin producing PKS is unable to produce polyketides in absence of the endogenous KR, EncD, suggests that for this PKS the reduction must take place during chain elongation rather than after as proposed for the majority of aromatic PKSs (Hertweck et al., 2004). The difference between the two proposed models for timing of ketoreduction is very important to determine the length of highly reactive, growing polyketide chain to be protected by spontaneous intramolecular reactions. The ketoreduction triggers the formation of the first ring that removes the need to control the regiochemistry of the subsequent cyclisations (Leeper and Staunton, 1984). If the ketoreduction takes place during the chain extension, as proposed by Hertweck et al., only fewer keto groups need protection suggesting that the control of the chain length is dictated by the formation of the first ring and successive chain extension steps for the production of the final product (Staunton and Weissman, 2001).

1.3.2.3 Cyclases

The high reactivity of the keto groups must be suppressed during polyketide assembly in order to produce a polyketide with the appropriate folding pattern (Harris and Harris, 1986). Ectopic expression of the spore pigment whiE ‘minimal’ PKS in S. coelicolor CH999 resulted in to the accumulation of more than 30 polyketides with different chain lengths and cyclisation regiochemistries (Shen et al., 1999). Also, co-expression of the cyclase whiE-ORFVI with the ‘minimal’ whiE PKS led to the production of a polyketide of right chain length and C9/C14 cyclisation reaction (Yu et al., 1998). The expression of frenolicin (fren) ‘minimal’ PKS in S. coelicolor CH999 together with the cyclase TcmN, from the tcm cluster, resulted in the accumulation of the expected nonaketide as predicted by the ‘design rules’ but no octaketides were detected. In contrast, S. coelicolor CH999 expressing only the ‘minimal’ fren PKS
produced only octaketides. These results suggested that in absence of a CYC, the fren ‘minimal’ PKS produces polyketides of only one chain length either octaketides or nonaketides with TcmN cyclase (Kramer et al., 1997).

These results suggested that the ‘minimal’ PKS does not dictate alone either the chain length or the first cyclisation event but must rely on auxiliary components, such as cyclases, to stabilize the highly reactive growing chain from spontaneous intramolecular reactions. Typically the first cyclisation, via aldol condensation, occurs between C7 and C12 and is then followed by an enzymatically triggered dehydratation to give rise the first aromatic ring. There is some debate as to whether or not the ‘minimal’ PKS is responsible for the closure of the first ring in the aromatic polyketide biosynthesis. Structural analysis of the amphipathic tunnel at the interface of the act KSα/KSβ heterodimer has revealed that the length of this cavity may influence the first ring closure. During the assembly of the actinorhodin polyketide backbone, at the stage of octaketide, the nascent chain would fill the tunnel only buckling and possibly initiating the first cyclisation (Keatinge-Clay et al., 2004).

The downstream enzymes, such as ketoreductases and cyclases do not have access to the polyketide chain while it is buried in the amphipathic tunnel, but the addition of these enzymes to the ‘minimal’ PKS could influence cyclisation and even the chain length patterns (Kramer et al., 1997; McDaniel et al., 1993). The tcm ‘minimal’ PKS chiefly produces the aberrant C7-C12 cyclised decaketides instead of the right C9-C14. However, the addition of the cyclase TcmN results in exclusive accumulation of the latter cyclized product suggesting a crucial role played by the cyclase (McDaniel et al., 1994). Although a large number of CYC genes have been identified in aromatic PKSs, to date only for few cyclases structural data are available: TcmF2 cyclase from tcm cluster (Thompson et al., 2004) and SnoaL and AknH in the biosynthesis of the aromatic polyketides nogalamycin in Streptomyces nogalater and aclacinomycin in Streptomyces galilaeus, respectively (Sultana et al., 2004; Kallio et al., 2006). Structural analysis of TcmF2 cyclase has revealed that the enzyme acts as homodimer influencing the chemical outcome of the cyclisation rather than providing a specific catalytic group (Thompson et al., 2004).
1.3.2.4 Aromatases

The cyclisation process in the biosynthesis of aromatic polyketides is often supported by aromatases that dehydrate cyclic alcohols leading to aromatic ring formation. *actVII* aromatase is involved in the two dehydration reactions after the ketoreduction and subsequent intramolecular C7-C12 aldol reaction that determines the closure of the first ring in the actinorhodin polyketide backbone. Remains some confusion over the classification aromatases/cyclases (ARO/CYCs) as these activities can be either combined on the same didomain enzyme, or be separated as in monodomain enzyme (Rawlings, 1997). Sequence analysis reveals that didomain proteins appear to have two homologues halves or domains at the N- and C-terminal thought to catalyse the enzymatic reactions and the N-terminal domain exhibits sequence similarity with the monodomain ARO/CYCs. Didomain ARO/CYCs, such as ActVII, catalyse the formation of the first six member carbocyclic ring derived from C9 reduced polyketides and have aromatase activity. Instead, monodomain ARO/CYCs, such as TcmN, catalyse exclusively C9-C14 aldol condensation instead of the C7-C12 observed in polyketides that have undergone to reduction at C9 and are speculated to have an aromatase activity (McDaniel *et al.*, 1995; Shen *et al.*, 1995). Didomain ARO/CYCs possess a hierarchy with respect to the substrates with different chain lengths. For example, an ARO/CYC from a decaketide synthase, such as griseusin producing PKS, catalyses aromatisation of octaketides, nonaketides and decaketides whereas an octaketide ARO/CYC, such as *actVII*, only recognises octaketides (McDaniel *et al.*, 1995)

Interestingly in aromatic PKS clusters it has not being identified any gene encoding for a protein responsible for the release of the polyketide chain from the ACP such as thioesterase II present in modular PKSs. Chain release is therefore, presumably a function of one or more component of the multiprotein complex that results in the hydrolysis of the ACP thioester. Examples of the genetic organization of typical aromatic PKS clusters are shown in Figure 1.5.

36
Figure 1.5 Genetic organization of typical Type II PKS clusters
1.3.3 Type III Polyketide synthases

Type III PKSs are a peculiar class of synthases found widespread in bacteria, fungi and plants. This class is defined by their amino acid sequence homology with the chalcone synthase (CHSs) which is the most well known representative in higher plants (Austin and Noel, 2003). Type III PKS have a unique biosynthetic mechanism different from the other class of synthases. While type I and II PKSs rely on the ACP and its prosthetic group to activate the acyl-CoA substrate and to deliver the growing acyl chain to each enzymatic activity, type III PKSs utilise CoA thioesters directly. Also, CHSs are homodimeric proteins that carry out the complete series of condensation, cyclisation and aromatisation reaction using only a single active site typically leading to the formation of a six-member ring lactone (Ferrer et al., 1999). The mechanism by which this class of synthases controls chain lengths and folding patterns is still obscure but recent structural studies on the 1,3,6,8-tetrahydroxynaphthalene synthase (THNS) have elucidate the structure of the internal cavity of the enzyme and its possible involvement in the choice of the starter unit and in controlling growing chain stability (Austin et al., 2004). Although of a great interest, this class of PKSs will not be considered further here.

1.4 Combinatorial biosynthesis

Modern medicine is constantly looking to combat the shifting spectrum of diseases. The emergence of drug-resistance and the redundancy of many existing antibiotics has spurred the search for alternatives. One strategy involves the finding of different organisms producing natural products with new or improved biological activities. Numerous Gram-positive and Gram-negative bacteria, fungi and symbionts from macroorganism such as lichens sponges and soft corals have been found to contain a wealth of metabolic diversity useful for human health. However, the ability to discovery compounds with desired properties from natural sources has become increasingly difficult. Therefore, the scientific community have turned towards a new approach that referred as combinatorial biosynthesis to produce novel natural products for drug development (Hranueli et al., 2005).

By analogy to combinatorial chemistry, where highly diverse libraries of new compounds are assembled by linking of simple chemical building blocks, combinatorial biosynthesis aims to generate similarly rich libraries of compounds by mixing and matching pieces of known natural products. The strategy relies upon manipulating the
genes that encode for enzymes responsible for synthesising these metabolites; using conventional methods of genetic engineering, enzymes are deleted, added or substituted, or their substrate specificity altered. Recently a theoretical study has shown that modular PKSs have the potential to generate hundreds of millions of compounds through reassembly of the catalytic units in novel combinations (Gonzalez-Lergier et al., 2005).

1.4.1 Engineering Type I PKSs
Type I PKS systems are amenable to combinatorial biochemistry by genetic manipulation, due to the direct relationship between genetic structure, protein structure and polyketide chain assembly. Therefore, rational genetic modifications in modular PKS lead to a final product with alterations in chain assembly and chemical group properties in predicted specific points of the polyketide structure. Type I PKSs have been extensively engineered exploiting the remarkable modularity of these multienzymes and common strategies include: (a) altering individual domains within modules, by deletions, substitutions and additions; (b) joining of intact heterologous modules and/or multimodular subunits; (c) feeding of synthetic precursors (d) modifying the enzymes involved in post-PKS processing (Weissman and Leadlay, 2005).

1.4.1.1 DEBS1-TE
The erythromycin PKS has been the focus of the majority of the genetic engineering strategies to explore the structural plasticity and substrate tolerance of modular PKSs. Due to the large size of these PKSs, a truncated model system for the analysis of genetic modifications, DEBS1-TE, has been designed to simplify the interpretation of individual domain activities. This was created by relocating the TE domain of DEBS3 to the C-terminal of DEBS1 to produce a bimodular PKS which is responsible for the production of a tryketone lactone after just two extension cycles (Fig. 1.6) (Cortes et al., 1995). As well as being a good model for engineering PKSs, DEBS1-TE has solved some functional issues of polyketide synthesis both in vivo and in vitro following rigorous purification of the hybrid protein.
Figure 1.6 DEBS1-TE
A truncated bimodular PKS obtained by relocation of TE to the C-terminal of DEBS1 and triketide lactone produced by the host *S. coelicolor* CH999.
It has been used to elucidate chain initiation (Hong et al., 2005), chain extension on modular PKS in vitro (Wiesmann et al., 1995), to demonstrate loading module tolerance for alternative substrates in vivo (Weissman et al., 1998; Regentin et al., 2004) and in vitro (Lau et al., 2000) and to investigate the molecular basis for control of stereochemistry in chain extension, focusing upon the KR domain (Holzbaur et al., 1999).

Hybrid modular PKSs obtained by altering domain architecture have been achieved mostly by the addition or deletion of the reductive subunits (KR, DH, ER), resulting in different degrees of β-keto group reduction and interesting changes in the stereochemistry of the polyketide (Weissman et al., 1997; Böhm et al., 1998; Holzbaur et al., 1999). Many derivatives of DEBS1-TE with other modifications, such as modules, domain and residue mutations, substitution, and deletions have also used to further investigate the functions and the mechanisms operating within DEBS (for a review see Rawlings, 2001).

1.4.1.2 Alteration of AT domain

Many efforts have been attempted to change the substrate specificity of the AT domain which plays a critical gatekeeper role in polyketide chain assembly (Hans et al., 2003; Kato et al., 2002). The substitution of an AT domain with an altered specificity represents a challenging task, as the domains are normally located in the core of the module; however, several experiments reported successfully replacements of AT domains with production of predicted new polyketides. AT2 of the rapamycin producing PKS, which is specific for malonyl-CoA, has been swapped with the one of module 1 from DEBS1-TE, which is specific for methylmalonyl-CoA. The resulting hybrid synthesised a predicted triketide lactone that lacked a methyl group at the C4 position of the lactone ring (Oliynyk et al., 1996). Erythromycin analogues have been created by replacing all the methylmalonyl specific AT domains in DEBS with heterologous malonyl specific ATs. Similar experiments have been carried out with the other modular PKSs (for a review see McDaniel et al., 2005).

1.4.1.3 Swapping Loading modules

The majority of modular PKSs identified so far contain an additional domain, called KS\textsubscript{Q}, located at the N-terminal of the AT0ACP0 loading domain. This domain closely resembles a KS domain whereas the active site cysteine has been replaced by glutamine
Tridomain loading modules mediates chain initiation via decarboxylation by the K\(S_Q\) of CoA thioester of dicarboxylic acids recruited by the AT0 (Bisang et al., 1999). Swapping loading domains results in modifications of the starter unit incorporated with accumulation of new compounds. Generally tridomain loading modules appear to have stricter substrate specificity than AT0ACP0 loading modules. Substitution of the DEBS loading domain with the correspondent domain present in the avermectin producing PKS, resulted in accumulation of erythromycin analogues that incorporated starter units typical of natural avermectins but also acetate and propionate (Marsden, 1998). By contrast, introducing the monensin K\(S_Q\) loading module in place of the native DEBS AT0ACP0 resulted in a complete switch of specificity to an acetate starter (Long et al., 2002).

1.4.1.4 Incorporation of alternative starter units

Production of novel polyketides can be achieved by incorporation of non-native starter units by expression of the PKS in a heterologous host where the substrate supply is different to the native host. While of DEBS1 expressed in the native host, S. erythraea, produces the expected triketide lactone from propionyl-CoA as starter unit (Wiessmann et al., 1995), heterologous expression in S. coelicolor leads to the production of mixture of lactones primed with acetate and propionate (Kao et al., 1995).

Production of a variety of new compounds has been accomplished coupling genetic manipulations of PKSs with feeding the growing culture of alternative starter units, a process referred as 'mutasynthesis'. Precursor directed biosynthesis of 6-dEB in S. coelicolor was initially performed utilising a null mutation in the first KS domain in DEBS1 (Desai et al., 2002). This mutation, KS1\(^0\), inhibits biosynthesis blocking the first round of condensation; however, the addition of synthetic thioesters such as diketide-S-N-acetyl cysteamine (SNAC), that mimic the structure of the intermediate generated from the first extension cycle and that can be loaded onto the successive KS domain, restores polyketide production (Jacobsen et al., 1997; Weissman et al., 1998). Although using different diketides could be a powerful approach for production of novel polyketides, the yield of these products are reduced possibly due to the inefficient loading or degradation of the synthetic precursor. Deletion of the entire first extension module from DEBS1 has been reported to increase the utilization of synthetic diketides (Regentin et al., 2004; Ward et al., 2007).
1.4.1.5 Modification of post-PKS tailoring steps

Post-PKS enzymes carry out essential modifications on the polyketide backbone assembled by the synthases which are the keys for the structural diversity and biological activity of this class of natural products. Oxidoreductases are a very broad group of enzymes such as oxygenases, oxidases, reductases and dehydrogenases. These enzymes introduce oxygen-containing functionalities which are important for the stereochemistry, solubility and the reactivity of the final products. Also, they allow successive chemical modifications, such as glycosylation, to occur. Target gene inactivation is a very useful in combinatorial biosynthesis to generate novel compounds with altered biological activities. Inactivation of CYP450 monooxygenase EpoK, involved in the final epoxidation step in epothilone A and B biosynthetic pathway, resulted in the production epothilone C and D, two antitumor drugs more effective than the parental compounds (Tang et al., 2000). Recombination of oxygenase encoding genes is also a method used to generate hybrid product in a mix-and-match fashion. Expression in S. lividans of a combination of ole genes from the oleandomycin PKS with DEBS genes, resulted in the accumulation of a 'un-natural' 6-dEB analogue with a dihydroxy functionality on C8 (Shah et al., 2000).

Glycosyltransferases are responsible for attachment of sugar moieties which often play an essential role in the biological activity of many natural products (Mendez and Salas, 2001; Floss, 2001). These enzymes display a relaxed substrate specificity and therefore are very important for combinatorial biosynthesis. The best example of glycosyltransferase with relaxed substrate specificity is DesVII from the pikromycin and methymicin producer S. venezuelae; its flexibility, towards polyketide acceptor with different sizes of macrolactone rings and NDP-sugar donor substrate, has allowed generation of novel polyketides (Yoon et al., 2002; Tang and McDaniel, 2001).

1.4.2 Combinatorial potential of Type I PKS

The construction of hybrid modular PKSs to assess the truly combinatorial potential of these multienzymes needs to satisfy a number of essential requirements. First, the new 'unnatural' compounds deriving from manipulation of the PKSs should not be lethal for the host which in turn must provide all the building blocks necessary for the hybrid multienzymes to carry out the biosynthesis. Substrate tolerance is a crucial issue especially when domains are relocated as they must be able to recognize un-natural polyketoacyl chain to process. Particular useful for this reason is the identification of
domains with a relaxed specificity that would be able to utilize different substrates (Dutton et al., 1991; Jacobsen et al., 1997; Weissman et al., 1998; Cane et al., 2002).

In principle it should be very easy to try to engineer modular PKSs and observe the outcome. Experiments of this type include additions, deletions and swaps of domains (Gaisser et al., 2003; Petkovic et al., 2003; Starks et al., 2003), replacing multimodular subunits (Kim et al., 2002), and prematurely truncating PKSs by relocating the chain-releasing TE domain to an earlier position (Martin et al., 2003). Domain swapping and standard methods of site-specific mutagenesis experiments have been carried out in order to switch the specificity of AT domains (Del Vecchio et al., 2003). Modification of the gene sequence flanking modules or domains by the addition of unique restriction sites, has been revealed an useful and powerful approach to facilitate the manipulation of modular PKSs with the production of 14 modules from eight PKS clusters and associated them in 154 bimodular combinations (Menzella et al., 2005). While several experiments have reported the successful production of 'unnatural' polyketides from engineered type I PKSs, in most cases the yield of these new compounds is substantially low. The inability of the acceptor KS module to accept the ketide from the upstream module might be one of the explanations for the observed inactive 'unnatural' engineered modular PKS. To overcome this problem, Santi and co-workers co-expressed donor and acceptor modules that contained cognate KSs from DEBSs naturally interacting with each other. This strategy revealed to be successful since the engineered bimodular PKSs produced the expected polyketides (Chandran et al., 2006).

Now it is generally recognized that to achieve truly combinatorial biosynthesis and to allow more efficient directed alterations to synthases with the aim of producing particularly desirable new structures, is very important not to perturb the overall structure of the multifunctional proteins. Therefore, it is required to know more about fundamental aspects of polyketide construction: the three-dimensional (3D) structure of domains, modules and subunits; how domains and modules recognise each other, how they link or dock together and details of how selective individual domains are for their substrates.
1.4.2.1 Structure of Type I PKS multienzyme

In 1996, Staunton and Leadlay proposed a model for the DEBS enzymes which could be extended to other type I PKSs (Fig. 1.7). Different experiments provided direct evidences for the the proposal of this model. Proteolysis of purified DEBS multienzymes followed by characterization of the obtained fragments showed that the DEBS multienzymes are present in duplicate and two identical copies form a dimer with tight interaction between each other (Aparicio et al., 1994; Staunton et al., 1996). Initial proteolysis resulted in of DEBS1 resulted in the release of loading domain as monomer and the other two modules as dimeric units. Extended proteolysis was required to chop the proteins between domains and revealed that individual domains or group of domains, such as KS-AT didomains and TE, remained as dimer while the reductive activities appeared to be singletons. N-terminal sequencing of DEBS proteins and their chopped fragments resulted in just one sequence suggesting a head-to-head orientation for the polypeptide homodimer partners.

The Cambridge model shows the DEBS proteins as homodimers aligned head-to-head in which the subunits of the twin polypeptides twist around each other to form a double helix. The KS, AT, ACP and TE are located at the core of the structure, directly against their counterparts in the opposite protein, while the reductive domains (KR, DH, ER) ‘loop out’ on the periphery of the structure. The KSs are in functional contact with the ACPs of the opposite subunits reflecting the mechanism of the assembly process (Fig. 1.7). A variant for the Cambridge model has been proposed by Cane and Walsh in which the KS and ACP pairs form the core of a tetrahedric structure with a α2β2 heterodimer (Fig. 1.7). The Harvard model proposes that the intervening core AT and auxiliary reductive domains are located on the strands of variable lengths at the edge of the tetrahedron but still in reach of the phosphopantetheine arm of the ACP (Cane and Walsh, 1999). The crystal structure of the TE domain of DEBS and PICS PKS solved at 2.8 Å and 1.3-3.0 Å resolution have demonstrated the homodimeric nature of this domain consistent with both the models suggested for modular PKSs (Tsai et al., 2001; Tsai et al., 2002).

The proposed models of type I PKS have been compared by analogy with the structure of animal I FAS. Mammalian FAS is composed of two identical multifunctional polypeptides initially believed to be arranged each other in head-to-tail fashion (Wakil, 1989).
Figure 1.7 Helical model structures of Type I PKSs
A) Original Cambridge model (Staunton et al., 1996) and B) the variation Harvard model (Cane and Walsh, 1999). For clarity, domains are not drawn to scale.
Combination of complementation experiments, chemical crosslinking and homology modeling of the KS domains turned ‘back to front’ the proposed model suggesting a head-to-head arrangement for the subunits of the mammalian FAS homodimer (Witkowski et al., 2004). The revised model shows the KSs dimer forming the central core with the ACP and with the other domains coiled around so that full range of interactions is permitted. Recently the X-ray crystal structure of the porcine FAS, homologous to human FAS, and a 194 kDa homodimeric fragment from module 5 of DEBS3, have been solved at 4.5 and 2.7 Å resolution respectively (Maier et al., 2006; Tang et al., 2006). In both systems, the KS domains form a homodimer separated from the monomeric AT domains by a distance that could not be just covered by the swinging phosphopantetheine arm of the ACP. These structures suggest that both FAS and PKS complexes have flexible structure to facilitate the directional transfer of the growing chain between the different catalytic sites during the biosynthesis.

1.4.2.2 Protein interactions on DEBS

Structural and functional studies have revealed the importance of the amino acid sequences between successive domains, modules and DEBS polypeptides. Since productive biosynthesis relies on efficient communication between individual domains, these linkers are crucial to establish intra and inter-module interactions within and between contiguous DEBS polypeptides (Gokhale et al., 1999; Gokhale and Khosla, 2000; Wu et al., 2002). Intramodular KS-AT linkers have been reported to consist of a new protein fold with three stranded β-sheet wrapped by two α-helices on one side and one on the other side. Also the characterization of the linker downstream the AT domain in module 3 and 5, which is then followed by a KR domain, has shown to be proline rich and to interact with the KS as wraps around both the AT domain and KS-to-AT linker (Khosla et al., 2007). The poor yield of new ‘unnatural’ products observed in AT swapping experiments could then be attributed to the perturbation of the interactions involving the flanking linkers that results in impaired chain transfer (Hans et al., 2003). Intermodular linkers play also a very important role for efficient biosynthesis catalysed by modular PKSs in particular when successive modules are noncovalently connected (Wu et al., 2001).

The interpolypeptide linkers in DEBS, renamed ‘docking domains’ (Broadhurst et al., 2003), correspond to short amino acid sequences located at the protein termini that do
not belong to the nearby domain and play a role in docking one multienzyme to another. They are entirely modular; intentional mismatching of these domains result in reduced efficiency of chain transfer while swapping matching pairs of docking domains does not appear to be so influent on the assembly process (Tsuji et al., 2001; Wu et al., 2002). Structural analysis of the pair docking domains found at the interface between DEBS1 and DEBS2 and between DEBS2 and DEBS3 has revealed that these linkers are organized in two four-α-helices bundles. The first intertwined bundle is composed of the first two helices at the C terminal of DEBS2 and is presumed to help dimeric interaction of the DEBS multienzymes. The second bundle is composed by the pair of each third helix from the C-terminal of DEBS2 that wraps around the coiled coil formed by the N-terminal of DEBS3. This bundle is likely to be essential for the docking of the DEBS polypeptides and studies involving swapping of the second helical bundle have confirmed the hypothesis (Weissman, 2006) (Fig. 1.8).

Structural and functional studies on the entire modular PKS and on the linkers have highlighted the importance of protein-protein interactions which role is essential for efficient biosynthesis to occur. Attempts to harness the combinatorial potential of modular PKS have shown that the fidelity of the assembly line depends on protein interactions that allow cooperating subunits to recognise their appropriate partner and to discriminate against improper associations. More structural studies that would able to elucidate the insight mechanisms that govern protein interactions are warranted to truly achieve efficient production of new ‘unnatural’ molecules from modular PKSs.

1.4.3 Engineering Type II PKSs
Aromatic PKSs are responsible for the production of an extraordinary variety of compounds with different biological activities. Several attempts have been made to exploit the combinatorial potential of the aromatic PKSs with the aim of producing novel metabolites. In 1990, Bartel et al. reported the first successful experiment of hybrid type II PKSs with the production of novel polyketides using a mix and matching strategy of PKS genes. Systematic studies conducted by Khosla, Hopwood and co-workers to define the role of each enzyme in the mechanism of biosynthesis of aromatic PKSs led to define a series of ‘design rules’ useful for rational production of novel molecules (McDaniel et al., 1995; Hopwood, 1997) (Fig. 1.9).
Figure 1.8 DEBS2-DEBS3 docking domains
The helix bundle is formed by the C-terminal helices of the DEBS2 (green and red) which are presumed to be important for dimerisation as well as the bundle formed by helices 4 at the N-terminal of DEBS3 (shown in orange and light green). The intertwined bundle composed by helices 3 wraps around the N-terminal helices of DEBS3 and is thought to be important for docking.
<table>
<thead>
<tr>
<th>Structural feature</th>
<th>Design rules and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starter unit selection</td>
<td>‘minimal’ PKS can influence the choice</td>
</tr>
<tr>
<td>Chain length</td>
<td>Controlled by the ‘minimal’ PKS, CLF plays a crucial role; KS/CLF pairs must be homologue otherwise may be not functional</td>
</tr>
<tr>
<td>Ketoreduction</td>
<td>Specific KR is required. Often reduction at the C9; Some PKS clusters, such as <em>tcn</em>, lack KR</td>
</tr>
<tr>
<td>First ring cyclisation</td>
<td>Controlled by ‘minimal’ PKS but regiospecificity can be influenced by other subunits in particular KR; Usually observed C7/C12 and C9/C14</td>
</tr>
<tr>
<td>First ring aromatisation</td>
<td>Uncatalysed for unreduced polyketides; Requires an ARO with chain length specificity for reduced molecules</td>
</tr>
<tr>
<td>Second ring cyclisation</td>
<td>Requires an appropriate CYC; Some CYCs have chain length specificity; For long unreduced molecules, it is catalysed by the ‘minimal’ PKS</td>
</tr>
<tr>
<td>Second ring aromatisation</td>
<td>ARO is responsible for aromatisation of unreduced chain for which the first aromatisation is spontaneous</td>
</tr>
<tr>
<td>Further cyclisations</td>
<td>‘Mix and Matching’ experiments performed so far have not included other proteins known to be essential for successive reactions leading to the production of the final polyketides</td>
</tr>
</tbody>
</table>

**Figure 1.9** Examples of ‘design rules’ for the production of novel ‘unnatural’ polyketides (Hopwood *et al.*, 1997)
Also, reconstituted PKS complexes in vitro, that included the act KR, the griseusin ARO/CYC and the tcm ARO/CYC together with the act or tcm ‘minimal’ PKS, were found to produce the polyketides expected from the ‘design rules’ (Zawada and Khosla, 1999).

The recently solved structure of the act heterodimer KSα/KSβ has elucidated how the ‘minimal’ PKS effectively controls chain assembly and the regiochemistry of the first ring closure (Keatinge-Clay et al., 2004). However, several studies have demonstrated the importance of the cooperation of the ‘minimal’ PKS with the tailoring enzymes for an efficient control of chain length and regiochemistry of aromatic polyketides (McDaniel et al., 1994; Kramer et al., 1997; Kantola et al., 1997; Shen et al., 1999; Petkovic et al., 1999). Substitution of the genes encoding for the cyclases (zhuI and zhuJ) of the estrogen receptor antagonist R1128 polyketide synthase with the genes encoding for the act cyclase, aromatase and ketoreductase, led to the production of new un-natural compounds with more cytotoxic activity than R1128 against the breast cancer cell line MCF-7 and with inhibitory activity against glucose-6-phosphate translocase, an attractive target for the treatment of type II diabetes (Tang et al., 2004a).

Unfortunately, rational biosynthesis of novel aromatic polyketides can not be fully achieved only using the ‘design rules’ because many components, in particular downstream enzymes, have more than one role and some have ‘contest dependent behavior’ (Meurer et al., 1997). Attempts to modify the fused ring system in aromatic PKSs from linear (daunorubicin and tetracenomycin) to angular (jadomicyn) and viceversa have revealed that production of an aromatic polyketide with a particular folding pattern can not be simply accomplished by introducing enzymes with known functions belonging to other PKSs in a mix and matching fashion (Meurer et al., 1997; Wohlert et al., 2001). In addition, post-PKS enzymes can also influence the outcome of the polyketide biosynthesis; ablation of the gene otcC, coding for a oxygenase in the pathway leading to the production of oxytetracyclin in Streptomyces rimosus, results in accumulation of novel aromatic compounds with altered chain length (Peric-Concha et al., 2005).

Increased diversity in aromatic polyketides could be obtained using PKSs that incorporate non acetate starter units. These PKSs include additional proteins that form an initiation module responsible for the production of the first condensation product.
subsequently used by the elongation module to carry out all the assembly reactions. Typically these systems contain a protein homolog to the type III KS, FabH, of *E. coli* that is involved in the selection of the starter unit and catalyses the first round of condensation (Jackowski *et al.*, 1989). Also, a protein homolog to a MAT is found in these PKS clusters, which seems to have a role as an acylthioesterase editing acetyl-ACP that could be aberrantly used to prime the KS with acetate. An additional ACP can also be a part of the initiation module (Tang *et al.*, 2004b).

Aklanonic acid (AA) is an anthraquinone intermediate in the biosynthesis of several clinically important aromatic polyketides including daunorubicin and aclacinomycin A (Hopwood, 1997). Biosynthetic efforts have been directed in generating of new AA analogues by engineering the R1128 initiation module with the elongation module of dodecaketide producing PKSs, antibiotic pradimicin from *Actinomadura hibisca* and spore pigment *whiE* from *S. coelicolor*. Production of non-acetyl primed AA analogues was obtained only using the pradimicin elongation module while the correspondent from *whiE* PKS failed to generate new AA derivative polyketides (Lee *et al.*, 2005). These efforts to harness the combinatorial potential of aromatic PKSs have shown that, as for modular PKSs, the cooperation of single enzymatic activities by protein interactions is the key for rational design of new ‘unnatural’ polyketides.

### 1.5 DNR/DXR Polyketide synthase

Daunorubicin (DNR) and its C-14 hydroxylated derivative doxorubicin (DXR) are currently two of the most widely used anticancer drugs (Minotti *et al.*, 2004). These two anthracyclines are produced by the filamentous soil bacterium *Streptomyces peucetius* through a pathway involving a type II PKS. The aromatic PKS catalyses the assembly of polyketide chain, by condensation of a propionate starter unit (from propionyl-CoA) with nine acetate extender units (derived from malonyl-CoA) and its subsequent chemical modifications (ketoreduction, aromatisation and cyclisation) leading to the production of the 21 carbon decaketide aklanonic acid which is the first free enzyme intermediate in the biosynthetic pathway. After the polyketide stage is complete, AA undergoes successive modifications, such as glycosylation with daunosamine, to produce DNR which is hydroxylated at C14 to yield DXR (Hutchinson and Colombo, 1999) (Fig. 1.10).
Figure 1.10 Biosynthesis of Daunorubicin and Doxorubicin
A) Dps proteins produce aklanonic acid which is subsequently converted by tailoring enzymes in DNR and DXR.
B) Genetic organization of the *dps* cluster.
The DNR/DXR PKS stands out amongst other type II PKSs for the choice of propionate as starter unit for the assembly of the polyketide instead of the more commonly used acetate. The biosynthetic cluster ($dps$), shown in Figure 1.10 is composed of eight genes, $dpsABCDEFGY$ (Grimm et al., 1994).

DpsA and DpsB are the $\alpha$ and $\beta$ subunits of the KS respectively that catalyse the elongation reactions for the assembly of the nascent polyketide chain which is tethered to the ACP, DpsG. Typically aromatic PKS gene clusters, including act PKS in S. coelicolor, present a genetic organization with the gene encoding for the ACP immediately downstream the ones for the subunits of the KS and convergently transcribed (Hopwood, 1997). Instead in the DNR/DXR PKS gene cluster, $dpsG$ is located 6.8 kbp upstream $dpsA$ and is divergently transcribed (Grimm et al., 1994; Ye et al., 1994). Downstream $dpsA$ and $dpsB$ there are two genes $dpsC$ and $dpsD$ which homologues have been found in other aromatic PKSs clusters that do not utilize acetate as starter unit (Raty et al., 2002; Marti et al., 2000).

$dpsC$ encodes for a protein that is very similar to the E.coli FabH a type III ketosynthase III (KSIII). FabH is an essential protein and catalyses the first condensation reaction that initiates chain elongation in the biosynthesis of type II fatty acid synthase (Jackowski et al., 1989; Revill et al., 2001; Lai and Cronan, 2003). The enzymatic activity of FabH involves a ping pong mechanism to catalyse the condensation of the starter unit, typically acetyl-CoA, with the first malonate extender unit. The acetyl group is firstly transferred to its active site sulphhydryl with release of CoA and then FabH binds malonyl-ACP condensing this intermediate with the active site acyl group to produce acetoacetyl-ACP (Jackowski et al., 1989). Compared to other KSIIIs that bind the acyl group via thioester linkage with the active site cysteine, DpsC display the unusual characteristic to contain a serine residue (Ser118) in place of cysteine forming ester bond with its substrate propionate (Bao et al., 1999a). In addition, gel filtration chromatography revealed that the active form of DpsC appears to be monomeric compared to the homodimeric organization of its homologues FabH and ZhuH, from oestrogen receptor antagonist R1128 PKS (Musayev et al., 2005; Qiu et al., 2001; Pan et al., 2002). $dpsD$ encodes for a protein that shows high similarity with a malonyl-CoA:ACP acyltransferase of bacterial FASs and also contains the essential serine in the consensus sequence, suggesting a role as acyltransferase (Grimm et al., 1994).
The role of DpsD is quite obscure since heterologous expression of the \textit{dps} genes showed that its gene can be omitted without any influence on the production of AA (Grimm \textit{et al.}, 1994).

Initial experiments have shown that heterologous expression of all genes involved in the aklanonic acid biosynthesis, lead to production of the polyketide in absence of \textit{dpsC} and \textit{dpsD} (Rajgarhia and Strohl, 1997) suggesting that these genes were dispensable for the biosynthesis of the anthracycline. A later study by Bao \textit{et al.} using both \textit{in vivo} and \textit{in vitro} data demonstrated that DpsC plays a primary role in propionyl starter unit discrimination (Bao \textit{et al.}, 1999b). Other experiments showed that the \textit{dpsCD'} double mutant \textit{Streptomyces sp.} strain C5VR5 accumulated compounds deriving from the utilization of an acetate starter unit in place of propionate (Rajgarhia \textit{et al.}, 2001).

Additionally, \textit{Streptomyces lividans} TK24 transformants carrying all the genes that support aklanonic acid biosynthesis produced the anthracycline whereas deletions of both \textit{dpsC} and \textit{dpsD} resulted in a production reduced to only 20\% and desmethylaklanonic acid, which derives from the incorporation of acetate as starter unit, was the major product. Presence of DpsD but not of DpsC resulted in the formation of both desmethylaklanonic and aklanonic acid in a 60:40 ratio. However, expression of \textit{dpsC} independently from the presence of DpsD, resulted in production of only aklanonic acid. These results suggested that promiscuous starter unit selection can occur in absence of \textit{dpsC} and \textit{dpsD} and that DpsC contributes to the selection of propionyl-CoA as starter unit but it does not alone dictate it (Rajgarhia \textit{et al.}, 2001). In addition, cell free extract of \textit{S. lividans} TK24 carrying all the \textit{dps} genes but excluding \textit{dpsD} incubated with both propionyl-CoA and malonyl-CoA failed to produce any compound (Rajgarhia \textit{et al.}, 2001). These results did not confirm the \textit{in vivo} data that indicated that the presence of DpsC could be sufficient for the production of only aklanonic acid. This may suggest that in absence of DpsD, the PKS complex might be unstable and unable to carry on polyketide biosynthesis. \textit{In vivo} the absence of DpsD could be compensated by other cellular components. Indeed, cross-talk between the \textit{S. coelicolor} FAS malonyl-CoA: ACP acyltransferase and the \textit{act} PKS complex has previously been suggested (Revill \textit{et al.}, 1995; Summers \textit{et al.}, 1995).

\textit{dpsE} encodes for a ketoreductase that specifically acts on \(\beta\)-keto group on the C9 of the nascent acyl chain (Grimm \textit{et al.}, 1994). \textit{dpsF} encodes for an aromatase that catalyses enzymatic dehydrations after the first aldol condensation between C7 and C12 of the
polyketide backbone to yield the aromatic ring. \(dpsY\) encodes for a cyclase which is involved in the formation of the second and third ring of AA (Grimm et al., 1994; Lomovskaya et al., 1998; Hautala et al., 2003). The important role of DpsY in the biosynthesis of AA had already been reported by previous studies revealing that the disruption of the gene in \(S.\ peucetius\) leads to the accumulation of UWM5, the C-19 ethyl homologue of SEK 43, a known shunt product in the biosynthesis of aromatic polyketides (Grimm et al., 1994; Lomovskaya et al., 1998).

1.5.1 Clinical use of DNR and DXR

Anthracycline are a class of compounds that display the second widest spectrum of activity of any antitumour agents second only to the alkylating agents (Minotti et al., 2004). DNR and DXR (known also as adriamycin) were the first isolated anthracyclines in early 1960s and to date are amongst the most effective chemoterapic agents against human cancers (Rabbani et al., 2005). These two compounds share the same biosynthetic pathway and similar chemical structure; the only difference is that the side chain of DNR terminates with a methyl which is, instead, hydroxylated in DXR. This minor difference confers different spectrum of activity to these two anthracyclines. DOX is essentially used for solid tumours and aggressive lymphomas while DNR is active against myelogenous leukemias (Buchner et al., 2004; Carrion et al., 2004). Despite the extensive clinical utilisation, there is still some controversy about the mechanism of action of DXR and DNR for the anticancer activity.

There are evidences that nuclear DNA is the primary target of anthracyclines that act essentially as intercalating agents by insertion of the anthraquinone ring between DNA bases and its stabilization due to hydrophobic and stacking interactions but also for hydrogen bonds (Chaires et al., 1996). Intercalation into the DNA interferes with the activity of the topoisomerases (Fig. 1.11). These enzymes modify the topology of DNA leading to formation of transient single strand (type I) or double strand (type II) DNA breaks subsequently resealed after changing the twisting status of the double helix (Binaschi et al., 2001). Anthracyclines stabilize the intermediate in the reaction in which the DNA strands are cut and covalently bound to the topoisomerase II impairing DNA repairing.
Figure 1.11 Structurally important parts of anthracyclines (A) and novel analogues of DNR/DXR (B)
This effect, known as ‘topoisomerase poisoning’, leads to growth arrest triggering cell apoptosis (Zunino et al., 2001). Covalent modification of the DNA by interstrand crosslinking (Purewal et al., 1993; Parker et al., 1999) and generation of free radicals with consequent DNA and membrane damage also have been proposed as mechanisms of actions for anthracyclines (Minotti et al., 2004).

1.5.2 DNR and DXR analogues

The clinical use of DNR and DXR is limited due to dose-related cardiac toxicity, related to the high cardiac concentrations of DXR and its oxygenated metabolites, and tumour resistance (Minotti et al., 2004). Several attempts have been made to generate novel anthracyclines with improved biological activity and/or fewer side effects. Cytotoxic activity is complex and drug base interactions are dictated in part by the structure of the anthracyclines. Even small changes in the structure are known to significantly alter the pharmacokinetic properties of anthracyclines (Minotti et al., 2004). More than 2000 DNR analogues have been made to improve activity and/or cardiac tolerability (Weiss, 1992). Distinct functional domains of the anthracycline molecule have been defined and structure function relationships have been established that drug intercalation is necessary but not sufficient for topoisomerase poisoning (Capranico et al., 1990). Moreover, the removal of the 4-methoxy and 3 amino substituents of the sugar moiety markedly influence the sequence selectivity of anthracycline DNA cleavage (Binaschi et al., 2000).

The sugar moiety of anthracyclines serves as minor groove binder of DNA and the occurrence of the single positive charge contribute to the binding (Capranico et al., 1994; Capranico et al., 1995). Modifications of the sugar structures have led to the second generation of DXR analogues as conformation of the monosaccharide such epirubicin and pirarubicin (Fig 1.11). The structure of the terminal sugar and its chain length have been shown to significantly influence DNA binding, topoisomerase poisoning and anticancer efficacy (Pratesi et al., 1998). More interest has raised for third generation anthracyclines, bearing disaccharides in place of the monosaccharide that have exhibited an increased anticancer activity (Zhang et al., 2005) (Fig. 1.11).

However, none of all the DNR analogues so far has given an effective improvement about optimizing the cytotoxic effect, increasing the DNA damage, without expose patients to overt cardiac damage and so the search for a better anthracycline continues.
1.6 Aim of the project

Biosynthesis of aromatic polyketides relies on the formation of multiprotein complexes in which each protein cooperates with the others and contributes with a structural role, rather than just with an enzymatic activity, to define the chemical properties of the final product such as choice of the starter unit, chain length and regiochemistry. The ‘design rules’ defined by Hopwood and collaborators have helped addressing each activities to specific proteins but successive attempts to generate novel molecules have highlighted the importance of a better understanding of the insight molecular mechanisms that rule the biosynthesis catalysed by type II PKSs. Protein interactions, responsible for crosstalk between different subunits necessary for effective and productive biosynthesis, are the key for successful use of combinatorial biosynthesis in the production of new ‘unnatural’ compounds.

DNR and DXR are currently two of the most effective and widely used anticancer drugs but the related cardiotoxicity and tumour resistance have limited their clinical use. There is a tremendous interest in identifying novel DNR/DXR derivatives that could retain anticancer efficacy without causing severe cardiac problems. New generation analogues have been designed by chemical modifications of the tetracyclic ring, side chain or the aminosugar moiety but yet the identification of a ‘better’ anthracycline, through synthetic approaches, is far to come. Also, chemical modification of single functional groups of a natural compound may require different steps which lead to a poor yield of the final product. Combinatorial biosynthesis is a relatively unexplored field for molecular design with few successfully examples of production of new polyketides obtained and offers great promise to generate novel anthracyclines with more suitable characteristics as anticancer drug in industrial relevant yield. However, the use of this approach has probably been hampered because our lack in understanding of the protein interactions required to form a functional multienzyme complex.

Using the DNR/DXR PKS as model, the overall aim of this project was to investigate protein interactions of a type II PKS complex. Chapter 3 describes the investigation of the in vivo interactions between the proteins forming the ‘minimal’ DNR/DXR PKS using the GAL4 yeast two-hybrid system. This powerful and sensitive technique was also used to identify the network of interactions involving the entire PKS complex including the downstream enzymes (Chapter 4). The application of a relatively novel technique for purification of multiprotein complexes, Tandem Affinity Purification
(TAP), to the investigation of the protein interactions involving the ‘minimal’ PKS is reported in Chapter 5. Attempts of a bioinformatic approach to study from a structural point of view the protein interactions detected using the GAL4 yeast two-hybrid system, using a docking simulation program, are described in Chapter 6.
CHAPTER 2

Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

Chemicals were supplied by Sigma-Aldrich (Saint Louis MI, USA) unless otherwise stated and were of the highest analytical grade available. Agarose was of electrophoresis grade. Deoxyribonucleotides were purchased from Eppendorf (Hamburg, Germany). Distilled water, further purified using the MilliQ water purification system was used for all applications.

2.1.1.2 Oligonucleotides

All oligonucleotide primers were purchased desalted from Thermo Electron Corporation (Waltham MA, USA). Oligonucleotides were resuspended in sterile 20 mM Tris-Cl buffer pH 8.5 and diluted to 200 pmol before storage at -20 °C. The sequences of primers used in this study are listed within the text of the relevant chapters.

2.1.2 Buffers and solutions

Buffers and solutions were prepared as described by Sambrook and Russell (2001), Kieser et al. (2000) and ‘Yeast Protocols Handbook’ (2001) (Clontech). Other buffers and solutions are listed below and in the relevant sections.

<table>
<thead>
<tr>
<th>Agarose gel loading buffer (6X)</th>
<th>Electrophoresis buffer agarose gel (TAE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25% (w/v) Bromophenol blue</td>
<td>40 mM Tris-acetate</td>
</tr>
<tr>
<td>0.25% (w/v) Xylene cyanol FF</td>
<td>1 mM EDTA</td>
</tr>
<tr>
<td>30% (v/v) Glycerol</td>
<td></td>
</tr>
</tbody>
</table>

Trace element solution (T³⁻)

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ZnCl₂ 40 mg</td>
</tr>
<tr>
<td>FeCl₃×6H₂O 300 mg</td>
</tr>
<tr>
<td>CuCl₂×2H₂O 10 mg</td>
</tr>
<tr>
<td>MnCl₂×4H₂O 10 mg</td>
</tr>
<tr>
<td>Na₂B₄O₇×10H₂O 10 mg</td>
</tr>
<tr>
<td>(NH₄)₆Mo₇O₂₄×4H₂O 10 mg</td>
</tr>
</tbody>
</table>
**TES buffer (1)**

TES 53.7 g

Titrate with NaOH to pH 7.2

**Tris maleic acid buffer**

Tris 1 M

Titrate with maleic acid to pH 8

### 2.1.3 Culture media

All media ingredients were purchased from Oxoid (Hampshire, UK). Yeast Nitrogen Base and drop-out (DO) supplements were supplied by BD Biosciences (Palo Alto CA, USA) for preparation of Synthetic Defined (SD) DO medium. All media were sterilised by autoclaving at 121 °C for 15 minutes. Bacteriological agar to a final concentration of 2% (w/v) was added to obtain solid media. All volumes are to 1 litre.

**LB broth**

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>10 g</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>5 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>10 g</td>
</tr>
</tbody>
</table>

**YPAD**

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptone</td>
<td>20 g</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>10 g</td>
</tr>
<tr>
<td>Adenine hemysulphate</td>
<td>0.003% (w/v)</td>
</tr>
<tr>
<td>Glucose</td>
<td>2% (w/v)</td>
</tr>
</tbody>
</table>

Titrate with NaOH to pH 5.8

**SD drop-out -2 (-Leu/-Trp)**

Yeast Nitrogen Base (with (NH₄)₂SO₄ w/o amino acids) 6.7 g

- Leu/-Trp DO supplement 0.64 g
  - Glucose 2% (w/v)

Titrate with NaOH to pH 5.8

**SD drop-out -4 (-Leu/-Trp/-Ade/-His)**

Yeast Nitrogen Base (with (NH₄)₂SO₄ w/o amino acids) 6.7 g

- Leu/-Trp/-Ade/-His DO supplement 0.6 g
  - Glucose 2% (w/v)

Titrate with NaOH to pH 5.8
YEME
Yeast Extract 3 g
Bacto-Peptone 5 g
Malt Extract 3 g
Glucose 1%
Sucrose 34%
MgCl₂ 5 mM

For preparing protoplasts also:
Glycine 20%

Regeneration medium for *Streptomyces coelicolor* protoplasts (R5 Medium)
Sucrose 10.3%
K₂SO₄ 0.25 g
MgCl₂ 10.12 g
Glucose 1%
Casamino acids 0.1 g
Trace Element solution 2 ml
Yeast Extract 5 g
TES 5.73 g
H₂O up to 1000 ml

100 ml of the above solution was poured in 250 ml Erlenmeyer flask containing 2.2 g of bacteriological agar and autoclaved. Prior use, medium was re-melted and adjusted to:
KH₂PO₄ 0.005%
CaCl₂ 20 mM
L-proline 0.3%
NaOH 70 mM
2.1.4 Antibiotic concentrations
Antibiotics were filter sterilised using a 0.22 μm sterile syringe filter (Millipore) (Billerica MA, USA) and stored as 200 µl aliquots at -20 °C before use.

<table>
<thead>
<tr>
<th>Antibiotic (solvent)</th>
<th>Stock concentration (mg/ml)</th>
<th>Working concentration (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin (water)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Kanamycin (water)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Chloramphenicol (ethanol)</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Thiostrepton (DMSO)</td>
<td>50</td>
<td>50 (agar), 15 (liquid),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 (selection)</td>
</tr>
</tbody>
</table>

2.1.5 Source of enzymes
All restriction enzymes used were purchased from New England Biolabs (NEB) (Ipswich, UK). RNase A (bovine pancrease) and lysozyme (chicken egg white) were purchased from Sigma. DNA polymerases Taq and Pfu were purchased from NEB and Promega (Madison WI, USA) respectively.

2.2 Growth and manipulation of microorganisms
All bacteriological procedures involving Saccharomyces cerevisiae and Streptomyces spp were performed in a class II laminar-flow hood (NUAIRE NU437-300E). Standard aseptic techniques were used with E. coli.
2.2.1 Strains

Bacterial strains used in this study:

<table>
<thead>
<tr>
<th>Organism</th>
<th>Use and Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> TOP10</td>
<td>Routine cloning, <em>mcrA, Δ (mrr-hsdRMS-mcrBC), φ80 Δ lac Δ M15, Δ lacX74, deoR, recA1, araD139 Δ (ara, leu), 7697, galU, galK, λ, rpsL, endA1, mupG</em></td>
<td>Invitrogen, Carlsbad CA, USA</td>
</tr>
<tr>
<td><em>Escherichia coli</em> K12 ER2925</td>
<td>Preparation of unmethylated plasmidic DNA, <em>ara-1, leuB6, fhuA31, lacy1, tsx78, glnV44, galK2, galT22, mcrA, dcm-6 hisG4, rfbD1, R(zgb210::Tn10)TetS, endA1, rpsL136, dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2</em></td>
<td>NEB</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em> AH109</td>
<td>Yeast two-hybrid system, <em>MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1 UAS-GAL1 TATA-HIS3, GAL2 UAS-GAL2 TATA-ADE2; URA3::MEL1 UAS-MEL1 TATA-lacZ</em></td>
<td>James <em>et al.</em>, 1996</td>
</tr>
<tr>
<td><em>Streptomyces coelicolor</em> A3(2) M145</td>
<td>Tap Tag experiments, Prototrophic SCP1* SCP2*</td>
<td>Kieser <em>et al.</em>, 2000</td>
</tr>
</tbody>
</table>
2.2.2 Plasmids

Sources of plasmids used for the construction of further vectors in this study:

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Use and Description</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Commercial vectors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pET100D-TOPO</td>
<td>Routine <em>E. coli</em> cloning&lt;br&gt;(amp^\prime, kan^\prime)</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pGEM(^\circ)-T Easy</td>
<td>Routine <em>E. coli</em> cloning&lt;br&gt;(amp^\prime)</td>
<td>Promega</td>
</tr>
<tr>
<td>pGADT-7</td>
<td>Yeast two-hybrid system&lt;br&gt;Gal4 TAD, HA epitope, LEU2, Amp(^\prime)</td>
<td>Clontech</td>
</tr>
<tr>
<td>pGBK7-7</td>
<td>Yeast two-hybrid system&lt;br&gt;Gal4 DNA-BD, c-Myc epitope, TRP1, Kan(^\prime)</td>
<td>Clontech</td>
</tr>
<tr>
<td>pVA3-1</td>
<td>Yeast two-hybrid system&lt;br&gt;mouse p53 Gal4 DBD, TRP1, Amp(^\prime)</td>
<td>Clontech</td>
</tr>
<tr>
<td>pTD1-1</td>
<td>Yeast two-hybrid system&lt;br&gt;SV 40 large T antigen, Gal4 TAD, LEU2, Amp(^\prime)</td>
<td>Clontech</td>
</tr>
<tr>
<td><strong>Non-commercial vectors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBS1761</td>
<td>Tap Tagging of proteins at the N-terminal&lt;br&gt;(TAP) tag, Amp(^\prime)</td>
<td>EUROSCARF University of Frankfurt, Germany</td>
</tr>
<tr>
<td>pWHM1012</td>
<td><em>E. coli</em>-Streptomyces shuttle vector&lt;br&gt;(dps) gene cluster, Amp(^\prime)</td>
<td>Bao et al., 1999b</td>
</tr>
</tbody>
</table>
2.2.3 Manipulation of *Escherichia coli* cultures

The growth and cultivation of *E. coli* used standard procedures as described by Sambrook and Russell (2001). *E. coli* strains were stored as glycerol cultures by addition of 500 μl sterile 50% (v/v) glycerol to 500 μl well-grown cultures and stored at -80 °C.

2.2.3.1 Preparation and transformation of chemically competent *E. coli* cells

A single bacterial colony was inoculated in 5 ml of LB broth and grown up at 37 °C for 12 hours. The culture was diluted 1:100 in 100 ml of LB broth and grown up to exponential phase (OD_{600nm} 0.4-0.6). The cells were left on ice for 30 minutes and then centrifuged for 5 minutes at 1,500 x g at 4 °C. The cells were then resuspended in 50 ml of a filter-sterile chilled solution 50 mM CaCl\(_2\) and incubated on ice for 15 minutes. After centrifugation for 5 minutes at 1,500 x g at 4 °C, the cells were resuspended in 8.5 ml of 50 mM CaCl\(_2\) and incubated on ice for 1 hour. 1.5 ml of 100% (v/v) glycerol was added to the cell suspension that was then aliquoted and stored at -80 °C.

Transformation was carried out adding plasmid DNA to 200 μl of competent cells and incubating it on ice for 30 minutes. Thermal shock was performed incubating the cells at 42 °C for 3 minutes followed by adding 800 μl of LB broth to the cells. After 1 hour incubation at 37 °C, the cells were plated out on LB agar plates containing a selective antibiotic.

2.2.4 Manipulation of *Saccharomyces cerevisiae* cultures

The growth and the genetic manipulation of *S. cerevisiae* cultures were carried out according to ‘Yeast Protocols Handbook’. *S. cerevisiae* cultures were stored in 15% (v/v) glycerol at -80 °C.

2.2.4.1 Preparation and transformation of *S. cerevisiae* competent cells.

A single yeast colony was inoculated in 5 ml of YPAD broth and grown up for 12 hours at 30 °C. The culture was diluted to an OD\(_{600}\) of 0.15 in 50 ml of YPAD broth and incubated at 30 °C until an OD\(_{600}\) of 0.6 was reached. The cells were centrifuged for 4 minutes at 2,500 x g at room temperature and washed with 50 ml of sterile distilled water. After centrifugation for 4 minutes at 2,500 x g at room temperature, the cells were washed with 12.5 ml of LiSorb solution. The cells were centrifuged twice for 4 minutes at 2,500 x g at room temperature to remove completely the remaining LiSorb
and re-suspended in a solution containing 1 ml of LiPEG, 300 µl of LiSorb and 1 mg of Salmon Sperm DNA.

The transformation was performed by adding up to 2 µg of plasmidic DNA to 130 µl of the cell suspension. After a vigorous mixing by vortexing, the transformation reaction was incubated for 25 minutes at room temperature. 15 µl of DMSO was added and mixed by pipetting and a thermal shock applied by incubating the cells at 42 °C for 10 minutes. The cells were centrifugated at 3,500 x g for 3 minutes at room temperature and the supernatant gently removed. The cells were resuspended in 200 µl of sterile water and 100 µl plated out on drop-out medium (−2 and −4). The plates were then incubated for up to 14 days at 30 °C.

**LiPEG**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₃COOLi</td>
<td>100 mM</td>
</tr>
<tr>
<td>Tris-Cl</td>
<td>10 mM</td>
</tr>
<tr>
<td>EDTA</td>
<td>1 mM</td>
</tr>
<tr>
<td>PEG3530</td>
<td>40% (w/v)</td>
</tr>
<tr>
<td>Filtered</td>
<td>(0.22 µm)</td>
</tr>
</tbody>
</table>

**LiSorb**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₃COOLi</td>
<td>100 mM</td>
</tr>
<tr>
<td>Tris-Cl</td>
<td>10 mM</td>
</tr>
<tr>
<td>EDTA</td>
<td>1 mM</td>
</tr>
<tr>
<td>Sorbitol</td>
<td>1 M</td>
</tr>
<tr>
<td>Filtered</td>
<td>(0.22 µm)</td>
</tr>
</tbody>
</table>

### 2.2.5 Manipulation of *Streptomyces coelicolor* cultures

Cultures of *S. coelicolor* were manipulated, grown and stored as described by Kieser *et al.* (2000). Conical flasks equipped with springs were used and cultures typically incubated at 30 °C with shaking at 200 rpm. *S. coelicolor* strains were stored as spore suspensions in 20% (v/v) glycerol at -20 °C.

#### 2.2.5.1 Growth of *S. coelicolor* cultures

Mycelial colonies were grown by streaking a 10 µl loop of stored cultured onto R5 agar plates and incubated at 30 °C for 48 hours. A 4 cm² surface area of mycelia was excised, crushed and used to inoculate 50 ml of YEME which was then incubated in baffled flasks with shaking at 30 °C for 48 hours at 200 rpm. For large-scale cultures, 50 ml of the above culture was used to inoculate 500 ml of YEME and incubated at 30 °C at 200 rpm for 48 hours or until the culture was densely grown.

For TAP tag experiments, *S. coelicolor* A3(2) carrying plasmid pNC147 was grown in 50 ml of YEME medium containing thiostrepton (15 µg/ml) in 250 ml baffled flasks at
30 °C at 200 rpm for 48 hours. This seed culture was used to inoculate 500 ml of YEME medium (thiostreptone 15 μg/ml) in baffled flaks and incubated 30 °C at 200 rpm for 20 hours.

2.2.5.2 Preparation and transformation of *S. coelicolor* A3(2) protoplasts

The preparation of *S. coelicolor* protoplasts and transformation with plasmid DNA were performed as described by Keiser *et al.* Briefly, 0.1 ml of spore suspension was used to inoculate 25 ml of YEME medium in a baffled flask and incubated at 30 °C for 36-40 hours with shaking at 200 rpm. The culture was centrifugated at 1,000 x g for 10 minutes and the pellet washed twice in 15 ml 10.3% (w/v) sucrose. The mycelium was then resuspended in 4 ml of buffer P containing lysozyme at the concentration of 2 mg ml⁻¹ and incubated at 30 °C for up to 40 minutes. Additional 5 ml of P buffer was added and the suspension filtered through cotton wool. Protoplasts were gently centrifuged at 1,000 x g for 7 minutes, resuspended in 1 ml of P buffer and divided in aliquots of 200 μl that were either used directly or stored at -80 °C for future use.

The transformation of *S. coelicolor* protoplasts was carried out using the rapid small-scale procedure with 50 μl of protoplast suspension used for each transformation. Up to 1 μg of plasmid DNA was added to the protoplast suspension that was mixed gently by tapping the tube. 200 μl of T-buffer was added and mixed by pipetting 4 times. The protoplast suspension was equally divided and plated out on two dried R5 plates. The plates were incubated at 30 °C for 20 hours before doing the selection by flooding and then were kept at 30 °C for 3 days to score resistant colonies.

**T (transformation) buffer**

10.3% (w/v) Sucrose 2.5 ml
33% (w/v) PEG₁₀₀₀ 7.5 ml
Trace element solution 0.02 ml
2.5% (w/v) K₂SO₄ 1 ml

Before use, to 9.3 ml of the above solution was added:

5M CaCl₂ 0.2 ml
Tris-maleic acid buffer (pH 8.0) 0.5 ml
**P (protoplast) buffer**

- Sucrose 103 g
- K$_2$SO$_4$ 0.25 g
- MgCl$_2$6H$_2$O 2.02 g
- Trace element solution 2 ml
- Distilled water to 800 ml

Before use, to 80 ml of the above solution was added:

- KH$_2$PO$_4$ (0.5%) 1 ml
- CaCl$_2$H$_2$O (3.68%) 10 ml
- TES buffer (5.73% adjusted with NaOH to pH 7.2) 10 ml

### 2.3 DNA manipulation techniques

Recombinant DNA manipulation techniques were carried out as described by Sambrook and Russell (2001).

#### 2.3.1 Plasmid DNA preparation

Plasmid DNA was isolated from 5 ml of an overnight culture of *E. coli* using the alkaline lysis technique described by Sambrook and Russell (2001). For a higher yield of plasmid DNA, a single colony was inoculated into 50 ml of LB broth containing the appropriate general antibiotic and the purification of the DNA was carried out by using DNA MIDI-PREP (QIAGEN GmbH) (Hilden, Germany) according to the manufacturer's instructions.

#### 2.3.2 DNA restriction digestion

For analytical reactions, 300 ng of DNA was digested with restriction enzyme(s) according to the manufacturer's instructions in a final volume of 10 µl. For preparative amounts of a DNA fragment, between 1 and 5 µg of plasmid was digested with 5-20 units of restriction enzyme according to the manufacturer’s guidelines before analysis by agarose gel electrophoresis.

#### 2.3.3 Agarose gel electrophoresis and recovery of DNA

Horizontal slab gels were cast using various concentrations of (0.7-2.0% w/v) agarose (Molecular Biology Certified Agarose) in TAE buffer. Ethidium bromide was added to the molten agarose prior to casting of the gels at a final concentration of 0.1 µg ml$^{-1}$. 

---

70
Samples were mixed with loading buffer and loaded into the gel and run at between 60 to 115 V (1-5 V cm⁻¹) for 1-4 hours. Molecular weight markers (1 Kbp Ladder or 100 bp DNA ladder-Promega) were loaded and run next to the samples. DNA was visualised at 302 nm using a UV transilluminator and recorded using a UVP gel documentation system (UVP Inc., USA). DNA fragments were extracted from the agarose gel using the QIAGEN II Gel Extraction Kit (QIAGEN GmbH) according to the manufacturer's protocols.

2.3.4 Ligation of DNA
DNA fragments (insert) were mixed with the DNA plasmid (vector) in a molar ratio 5:1 and incubated in a volume of 15 μl with 5 units of T4 DNA Ligase and in presence of the appropriate reaction buffer. The reaction mixture was incubated either overnight at 16 °C to ligate blunt ended DNA fragments, or for 2 hours at 37 °C for cohesive ends.

2.3.5 DNA sequencing
DNA was submitted at a concentration of 100 ng/μl and purity of 1.8 OD (absorbance ratio A₂₆₀/A₂₈₀) for sequencing at the Department of Biochemistry, University of Cambridge, UK.

2.3.6 Polymerase Chain Reaction (PCR)
An Eppendorf “Mastercycler gradient” was used for DNA amplification by PCR according to the DNA polymerase manufacturer’s guidelines. 50 ng template DNA was added to a PCR mixture in a total of 50 μl containing PCR buffer, 5 μM oligonucleotides, 0.2 mM dNTP and 5 U of either Taq or Pfu DNA polymerase. DMSO at final concentration of 5% (v/v) was included in the reaction mixture, since this is beneficial for product amplification from high G+C templates, as it lowers the template denaturation temperature.

Typical PCR program: a first denaturation carried out for 5 minutes at 95 °C, 30 denaturation cycles (1 minute at 95 °C), annealing (1 minute at 56-62 °C, depending on the C/G % of the primers) and elongation (72 °C with extension time depending on the length of the DNA to amplify generally 1 minute per kilobases) plus 5 minutes at 72 °C for the termination of the chains. The products of the reaction were visualized by UV on agarose gel.
2.4 Yeast two-hybrid system methods

2.4.1 Extraction of proteins from *S. cerevisiae*

Total proteins from yeast transformants were extracted using 1 ml of overnight culture. The cells were collected by centrifugation and re-suspended in 1 ml of cold sterile distilled water. 150 µl NaOH/β-ME buffer were added to the suspension that was then vortexed for 30 seconds and incubated on ice for 15 minutes. After incubation, the cells were mixed and 150 µl of 55% (w/v) TCA (in water) was added and incubated on ice for 10 minutes. The protein extract was collected by centrifugation at 12,000 x g for 10 minutes at 4 °C. The pellet was resuspended in 300 µl of SU buffer. The protein extract was heated at 65 °C for 3 minutes prior loading to SDS/PAGE.

<table>
<thead>
<tr>
<th>NaOH/β-ME buffer</th>
<th>SU buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.85 M NaOH</td>
<td>5% (w/v) SDS</td>
</tr>
<tr>
<td>7.5% (v/v) β-Mercaptoethanol</td>
<td>8 M Urea</td>
</tr>
<tr>
<td></td>
<td>125 mM Tris-Cl pH 6.8</td>
</tr>
<tr>
<td></td>
<td>0.1 mM EDTA</td>
</tr>
<tr>
<td></td>
<td>0.005% (w/v) Bromophenol Blue</td>
</tr>
<tr>
<td></td>
<td>15 mg/ml DTT</td>
</tr>
</tbody>
</table>

2.4.2 *LacZ* assay

*LacZ* assay was performed as described in the ‘Yeast Protocols Handbook’. Briefly, yeast transformants were streaked onto SD-2 drop-out medium agar plates and incubated for 3 days at 30 °C. A sterile Whatman #5 filter was soaked in 5 ml of Z-buffer/X-gal solution. A dry sterile filter was placed over the surface of the colonies to be assayed. The filter was rubbed gently with forceps in order to lift the colonies. The filter was taken off the agar plate and transferred (colonies facing up) to liquid nitrogen. After the filter was completely frozen (generally 30 seconds), it was allowed to thaw at room temperature. The filter was then placed, colonies side up, on pre-soaked filter and incubated at 30 °C for up to 8 hours. Periodically, the filter was checked for the appearance of blue colonies.
**Z-buffer (pH 7.0)/X-gal solution**

- 60 mM Na$_2$HPO$_4$
- 40 mM NaH$_2$PO$_4$
- 10 mM KCl
- 1 mM MgSO$_4$
- 0.27% (v/v) β-Mercaptoethanol
- 0.167% (w/v) X-gal

### 2.5 Protein chemistry

#### 2.5.1 Protein quantification

Protein concentrations were determined in duplicate by the dye-binding methods of Bradford (Bradford, 1976). Standard curves were constructed using duplicate samples of BSA over a concentration range of 0-30 μg ml$^{-1}$.

#### 2.5.2 SDS-Polyacrylamide Gel Electrophoresis

SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed using vertical polyacrylamide slab gels as described by Laemmli (1970). Resolving gels at 10%, 12% and 15% (w/v) acrylamide were used. Stacking gels were made of 5% (w/v) acrylamide. The composition of the acrylamide stock solution and of the stacking and resolving gels are given below:

<table>
<thead>
<tr>
<th>Resolving gel mixture (5 ml)</th>
<th>10% (ml)</th>
<th>12% (ml)</th>
<th>15% (ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$O</td>
<td>1.9</td>
<td>1.6</td>
<td>1.1</td>
</tr>
<tr>
<td>30% mix acrylamide</td>
<td>1.7</td>
<td>2.0</td>
<td>2.5</td>
</tr>
<tr>
<td>1.5 M Tris-Cl pH 8.8</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>10% (w/v) SDS</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>10% (w/v) ammonium persulphate</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>TEMED</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

73
Protein loading buffer was added to samples prior to heat denaturation by boiling for 5 minutes before loading onto the gel. Samples were run alongside an appropriate SDS molecular weight marker (Precision Plus Protein™ Dual Color Standard and Precision Plus Protein™ Unstained Standard) (BIORAD) (Hercules CA, USA). Electrophoresis was carried out at a constant current of 45 mA for 50 minutes or until the bromophenol blue dye front reached the bottom of the gel. Gels were stained with Coomassie staining solution for 30 minutes and then destained in Coomassie destaining solution until the background was colourless. Gel were soaked in 1-2% (v/v) glycerol solution for 15 minutes and photographed or dried.

Protein loading buffer (2X)

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 mM Tris-Cl pH 6.8</td>
<td>25 mM Tris</td>
</tr>
<tr>
<td>4% (w/v) SDS</td>
<td>192 mM Glycine</td>
</tr>
<tr>
<td>0.2% (w/v) Bromophenol blue</td>
<td>0.1% (w/v) SDS</td>
</tr>
<tr>
<td>20% (v/v) Glycerol</td>
<td></td>
</tr>
<tr>
<td>200 mM DTT</td>
<td></td>
</tr>
</tbody>
</table>

Coomassie staining solution

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1% (w/v) Coomassie Brilliant Blue R-250</td>
<td></td>
</tr>
<tr>
<td>50% (v/v) Methanol</td>
<td>40% (v/v) Methanol</td>
</tr>
<tr>
<td>10% (v/v) glacial acetic acid</td>
<td>10% (v/v) Glacial acetic acid</td>
</tr>
</tbody>
</table>

Coomassie destaining solution

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% (v/v) Methanol</td>
<td></td>
</tr>
<tr>
<td>10% (v/v) Glacial acetic acid</td>
<td></td>
</tr>
</tbody>
</table>

2.5.3 Western Blot analysis

Proteins were separated by SDS-PAGE as described in Section 2.5.1; the gel, the nitrocellulose membrane and 3 MM paper sheets were soaked in a transfer buffer (TG) for 1 minute. Blotting was performed using a MINI TRANS-BLOT CELL ASSEMBLY
(BIORAD) following the manufacturer’s instructions. After blotting, the membrane was washed in T-TBS and transferred to the blocking solution of 2% (w/v) dry milk in T-TBS for 2 hours at room temperature. The blocking solution was removed and the membrane incubated in a solution 1% (w/v) dry milk in T-TBS containing the primary antibody, diluted following the manufacturer's instructions, for 2 hours at room temperature. The membrane was washed twice with T-TBS for 5 minutes, transferred into a solution of T-TBS containing the secondary antibody, diluted as suggested by manufacturer's instructions, and incubated for 3 hours at room temperature. The membrane was washed in 4 times in T-TBS for 5 minutes and once quickly in TBS and then developed by colorimetric assay using 3,3’5,5’ Tetramethylbenzidine following the manufacturer’s instructions.

<table>
<thead>
<tr>
<th>TBS  (pH 7.5)</th>
<th>TG (pH 7.5)</th>
<th>T-TBS (pH 7.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 mM Tris</td>
<td>25 mM Tris</td>
<td>50 mM Tris</td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td>192 mM Glycine</td>
<td>150 mM NaCl</td>
</tr>
<tr>
<td>20% (v/v) Methanol</td>
<td></td>
<td>0.05% (v/v) Tween 20</td>
</tr>
</tbody>
</table>

2.5.4 Source of antibodies
The antibodies used in this work were obtained from Abcam Ltd (Cambridge, UK) to perform Western blots following manufacturer’s guidelines. The primary antibodies were mouse monoclonal anti-c-Myc and mouse monoclonal anti-HA tag. The secondary antibody was rabbit polyclonal to mouse IgG H&L horseradish peroxidase (HRP) conjugated anti-IgG.

2.6 RNA methods
2.6.1 Extraction of RNA from S. coelicolor
RNA was extracted from cells using the FastRNA® Pro Blue kit from Q-BIOgene (Morgan Irvine, CA, USA) according to the manufacturer’s guidelines. Briefly, 10 ml of overnight culture were collected and centrifuged at 1,500 x g for 15 minutes at 4 °C. The supernatant was discarded and the cells were suspended in 1 ml of RNApro™ solution. 1 ml of suspended cells was transferred to a blue-cap tube containing Lysing Matrix B. The sample was processed in the FastPrep™ FP 120 (Thermo Electron Corporation) for 40 seconds at a setting 6.0 and then centrifuged at 12,000 x g for 5 minutes at 4 °C. The supernatant was transferred to a new tube and 300 μl of chloroform added mixing the sample for 10 seconds by vortexing. After incubation of 5
minutes at room temperature, the sample was centrifuged at 12,000 x g for 5 minutes at 4 °C. The upper phase was recovered, transferred to a new tube and 500 µl of absolute ethanol were added. The solution was mixed by inverting 5 times and incubated at -20 °C overnight. The sample was centrifuged at 12,000 x g for a minimum of 15 minutes at 4 °C and the pellet was washed with 500 µl of cold 75% (v/v) ethanol. The alcohol was removed from the sample by air drying for 5 minutes at room temperature. RNA was dissolved in 100 µl of RNase-free water incubating it for 5 minutes at room temperature to facilitate resuspension. RNA integrity was determined by analyzing 4.5 µl of the sample by electrophoresis as described in the following section 2.6.2.

2.6.2 Electrophoresis of RNA through Formaldehyde gels
Appropriate amount of agarose was melted in water (1% w/v) and cooled to 60 °C. 5x Formaldehyde gel-running buffer and formaldehyde were added to a final concentration of 1x and 2.2 M respectively. Ethidium bromide was added to the molten agarose prior to casting of the gels at a final concentration of 0.3 µg ml⁻¹. Horizontal slab gels were cast in a chemical hood in Formaldehyde gel running buffer. The RNA sample was prepared as follow:

**RNA sample for loading**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA (up to 30 µg)</td>
<td>4.5 µl</td>
</tr>
<tr>
<td>5x Formaldehyde gel running buffer</td>
<td>2.0 µl</td>
</tr>
<tr>
<td>Formaldehyde 36.5-38% (v/v)</td>
<td>3.5 µl</td>
</tr>
<tr>
<td>Formamide 99% (v/v)</td>
<td>10.0 µl</td>
</tr>
</tbody>
</table>

The RNA sample was then incubated for 15 minutes at 65 °C and then chilled on ice. 2 µl of formaldehyde gel-loading buffer was added to RNA sample prior loading.

**5x Formaldehyde gel-running buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 M MOPS (pH 7.0)</td>
<td></td>
</tr>
<tr>
<td>40 mM sodium acetate</td>
<td></td>
</tr>
<tr>
<td>5 mM EDTA (pH 8.0)</td>
<td></td>
</tr>
</tbody>
</table>

**Formaldehyde gel-loading buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% (v/v) Glycerol</td>
<td></td>
</tr>
<tr>
<td>1 mM EDTA (pH 8.0)</td>
<td></td>
</tr>
<tr>
<td>0.25% (w/v) Bromophenol Blue</td>
<td></td>
</tr>
<tr>
<td>0.25% (w/v) Xylene cyanol FF</td>
<td></td>
</tr>
</tbody>
</table>
2.6.3 DNase treatment of RNA

Removal of any contaminating genomic or plasmid DNA from RNA samples was achieved using the DNA-free™ Kit from AMBION (Austin TX, USA) following manufacturer’s instructions. This protocol is designed to purify the RNA samples from contaminating DNA that can interfere with subsequent RT-PCR reactions. The procedure involved addition of 0.1 volumes of 10x DNase I buffer and 1 µl of DNase enzyme to the RNA sample followed by incubation at 37 °C for 30 minutes. For vigorous DNase treatment, the incubation with DNase was repeated twice adding 0.5 µl of enzyme each time. The reaction was stopped by adding 0.1 volumes of DNase Inactivation Reagent and incubating 2 minutes with occasional mixing at room temperature. The sample was centrifugated for 2 minutes at 10,000 x g and the supernatant recovered.

2.6.4 Reverse transcriptase (RT-PCR)

RT-PCR was performed using a cMaster™ RTplus PCR system (Eppendorf) according to the manufacturer’s guidelines. The kit is designed to perform high performance first strand synthesis of full length cDNA ranging from 0.1-12.5 kb. In addition the system allows reverse transcription of RNA with stable secondary structure at elevated temperatures (37-60 ºC). A standard procedure involved preparation of the Mastermix 1 solution containing the RNA template which was incubated for 5 minutes at 65 ºC and subsequently held on ice. The Mastermix 2 solution was then added to the Mastermix 1 and the sample incubated at 50 ºC for 60 minutes to synthesise full length cDNA. The reaction was stopped by heating the sample at 85 ºC for 5 minutes. PCR reactions were performed transferring 10 µl of Mastermix 3 to 40 µl of Mastermix 4 solution in an Eppendorf “Mastercycler gradient” thermal cycler following standard PCR amplification reaction programs. A minus reverse transcriptase control was included in all the PCR amplification reaction as negative control to discount DNA contamination.

<table>
<thead>
<tr>
<th>Mastermix 1 (10 µl)</th>
<th>Mastermix 2 (10 µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RNA template</strong></td>
<td><strong>RTplus PCR Buffer with 25 mM Mg^{2+}</strong></td>
</tr>
<tr>
<td>5.0 µl</td>
<td>4.0 µl</td>
</tr>
<tr>
<td>dNTPs (10 mM)</td>
<td><strong>cMaster RT Enzyme</strong></td>
</tr>
<tr>
<td>2.0 µl</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Oligo (25 mM)</td>
<td><strong>Prime RNase Inhibitor Solution</strong></td>
</tr>
<tr>
<td>2.0 µl</td>
<td>0.5 µl</td>
</tr>
<tr>
<td>DMSO</td>
<td><strong>RNase-free water</strong></td>
</tr>
<tr>
<td>0.5 µl</td>
<td>4.5 µl</td>
</tr>
<tr>
<td>RNase-free water</td>
<td></td>
</tr>
<tr>
<td>0.5 µl</td>
<td></td>
</tr>
</tbody>
</table>
Mastermix 3 (10 µl) | Mastermix 4 (40 µl)
---|---
First step RT reaction mix | RT<sub>plus</sub>PCR Buffer with 5 mM Mg<sup>2+</sup> 5.0 µl
1.0 µl | cMaster PCR Enzyme 0.4 µl
dNTPs (10 mM) | RNase-free water 34.6 µl
1.0 µl | RNase-free water 3.5 µl
Forward primer (25 mM) | DMSO 0.5 µl
1.0 µl |
Reverse primer (25 mM) | RNase-free water 3.5 µl
1.0 µl |
DMSO | RNase-free water 3.5 µl
0.5 µl |

2.7 Tap tag method

2.7.1 Growth of *S. coelicolor* cells

*S. coelicolor* strain A3(2) carrying plasmid pNC147 was grown in 50 ml of YEME medium containing thiostrepton (15 µg/ml) in 250 ml baffled flasks at 30 °C with shaking at 200 rpm for 48 hours. This seed culture was used to inoculate 500 ml of YEME medium (thiostreptone 15 µg/ml) in baffled flasks and incubated at 30 °C with shaking at 200 rpm for 20 hours.

2.7.2 Preparation of the protein extracts from *S. coelicolor* cells

The mycelium was collected, centrifugated at room temperature at 5,000 x g and washed twice with 0.09 M Tris-Cl pH 7.9. The packed cell volume (PCV) was measured and the cell stored at -80 °C. The frozen pellet was thawed on ice and a PCV volume of buffer A was added. The disruption of the mycelium was carried out by sonication using a Soniprep 150 MSE (UK) for 2 minutes, alternating 40 seconds of sonication and 40 seconds of cooling on ice. The extract was centrifuged at 25,000 x g for 30 minutes at 4 °C and the supernatant transferred to a new tube and centrifugated at 100,000 x g for 1 hour at 4 °C. The supernatant was collected and filtered with a Millipore 0.22 µm filter to eliminate any fine particle from the solution.
2.7.3 Purification of Tap tagged protein

2.7.3.1 Binding to IgG Sepharose
The composition of the protein extract was adjusted to contain 10 mM Tris-Cl pH 8, 150 mM NaCl. 200 μl of IgG sepharose beads suspension (IgG Sepharose™ 6 Fast Flow) (Amersham Bioscience, Uppsala, Sweden) was washed twice using 5 ml of IPP150. The extract was added to the washed beads and rotated for 2-4 hours at 4 °C.

2.7.3.2 Cleavage with Tobacco etch virus (TEV) protease
The IgG beads were washed three times with 10 ml of IPP150 and twice with 5 ml of tobacco etch virus (TEV) protease cleavage buffer. The washed beads were resuspended in 1 ml of TEV cleavage buffer containing 100 units of TEV protease (kind gift from Mr. Pakorn Wattana-Amorn). Cleavage was performed rotating the slurry overnight at 4 °C.

2.7.3.3 Binding to Calmodulin Sepharose
200 μl of calmoduline sepharose beads suspension (Calmodulin Affinity Resin) (Stratagene) (Cedar Creek TX, USA) were washed twice with 5 ml of IPP150 calmodulin binding buffer. IgG beads post-cleavage were centrifugated and the supernatant recovered. The beads were washed three times with 1 ml of IPP150 calmodulin binding buffer and at each step the supernatant was recovered. The supernatants deriving from the centrifugations of the IgG beads were collected in the same tube. The washed calmodulin beads were transferred to this solution and CaCl₂
was added to a final concentration of 4 mM. Binding to calmodulin sepharose was performed by rotating the slurry for 1 hour at 4 °C.

2.7.3.4 Elution from Calmodulin Sepharose
The calmodulin beads were washed three times with 10 ml of IPP150 calmodulin buffer. Elution was performed by washing the beads five times with 200 µl of IPP150 calmodulin elution buffer and, at each step, the supernatant was recovered for further analysis. Samples to be analysed by mass spectrometry were precipitated with TCA and resuspended in 100 mM of NH₄HCO₃.

**IPP150**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM Tris-Cl pH 8.0</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>0.1% (v/v) NP40</td>
<td></td>
</tr>
</tbody>
</table>

**IPP150 Calmodulin binding buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM β-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td>10 mM Tris-Cl pH 8.0</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>0.1% (v/v) NP40</td>
<td></td>
</tr>
<tr>
<td>1 mM Magnesium acetate</td>
<td></td>
</tr>
<tr>
<td>1 mM Imidazole</td>
<td></td>
</tr>
<tr>
<td>2 mM CaCl₂</td>
<td></td>
</tr>
</tbody>
</table>

**TEV protease cleavage buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM Tris-Cl pH 8.0</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>0.1% (v/v) NP40</td>
<td></td>
</tr>
<tr>
<td>0.5 mM EDTA pH 8.0</td>
<td></td>
</tr>
<tr>
<td>1 mM DTT (added prior use)</td>
<td></td>
</tr>
</tbody>
</table>

**IPP150 Calmodulin binding buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM β-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td>10 mM Tris-Cl pH 8.0</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>0.1% (v/v) NP40</td>
<td></td>
</tr>
<tr>
<td>1 mM Magnesium acetate</td>
<td></td>
</tr>
<tr>
<td>1 mM Imidazole</td>
<td></td>
</tr>
<tr>
<td>22 mM EGTA</td>
<td></td>
</tr>
</tbody>
</table>

2.8 Protein identification by mass spectrometry

2.8.1 Excision and destaining of protein bands following SDS-PAGE
Protein samples eluted by the TAP purification method were separated by SDS-PAGE as described in Section 2.5.1. Gels were stained using Coomassie staining solution for 30 minutes, and then destained in Coomassie destaining solution until the background was colourless. The gel was subsequently rinsed twice in H₂O for 5 minutes on orbital shaker. The protein bands of interest were excised out from the gel keeping the volume to a minimum by cutting as close to the band as possible. Also, a piece of gel from a
non-protein containing region was excised for use as a control. Each slice of gel was dice into smaller pieces (1 mm³) and placed in a 1.5 ml siliconized tube. 200 μl of 50 mM NH₄HCO₃/50% (v/v) Acetonitrile (ACN) solution was added to the diced piece of gel containing the protein of interest to destain. The sample was vortexed for 10 minutes, centrifuged briefly at 12,000 x g and the supernatant discarded. The previous steps were repeated twice until the all stain was removed.

2.8.2 In-gel trypsin digestion

2.8.2.1 Reduction and alkylation of proteins

100 μl of 100% ACN was added to the gel particles and incubated for 5 minutes with occasional mixing. The sample was then centrifuged briefly at 12,000 x g and the supernatant discarded. After repeating this step, the gel particles were covered with 100 μl of 10 mM DTT and incubated at 50 °C for 30 min. The sample was then centrifuged at 12,000 x g for 30 seconds and the supernatant removed. 100 μl of 100% (v/v) ACN was added to the sample to shrink the gel and remove excess of liquid. The gel particles were treated with 100 μl of 55 mM iodoacetamide and incubated for 1 hour at room temperature in the dark. The sample was then centrifuged to remove all the liquid. The particles were washed with 200 μl of 50 mM NH₄HCO₃ for 15 minutes with occasional mixing and then centrifuged to remove the excess of liquid. The sample was incubated for 5 minutes with 100 μl of 100% (v/v) ACN and the pellet recovered by brief centrifugation. Gel particles were subsequently dried by rotary evaporation for a minimum of 30 minutes at 30 °C.

2.8.2.2 Protein digestion and extraction of digested peptides

Dried gel particles were treated with 10 μl of trypsin solution (40 ng μl⁻¹) and rehydrated for 30 minutes at room temperature. 25 μl of 50 mM NH₄HCO₃ was added to the sample and incubated at 37 °C overnight. Tryptic digestion was treated with 25 μl of a 5% (v/v) formic acid in 50% (v/v) ACN and vortexed for 10 minutes. The supernatant containing peptides from the tryptic digestion was recovered by centrifugation at 12,000 x g for 1 minute and transferred to a clean siliconized tube. These steps were repeated washing the gel pieces in succession with 30 μl of 100% (v/v) ACN, 30 μl of 5% (v/v), formic acid in 50% (v/v) ACN and twice with 30 μl 100% (v/v) ACN collecting the supernatant from each step to the same siliconized tube containing the extracted peptides from the first step. The volume of the sample containing the extracted peptides
was reduced to 10-15 µl by rotary evaporation, dried and then stored at -80 °C until analysis.

2.8.3 Mass spectrometry analysis
The protein samples for subsequent mass spectrometry analysis were submitted to the School of Chemistry University of Bristol and to the Structural Chemistry Facility at the School of Pharmacy according to their guidelines.

2.9 Computer Software
Sequence analysis was performed using the BioEdit program and the sequences checked using the program BLASTn (http://www.ncbi.nlm.nih.gov/BLAST/). Multiple sequence alignment were obtained using ClustalW (http://www.ebi.ac.uk/clustalw/). Plasmids were displayed using the software BVTech Plasmid (http://www.biovisualtech.com/). DeepView-SwissPdbViewer (Guex et al., 1997) and Chimera UCSF (Pettersen et al., 2004) were used for viewing structural files. Structure images were produced using Chimera UCSF. Docking simulations were performed using Hex 4.5 program (Ritchie and Kemp, 2000).
CHAPTER 3

Investigation of protein interactions of the ‘minimal’
DNR/DXR PKS using yeast two-hybrid system

3.1 Introduction to protein interactions

The diversity of biological processes is due to the dynamic associations of cellular components including non-covalent protein-protein and protein-ligand interactions (Parrish et al., 2006). Protein interaction networks dictate formation of multiprotein complexes and influence their biological functions. For a number of metabolic pathways, several enzymes that catalyse sequential reactions are often associated non-covalently to form a multienzyme complex. Such complexes afford increased reaction rates and protect labile intermediates from decomposition by channelling intermediates directly from one active site to another (Sharon and Robinson, 2007). The challenge for the modern molecular biology is to understand the nature of these interactions including the reactive groups involved at the active sites of the enzymes, the amino acids involved in surface binding sites, the specific order in which the large protein complexes are assembled and the overall topology of the complex.

Many genetic and biochemical methods are now available to achieve this task. Classical in vitro techniques include chemical cross-linking where interacting proteins are covalently bound using a reagent that contains reactive ends to specific functional groups (primary amines, sulphydryls, etc.) on proteins. Examples of the most commonly used cross-linking reagents include N-hydroxysuccide (NHS) esters (Lomant and Fairbanks, 1976) that create stable amide and imide bonds with primary or secondary amines and maleimides that are specific for sulphydryl groups. Crosslinking has been used to investigate the architecture of macromolecular complexes, such as E. coli F1-ATPase (Bragg and Hou, 1980) and recently has been used to elucidate the structural organization of animal fatty acid synthase (Witkowski et al., 2004). However, chemical agents could induce crosslinking between proteins that normally do not interact leading to false results. Protein affinity chromatography also represents very sensitive tool to identify protein interactions. A protein covalently bound to a matrix such as Sepharose can be used to select ligand proteins that are bound and retained from a cell extract. This method has been used to detect the phage and host proteins that interact with the E. coli RNA polymerase (Ratner, 1974). The major advantage of affinity chromatography is to
allow a relatively easy purification of the interacting protein partners. Successful affinity chromatography requires concentration of pure proteins well above the K_d of the interaction to avoid interferences due to any contaminant in the preparation. The high sensitivity of this method could represent a potential limit for the high incidence of false positive. Indirect interactions might be detected since proteins could interact through a second protein that binds the test protein. In addition, proteins might interact with high specificity even if in vivo this event never occurs (Lazarides and Lindberg, 1974).

Co-immunoprecipitation is the classical method used to detect protein-protein interactions. To identify interaction partners of a known protein, an antibody specific for the target protein is used to co-precipitate any other interacting partners from a crude extract which can then be identified by mass spectrometry. Advantages of this method include the possibility to co-precipitate elaborate complexes that in vitro might be difficult to obtain due to their instability. Also, as the interactions are detected in vivo, the proteins are likely to be in their native state of post-translational modification. However, this method is less sensitive than others and potential drawbacks include precipitation of proteins that interact with the antibody and not with the test protein and, most importantly, detection of interactions that occur as consequence of lysis but do not take normally place in the cell. All the techniques cited are widely used to detect protein-protein interactions but present the disadvantage to require further downstream investigations such as mass spectrometric analysis to determine the nature of the interacting partners.

Fluorescence resonance energy transfer (FRET) is a high-resolution technique that also allows protein-protein interactions to be investigated in vivo. FRET is a distance-dependent physical process where energy transfer occurs between an excited molecular fluorophore (donor) and another fluorophore (acceptor). Interactions can be assessed measuring fluorescence signals deriving from physical proximity, typically 3-6 nm, of proteins labeled with the different fluorophores. The efficiency of FRET is dependent on the inverse sixth power of intermolecular separation (Lakowicz, 1999) making it a very sensitive tool for investigating protein interactions. The emission and absorption spectra of the two fluorophores are directly correlated to their reciprocal distance. Putative interacting proteins are normally expressed in the cell as chimeras fused to green-fluorescent proteins (GFPs) (Prasher et al., 1992; Chalfie et al., 1994) which are the most widely donor and acceptor fluorophores in FRET analysis. The use
of GFPs in FRET analysis requires sufficient separation in excitation spectrum for donor and acceptor GFP and an overlap, minimum 30%, between the emission spectrum of the donor and the absorption spectrum of the acceptor to obtain efficient energy transfer (Hanson and Kohler, 2001).

Phage display was one of the first genetic techniques developed to investigate protein interactions (Smith, 1985). The method relies on the ability of filamentous bacteriophages to display on the surface recombinant proteins that can be selected by interacting the phage with selected immobilized ligands. Libraries of fusion phages can be constructed and screened to identify proteins that bind to a specific antibody. This method is very sensitive and selective although disadvantages include the expression in a bacterial host that may preclude the correct protein folding or post-translational modifications necessary for the interactions of some mammalian proteins. In addition, since the binding assays are conducted in vitro some artificial false positive results can occur. Yeast two-hybrid system represents one of the most powerful genetic tool for detecting of protein interaction in vivo (Fields and Song, 1989) and was extensively used during this project.

3.2 Introduction to the yeast two-hybrid system

3.2.1 Yeast transcription factors

Yeast cells grown in a medium containing galactose as the sole carbon source require expression of special set of genes (GAL) to metabolize this sugar (Johnston, 1987). The induction of these genes requires a transcriptional activator protein called GAL4 that binds to specific enhancer-like sequences, which in yeast are called upstream activation sequences (UASs) located near the promoter (Johnston, 1987; Giniger et al., 1985). Transcriptional activator proteins have the ability to bind the DNA at specific sequences allowing interaction with the transcription machinery to activate gene expression. Two different domains are essential for the functions of the transcriptional activator: a DNA-binding domain (DBD) and a transcriptional activator domain (TAD). GAL4 DBD is located at the N-terminal of the protein and binds to the UAS (Giniger et al., 1985) whereas at the C-terminal sequence of the protein there are two TADs (TAD1 and TAD2) (Ma and Ptashne, 1987a) that direct RNA polymerase II to transcribe the gene downstream of the UAS (Fig.3.1A). A study in 1985 by Brent and Pashne demonstrated that the DBD and the TAD are physically separable and functionally independent domains. The authors constructed a hybrid protein LexA-GAL4 that contained the DBD
Figure 3.1 Properties of yeast transcriptional activator GAL4
A) GAL4 activates transcription of the \textit{gal} genes by binding to the upstream activation sequences (UAS) located upstream the minimal promoter with the DNA-binding domain (DBD). The two transcriptional activation domains are represented with TAD and interact with the RNA polymerase.
B) The hybrid protein LexA-GAL4c contains the LexA DBD and the C-terminal of GAL4 (GAL4c) including the TAD. The hybrid protein is able to activate transcription of the genes located downstream the LexA operator (LexA OP) suggesting that the two domains are independent and physically separable.
C) An hybrid protein containing the TAD B42 fused to the yeast repressor GAL80 is brought to the DNA by the GAL4 derivative protein (GAL4d) that contains the DBD and the domain that interacts with GAL80. This demonstrated that the activator’s two essential domains DBD and TAD can reside on separate molecules and be artificially brought in physical proximity by protein interactions.
of the bacterial protein LexA (Ebina et al., 1983) fused to the C-terminal portion of GAL4 (Brent and Ptashne, 1985). This hybrid protein was able to activate expression of a reporter gene in yeast cells if the LexA DNA binding sites were located upstream the gene promoter (Fig. 3.1B). In 1988 Ma and Pashne demonstrated that the TAD and the DBD do not need to be covalently bound to activate transcription (Ma and Ptashne, 1988). In this study, a hybrid protein GAL80-B42 was constructed, that contained as TAD the E. coli acidic peptide B42 (Ma and Ptashne, 1987b) fused to a yeast repressor protein that interacts with GAL4, GAL80. They showed that GAL4-B42, a protein which is not able to bind to DNA, activated transcription in yeast only if the cell also carried a copy of GAL4 containing the DBD and able to interact with GAL80 (Fig. 3.1C). This result reported for the first time that is possibly to artificially bring together, through protein interaction, the TAD and the DBD of a transcriptional activator located on separate proteins.

3.2.2 Yeast two-hybrid system
Eukaryotic transcription factors are trans-acting protein composed of physically separable and functionally independent domains. In 1989 Fields and Song tested the possibility to exploit the modular characteristics of transcription factors to detect protein- protein interactions. They constructed two-hybrid proteins where the DBD and the TAD of GAL4 were fused respectively to the serine-treonine specific kinases, SNF1 and SNF4, two known interacting proteins in yeast (Celenza et al., 1989). They demonstrated that the two proteins could activate transcription of the GAL genes only if present together in the yeast cell. Activation of the GAL genes was also observed by replacing SNF1 and SNF4 with any other pair of interacting proteins or protein domains (Fields and Song, 1989). These results marked the birthday of the modern yeast two-hybrid system as a generic tool to investigate protein interactions. More recently another yeast two-hybrid system has been developed (Golemis et al., 1996) where the GAL4 DBD is substituted by the prokaryotic repressor LexA (Ebina et al., 1983) and the TAD by the B42 peptide from E. coli. In this system the binding sequences of LexA are located upstream the ‘minimal’ promoter of the reporter gene.

In the GAL4 yeast two-hybrid system developed by Fields and Songs in 1989, the DBD and the TAD derived respectively from the amino- and carboxy- terminal domains of the protein (a.a. 1-147 and 768-881) whereas GAL4 responsive elements represented
the DBD binding sites. Two different cloning vectors are used to produce separate fusions of the DBD and TAD to genes that encode for proteins being tested for possible interactions. The protein fused to the DBD is called the bait while the prey is the protein fused to the TAD. The hybrid proteins are co-expressed in yeast targeted to the yeast nucleus. If the proteins fused to the GAL4 domains interact with each other, they will bring into physical proximity the GAL4 DBD and the TAD activating the transcription of the downstream reporter gene (Fig. 3.2A). If no interaction takes place in the yeast nucleus, no expression of the reporter gene will occur (Fig. 3.2B).

3.2.3 Reporter genes and yeast strain
Most reporter genes are under the control of an artificial promoter that contains a TATA sequence and an appropriate UAS sequence to provide a binding site for the DBD. LacZ represent the most common used reporter gene because the β-galactosidase activity can be detected using a variety of assays. LacZ assays performed in liquid cultures also provide quantitative data that could be used to determine the strength of an interaction. Most studies use a second reporter gene to confirm an interaction and more importantly to eliminate initial positives that could be revealed as false using other techniques. False positives could be frequent when a large number of possible interactions are screened. In such cases, it is necessary to use two reporter genes under the control of two different promoters, one of which should be a nutritional reporter gene such as HIS3 or ADE2 for the GAL4 system or URA3 for the LexA systems. The classical GAL4 two-hybrid system utilizes two genomic integrated nutritional reporter genes (HIS3 and LacZ) under the control of a GAL4-UAS binding sequence. The yeast strain used in the GAL4 system must have gal4 and gal80 genes deleted in order to prevent interference of the endogenous proteins with the regulatory elements of the yeast two-hybrid system (Johnston et al., 1994).

The system developed by Brent based on LexA (1993) used an integrated LEU2 as nutritional marker and LacZ gene on a plasmid with both the reporter genes under the control of multiple LexA operators (Gyuris et al., 1993). The presence of LacZ on a high copy plasmid allowed β-galactosidase assays to be performed directly on the agar plates supplemented with X-gal directly in the growth medium. However, such multiply copies of LacZ can lead to a high background of false positives. Yeast strains such as EGY48 (Gyuris et al., 1993) possess a genome integrated copy of LacZ reporter gene to solve this problem.
Figure 3.2 GAL4 Yeast two hybrid system
A) Interacting proteins will bring together the GAL4 DBD and TAD activating transcription of the reporter gene located downstream the GAL4 UAS and minimal promoter.
B) Non-interacting proteins will be unable to closely associate the GAL4 DBD and TAD failing to activate transcription of the reporter gene located downstream the GAL4 UAS and minimal promoter.
Another characteristic of the LexA based system is that the fusion protein can be express under the control of galactose responsive elements such as GAL UAS. Inducible expression is an advantage if the hybrid proteins are toxic to the yeast cells, and also yeast cells do not have to be \textit{gal}4 or \textit{gal}80 deleted mutants. Generally the multiple copies of LexA operators used to control gene reporter expression make the LexA system more sensitive than the system based on GAL4. However, the standard GAL4 system has been used more extensively. It is difficult to predict which system is the best to test for interactions and it is possible that different results could be obtained using the two systems.

3.2.4 MATCHMAKER™ GAL4 yeast two-hybrid system 3 (Clontech)

Since the experiment of Fields and Song in 1989, reporting for the first time the use of a yeast two-hybrid system as a powerful technique for detecting \textit{in vivo} protein interactions, much progress has been made to overcome some of the limitations of the original method. The main problem of the yeast two-hybrid system is the incidence of false positive results. Major improvements have been made to sensibly reduce this problem especially regarding the plasmid vectors for the construction of the hybrid proteins and the yeast strain used as host. In this thesis, the commercially available kit based on GAL4 from Clontech MATCHMAKER™ yeast two-hybrid system 3 was used for investigation of protein interactions.

3.2.4.1 Yeast strain AH109

The particular yeast strain AH109 present in MATCHMAKER™ yeast two-hybrid system 3 provides both a high level of sensitivity and an extremely low background of false positives (James \textit{et al.}, 1996). False positives are clones where autoactivation of the reporter genes occur without any specific interaction between the 'bait' and the 'prey'. Autoactivation is considered to be promoter specific (Bartel \textit{et al.}, 1993a; Bartel \textit{et al.}, 1993b) and different reporter genes controlled by the same promoter will be more prone to lead to false positives results. AH109 contains two reporter genes under the control of two different inducible promoters. The strain contains other two reporter genes that can be use alternatively and are under the control of a different inducible promoter. These special genetic arrangements improve the ability to discriminate between false positives and real interactions. The genotype of strain AH109 is illustrated in section 2.2.1.
HIS3 and ADE2 are nutritional markers under the control of GAL1 and GAL2 promoters respectively (Fig. 3.3). The presence of these two reporter genes allows the stringency of the selection to be modulated and thus gives an indication of the strength of the interaction. LacZ and MEL1 genes are the other two reporter genes integrated in the yeast genome under the control of same promoter, P_{MEL1} (Fig. 3.3). MEL1 is an endogenous gene encoding for α-galactosidase which is secreted and can be assayed directly on agar plate containing X-α-Gal using white/blue selection. Both α- and β-galactosidase activity can be used to detect transient protein-protein interactions that would be too weak to support the expression of the nutritional reporter genes ADE2 and HIS3 in a medium lacking both adenine and histidine.

3.2.4.2 GAL4 expression vectors

The plasmid vectors used in MATCHMAKER™ yeast two-hybrid system 3 to express the test proteins in yeast are designated pGBK7T for the DBD and pGADT7 for the TAD fusion proteins (Fig. 3.4 and 3.5). These vectors are designed for a high level protein expression and allow easy detection of the hybrid proteins due to the presence of different epitope tags, c-Myc at the N-terminal of the ‘bait’ and Hemagglutinin (Ha) at the C-terminal of the ‘prey’.

High level expression is provided by the constitutive alcohol dehydrogenase (ADHI) promoter (P_{ADHI}) and nuclear localization sequences (NLSs) that target the hybrid protein to the yeast nucleus. Mutiple cloning sites (MCSs) have been specially designed to allow sub-cloning of genes between the two plasmids. pGBK7T and pGADT7 have different bacterial antibiotic resistance markers, Kan^r and Amp^r respectively, for selection during sub-cloning in E. coli. Cotransformation of both plasmids in the yeast can be monitored using the different nutritional markers, LEU2 encoded by pGADT7 and TRP1 encoded by pGBK7T. Yeast strains carrying pGBK7T display an increased transformation efficiency compared to strains containing other DBD vectors (Louvet et al., 1997) and this characteristic facilitates the introduction of TAD fusions.
Figure 3.3 Reporter genes in *Saccharomyces cerevisiae* AH109

*HIS3* and *ADE2* reporter genes are under the control of the *GAL1* and *GAL2* promoter elements respectively. *LacZ* and *MEL1* instead are under the control of the same *MEL1* promoter elements. *GAL1*, *GAL2* and *MEL1 UAS* are recognised by *GAL4*. 
**Figure 3.4 pGBKT7 plasmid**

This plasmid vector was used to generate ‘bait’ fusions. Important features of the polylinker are shown: T7 and 3’DNA-BD sequencing primers were used for DNA sequencing, c-Myc epitope to detect hybrid protein in immunoblotting assays, _Ndel_ and _EcoRI_ recognition sites for cloning of the _dps_ genes. The MCS also displays three stop codons for the termination of the translation of the GAL4 DBD hybrid proteins.
Figure 3.5 pGADT7 plasmid
This plasmid vector was used to generate ‘prey’ fusions. Important features of the polylinker are shown: T7 and 3’AD sequencing primers were for DNA sequencing, Ha epitope to detect hybrid proteins in immunoblotting assays, Ndel and EcoRI recognition sites for cloning of the $dps$ genes. The MCS also displays three stop codons for the termination of the translation of the TAD GAL4 hybrid proteins.
MATCHMAKER™ yeast two-hybrid system 3 includes controls to be performed in parallel with the ‘bait’ and ‘prey’ interaction assays. Positive controls are performed using the plasmids pTD1-1 and pVA3-1 (Fig. 3.6) which encode the TAD fused to the SV40 large T-antigen and the DBD fused to the murine p53 protein respectively. Murine p53 protein and SV40 large T-antigen are two proteins known to interact in yeast two-hybrid experiments (Li and Fields, 1993). Plasmid pCL1 contains a wild type full length GAL4 protein and is used as positive control of the transformation. Negative controls are performed by co-transforming pGADT7 and pGBKKT7 empty vectors into yeast.

3.2.5 Sensitivity and limitations of the yeast two-hybrid system
Since its introduction in 1989, the yeast two-hybrid system has been used and considered an invaluable tool for the investigation of protein interactions. It provides a sensitive method for detecting both strong and relatively weak and transient interactions, often the most interesting in signaling cascade or multienzymatic reactions that could not be detected using classical biochemical assays. Compared to the other standard biochemical methods for detecting protein interactions, the yeast two-hybrid system has the peculiar characteristic that the assays are conducted in vivo thus proteins are more likely to be in their native conformations. Another advantage of the technique is that the interactions are detected in the yeast nucleus without requiring biochemical purification. The intrinsic sensitivity of the two-hybrid system is presumably attributable to the amplification of positive signals in vivo (i.e. transcriptional, translational and enzymatic). Furthermore, the sensitivity of the classical yeast two-hybrid system has been improved by developing new yeast strain and expression vectors.

The high sensitivity of the assay has allowed its use for a plethora of applications: from investigation of interactions among the subunits of multiprotein complexes (Lalo et al., 1993; Brown et al., 1994) to map specific domains within proteins responsible for an interaction (Van Aelst et al., 1993; Kalpana and Goff, 1993; Takacs et al., 1993). Most significantly the yeast two-hybrid system has been used for identification of new interacting proteins by screening expression libraries (Hannon et al., 1993; Vojtek et al., 1993).
Figure 3.6 pTD1-1 and pVA3-1 plasmids
These plasmids were used to perform positive controls. pTD1-1 encodes for SV40 T-Large antigen whereas pVA3-1 encodes murine p53. SV40 T-Large antigen and murine p53 interact in yeast nucleus.
In addition, point mutations can be assayed to identify specific amino acid residues critical for the interaction (Li and Fields, 1993). Binding constant determined by surface plasmon resonance have been compared to the results from the two-hybrid system suggesting that dissociation constant above 70 μM can be detected (Yang et al., 1995).

Potential drawback of the high sensitivity of the assay is the incidence of false positive results that can occur in particular when large libraries are screened. Yeast strains specially engineered have been developed to reduce the occurring of these events. Main problem of the yeast two-hybrid system is the potential transcriptional activation effect that some proteins might show when present in the yeast. Also, this method is limited to proteins that are able to fold properly, exist stably and retain activity as fusion protein in the yeast nucleus. The conditions in the yeast cells may not allow post-translational modifications such as glycosylation, required for interaction of some especially mammalian proteins (Fields and Sternglanz, 1994). Limitations regard proteins that are not normally located in the nucleus; however interactions involving extracellular proteins have been detected (Ozenberger and Young, 1995; Kuo et al., 1997). In some cases the DBD or the TAD domain may occlude the normal site of interaction or interfere with the native folding of the hybrid protein impairing the interaction (Li and Fields, 1993; Van Aelst et al., 1993). Many proteins, including those not normally involved in transcription, may activate transcription when fused to the DBD interfering with the assay (Van Aelst et al., 1993). However, it is often possible to delete portions of the ‘bait’ that activates transcription while retaining other function of the protein (Bartel et al., 1993b).

3.2.6 Type II PKS and DNR/DXR PKS

DNR/DXR PKS is responsible for the production of the 21-carbon atom polyketide aklanonic acid that is the first intermediate in the biosynthetic pathway of the two anticancer drugs (Grimm et al., 1994). The ‘minimal’ DNR/DXR PKS includes the subunits of the KS, DpsA and DpsB, the type III KS, DpsC, the ACP, DpsG, and the putative MAT, DpsD. In this chapter, possible protein interactions involving these proteins were investigated using the Clontech MATCHMAKER™ yeast two-hybrid system 3.
3.3 Results

3.3.1 PCR amplification of dps genes

The genetic source for the dps genes was the plasmid pWHM1012 (Bao et al., 1999b) obtained as a gift from Prof. Ben Shen, University of Wisconsin-Madison, USA. This plasmid contains clustered together dpsABCDGEFY genes and has been already used for heterologous expression of Dps proteins and for the production of aklanonic acid in Streptomyces lividans (Bao et al., 1999b). PCR amplification of dps genes from the plasmid template was performed using the high fidelity Pfu DNA polymerase with specific oligonucleotide primers specially designed for convenient and easy cloning into the plasmid vectors pGBKT7 and pGADT7 (Table 3.1). Each forward primer carried an NdeI restriction site just upstream the translational start codon while the reverse primers contained an EcoRI restriction site located downstream the dps gene stop codon. The recognition sequences for the endonuclease restriction enzymes NdeI and EcoRI were unique to the polylinker sites of each plasmid, pGBKT7 and pGADT7, but not present in any of the dps gene sequences. The size of the expected amplified product, the sequence of the oligonucleotide primers and the PCR amplification conditions for the different dps genes are listed in Table 3.1.

3.3.2 Cloning strategies for the dps genes

Two different cloning strategies were used to achieve cloning of the dps genes in both the yeast two-hybrid system plasmid vectors pGBKT7 and pGADT7. A single step strategy (Fig. 3.7) involved direct digestion of the amplified PCR products using the restriction endonucleases NdeI and EcoRI following addition of extrabases at the 5' end of both forward and reverse oligonucleotide primers to allow the enzyme recognition to occur. The number of the extra bases added upstream of the restriction site was dependent on the endonuclease: NdeI requires at least 7 nucleotides flanking the recognition sequence while EcoRI needs at least 2. An alternative double step strategy was also used to clone some of the dps genes that were not directly digested after the PCR amplification but cloned in the commercially available cloning vector pET100/D-TOPO® (Fig. 3.8). This vector is designed to facilitated rapid directional cloning of blunt end PCR products by adding four bases to the forward primers (CACC).
<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5' to 3'</th>
<th>Recognition sites</th>
<th>Annealing (°C)</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>dpsA</em>-forward</td>
<td>caccatatgaacgcgggctcctcacc</td>
<td>NdeI</td>
<td>58</td>
<td>1259</td>
</tr>
<tr>
<td><em>dpsA</em>-reverse</td>
<td>gaattgcctacccagagacgggccttcacc</td>
<td>EcoRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>dpsB</em>-forward</td>
<td>caccatatgacaggtagccgggctcctcacc</td>
<td>NdeI</td>
<td>60</td>
<td>1278</td>
</tr>
<tr>
<td><em>dpsB</em>-reverse</td>
<td>gaattgcctacccagagacgggctcctcacc</td>
<td>EcoRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>dpsG</em>-forward</td>
<td>caccatatgacaggtagccgggctcctcacc</td>
<td>NdeI</td>
<td>58</td>
<td>255</td>
</tr>
<tr>
<td><em>dpsG</em>-reverse</td>
<td>gaattgcctacccagagacgggctcctcacc</td>
<td>EcoRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>dpsC</em>-forward</td>
<td>caccatatgacaggtagccgggctcctcacc</td>
<td>NdeI</td>
<td>56</td>
<td>1059</td>
</tr>
<tr>
<td><em>dpsC</em>-reverse</td>
<td>gaattgcctacccagagacgggctcctcacc</td>
<td>EcoRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>dpsD</em>-forward</td>
<td>caccatatgacaggtagccgggctcctcacc</td>
<td>NdeI</td>
<td>62</td>
<td>1013</td>
</tr>
<tr>
<td><em>dpsD</em>-reverse</td>
<td>gaattgcctacccagagacgggctcctcacc</td>
<td>EcoRI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.1** Amplification of the *dps* genes
Oligonucleotide primers used for the amplification of *dps* genes by PCR; the recognition sequence for restriction enzymes is underlined. Annealing temperature and the size of the expected amplification product for each *dps* gene is indicated.
Figure 3.7 Single step cloning strategy
pGBKT7 derivative plasmids were obtained by ligation of DNA fragments derived from NdeI/EcoRI double digestion of the amplified dps gene and pGBKT7.
Figure 3.8 Double step cloning strategy
Cloning of the *dps* gene amplified by PCR into pET100/D-TOPO®. Topoisomerase catalyses the insertion in the plasmid vector of amplified product containing the sequence CACC.
The *dps* gene fragments generated by the *NdeI/EcoRI* double digestion from both strategies were cloned into pGBK7 (Fig. 3.7 and 3.9). Once cloned in pGBK7, the *dps* genes were subsequently sub-cloned to the pGADT7 vector to generate the corresponding TAD fusions. Successful ligation products were checked by PCR amplification and restriction digestion patterns using suitable endonucleases to verify the insertion of the *dps* gene (Fig. 3.10). Strategies used and plasmids generated during this work are described in Table 3.2.

### 3.3.3 DNA sequencing
The essential requirement of fully functionally hybrid proteins to perform the assays needed DNA sequencing of the in the linking junction between the coding sequence of the GAL4 domains and the *dps* genes proteins. Also, because the *dps* genes were amplified by PCR, although a high fidelity DNA polymerase was used to perform the amplification reactions, their entire coding sequences were checked to verify the absence of mutations that could alter the primary structure of the proteins. Analysis of the joining sequence between the GAL4 domains and the *dps* genes, for all the plasmids submitted, revealed that the two coding sequences were in the same translational frame. Sequencing results of the *dps* genes did not show any relevant mutation.

### 3.3.4 Expression of recombinant proteins
Expression of the hybrid proteins in the yeast cells was verified before starting the yeast two-hybrid system assays. This was because *dps* genes, like any other Streptomyces genes, displays an average of 74% G+C content determining the occurrence of rare codon in other organisms with considering lower G+C content such as *E. coli* or *S. cerevisiae* (Andersson and Kurland, 1990). Under conditions of high level expression, the limited availability of charged cognate tRNA for rare codons might limit protein synthesis (Robinson *et al.*, 1984; Makoff *et al.*, 1989). Thus, detection of the recombinant proteins represented a critical control before proceeding with the interaction assays.
Figure 3.9 Double step cloning strategy
pGBKT7 derivative plasmids obtained by ligation of DNA fragments derived from Ndel/EcoRI double digestion of pET100/D-TOPQ® derivative plasmids carrying dps gene and pGBKT7.
Figure 3.10 *NdeI/EcoRI* digestion of the pGBK7 and pGAD7 derivative plasmids carrying *dps* genes (A and B) with the excision of the *dps* gene insert (arrowhead)

A)  
M: 1 kb ladder  
1: pNC107 (*dpsA*)  
2: pNC108 (*dpsB*)  
3: pNC109 (*dpsC*)  
4: pNC110 (*dpsD*)  
5: pNC111 (*dpsG*)

B)  
M: 1 kb ladder  
6: pNC102 (*dpsA*)  
7: pNC106 (*dpsB*)  
8: pNC104 (*dpsC*)  
9: pNC105 (*dpsD*)  
10: pNC103 (*dpsG*)
<table>
<thead>
<tr>
<th>Gene</th>
<th>Cloning strategy</th>
<th>Plasmid (features/backbone)</th>
</tr>
</thead>
</table>
| \(dpsA\) | Double step | pNC99  
\(dpsA\) \(//\) pET100/D-TOPO\(^\text{®}\) |
|      |                  | pNC107  
\(DBD::dpsA\) \(//\) pGBK7 |
|      |                  | pNC102  
\(TAD::dpsA\) \(//\) pGADT7 |
| \(dpsB\) | Double step | pNC100  
\(dpsB\) \(//\) pET100/D-TOPO\(^\text{®}\) |
|      |                  | pNC108  
\(DBD::dpsB\) \(//\) pGBK7 |
|      |                  | pNC106  
\(TAD::dpsB\) \(//\) pGADT7 |
| \(dpsC\) | Single step | pNC109  
\(DBD::dpsC\) \(//\) pGBK7 |
|      |                  | pNC104  
\(TAD::dpsC\) \(//\) pGADT7 |
| \(dpsD\) | Single step | pNC110  
\(DBD::dpsD\) \(//\) pGBK7 |
|      |                  | pNC105  
\(TAD::dpsD\) \(//\) pGADT7 |
| \(dpsG\) | Double step | pNC101  
\(dpsG\) \(//\) pET100/D-TOPO\(^\text{®}\) |
|      |                  | pNC111  
\(DBD::dpsG\) \(//\) pGBK7 |
|      |                  | pNC103  
\(TAD::dpsG\) \(//\) pGADT7 |

**Table 3.2** Cloning strategies and features of the plasmids generated during this study
Plasmids carrying DBD and TAD fusions were independently transformed into the yeast strain AH109 and protein extracts were analysed by immunoblotting assays. ‘bait’ fusions encoded by the plasmids deriving from a pGBK7 backbone, were detected using antibodies targeted to the c-Myc epitope while ‘prey’ fusions that were encoded by the plasmids deriving from pGADT7 backbone, were detected using anti-Ha antibodies. Protein extract from untransformed yeast strain AH109 was used as negative control. Immunoblotting assays verified that all the plasmids for the ‘bait’ and ‘prey’ encoded full-length hybrid proteins in yeast (Fig. 3.11).

3.3.5 Yeast two-hybrid system assay

The matrix in Figure 3.12 shows all possible combinations of ‘bait’ and ‘prey’ used to test protein interactions. Heterotypic interactions could also be detected using two different combinations of ‘bait’ and ‘prey’ e.g. interaction between DpsA and DpsB could be tested using the combinations DBD-DpsA/TAD-DpsB and DBD-DpsB/TAD-DpsA. However, homotypic interactions could only be tested using only one combination of ‘bait’ and ‘prey’. Yeast cells able to grow on the -2 drop-out medium which lacks of tryptophan and leucine, contained the plasmids for the ‘bait’ and the ‘prey’ that carry the nutritional markers TRP1 and LEU2 respectively. The number of colonies obtained indicated the efficiency of the transformation and represented a control of the reaction. The average of number of colonies for each transformation was 150 cfu μg⁻¹ DNA of each plasmid.

Transformants able to grow on the -4 drop-out medium were submitted to a very high stringency condition due to the absence of leucine, tryptophan but more importantly of adenine and histidine. Growth of transformants on the -4 drop-out medium indicated that the ‘bait’ and ‘prey’ interacted with each other in yeast nucleus allowing expression of the ADE2 and HIS3 reporter genes required for growth in absence of adenine and histidine. These interactions were classified as strong interactions. Transformants carrying ‘bait’ and ‘prey’ that did not support the expression ADE2 and HIS3 reporter genes and consequently failed to show any growth on the -4 drop-out medium, were tested for β-galactosidase activity. If the assay resulted positive, the interaction between ‘bait’ and ‘prey’ was classified as weak interaction.
Figure 3.11 Western blot assays performed on protein extract of yeast transformants carrying hybrid Dps proteins fused to DBD (A) and to TAD (B) indicated by arrowheads.

A

- M: Precision Plus™ Dual Color
- 1: DpsA-DBD (60.6 kDa)
- 2: DpsB-DBD (60.3 kDa)
- 3: DpsC-DBD (54.1 kDa)
- 4: DpsD-DBD (51.8 kDa)
- 5: DpsG-DBD (26.3 kDa)
- 6: DBD (16.9 kDa)
- C: AH109 protein extract

B

- M: Precision Plus™ Dual Color
- 7: DpsA-TAD (60.5 kDa)
- 8: DpsB-TAD (60.2 kDa)
- 9: DpsC-TAD (54 kDa)
- 10: DpsD-TAD (51.7 kDa)
- 11: DpsG-TAD (26.2 kDa)
- 12: TAD (16.8 kDa)
- C: AH109 protein extract
<table>
<thead>
<tr>
<th>‘bait’</th>
<th>A (KSα)</th>
<th>B (KSβ)</th>
<th>C (KSIII)</th>
<th>D (MAT)</th>
<th>G (ACP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘prey’</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>A (KSα)</td>
<td>AA</td>
<td>AB</td>
<td>AC</td>
<td>AD</td>
<td>AG</td>
</tr>
<tr>
<td>B (KSβ)</td>
<td>BA</td>
<td>BB</td>
<td>BC</td>
<td>BD</td>
<td>BG</td>
</tr>
<tr>
<td>C (KSIII)</td>
<td>CA</td>
<td>CB</td>
<td>CC</td>
<td>CD</td>
<td>CG</td>
</tr>
<tr>
<td>D (MAT)</td>
<td>DA</td>
<td>DB</td>
<td>DC</td>
<td>DD</td>
<td>DG</td>
</tr>
<tr>
<td>G (ACP)</td>
<td>GA</td>
<td>GB</td>
<td>GC</td>
<td>GD</td>
<td>GG</td>
</tr>
</tbody>
</table>

Figure 3.12 Matrix of possible protein interactions of the ‘minimal’ DNR/DXR PKS
++ indicates strong interactions revealed by nutritional selection
+ indicates weak interactions revealed by β-galactosidase assay
Different positive and negative controls were set up in parallel with the interaction assays in order to decrease the incidence of false results. Firstly competent yeast cells used for the transformation reactions were plated both on a -2 and -4 drop-out medium to verify that were unable to grow. Double negative controls were performed by co-transforming the empty vectors pGADT7 and pGBK7T7. Co-transformation of the ‘bait’ fusion together with empty pGADT7 vector and of the ‘prey’ fusion together with the empty pGBK7T7 vector represented single negative controls to verify absence of auto-activation of the reporter genes. Positive controls were performed by co-transforming plasmids pVA3-l and pTD1-l encoding for murine p53 protein and SV40 large T-antigen respectively, known to interact in the yeast two-hybrid system assay. Each interaction assay was tested independently three times along with all the controls.

3.3.5.1 Nutritional selection

Co-transformants carrying the α-subunit of the ketosynthase, DpsA, as ‘bait’ and ‘prey’, (DBD-DpsA/TAD-DpsA) were able to grow on -4 drop-out medium indicating strong homotypic interactions between DpsA polypeptides. DpsA was found to interact strongly with the β-subunit of the ketosynthase, DpsB. Interestingly this interaction could be detected only using the particular combination DBD-DpsB/TAD-DpsA (pNC108/pNC102) but not testing the equivalent DBD-DpsA/TAD-DpsB (pNC107/pNC106). Also, a strong interaction was also revealed by nutritional selection between DpsD and DpsA co-transforming pNC105 (TAD-DpsD) and pNC107 (DBD-DpsA). Again, the equivalent combination tested in reverse using pNC110 (DBD-DpsD) and pNC102 (TAD-DpsA) failed to support expression of the nutritional reporter genes. Homo and heterotypic strong interactions involving the other proteins of the ‘minimal’ DNR/DXR PKS as either ‘bait’ or ‘prey’ could not be detected. All the interactions were performed in parallel with the controls; single and double controls resulted in growth on -2 but not on -4 drop-out medium while the positive control, using plasmids pVA3-1 and pTD1-1, was positive for both media. Results obtained by the nutritional selection are summarized in Figure 3.12.

3.3.5.2 β-galactosidase activity

Expression of the reporter gene LacZ was monitored to detect possible weak protein interactions testing the combinations shown in Figure 3.12. β-galactosidase activity was observed when the ACP, DpsG, was used as ‘bait’ in combination with the other proteins DpsA, DpsB, DpsC and DpsD used as ‘prey’ (Fig. 3.13 and 3.14).
Figure 3.13  *LacZ* assays performed on yeast transformants carrying the indicated combination of prey/bait (TAD-Dps/DBD-Dps) blue colonies indicates positive

A)  DpsA ‘prey’
C+: p53/SV40 Large T antigen
1: TAD-DpsA/DBD-DpsA
2: TAD/DBD-DpsA
3: TAD-DpsA/DBD-DpsB
4: TAD/DBD-DpsB
5: TAD-DpsA/DBD-DpsC
6: TAD/DBD-DpsC
7: TAD-DpsA/DBD-DpsD
8: TAD/DBD-DpsD
9: TAD-DpsA/DBD-DpsG
10: TAD/DBD-DpsG
11: TAD-DpsA/DBD

B)  DpsB ‘prey’
C+: p53/ SV40 Large T antigen
12: TAD-DpsB/DBD-DpsA
13: TAD/DBD-DpsA
14: TAD-DpsB/DBD-DpsB
15: TAD/DBD-DpsB
16: TAD-DpsB/DBD-DpsC
17: TAD/DBD-DpsC
18: TAD-DpsB/DBD-DpsD
19: TAD/DBD-DpsD
20: TAD-DpsB/DBD-DpsG
21: TAD/DBD-DpsG
22: TAD-DpsB/DBD

110
Figure 3.14 LacZ assays performed on yeast transformants carrying the indicated combination of prey/bait (TAD-Dps/DBD-Dps) blue colonies indicates positive
A) DpsC 'prey'
C+: p53/ SV40 Large T antigen
1: TAD-DpsC/DBD-DpsA
2: TAD /DBD-DpsA
3: TAD-DpsC/DBD-DpsB
4: TAD /DBD-DpsB
5: TAD-DpsC/DBD-DpsC
6: TAD /DBD-DpsC
7: TAD-DpsC/DBD-DpsD
8: TAD /DBD-DpsD
9: TAD-DpsC/DBD-DpsG
10: TAD /DBD-DpsG
11: TAD-DpsC/DBD
C-: TAD /DBD

B) DpsD 'prey'
C+: p53/ SV40 Large T antigen
12: TAD-DpsD/DBD-DpsA
13: TAD /DBD-DpsA
14: TAD-DpsD/DBD-DpsB
15: TAD /DBD-DpsB
16: TAD-DpsD/DBD-DpsC
17: TAD /DBD-DpsC
18: TAD-DpsD/DBD-DpsD
19: TAD /DBD-DpsD
20: TAD-DpsD/DBD-DpsG
21: TAD /DBD-DpsG
22: TAD-DpsD/DBD
C-: TAD /DBD
However, all these weak heterotypic interactions could not be detected when DpsG was tested in reverse as ‘prey’ using pNC103 (TAD-DpsG) (Fig. 3.15). Co-transformants carrying pNC111 and pNC103, that express DpsG as ‘bait’ and ‘prey’ respectively, did not show activation of LacZ reporter gene (Fig. 3.15). Other homotypic interactions involving DpsB, DpsC and DpsD polypeptides were not observed using this assay (Fig. 3.13B and 3.14). DpsC did not show interactions with any of the other Dps protein except with DpsG although only when it was tested as ‘prey’ (Fig. 3.14A and 3.15). Weak interactions were observed between the putative MAT, DpsD and the β-subunit of the ketosynthase, DpsB. For this particular interaction equivalent combinations of ‘bait’ and ‘prey’ showed the same result (Fig. 3.13B and 3.14B).

All the relative controls gave the expected results for β-galactosidase activity: co-transformation of pVA3-l and pTD1-l showed lacZ expression while both double and single negative controls failed. β-galactosidase activity also confirmed the protein interactions detected by nutritional selection.

3.4 Discussion

The ‘minimal’ DNR/DXR PKS has been intensively analysed during these experiments using the yeast two-hybrid system assays. dpsABCDG genes encoding for proteins forming the ‘minimal’ DNR/DXR PKS were amplified from pWHM1012 plasmid and independently cloned in pGBKT7 and pGADT7 vectors to express hybrid Dps proteins fused to GAL4 DBD and TAD domains respectively. The expression plasmids were used to transform Saccharomyces cerevisiae AH109 strain to investigate putative interactions testing all the possible combinations (Fig. 3.12). All the combinations for each interaction were independently performed three times along with all the relative negative and positive controls.

Strong interactions were assessed by nutritional selection using ADE2 and HIS3 reporter genes on -4 drop-out medium. The results from these assay, summarised in the Figure 3.12, suggested interaction between the two subunits of the KS, DpsA and DpsB. This interaction was expected since it has been already described for the tcm and act KSα/KSβ heterodimers (Gramajo et al., 1991; Carreras and Khosla, 1998).
Figure 3.15 *LacZ* assays performed on yeast transformants carrying the indicated combination of prey/bait (TAD-Dps/DBD-Dps) blue colonies indicates positive DpsG 'prey'.

C+: p53/ SV40 Large T antigen
1: TAD-DpsG/DBD-DpsA
2: TAD /DBD-DpsA
3: TAD-DpsG/DBD-DpsB
4: TAD /DBD-DpsB
5: TAD-DpsG/DBD-DpsC
6: TAD /DBD-DpsC
7: TAD-DpsG/DBD-DpsD
8: TAD /DBD-DpsD
9: TAD-DpsG/DBD-DpsG
10: TAD /DBD-DpsG
11: TAD-DpsG/DBD
It has been shown that the KSα and KSβ from the act system when expressed and purified, co-elute as heterodimer suggesting strong non-covalent interactions between the two subunits of the ketosynthase (Carreras and Khosla, 1998). Thus, the interaction between DpsA and DpsB assessed by yeast two-hybrid system, could be considered as an intrinsic positive control for the technique used in this study. Strong interactions were found involving DpsA polypeptides. Interestingly, the nutritional selection screening of the possible interactions within the 'minimal' PKS, revealed a strong interaction between DpsA and the putative MAT, DpsD.

The putative model of architecture for the 'minimal' DNR/DXR PKS displays a head-to-tail arrangement for the three proteins: DpsA stabilizes the complex due to its strong homotypic interactions whereas DpsB and DpsD interact with each DpsA polypeptide and are aligned in opposite orientations (Fig. 3.16).

LacZ assays revealed that DpsB and DpsD interacted transiently with each other but were not involved in homotypic interactions. The strong interaction displayed by DpsD with DpsA is particular intriguing because the role of DpsD in the production of AA, the first enzyme free intermediate in the DNR/DXR biosynthetic pathway, is not clear since dpsD appear not to be required for in vivo synthesis in heterologous host (Grimm et al., 1994). The proposed model might explain the data reported by Rajgarhia et al. during expression of dps genes in heterologous host (Rajgarhia et al., 2001). In this study, expression of the complete set of dps genes in Streptomyces lividans TK24 supported production of AA whereas a dpsCD double mutant accumulated as major product desmethylaklanonic acid, derived from the utilisation of acetate as starter unit, with the AA level reduced to only 20%. Expression of dpsD but not of dpsC also resulted in production of desmethylaklanonic acid and AA in a ratio 60:40. However, expression of dpsC independently from the presence of DpsD, reverted the production of AA to a level similar to wild type. These results suggested that promiscuous starter unit selection can occur in absence of dpsC and dpsD and that DpsC contributes to the selection of propionyl-CoA as starter unit but it does not alone dictate it (Rajgarhia et al., 2001).
Figure 3.16 Putative head-to-tail arrangement for the proteins forming the core of the 'minimal' DNR/DXR PKS

This arrangement of proteins has been suggested by the investigation of protein interactions of 'minimal' DNR/DXR PKS. DpsA establishes homo and heterotypic strong interactions with DpsB and DpsD which are aligned in antiparallel fashion. DpsB and DpsD interact transiently with each other.
Figure 3.17 shows that in absence of DpsC and DpsD, DpsA can be loaded with an acetate primer deriving from the decarboxylation reaction carried out by DpsB of the malonyl-DpsG. DpsB has been shown to act as decarboxylase by a previous study (Bao et al., 1999b). However, also propionate could be loaded on the DpsA from propionyl-CoA without prior decarboxylation and incorporated as starter unit (Rajgarhia et al., 2001). Differences in the amount of desmethylaklanonic and AA (ratio 80:20) observed might have been due to the different kinetic rates for these two loading reactions. Acetylation of the DpsA might be faster than the loading of propionate due to the smaller dimension of the acyl residue resulting in a higher incorporation rate of acetate as starter unit. Also, the cellular concentration of acetyl-CoA might be higher compared to propionyl-CoA explaining the difference in the use of the two alternative starter units.

As proposed in the model, the presence of DpsD might interfere with loading of acetate as a starter unit slowing down this priming reaction (Fig. 3.18). Therefore, incorporation of propionate over acetate may result to be more favourable as confirmed by the metabolite production analysis that displayed an increase in the amount of AA detected in the culture medium. DpsD, as shown by the metabolite production, might contribute in the selection of the right starter unit, propionate, interfering with the loading reaction of the acetate onto DpsA. Studies on the type II PKS responsible for the production of the family of anthraquinone compounds nominated R1128 (Marti et al., 2000), revealed new insight in the priming reaction and about the function of homologues of DpsD. R1128 PKS possesses, the unusual feature amongst the type II PKSs, as the DNR/DXR PKS, to utilize a non-acetate starter unit and contains ZhuC, a protein homolog to DpsD. ZhuC was found to have enzymatic activity as acetyl-CoA thioesterase purging the acetate primers that otherwise would dominate in the polyketide chain initiation (Tang et al., 2004b). The expression of DpsC in heterologous host along with all the dps genes, restored the wild type fenotype with production of AA virtually at 100% (Fig. 3.19).

The ACP, DpsG, was found to interact transiently with all the Dps proteins but did not show any homotypic interaction. Indeed, it has been demonstrated that the interactions established by the ACP involve hydrophobic/electronegative residues along the structure formed by helix a2-loop-helix a3. Negatively charged residues recognise
In the absence of both DpsC and DpsD, acetate starter units, deriving from decarboxylation of malonyl-DpsG, could be loaded faster than propionate leading to more acetate being incorporated in the nascent polyketide as indicated by the ratio of aklanonic acid and desmethylaklanonic acid. The proposed structure of the heterotetramer formed by DpsA and DpsB supports this hypothesis.
Figure 3.18 Correlation of the *in vivo* data with the putative head-to-tail model

The presence of DpsD could interfere with the loading of acetate resulting in a slower rate of acetate incorporation, allowing more propionate to be incorporated as observed by the increased percentage of aklanonic acid produced. The putative arrangement for DpsA, DpsB and DpsD could explain the hypothesised slower rate of loading of acetate due to the partial inhibition for the presence of DpsD.
Figure 3.19 Correlation of the *in vivo* data with the putative head-to-tail model. The presence of DpsC influences the choice and the loading of the right starter unit, propionate, restoring the exclusive production of aklanonic acid.
positive residues embedded in hydrophobic patches of its partner enzymes in type II FAS and PKS (Zangh et al., 2003a, Zhang et al., 2001). A model for the S. coelicolor actinorhodin 'minimal' PKS complex has been described where the ACP dissociates from the KSα/KSβ heterodimer after each round of condensation (Fig. 1.4) (Dreier and Khosla, 2000). This model would also be consistent with the observation that the ACP interacts weakly with the other components of the 'minimal' complex. However, no homotypic interaction between two ACPs was revealed by the assays. Indeed, such interaction which facilitates the inter-ACP transfer of malonate has been suggested for several type II PKS ACPs (Zhou et al., 1999; Arthur et al., 2006).

β-galactosidase activity assays showed that the type III ketosynthase, DpsC, was found to interact transiently only with DpsG. dpsC gene encodes a protein homolog to E. coli FabH. FabH is essential for the biosynthesis of type II fatty acid synthase catalysing the first condensation reaction yielding acetoacetyl-ACP. The enzyme is firstly acetylated on its active site cysteine and then binds malonyl-ACP to catalyse via a ping pong mechanism biosynthesis of acetoacetyl-ACP (Jackowski et al., 1989). The ping pong nature of the mechanism involved in the transfer and condensation reaction catalysed by the homolog FabH could explain the weak interaction found between DpsC and the ACP, DpsG.

FabH homologues from different bacteria such as Staphylococcus aureus and Mycobacterium tuberculosis have been characterized and their crystal structure revealed to be active as homodimer (Musayev et al., 2005; Qiu et al., 2005; Pan et al., 2002). No such homotypic interaction has been found involving DpsC neither strong nor weak. Purification studies have shown that, in contrast with the homodimeric structure of FabH homologues, the active form of DpsC is in fact monomeric (Bao et al., 1999a). Interestingly, DpsC appears to be acylated with propionate onto Ser-118 while the other E. coli and S. coelicolor KSIII enzymes are acetylated on the correspondent position via thioester linkage on a cysteine residue (Bao et al., 1999a). Metabolic studies revealed that DpsC coupled with E. coli, ACP, AcpP, exhibited two different abilities: a typical KSIII like activity catalysing the condensation between propionyl-CoA and malonyl-AcpP to yield β-ketovaleryl-AcpP, but also as acyltransferase transferring propionate from propionyl-CoA onto the AcpP (Bao et al., 1999a). Therefore, although DpsC shows high similarity with other KSIII homologues, it can not be ruled out its function as acyltransferase transferring the propionyl group to either DpsG or DpsA.
From this study it appears that DpsC interacts only with DpsG, thus it can not be discriminate whether it acts as KSIII or as acyltransferase transferring a β-ketovaleryl or propionyl group respectively onto DpsG.

During the yeast two-hybrid assays, swapping of proteins as ‘bait’ and ‘prey’ resulted in different outcomes. This phenomenon was particularly evident for the ACP, DpsG, when fused to the GAL4 TAD used as ‘prey’. ACPs are relatively small proteins (9 kDa) (White et al., 2005) and it could be speculated that the GAL4 TAD (~17 kDa) might assume a conformation that interferes with the normal folding of DpsG impairing the ability to establish interactions with other protein partners during assays. In addition, the mainly acidic nature of the exposed amino acids on the ACP surface may have caused improper folding of the GAL4 domain interfering with its ability to interact with the proteins of transcription machinery. Also, the strong interactions either already described in literature for homologues of Dps protein such as act KSα/KSβ or novel, such as DpsA/DpsD, were detected only testing one particular combination of ‘bait’ and ‘prey’. False negatives can be encountered performing yeast two-hybrid system assay leading to miss valid interactions. The reason for missing interactions are not clear but may include steric constraints on the ‘bait’ or ‘prey’ interaction properties or protein misfolding induced by the expression of the protein of interest as fusion of the DBD or TAD (Golemis and Brent, 1992; Toby and Golemis, 2001). During this study, the interaction assays, performed testing the alternative combinations of ‘bait’ and ‘prey’, have reduced the unlikely event of missing valid interactions.

The investigation of the interactions between the proteins forming the ‘minimal’ DNR/DXR PKS using the yeast two-hybrid system has revealed very interesting results that are supported by other previous analysis conducted both in vitro and in vivo. On the base of these results a putative architecture for the ‘minimal’ complex is proposed where the two subunits of the KS, DpsA and DpsB, exist as heterotetramer stabilized by the strong interactions occurring between two DpsA polypeptides. The structure assumes a head-to-tail alignment with DpsD that interacts strongly with DpsA and only weakly with DpsB. Two alternative putative mechanisms for the chain initiation in the aklanonic acid biosynthesis can be proposed based on the protein interactions assessed by yeast two-hybrid assays. In the priming reaction the propionyl-CoA is loaded onto DpsC that could either behave as KSIII, catalysing the first condensation with malonyl-DpsG producing β-ketovaleryl-DpsG, or as acyltransferase, transferring the propionyl
group onto DpsG, so that the first condensation reaction is carried out by the KS subunits DpsA and DpsB (Fig. 3.20). The putative MAT, DpsD, forming the core of the 'minimal' PKS, might have a structural role as steric inhibitor of the aberrant loading of acetate as starter unit or act as acylthioesterase providing a specific catalytic residue as indeed indicated by a previous study (Tang et al., 2004b). The protein interactions involving also the auxiliary enzymes present in the DNR/DXR PKS, DpsEFY, were subsequently investigated and the results are described in the next chapter.
Figure 3.20 Proposed priming mechanisms for the ‘minimal’ DNR/DXR PKS
In the priming reaction DpsC could act either as a type III ketosynthase (KSIII) catalysing the first condensation between propionate and malonate or as acyltransferase (AT) transferring the starter unit onto DpsG.
CHAPTER 4
Investigation of protein interactions of the entire
DNR/DXR PKS using yeast two-hybrid system

4.1 Introduction
Production of aromatic polyketides relies on the condensing reactions carried out by the ‘minimal’ PKS but also on the activity of other proteins responsible for the successive chemical modifications of the polyketide carbon skeleton. Typical downstream enzymes present in type II PKS clusters include ketoreductases, cyclases and aromatases that define the folding pattern of the nascent poly-β-keto intermediate (Hopwood, 1997). DpsE is the enzyme present in the DNR/DXR PKS that catalyses the reduction of the keto group at the C9 of the nascent carbon chain in aklanonic acid biosynthesis (Grimm et al., 1994). The DNR/DXR PKS contains the aromatase DpsF that carries out the dehydration and cyclisation reactions associated with the formation of the first aromatic ring whereas the cyclase DpsY is responsible for the generation of the second and third ring in aklanonic acid (Grimm et al., 1994; Lomovskaya et al., 1998; Hautala et al., 2003).

The influence of the downstream enzymes on the folding pattern and even on the chain length of the aromatic polyketides, has been reported by different experiments (Kramer et al., 1997; Shen et al., 1999). Polyketide biosynthesis directed by type II PKSs relies not just on proteins with enzymatic activities but on the association of these components in multiprotein complexes. The protein interactions necessary for the assembly of a type II PKS complex are very important for the biosynthetic machinery to work properly leading to the production of aromatic polyketide with a specific chain length and cyclisation pattern. In this chapter, possible protein interactions involving the DNR/DXR PKS including also the downstream enzymes, DpsEFY, have been investigated using the GAL4 yeast two-hybrid system (Fields and Song, 1989).

4.2 Results
4.2.1 PCR amplification and cloning of the dps genes
PCR amplification of dpsEFY genes was achieved by using the plasmid template pWHM1012 and the high fidelity Pfu DNA polymerase. Oligonucleotide primers were
specially designed for convenient and easy cloning into plasmid vectors pGBKT7 and pGADT7 by using the one-step cloning strategy already described in section 3.3.2 (Fig. 3.7). Each forward primer was designed to anneal at the 5' end of the \(dps\) gene and contained an \(NdeI\) restriction site upstream the translational start codon. Reverse primers were specific for the 3' of the \(dps\) gene end and carried an \(EcoRI\) restriction site downstream stop codon. Both \(NdeI\) and \(EcoRI\) restriction sites were present in the MCS of the plasmid vectors for the yeast two-hybrid system but not in the sequence of any \(dps\) gene. The size of the PCR products, oligonucleotide primers and annealing temperatures used to amplify the different \(dps\) genes in this chapter are listed in Table 4.1. Extrabases present at the 5' end of each oligonucleotide primer allowed direct digestion of the amplified product. Double digested amplification products were ligated in the pGBKT7 vector for the 'bait' and successively cloned in pGADT7 to obtain also the 'prey'. Plasmids generated during this work are listed in Table 4.2 and were analysed by PCR and restriction digestion pattern (Fig. 4.1). DNA sequencing results revealed that the \(dps\) genes had been cloned in the same translational frame of the GAL4 domain genes and also that mutations had not been introduced by the DNA polymerase used to amplify the \(dps\) genes.

### 4.2.2 Expression of recombinant proteins

Expression of full length hybrid proteins was assessed by immunoblotting assays performed on protein extracts from yeast transformants carrying GAL4 DBD and TAD fusions. 'Bait' proteins were detected using antibodies targeted to the c-Myc epitope encoded by the pGBKT7 derivative plasmids while presence of 'prey' hybrid proteins was assessed using antibody against Ha epitope encoded by pGADT7 derivative plasmids. Protein extract from untransformed \(S. cerevisiae\) AH109 strain was used as negative control. Immunoblotting assays revealed the presence of DpsE, DpsF and DpsY as full-length hybrid proteins fused to GAL4 DBD and TAD domains in the yeast cells (Fig. 4.2).
Table 4.1 Amplification of the *dps* genes
Oligonucleotide primers used for the amplification of *dps* genes by PCR; the recognition sequence for restriction enzymes is underlined. Annealing temperature and the size of the expected amplification products for each *dps* gene are indicated.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5' to 3'</th>
<th>Recognition sites</th>
<th>Annealing (°C)</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>dpsE</em>-forward</td>
<td>caccatacatatgtccagggccgaccg</td>
<td><em>NdEL</em></td>
<td>60</td>
<td>785</td>
</tr>
<tr>
<td><em>dpsE</em>-reverse</td>
<td>ccgaattccggtcagtagctgccgaggg</td>
<td><em>EcoRI</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>dpsF</em>-forward</td>
<td>caccatacatatgtccagttgcctgccgactcg</td>
<td><em>NdEL</em></td>
<td>60</td>
<td>949</td>
</tr>
<tr>
<td><em>dpsF</em>-reverse</td>
<td>cggaattccgctcaggcctggactcg</td>
<td><em>EcoRI</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>dpsY</em>-forward</td>
<td>caccatacatatgtcttgtggccgagcagg</td>
<td><em>NdEL</em></td>
<td>60</td>
<td>818</td>
</tr>
<tr>
<td><em>dpsY</em>-reverse</td>
<td>cggaattcgttcatctgtcgaacgagg</td>
<td><em>EcoRI</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Plasmid (features/backbone)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| \textit{dpsE} | pNC112  
DBD::\textit{dpsE} // pGBK7 |
|       | pNC116  
\textit{TAD}::\textit{dpsE} // pGAD7 |
| \textit{dpsF} | pNC113  
DBD::\textit{dpsF} // pGBK7 |
|       | pNC117  
\textit{TAD}::\textit{dpsF} // pGAD7 |
| \textit{dpsY} | pNC114  
DBD::\textit{dpsY} // pGBK7 |
|       | pNC115  
\textit{TAD}::\textit{dpsY} // pGAD7 |

\textbf{Table 4.2} Cloning strategies and features of the plasmids generated during this study
Figure 4.1 NdeI/EcoRI digestion of the pGBKT7 and pGADT7 derivative plasmids carrying \textit{dps} genes (A and B) with the excision of the \textit{dps} gene insert (arrowhead)

A)  
M:1 Kb ladder  
1:pNC112 (\textit{dpsE})  
2:pNC113 (\textit{dpsF})  
3:pNC114 (\textit{dpsY})

B)  
M:1 Kb ladder  
4:pNC116 (\textit{dpsE})  
5:pNC117 (\textit{dpsF})  
6:pNC115 (\textit{dpsY})
Figure 4.2 Western blot assays performed on protein extract of yeast transformants carrying hybrid Dps proteins fused to DBD (A) and to TAD (B) indicated by arrowheads

A) M: Precision Plus™ Dual Color
1: DpsE-DBD (44.0 kDa)
2: DpsF-DBD (51.7 kDa)
3: DpsY-DBD (46.4 kDa)
4: DBD (16.9 kDa)
C: AH109 protein extract

B) M: Precision Plus™ Dual Color
5: DpsE-TAD (43.9 kDa)
6: DpsF-TAD (51.6 kDa)
7: DpsY-TAD (46.3 kDa)
8: TAD (16.8 kDa)
C: AH109 protein extract
4.2.3 Yeast two-hybrid system assay

Analysis of the possible protein interactions that occur between all the proteins required for aklanonic acid biosynthesis was performed using the matrix shown in Figure 4.3. As described in Chapter 3, heterotypic interactions were investigated using two different combinations of ‘prey’ and ‘bait’ to minimize any influence of either GAL4 DBD or TAD in the interaction assay. However, homotypic interactions could only be analysed using one combination of ‘bait’ and ‘prey’.

Up to 1 μg of each plasmid was used to transform *S. cerevisiae* AH109 strain and the average number of colonies obtained on -2 drop-out medium was 150 cfu μg⁻¹. Strong interactions were detected using nutritional selection on -4 drop-out medium and weak interactions were assessed by β-galactosidase activity on colonies able to grow on -2 but not on -4 drop-out medium. All the controls described in Chapter 3 were set up at each round of transformation to decrease the incidence of false results and each interaction assay was repeated independently three times.

4.2.3.1 Nutritional selection

Expression of *ADE2* and *HIS3* reporter genes was monitored on -4 drop-out medium to detect strong interaction occurring between the ‘bait’ and the ‘prey’. The cyclase DpsY was found to interact strongly with the two subunits of the KS, DpsA and DpsB, with the putative MAT, DpsD, and with the ketoreductase DpsE. Interestingly, these strong heterotypic interactions could be only detected using pNC116 where DpsY is fused to the GAL4 TAD as ‘prey’ but not testing the equivalent combinations with the DBD fusion as ‘bait’. DpsY also established a strong homotypic interaction.

Strong interaction between DpsE and DpsD were detected co-transforming pNC112 (DBD-DpsE) and pNC105(TAD-DpsD) but not swapping domains using pNC110 (DBD-DpsD) and pNC115(TAD-DpsE). DpsE tested as ‘prey’ did not show any heterotypic strong interaction. Also, strong homotypic interaction involving DpsE were not detected. The cyclase DpsF tested either as ‘bait’ or ‘prey’, using pNC113(DBD-DpsF) or pNC117(TAD-DpsF), did not show any either strong homo or heterotypic interactions. All the interaction assays were performed along with the relative positive and negative controls. No autoactivation effects were detected, negative controls showed growth only on -2 drop-out medium while co-transformation of pVA3-1 and pTD1-1 resulted in growth also on -4 drop-out medium.
<table>
<thead>
<tr>
<th>'bait'</th>
<th>A (KSα)</th>
<th>B (KSβ)</th>
<th>C (KSIII)</th>
<th>D (MAT)</th>
<th>E (ACP)</th>
<th>F (ARO)</th>
<th>Y (CYC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>'prey'</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>A (KSα)</td>
<td>AA</td>
<td>AB</td>
<td>AC</td>
<td>AD</td>
<td>AG</td>
<td>AE</td>
<td>AF</td>
</tr>
<tr>
<td>B (KSβ)</td>
<td>BA</td>
<td>BB</td>
<td>BC</td>
<td>BD</td>
<td>BG</td>
<td>BE</td>
<td>BF</td>
</tr>
<tr>
<td>C (KSIII)</td>
<td>CA</td>
<td>CB</td>
<td>CC</td>
<td>CD</td>
<td>CG</td>
<td>CE</td>
<td>CF</td>
</tr>
<tr>
<td>D (MAT)</td>
<td>DA</td>
<td>DB</td>
<td>DC</td>
<td>DD</td>
<td>DG</td>
<td>DE</td>
<td>DF</td>
</tr>
<tr>
<td>G (ACP)</td>
<td>GA</td>
<td>GB</td>
<td>GC</td>
<td>GD</td>
<td>GG</td>
<td>GE</td>
<td>GF</td>
</tr>
<tr>
<td>E (KR)</td>
<td>EA</td>
<td>EB</td>
<td>EC</td>
<td>ED</td>
<td>EG</td>
<td>EE</td>
<td>EF</td>
</tr>
<tr>
<td>F (ARO)</td>
<td>FA</td>
<td>FB</td>
<td>FC</td>
<td>FD</td>
<td>FG</td>
<td>FE</td>
<td>FF</td>
</tr>
<tr>
<td>Y (CYC)</td>
<td>YA</td>
<td>YB</td>
<td>YC</td>
<td>YD</td>
<td>YG</td>
<td>YE</td>
<td>YF</td>
</tr>
</tbody>
</table>

**Figure 4.3** Matrix of possible protein interactions of the entire DNR/DXR PKS
++ indicates strong interactions revealed by nutritional selection
+ indicates weak interactions revealed by β-galactosidase assay
4.2.3.2 β-galactosidase activity

Weak interactions occurring between all the Dps proteins were assessed monitoring LacZ reporter gene expression (Fig. 4.4, 4.5 and 4.6). β-galactosidase activity was detected in all the interactions involving DpsY as 'prey' except with DpsG and DpsF that were indeed detected by swapping domains (Fig. 4.5A, 4.5B and 4.4B). Surprisingly, using DpsY as 'bait' and testing the interactions with DpsB, used as 'prey', did not result in any β-galactosidase activity observed (Fig. 4.5B). The same interaction testing the equivalent combination using DpsY as 'prey' and DpsB as 'bait', showed expression of LacZ reporter gene and had already been assessed by nutritional selection (Fig. 4.5A).

When DpsF was used as 'prey', weak interactions with all the Dps proteins except with DpsB and DpsC were detected (Fig. 4.4B). However, DpsF tested as 'bait' was found to establish transient interactions also with DpsB and DpsC (Fig. 4.6B). Weak interactions between DpsF and DpsG could only be detected testing the combination DBD-DpsG and TAD-DpsF but not swapping domains (Fig. 4.4B and 4.6B). β-galactosidase activity was also observed testing the homotypic interaction involving DpsF (Fig. 4.4B). Weak interactions were found when pNC115, that encodes DpsE as 'prey', was co-transformed with pNC110(DBD-DpsD), pNC111(DBD-DpsG), pNC113(DBD-DpsF) and pNC114(DBD-DpsY) (Fig. 4.4A). These interactions were also confirmed testing the equivalent combination of 'bait' and 'prey' (Fig. 4.4B, 4.5A and 4.6A). Interactions between DpsE and the proteins DpsB and DpsC, that resulted negative tested using DpsE as 'prey', were indeed detected by swapping domains (Fig. 4.4A and 4.6A). The LacZ assays for detection of the possible interaction involving DpsE and DpsA resulted negative testing the two equivalent combinations of 'bait' and 'prey' (Fig. 4.4A and 4.6A)

The β-galactosidase activity assays were performed along with the relative negative and positive controls described in Chapter 3. Double and single negative controls did not result in any expression of the reporter gene while interaction between murine p53, encoded by pVA3-1, and SV40 large antigen, encoded by pTD1-1, resulted positive for β-galactosidase activity.
Figure 4.4 LacZ assays performed on yeast transformants carrying the indicated combination of prey/bait (TAD-Dps/DBD-Dps) blue colonies indicates positive
A) DpsE ‘prey’  
C+: p53/ SV40 Large T antigen  
1: TAD-DpsE/DBD-DpsA  
2: TAD /DBD-DpsA  
3: TAD-DpsE/DBD-DpsB  
4: TAD /DBD-DpsB  
5: TAD-DpsE/DBD-DpsC  
6: TAD /DBD-DpsC  
7: TAD-DpsE/DBD-DpsD  
8: TAD /DBD-DpsD  
9: TAD-DpsE/DBD-DpsG  
10: TAD /DBD-DpsG  
11: TAD-DpsE/DBD-DpsE  
12: TAD /DBD-DpsE  
13: TAD-DpsE/DBD-DpsF  
14: TAD /DBD-DpsF  
15: TAD-DpsE/DBD-DpsY  
16: TAD /DBD-DpsY  
17: TAD-DpsE/DBD  
C-: TAD /DBD  
B) DpsF ‘prey’  
C+: p53/ SV40 Large T antigen  
18: TAD-DpsF/DBD-DpsA  
19: TAD /DBD-DpsA  
20: TAD-DpsF/DBD-DpsB  
21: TAD /DBD-DpsB  
22: TAD-DpsF/DBD-DpsC  
23: TAD /DBD-DpsC  
24: TAD-DpsF/DBD-DpsD  
25: TAD /DBD-DpsD  
26: TAD-DpsF/DBD-DpsG  
27: TAD /DBD-DpsG  
28: TAD-DpsF/DBD-DpsE  
29: TAD /DBD-DpsE  
30: TAD-DpsF/DBD-DpsF  
31: TAD /DBD-DpsF  
32: TAD-DpsF/DBD-DpsY  
33: TAD /DBD-DpsY  
34: TAD-DpsF/DBD  
C-: TAD /DBD
Figure 4.5 *LacZ* assays performed on yeast transformants carrying the indicated combination of prey/bait (TAD-Dps/DBD-Dps) blue colonies indicates positive

A) DpsY 'prey'
C+: p53/SV40 Large T antigen
1: TAD-DpsY/DBD-DpsA
2: TAD/DBD-DpsA
3: TAD-DpsY/DBD-DpsB
4: TAD /DBD-DpsB
5: TAD-DpsY/DBD-DpsC
6: TAD/DBD-DpsC
7: TAD-DpsY/DBD-DpsD
8: TAD/DBD-DpsD
9: TAD-DpsY/DBD-DpsG
10: TAD /DBD-DpsG
11: TAD-DpsY/DBD-DpsE
12: TAD/DBD-DpsE
13: TAD-DpsY/DBD-DpsF
14: TAD/DBD-DpsF
15: TAD-DpsY/DBD-DpsY
16: TAD/DBD-DpsY
17: TAD-DpsY/DBD
C+: TAD/DBD

B) DpsY 'bait'
C+: p53/SV40 Large T antigen
18: TAD-DpsA/DBD-DpsY
19: TAD-DpsA/DBD
20: TAD-DpsB/DBD-DpsY
21: TAD-DpsB/DBD
22: TAD-DpsC/DBD-DpsY
23: TAD-DpsC/DBD
24: TAD-DpsD/DBD-DpsY
25: TAD-DpsD/DBD
26: TAD-DpsG/DBD-DpsY
27: TAD-DpsG/DBD
28: TAD/DBD-DpsY
C-: TAD/DBD
Figure 4.6 LacZ assays performed on yeast transformants carrying the indicated combination of prey/bait (TAD-Dps/DBD-Dps) blue colonies indicates positive

A) DpsE ‘bait’
C+: p53/ SV40 Large T antigen
1: TAD-DpsA/DBD-DpsE
2: TAD-DpsA/DBD
3: TAD-DpsB/DBD-DpsE
4: TAD-DpsB/DBD
5: TAD-DpsC/DBD-DpsE
6: TAD-DpsC/DBD
7: TAD-DpsD/DBD-DpsE
8: TAD-DpsD/DBD
9: TAD-DpsG/DBD-DpsE
10: TAD-DpsG/DBD
11: TAD/DBD-DpsE
C-: TAD/DBD

B) DpsF ‘bait’
C+: p53/ SV40 Large T antigen
12: TAD-DpsA/DBD-DpsF
13: TAD-DpsA/DBD
14: TAD-DpsB/DBD-DpsF
15: TAD-DpsB/DBD
16: TAD-DpsC/DBD-DpsF
17: TAD-DpsC/DBD
18: TAD-DpsD/DBD-DpsF
19: TAD-DpsD/DBD
20: TAD-DpsG/DBD-DpsF
21: TAD-DpsG/DBD
22: TAD/DBD-DpsF
C-: TAD/DBD
4.3 Discussion

Possible interactions involving the entire set of proteins forming the DNR/DXR PKS, DpsABCDGEFV were tested using a GAL4 yeast two-hybrid system. dpsEFY genes were amplified by PCR and cloned in the plasmid vectors pGBK7 and pGADT7 for expression of hybrid proteins in yeast fused to GAL4 DBD and TAD respectively. The plasmids obtained were used to transform S. cerevisiae strain AH109 to carry out protein interaction assays testing all the possible combinations of ‘bait’ and ‘prey’ (Fig. 4.3).

A strong interaction was assessed by nutritional selection involving the ketoreductase DpsE with one of the component of the ‘minimal’ PKS, DpsD. In contrast to the homolog KR from the act PKS that has a tetrameric structure (Hadfield et al., 2004), DpsE did not show any homotypic interaction by using yeast two-hybrid system. Nutritional selection revealed the strong homo and heterotypic interactions involving the cyclase DpsY. A previous study has reported that disruption of DpsY in S. peucetius leads to the accumulation of UWM5 which is similar to SEK43, a known shunt product of the biosynthetic pathway of aromatic polyketides (Grimm et al., 1994; Lomovskaya et al., 1998).

An important role for DpsY in DNR/DXR PKS might be suggested by the strong interactions assessed using the yeast two-hybrid system. DpsY was found to interact strongly with the two subunits of the ketosynthase, DpsA and DpsB, the putative MAT, DpsD, and the ketoreductase DpsE. In addition, a strong homotypic association was detected involving DpsY suggesting a quaternary structure for the cyclase with a multimeric arrangement. Structural studies revealed that TcmF2 cyclase from the tcm PKS cluster in S. glaucescens exists as homodimer (Thompson et al., 2004) whereas AknH and SnoaL, responsible for the last cyclisation step in aclacinomycin and nogalamycin biosynthesis in S. galilaeus and S. nogalater respectively, are tetramers (Kallio et al., 2006; Sultana et al., 2004). These results might suggest a particular structural role for the cyclase in addition to its own enzymatic activity in the formation of the second and third ring in the biosynthesis of aklanonic acid (Hautala et al., 2003).

The interactions identified by nutritional selection could be used to hypothesise that the DNR/DXR PKS complex assumes a cylinder-like architecture as shown in Figure 4.7.
Figure 4.7 Proposed architecture for the DNR/DXR PKS
The model proposed for the arrangement of the DNR/DXR PKS has a cylinder-like architecture. The nascent growing chain is synthesised in the amphipathic tunnel composed by DpsA, DpsB and DpsD. When the condensation reactions have been carried out, the polyketide chain may fill the tunnel buckling and the C9 results exposed to DpsE that interacts with DpsD. DpsY has a crucial structural function interacting with all the components of the PKS and surrounding the internal core forming the barrel of the cylinder. Dps proteins found to associate transiently with their interacting partners are not shown. For clarity, the indicated stoichiometry of the complex is only speculative.
In this model, DpsY would be physically located on the outer layer of the barrel with the internal core composed of DpsA, DpsB and DpsD. It can be speculated that the cylinder shape of the complex could influence both chain length and the regiochemistry of the growing, highly reactive, poly-oxy-acyl chain. Chain length and regioselectivity are the main characteristics to be determined during the biosynthesis of the carbon backbone of an aromatic polyketide. Previous studies have proved that the length of the growing carbon chain is influenced not only by the ‘minimal’ PKS but also by the presence of downstream enzymes, as reported for the whiE PKS from S. coelicolor (Shen et al., 1999) and fren PKS (Kramer et al., 1997), which control the folding of the final polyketide product. Therefore, it is likely that a complex association of many proteins is needed for effective and controlled reactions to be carried out by a type II PKS.

The 2.0 Å resolution X-ray solved crystal structure of the act heterodimer KSα/KSβ (Keatinge-Clay et al., 2004) has suggested that the polyketide chain is synthesised at the interface where the two proteins form an amphipathic tunnel. In the amphipathic tunnel, the growing polyketide chain is fully extended separating the keto groups preventing any spontaneous intramolecular reaction (Harris and Harris, 1986). Once the carbon chain has reached the appropriate length presumably emerging from the internal cavity, the downstream enzymes can act on the polyketide backbone. Hence, in the biosynthesis of the carbon skeleton of an aromatic polyketide, the chain length is mainly dictated by the physical dimensions of the tunnel where it has been synthesised. Also, the structural details of the interface of the act KSα/KSβ heterodimer revealed that at the stage of heptaketide, the extended chain is at the end of the tunnel and the final condensation causes the chain to buckle allowing the first cyclisation to occur (Keatinge-Clay et al., 2004). Although the auxiliary enzymes, such as ketoreductases and cyclases do not have access to the carbon skeleton during the condensation reactions, the addition of these components to the ‘minimal’ PKS could influence cyclisation and even the chain length patterns (McDaniel et al., 1993; Kramer et al., 1997).

The observations from the actinorhodin producing PKS support the model shown in Figure 4.7. The amphipathic tunnel at the interface of the DpsA/DpsB heterodimer would keep the highly reactive growing carbon chain in a stable environment throughout all the condensation reactions preventing any aberrant cyclisation reaction.
The internal core of the complex is also composed of the putative MAT, DpsD, that interacts strongly with the ketoreductase DpsE. Once the nine condensation reactions have been completed, the 21-carbon atom skeleton of aklanonic acid might then become exposed to DpsE that catalyses the reduction of the keto group at C9. After the formation of the first aromatic ring by the action of DpsF, successive cyclisation reactions could be carried out by DpsY that form the outer layer of the barrel. As proposed in this model, DpsY would possess not only enzymatic activity but also an essential structural function for the assembly of the complex. Interestingly as it had already been assessed during the investigation of the protein interactions for the 'minimal' PKS, DpsG was found to interact with all the downstream enzymes reflecting the particular role of the ACP to shuttle the acyl chain substrate to the different components of the PKS in order to carry out the specific enzymatic modifications (White et al., 2005).

As described in Chapter 3, the yeast two-hybrid system used in this study has revealed limitations also during the investigation of the protein interactions for the whole DNR/DXR PKS. The crystal structure of the KR from the act PKS, ActIII, showed a tetrameric quaternary structure as dimer of dimers (Hadfield et al., 2004). The yeast two-hybrid system assays involving DpsE did not result in any either strong or transient homotypic interaction for this protein. The GAL4 DBD and/or TAD polypeptides fused to the ketoreductase might have impaired the establishment of such interaction.

The investigation of protein interactions of the DNR/DXR PKS described in this chapter has suggested a putative overall architecture for the complex (Fig. 4.7). The PKS complex assumes a cylinder-like structure with the cyclase DpsY having an important role in maintaining the complex in biologically active configurations. The internal core of this complex is formed by the subunits of the KS, DpsA and DpsB and the putative MAT, DpsD. This core is kept together mainly by the strong homotypic interactions between DpsA polypeptides. The interface DpsA/DpsB provides the amphipathic tunnel where the 21-carbon atom chain is synthesised via Claisen condensation. The final condensation allows the growing acyl chain to become susceptible to reduction at C9 by DpsE that is located at both ends of the tunnel interacting with DpsD. Successive aromatisation and cyclisation reactions are carried out respectively by DpsF and DpsY (Hutchinson and Colombo, 1999). The other proteins, that from the nutritional selection did not appear to take part to the formation
of the cylinder-like complex, should be indeed in close association with each other and with the complex itself. The results of the investigation of protein interactions involving the DNR/DXR PKS are corroborated by other studies conducted on different of type II PKSs (Keatinge-Clay et al., 2004; Kramer et al., 1997). Structural analysis of the overall architecture of type II PKS complexes could be useful to guide biosynthetic chemists in designing new pharmaceuticals. In order to investigate protein interactions involving the DNR/DXR PKS by using a different \textit{in vivo} technique, the project moved on the application of Tandem Affinity Purification (TAP) method on the ‘minimal’ PKS complex. Attempts to verify the suitability of this method for the purification the ‘minimal’ DNR/DXR PKS are described in the next chapter.
CHAPTER 5

Tandem Affinity Purification of the ‘minimal’ DNR/DXR PKS

5.1 Purification of protein complexes

A large part of the proteomic research is focused on the analysis of the protein-protein interactions. Isolation of intact multiprotein complexes is of crucial importance for the study of those interactions in living cells. Purification of protein assemblies is a challenging task for scientists because some complexes might be unstable during the purification steps. Many protein purification protocols use compounds or procedures that could interfere with weak interactions, especially those involved in the formation of transient complexes. Fusion of 6-histidine residues (6x His-Tag) to a protein of interest is routinely used for purification by affinity chromatography. The imidazole ring of the amino acid form complexes with the Nickel ions exposed on the surface of the resin. Typically, the crude protein extract is passed through a column containing Nickel-bound resin to allow retaining of the protein fused to the His-Tag. Elution of the protein bound to the resin is performed using high concentration of imidazole that compete with the His tagged protein for the Nickel ions releasing fusion protein from of the resin (Hengen, 1995). The use of high concentration of imidazole, up to 1 M, can perturb weak interactions between protein partners causing the protein complex to become unstable. As a result of this instability under purification conditions, isolation of many biologically important complexes in which protein partners interact only transiently, might be difficult to achieve (Rigaut et al., 1999).

5.1.2 Tandem Affinity Purification

Tandem Affinity Purification (TAP) is a relatively recent technique that allows purification of protein complexes under native conditions in high yield (Rigaut et al., 1999). The method consists of two sequential affinity purification steps using two different tags contained in the unit named TAP tag. The two IgG-binding sites of protein A of Staphylococcus aureus (ProtA) and the calmodulin-binding peptide (CBP) are the two units present in TAP tag to perform the two affinity purification under native conditions. TAP tag can be fused to either the N- or C- terminal of the target protein of interest (Fig. 5.1).
**Figure 5.1** Overview of the TAP method

A) N-terminal and C-terminal TAP tag available. ProtA: IgG binding site of Protein A from *Staphylococcus aureus*; TEV: Tobacco etch virus protease cleavage sequence; CBP: Calmodulin binding peptide; EK: Enterokinase cleavage sequence (present only in N-terminal TAP tag)

B) TAP purification method: The crude cell extract is first passed on a IgG bound resin that retains the target protein and its interacting partners. The complex is released from the resin using TEV protease and passed on a calmodulin resin to purify the complex from remaining TEV protease. The elution of the purified complex is then accomplished using EGTA.
The choice of the tags to fuse to the target protein is crucial to achieve efficient recovery of protein complex without impairing protein interactions and activity. Typically in TAP tag, CBP is located closer to the protein of interest whereas ProtA is always at the end of the tag, either at the N- or C-terminal depending if it is an amino- or carboxy-terminal fusion. The TAP method involves initial affinity purification using a matrix-bound IgG which retains the protein complex through high affinity and very specific interactions with the ProtA epitope present in the Tap tagged protein. This interaction is very stable and the release could only be performed at very low pH which represents a denaturing condition. To achieve releasing of the protein complex without perturbation of the association between interacting partners, TAP tag contains a proteolytic cleavage site between ProtA and CBP that allows the elution of the complex by protease digestion. The cleavage site is for the protease encoded by the tobacco etch virus (TEV) which is suitable for the purification process as the proteolytic digestion is very efficient and the enzyme has a highly stringent sequence specificity compared to other proteases (Dougherty et al., 1989a). In addition, the use of TEV protease is generally safe for purification of protein assemblies as its consensus site is not commonly found amongst proteins (Dougherty et al., 1989b).

After the TEV protease cleavage, the eluate is incubated for the second affinity purification with calmodulin-coated beads in presence of bivalent calcium ions. The CBP tag still fused to the protein of interest allows binding to the calmodulin resin. After washing to remove contaminants and TEV protease still carried over from the previous affinity step, the bound material is released by adding the chelating agent EGTA that forms complex with bivalent calcium ions impairing the binding of the CBP to calmodulin. An additional proteolytic cleavage site for an enterokinase is located in the TAP tag for N-terminal fusions immediately downstream CBP, allowing complete removal of the tag after the purification steps. TAP method requires fusion of the DNA encoding for the TAP tag in the same translational frame of the coding gene of the protein of interest in an appropriate expression vector. The recombinant vector is subsequently transformed in the host cell where could be integrated in the genome or existing as independent unit as plasmid resulting in stable or transient expression respectively. An overview of the TAP purification strategy is also shown in Figure 5.1.
5.1.2.1 Sensitivity and limitations of the TAP method

The TAP method has been initially developed in yeast but its use has been successfully extended to other organisms such as plants and bacteria and mammalian cells (Rohila et al., 2006; Takebe et al., 2006; Tsai and Carstens, 2006). In yeast, TAP method has been tested isolating the already described complex formed by the protein U1 snRNP (Gottschalk et al., 1998) which is essential for recognition of the pre-mRNA 5'-splice site and the subsequent assembly of the spliceosome (Rigaut et al., 1999). One of the main advantages of this technique is the simplicity of the entire process, the high yield of protein recovered from a relatively small amount of cells and the wide applicability to different organisms.

Different variants of the TAP tag have been developed to overcome potential problems with the original affinity components (Terpe, 2003). The use of EGTA as chelant agents might interfere with some protein interaction and a tag, called TAPa, has been specially designed to contain 9 copies of the c-Myc epitope or 6 histidine residues in place of the CBP (Rubio et al., 2005). In addition, this variant has a proteolytic cleavage site for the rhinovirus 3C protease instead of TEV protease which is more active and specific at low temperature. This allows releasing of the bound material with higher yield from the IgG resin at more suitable condition for the stability of some complexes. TAP tag has also been modified for mammalian cells by using either an octapeptide, named FLAG tag, as replacement of the CBP tag (Tsai and Carstens, 2006) or substituting the original tags with proteinG and streptavidine-binding peptide (Burckstummer et al., 2006) to obtain an higher yield of recovery after purification.

Protein complexes isolated by TAP method have a high degree of purity allowing subsequent identification of the subunits by mass spectrometry. Also, information about the stoichiometry of the entire protein complex can be obtained by mass spectrometry analysis (Puig et al., 2001). Ideally, TAP method should be performed expressing proteins of interest at, or close to, their natural level. Protein overexpression is not possible for complexes of unknown composition as might lead to formation of non-specific interactions with host proteins (Swaffield et al., 1995). However, identification of partners of target protein, whose gene was under the control of the strong promoter P_{BAD} from the araBAD operon, has been successfully achieved with the identification of new interacting partners of the *E. coli* ACP (Gully et al., 2003).
The high sensitivity of the TAP method to accomplish isolation of protein assemblies in native conditions allows to test the activity of the purified complex and to reveal the effect of mutations over stability and activity of the complex. Compared to the yeast two-hybrid system, TAP method displays the advantage to identify all the interacting partners of the protein of interest in one single experiment. An intrinsic limitation of the TAP method, that is shared with all the tagging methods, is the possibility that the tag could not be sufficiently exposed on the surface of the complex to permit efficient binding in the different affinity purification steps resulting in loss of material. Also, both the expression level of the protein and the activity of the complex could be affected by the presence of the tag. In order to solve this problem, TAP tag might be located either at N- or C-terminal of the target protein (Puig et al., 2001).

5.1.2.2 TAP method for ‘minimal’ DNR/DXR PKS
The results obtained using the yeast two-hybrid experiments suggested that the ‘minimal’ DNR/DXR PKS might be organized as a core complex with two DpsA polypeptides interacting with each other and each of it with DpsB and DpsD orientated in an antiparallel fashion. The other proteins forming the ‘minimal’ PKS, DpsC and DpsG interact transiently with the core complex according with their mechanism of action (Jackowski et al., 1989; White et al., 2005). To verify protein interactions previously assessed via yeast two-hybrid system and to verify the applicability of the TAP method to the isolation of aromatic PKS complexes, the outcomes of the attempts to purify the ‘minimal’ DNR/DXR PKS from the host organism S. coelicolor A(3)2 are described in this chapter.

5.2 Results
5.2.1 Cloning strategy for TAP tagging of DpsA and expression of the ‘minimal’ DNR/DXR PKS
From the yeast two-hybrid results described in Chapter 3, DpsA appeared to be very important for the formation of the ‘minimal’ PKS which might be considered a homodimer of a heterotrimer composed of DpsA, DpsB and DpsD (Fig. 3.16). In order to use the TAP method for the purification of the ‘minimal’ DNR/DXR PKS, the TAP tag was fused at the N-terminal of DpsA. The vector used for expression of the genes \(dpsABCDG\) encoding for the ‘minimal’ constituents of the PKS complex was a derivative of the plasmid pWHM1012 (Bao et al., 1999b). The cloning strategy involved amplification by PCR of the TAP cassette present in the DNA template.
plasmid pBS1761 which is suitable for N-terminal fusion of target proteins. The oligonucleotides used to prime the PCR reaction for the amplification reaction were specially designed to accomplish cloning of the DNA fragment encoding for the TAP tag in the same translational frame of \( \textit{dps}A \). Important elements present in the forward and reverse oligonucleotide primers and their function for the cloning strategy are shown in Figure 5.2.

The high fidelity \( \textit{Pfu} \) DNA polymerase was used to amplify TAP tag with TAP forward and TAP reverse oligonucleotide primers and an annealing temperature of 46 °C, obtaining a DNA fragment of 580 bp. The amplified product was cloned in the commercial vector pGEMT-easy generating the plasmid designated pNC140. DNA sequencing using T7 forward primer of the coding sequence of TAP tag in pNC140, was performed to verify absence of any relevant mutations that may have been introduced by the \( \textit{Pfu} \) DNA polymerase during the PCR reaction. pNC140 was incubated with \( \textit{NdeI} \) restriction endonuclease and the DNA fragment of 580 bp was cloned in pNC102 (Chapter 3) immediately upstream the translational start codon of \( \textit{dps}A \) gene using the \( \textit{NdeI} \) restriction site. The plasmid obtained was designated pNC145 (Fig. 5.3).

Since the cloning of \( \textit{TAP tag} \) in pNC102 was performed using a single restriction site, \( \textit{NdeI} \), the DNA fragment might have been introduced in two possible orientations. In order to check the orientation of the TAP tag insert, pNC145 was double digested using \( \textit{BclI} \) and \( \textit{XhoI} \) restriction endonucleases. \( \textit{XhoI} \) restriction site was present at the 3’end of the coding region of \( \textit{dps}A \) whereas a \( \textit{BclI} \) recognition sequence was introduced downstream the \( \textit{NdeI} \) restriction site in the TAP forward primer. The double digestion yielded a fragment of 1.8 kbp which was the size expected from the insertion of TAP tag with the right orientation upstream \( \textit{dps}A \) gene. Prior digestion using \( \textit{BclI} \) restriction enzyme, pNC145 was transformed in \( \textit{E. coli} \) K12 ER2925 strain to obtain unmethylated plasmid as this particular endonuclease is sensitive to methylation on guanine of its recognition sequence. The 1.8 kbp \( \textit{TAP tag::dps}A \) fragment obtained from the double digestion \( \textit{BclI/XhoI} \) was used to replace the \( \textit{BglII/XhoI} \) fragment present in plasmid pNC141. pNC141 is a derivative of pWHM1012 generated by intramolecular ligase reaction of the 5.6 kbp fragment derived from the digestion with \( \textit{XhoI} \) restriction endonuclease (Fig. 5.4).
TAP forward primer

\[ \text{BclI} \]

CATATGACATTAGAAAAGGAGGAACCCCGGATGATAACTTCG

\[ Ndel \] STOP RBS START

tcmJ*

TAP reverse primer

\[ \text{Ndel} \]

CATATTACGCTTATCGTCATCATCAAGTGC

Figure 5.2 Oligonucleotide primers for amplification of TAP tag from pBS1761. Important features of the two oligonucleotide primers are indicated. Ndel restriction sites were used for cloning of the amplified TAP tag in pNC102. BclI restriction site was inserted to generate pNC146. A stop codon was inserted to terminate translation of \textit{tcmJ} whereas a canonical \textit{Streptomyces} RBS was introduced to allow translation of TAP tag::dpsA fusion gene. In blue color is shown the TAP tag gene sequence while in orange the Ndel restriction site located at the translational start codon of DpsA present in pNC147.
Figure 5.3 Cloning strategy: preparation of pNC145

pNC140, a pGEMT-easy derivative carrying amplified TAP tag, and pNC102 (Chapter 3) were digested using NdeI endonuclease. The 580 bp TAP tag fragment, obtained from pNC140, and the 9.2 kbp, derived from pNC102, were ligated together generating pNC145 containing TAP tag upstream dpsA. For clarity arrows indicate full-length genes, rectangles other elements or part of genes.
**Figure 5.4** Cloning strategy: preparation of pNC141

pWHM1012 was digested using XhoI endonuclease. The 5.6 kbp fragment, containing *ermE*p promoter, *tcmJ* and *dpsA* plus elements for propagation and selection in *E. coli*, *oriC* and *Amp*', was ligated on itself to generate pNC141. For clarity arrows indicate full-length genes, rectangles other elements or part of genes.
The 5.6 kbp \textit{XhoI} fragment carried the sequence region of pWHM1012 that contains the promoter \textit{ermE*p} (Bibb \textit{et al.} 1994) and \textit{dpsA} gene up to the \textit{XhoI} restriction site plus the genetic elements necessary for the existence as plasmid in \textit{E. coli} such as origin of replication, OriC, and the selection marker, \beta-lactamase resistance gene.

pWHM1012 contains other than \textit{ermE*p} and the complete set of \textit{dps} genes, \textit{tcmJ} gene from the \textit{tcm} PKS cluster which was carried along during the cloning steps to generate the plasmid solely for convenience (Bao \textit{et al.}, 1999b). \textit{tcmJ}, located upstream \textit{dpsA} and also present in pNC141, contains a \textit{BglII} restriction site in the coding region. \textit{BglII} and \textit{BclII} are two compatible restriction enzymes that generate the same overhang protruding sequence. This allowed successful replacement of the 1.3 kbp \textit{BglII/XhoI} fragment from pNC141 by the 1.8 kbp \textit{TAP tag::dpsA BclII/XhoI} fragment from pNC145 obtaining plasmid pNC146 (Fig. 5.5). pNC146 was subsequently linearised by digestion with \textit{XhoI} restriction endonuclease and ligated with a 7.9 kbp fragment obtained from \textit{XhoI} digestion of plasmid pNC143. pNC143 is another derivative of pWHM1012 which lacks of \textit{dpsEFY} genes compared to the parental plasmid (Fig. 5.6). Ligation of the 7.9 kbp fragment from pNC143 with the linear pNC146 resulted in the construction of the plasmid pNC147 that encodes for the N-terminal fusion of TAP tag with DpsA and for the other constituents of the ‘minimal’ PKS, DpsBCDG (Fig. 5.7).

5.2.2 \textbf{Considerations upon tcmJ and design of oligonucleotide primers}

\textit{tcmJ} gene is located upstream \textit{dpsA} both in pWHM1012 and pNC147. Its presence in the plasmid and presumably expression did not have any influence on the production of aklanonic acid in heterologous host (Bao \textit{et al.}, 1999b). In order to create N-terminal fusion of TAP tag with DpsA, a \textit{BglII} restriction site present in \textit{tcmJ} coding region was used to introduce the \textit{TAP tag} gene. The cloning would have resulted in an in-frame fusion of TcmJ with TAP tag that might have interfered with the purification of the ‘minimal’ DNR/DRX PKS complex. To avoid this possibility, the forward oligonucleotide primer for the \textit{TAP tag} gene amplification by PCR was designed to contain a stop codon downstream the \textit{BclII} restriction site to terminate the translation of \textit{tcmJ} mRNA. A canonical GGAGG Streptomyces ribosome binding site (RBS) was introduced in the TAP forward primer separated by 8 nucleotides upstream the translational start codon of the TAP tag, to allow efficient protein synthesis of TAP tag-DpsA fusion and the other ‘minimal’ Dps proteins (Fig. 5.8).
Figure 5.5 Cloning strategy: preparation of pNC146
The 1.8 kbp fragment obtained from BclI/XhoI double digestion of pNC145 was used to replace the 1.3 kbp BgIII/XhoI fragment in pNC141. The plasmid obtained, pNC146, contained TAP tag upstream truncated version of dpsA. For clarity arrows indicate full-length genes, rectangles other elements or part of genes.
Figure 5.6 Cloning strategy: preparation of pNC143

pWHM1012 was digested using EcoRI endonuclease. The 7.2 kbp fragment containing elements for propagation and selection in *E. coli*, oriC and Amp', and in *S. coelicolor*, oriS and *Tsr*', was ligated with the 5.8 kbp fragment carrying *ermE* promoter, *tcmJ* and *dpsABCDG* genes to generate pNC143. For clarity arrows indicate full-length genes, rectangles other elements or part of genes.
Figure 5.7 Cloning strategy: preparation of pNC147

pNC146 and pNC143 were digested using *Xhol* endonuclease. The 6.2 kbp fragment from pNC146 was ligated to the 7.4 kbp fragment from pNC143 to generate pNC147. pNC147 was used to express the 'minimal' DNR/DXR PKS for TAP purification. For clarity arrows indicate full-length genes, rectangles other elements or part of genes.
Figure 5.8 Genetic organization around \( tcmJ \) and \( dpsA \) genes in pWHM1012(A) and in pNC147(B).

In pWHM1012  \( tcmJ \) and \( dpsA \) are separated by a short nucleotide sequence. In pNC147  \( tcmJ \) (\( \Delta tcmJ \)) gene is prematurely truncated by a TGA stop codon. \( BglII \) consensus site is disrupted (\( BglII^*/BglII^* \)) by cloning of the \( BclI \) protruding sequence indicated in green. Ribosome binding site (RBS) allows translation of \( TAP \) tag::\( dpsA \) fusion gene. \( BclI \) consensus site, TGA stop codon and RBS were present in TAP primer forward. In orange is indicated the \( NdeI \) consensus sequence introduced by TAP primer reverse that inserted a CAT codon (H) at the linking junction of \( TAP \) tag and \( dpsA \) and mutated the \( dpsA \) start codon from GTG to ATG.

For clarity, the drawn is not on scale. Dashed line indicates <600 bp sequence separation, dashed line plus // indicates >600 bp.
5.2.3 Expression of the ‘minimal’ DNR/DXR PKS

The cloning strategy previously illustrated yielded the plasmid pNC147 which resembles the genetic organization present in pWHM1012. From the analysis of the regulatory and coding regions in pWHM1012, the expression of the *dps* genes in pWHM1012 appear to derive from the translation of a single polycistronic mRNA transcribed from the constitutive promoter *ermE*p (Fig. 5.9). Effective expression of all *dps* genes carried by pWHM1012 with correspondent production of aklanonic acid was also reported in 1999 by Bao et al. (Bao et al., 1999b). The genetic structure of pNC147 is very similar to pWHM1012 with the addition of TAP tag gene fused at the 5’ end of *dpsA* and the premature termination of the translation of the *tcmJ* mRNA. Also a canonical Streptomyces RBS was inserted eight nucleotides upstream the translational start codon TAPtag::dpsA to provide effective translation of the fusion gene.

These genetic modifications determined alterations of the sequence of the long polycistronic mRNA transcribed from *ermE*p promoter responsible for the expression of *dpsBCDG* and the TAP tag::dpsA hybrid gene. The stability of bacterial mRNA is highly dependent on the secondary structure of the nascent transcript in particular for clustered genes which expression relies on the synthesis of one unique polycistronic mRNA. Transcription and translation can be impaired by particular secondary structures formed during the transcription of a gene resulting in degradation of the mRNA with subsequent negative effects on the expression of the downstream genes (polar effect) (Carter and Newton, 1969). To discount this possibility, RT-PCR was performed to verify the effective synthesis of a long polycistronic mRNA necessary for the expression of the *dps* genes encoding for the ‘minimal’ PKS. To prime the reverse transcriptase reaction two different reverse oligonucleotide primers were used: *dpsG*-reverse and *dpsB*-reverse. cDNA obtained by the reverse transcriptase reaction primed with *dpsG*-reverse primer was used to perform standard PCR reaction to verify the presence of *dpsG*, *dpsD* and *dpsC* genes using the couple of primers *dpsG*-forward/*dpsG*-reverse, *dpsD*-forward/*dpsD*-reverse and *dpsC*-forward/*dpsC*-reverse respectively. The product from the reverse transcriptase reaction primed with *dpsB*-reverse was used to perform standard amplification by PCR reaction to verify the presence of *dpsB* and *dpsA* genes with the couple of oligonucleotide primers *dpsB*-forward/*dpsB*-reverse and *dpsA*-forward/*dpsA*-reverse respectively (Fig. 5.9).
Figure 5.9 Overview of the transcription of the gene cluster in pNC147 and RT-PCR. Transcription of TAP tag::dpsA and DpsBCD G genes from ermE*p promoter led to the synthesis of a long polycistronic mRNA. dpsB-reverse and dpsG-reverse oligonucleotide primers were used to prime cDNA synthesis. PCR amplification of the dps genes was performed using oligonucleotide primers listed in Table 3.1.
Each amplification reaction resulted positive as shown by the size of expected molecular weight of bands obtained after DNA electrophoresis on agarose gel, suggesting transcription of the \textit{dps} genes in \textit{S. coelicolor} A3(2) transformed with plasmid pNC147 (Fig. 5.10). RNA preparation used for the RT-PCR experiments was also submit to PCR omitting the first step of cDNA synthesis by reverse transcriptase reaction. This negative control, performed using all couple of primers for amplification of the \textit{dps} genes, was necessary to check purity of RNA preparation from any plasmidic DNA contamination. These negative controls did not show any band on agarose gel suggesting that the amplification products obtained by RT-PCR reactions were specifically due to the presence of mRNA encoding for the \textit{dpsABCDG} genes in the initial RNA preparation.

Expression of the chimeric protein TAP tag-DpsA was revealed by immunoblotting using anti c-Myc monoclonal primary antibodies. The ProtA epitope contained in TAP tag is a binding site for IgG antibodies and TAP tag fusion proteins can be detected by immunoblotting assay using virtually every IgG derivative antibody. Anti-c-Myc and horseradish alkaline peroxidase-conjugated were used, as primary and secondary antibodies respectively, to perform Western Blot experiments. Protein extract from \textit{S. coelicolor} A(3)2 transformed with pNC147 showed a band of approximately 65 kDa corresponding to the full-length hybrid protein TAP tag-DpsA. Protein extracts from \textit{S. aureus}, that contained Protein A, and from untransformed \textit{S. coelicolor} A(3)2 were used as positive and negative controls respectively (Fig. 5.11). Comparison of crude protein extracts from transformed and un-transformed host \textit{S. coelicolor} A(3)2 did not show in any visible difference, particularly regarding overexpression of TAP tag::\textit{dpsA} fusion gene and \textit{dpsBCDG} genes, between the two preparations on SDS-PAGE.

### 5.2.4 TAP method

The procedure followed for purification using the TAP method is described in section 2.7 with protein extracts from up to 2 litres of mycelium. During the preliminary purification attempts was assessed that already in the first elution step with 22 mM EGTA, proteins were efficiently released from the calmodulin beads. The amount of protein eluted was maximum at the first step and gradually decreased up to the last fifth elution step. All five eluate samples analysed by SDS-PAGE appeared to contain two proteins in major quantity of estimated molecular weight of approximately 48 and 44 kDa (Fig. 5.12 and 5.13).
Figure 5.10 RT-PCR on RNA extracted from *S. coelicolor* A(3)2 transformed with pNC147 (A) and wild type (B)

A)  
M: 1 kb ladder  
1: RT-PCR *dpsA* primers  
2: PCR *dpsA* primers  
3: Negative control PCR *dpsA* primers  
4: RT-PCR *dpsB* primers  
5: PCR *dpsB* primers  
6: Negative control PCR *dpsB* primers  
7: RT-PCR *dpsC* primers  
8: PCR *dpsC* primers  
9: Negative control PCR *dpsC* primers  
10: RT-PCR *dpsD* primers  
11: PCR *dpsD* primers  
12: Negative control PCR *dpsD* primers  
13: RT-PCR *dpsG* primers  
14: PCR *dpsG* primers  
15: Negative control PCR *dpsG* primers

B)  
M: 1 kb ladder  
16: RT-PCR *dpsA* primers  
17: PCR *dpsA* primers  
18: Negative control PCR *dpsA* primers  
19: RT-PCR *dpsB* primers  
20: PCR *dpsB* primers  
21: Negative control PCR *dpsB* primers  
22: RT-PCR *dpsC* primers  
23: PCR *dpsC* primers  
24: Negative control PCR *dpsC* primers  
25: RT-PCR *dpsD* primers  
26: PCR *dpsD* primers  
27: Negative control PCR *dpsD* primers  
28: RT-PCR *dpsG* primers  
29: PCR *dpsG* primers  
30: Negative control PCR *dpsG* primers
**Figure 5.11** Immunoblotting assay on crude protein extract using anti-c-Myc IgG conjugated antibody

M: Precision Plus\textsuperscript{TM} Dual Color

1: *S. coelicolor* A3(2) transformed with pNC147

2: *S. coelicolor* A3(2) wild type

3: *Staphylococcus aureus* crude extract
**Figure 5.12** Analysis of the TAP purification steps on protein extract from *S. coelicolor* A3(2) transformed with pNC147 on SDS-PAGE at 10% CBP-DpsA and DpsB are visible in all lanes from 2-8

M: Precision Plus™ Unstained
1: Flow-through after binding on IgG resin
2: Flow-through after TEV protease cleavage
3: Flow-through after binding on calmodulin affinity resin
4: I elution from calmodulin affinity resin
5: II elution from calmodulin affinity resin
6: III elution from calmodulin affinity resin
7: IV elution from calmodulin affinity resin
8: V elution from calmodulin affinity resin
**Figure 5.13** Analysis of the TAP purification steps on protein extract from *S. coelicolor* A3(2) transformed with pNC147 on SDS-Page at 10% (A) and 15% (B)

M: Precision Plus™ Unstained  
1: Crude extract  
2: Flow-through after binding on IgG resin  
3: Flow-through after TEV protease cleavage  
4: Flow-through after binding on calmodulin affinity resin  
5: Elution from calmodulin affinity resin
The size of the bands obtained on gel suggested that the eluted proteins were DpsA and DpsB. After the purification steps, DpsA is still fused to a peptide of 4.76 kDa that derives from the cleavage of TAP tag by TEV protease for the release from the IgG matrix. As a result of this, DpsA (43.67 kDa) was expected to be present as fusion protein with a molecular weight of 48 kDa whereas DpsB was expected to be revealed by a band on gel that corresponds to its molecular weight (43.67 kDa). Mass spectrometry analysis confirmed the presence of DpsA as chimeric protein of 48 kDa and DpsB as a band of 43.67 kDa at the end of the purification process.

5.3 Discussion

The work described in this chapter regards the attempts to purify the ‘minimal’ DNR/DXR PKS using a methodology named Tandem Affinity Purification (TAP) (Rigaut et al., 1999). This technique allows purification of protein complexes in native conditions by using two different affinity purification steps. This can be achieved targeting a protein with a special tag, called TAP tag, which contains two different epitopes to perform the two affinity chromatography steps. The process will purify the protein complex which contains the target protein. Because the purification steps are conducted in native condition the complex will eventually retain its own biological activity once purified (Puig et al., 2001). The adaptability of this method for the aromatic PKS complexes was suggested by other reported studies involving E.coli ACP, that resulted in co-purification of most of the key enzymes of type II FAS (Gully et al., 2003; Butland et al., 2005), and type II FAS system from plants (Brown et al., 2006).

Investigation of protein-protein interactions of the ‘minimal’ DNR/DXR PKS using GAL4 yeast two-hybrid system (Chapter 3) resulted in strong interactions found involving DpsA both in strong homo and heterotypic interactions with DpsB and DpsD. The other proteins of the ‘minimal’ PKS resulted to interact transiently with their partners. These data led to the speculation that the ‘minimal’ PKS might be constitute by a core composed of two DpsA polypeptides each one interacting with DpsB and DpsD. The core of the ‘minimal’ PKS should be centred on the interacting DpsA polypeptides with DpsB and DpsD aligned in antiparallel orientation but facing each other (Fig 3.16). The remaining proteins, DpsC and DpsG, interact only transiently with core as was expected by the well documented mechanism of action of their homologues (White et al., 2005).
The attempts of purification of the ‘minimal’ PKS were performed targeting DpsA which from the yeast two-hybrid results was found to be the key protein to hold together the core of the ‘minimal’ PKS. Analysis of the crystal structure of the KSα/KSβ from the act cluster (PDB accession code: 1TQY) (Keatinge-Clay et al., 2004) revealed that both N- and C-terminal of the KSα are not involved in the interaction and thus either could have been used in DpsA to fuse the TAP tag minimizing the possibility of perturbation of the protein association. Also, the amino and carboxy termini of the protein are not buried in the complex and possibly sufficiently exposed to allow binding during the affinity purification steps.

The expression of the ‘minimal’ DNR/DXR PKS with the hybrid protein TAP tag-DpsA was performed in the convenient host S. coelicolor A3(2) using a plasmid that carried all the genes necessary for the analysis under the control of the strong constitutive promoter ermE*p (Bibb et al., 1994). Canonical TAP purification method requires expression of the target protein under the control of the native promoter or close to physiological level (Rigaut et al., 1999). This requirement is necessary to avoid non-specific interaction of the tagged protein with other cellular components that would affect the downstream analysis of interaction networks established by the target protein. Example of diversion from this general rule have been reported in literature with new interacting partners of the ACP in E. coli have been identified using a transient expression driven by a strong promoter (Gully et al., 2003).

In this study, the TAP method was used to verify the interactions already assessed by yeast two-hybrid system and possible non-specific interactions would have been identified by mass spectrometry analysis on the eluate complex. In addition, plasmid carrying ermE*p promoter to drive expression of the dps genes has already been used successfully for production of akalanonic acid in Streptomyces lividans (Bao et al., 1999b) suggesting that the Dps subunits, expressed under the control of a constitutive promoter, are still able to establish appropriate interactions to form a functional protein complex.

Expression of the dps genes was assessed by RT-PCR that revealed the transcription of a long polycistronic mRNA under the control of the ermE*p promoter. The presence of a full-length hybrid protein TAP tag-DpsA was detected by immunoblotting assay using IgG antibody that binds the ProtA epitope located at the N-terminal of the TAP tag.
Analysis of the eluates collected at the end of the tandem affinity purification by SDS-PAGE and the successive preliminary characterization by mass spectrometry, revealed that the complex obtained contained DpsA and DpsB. No detection of DpsC, DpsD or DpsG was obtained from the analysis of the SDS-PAGE at the end of the purification step (Fig. 5.13).

The interaction between DpsA and DpsB was expected since it has been revealed from the analysis of the crystal of the KSα/KSβ from the act PKS (Keatinge-Clay et al., 2004). Also, co-purification by gel chromatography of act KS subunits that eluted as αβ2 heterotetramer, had already been described earlier (Carreras and Khosla, 1998). Failure to detect any other interactions with the other components of the ‘minimal’ PKS might have been due to the presence of the TAP tag at the N-terminal of DpsA. The analysis of the crystal structure of the act KSα/KSβ complex revealed that the N-terminal of DpsA is sufficiently exposed and not involved in crucial interactions with DpsB. However, the presence of the TAP tag in this region of the protein might have impaired interaction with the other ‘minimal’ components such as DpsD. Failure of the detection of the DpsC and DpsG was taken into account as suggested from the weak interactions in which they were found to be involved by yeast two-hybrid system and according to their proposed mechanism of action as KSIII and ACP respectively (Jackowski et al., 1989; White et al., 2005).

The absence of DpsD at the end of the purification process appeared not be consistant with the protein interaction data assessed by the yeast two-hybrid system (Chapter 3). This result does not necessary mean that the strong interactions found between DpsA and DpsD using yeast two-hybrid system should be classified as false positives. TAP method and yeast two-hybrid system are two different but successfully used techniques to investigate protein-protein interactions (Field and Song, 1989; Puig et al., 2001). While yeast two-hybrid system explores interactions as binary combination of proteins, TAP method involves purifications of the macromolecular complex containing the target protein. Protein complexes are dynamic structures where the strength of the association might not be the same between all the different interacting partners. The strong interaction between the KSα and KSβ has allowed researchers to purify the crystal revealing the molecular insights of this complex central in the biosynthesis of aromatic polyketides (Keatinge-Clay et al., 2004). Also, KSα/KSβ interaction was also detected by gel filtration in previous studies that involved tcm and act systems.
(Gramajo et al., 1991; Carreras and Khosla, 1998). The interaction of DpsA with DpsD might be weaker and less intimate than with DpsB resulting in a different strength of the association that may not resist to the conditions, although mild, used during the TAP purification process.

Analysis of the protein interactions assessed for the entire DNR/DXR PKS has led to hypothesise a putative complex architecture where DpsY plays a primary structural role. It can be speculated that the association of DpsD with DpsA might be stabilized by the presence of DpsY during the biosynthesis of aklanonic acid. If this is correct, the absence of DpsY during the expression and purification of the ‘minimal’ PKS might have been the reason of the failure in co-purifying DpsD together with DpsA and DpsB by TAP method. The association of DpsA and DpsD could have been instable without the cyclase leading to the loss of DpsD during the purification steps. On the other hand, the interaction between DpsA and DpsD could indeed be detected using yeast two-hybrid system probably due to the high sensitivity of the technique (Yang et al., 1995).

Analysis of the protein content of the samples derived from the different washing steps by SDS-PAGE did not indicate clearly the presence of DpsD and any of the other proteins of the ‘minimal’ PKS, DpsC or DpsG (Fig. 5.14). Therefore, by using SDS-PAGE analysis it was difficult to determine the critical purification step that led to the loss of those proteins.

The experiments reported here represent a first and preliminary attempt of application of purification of complexes from aromatic PKSs by Tandem Affinity Purification (TAP). The method warrants improvements and different approaches to accomplish purification of the complex. A different approach could be performed fusing the TAP tag either at N- or C-terminal and analysing the outcome of the purification process. This would allow to assess whether the location of the TAP tag influences the interaction with other proteins and to map interaction domains. Protein interactions that have not been revealed using DpsA as target, might be detected fusing the TAP tag with a different protein. DpsD might be used as protein target for fusion with TAP tag in order to establish its associations with other Dps proteins.
Figure 5.14 Analysis of the washing steps during the TAP purification of protein extract from *S. coelicolor* A3(2) transformed with pNC147 on SDS-Page at 10%(A) and 15%(B).

**M:** Precision Plus™ Unstained

1: I wash from IgG affinity beads
2: II wash from IgG affinity beads
3: III wash from IgG affinity beads
4: I wash from calmodulin resin
5: II wash from calmodulin resin
6: III wash from calmodulin resin
Another approach to the analysis of the protein interaction of type PKSs could be targeting particular proteins that play crucial structural role. The hypothesis formulated in Chapter 4 for the arrangement of the DNR/DXR PKS relied on the observation that the DpsY was involved in strong interactions suggesting an essential role in the architecture of the entire complex. Using DpsY as target protein might allow purification of its interacting partners and possibly of the entire DNR/DXR PKS complex as suggested by the model shown in Figure 4.7. Coupled with mass spectrometry this approach might be used to verify the protein interactions detected and more important to define the stoichiometry of the complex elucidating the overall structure of the PKS complex. The possible influence of the threedimensional structure of the Dps proteins on the outcome of the TAP analysis and on the interactions established with each other led to to the structural study of those proteins described in the next chapter.
CHAPTER 6

Investigation of Dps proteins interactions by computational methods

6.1 Introduction

Since Watson and Crick first published the double helical structure of DNA in 1953, DNA and its products, proteins, have been the focus of the scientific community. Sequencing projects have made available the genome of many organisms and have allowed the annotation of previously unknown open reading frames. The availability of entire genome sequences and putative gene products have elucidated the importance of proteomics in the study of biological processes (Pandey and Mann, 2000). Proteomics is the large-scale study of proteins to investigate their location, quantity and activity and to address protein functions in the cellular processes. The abundance of proteomic data has highlighted that the activity of most proteins is modulated by the formation of highly organized multiprotein complexes which are essential for many biological activities (von Mering et al., 2002; Rives et al., 2003; Gavin et al., 2006; Krogan et al., 2006).

Protein associations with diverse partners under different conditions result in different biological outcomes, depending on the nature of the interacting proteins. Since protein interactions mediate the formation of many biological complexes, interaction proteomics has gained more attention as a comprehensive view of the global organism’s proteome and as well to define specific interactions. The analysis of the protein interaction networks, called ‘interactome’, might be used to predict function for interacting partners of a known protein by the so called ‘guilt-by principle’ (Oliver, 2000). Also, the function of a novel protein complex can be predicted using a ‘majority rule’ principle, whereby the most common function held by the majority of the members defines the overall function of the complex. In addition to useful information for a better understanding of the biological processes, the interactome has also rapidly emerged as an attractive and effective target for modern medicine for tackling many diseases (Bark and Hook, 2007) and to design novel drugs (Strong and Eisenberg, 2007).
For the model yeast *Saccharomyces cerevisiae* protein-protein interaction maps have been systematically analysed by using an impressive set of experimental techniques including yeast two-hybrid system (Fields and Song, 1989) and purification of protein complexes (Rigaut et al., 1999) coupled with mass spectrometry to identify interacting partners (Ho et al., 2002; Gavin et al., 2002). The accumulated and annotated proteomic data have urged the necessity of a more rapid and less expensive methods for prediction of protein-protein interactions. Bioinformatic tools are rapidly gaining more attention amongst protein scientists for the computational methods that have been developed to help prediction of docked complexes (Smith et al., 2002).

6.1.2 Protein databases

The construction and constant maintenance of protein databases is amongst the most useful services that computer scientists may offer to biologists. The need for computer readable biological databases arises from the extraordinary quantity of experimental data that are widely disseminated in literature and are hard to search by scientists. Development of new techniques that allow genomic and proteomic analysis and the growing number of experimental data require a constant update of existing and creation of new biological databases (Peters and Sikorski, 1997; Xenarios and Eisenberg, 2001).

Three-dimensional structures of proteins derive typically from nuclear magnetic resonance (NMR) spectrometry or X-ray crystallography solutions of *in vitro* purified complexes. These data are subsequently deposited in the database PDB (Protein Data Bank) (http://www.rcsb.org/pdb/home/home.do) managed by Research Collaboratory for Structural Bioinformatics (RCSB). The main goal of PDB database is to provide a single world-wide archive of structures of biological macromolecules available for scientists (Westbrook et al., 2003; Kouranov et al., 2006). SWISS-PROT (http://www.expasy.org/sprot/ and http://www.ebi.ac.uk/swissprot/) is a protein sequence and knowledge database (also called Knowledgebase) that provides scientific community with high quality annotations for proteins (O'Donovan et al., 2002). The core data include basic protein information such as amino acid sequence, description and function. Also biological relevant domain/site(s), subcellular locations and splicing isoforms are amongst the further information available from the database (Boeckmann et al., 2003). Universal Protein Knowledgebase (UniProt) includes SWISS-PROT and PIR (Protein Information Resource) (Wu et al., 2003) to provide scientific community with one single comprehensive database of protein sequences and functional
information (Apweiler et al., 2004). Several databases including BIND (Biomolecular Interaction Network Database) (Bader et al., 2003) and BioGRID (General Repository for Interaction Datasets) (Stark et al., 2006) have also been organized to facilitate visualization and availability of interactomic data.

A different approach can be used to support the methods aimed to predict protein domains that are responsible for interactions. The automatic analysis of protein sequences is possible through the use of ‘protein signatures’ which are methods for definition of a domain or characteristic region of a protein family in a protein sequence. Pfam is a database of protein domain families available at http://www.sanger.ac.uk/Software/Pfam/ which contains multiple sequence alignments for each family to identify particular domains in new protein sequences (Finn et al., 2006). The latest Pfam release (22.0) in July 2007 contains over 9318 proteins families (http://www.sanger.ac.uk/Software/Pfam/). Pfam helps annotation of protein families with the aim to identify previously undetected structural homologues that have arise from a single evolutionary origin (Finn et al., 2006). Other signature and sequence cluster databases have been created including PROSITE (http://expasy.org/prosite) (Hulo et al., 2006) which contains signature as patterns and profiles, ProDom (www.toulouse.inra.fr/prodom.html) (Bru et al., 2005) and SMART (Letunic et al., 2006). These databases are most effective when used together rather than in isolation. InterPro integrates into one resource ten proteins signature databases to allow a comprehensive annotation of structural and biological information for proteins (Mulder et al., 2007).

6.1.3 Docking methods
The creation of databases that annotate global characteristics of protein domains has represented an important step towards the prediction of protein-protein interactions which is the most ambitious target of modern bioinformatics (Franzot and Carugo, 2003; Valencia and Pazos, 2002). Computational methods are much faster and less expensive than most experimental analysis and different algorithm have been developed to achieve prediction of protein interactions.
The simplest scoring function is shape complementarity which defines the surface characteristic of separated components and their potential complementarity. The free energy change ($\Delta G$) of formation of a complex relative to its unbound components is the other important parameter since the association of separated subunits correspond to a minimum of $\Delta G$. Also, electrostatic and van der Walls interactions must be considered as well as water mediated interactions (Tobias, 2001). Protein docking simulation is a process composed of different stages to define a list of proposed orientations for docked complexes. The first step is normally the treating of single protein as rigid bodies searching six dimensional protein orientations with shape complementarity as major scoring function. Typically, during this process a degree of ‘softness’ is introduced to allow minor conformational changes (Jiang and Kim, 1991). The orientations found are then ‘filtered’ on the base of the energy function to discriminate nearly-correct orientations. The third step is the refining of the energy function introducing flexibility of sidechains and possibly backbone. The second and third steps are often combined in one single refining step (Smith and Sternberg, 2002).

The success of a program in the prediction of the structure of a docked complex is typically measured by the root-mean-square deviation (RMSD) between experimentally observed positions of heavy atoms and those predicted by the program (Cole et al., 2006). Predicted structures of protein complexes are typically presented at CAPRI (Critically Assessesment of PRedicted Interaction) a community-wide experiment where predicted arrangement of proteins are tested over the experimentally solved structure of the crystal complex (Janin et al., 2003; Janin, 2005). Although high resolution docking programs often provide correct orientations, there remain several limitations. The main problem of docking simulation algorithms is to develop a scoring function/energy that can discriminate correctly from incorrectly docked orientations and developing a search method able to find the nearly correct docked orientation (Smith and Sternberg, 2002). Therefore, new improved algorithms are needed to be developed and tested to provide better energy function making docking simulation program more reliable.
6.1.3.1 The Hex Program

One of the most common techniques used as a first step in docking simulations is the fast Fourier transform (FFT) method (Harrison et al., 1994). The FFT approach uses a Cartesian grid to search the three-dimensional space of relative protein orientations using shape complementarity. The molecules during the docking simulation have only three translational degrees of freedom and to perform the search a very large Cartesian grid is required to accommodate the different translations (Ritchie and Kemp, 2000). The Hex program (4.5 latest release) is a docking simulation program developed by Davie Ritchie and Graham Kemp at the University of Aberdeen, Scotland, UK (Ritchie and Kemp, 2000). Hex uses spherical polar Fourier correlations for the calculation modelling each molecule using a 3D parametric function based on surface shape, electrostatic charge and potential distribution. Each property is represented with a vector that is used to derive an expression of docking score as result of the six degree of freedom during a rigid body search. The docking score obtained is a function of the energy associated with the interaction which tends to be minimised to obtain nearly-correct orientations of protein in a docked complex. Another advantage of the method is the possibility to constrain the search around a particular binding site if it is known or predicted.

Hex supports three basic docking search modes that can be selected using the Search mode selection in the Docking control panel. As default, the program performs a full six dimensional search over the full rotational ranges for ligand and receptor with Full Rotation mode which is required for most docking problems. Other search modes are Ligand Translation in which the receptor is fixed and the ligand is translated over the receptor and Ligand Orbit mode, in which the ligand is translated as in the previous mode but also rotated about its own principle axis. Also, the search mode can be done using either by Shape+Electrostatic or Shape only in the Correlation type menu.

A full six dimensions steric scan at the beginning of the docking simulation is typically the most extensive part of the calculations. The Radial Filter selection can accelerate this step avoiding scanning orientations in which molecules overlap significantly or are too far apart to establish realistic interactions. The Re-entrant and Star-like filters exclude to scan orientations in which the molecules are respectively too far part or too close. Following the basic docking algorithm in Hex, the putative orientations may be refined using the Post Processing options which include Bumps+Volumes, the number
of steric clashes between non-bonded pairs of heavy atoms, *MM Energies*, single molecular mechanism energy and *MM minimisation*, Newton-like energy. The results are grouped and ordered first by energy and the lowest energy solution is made the seed orientation for the first cluster. Each cluster is composed of similar docking solutions which are in the 2 Å range of a main-chain α-C RMSD from the seed orientation.

Hex 4.5 shares in common with other dockings algorithms the same limitations described above. In addition, the rigid body search performed for predicting orientation might results in false positive conformations. However, the discrimination between near-native orientations from non-native orientations might be performed using biochemical knowledge to filter the solutions (Gabb *et al.*, 1997). In the CAPRI rounds 3-5, the prior knowledge of the binding site was sufficient to help Hex to find acceptable solution (0.96 RMSD) for the blind prediction for the docking complex cohesin-dockerin (Mustard and Ritchie, 2005).

**6.1.4 First approach mode and DeepView modelling**

The use of docking algorithms to investigate protein interactions requires knowledge of the tertiary structure of the putative interacting proteins. However, the number of structurally characterised proteins deposited in PDB is extremely low compared to the annotated primary structure of proteins present in databases (Westbrook *et al.*, 2003; Bairoch *et al.*, 2005). Different computational methods have been developed to help scientists to predict tertiary structure of proteins from their amino acid sequences. Comparative modelling is widely recognised as a reliable method to generate three-dimensional model of a target protein from its primary structure (Tramontano and Morea, 2003; Moult, 2005). An essential requirement for this method is the identification of at least one experimentally solved 3D structure of a protein homologue that is used as template.

Predicting tertiary structures by homology modelling involves four steps: a) selection of the template(s), b) alignment of the target sequence and template(s), c) model building and d) model quality evaluation (Schwede *et al.*, 2003). The quality of the predicted structures of a target protein is strictly dependent on the suitability of the template. Therefore, the most frequent source of errors in comparative modelling is inaccurate target-template alignments (Tramontano and Morea, 2003). Multiple sequence alignment is a useful tool in comparative modelling projects. Typically, the input file for
amino acid sequences of the target and the template(s) are analysed into a standard ClustalW (http://www.ebi.ac.uk/clustalw/) format. Since deposited structures are not identical to their primary structures due to missing atom coordinates, mutations or sequence tags introduced for an easier purification and crystallization, the alignment is submitted to the sequence of the actual template primary sequence extracted from the PDB template library (Arnold et al., 2006).

First approach mode (http://swissmodel.expasy.org//SM_FIRST.html) from Swiss-Model offers a simple on-line interface for building models of target proteins from their amino acid sequences (Schwede et al., 2003). The server automatically selects up to five most suitable templates that have a sequence identity of more than 25% with the submitted target sequence. Since the choice of structural templates influences the accuracy of the predicted model for the target protein, the reliability of this method decreases when the target-template pairs share less than 50% of sequence identity (Rost, 1999). In this scenario, the model obtained from first mode approach requires manual adjustments of the alignment by introducing gaps or mutations (Schwede et al., 2003).

Visual inspections and manual manipulation of the alignments can improve significantly the quality of the resulting model (Bates et al., 2001). The program DeepView-SwissPdbViewer (Guex and Peitsch, 1997) can be used to search for suitable template structures and download the corresponding PDB files directly from the DeepView server. Target sequence can be superimposed onto model templates using the integrated tools for sequence alignments and structural superimposition algorithms. In addition, DeepView allows to analyse and edit modelling project files displaying the superimposed template structures and the target-templates alignments. The program also offers the control over important modelling parameters such as placement of insertion and deletions in the context of the three-dimensional structures.
6.1.5 Use of computer-aided tools for the investigation of protein interaction between Dps proteins

Investigating protein interactions between the proteins forming the DNR/DXR PKS (Chapter 3 and 4), has led to speculate a possible association of Dps proteins in a multiprotein complex with a cylinder-like shape (Fig. 4.7). This putative structural arrangement for the PKS complex relied on strong protein interactions assessed by yeast two-hybrid system reported in Chapter 3 for the ‘minimal’ and Chapter 4 for the entire PKS. The abundance of freely available protein databases and docking simulation programs has allowed many scientists to use bioinformatics tools to identify potential interacting partners. This chapter describes the attempts of using some of the computer based facilities available for biologists to investigate the protein interactions assessed in vivo using the yeast two-hybrid system which led to the proposal of a putative arrangement of the DNR/DXR PKS complex (Fig. 4.7).

6.2 Results

6.2.1 Prediction of tertiary structures of Dps proteins

To date no three-dimensional data are available for Dps proteins as no structural studies have been carried out on the components of the DNR/DXR PKS. Different methods have been used to evaluate the putative structures of the DpsABDEY proteins involved in the strong protein interactions assessed by yeast two-hybrid system (Chapter 4).

6.2.1.1 DpsA and DpsB

The structural models for DpsA and DpsB, respectively the α and β subunits of the KS present in the DNR/DXR PKS, have been obtained by using first approach mode. The protein structures used as templates to build the model for DpsA were the chains A, C, E and G (1TQYA, 1TQYC, 1TQYE and 1TQYG) present in the deposited structure in PDB (Accession code:1TQY) which correspond to the ortholog act KSα. The proposed model covers DpsA amino acid sequence from Arg3 to Arg419 with expected similar structural characteristics including the position of the catalytic Cys169 (Fig. 6.1). As for DpsA, the structure for DpsB was modelled using as templates the 3D structures of chains B, D, F and H (1TQYB, 1TQYD, 1TQYF and 1TQYH) present in PDB (Accession code:1TQY) that represent the act KSβ. The model covers DpsB primary structure from the Arg26 to Ala424 and the highly conserved residue Gln161 in act KSβ (Keatinge-Clay et al., 2004) occupies the position 183 in the amino acid sequence of DpsB (Fig. 6.2).
Figure 6.1 Predicted three-dimensional structure for DpsA
A) 3D structure of DpsA; B) primary and secondary structure of DpsA, the amino acid sequence covered by the model is indicated in grey, in green α helix, in orange β-sheets and red active site; C) 3D structure with secondary structures in evidence; D) in evidence the active site Cys169 in helix α11.
Figure 6.2 Predicted three-dimensional structure for DpsB
A) 3D structure of DpsB; B) primary and secondary structure of DpsB, the amino acid sequence covered by the model is indicated in grey, in green α-helix, in orange β-sheets and red active site; C) 3D structure with secondary structures in evidence; D) in evidence the active site Gln183 in helix α9.
Pfam analysis of the primary structures of DpsA and DpsB revealed that they belong to the thiolase family (Pfam: PF00108) with the active site of beta-ketoacyl synthase subunits located in the central region of the polypeptides. The homodimer of DpsA was modelled in DeepView using as template the chains A and C of the deposited PDB structure 1TQY. In a similar fashion, the heterodimer of DpsA/DpsB was modelled using as template the structure of the asymmetric unit for the act heterodimer KSa/KSβ, 1TQYA/1TQYB.

6.2.1.2 DpsD
The structural model for the putative MAT, DpsD was obtained using SwissProt-DeepViewW using as template the ortholog malonyl/acyl transferase FabD from E.coli (PDB accession code: 1MLA) (Fig. 6.3). The putative model for DpsD spans the amino acid sequence from Gly14 to Ser326. Pfam analysis of the primary structure of DpsD confirmed that the protein belongs to the family of enzymes involved in polyketide and fatty acid biosynthesis (INTERPRO description entry: IPR 001227)

6.2.1.3 DpsE
The model for the tertiary structure of the ketoreductase, DpsE, was obtained by using first approach mode. The templates used for DpsE modelling were the deposited structure of the chains forming the homodimer of the act KR (PDB accession code: 1X7HA, 1X7HB, 1X3A, 1X3B). The proposed model for DpsE covers its entire primary structure (Fig. 6.4). Pfam analysis of the DpsE primary sequence revealed that the protein belongs to the family of short-chain dehydrogenases/reductases family (SDR) (Jörnvall et al., 1995) which are normally NADH/NADPH dependent oxidoreductase (INTERPRO description entry: IPR002198).

6.2.1.4 DpsY
The putative 3D structure for the cyclase, DpsY, was modelled in DeepView-SwissPdbViewer using as template the deposited structure of a predicted metal-dependent hydrolase from Bacillus stearothermophilus (PDB accession code:1R61) (Maderova et al., 2004 unpublished). This protein template was chosen by analysing the primary structure of DpsY in the Pfam database that was confirmed by analysis in ClustalW. The model for DpsY covers its amino acid sequence from Arg34 to Asn237 (Fig. 6.5).
Figure 6.3 Predicted three-dimensional structure for DpsD
A) 3D structure of DpsD; B) primary and secondary structure of DpsD, the amino acid sequence covered by the model is indicated in grey, in green α helix, in orange β-sheets and red active sites; C) 3D structure with secondary structures in evidence C) in evidence the catalytic dyad His100-Ser101 between β2 and α6.
Figure 6.4 Predicted three-dimensional structure for DpsE
A) 3D structure of DpsE; B) primary and secondary structure of DpsE, the amino acid sequence covered by the model is indicated in grey, in green α helix, in orange β-sheets and red active sites; C) 3D structure with secondary structures in evidence D) in evidence the catalytic tetrad Asn114-Ser144-Tyr157-Lys161.
Figure 6.5 Predicted three-dimensional structure for DpsY
A) 3D structure of DpsY; B) primary and secondary structure of DpsY, the amino acid sequence covered by the model is indicated in grey, in green α helix, in orange β-sheets and red putative active sites; C) 3D structure with secondary structures in evidence D) in evidence the conserved signature with the putative active residues His80-His84-Asp86-His90.
Putative homodimeric arrangement for DpsY was also built by homology modelling in SwissProt DeepView using as template the monomer of the predicted metal-dependent hydrolase. Pfam analysis revealed that DpsY belongs to a family of cyclases found in antibiotic producing clusters but also in non antibiotic producing organisms such *B. stearothermophilus*. The active site for the proteins classified in this family is thought to be the conserved signature HXGTHXDXPXH (INTERPRO description entry IPR007325). The homodimer of also DpsY was modelled in DeepView-SwissPdbViewer using as templates the chains A and B of the deposited PDB structure of the metal dependent hydrolase from *B. stearothermophilus*. The model for the homodimer of DpsY was used in all the tested docking simulation (Fig. 6.6).

6.2.2 Docking simulations

The docking simulations described in this chapter were performed by using Hex 4.5 program (Ritchie and Kemp, 2000). Docking simulation between the two subunits of the KS DpsA/DpsB was performed using DpsA as receptor and DpsB as ligand. The search mode used the default *Full Rotation* with correlation type *Shape+Electrostatic*, filtering by *Re-entrant* and with *Bumps+Volumes* as refinement. The orientation chosen amongst the first four (-608, -573, -568, -552 kJ mol⁻¹) was the fourth one (-552 kJ mol⁻¹) based upon the published crystal structure for the *act* orthologs (Keatinge-Clay *et al.*, 2004) (Fig. 6.7). Hex 4.5 failed to produce a significant simulation for the homotypic interactions of DpsA and DpsY. This problem was overcome as described before for DpsA (section 6.2.1.1) and DpsY (section 6.2.1.4).

Docking simulation between DpsA/DpsY was performed using the dimer of DpsY as receptor and DpsA as ligand. The search mode used the default *Full Rotation* with correlation type *Shape+Electrostatic*, filtering by *Re-entrant* and using *Bumps+Volumes* as refinement. A docking simulation between DpsB/DpsY was performed in a similar fashion. Hex 4.5 displayed only one solution with an associated energy of -187 kJ mol⁻¹ for DpsA/DpsY docking which interface was then analysed. The solution for DpsB/DpsY docking with the lowest associated energy (-187 kJ mol⁻¹) was chosen for analysis of the interface (Fig. 6.8). The docking between the two subunits of the KS dimer (DpsA/DpsB) and the homodimer of DpsY was also performed using the DpsA/DpsB dimer as receptor and DpsY homodimer as ligand. The search mode used the default *Full Rotation* with correlation type *Shape+Electrostatic,*
Figure 6.6 Predicted three-dimensional structure for DpsY homodimer
A) 3D DpsY homodimer; B) position of the conserved signature in each monomer (red); C) in evidence the conserved signature with the putative active sites exposed on the surface (yellow His, green Asp).
Figure 6.7 Docking simulation for DpsA/DpsB
A) Orientation of the docking with characteristic ‘grasping loop’;
B) Amphipatic tunnel where aklanonic acid backbone is built in analogy with the act PKS (Keatinge-Clay et al., 2004).
Figure 6.8 Docking simulations for DpsA and DpsB with DpsY homodimer
A) Orientation of the docking DpsA/DpsY homodimer;
B) Orientation of the docking DpsB/DpsY homodimer;
C) Orientation of the docking DpsA/DpsB heterodimer with DpsY homodimer.
with no filtering and using *Bumps+Volumes* as refinement. The best orientation with the lowest energy (-287 kJ mol⁻¹) was chosen for analysis of the interface (Fig. 6.8). Docking simulation between DpsA/DpsD was performed using DpsA as receptor and DpsD as ligand. The search mode used the default *Full Rotation* with correlation type *Shape+Electrostatic*, with no filtering and using *Bumps+Volumes* as refinement. Similar protein orientations were obtained with energies in the range of (-379 to -352 kJ mol⁻¹) (Fig. 6.9). Simulation of docking between DpsA and DpsD was also performed using DpsA homodimer as receptor and DpsD as ligand. The search mode used the default *Full Rotation* with correlation type *Shape+Electrostatic*, with no filtering and using *Bumps+Volumes* as refinement. Similar orientations were obtained with energies in the range of (-436 to -421 kJ mol⁻¹) (Fig. 6.10).

Docking simulations DpsE/DpsY, DpsD/DpsY and DpsD/DpsE were performed using search mode the default *Full Rotation* with correlation type *Shape+Electrostatic*, filtering by *Re-entrant* and with *MM energies* as refinement. For DpsE/DpsY simulation, the homodimer of DpsY was receptor and DpsE as ligand and the orientation with the lowest associated energy (-237 kJ mol⁻¹) was chosen for the analysis of the interface characteristics (Fig. 6.11). DpsD/DpsY docking was simulated using the homodimer of DpsY as receptor and DpsD as ligand. The orientation with the lowest associated energy (-459 kJ mol⁻¹) was chosen for the analysis of the interface (Fig. 6.12). DpsD/DpsE docking was simulated using DpsD as receptor and DpsE as ligand and the orientation with the lowest associated energy (-436 kJ mol⁻¹) was subsequently analysed (Fig. 6.13).

### 6.3 Discussion

In this chapter, the strong interactions between Dps proteins previously assessed by yeast two-hybrid system (Chapters 3 and 4) were analysed using the Hex software developed by Ritchie and Kemp in 2000 and successively modified up to the actual released version 4.5. Bioinformatic approaches to the study of protein associations is a relatively new field amongst biological scientist community. The completion of genomic sequencing projects from different organisms has provided annotation in databases of novel unknown DNA and protein sequences. The docking simulation programs can support scientist in the investigation of possible protein interactions involving novel unknown proteins with the aim to understand the map of interactions called 'interactome'.
Figure 6.9 Docking simulation for DpsA/DpsD
A) Orientation of the docking DpsA/DpsD;
B) Zoom in of the interface (in yellow hydrophobic residues). Two salt bridges are supposed to be involved: Arg2-Asn163 and Glu250-Arg182.
**Figure 6.10** Docking simulation for DpsA homodimer and DpsD

A) Orientation of the docking DpsA homodimer with DpsD;

B) Zoom in of the interface (in yellow hydrophobic residues). A salt bridge is supposed to be involved in the interaction (Asp294-Asn154).
Figure 6.11 Docking simulation for DpsE/DpsY
A) Orientation of the docking DpsE/DpsY;
B) Zoom in of the interface (in yellow hydrophobic residues) with the salt bridge supposed to be involved in the interaction (Lys216-Asp147);
C) In evidence the active sites of the two proteins.
Figure 6.12 Docking simulation for DpsD/DpsY
A) Orientation of proteins in the docking DpsD/DpsY;
B) Zoom in of the interface (in yellow hydrophobic residues) with two salt bridges supposed to be involved in the interaction Arg220-Asp213 and Gln133-Glu37;
C) In evidence the conserved signature of DpsY.
Figure 6.13 Docking simulation for DpsD/ DpsE
A) Orientation of protein for the docking DpsD/DpsE;
B) Zoom in of the interface (in yellow hydrophobic residues) with a salt bridge supposed to be involved in the interaction (Arg182-Asp6);
C) In evidence the active site of DpsE (red) and the regions of DpsD and DpsE involved in the interaction with DpsY (light blue).
The interactions involving DpsABDEY had been previously assessed using a yeast two-hybrid system and used to hypothesise a model for the architecture of the DNR/DXR PKS complex. The investigation of these interactions using a docking method required the prediction of three-dimensional structures as no structural information on any of the Dps components is currently available deposited in PDB. Tertiary structure predictions were performed by using essentially two different methods, first mode approach and DeepView SwissProt, from the SWISS-MODEL server.

The models for the structures of the two KS subunits, DpsA and DpsB, were predicted using as template the homolog subunits from the act PKS which share high similarity in 3D structure (Keatinge-Clay et al., 2004). The proposed 3D structures for DpsA and DpsB have similar secondary and tertiary structures with an alternating α-helix and β-sheet arrangement (αβαβα) which is also observed in thiolase from S. cerevisiae (Mathiue et al., 1994) and is similar to the corresponded condensing enzyme FabF for the type II fatty acid synthase in E. coli (Huang et al., 1998). In addition, the active site residues Cys169 for DpsA and Gln183 for DpsB, are located in correspondent positions, at the N-terminal of α11 and N-terminal of α12 respectively, of the three-dimensional structures (Fig. 6.1 and 6.2).

The DpsD model was predicted using DeepView using as template the ortholog malonylacyltransferase encoded by the fabD gene in E. coli involved in type II FAS (Serre et al., 1995). The proposed model for DpsD shows an overall structural similarity related to the α/β hydrolase superfamily with two distinctive subdomains (Keatinge-Clay et al., 2003). The large subdomain consists of an α/β hydrolase core represented by a central parallel β-sheet that is wrapped by twelve α helices some of which form a helical flap. The small subdomain is constituted by four antiparallel β-strands and two α helices resembling a ferrodoxin-like fold (Welch et al., 1998). The corresponded catalytic dyad His91-Ser92 present in E. coli FabD is located at the positions 100 and 101 in the primary structure of DpsD. The catalytic dyad lies in a nucleophilic elbow between β2 and α6 in the centre of the highly conserved signature GHSXG (Joshi and Wakil, 1971) (Fig. 6.3).
The three-dimensional structure of DpsE resulted very similar to the ortholog act KR. The overall structure is arranged in a Rossman fold (Rossmann and Argos, 1981) typical of short chain dehydrogenases/reductases (SDR) a large family of proteins that use NADPH or NADH as cofactor (Jornvall et al., 1995; Opperman et al., 2003). The coenzyme NADP⁺ binding site contains the catalytic residues, Ser144 (between β5 and α9), Tyr157(N-terminal of α8), Lys161(N-terminal of α8), and Asn114(α5) amongst other conserved residues (Korman et al., 2004; Hadfield et al., 2004) (Fig. 6.4).

The structural model for DpsY was built by homology modelling with the known structure for the putative metal-dependent hydrolase from B. stereothermophilus (Maderova et al., unpublished) using DeepView. The predicted structure revealed an internal cavity composed of eight β-sheets: the bottom of the cavity being formed by 4 antiparallel β-sheets (β5-β9-β10-β1), whereas the top is formed by the parallel strands β6-β7-β8 and β4 running in antiparallel orientation alongside with β8. Exposed on the surface there are six α helices with α3, α4, α5, α6 surrounding the top of the central cavity formed by the β-strands. Interestingly β2-β3 are the only strands not buried in the internal cavity but form a peculiar exposed structure alongside with the small helices α2 and α3 (Fig. 6.5). Each DpsY subunits exhibits a fold that belongs to the class of α/β proteins of the putative cyclase superfamily characterized by ‘swivelling’ β/β/α domains in the SCOP (Structural Classification of Proteins) database (http://scop.mrc-lmb.cam.ac.uk/scop/) (Murzin et al., 1995) (Fig. 6.5 and 6.6).

This folding is different to the tertiary structure of the TcmF2 cyclase from the tcm PKS in Streptomyces glaucescens. The 1.9 Å resolution X-ray structure revealed that this protein is assembled as a tight dimer with a βαβ ferrodoxin-like fold formed by four β strands of anti-parallel sheets and a layer of α-helices which creates a cavity thought to contain the active site (Thompson et al., 2004). In the dimer, the two subunits assemble to form an elongated structure with an extensive interface mainly constituted by juxtapositioning of the β-sheets. The orientation of the strands allows the dimer to interact tightly with a closed packed interface mainly via hydrophobic interactions (Thompson et al., 2004). A similar arrangement of subunits via ferrodoxin-like fold has also been suggested from the 1.3 Å resolution solved crystal structure of the ActVA-Orf6 monooxygenase of the act PKS (Sciara et al., 2003). As for TcmF2 cyclase, the dimer of this enzyme is assembled through the close associations of the β structures (Sciara et al., 2003).
Despite differences in the overall folding of the subunits, the central cavity formed by the β-sheets present in TcmF2 cyclase resembled the one present in the putative model for DpsY. From the structure of the dimer, it appears that the β structures present in the central core of DpsY might also be involved in the dimerisation of the two monomers. The main difference between the structures of the two cyclases is the flap composed by the β2-β3 strands that is exposed on the surface of the dimer of DpsY which is not present in the TcmF2 (Fig. 6.5 and 6.14).

The putative conserved signature HXGTHXDXPXH, also present in other cyclase such as AknW in *S. galilaeus* (Raty *et al.*, 2000), that is supposed to contain the catalytic site, is located in the turn that stretches from β3 to α3. Analysis of the location of the possible putative candidate as catalytic residues, histidine or aspartate, revealed that their sidechains are not orientated towards the internal cavity, but are exposed. This peculiar location on the surface of the protein was confirmed by the analysis of the DpsY dimer (Fig. 6.6). It might be speculated that the role of the internal cavity is not to accommodate substrates but mainly to help dimerisation of monomers via hydrophobic interactions, as observed for TcmF2 cyclase (Thompson *et al.*, 2004), whereas the catalytic reaction take place on the surface of the protein (Fig. 6.6). Since no site-directed mutagenesis studies have been reported in literature on the residues contained in the putative conserved signature HXGTHXDXPXH, it can not be ruled out that the active site lies elsewhere in the protein.

On the other hand, in the TcmF2 cyclase, the internal cavity was speculated to contain the active site with the possible putative candidates His26, Asp27, Arg40 and His51 as catalytic residues. However, site-directed mutagenesis of these residues did not show any significant decrease in enzymatic activity suggesting that the enzyme might not provide a specific catalytic residue, but instead, influences the outcome of the cyclisation reaction. It has been suggested that the side chains of these residues are involved in establish interactions with either the substrate or the water molecule providing specificity or direction for a water-facilitate reaction (Thompson *et al.*, 2004).
Figure 6.14 Comparison of the tertiary and quaternary structure of TcmF2, SnoAL, AknH cyclases and ActVA monooxygenase

The tertiary structure results very similar between each other and with the predicted structure of DpsY. TcmF2 and ActVA have a dimeric arrangement whereas the cyclases SnoAL and AknH form tetramers.
The use of β-sheets to maintain dimerisation between subunits has also been observed in solved structures for the cyclases from other type II PKS clusters (Kallio et al., 2006; Sultana et al., 2004). AknH is involved in the closure of the fourth carbon ring in aclacinomycin biosynthesis in Streptomyces galilaeus leading to the conversion of aklanonic acid methyl ester into aklaviketone (Kantola et al., 2000). SnoaL catalyses a similar reaction in the biosynthesis of nogalamycin in Streptomyces nogalater (Torkkell et al., 2000). The structure of these cyclases revealed a common quaternary organization with the monomers assembled as tetramers interacting through the hydrophobic β-structures (Kallio et al., 2006; Sultana et al., 2004) (Fig. 6.14). The topological similarity of all these enzymes involved in the late steps of the type II polyketide biosynthesis suggests that this structural organization might be suitable for enzymes that catalyse adjacent successive chemical modifications of polyaromatic substrates (Thompson et al., 2004) (Fig. 6.14).

The docking simulation for the subunits of the KS dimer DpsA/DpsB was performed in order to check whether the Hex 4.5 program was able to predict a putative orientation for the two monomers near to the native structure as revealed by the X-ray solution at 2.0 Å resolution for the ortholog act KS dimer (Keatinge-Clay et al., 2004). The program found amongst the first solution, one which showed similar characteristics to the act KSα/KSβ heterodimer. Stroud and his co-workers reported that the two KS subunits interact via tight complementary contacts that bury over 1/5 of the surface area of each monomer forming an amphipathic tunnel. The cavity at the interface of the two monomers is where the polyketide backbone is synthesised and its 17.0 Å length influences the chain length of the growing polyketide (Keatinge-Clay et al., 2004). A characteristic structure formed by the helices α8 of the KSβ and α7 of the KSα called ‘grasping loop’ is responsible to make tight interactions between the two subunits. In particular, for the act PKS two residues, Tyr118 on the KSα and the Phe116 on the KSβ, were found to be involved in establishing close interactions. The active site of the KSα, Cys169, and Phe116 are thought to represent the gating residues that regulate chain length marking respectively the beginning and the end of the amphipathic tunnel (Keatinge-Clay et al., 2004).
The docking simulation for the DpsA/DpsB heterodimer shows similar features compared to the above mentioned structure. Tyr118 of the act KSα is conserved in the same position in the amino acid sequence of DpsA whereas Phe116 of act KSβ is conserved at the position 131 of DpsB primary structure. In a similar fashion to the act system, it can be speculated that the aklanonic acid backbone is built in the tunnel which length is determined by the distance between the two residues Cys169 and Tyr118 on DpsA (Fig. 6.7).

The proposed orientation for the interaction between DpsA/DpsD showed an association of the two proteins involving the helices α14(Ala249-Arg255) and α6(Ala96-Val100), the C-terminal of α11(Ala167-Glu185). DpsD contributes to the interaction with the helix α8(Asp182-Glu193), the loop that connects the C-terminal of the α8(Gly149-Ala160) to the N-terminal of the β-sheet β4 (Pro165-Leu170) and the helix α6(Leu116-Ala133). Two salt bridges have been identified to play a role in the association of the two interacting protein: Arg4 and Glu250 of DpsA with Asn163 and Arg182 in DpsD respectively (Fig. 6.9).

DpsA was assessed by yeast two-hybrid assays to establish homotypic interaction (Chapter 3). In order investigate whether this association would have any influence on the orientations of the interaction with DpsD, a docking simulation was also performed between DpsA as a homodimer and DpsD. For this docking simulation the best orientations (-434 kJ mol⁻¹) showed that the docking region of the two proteins is slightly different than the previously assessed simulation between monomeric DpsA and DpsD (Fig. 6.9). DpsD appeared to contribute to the interaction mainly with the helix α8(Gly149-Ala160). N- and C-terminal of DpsA are involved in this interaction as well as the C-terminal of the helices α15(Ala283-Ala296) and α17(Lys314-Arg334). A salt bridge between Asp294 in α15 of DpsA and Asn154 in α8 of DpsD might also be involved in the interaction (Fig. 6.10).

The docking simulation of either DpsA or DpsB with the homodimer of DpsY displayed similar orientations with the proteins very distant from each other and very little can be speculated about these docking orientations. Interestingly, Hex 4.5 could find only one solution for the docking DpsA/DpsY and this might indicate difficulty for the program to simulate such interaction (Fig. 6.8). To overcome this apparent problem, different strategies were used. First, a docking was simulated between the homodimeric DpsA
and the homodimer of DpsY but again, the program was unable to produce a meaningful result. Second, a docking simulation was tested between heterodimer DpsA/DpsB and homodimer of DpsY was performed revealing an orientation for the proteins very similar to the one found testing monomeric DpsA with homodimeric DpsY (Fig. 6.8).

The docking simulation for the interaction between DpsY and DpsE showed an orientation in which the two proteins interact mainly via hydrophobic interactions and one salt bridge between Lys216 of DpsE, located in helix a10(Glu210-Lys220), and Glu147 of DpsY, present in a loop that connects a5(Gly143-Pro146) to a6(Ser159-Phe168), may also take part in establishing interactions. Helix a10 in DpsE is on top of the nucleotide binding site that accommodates the catalytic tetrad (Asn114-Ser144-Tyr157-Lys161) whereas Glu147 on DpsY is close to the putative consensus sequence for the active site of the cyclase. Also, the region of DpsE involved in the interaction is identical to the one proposed in the ortholog act KR for the binding of the ACP and it is close to the catalytic tetrad (Korman et al., 2004) (Fig. 6.11).

An attractive feature of this arrangement is the physical proximity of the catalytic sites of the two proteins which can be speculated as correct since the two proteins catalyse successive chemical modifications on the aklanonic acid backbone (Grimm et al., 1994). Following reduction at C9 of the 21 carbon atom polyketide, the CYC/ARO DpsF determines the closure and the aromatisation of the ring and then DpsY carries out the cyclisation of the second and third ring (Hutchinson and Colombo, 1999). The ACP, to which the aklanonic acid backbone is tethered during the biosynthesis, delivers the substrate interacting with DpsE and after interaction with DpsF to allow closure/aromatisation of the first ring, might interact with DpsY in the region close to the interface with DpsE to catalyse the second and third ring closure.

The proposed orientation for the docking between DpsD and the dimer of DpsY revealed an interesting interface between these two proteins. DpsD appeared to interact mainly via interactions established with the two helices a10(Ala212-Pro226) and a7(Leu116-Ala134) with the N-terminal of the b9(Thr211-Ala220) in DpsY. In addition, the N-terminal of the DpsY, that in the model starts with the arginine residue in position 34 in the amino acid sequence, points towards the interface with DpsD. Two salt bridges were found by analysing the nature of the interface. The first involves
Gln133 in α7 of DpsD and Glu37 present at the N-terminal of the model for DpsY. The other salt bridge is formed by the Arg220 in α10 of DpsD and Asp213 in β9 of DpsY (Fig. 6.12).

Fascinating feature of this proposed orientation includes the different region that DpsY appeared to use to establish contacts with DpsD and DpsE. Also, the conserved signature HXGTHXDXPXH, thought to contain the catalytic site(s), in DpsY is not involved in the interaction with DpsD but, indeed, it is on the opposite part of the protein (Fig. 6.15). The proposed orientation of DpsD/DpsY appears to be compatible with the orientation of DpsE/DpsY docking because the interacting region of the cyclase with the KR is different and positioned in the correct orientation for DpsD to dock with DpsY (Fig. 6.15).

The simulation for the docking DpsD/DpsE also revealed a particular and complicated interface. DpsD takes part in the interaction with the antiparallel β-sheets, β3(Leu139-Glu145) and β6(Ala199-Ser202), the α9(Arg182-Glu193) and the loop that connects α7(Leu116-Ala134), previously found implicated in the interaction with DpsY, to β3. Also, the region between β6(Ala199-Ser202) and α10(Ala212-Pro226) takes part in establishing the association. DpsE contributes to the interface with its N-terminal region, the small helix α12(Asp242-Asp245), the C-terminal of the helix α11(Pro229-Ala240) and the loop between helix α1(Gly17-Asp29) and β2(Arg32-Ala37). A salt bridge could be detected in the close proximity of the two proteins formed by Arg182 on DpsD and Asp6 on DpsE. Hydrophobic amino acid residues are also involved in establishing the interaction (Fig. 6.13). Intriguing characteristic for the docking DpsD/DpsE is the putative orientation of the two proteins that appears to be compatible with the docking orientations for DpsD/DpsY and DpsE/DpsY (Fig. 6.15). Also, in the interaction DpsD/DpsE the KR domain that contains the catalytic tetrad (Asn114-Ser144-Tyr157-Lys161) is not involved in the interaction with the MAT neither is the previously described α10 helix involved with the interaction with DpsY.
Figure 6.15 Comparison of the orientation of proteins for DpsD/DpsY and DpsE/DpsY docking simulations. DpsD and DpsE interact with different regions of DpsY. In red are indicated the putative catalytic region of DpsY and the nucleotide binding site of DpsE.
Although all docking simulation algorithms, including Hex 4.5 used in the chapter, share the same limitations such as the incapability to produce an energy function to discriminate near-native from non-native orientations, they might represent a valid approach to predict-protein interactions. Computational methods have been previously used to guide experimental designs that have allowed identification of the recognition helix on the ACP of type II FAS in *E. coli* and the common features on its target partners responsible for protein interactions (Zhang et al., 2003b; Hadfield et al., 2004). During this study, the analysis of the docking simulations has revealed interesting molecular insights for the interactions of the Dps proteins in particular for the interaction involving DpsE, DpsD and DpsY.

The solved structures of homologues and the classification in protein families for the Dps proteins has represented the biochemical knowledge that is essential for the critical analysis of putative orientations obtained by docking simulations (Gabb et al., 1997) (Mustard and Ritchie, 2005). The importance of the biochemical knowledge to filter docking solutions has emerged from the simulation of DpsA/DpsB interaction where the solved structure of the *act* heterodimer KSα/KSβ has allowed the selection, amongst the orientations with the lowest energies, of an orientation similar to the one reported for the *act* KS heterodimer (Keatinge-Clay et al., 2004) (Fig. 6.7). For the other proposed orientations, the lowest energy and/or the previous assessed results from the yeast two-hybrid system were used to choose the putative best solutions.

Some of the proposed orientations involving DpsD were of obscure meaning in particular the interaction with DpsE and DpsA, in which both appear to involve the small subdomain with ferredoxin-like fold. It might be speculated that this region represent a favourable spot for establishing interactions and by using a docking simulation it can not be discriminated whether this domain is involved in the interaction with DpsE or DpsA. The nature of the interaction of DpsD with DpsA could have elucidated better the arrangement of the protein forming the ‘minimal’ core complex, DpsABD. However, using monomeric and homodimeric state DpsA, different regions of this protein were found to be involved in the interaction with DpsD. Despite these different outcomes, the predicted regions of DpsA did not result the same involved in the docking with DpsB. Elucidation of the mechanism of action of the putative MAT might give new insight in the interactions established by the protein.
The combination of orientations for docking DpsD/DpsE and DpsD/DpsY and DpsE/DpsY appear to be intriguing for a mechanistic explanation of the last chemical modification steps of the aklanonic acid backbone. Also, it might be important for a better understanding the architecture of the DNR/DXR PKS complex (Fig. 6.13, 6.15 and 6.16). It can be speculated that DpsD in the complex has a structural role to establish contacts with DpsY and DpsE (Fig. 6.16). The position of DpsD in the orientations of the docking simulation with DpsE and DpsY might be important to position the active site of DpsE close to the DpsY putative catalytic domain. This orientation has been confirmed by the docking simulation DpsE/DpsY and also it would have been expected as the DpsE and DpsY catalyse adjacent reactions in conjunction with DpsF (Hutchinson and Colombo, 1999). In addition, the combination of the three docking simulations DpsD/DpsE, DpsD/DpsY and DpsE/DpsY appear to be compatible with the orientations of the single proteins suggested in the putative model for the architecture of the DNR/DXR PKS (Fig. 4.7, 6.15 and 6.16).

The particular structural role of DpsY suggested by the yeast two-hybrid analysis (Chapter 4) could not be well defined using the investigated docking orientations. This is mainly due to the unclear orientations found for the docking with either the monomeric forms of DpsA and DpsB or with the heterodimer DpsA/DpsB or DpsA homodimer. A limitation of the analysis herein described regards the structure of DpsY used for the docking simulations. Due to the lack of better templates available in the database to use for homology modelling, the predicted structure for the cyclase could only be modelled using the smaller hydrolase from *B. stearothermophilus* not including thirty amminoacids residues both at the N- and C-terminal that might be important to play a role in establishing interactions with other Dps components. However, the putative 3D model for the DpsY homodimer showed that the cyclase presents hydrophobic residues on its surface and this might be important either to associate other Dps proteins for the assembly of the DNR/DXR PKS complex or to localize the complex to an hydrophobic structure such as the cellular membrane which has been hypothesised by a previous study for the *tcm* and the *act* PKSs (Gramajo et al., 1991; Hesketh and Chater, 2003) (Fig. 6.16).
Figure 6.16 Analysis of the surface of DpsD and DpsY
A) DpsD involves different structures for establishing interactions with DpsY and DpsE.
B) The surface of DpsY presents uniform distribution of hydrophobic residues (yellow). Hydrophobic residues on the β-strands in the central cavity help dimerization of the subunits. Hydrophobic patches on the surface might be important to associate other Dps components or to localize the complex to the membrane.
CHAPTER 7

Concluding Remarks

Type II PKSs are the focus of many research projects as they are responsible for the biosynthesis of an extraordinary variety of compounds with biological activities useful for agriculture and human health (Hopwood, 1997). The 21 carbon atom decaketide aklanonic acid, produced by the DNR/DXR PKS, represents the target of many studies since it is the precursor of the anticancer drugs daunorubicin and doxorubicin (Lee et al., 2005). Despite the efficacy against different kinds of cancer, their use is limited due to their side effects. Several efforts are being conducted to synthesise new analogues of these anthracyclines that could have improved biological activity with fewer side effects (Minotti et al., 2004). One of the most powerful methods to achieve this task could be to harness the combinatorial potential of type II PKSs. Type II PKSs are formed by monofunctional proteins that associate in a complex and the characteristics of the molecules produced by these systems appear to be highly dependent more on the interactions established within the complex than on the specific enzymatic activity of the single proteins (Shen et al., 1999; Peric-Concha et al., 2005) as demonstrated also by 'contest-dependent behaviour' (Meurer et al., 1997; Kramer et al., 1997). Important features of the type II PKSs have been elucidated over the past decade but up to date the complexity of these systems has not allowed scientists to fully understand their mechanism of action such as determination of chain length and folding pattern of the final product (Hertweck et al., 2007) The work presented in this thesis aimed to extend our knowledge of the overall structure of the type II PKSs complexes by investigating interactions between the eight proteins that form the DNR/DXR PKS, using a broad range of approaches many of which have never been applied to this problem.

One of the main goals of this work was to investigate of protein interactions in vivo using a GAL4 yeast two-hybrid system. Protein interactions involving the components of the ‘minimal’ DNR/DXR PKS, DpsABCDG, were initially analysed testing all the binary combinations for possible interactions. The outcome of this investigation revealed intriguing results. DpsA and DpsB, were found to establish strong interactions as previously described for the orthologs act and tcm KSα/KSβ heterodimers (Keatinge-Clay et al., 2004; Gramajo et al., 1991; Carreras and Khosla, 1998) and
strong homotypic interaction was revealed for DpsA but not for DpsB. The KSIII, DpsC, established weak interactions only with DpsG which, in turn, was found to interact transiently with all the components of 'minimal' PKS. DpsD, which role in the aklanonic acid biosynthesis in S. peucetius appears to be obscure (Grimm et al., 1994), showed strong and transient associations with DpsA and DpsB respectively.

These results led to hypothesise the association of DpsA, DpsB and DpsD in an oligomeric structure stabilised by strong and transient interactions. The putative structure might be considered as a dimer of a heterotrimer with a head-to-tail arrangement. DpsA establishes homotypic and heterotypic interactions with DpsB and DpsD which are aligned in opposite orientations but interact weakly with the opposite partner (Fig. 3.18). This structure with a head-to-tail arrangement can be correlated to the metabolic pattern of a heterologous host expressing different sets of DNR/DXR PKS genes (Rajgarhia et al., 2001) but more importantly suggests a function for DpsD which is dispensable for the in vivo production of aklanonic acid in S. peucetius (Grimm et al., 1994). DpsD might suppress decarboxylation of malonyl-DpsG by DpsB preventing aberrant loading of acetate onto DpsA in the priming reaction, due to its physical position in the head-to-tail structure although a previous study has suggested a specific enzymatic role as acetyltiosterase to purge acetate starter deriving from acetyl-CoA (Tang et al., 2004b). The only interaction found involving DpsC with DpsG corroborates the suggestion of a KSIII function for this protein (Rajgarhia et al., 2001) although it can not be ruled out a function as acyltransferase (Bao et al., 1999a).

Successively, the analysis of the protein interactions involving the entire set of Dps proteins was carried out in a similar fashion using a GAL4 yeast two-hybrid system. DpsY showed the most interesting interactions. This cyclase was found to establish strong homotypic and heterotypic interactions with all the components of the 'minimal' PKS and also with DpsE. The biological significance of these results might be explained by suggesting a structural key role for DpsY in the scaffolding of the complex. The DNR/DXR PKS complex might have a cylinder-like architecture with the cyclase representing the barrel of the cylinder surrounding the internal core composed of DpsA, DpsB and DpsD. DpsE is located at both ends of the tunnel as suggested by the strong interaction with DpsY and DpsD (Fig. 4.7).
The intriguing features of the proposed architecture might elucidate how molecular mechanism for chain length determination and successive modifications of the polyketide backbone are achieved. The nascent polyketide might be synthesised in the internal tunnel formed by DpsA and DpsB, as also suggested by the structural analysis of the ortholog act KSα/KSβ (Keatinge-Clay et al., 2004), with DpsD ‘proofreading’ of the chain initiation reaction. The internal tunnel also might offer an optimal environment to stabilise the highly reactive poly-β-keto chain from spontaneous aberrant cyclisation reactions (Harris and Harris, 1986). Once the polyketide chain reaches composite length of 21 carbon atom, it protrudes from the tunnel and becomes exposed to DpsE, DpsF and DpsY that carry out further chemical modifications (ketoreduction, aromatisation and cyclisations) for the production of aklanonic acid (Hutchinson and Colombo, 1999). The other components, DpsCGF, of the DNR/DXR PKS might be in close proximity to the complex although interacting only transiently. The ‘workhorse’ role of the ACP to shuttle the substrate to all the different proteins was confirmed by the transient interaction established by DpsG with all the Dps proteins.

Protein interactions assessed using yeast two-hybrid system, such the strong association DpsA/DpsB and the transient associations involving DpsG, were expected and supported by the already known mechanism of action of their homologues (Keatinge-Clay et al., 2004; Gramajo et al., 1991; Carreras and Khosla, 1998; White et al., 2005). However, the investigation failed to detect some predicted homotypic interactions for DpsG and DpsE. The acyl carrier protein is speculated to establish transient homotypic interaction as suggested by the observed inter-transfer of malonate between the tcm PKS ACP and the E. coli FAS ACP (Zhou et al., 1999). On the other hand, the construction of a chimeric protein in this study, where DpsG is fused to the much bigger polypeptides, GAL4 DBD and TAD (~ 17 kDa each), and the particular acidic nature of the surface of the ACPs, might have impaired the ability to detect homotypic interactions. As the nature of the surface of the ACP is mainly negatively charged, hydrophobic interactions involving the phosphopantetheinyl arm of the holo-enzyme might be important for establishing homotypic interactions. The conversion of the apo-ACP to the active holo-ACP is achieved post-translationally by the enzymatic addition of the 4'-phosphopantethenyl moiety from coenzyme A onto Ser42 catalysed by a specific PPTase (Lambalot et al., 1996). During this study, the investigation of protein interactions involving the ACP were performed without co-expressing in the host.
S. cerevisiae, any PPTase and, therefore, it is reasonable to speculate that DpsG was present as apo-protein. This problem might be overcome by co-expressing a specific PPTase in S. cerevisiae that could convert the ACP to its active phosphopantetheinyl form during the yeast two-hybrid assays.

The solved crystal structure of the act KR, ActIII, revealed a tetrameric arrangement for the subunits (Hadfield et al., 2004) but despite the high similarity with DpsE, in this study no homotypic interactions were detected for the dps KR. The fusion of the GAL4 DBD and TAD might have had influenced the native folding of the protein impairing the possibility to establish homotypic interaction. To verify whether the N-terminal of DpsE where the GAL4 domains were fused, was involved in establishing homotypic interactions, the deposited PDB file (code: 1X7H) for the ortholog ActIII was visually inspected by using DeepView. This analysis did not reveal any involvement of the N-terminal of the act KR in the interaction and, therefore, it can be excluded that the location of the GAL4 domains was directly responsible for failure in detecting homotypic interactions for DpsE. The problem due to the presence of the GAL4 domains at the N-terminal of DpsE might have impaired detection of other interacting partners and/or assessment of the strength of the interactions.

The study subsequently moved on, in Chapter 5, to investigate protein interactions on the ‘minimal’ DNR/DXR PKS by using Tandem Affinity Purification method that allows identification of all the interacting partners of the target protein in one goal. From the yeast two-hybrid system assays, DpsA was found to have a key role for the head-to-tail structure and, therefore, was chosen as target protein to fuse the TAP tag. The ‘pull down’ experiments resulted in the purification of the two subunits of the KS, DpsA and DpsB, but failed to co-purify also DpsD which, from the previous yeast two-hybrid system results reported in Chapter 3, was predicted to interact strongly with DpsA.

Several reasons might be used to explain the failure of detecting DpsD at the end of the TAP purification. Firstly, TAP tag might affect the binding of DpsD to DpsA, secondly, the association DpsA/DpsD might be sensitive to the conditions, although mild, used during the purification steps and, thirdly, the interaction DpsA/DpsD might rely on other components of the complex such as DpsY. Different approaches to resolve these problems have not been undertaken during this study due to time limitations, however,
successive studies might investigate: a) fusion of the TAP tag either at the C-terminal of DpsA or to DpsD, b) modifications of the conditions used in the purification procedure, c) fusion of the TAP tag to DpsY to try to pull out the entire complex as the cyclase was found to interact with all the Dps components. This study aimed to verify the suitability of this method to PKS systems. The ‘minimal’ PKS was chosen to be analysed by using TAP method since contains fewer proteins and, therefore, it can be considered a simpler model of PKS complexes. Purification of the entire PKS complex might be achieved targeting the cyclase DpsY which was assessed by yeast two-hybrid assays to interact strongly with most of the Dps proteins. Purification of the DNR/DXR PKS, by using TAP method coupled with mass spectrometry analysis might give a better understanding of the overall architecture of the multiprotein complex.

Structural investigation of the Dps proteins and docking simulations were undertaken in the attempt to elucidate conformation of the proteins and predict the docking orientations of proteins involved in the previously observed interactions. In particular, the strong interactions assessed by yeast two-hybrid system involving the entire PKS complex were investigated by a in silico docking simulation program, Hex 4.5 (Ritchie and Kemp, 2000). The study has resulted with the proposal of three-dimensional structures for DpsABDE and Y. The 3D models for the two subunits of the KS, the putative MAT and the KR, resembled the tertiary structure of the orthologs from act PKS or E. coli FAS II clusters as expected from the high similarity in their primary structure. The 3D model of cyclase DpsY was built by using as template the structure of a predicted metal-dependent hydrolase from B. stearothermophilus (Maderova et al., unpublished) belonging to the same Pfam family since the identity score in amino acid sequence with other PKS orthologs resulted very low. Interestingly, the model structure shows a homodimer with similar tertiary and quaternary structures of the TcmF2 cyclase from tcm PKS and the ActVA-Orf6 monooxygenase from act PKS (Sciara et al., 2003). A similar conformation for the DpsY monomer was also found for AknH and SnoaL cyclases from the biosynthetic pathway of aclacinomycin and nogalamycin respectively (Kallio et al., 2006; Sultana et al., 2004). However, these cyclases show an arrangement as tetramers rather than dimers. The common structural organization shared by these proteins might be required to catalyse successive chemical modifications in the last steps of the biosynthesis of aromatic polyketides (Thompson et al., 2004).
The docking simulations indicated an intriguing disposition for the proteins DpsD, DpsE and DpsY. DpsD in the complex might have a structural role in the PKS complex to bring the active sites of DpsE and DpsY in close proximity allowing proteins to carry out adjacent modifications of the aklanonic acid backbone in conjunction with DpsF (Hutchinson and Colombo, 1999). The docking program was able to predict a docking orientation for DpsA/DpsB similar to the observed for the solved structure of the ortholog act KSα/KSβ heterodimer (Keatinge-Clay et al., 2004). However, Hex 4.5 did not give clear indication of orientations for homotypic interactions involving DpsA and DpsY, and for the interactions between DpsY and the KS subunits, DpsA and DpsB. Also, the docking simulations for interactions between DpsA and DpsD using monomeric or homodimeric state DpsA gave different outcomes for the putative orientations of the two proteins. The docking simulations for DpsA/DpsD resulted in different regions of DpsA involved in the interaction depending on the homodimeric or monomeric state of the KSα. These regions did not include the part of DpsA involved in establishing interactions with DpsB and, therefore, this is an encouraging support to the putative arrangement of proteins DpsABD suggested by the yeast two-hybrid results. However, the proposed head-to-tail organization for the ‘minimal’ PKS could not be verified by using Hex 4.5 since this would have involved the docking simulation of two heterotrimers each composed of the DpsABD proteins.

The docking simulations also revealed an intriguing result for the DpsA/DpsD association that might explain the failure of co-purifying DpsD together with DpsA and DpsB by TAP method. The docking simulation performed between the DpsA homodimer and DpsD resulted in a putative interface between the two proteins that involves also the N-terminal of DpsA. The choice of fusing the TAP tag to the N-terminal of DpsA was led by initial analysis of the ortholog act KSα/KSβ heterodimer prior the cloning strategy for the TAP method. The study of the deposited structure for the heterodimer revealed that the N-terminal of KSα was sufficiently exposed and not involved in the association. However, while the N-terminal fusion of the TAP tag could not have influenced the interaction with DpsB, might have influenced the binding of DpsD that was not detected at the end of the purification method. However, the involvement of the N-terminal region of DpsA in the binding of DpsD was only suggested by the docking simulation performed by using DpsA as homodimer but not when used as monomer. While caution must be used to interpret the data from the docking simulations because of the difficulty in confidently distinguishing biologically
relevant interactions from artefacts, the yeast two-hybrid system results and biochemical knowledge of the Dps proteins homologues were used for the critical analysis of the docking orientations suggested using the program Hex 4.5. Despite the intrinsic limitations relative to the methods used in this study, the results herein described might represent an important step forward into the knowledge of the DNR/DXR PKS but also for all the type II PKSs clusters.

This study represents the first systematic and comprehensive investigation in vivo of protein interactions on a type II PKS. The analysis of the protein interactions involving the DNR/DXR PKS has given new insight of the architecture of the complex that is responsible for the biosynthesis of aklanonic acid. In particular, an intriguing role for DpsD is proposed as steric inhibitor of the decarboxylation of malonyl-CoA in the priming reaction and also to align in close proximity the active sites of DpsY and DpsE. Lack of DpsD still results in aklanonic acid production in vivo as it might be replaced by a homolog from the type II FAS (Revill et al., 1995) but the starter selection results promiscuous with the incorporation of also acetate rather than the usual propionate. A scaffolding role of DpsY also has been suggested for the assembly of the entire PKS complex. The proposed arrangement for the DNR/DXR PKS supports previously described features of the type II PKSs, such as act PKS, but more importantly contributes to a better understanding of the overall architecture of such complexes. The ultimate definition of the architecture and stoichiometry of the complex could be obtained by purification and analysis of the entire complex by mass spectrometry and crystallographic techniques. Also, this work emphasises the importance of protein interactions in type II PKS complex on the biosynthetic mechanisms that lead to the production of the final product. The outcome of this study should give a fresh impetus to future combinatorial biosynthetic approaches to help biologist and chemists to generate new daunorubicin/doxorubicin analogues with less toxicity while retaining biological activity.
REFERENCES


Bentley, S. D. and and Pathogen Sequencing Unit, The Sanger Center Wellcome Trust Genome Hixton Cambridgeshire UK. The Streptomyces coelicolor A3(2) genome project at the Sanger Centre. Abstract and Presentation at ISBA, Vancouver, Canada. 2001.


229


