Cytochrome P450 monooxygenases from extremophiles

by

Walter Joseph Müller

Submitted in fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

In the

Department of Microbial, Biochemical and Food Biotechnology

Faculty of Natural Sciences

University of the Free State

Bloemfontein

Republic of South Africa

January 2012

Promoter:	Prof. M.S. Smit
Co-promoters:	Prof. E. van Heerden
	Prof. J. Albertyn
	Dr. L.A. Piater

DECLARATION

I declare that this thesis hereby submitted by me for the Doctor of Philosophy degree at the University of the Free State is my own independent work and has not previously been submitted by me at another university/faculty. I further cede copyright of the thesis in favour of the University of the Free State.

_____

Walter Joseph Müller (2000016416)

January 2012

This thesis is dedicated to my father. He has sacrificed so much so that I could have so much more…

ACKNOWLEDGEMENTS

I sincerely wish to express my gratitude to the following people and organizations who contributed to this thesis. It may take a village to raise a child but it takes an army of colleagues to launch, conduct and complete a Ph.D.-study! With many thanks to:

- Prof. Martie Smit for her insatiable curiosity and passion for cytochrome P450 monooxygenases.

- Prof. Esta van Heerden for providing a wonderfully conducive research environment and for financial support.

- Dr. Lizelle Piater for keeping her finger on the pulse of this study even after her departure from our group.

- The National Research Foundation (South Africa) and TIA/BIOPAD for continued financial support.

- Dr. Matthias Schlesner (Max Planck Institute of Biochemistry, Martinsried, Germany) for providing the *Halobacterium salinarum* R1 strain, protocols, suicide vector and helpful advice.

- The research group of Prof. Shiladitya DasSarma (UMBI, University of Maryland, Baltimore, USA) for their warmth and sharing of knowledge during my research visit. A special word of thanks to Melinda Capes for her assistance during the microarray experiments but most of all for allowing me to be her shadow during my stay there.

- Prof. Tom Kieft (New Mexico Institute of Mining and Technology, New Mexico, USA) for the generous gift of the *Thermus* sp. NMX2.A1 strain.

- Charlene Randall and Abhita Jugdave for reading and critically evaluating this thesis.

- All the members of the EXBOC research group for their support and friendship – especially Robert Jordan, Abhita Jugdave, Suzanne Litthauer, Wilmari Meyer and Karin Botha.

- Dr. Ramakrishna Gudiminchi, Newlande van Rooyen, Ruan Ells, Chantel Swart, Du Toit Schabort and Antonie Meyer for their friendship, advice and support.

- My family who had to make so many sacrifices and who had to endure my absence so that I could pursue this study: thank you to my father (Joe), mother (Engela), twin sister (Beaula) and brother (Stefan) for understanding although it was not always easy to do so. I love all of you dearly.

- Thank you to my extended family which includes the Thompson-, Burger- and the Theocharous family. Thank you for your love - it kept my grounded.

- Eleanor Fourie for looking after my heart while I was looking after my career. Thank you for being my angel, my friend and my rock. I am so blessed to have you in my life.

Lastly to the Heavenly Father: Thank you Lord for providing me with health, wisdom, motivation, perseverance and inner strength but most of all thank you for Your mercy and love.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

%	Percentage
^{o}C	Degrees Celsius
16S rRNA	Small subunit ribosomal ribose nucleic acid
bat	bacterioopsin gene activator
blh	brp-like homolog
bop	bacterioopsin gene
brp	bacterioopsin related protein
brz	bacteriorhodopsin-regulating zinc finger protein
BLAST	Basic local alignment search tool
bp	Base pairs
BO	Bacterioopsin
BR	Bacteriorhodopsin
cDNA	complimentary deoxyribose nucleic acid
CO	Carbon monoxide
CYP450	Cytochrome P450 monooxygenase
DNA	Deoxyribose nucleic acid
DSMZ	Deutsche Sammilung von Mikroorganismen und Zelkulturen
EDTA	Ethylenediaminetetraacetic acid
Fdx	Ferredoxin
FNR	Ferredoxin reductase
GDH	Glucose dehydrogenase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IS	Insertion element
kDa	kilo Dalton
LB	Luria Bertani
Mb	Mega bases
PM	Purple membrane
µL	microliter
min	minute

mL	milliliter
mM	Millimolars
MOPS	3-(N-morpholino) propanesulfonic acid
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National centre for biotechnology information
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
p.s.i	pound per square inch
RNA	Ribonucleic acid
r.p.m	revolutions per minute
s	seconds
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TLC	Thin layer chromatography
Tris	2-Amino-2-(hydroxymethyl)-1, 3-propandiol
TYG	Tryptone, yeast extract, glucose
UV	Ultraviolet
V	Volts
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranosidehosphate

The noblest pleasure is the joy of understanding.

-Leonardo da Vinci-

Chapter 1

Cytochrome P450 Monooxygenases from Extremophiles – A Literature Review

1. Introduction

Often in the not so distant past, scientists have imposed anthropogenic views on biology to conform to their parameters of what life is and where life is possible. Biological niches, with previously considered insurmountable physical and chemical barriers that were thought to be non-conducive to life, are now considered home to several unique and fascinating extremophiles (Rothschild & Mancinelli, 2001; Cavicchioli, 2002). Extremophiles are living organisms found in all three kingdoms of life that not only tolerate their extreme environments but also flourish under these "inhospitable" conditions that define their environments. Extremophiles thrive in almost every conceivable niche on earth: ice, boiling water, acid, the water core of nuclear reactors, desiccated salt crystals, volcanoes, beds of ultra-deep oceans and toxic waste (Ferreira *et al*., 1997; Madigan, 2000; Cavicchioli, 2002; Seckbach & Oren, 2004).

The cornerstone of traditional microbiology is *in vitro* culturing and the study of microorganisms as axenic cultures. This form of microbiology is very workable when one moves within this particular framework in the laboratory, but to culture and study most extremophiles is quite a different matter.

Our limited definition of where life is possible and the fact that extremophiles pose a major challenge to culture in the laboratory contributed to the fact that the study of extremophiles has been neglected in the past. Donn Kushner (1978) made mention of this when he published one of the very first books devoted entirely to the biology of extremophilic microorganisms:

‘‘Indeed, many organisms that live in extreme environments have been unfairly neglected, partly because of the difficulty in studying them and obtaining publishable results. Admittedly, it is trying to study microorganisms whose growth media fills the laboratory with steam, or the centrifuge heads with salt, or which grow so slowly that

weeks, instead of hours, may be required for experiments and whose genetics are unknown or almost impossible to study. Those who have persisted have found their rewards, both in the satisfaction and leisure for contemplation available to the student of an out-of-the-way field, and in the fascination afforded by the microorganisms themselves and the very clever ways they have found to adapt to such a wide range of environmental conditions.''

Fortunately, in this day and age scientists have made wonderful advances in the field of genomics and especially metagnomics which, in most instances, circumvents the need to culture microorganisms. Consequently the study of extremophiles and their uniquely "adapted" proteins have become much more accessible. An example of proteins that were unexpectedly discovered in extremophiles *via* the genomics approach are cytochrome P450 monooxygenases (CYP450s). CYP450s are found ubiquitously in eukaroytes and bacteria but the first CYP450 from an extremophile was discovered by accident in the hyperthermo-acidophilic archaeon *Sulfolobus solfataricus* when a gene library was screened for a thymidylate synthase gene (Wright *et al*., 1996). Given the fact that the temperature and pH optima of most *Sulfolobus* strains are 75 – 80˚C and 2.0 – 3.0 respectively, the discovery of a CYP450 in an extremophile such as *S. solfataricus* was a surprise, since CYP450s are notoriously unstable (Urlacher *et al*., 2004; Munro *et al*., 2007) and usually require co-factors like NADH and NADPH - both of which are sensitive to high temperature and acidic pH levels (Wu *et al*., 1986).

Currently, there are several known CYP450 genes from extremophiles, which have been discovered (mostly) by whole genome sequencing projects but the physiological role of many of them remain unknown. This literature review will give an overview of CYP450s in general before discussing in detail the handful of described CYP450s from extremophiles. Topics pertaining to crystal structure, redox partners and the CYP450's physiological role in their native hosts will be covered. The CYP450s from extremophiles that are discussed are:

- The CYP119s from the hyperthermo-acidophilic archaea *Sulfolobus acidocaldarius* and *Sulfolobus tokodaii*

- CYP231A2 from the thermo-acidophilic archaeon *Picrophilus torridus* and

- CYP175A1 from the thermophilic, gram negative bacterium *Thermus thermophilus* HB27.

1.1 General aspects of cytochrome P450 monooxygenases

The cytochrome P450 monooxygenases (CYP450s) constitute a highly diversified, ever growing superfamily of soluble and membrane-bound heme-thiolate proteins that are distributed in all three domains of life (Lewis, 1996; Momoi *et al*., 2006; Urlacher & Eiben, 2006). CYP450s catalyze the following general reaction:

$$RH + O_2 + NAD(P)H + H^+ \longrightarrow ROH + H_2O + NAD(P)^+$$ (Bernhardt, 2006).

CYP450s contain a heme (iron-protoporphyrin IX) prosthetic group that is the active center for catalysis (Schneider *et al*., 2007). In addition, the heme iron is also coordinated to the thiolate of the absolutely conserved cysteine residue that acts as the fifth ligand. Resting CYP450s are in the ferric form and partially six-coordinated with a molecule of solvent (Werck-Reichhart & Feyereisen, 2000). CYP450s contain b-type heme (Fig. 1.1) *i.e*. the heme is non-covalently bound to the protein. Heme diversity (*e.g*. a-type, c-type and d_1-type heme) essentially arises due to the manner in which the vinyl- and methyl groups are linked to the overall protein molecule (Schneider *et al*., 2007).

The heme group in CYP450s comprises four pyrrole rings linked by four methyl bridges (α, β, γ, δ) that form a tetrapyrrole ring. Pyrrole rings I and IV each carries a methyl- and a propionate group while pyrrole rings II and III each carries a vinyl- and methyl group. The ferric or ferrous iron is coordinated by four pyrrole nitrogens (Schneider *et al*., 2007).

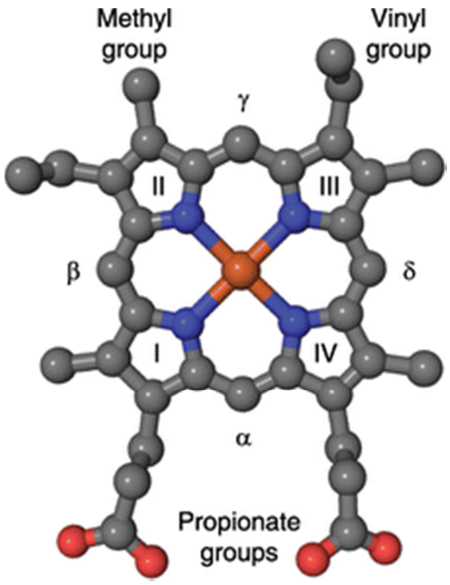


Fig. 1.1 Ball and stick model of b-type heme. Heme comprises four pyrrole rings (I – IV) linked by methyl bridges (α, β, γ, δ) that form a tetrapyrrole ring. Pyrroles I and IV carry two propionate groups and the ferric or ferrous iron (orange) are coordinated by four pyrrole nitrogens (blue) (Schneider *et al*., 2007).

Carbon monoxide (CO) is able to bind to the sodium dithionite-reduced ferrous iron of the heme to yield a CO-bound complex that displays a typical absorption maximum, with a characteristic Soret peak at 450 nm. This unusual spectral property was first described for the red pigments of rat liver microsomes (Klingenberg, 1958) and these hemo-proteins were called ‘P450’ - ‘P’ indicating ‘pigment’ and ‘450’ indicating the wavelength of the absorption maximum of the CO-bound complex. Figure 1.2 provides an example of such a spectrum. Bound CO causes inhibition of CYP450 activity which can be reversed by light, with maximum efficiency at 450 nm. Binding of other ligands, substrates or inhibitors induce absorbance shifts of the Soret peak in CYP450s. Consequently these spectral properties have given rise to differential spectrophotometry which can be used to monitor and assess the binding of ligands in the CYP450 active site. For example: substrates that displace the six-coordinated solvent in resting (ferric state) CYP450s, usually induce a spectral shift from 420 nm to 390 nm *i.e*. to the blue region of light. This is an indication of the low- to high-spin transition of the iron (Werck-Reichhart & Feyereisen, 2000).

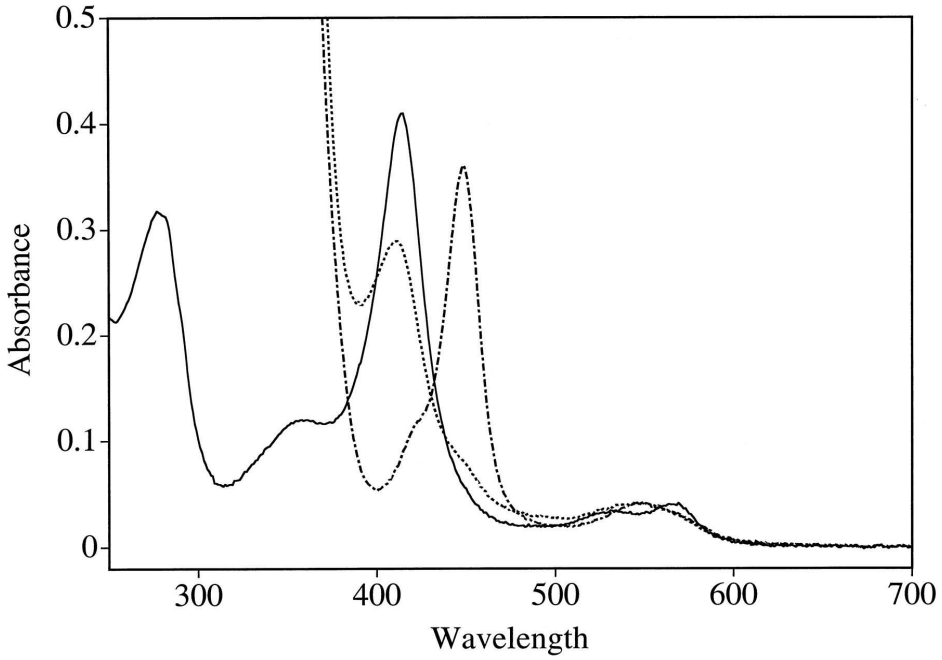


Fig. 1.2 Carbon monoxide difference spectra of the CYP119 from *Sulfolobus solfataricus*. ***Solid line***: substrate-free ferric protein, ***Dashed line***: dithionite-reduced ferrous protein and ***Dashed and solid line***: ferrous CO complex (note the Soret peak at 450 nm). (Adapted from: Koo *et al*., 2000).

Monooxygenases are divided into two classes namely: internal and external. Internal monooxygenases extract two reducing equivalents from the substrate to reduce one atom of dioxygen to water, whereas external monooxygenases utilize an external reductant (Bernhardt, 2006). Figure 1.3 illustrates the assignment of CYP450s into enzyme groups, eventually being classified as external monooxygenases (Hannemann *et al*., 2007). CYP450s are external monooxygenases that catalyse the incorporation of a single atom of molecular oxygen into X-H bonds (X: -C, -N, -S) with the concomitant reduction of the other oxygen atom to water (Hannemann *et al*., 2007).

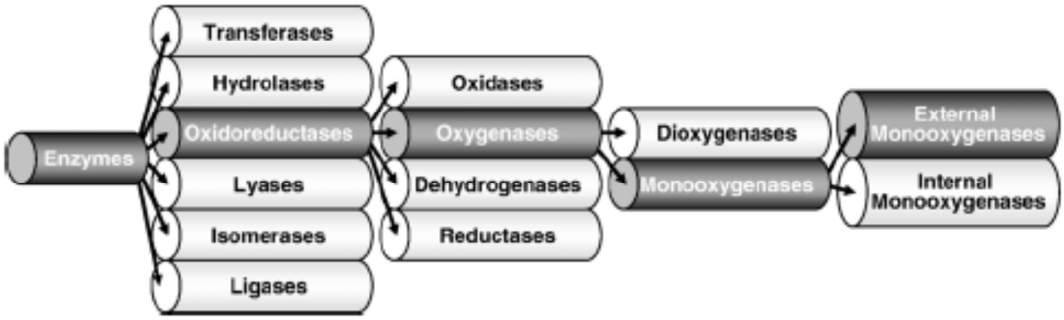


Fig. 1.3 Assigning CYP450s to enzyme groups (Hannemann *et al*., 2007). CYP450s belong to the dark grey colored subdivisions.

1.1.1 CYP450s have common protein architecture

Sequence identity among CYP450 proteins is often extremely low (in some cases < 20%), they have broad substrate ranges and catalyze a plethora of chemical reactions. However, in spite of all of this, CYP450s all share a general topography and structural fold which is highly conserved. This general topography hints at a common mechanism of oxygen activation (Werk-Reichhart & Feyereisen, 2000; Hannemann *et al*, 2007). However, CYP450s also possess highly variable regions that represent their flexible substrate recognition regions and thus their consequent versatile ability to attack an enormous variety of substrates (Hannemann *et al*., 2007). The highest structural conservation is found in the core of the protein around the heme that reflects a common mechanism of electron and proton transfer as well as oxygen activation. The conserved core comprises a four-helix bundle (D, E, I and L), helices J and K, two sets of β-sheets and a coil called the 'meander' (Fig. 1.4).

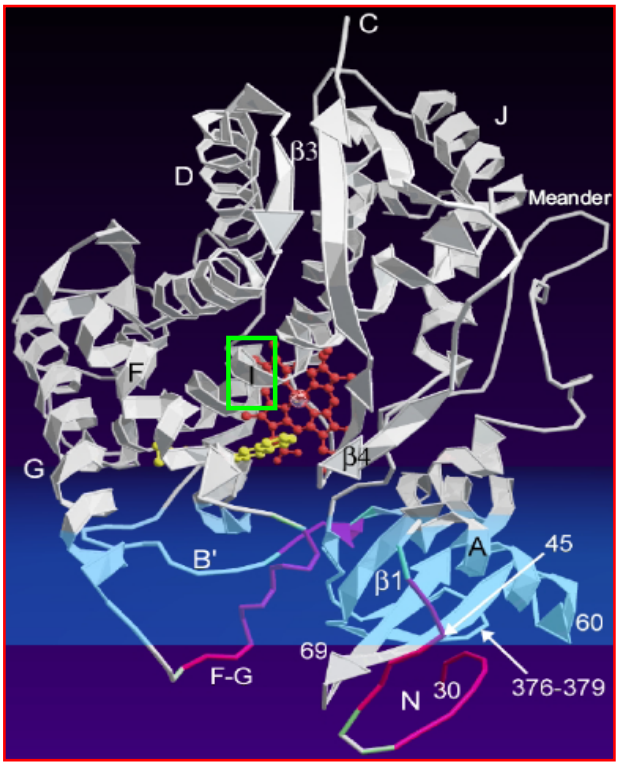


Fig. 1.4 A ribbon representation of the distal face of a folded CYP2C5 microsomal protein illustrating the protein architecture of CYP450s. The heme prosthetic group is indicated as a red ball and stick model while the bound substrate is indicated in yellow. Helices and sheets are labeled. The central part of the I-helix is indicated by a green border (Werck-Reichhart and Feyereisen, 2000).

The conserved core comprises of: firstly, the heme-binding loop which contains the most characteristic CYP450 consensus sequence (Phe-X-X-Gly-X-Arg-X-Cys-X-Gly) located on the proximal face of the heme just before the L-helix (not labeled on the distal face of Fig. 1.4) with the absolutely conserved cysteine that serves as the fifth (axial) ligand to the heme iron; secondly, the almost absolutely conserved Glu-X-X-Arg motif on helix K (also on the proximal side of the heme) which is probably needed to stabilize the core structure, and finally, the central part of the I-helix which contains the 6-letter CYP450 signature sequence: Ala/Gly-Gly-X-Asp/Glu-Thr-Thr/Ser, which corresponds to the proton transfer groove on the distal side of the heme (Werk-Reichhart & Feyereisen, 2000).

1.1.2 Catalytic mechanism of CYP450s

The catalytic mechanism of CYP450s can roughly be divided into four steps namely (1) substrate binding, (2) reduction of the substrate-hemoprotein-complex to a ferrous state, (3) binding of molecular oxygen and (4) a second reduction step resulting in activated oxygen species (Fig. 1.5) (Werck-Reichhart & Feyereisen, 2000; Munro *et al*., 2007). The short-lived activated oxygen species are responsible for the attack on substrates and comprises a mixture of two electrophilic iron–peroxo and iron-oxo oxidants (Fig. 1.5 labeled as [**compound 0**] and [**compound I**] respectively). Both these oxidants are formed by protonation of the two-electrons-reduced dioxygen – a process that occurs when a water channel is formed in the groove of the I-helix upon O_2 binding.

The oxo-species is the most abundant and is formed upon the cleavage of the O-O bond where one atom of oxygen leaves with the two electrons and two protons as water. The oxo-species inserts oxygen while the iron-hydroperoxo-species insert OH^+ to yield protonated alcohols. It should be noted that the end result of CYP450 catalysis is not always insertion of oxygen but can be *e.g*. dealkylation, dehydration or carbon-carbon bond cleavage (Werck-Reichhart & Feyereisen, 2000).

The incorporation of oxygen into X-H bonds can be achieved by a variety of chemical procedures *e.g*. by epoxidation/hydrolysis, the addition of water, nucleophilic substitution and reduction. Very often these reactions are not stereo-selective and do not allow for the distinction of carbon atoms carrying the same type of activation between, for example, two double bonds. In addition, the hydroxylation of *e.g*. non-activated carbon atoms can only be achieved by radical reactions which are, as a rule, not sufficiently selective to result in a chiral hydroxyl group at the desired position (Urlacher *et al*., 2004). Contrasting to this, CYP450s are capable of introducing molecular oxygen regiospecifically and enantioselectively into allylic positions, double bonds and non-activated C-H bonds (Urlacher *et al*., 2004; Urlacher & Eiben, 2006; Mandai *et al*., 2009a).

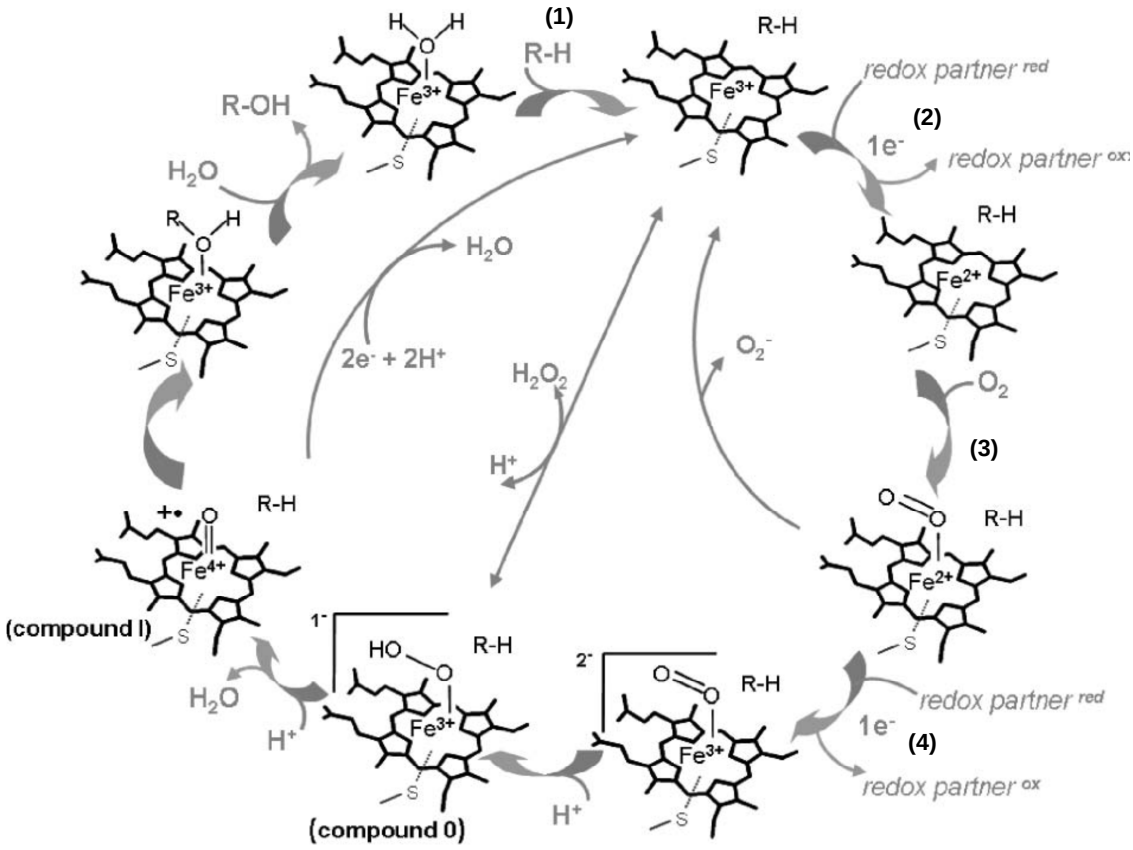


Fig. 1.5 Catalytic mechanism of CYP450s depicting the first atom of oxygen being reduced to water and insertion of the second oxygen atom into a substrate to yield a hydroxylated product. The very reactive ferric hydroperoxo species (**compound 0**) inserts OH^+, while the electrophilic oxidant, the ferryl-oxo enzyme species (**compound I**) attacks the substrate and effects its hydroxylation. (Adapted from: Munro *et al*., 2007).

Due to these unique chemical traits, CYP450s are involved in a plethora of reactions *e*.*g*. biotransformation of drugs, bioconversion of xenobiotics, metabolism of chemical carcinogens, biosynthesis of physiologically important compounds such as steroids, fatty acids, eicosanoids, fat-soluble vitamins and bile acids, as well as the conversion of *n*-alkanes, terpenes and aromatic compounds. CYP450s are also responsible for the degradation of several recalcitrant herbicides and insecticides. CYP450s catalyze many different types of reactions that include aliphatic hydrocarbon hydroxylation, heteroatom oxygenation, dealkylation, epoxidation, aromatic hydroxylation, reduction and dehalogenation (McLean *et al*., 2005; Bernhardt, 2006). As a result of this catalytic

versatility, microbial and mammalian CYP450s have been targeted as biocatalysts for the industrial production of fine chemicals, fragrances, and pharmaceutical compounds and used as bioremediation agents (Budde *et al*., 2005; Urlacher & Eiben, 2006; Mandai *et al*., 2009a).

Despite this impressive and diverse chemical repertoire, all CYP450s share some fundamental properties that hamper their commercial implementation: nearly all CYP450s are dependent on at least equimolar amounts of expensive NAD(P)H for each reaction cycle, have low catalytic activity and rely on complex electron transfer systems to reduce the heme in the monooxygenase. Although these limitations can be overcome by using whole-cell systems, other hurdles *e.g*. substrate limitation, product or substrate toxicity and product degradation also hamper whole-cell systems (Urlacher *et al*., 2004; Urlacher & Eiben, 2006).

From an industrial perspective bacterial CYP450s have enjoyed intense focus and scrutiny although they constitute a very small percentage of the CYP450 superfamily. It is only quite recently that mammalian CYP450s have been investigated with respect to industrial and biotechnological applications (Urlacher & Eiben, 2006). Although bacterial monooxygenases display higher stability, activity and better expression rates in recombinant hosts, the substrate range and reactions catalyzed by eukaryotic CYP450s are more amenable to industrial applications (Urlacher & Eiben, 2006). Considering the fact that eukaryotic CYP450s form the largest portion of the CYP450 group, this creates several new and exciting industrial prospects.

1.1.3 Nomenclature of CYP450s

CYP450s are subdivided and classified according to the guidelines set by a nomenclature committee. These guidelines include: amino acid identity, phylogenetic criteria and gene organization (Nelson *et al*., 1996). The root symbol ‘CYP’ is followed by a number which represents families (generally groups of proteins with > 40% amino acid sequence identity), a letter for subfamilies (> 55% identity) and a number for the specific protein (Werck-Reichhart & Feyereisen, 2000).

1.2 Classification of CYP450s: from reducing equivalents to electron transport

CYP450s can be divided into ten classes depending on how the electrons are delivered from the donor, mostly NAD(P)H, to the prosthetic heme group in the catalytic site. The classification is also dependent on the cellular localization of the redox partners and the CYP450 (Hannemann *et al*., 2007).

1.2.1 Class I systems

Class I CYP450 systems comprise mostly bacterial CYP450 systems (Fig. 1.6 A) as well as the mitochondrial CYP450 systems from eukaryotes (Fig. 1.6 B). Both systems require a FAD-containing reductase, which transfers reducing equivalents from a pyrimidine nucleotide (*i.e*. NADH or NADPH) to a ferredoxin protein which in turn reduces the CYP450. In bacteria all three proteins are soluble whereas in eukaryotes only the ferredoxin is a soluble protein of the mitochondrial matrix. The reductase and CYP450 are membrane-associated or membrane-bound to the inner mitochondrial membrane, respectively (Bernhardt, 2006; Hannemann *et al*., 2007).

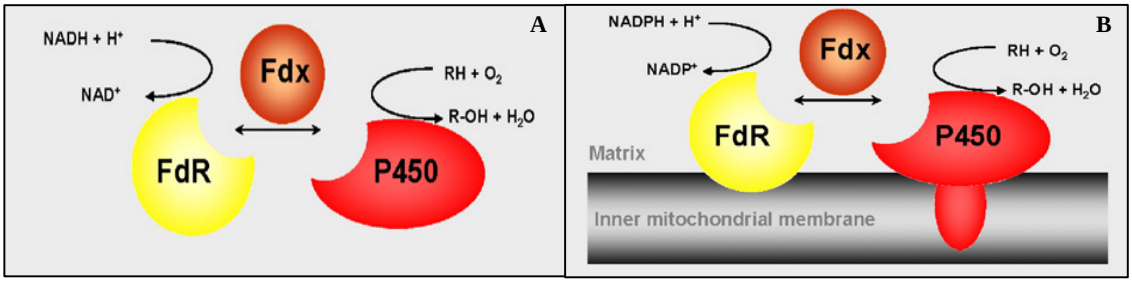


Fig. 1.6 Electron transfer mechanisms of Class I CYP450s in (**A**) bacteria and (**B**) eukaryotes. **Key:** FdR = FAD-containing ferredoxin reductase and Fdx = ferredoxin (Taken from Hannemann *et al*., 2007).

1.2.2 Class II systems

This class of CYP450 proteins are the most common in eukaryotes and in its simplest form the system comprises two integral membrane proteins (Fig. 1.7) that are found in the endoplasmic reticulum (ER): the CYP450 and NADPH CYP450 reductase containing the prosthetic groups FAD and FMN, which transfers both the required redox equivalents from NADPH to one of the many CYP450 isozymes. The reductase has evolved as a fusion of two ancestral proteins and displays, in its N-terminus, homology with the FMN-containing bacterial flavodoxins, while the C-terminus is homologous to the FAD-containing ferredoxin $NADP^+$ reductase and NADH-cytochrome *b5* reductase (Smith *et al*., 1994).

Apart from the vast number of eukaryotic class II CYP450s, only one prokaryotic class II monooxygenase system has been described in *Streptomyces carbophilus*. This prokaryotic system is composed of the CYP450 (CYP105A3) and a NADH-dependent CYP450 reductase containing both FAD and FMN. The proteins are located in the soluble fraction and this particular system has interestingly found industrial application since it catalyses the hydroxylation of mevastatin to pravastatin which is a tissue selective inhibitor of cholesterol biosynthesis (Serizawa and Matsuoka, 1991).

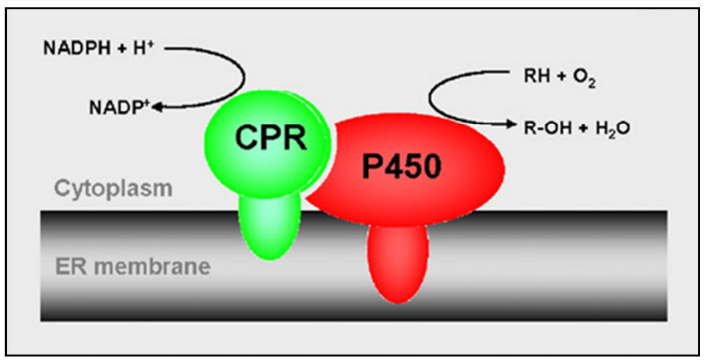


Fig. 1.7 Electron transfer mechanism of microsomal Class II CYP450 systems in eukaryotes. **Key:** CPR = NADPH-cytochrome CYP450 reductase (Taken from Hannemann *et al*., 2007).

1.2.3 Class III systems

Class III CYP450 systems were discovered in 2002 in the bacterium *Citrobacter braakii* (Hawkes *et al*., 2002). Like the Class I system, Class III CYP450 systems also rely on three protein components for electron transfer although they do not utilize an iron-sulfur protein (ferredoxin) but instead a flavodoxin which has been designated as cindoxin. Thus, the electrons are delivered *via* the redox centers FAD and FMN and not *via* FAD and an iron-sulfur-cluster as is the case with Class I CYP450 systems. Flavodoxin reductase and flavodoxin from *Escherichia coli* have been known to be able to substitute the endogenous interaction partners of heterologously expressed CYP450s (Barnes *et al*., 1991; Jenkins *et al*., 1994). The cytochrome from *C*. *braakii* (P450cin) is however the first example of a CYP450 known to naturally use a flavodoxin as redox partner.

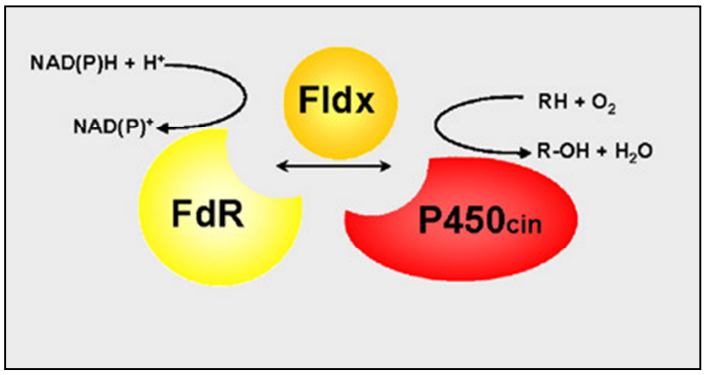


Fig. 1.8 Class III CYP450 system in *C*. *braakii*. **Key:** FdR = NAD(P)H-dependent FAD-containing ferredoxin reductase; Fldx = FMN-containing flavodoxin (cindoxin) and P450cin = cytochrome P450 (CYP176A1) from *C*. *braakii* (Taken from Hannemann *et al*., 2007).

1.2.4 Class IV systems

The CYP450 that represents this class, CYP119, was the first ever described thermostable CYP450 and as a result has been studied extensively (see sections 1.3.1.1.1 – 1.3.1.1.4). This soluble CYP450 was isolated from an extreme acidothermophilic archeaon and is thus also the first archaeal CYP450 ever described

(Wright *et al*., 1996; Nishida & Ortiz de Montellano, 2005). This eletron transfer system is unique since reducing equivalents are not obtained from NAD(P)H and CYP119 does not receive its electrons *via* NAD(P)H-dependent flavoproteins. Instead, initial reducing equivalents are provided by a 2-oxo-acid namely pyruvate and the flavoprotein is replaced with a 2-oxo-acid:ferredoxin oxidoreductase (Fig. 1.9) (Puchkaev *et al*., 2002; Puchkaev & Ortiz de Montellano, 2005) (see section 1.3.1.1.4 for more detail).

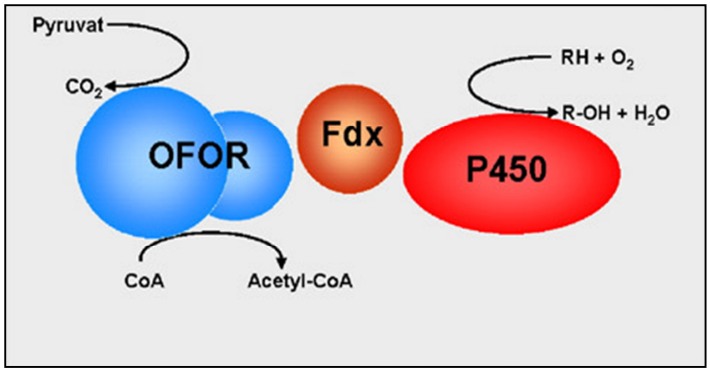


Fig. 1.9 Class IV CYP450 system in *S. solfataricus*. Note that the initial reducing equivalents are provided by pyruvate. **Key:** OFOR = 2-oxo-acid:ferredoxin oxidoreductase and Fdx = ferredoxin (Taken from Hannemann *et al*., 2007).

1.2.5 Class V systems

Class V CYP450 systems have an unique primary structural organization and consists of two separate protein components: a putative NAD(P)H-dependent reductase and a cytochrome P450-ferredoxin fusion protein (Fig. 1.10). In this system, the CYP450 heme-monooxygenase domain is fused at the C-terminus to a [3Fe-4S] type ferredoxin domain *via* an alanine-rich linker region, which is thought to act as a flexible hinge that allows interactions between the two domains. To date, the only example of such a system is the sterol 14α-demethylase CYP51 (MCCYP51FX) from *Methylococcus capsulatus* (Jackson *et al*., 2002).

Jackson and co-workers verified the 14α-demethylase activity with an *in vitro* assay using purified, heterologously expressed MCCYP51FX in which lanosterol was used as

substrate, NADPH as primary electron donor and spinach ferredoxin reductase as a surrogate reductase. Although the ferredoxin reductase was shown to be essential in the catalysis, additional saturating amounts of spinach ferredoxin had no significant impact on activity. This finding illustrates the functionality of the ferredoxin domain of MCCYP51FX so that an electron flow, as in the Class I systems, can be assumed. It has been suggested that there is a close evolutionary link between Class I and Class V CYP450 systems based on high homologies of the cytochrome domain (49% on amino acid level) and ferredoxin (42% on amino acid level) of the *Mycobacterium tuberculosis* CYP51 to that of the MCCYP51FX system.

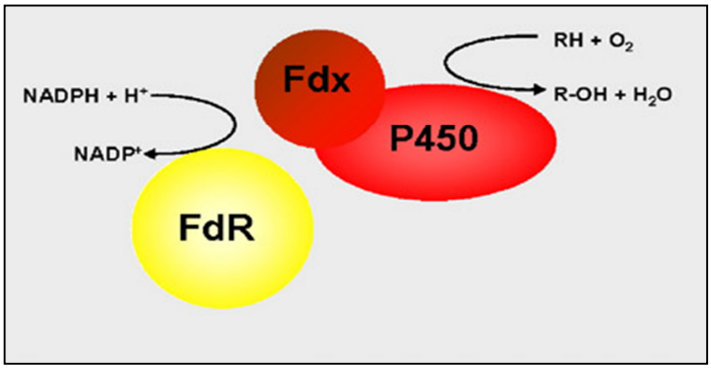


Fig. 1.10 Class V CYP450 system in *M. capsulatus*. **Key:** FdR = putative NAD(P)H-depenent reductase and Fdx+P450 = ferredoxin-cytochrome CYP450 fusion (Taken from Hannemann *et al*., 2007).

1.2.6 Class VI systems

This class of the CYP450 systems is almost a hybrid between Class III (P450cin) and Class VIII (P450BM3) systems since it is composed of a putative NAD(P)H-dependent flavoprotein reductase and a flavodoxin-P450-fusion protein. The first example of such a Class VI CYP450 was uncovered in *Rhodococcus rhodochrous* strain 11Y where the CYP450 (designated as XplA) is fused to a flavodoxin domain at its N-terminus (Fig. 1.11). The functional protein has been shown to degrade the widely used military explosive chemical called hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX). Rylott *et al*., (2006) illustrated the functionality of the fused flavodoxin domain of XplA by degrading

RDX in an assay that utilized purified, soluble XplA with added NADPH and ferredoxin reductase. Neither ferredoxin nor flavodoxin was added in the assay. Homologues of XplA have been reported in *Rhodococcus* strains DN22 and YH1 isolated from RDX-contaminated soil in Australia and Israel respectively (Rylott *et al*., 2006).

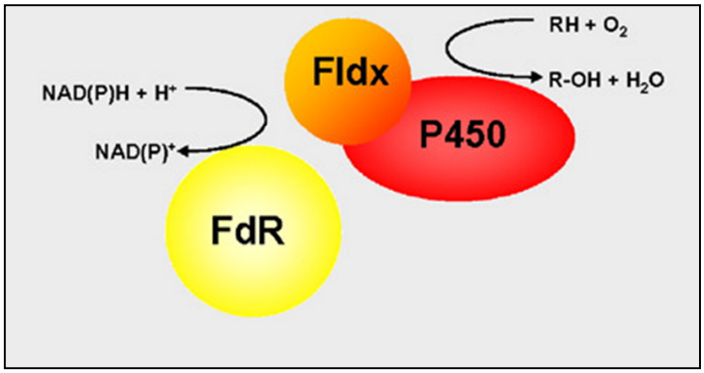


Fig. 1.11 Class VI CYP450 system as described in *R. rhodochrous* strain 11Y. **Key:** FdR = putative NAD(P)H-dependent flavoprotein reductase and Fldx+P450 = flavodoxin fused to the CYP450 (Taken from Hannemann *et al*., 2007).

1.2.7 Class VII systems

This particular group of CYP450s constitutes a completely novel class of the CYP450 systems since it is quite unique in its structural organization: the C-terminal of the CYP450 domain is fused to the domain of a phthalate dioxygenase reductase (Fig. 1.12). The first reported Class VII CYP450 was from *Rhodococcus* sp. strain NCIMB 9784 (a CYP116B2) and designated as P450RhF (Roberts *et al*., 2002). The CYP450 domain of P450RhF displayed a high homology (55%) to the Class I CYP116 from *Rhodococcus erythropolis*, which resulted in the classification of P450RhF as CYP116B2. The isoform of CYP116B2 from *Rhodococcus ruber* DSM 44319 (CYP116B3) has also been cloned and expressed and displays 90% amino acid identity with P450RhF (Liu *et al*., 2006).

The CYP450- and reductase domains of P450RhF are separated by a short linker region of 16 amino acids and the reductase portion displays three distinct functional parts: a FMN-binding domain, a NAD(P)H-binding domain and a [2Fe-2S] ferredoxin domain.

The flavin co-factor of the reductase in this system is FMN rather than the expected FAD and the P450RhF system also has a clear preference for NADPH (Roberts *et al*., 2003; Hunter *et al*., 2005).

Other P450RhF sequence homologues have also been identified in pathogenic *Burkholderia* species, *Ralstonia metallidurans*, *Ralstonia eutropha* JMP134 and the filamentous ascomycete *Gibberella zeae* PH-1 with the aid of genome analyses (De Mot & Parret, 2002; Hunter *et al*., 2005).

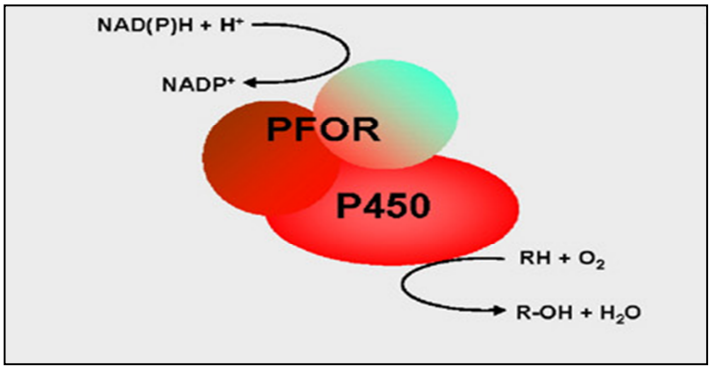


Fig. 1.12 Class VII CYP450 system in *Rhodococcus* sp. strain NCIMB. **Key:** PFOR = phthalate dioxygenase (Taken from Hannemann *et al*., 2007).

1.2.8 Class VIII systems

Class VIII CYP450s have been identified in several *Bacillus* sp. as well as in certain basidio- and ascomycetes fungi. These CYP450s are composed of a single polypeptide (Fig. 1.13) and are therefore catalytically self-sufficient monooxygenases (Urlacher *et al*., 2004; Budde *et al*., 2005; Bernhardt, 2006).

Probably the best studied Class VIII CYP450 is the CYP102A1 or P450BM3 from the soil bacterium *Bacillus megaterium*. P450BM3 is a cytosolic, 119 kDa polypeptide consisting of a heme-containing CYP450 oxygenase domain that is connected *via* a short protein linker to a diflavin reductase domain which contains one equivalent of the cofactors FAD and FMN (Miura & Fulco, 1974; Narhi & Fulco, 1986, Li & Poulos, 1999). P450BM3

catalyses the NADPH-dependent hydroxylation of medium- and long-chain saturated fatty acids at the ω-1, ω-2 and ω-3 positions. This CYP450 displays high stability and has one of the highest monooxygenase activities (Budde *et al*., 2005).

Two homologues of CYP102A1, namely CYP102A2 and CYP102A3, from *Bacillus subtilis* have also been characterized. Although these homologues also hydroxylate fatty acids at the same ω-positions as their *B*. *megaterium* counterpart, they display a strong preference for long-chain and branched-chain unsaturated fatty acids (Gustaffson *et al*., 2004). In addition, De Mot and co-workers (2002) also discovered two additional homologues in *Bacillus anthracis* (Ames Strain) and *Bacillus cereus*.

As mentioned previously, self-sufficient CYP450s also exist in eukaryotes. CYP505A1 (P450foxy) was originally isolated from the ascomycetous fungus *Fusarium oxysporum* (Nakayama *et al*., 1996; Kitazume *et al*., 2002). Like P450BM3, the P450foxy protein is also a single polypeptide and is 118 kDa in size. P450foxy also catalyzes the subterminal hydroxylation (ω-1 to ω-3) of fatty acids. Unlike P450BM3, P450foxy is not localized in the cytosol but is loosely bound to the cell membrane and does not possess any well defined membrane anchor region (Kitazume *et al*., 2000).

Another Class VIII CYP450 has been identified in the filamentous ascomycete *Gibberella moniliformis* (*Fusarium verticilioides*) which is a known producer of mycotoxins (Seo *et al*., 2001). This CYP450 (CYP505B1), also known as Fum6p, displays the same putative domain arrangement in a single protein as observed in CYP102A1 – 3 and CYP505A1 (Seo *et al*., 2001). The physiological function of Fum6p has not yet been elucidated but it is speculated that is may act as a polyketide hydroxylase that is involved in the biosynthesis of the mycotoxin fumonisin (Proctor *et al*., 2003).

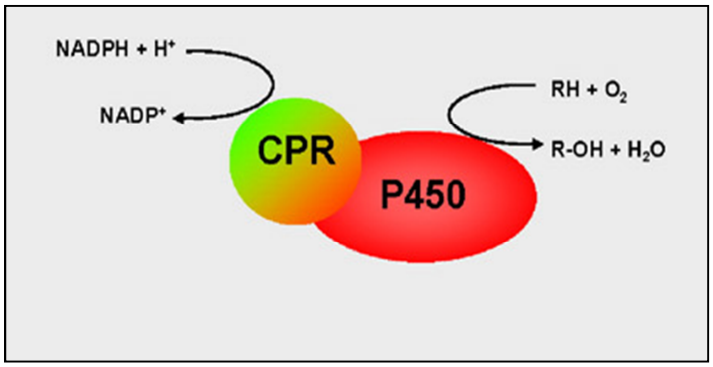


Fig. 1.13 Bacterial Class VIII CYP450 system as in *B. megaterium*. **Key:** CPR = NADPH-dependent CYP450 reductase (Taken from Hannemann *et al*., 2007).

1.2.9 Class IX systems

To date this class of CYP450s has only been identified in fungi. The CYP55 or P450nor is a nitric oxide reductase and was isolated from the filamentous fungus *Fusarium oxysporum*. Unlike other eukaryotic CYP450s, the P450nor localizes in both the mitochondrial and cytosolic fractions and thus makes P450nor the only soluble eukaryotic CYP450 described to date (Takaya *et al*., 1999). Electrons are donated to P450nor by NADH, and P450nor does not rely on any other electron transfer proteins (Fig. 1.14) to convert two molecules of nitric oxide into nitrous oxide (Nakahara *et al*., 1993). Fungal genome sequence data have identified several P450nor isozymes and two have already been cloned and expressed from *Cylindrocarpon tonkinense* (Kudo *et al*., 1996) and *Trichosporum cutaneum* (Zang *et al*., 2001).

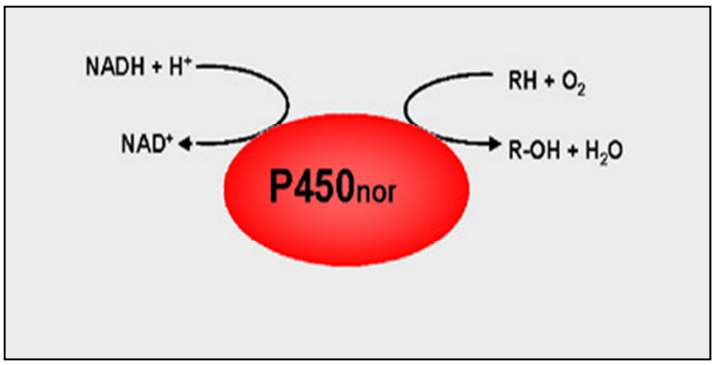


Fig. 1.14 Fungal Class IX CYP450 system as in *F. oxysporum*. Note the absence of any electron transfer proteins in the system (Taken from Hannemann *et al*., 2007).

1.2.10 Class X systems

This class of CYP450s deviates from classical monooxygenase behavior since they do not require any redox partners, molecular oxygen or an electron source like NAD(P)H to perform catalysis. Instead, these Class X CYP450s utilize an independent, intramolecular transfer system (Fig. 1.15). Thus far, Class X CYP450s consist of allene oxide synthase (CYP74A or AOS), fatty acid hydroperoxide lyase (CYP74B/C or HPL), divinyl ether synthase (CYP74D or DES), prostacyclin synthase (PGIS or CYP8A) and thromboxane synthase (TXAS or CYP5A).

The AOS, HPL and DES proteins form part of the lipoxygenase pathway in plants and are localized in the membranes of chloroplasts. The oxygen necessary for the reaction is obtained from the hydroperoxide of the substrate (Lau *et al*., 1993; Shibata *et al*., 1995; Froelich *et al*., 2001; Itoh & Howe, 2001).

The last two members of Class X CYP450s are found in mammals and are responsible for the rearrangement of C_{20} acyl peroxides in the arachidonic acid cascade. CYP8A and CYP5A utilize the same substrate to synthesize two different compounds that have opposite physiological roles. CYP8A catalyses the isomerization of prostaglandin H2 to form protacyclin, which is a potent vasodilator and platelet aggregation inhibitor. CYP5A on the other hand, catalyses the isomerization of prostaglandin H2 to form thromboxane

A2: a potent inducer of vasoconstriction and platelet aggregation (Ullrich *et al*., 2001; Wu & Liou, 2005).

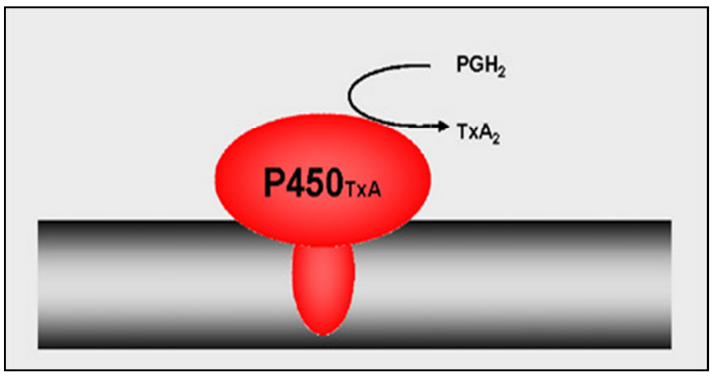


Fig. 1.15 An example of the membrane bound Class X CYP450 in mammals. **Key:** P450TxA = Thromboxane synthase (CYP5A); PGH_2 = prostaglandin H2 and TxA_2 = thromboxane A2 (Taken from Hannemann *et al*., 2007).

1.3 CYP450s from extremophiles

1.3.1 Described CYP450s from extremophiles – from thermostability to surviving in acid

To date, only four CYP450s from extremophiles have been studied in detail. Three of these CYP450s are from hyperthermo-acidophilic and thermo-acidophilic *Archaea*, namely *Sulfolobus acidocaldarius, Sulfolobus tokodaii* and *Picrophilus torridus* respectively, and the fourth CYP450 from a gram negative, thermophilic bacterium *Thermus thermophilus* HB27.

1.3.1.1 CYP119 from *Sulfolobus* spp.

1.3.1.1.1 On the origin of CYP119A1 – an *erratum*

The first CYP450 from an archaeon, CYP119A1, was initially cloned by Wright *et al*. (1996) from, according to them, *S*. *solfataricus*. A recent survey of the genome sequence of *S*. *solfataricus,* however, revealed that it did not harbour *CYP119A1* or any other CYP450 gene. *CYP119A1* was in fact located in *S*. *acidocaldarius,* which possesses a single CYP450 gene (*CYP119A1*) located on its chromosome (Ho *et al*., 2008). It is suspected that when the *S*. *solfataricus* strain was used by Wright *et al*. (1996), it was already contaminated with *S*. *acidocaldarius*. In addition, the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) in Germany also commented on this issue on their website (http://www.dsmz.de/microorganisms/html/strains/strain.dsm001616.html) stating that: "…cultures of DSM 1616 supplied until April 1989 were contaminated with *Sulfolobus acidocaldarius*" (Ho *et al*., 2008). Consequently, all papers relating to CYP119(A1) referred to *S*. *solfataricus* as the CYP450-harboring strain. For clarification and non-ambiguity purposes in this review, the gene name '*CYP119(A1)*' refers to the gene from *S*. *acidocaldarius* and not *S*. *solfataricus*.

Sulfolobus acidocaldarius DSM639 is the type strain of the archaeal genus *Sulfolobus* and was the first hyperthermoacidophile to be characterized (Brock *et al*., 1972). *S*. *acidocaldarius* grows optimally at 75 to 80˚C and pH 2.0 – 3.0 under strictly aerobic

conditions (Chen *et al*., 2005). *S*. *acidocaldarius* differs markedly from other *Sulfolobus* species *e.g*. *S*. *solfataricus* and *S*. *tokodaii* with regard to its genome stability and organization. In addition, the genome of *S*. *acidocaldarius* encodes for 2 292 proteins of which 400 are not present in the genomes of *S*. *solfataricus* and *S*. *tokodaii*. Moreover, *S*. *acidocaldarius* lacks some genes coding for sugar transporters which explains why *S*. *acidocaldarius* is only able to grow on a limited range of carbon sources in comparison to other *Sulfolobus* species (Chen *et al*., 2005).

To date, the only other *Sulfolobus* strain found to contain a CYP450 is *S*. *tokodaii* str. 7, which contains a CYP450 named P450st or CYP119A2 (Genbank accession number NP_377075.1) (Oku *et al*., 2004).

1.3.1.1.2 Crystal structure of CYP119

Crystal structures of CYP119A1 have been solved by Yano *et al*. (2000) and Park *et al*. (2000) to resolutions of 1.93 Å and 1.5 Å, respectively. Later, the crystal structure of CYP119A2 (P450st) from *S*. *tokodaii* str. 7 was initially solved to a resolution 3.0 Å (Oku *et al*., 2004) and recently to a resolution of 1.94 Å (Matsumura *et al*., 2011). This CYP450 has 64% amino acid identity to CYP119A1 and the r.m.s. for the aligned Cα-backbones of the two CYP119s is 1.3 Å (Oku et al., 2004).

CYP119A1 and CYP119A2 are respectively 368 and 367 amino acids in length and are considerably smaller than other well known mesophilic CYP450s *e.g*. $P450_{cam}$ (414 residues) (Poulos *et al*., 1985) and $P450_{eryF}$ (403 residues) (Cupp-Vickery & Poulos, 1995). The shorter length of the CYP119s is mainly attributed to the fact that the proteins lack several of the N-terminal residues which are present in $P450_{cam}$ and $P450_{eryF}$ thus conferring a more compact structure to CYP119. The CYP119s exhibit the typical CYP450-fold. CYP119A1 is given as an example (Fig. 1.16).

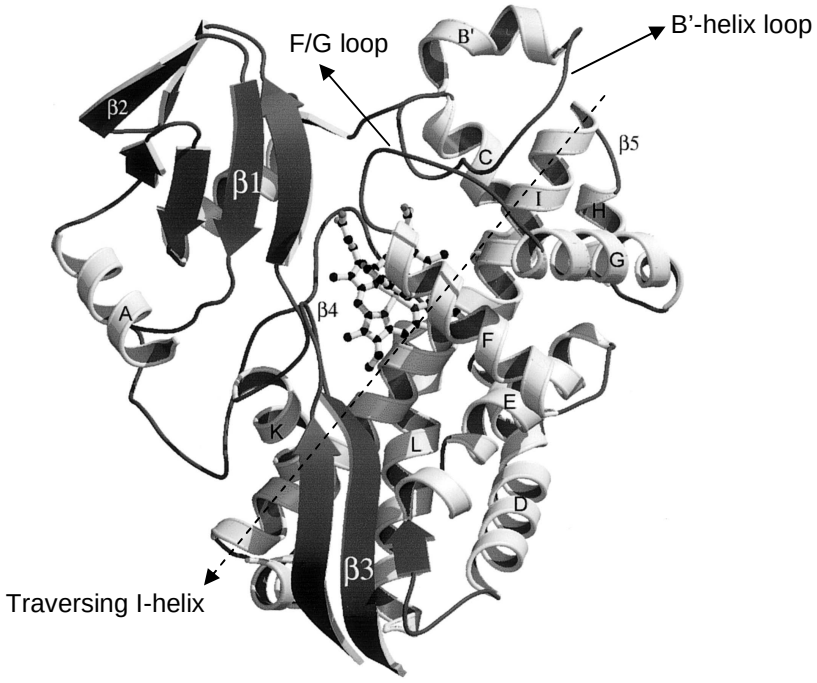


Fig. 1.16 Ribbon diagram of CYP119A1 complexed with 4-phenylimidazole as ligand to the heme group. α-Helices are indicated as white cylinders while β-sheets are indicated as grey arrows (Adapted from: Yano *et al*., 2000).

On the distal side of both CYP119A1 and CYP119A2, the I-helix traverses the entire protein and provides a framework for the catalytic residues including the highly conserved Thr-213 (facilitating the activation of oxygen) (Koo *et al*., 2000) as well as Thr-214 and Thr-215 – this triplet set of threonines also occurs in other CYP450s but is not present in the well-studied bacterial CYP450s (Ortiz de Montellano, 1995; Koo *et al*., 2000). Other distinct structural features of the CYP119s are that the B'-helix region, which is known to be important in substrate binding (Gotoh, 1992), does not form a complete helix (when compared to *e.g*. P450eryF) and that the B'-helix is preceded by a long loop that enables the helix to adopt a position away from the active site towards the molecular surface (Fig. 1.17). In addition, the F/G loops of $P450_{cam}$ and other CYP450s, extend towards the surface but in the CYP119s, the F/G loop points completely in the opposite direction and dips into the active site (Fig. 1.17), thus occupying the space normally taken by the B'-helix region (Yano *et al*., 2000).

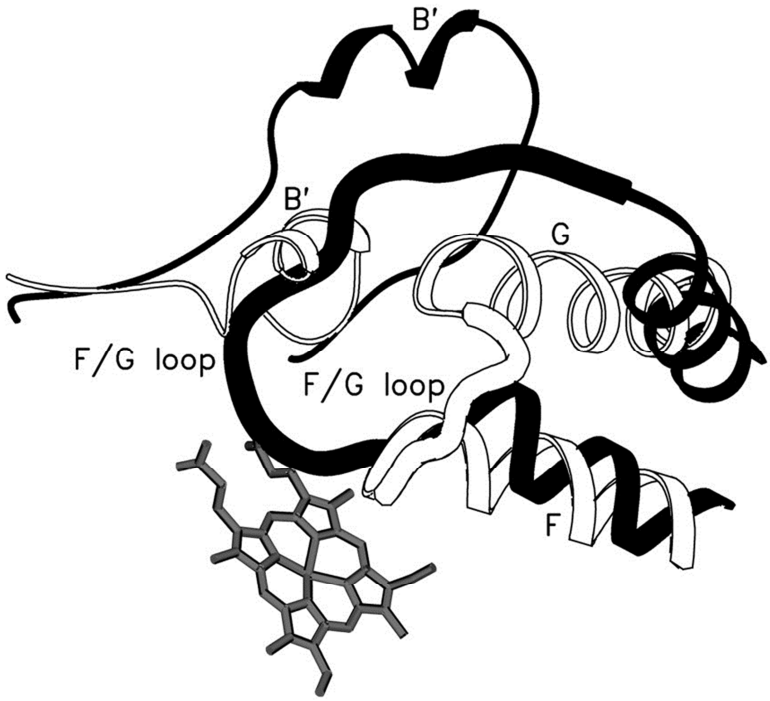


Fig. 1.17 A stereo-diagram illustrating the superimposed F/G loop region of CYP119A1, complexed with imidazole, (black) with the F/G loop region of $P450_{cam}$ (white) in the presence of the heme prosthetic group (Adapted from: Yano *et al*., 2000).

In CYP119A1 it was demonstrated that the F/G loop can undergo various conformational changes, depending on the ligand bound to the active site and the most significant change is usually seen in the F-helix. For example, when 4-phenylimidazole is used as a ligand, the F-helix includes residues 141 – 155 but when imidazole acts as the ligand, the helix stops at residue 151 – a loss of one full turn of the helix. As a consequence of the unraveling of the helix the F/G loop lengthens, enabling the loop to dip down into the active site thereby making more contacts with the smaller imidazole molecule. When 4-phenylimidazole is bound, the F/G loop moves to create more space for the larger ligand (Yano *et al*., 2000).

1.3.1.1.3 Factors contributing to the thermostability in CYP119

CYP119A1 has a T_m = 91˚C which is much higher than the optimum growth temperature of *S. acidocaldarius* (75 – 80˚C). Thermostability in CYP119 has been attributed to

several factors of which chief are: (a) the higher density of salt bridges, (b) relatively low density of alanines as well as a high incidence of isoleucine in the interior of the protein thus resulting in more efficient side-chain packing and (c) the presence of large and extended clusters of aromatic residues that are not present in other known mesophilic CYP450s (Chang & Loew, 2000; Yano *et al*., 2000). When inspecting the crystal structure of CYP119A1, two large clusters of aromatic residues are apparent (Fig. 1.18). These residues are, with two exceptions, identical in CYP119A2.

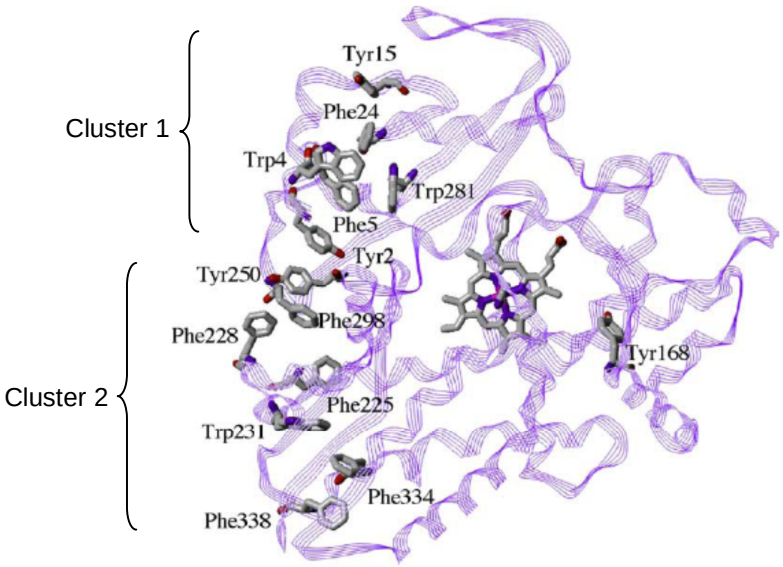


Fig. 1.18 Two aromatic residue clusters on the surface of CYP119 contributing towards thermostability (Nishida & Ortiz de Montellano, 2005).

The first cluster comprises six residues and are identical in both CYP119s namely Tyr-2, Trp-4, Phe-5, Phe-24, Trp-281 and Tyr-15. The second cluster comprises seven residues with two substitutions by other aromatic amino acids in CYP119A2 (shown in parentheses) namely Phe-225, Phe-228 (Tyr-229), Trp-231, Tyr-250 (Phe-250), Phe-298, Phe-334 and Phe-338 (Chang & Loew, 2000; Yano *et al*., 2000). In a study by Puchkaev *et al*. (2003) several point mutations were introduced into these aromatic residues of CYP119A1 to yield mutants: W231A, Y250A, W281A, Y2A/Y250A, W4A/W281A and Y168A. By utilizing circular dichroism, it was demonstrated that each of the CYP119A1 mutants displayed a *ca*. 10˚C lower melting point when compared to the melting point of wildtype CYP119A1 (Puchkaev *et al*., 2003). In addition, a study by

Maves & Sligar (2001), in which random mutagenesis was performed on CYP119A1, reported similar results with regard to the disruption of aromatic residue clusters and decreased melting points in comparison to the wildtype CYP119A1.

1.3.1.1.4 Electron donor partners of CYP119s

Although the crystal structure of CYP119A1 was well described, including the factors contributing to its thermostability, making it an ideal catalyst at high temperatures and acidic pH, the native electron partners and native substrate remained elusive. Initial attempts to oxidize styrene were achieved with purified preparations of CYP119A1, using H_2O_2 as electron donor to yield styrene oxide (*i.e*. an epoxidation reaction) at a rate of 0.6 $nmol.min^{-1}.nmol^{-1}$ CYP450 at 30˚C (Koo *et al*., 2000). Similarly, successful oxidation of *cis*- and *trans*-β-methylstyrenes was also achieved (Koo *et al*., 2000). When styrene is bound to CYP119A1, styrene has a binding constant (K_s) of 530 µM. CYP119A2 also epoxidizes styrene using H_2O_2 as electron donor with kinetic constants similar to that of CYP119A1 (K_s = 170 µM and k_{cat} = 0.0087 s^{-1} [0.5 min^{-1}] at 30˚C) (Matsumura *et al*., 2008). Mutants of CYP119A2 with Phe310 changed to Ala and Ala320 to Gln have higher binding constants but improved activity. A double-point mutant with the F310A and A320Q mutations gave improved styrene epoxidation with k_{cat} = 0.0287 s^{-1} (1.7 min^{-1}) at 30˚C although the binding constant was very high (K_s = 800 µM).

Although Koo *et al*. (2000) were unable to demonstrate styrene epoxidation with CYP119A1 when using putitaredoxin/putidaredoxin reductase for electron transfer, Mastumura *et al*. (2011) recently demonstrated that CYP119A2 can accept electrons directly from NAD(P)H to epoxidize styrene with k_{cat} = 0.61 x 10^{-4} s^{-1} (3.7 x 10^{-3} min^{-1}) at 25˚C. This activity was much lower than with H_2O_2. The reaction conditions (pH, buffer and NADH concentration) used with the two enzymes were completely different and it is thus still possible that CYP119A1 might also be able to accept electrons directly from NAD(P)H. Notable is the high K_m (13 mM) that CYP119A2 demonstrated for NADH. Koo *et al*. (2000) used only 1.5 mM NADH in their reactions, while Matsumura *et al*. (2011) used 4 mM. The only other CYP450s that are known to accept electrons directly from NADH, thus circumventing the need for redox partners, are those belonging to Class IX *i.e*. CYP55A1 (P450nor) and its homologues. The crystal structure of P450nor revealed

that the hydrophilic regions of the F/G helices are pushed by die hydrophilic region of the I-helix. This results in the heme pocket being more open to solvents when compared to P450cam (Fig. 1.19). It has been suggested that this region could be a possible NADH-binding site. In the crystal structure of P450st (also CYP119A1A) the distal pocket was also expansively open to solvents as is the case with P450nor (Matsumura *et al*., 2011).

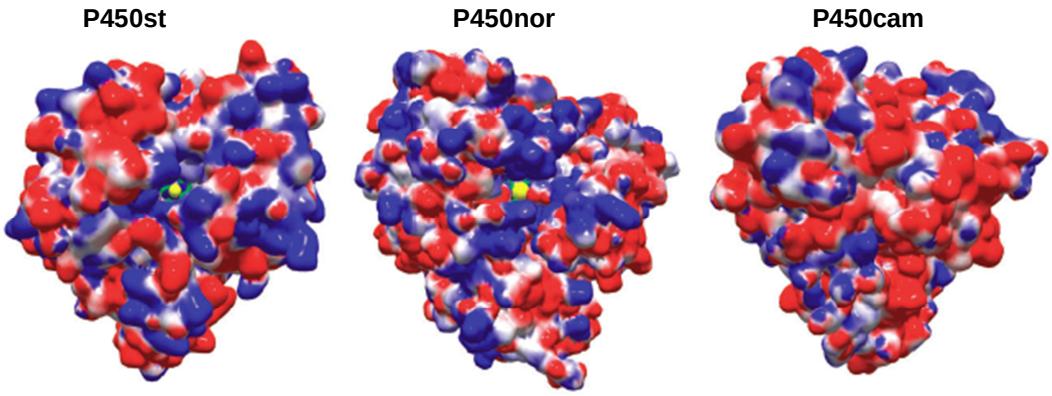


Fig. 1.19 Comparison of the distal heme pockets between P450st, P450nor and P450cam. The heme iron is illustrated in yellow and positive, negative and neutral potentials are shown in blue, red and white respectively (Adapted from Matsumura *et al*., 2011).

Fatty acids, especially lauric acid, binds more tightly to CYP119A1 than styrene (lauric acid K_s = 1.2 µM) (Koo *et al*., 2002). Fatty acid hydroxylation was initially achieved with the surrogate electron donor partners: putitaredoxin/putidaredoxin reductase (Koo *et al*., 2002). The best lauric acid hydroxylation results with these surrogate electron transfer systems were obtained at 37˚C with an optimized reaction comprising 1:20:1 of CYP119:putitaredoxin:putitaredoxin reductase with a K_m = 21 µM and V_{max} = 0.36 min^{-1} to yield hydroxylauric acid (Koo *et al*., 2002). Puchkaev *et al*. (2002) however, used a ferredoxin (Fdx) as well as 2-oxoacid:ferredoxin oxidoreductase (OFOR) from *S*. *tokodaii* str. 7 together with CYP119A1 from *S*. *acidocaldarius* to successfully oxidize lauric acid (Puchkaev *et al*., 2002). This system was unusual from a CYP450 point of view, since the ultimate source of electrons was never NADH nor NAD(P)H but rather a 2-oxoacid *e.g*. pyruvate. In addition, the Fdx receives electrons from an 2-oxoacid:ferredoxin oxidoreductase (OFOR) and not from the usual NAD(P)H-dependent flavoprotein. Three years later Puchkaev & Ortiz de Montellano, (2005) managed to use the purified, native

Fdx and OFOR from *S*. *acidocaldarius* as well as its purified CYP119A1 to reconstitute a thermostable system in which lauric acid was successfully oxidized at 70˚C and pH 4.5, and even remained catalytically active for 20 min at 90˚C.

The Fdx and OFOR in *S*. *tokodaii* str. 7 were first described by Zhang *et al*. (1996) for their role in central metabolism *i*.*e*. the CoA-dependent oxidation of pyruvate and 2-oxoglutarate in the tricarboxylic acid cycle. It was not until Puchkaev *et al*. (2002) published their results, that the Fdx and OFOR were regarded in a CYP450 context. Several OFORs in other archaea *e*.*g*. *Halobacterium salinarum*, *Pyrococcus furiosus* and *Archaeoglobus fulgidus* have been described [for a comprehensive review see Zhang *et al*. (1996)]. Due to the unique source of electrons (the oxoacids) and the OFOR (that is not flavin containing), this electron transfer system constitutes a whole new class of CYP450 electron donors namely Class IV (section 2.1.4). In addition, the Fdx and OFOR makes it possible to reconstitute high temperature CYP450 catalytic systems (Puchkaev & Ortiz de Montellano, 2005).

1.3.1.2 CYP231A2 from *Picrophilus torridus*

Picrophilus torridus is a thermoacidophilic euryarchaeon that flourishes optimally at 60˚C and pH 0.7, although there are some reports that it can grow at pH 0, making *P*. *torridus* one of the most acidophilic organisms known. The genome of *P*. *torridus* has been sequenced and reveals the presence of two CYP450 genes: *CYP231A2* (PTO1399) and *CYP232A2* (PTO0085) which translate into proteins with molecular weights of 39.56 and 44.36 kDa respectively (Futterer *et al*., 2004). To date, only CYP231A2 has been crystallized (Ho *et al*., 2008). The native substrates and redox partners of both CYP450s are unknown. CYP231A1 and CYP231A2 were previously annotated in the sequenced genome of *Ferroplasma acidarmanus* (see: http://drnelson.uthsc.edu/subfam.list.htm), another thermoacidophilic euryarchaeon that can grow at pH 0 (Dopson *et al*., 2004). However, CYP231A1 and CYP231A2 are now annotated in the genome of *P*. *torridus,* whereas CYP232A1 and CYP232A2 have been annotated in the genome of *F*. *acidarmanus*.

1.3.1.2.1 Spectroscopic characterization of the CYP231A2 active site

Since the optimal pH for growth of *P*. *torridus* is 0.7 and its intracellular pH 4.6, CYP231A2 provides the opportunity to study a CYP450 that has adapted to tolerate such acidic conditions. This was initially achieved by Ho *et al*. (2008) whereby the purified CYP231A2's active site integrity was assessed by performing CO-difference spectra under anaerobic conditions at pH 6.4 and pH 4.6, and comparing it to the CO-difference spectra of the prototypical mesophilic $P450_{cam}$ tested at the same pH values. A characteristic CO-difference spectrum, with a Soret peak at 450 nm, was achieved for CYP231A2 at pH 6.4 in as little as 2 min after CO addition (Fig. 1.20 B) along with 25% P420 species (indicating either inactive or denatured CYP450). However, upon exposing the sample to air, the spectra reverted back to its original ferric spectrum. The authors reported that $P450_{cam}$ exhibited higher stability at pH 6.4 than CYP232A2 since the CYP450 species dominated the spectra even after 60 min of incubation (spectra not shown in paper). However, $P450_{cam}$ was spectroscopically unstable at pH 4.6 and the protein precipitated. After reduction with sodium diothionite, which yielded exclusively P420 species, the precipitation continued. Contrasting to this, CYP231A2 remained spectroscopically stable and soluble at pH 4.6. Interestingly enough, upon the addition of dithionite and CO, P420 species were predominantly present together with a very small Soret peak at 450 nm (Fig. 1.20 E).

The spectral stability of the ferric form of CYP231A2 at pH 4.6 suggested that the heme environment was not drastically altered and that the heme remained bound in its normal position. However, when the pH was increased close to neutrality with NaOH, the spectrum changed to yield a much more prominent CYP450 peak (Fig. 1.20 F). When compared to $P450_{cam}$, it is apparent that CYP231A2 is much more stable at pH 4.6 and that it is able to revert to the P450 (thiolate) form from the P420 (putative thiolate) form when the solution is neutralized. Ho *et al*. (2008) noted that the above-mentioned results were consistent with (1) protonation of the proximal thiolate ligand to form P420 species at pH 4.6, (2) resistance of the protein structure to significant pH-dependent deformation and (3) reversible protonation-deprotonation of the proximal thiol ligand.

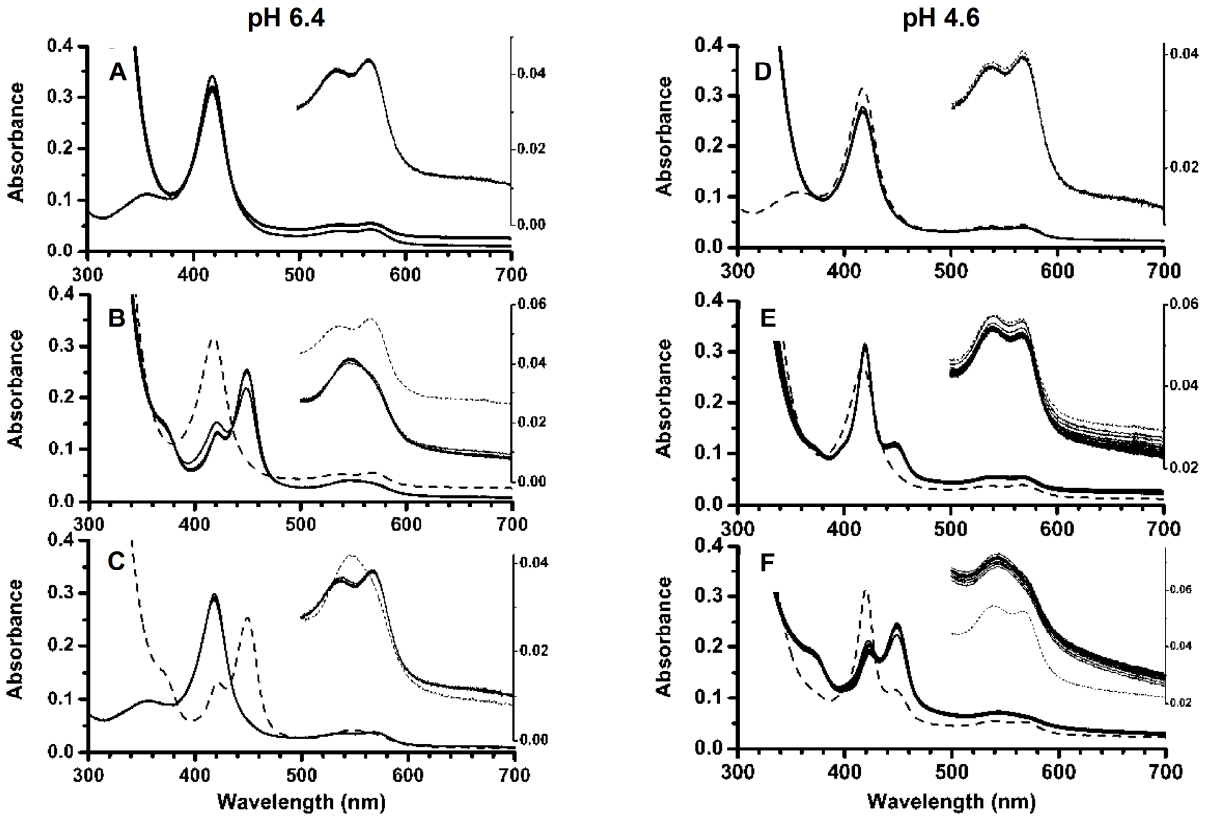


Fig. 1.20 Monitoring the formation of the CYP231A2 ferrous-CO complex at pH 6.4 (**A – C**) and pH 4.6 (**D – F**). (**A**) addition of sodium dithionite; (**B**) exposure of reduced sample to CO (dashed spectrum is of the protein before CO exposure); (**C**) exposure of sample to air (dashed spectrum is of the protein before exposure to air); (**D**) addition of sodium dithionite; (**E**) exposure of reduced sample to CO (dashed spectrum is of the protein before CO exposure); (**F**) addition of NaOH (dashed spectrum is of protein before NaOH addition) (Adapted from: Ho *et al.*, 2008).

1.3.1.2.2 Crystal structure of CYP231A2

CYP231A2 comprises 352 residues, making it smaller than both CYP119 (368 residues) and CYP175A1 (389 residues). The overall structure of CYP231A2 displays a typical CYP450 fold (Fig. 1.21). The topology and secondary assignments were based on the structure of $P450_{cam}$ from *P. putida* (Poulos *et al.*, 1985). CYP231A2 was refined to a resolution of 2.5 Å and comprises 12 α-helices that account for 45% of the total structure

and five antiparallel β-sheets constituting 13% of the structure. Normally, CYP450s have a N-terminal A-helix but this is absent in the structure of CYP231A2 since the authors truncated the gene by 27 base pairs during the cloning process. The N-terminus of CYP231A2 thus starts at β1 and the first visible residue is Leu-12 (Fig. 1.21).

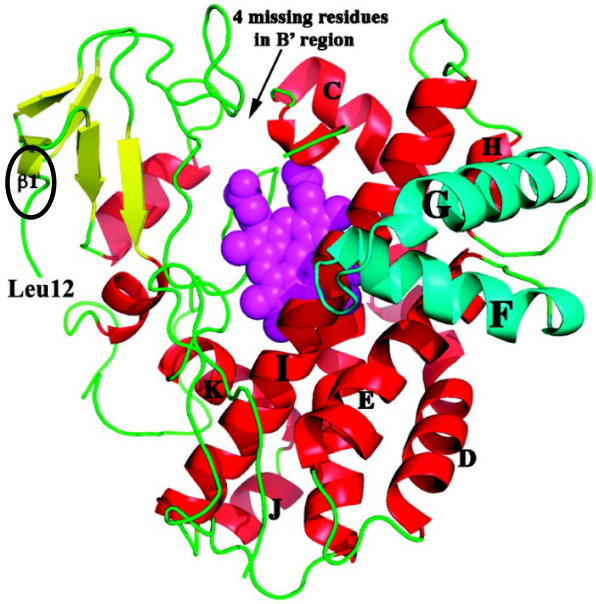


Fig. 1.21 Ribbon diagram of CYP231A2 (ligand free). Helices are labeled according to the accepted nomenclature derived from $P450_{cam}$ from *P. putida*. Note: there is no A-helix and the N-terminus starts just before the encircled β-sheet β1. (Adapted from: Ho *et al.*, 2008).

The heme prosthetic group is embedded between the I and L helices with Cys-304 serving as the proximal axial thiolate ligand. The propionate side chains of the heme are hydrogen bonded to the protein *via* residues Arg-68, Arg-247 and His-302 (Ho *et al.*, 2008). Like other CYP450s, CYP231A2 has a β-sheet-rich as well as α-helical-rich domain. Between these two domains lies a cleft that provides access from the protein surface to the distal side of the heme group. In most known CYP450 structures, this cleft is closed by direct contacts between the two domains which is often due to contacts between the F/G loop and β regions. However, this cleft is wide open in CYP231A2 (Fig. 1.21).

1.3.1.2.3 Factors contributing to thermostability

As observed with other CYP450s, it is very likely that the F/G helices and F/G loops undergo substantial movement when substrates or inhibitors bind, thus resulting in the closing down of the active site's entry channel. The ligand free form of CYP231A2 has a T_m = 65˚C. However, when 4-phenylimidazole was used as ligand, the T_m increased to 73˚C – illustrating that the ligand-bound form is much more thermally stable. The native substrate of CYP231A2 is unknown but it is envisaged that the substrate complex of CYP231A2 will also adopt a closed conformation and thus be thermally stable at the optimum growth temperature of *P*. *torridus* (Ho *et al*., 2008). CYP231A2 does not appear to have extensive salt bridge networks, which seems to aid in thermal stability (Yano *et al*., 2003), as is the case with CYP119A1 and CYP175A1 – in fact it only has one such network. Another factor that appears to contribute to thermostability is large networks of aromatic residues in the protein (Yano *et al*, 2000). CYP231A2 does possess more aromatic residue networks than CYP119 but not as extensive as is the case with CYP119. Overall, it seems as if CYP231A2 does not possess any of the characteristic features associated with enhanced thermal stability, except for being the smallest known CYP450.

1.3.1.3 CYP175A1 from *Thermus thermophilus* HB27

Thermus thermophilus belongs to the phylum *Deinococcus-Thermus* - a phylum in the domain *Bacteria* that include members that are known for their resistance to extreme stresses including radiation, oxidation, desiccation and especially high temperatures (Tian & Hua, 2010). To date, two stains of *T*. *thermophilus* have been described, namely strains HB27 and HB8 (Oshima & Imahori, 1974, Williams *et al*., 1995). *T*. *thermophilus* HB27 grows aerobically, is an obligate heterotroph, has a maximum growth temperature of 85˚C and displays yellow pigmentation when cultured (Chung *et al*. 2000; Balkwill *et al*., 2004; Henne *et al*., 2004). The 1.89 Mb genome of *T*. *thermophilus* HB27 has been sequenced, as well as its megaplasmid which is 232 605 bp in size (Henne *et al*., 2004). The megaplasmid harbours the *CYP175A1* gene which forms part of a carotenoid gene cluster involved in β-carotene biosynthesis, giving *T*. *thermophilus* HB27 its yellow colour (Hoshino *et al*., 1993; Blasco *et al*., 2004).

It has been demonstrated that the CYP450 from *T*. *thermophilus* HB27 (CYP175A1) is a β-carotene hydroxylase that is responsible for the hydroxylation of β-carotene at the 3- and 3'-positions of its β-ionone rings to yield the two xanthophylls, namely β-cryptoxanthin (containing one hydroxyl group) and zeaxanthin (containing two hydroxyl groups) (Blasco *et al*., 2004; Momoi *et al*., 2006; Tian & Hua, 2010). Although β -cryptoxanthin and zeaxanthin have antioxidant properties, there are also strong indications that these oxidized carotenoids might play a role in cellular resistance to environmental stresses *e*.*g*. the stabilization of membrane structure at high temperatures (Tian & Hua, 2010).

In *T*. *thermophilus* it has been demonstrated that thermo(bis)zeaxanthins (comprising one or two sets of: zeaxanthin, glucose and branched fatty acids) aided in the stabilization of egg phosphatidylcholine liposomes over a temperature range of 30˚C to 80˚C (Hara *et al*., 1999). Thermozeaxanthins have a hydrophobic-hydrophilic-hydrophobic structure which causes thermozeaxanthins to be anchored in the membrane in the following fashion: the zeaxanthin moiety is embedded in the lipid bilayer, the glucose is exposed to the surface and the branched fatty acid moiety curls back into the lipid bilayer (Fig. 1.22) (Yokoyama *et al*., 1995). It has been suggested that the presence of carotenoids in the membrane might reduce membrane fluidity and reinforces the membrane, thus contributing to the survival of *T*. *thermophilus* HB27 at high temperatures (Yokoyama *et al*., 1995; Tian & Hua, 2010).

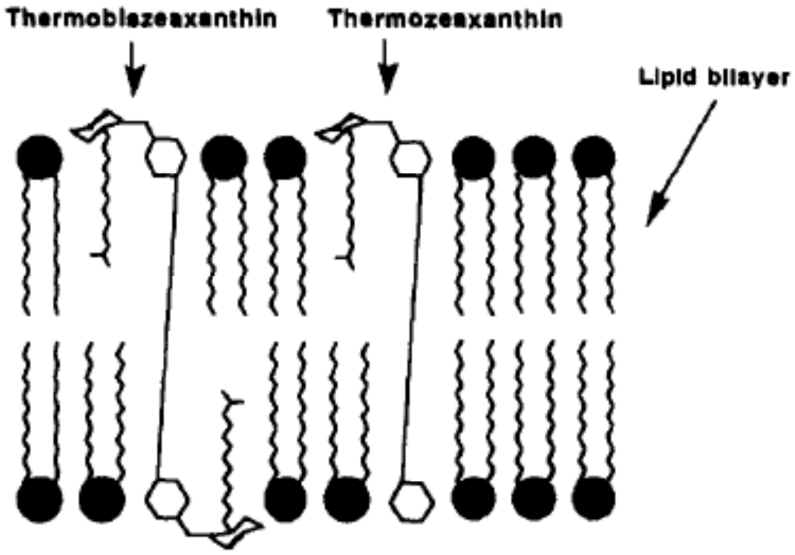


Fig. 1.22 The hydrophobic-hydrophilic-hydrophobic anchoring of thermo(bis)zeaxanthins from *T*. *thermophilus* HB27 in the lipid bilayer. Thermo(bis)zeaxanthins comprise zeaxanthin (embedded in the lipid bilayer), glucose (exposed to the surface of the cell) and the branched fatty acid that curls back into the lipid bilayer (Yokoyama *et al*., 1995).

1.3.1.3.1 Crystal structure of CYP175A1

The crystal structure of CYP175A1 was solved to a resolution of 1.8 Å by Yano *et al*. (2003) after heterologous expression in *E*. *coli*. At the time, CYP175A1 was the second thermostable CYP450 to have its crystal structure solved – the first being CYP119 from *S*. *acidocaldarius* (Yano *et al*., 2000). The CYP175A1 protein itself is 389 amino acids in size with a melting temperature (T_m) of 88˚C. CYP119 is slightly smaller than CYP175A1 and is 368 amino acids in size displaying a T_m = 91˚C (Yano *et al*., 2000). Interestingly, CYP175A1 displayed the highest amino acid identity (26%) to the P450BM3 from *Bacillus megaterium* and consequently P450BM3 was used as the search model for molecular replacement.

The overall structure of CYP175A1 exhibited a typical prism-like CYP450-fold (Fig. 1.23) comprising 17 α-helices and 11 β-strands. The core of CYP175A1 contained a four-helix

bundle consisting of helices: D, E, I and L as well as two α-helices namely J and K. The heme iron was embedded between the distal I-helix and proximal L helix with the cysteine (Cys-336) serving as the fifth axial thiolate ligand to the heme iron. The propionate side chains of the heme group interact with the protein *via* residues Trp-87, Arg-273 and Arg-334. The I-helix spanned the entire length of the molecule and also contained the highly conserved tryptophan residue at position 225 (Yano *et al*., 2003).

A region of particular importance in CYP175A1 is the substrate binding environment consisting of portions of the F and G helices (spanning from Ser-150 – Ser-170 and Asp-176 – Ile-198 respectively) as well as the B'-helix (spanning from Thr-65 – Ser-72) (Yano *et al*., 2003). Studies on the substrate-free and substrate-bound forms of P450BM3 indicated that the F and G helices and the loop connecting these helices display a large open/close motion (Li & Poulos, 1999). In addition, the F and G regions of CYP119 undergo large conformational changes, depending on the ligand bound to the active site (Yano *et al*., 2000). These two studies conducted on P450BM3 and CYP119 demonstrated the importance of the F/G region in substrate recognition. The F/G regions as well as the B'-helix from both CYP175A1 and P450BM3 were almost identical in size, location and sequence, which implied that CYP175A1 could possibly bind and hydroxylate palmitic acid like P450BM3. This was not the case since the structure of the substrate-bound form of P450BM3 superimposed onto the structure of CYP175A1 indicated steric crowding due to the presence of Gln-67, Trp-269 and Ile-270 in the active site of CYP175A1. P450BM3, on the other hand, contained Ala-74, Pro-329 and Ala-330 in the corresponding positions respectively, which have much smaller side chains, thus providing additional space for palmitoleic acid to fit into the substrate binding pocket (Yano *et al*., 2003). In addition, CYP175A1 does not possess the Arg-47 and Tyr-51 residues which both stabilize the carboxyl group of palmitoleic acid when bound to P450BM3 (Li & Poulos, 1997). Instead, the corresponding residues in CYP175A1 are Phe-41 and Leu-45.

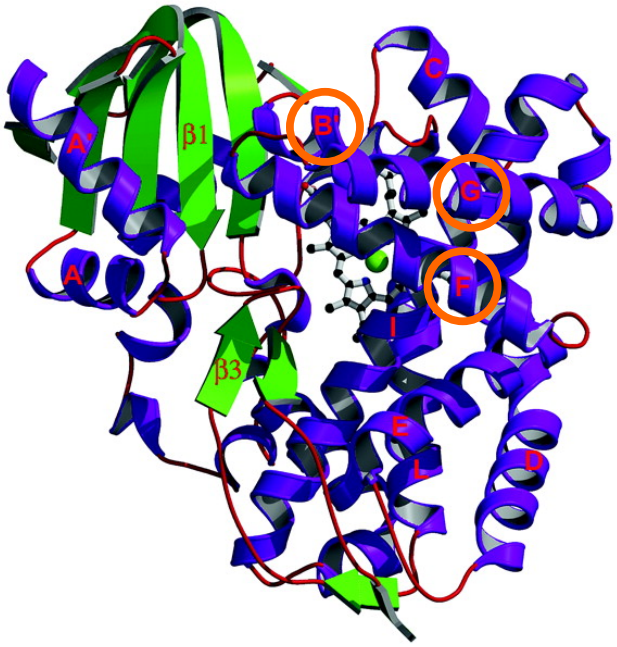


Fig. 1.23 Ribbon diagram of CYP175A1. α-Helices are indicated in purple while β-sheets are indicated in green. The F, G and B'-helices are encircled in orange on the diagram (Adapted from: Yano *et al*., 2003).

1.3.1.3.2 Structural factors conferring thermostability of CYP175A1

When CYP119 was crystallized, it was apparent that it had a compact structure and was quite small (368 amino acids) in comparison to other mesophilic CYP450 counterparts *e.g*. $P450_{cam}$ from *Pseudomonas putida* (414 amino acids) and $P450_{eryF}$ from *Saccharopolyspora erythraea* (403 amino acids) (Poulos *et al*., 1985; Cupp-Vickery & Poulos, 1995). The size of CYP175A1 is comparable to that of CYP119 although slightly larger at 389 amino acids (Yano *et al*., 2003). A distinct feature of CYP119, that was linked directly to its thermostability, was the presence of a large cluster of aromatic residues that is not found in other CYP450s. Large clusters of aromatic residues were however not found in CYP175A1 although both CYP175A1 and CYP119 have almost the same content of aromatic residues (10% and 9.8% respectively). Yano *et al*. (2003) however, did notice that both CYP119 and CYP175A1 possessed extensive salt bridge networks (Fig. 1.24) and is has been demonstrated that extensive salt bridge networks,

especially networks involving more than two charged residues, stabilize proteins at elevated temperatures (Elcock, 1998).

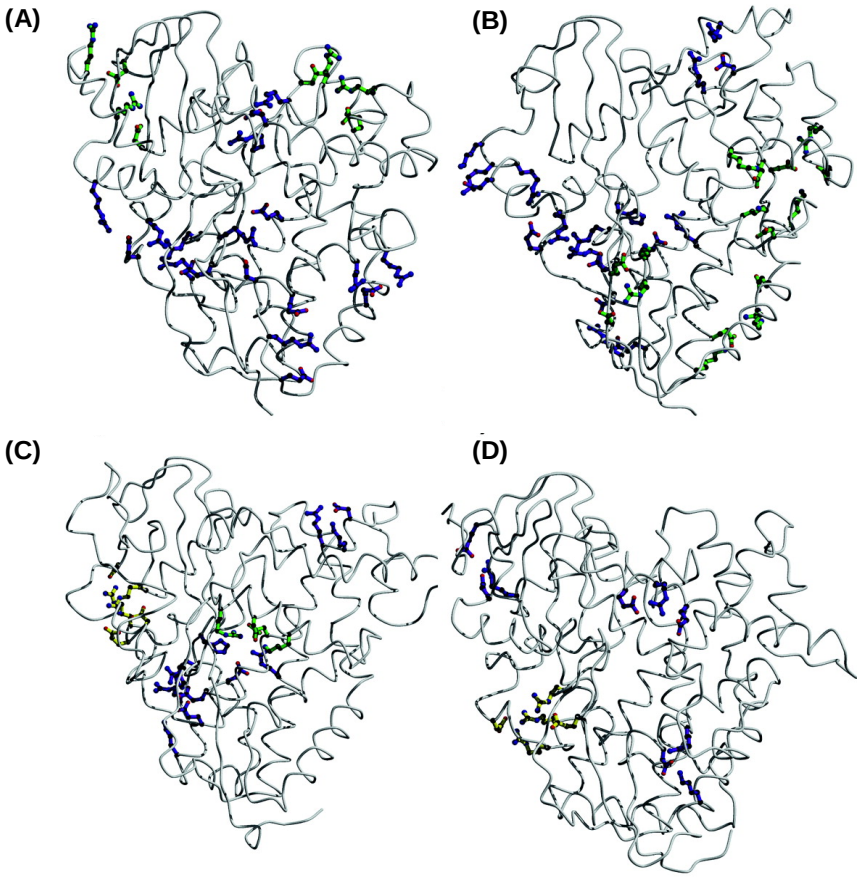


Fig. 1.24 Salt bridge networks in **(A)** CYP175A1, **(B)** CYP119, **(C)** $P450_{cam}$ and **(D)** P450BM 3. Three-, four- and five residue salt bridge networks are illustrated in blue, green and yellow respectively (Adapted from: Yano *et al*., 2003).

Yano *et al*. (2003) compared the number of salt bridge networks from CYP175A1 and CYP119 to that of some mesophilic CYP450s *e.g*. BM3, $P450_{cam}$ and $P450_{eryF}$ and found that CYP175A1 had eight and CYP119 ten salt bridge networks *versus* their mesophilic counterparts that had between four and six. In addition to the increased number of salt bridge networks in CYP175A1, the networks are also more evenly distributed over the surface of the structure when compared to the two mesophilic proteins (Fig. 1.24). Another significant feature regarding the salt bridge networks is the total fraction of salt

bridges that form part of a network: in the two thermostable CYP450s, *ca*. 60% of the total salt bridges are part of networks compared to *e.g*. the 36.8% from $P450_{cam}$.

1.3.1.3.3 Redox partners and co-factors of CYP175A1

Until very recently, the native redox partners of CYP175A1 were unknown but in spite of this, some researchers still achieved successful hydroxylation of β-carotene by using surrogate redox partners from mesophilic hosts in conjunction with heterologously expressed CYP175A1. In a study by Blasco *et al*. (2004), *CYP175A1* was heterologously expressed in *E*. *coli* together with genes necessary for β-carotene biosynthesis taken from the yellow pigmented bacterium *Erwinia uredovora* (Misawa *et al*., 1990). Blasco and co-workers (2004) managed to accumulate β-cryptoxanthin as well as zeaxanthin intracellularly, which implied that the catalytic activity of the thermophilic CYP175A1 was functionally supported by an unknown mesophilic redox partner(s) from *E*. *coli* – an organism that contains no endogenous CYP450s. The authors suspected that the soluble flavodoxin and flavodoxin reductase from *E*. *coli* probably served as redox partners that were able to transport the electrons to the heme iron of the monooxygenase. Momoi *et al*. (2006) used cell free extracts of *E*. *coli* containing heterologously expressed CYP175A1 and β-carotene suspension to form small amounts of β-cryptoxanthin at 37˚C. However, upon incubation of the cell free extract at 65˚C, the activity disappeared completely – an indication that the mesophilic redox partners from *E*. *coli* denatured. Momoi *et al*. (2006) also used purified preparations of flavodoxin and flavodoxin reductase from *E*. *coli* together with purified CYP175A1 and NADPH as electron donor to accumulate small amounts of β-cryptoxanthin as well as zeaxanthin at 37˚C. A fully reconstituted system, using purified preparations of putidaredoxin and putidaredoxin reductase from *Pseuodomonas* sp. and NADH as electron donor at 37˚C, also yielded β-cryptoxanthin and zeaxanthin with the turnover rate of CYP175A1 reaching 0.23 nmol β-cryptoxanthin.min^{-1}.$nmol^{-1}$ CYP175A1 under ideal conditions.

It was only in 2009 that Mandai *et al*. (2009a) reported on the native redox partners of CYP175A1, namely a thermostable and soluble ferredoxin (Fdx) and ferredoxin-$NAD(P)^+$ reductase. Cytochrome *c* reductase assays, using cytochrome *c* from equine heart and NADPH as electron donor, indicated that a ratio of 1:10 of FNR:Fdx displayed a five-fold

increase in activity as compared to the 1:1 ratio of redox partners. The reduction ability of FNR was also assessed with a ferricyanide assay using NADPH and NADH as electron donors. Kinetic data from this assay indicated that FNR had a stronger affinity for NADPH (K_m = 4.1 µM) as compared to NADH (K_m = 2440 µM) and the V_{max} values also reflected the FNR's preference for NADPH (V_{max} = 8318 nmol.min^{-1}.nmol^{-1} of FAD) as compared to NADH (V_{max} = 152 nmol.min^{-1}.nmol^{-1} of FAD). The authors concluded that the electron transport system of CYP175A1 belonged to Class I (see section 2.1.1) (Hannemann *et al*., 2007; Mandai *et al*., 2009a).

These thermostable redox partners were reconstituted with purified CYP175A1, NADPH and β-carotene at 65˚C to yield both β-cryptoxanthin and zeaxanthin (Mandai *et al*., 2009a). The rate of turnover for the reconstituted system at 65˚C was 12.4 nmol β-cryptoxanthin.min^{-1}.nmol^{-1} CYP175A1 – about 54 times greater than the turnover rate reported by Momoi *et al*. (2006). In addition to β-carotene hydroxylation at 65˚C, Mandai *et al*. (2009b) successfully hydroxylated testosterone at 65˚C using a fusion protein comprising CYP175A1, Fdx and FNR from *T*. *thermophilus* HB27. This study indicated it was possible to engineer a self-sufficient thermostable CYP450 for the oxidation of an unnatural substrate (Mandai *et al*., 2009b).

Table 1.1 Quintessential information pertaining to four CYP450s from four different extremophiles

	CYP119A1	***CYP119A2***	***CYP231A1***	***CYP175A1***
Natural host	***S. acidocaldarius***	***S. tokodaii***	***P. torridus***	***T. thermophilus* HB27**
Kingdom	*Archaea*	*Archaea*	*Archaea*	*Bacteria*
Temperature and pH optima of organism	75˚C - 80˚C pH 2.0 – 3.0	80˚C pH 2.5 – 3.0	60˚C pH 0.7	85˚C pH 7.0
Genome sequenced **GenBank accession**	Yes CP_000077	Yes NC_003106.2	Yes NC_005877.1	Yes NC_005838.1
CYP450 (s) location	2.22 Mb chromosome	2.69 Mb chromosome	1.54 Mb chromosome	232 605 bp megaplasmid
Number of CYP450s in genome and Locus tag	1 Saci_2081	1 ST1148	2 PTO1399 (*CYP231A2*)[a] PTO0085 (*CYP232A2*)	1 TTP0059
Heterologous expression in	*E. coli*	*E. coli*	*E. coli*	*E. coli*
P450 protein information	368 amino acids, Soluble, T_m = 91˚C	367 amino acids, Soluble, To be determined	352 amino acids, Soluble, T_m = 65˚C	389 amino acids, Soluble, T_m = 88˚C
Crystal structure resolution **Secondary structure information**	1.93 Å 13 α-helices and 12 β-sheets	1.94 Å 13 α-helices and 5 β-sheets	2.5 Å 12 α-helices and 5 antiparallel β-sheets	1.8 Å 17 α-helices and 11 β-sheets

Table 1.1 continued…

	S. acidocaldarius	***S. tokodaii***	***P. torridus***	***T. thermophilus* HB27**
Chief factor contributing to thermostability	Large network of aromatic residues	Large network of aromatic residues	Small size of protein	8 evenly distributed salt bridge networks
Redox partners **Co-factor / source of electrons**	Fdx and OFOR[b] Pyruvate	None required NAD(P)H	To be determined[c]	Fdx and FNR[d] NADH (K_m = 2440 µM) NADPH (K_m = 4.1 µM)
Substrates catalyzed	Lauric acid Styrene *cis*- and *trans*-β-methylstyrenes	Lauric acid Styrene	To be determined	β-carotene Testosterone
References	Wright *et al*., (1996) Koo *et al*., (2000) Yano *et al*., (2000) Chen *et al*., (2005)	Kawarabayasi *et al*., (2001) Suzuki *et al*., (2002) Oku *et al*., (2004) Matsumura *et al*., (2008) Matsumura *et al*., (2011)	Futterer *et al*., (2004) Ho *et al*., (2008)	Yano *et al*., (2003) Henne *et al*., (2004) Blasco *et al*., (2004) Mandai *et al*., (2009 a) Mandai *et al*., (2009 b)

[a] Although *P. torridus* has two known CYP450s, only CYP231A2 has been studied

[b] Fdx = Ferredoxin and OFOR = 2-oxoacid ferredoxin oxidoreductase

[c] *P. torridus* possesses four OFORs in its genome: pyruvate ferredoxin oxidoreductase, (α- chain) (Locus tag: PTO1360); 2-oxoacid ferredoxin oxidoreductase (β-subunit) (Locus tag: PTO1361); 2-oxoglutarate synthase, (α-chain) (Locus tag:PTO1000); 2-oxoglutarate ferredoxin oxidoreductase (β-subunit) (Locus tag: PTO0999)

[d] Fdx = Ferredoxin and FNR = Ferredoxin-$NAD(P)^+$ reductase

1.4 Concluding remarks

CYP450s form part of an ever-growing superfamily of heme-thiolate proteins that are found in all three kingdoms of life. These hemoproteins are external monooxygenases and are capable of incorporating a single atom of molecular oxygen into X-H bonds (X: -C, -N, -S) in a regio- as well as stereo-selective manner. CYP450s catalysis is not limited to only oxygen insertion but CYP450s are also capable of a plethora of other reactions *e.g*. dealkylation, dehydration, carbon-carbon bond cleavage and phenol ring coupling to name but a few (Lewis, 1996; Momoi *et al*., 2006; Urlacher & Eiben, 2006).

A large portion of well described CYP450s are from organisms that optimally function in environments that are governed by physico-chemical parameter(s) that fall within the anthropogenic definitions of 'life' according to biologists *i.e*. temperatures, pH, atmospheric pressure and oxygen content that are conducive and tolerable to human-, animal-, plant- and microbial life. Beyond this limited, anthropogenic view of life is a world in which a kaleidoscope of organisms (extremophiles) flourish in environments as diverse as Black Smokers at the bottom of ultra-deep oceans to volcanic basins and the Dead Sea saturated with salt (Ferreira *et al*., 1997; Madigan, 2000; Cavicchioli 2002; Seckbach & Oren, 2004). To date, the most well described CYP450s from extremophiles, namely two CYP119s, CYP231A2 and CYP175A1 are from thermoacidophilic archaea (*S*. *acidocaldarius, S. tokodaii* and *P*. *torridus*) and a thermophilic bacterium (*T*. *thermophilus* HB27) respectively (Wright *et al*., 1996; Blasco *et al*., 2004; Ho *et al*., 2008).

As far as the physiological roles of the above-mentioned CYP450s are concerned, very little is actually known except for CYP175A1 (which is a β-carotene hydroxylase). The native substrate(s) of CYP119 and CYP231A2 remain unknown although some researchers have been able to illustrate catalyses with CYP119. In addition, the native electron transport proteins for CYP175A1 and CYP119 are known – CYP175A1 utilizes a Class I electron transport system whereas CYP119 uses a Class IV electron transport system. The electron transport system for CYP231A2 remains unknown although electron transport proteins similar to that found in *S*. *acidocaldarius* and other archaea have been identified in the genome of *P*. *torridus*.

Thermostability is a common theme in CYP119, CYP231A2 and CYP175A1 but there does not seem to be a unifying mechanism with regard to their thermostability, except that they are all significantly smaller (352 - 389 amino acids) than prokaryotic CYP450s from mesophiles, which usually have more than 400 amino acids. Both the CYP119s as well as CYP175A1 utilize evenly distributed salt bridge networks to increase thermostability. The CYP119 also have a large conserved network of aromatic amino acid residues that contribute to their thermostability (Wright *et al*., 1996) (Yokoyama *et al*., 1995; Blasco *et al*., 2004). The thermostability of CYP231A2 is apparently largely due to its small size. The small size of CYP231A2 implies a decrease in surface to volume ratio due to shorter loops and the discarding of excess secondary structure not critical in the stabilization of CYP450 core structure (Ho *et al*., 2008). To date, our knowledge pertaining to the thermostability of these above-mentioned CYP450s have been deduced from their 3-D crystal structures and by site-directed mutagenesis studies. If we could fully understand the structural basis for thermal stability in the above-mentioned three CYP450s, it could be possible to engineer mesophilic CYP450s, that catalyze medically and industrially important reactions *e.g*. steroid hydroxylation or epoxidations, to be more thermostable.

The number of CYP450s from extremophiles from various and exotic extreme environments will no doubt grow as their metagenomes become more accessible to us. Consequently the potential for finding more novel CYP450s with interesting, useful and possibly industrially exploitable catalytic properties exists and will, no doubt, be taken advantage of in the future.

Literature cited

Balkwill, D.L., Kieft, T.L., Tsukuda, T., Konstandarithes, H.M., Onstott, T.C., Macnaughton, S., Bownas, J. and Fredrickson, J.K. (2004). Identification of iro-reducing *Thermus* strains as *Thermus scotoductus*. *Extremophiles* **8**, 37 – 44.

Barnes, H.J., Arlotto, M.P. and Waterman, M.R. (1991). Expression and enzymatic activity of recombinant cytochrome P450 17 alpha-hydroxylase in *Escherichia coli*. *Proceedings of the National Academy of Sciences U.S.A.* **88**, 5597 – 5601.

Bernhardt, R. (2006). Cytochrome P450 as versatile biocatalysts. *Journal of Biotechnology* **124**, 128 – 145.

Blasco, F., Kauffmann, I. and Schmid, R.D. (2004). CYP175A1 from *Thermus thermophilus* HB27, the first b-carotene hydroxylase of the P450 superfamily. *Applied Microbiology and Biotechnology* **64**, 671 – 674.

Brock, T.D., Brock, K.M., Belly, R.T. and Weiss, R.L. (1972). *Sulfolobus*: a new genus of sulfur oxidizing bacteria living at low pH and high temperature. *Archives of Microbiology* **84**, 54 – 68.

Budde, M., Maurer, S.C., Schmid, R.D. and Urlacher, V.B. (2005). Cloning, expression and characterization of CYP102A2, a sef-sufficient P450 monooxygenase from *Bacillus subtilis*. *Applied Microbiology and Biotechnology* **66**, 180 – 186.

Cavicchioli, R. (2002). Extremophiles and the search for extraterrestrial life. *Astrobiology* **2**, 281 - 292.

Chang, Y.T. and Loew, G.H. (2000). Homology modeling, molecular dynamics simulations, and analysis of CYP119, a P450 enzyme from the extreme acidothermophilic archaeon *Sulfolobus solfataricus*. *Biochemistry* **39**, 2484 – 2498.

Chen, L., Brügger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk H-P. and Garrett, R.A. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. *Journal of Bacteriology* **187**, 4992 – 4999.

Chung, A.P., Rainey, F.A., Valente, M., Nobre, M.F., and da Costa, M.S. (2000). *Thermus igniterrae* sp. nov. and *Thermus antranikianii* sp. nov., two new species from Iceland. *International Journal of Systematic and Evolutionary Microbiology* **50**, 209 – 217.

Cupp-Vickery, J.R. and Poulos, T.L. (1995). Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nature Structural Biology* **2**, 144 – 153.

De Mot, R. and Parret, A.H. (2002). A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends in Microbiology* **10**, 502 – 508.

Dopson, M., Baker-Austin, C., Hind, A., Bowman, J.P and Bond, P.L. (2004). Characterization of *Ferroplasma* isolates and *Ferroplasma acidarmanus* sp. nov., extreme acidophiles from acid mine drainage and industrial bioleaching environments. *Applied and Environmental Microbiology* **70**, 2079 – 2088.

Elcock, A.H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *Journal of Molecular Biology* **284,** 489 – 502.

Ferreira, A.C., Nobre, M.F., Rainey, F.A., Silva, M.T., Waite, R., Burghardt, J., Chung, A.P. and da Costa, M.S. (1997). *Deinococcus geothermalis* sp. nov. and *Deinococcus murrayi* sp. nov., two extremely radiation-resistant and slightly thermophilic species from hot springs. *International Journal of Systematic Bacteriology* **47**, 939 – 947.

Froehlich, J.E., Itoh, A. and Howe, G.A. (2001). Tomato allene oxide synthase and fatty acid hydroperoxide lyase, two cytochrome 450s involved in oxylipin metabolism, are targeted to different membranes of chloroplast envelope. *Plant Physiology* **125**, 306 – 317.

Futterer, O., Angelov, A., Liesegang, H. Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G. and Liebl, W. (2004). Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proceeding of the National Academy of Sciences U.S.A.* **101**, 9091 – 9096.

Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry* **267**, 83 – 90.

Gustafsson, M.C., Roitel, O., Marshall, K.R., Noble, M.A., Chapman, S.K., Pessegueiro, A., Fulco, A.J., Cheesman, M.R., von Wachenfeldt, C. and Munro, A.W. (2004). Expression, purification, and characterization of *Bacillus subtilis* cytochromes P450 CYP102A2 and CYP102A3: Flavocytochrome homologues of P450 BM3 from *Bacillus megaterium*. *Biochemistry* **43**, 5474 – 5487.

Hannemann, F., Bichet, A., Ewen, K.M. and Bernhardt, R. (2007). Cytochrome P450 systems - biological variations of electron transport chains**.** *Biochimica et Biophysica Acta* **1770**, 330 – 344.

Hara, M., Yuan, H., Yang, Q., Hoshino, T., Yokoyama, A. and Miyake, J. (1999). Stabilization of liposomal membranes by thermozeaxanthins: carotenoid-glucoside esters. *Biochimica Biophysica Acta* **1461**, 147 – 154.

Hawkes, D.B., Adams, G.W., Burlingame, A.L. and Ortiz de Montellano, P.R. (2002). Cytochrome P450(cin) (CYP176A), isolation, expression and characterization. *Journal of Biological Chemistry* **277**, 27725 – 27732.

Henne, A., Brüggemann, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., Johann, A., Lienard, T., Gohl, O., Martinez-Arias, R., Jacobi, C., Starkuviene, V., Schlenczeck, S., Dencker, S., Huber, R., Klenk, H.P., Kramer, W., Merkl, R., Gottschalk, G. and Fritz, H.J. (2004). The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nature Biotechnology* **22**, 547 – 553.

Ho, W.W., Li, H., Nishida, C.R., Ortiz de Montellano, P.R. and Poulos, T.L. (2008). Crystal structure and properties of CYP231A2 from the thermoacidophilic archaeon *Picrophilus torridus*. *Biochemistry* **47**, 2071 – 2079.

Hoshino, T., Fujii, R. and Nakahara, T. (1993). Molecular cloning and sequence analysis of the *crtB* gene of *Thermus thermophilus* HB27, an extreme thermophile producing carotenoid pigments. *Applied and Environmental Microbiology* **59**, 3150 – 3153.

Hunter, D.J., Roberts, G.A., Ost, T.W., White, J.H., Muller, S., Turner, N.J., Flitsch, S.L. and Chapman, S.K. (2005). Analyses of the domain properties of the novel cytochrome P450 RhF. *FEBS Letters* **579**, 2215 – 2220.

Itoh, A. and Howe, G.A. (2001). Molecular cloning of a divinyl ether synthase. Identification as a CYP74 cytochrome P-450. *Journal of Biological Chemistry* **276**, 3620 – 3627.

Jackson, C.J., Lamb, D.C., Marczylo, T.H., Warrilow, A.G., Manning, N.J., Lowe, D.J., Kelly, D.E. and Kelly, S.L. (2002). A novel sterol 14alpha-demethylase/ferredoxin fusion protein (MCCYP51FX) from *Methylococcus capsulatus* represents a new class of the cytochrome P450 superfamily. *Journal of Biological Chemistry* **277**, 46959 – 46965.

Jenkins, C.M. and Waterman, M.R. (1994). Flavodoxin and NADPH-flavodoxin reductase from *Escherichia coli* support bovine cytochrome P450c17 hydroxylase activities. *Journal of Biological Chemistry* **269**, 27401 – 27408.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T. and Kikuchi, H. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Research* **8**, 123 – 140.

Kitazume, T., Takaya, N. Nakayama, N. and Shoun, H. (2000). *Fusarium oxysporum* fatty-acid subterminal hydroxylase (CYP505) is a membrane-bound eukaryotic counterpart of *Bacillus megaterium* cytochrome P450BM3. *Journal of Biological Chemistry* **275**, 39734 – 39740.

Kitazume, T., Tanaka, A., Takaya, N., Nakamura, A., Matsuyama, S., Suzuki, T. and Shoun, H. (2002). Kinetic analysis of hydroxylation of saturated fatty acids by recombinant P450foxy produced by an *Escherichia coli* expression system. *European Journal of Biochemistry* **269**, 2075 – 2082.

Klingenberg, M. (1958). Pigments of rat liver microsomes. *Archives in Biochemistry and Biophysics* **75**, 376 – 386.

Koo, L.S., Tschirret-Guth, R.A., Straub, W.E., Moënne-Loccoz, P., Loehr, T.M. and Ortiz de Montellano, P.R. (2000). The active site of the thermophilic CYP119 from *Sulfolobus solfataricus*. *The Journal of Biological Chemistry* **275**, 14112 – 14123.

Koo, L.S., Immoos, C.E., Cohen, M.S., Farmer, P.J. and Ortiz de Montellano, P.R. (2002). Enhanced electron transfer and lauric acid hydroxylation by site-directed mutagenesis of CYP119. *Journal of the American Chemical Society* **124**, 5684 – 5691.

Kudo, T., Tomura, D., Liu, D.L., Dai, X.Q. and Shoun, H. (1996). Two isozymes of P450nor of *Cylindrocarpon tonkinense*: molecular cloning of the cDNAs and genes, expressions in the yeast, and the putative NAD(P)H-binding site. *Biochimie* **78**, 792 – 799.

Kushner, D.J. (ed.) (1978). Microbial Life in Extreme Environments, Academic Press, London.

Lau, S.M., Harder, P.A. and O'Keefe, D.P. (1993). Low carbon monoxide affinity allene oxide synthase is the predominant cytochrome P450 in many plant tissues. *Biochemistry* **32**, 1945 – 1950.

Lewis, D.F.V. (1996). Cytochrome P450: structure function and mechanism. Taylor & Francis, London.

Li, H. and Poulos, T.L. (1999). Fatty acid metabolism, conformational change, and electron transfer in cytochrome P-450(BM3). *Biochimica et Biophysics Acta* **144**, 141 – 149.

Liu, L, Schmid, R.D. and Urlacher, V.B. (2006). Cloning, expression, and characterization of a self-sufficient cytochrome P450 monooxygenase from *Rhodococcus ruber* DSM 44319. *Applied Microbiology and Biotechnology* **72**, 876 – 882.

Madigan, M.T. (2000). Bacterial habitats in extreme environments. In: Journey to Diverse Microbial Worlds. Adaptation to Exotic Environments (J. Seckbach, ed.), pp. 61 – 72. Kluwer Academic Publishers, Dordrecht.

Mandai, T., Fujiwara, S. and Imaoka, S. (2009a). A novel electron transport system for thermostable CYP175A1 from *Thermus thermophilus* HB27. *FEBS Journal* **276**, 2416 – 2429.

Mandai, T., Fujiwara, S. and Imaoka, S. (2009b). Construction and engineering of a thermostable self-sufficient cytochrome P450. *Biochemical and Biophysical Research Communications* **384**, 61 – 65.

Matsumura, H., Wakatabi, M., Omi, S., Ohtaki, A., Nakamura, N., Yohda, M. and Ohno, H. (2008). Modulation of redox potential and alteration in reactivity *via* the peroxide shunt pathway by mutation of cytochrome P450 around the proximal heme ligand. *Biochemistry* **47**, 4834 – 4842.

Matsumura, H., Matsuda, K., Nakamura, N., Ohtaki, A., Yoshida, H., Kamitori, S., Yohda M. and Ohno, H. (2011). Monooxygenation by a thermophilic cytochrome P450 *via* direct electron donation from NADH. *Metallomics* **3**, 389 – 395.

Maves, S.A. and Sigar, S.G. (2001). Understanding thermostability in cytochrome P450 by combinational mutagenesis. *Protein Science* **10**, 161 -168.

McLean, K.J., Sabri, M., Marshall, K.R., Lawson, R.J., Lewis, D.G., Clift, D., Balding, P.R., Dunford, A.J., Warman, A.J., McVey, J.P., Quin, A.-M., Sutcliffe, M.J., Scrutton, N.S. and Munro, A.W. (2005). Biodiversity of cytochrome P450 redox systems. *Biochemical Society Transactions* **33**, 796 – 801.

Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., Harashima, K. (1990). Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of the gene products expressed in *Escherichia coli*. *Journal of Bacteriology* **172**, 6704 – 6712.

Miura, Y. and Fulco, A.J. (1974). (Omega-2) hydroxylation of fatty acids by a soluble system from *Bacillus megaterium*. *Journal of Biological Chemistry* **249**, 1880 – 1888.

Momoi, K., Hofmann, U., Schmid, R.D. and Urlacher, V.B. (2006). Reconstitution of β-carotene hydroxylase activity of thermostable CYP175A1 monooxygenase. *Biochemical and Biophysical Research Communications* **339**, 331 – 336.

Munro, A.W., Girvan, H.M. and McLean, K.J. (2007). Variations on a (t)heme – novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Natural Products Reports* **24**, 585 – 609.

Nakahara, K., Tanimoto, T., Hatano, K., Usuda, K. and Shoun, H. (1993). Cytochrome P-450 55A1 (P-450dNIR) acts as nitric oxide reductase employing NADH as the direct electron donor. *Journal of Biological Chemistry* **268**, 8350 – 8355.

Nakayama, N., Takemae, A. and Shoun, H. (1996). Cytochrome P450foxy, a catalytically self-sufficient fatty acid hydroxylase of the fungus *Fusarium oxysporum*. *Journal of Biochemistry* (Tokyo) **119**, 435 – 440.

Narhi, L.O. and Fulco, A.J. (1986). Characterization of a catalytically self sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in *Bacillus megaterium*. *Journal of Biological Chemistry* **261**, 7160 – 7169.

Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J., Waterman, M.R., Gotoh, O., Coon, M.J., Estabrook, R.W., Gunsalus, I.C. and Nebert, D.W. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6**, 1 – 42.

Nishida, C.R. and Ortiz de Montellano, P.R. (2005). Thermophilic cytochrome P450 enzymes. *Biochemica et Biophysics Research Communications* **338**, 437 – 445.

Oku, Y., Ohtaki, A., Kamitori, S., Nakamura, N., Yohda, M., Ohno, H. and Kwarabayashi, Y. (2004). Structure and direct electrochemistry of cytochrome P450 form the thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* str. 7. *Journal of Inorganic Biochemistry* **98**, 1194 – 1199.

Ortiz de Montellano, P.R. (1995). Cytochrome P450: Structure, Mechanism and Biochemistry, 2nd Edition. (Ortiz de Montellano, P.R. ed), pp. 575 – 585, Plennum Press, New York (USA).

Oshima T. and Imahori, K. (1974). Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a non-sporulating thermophilic bacterium form a Japanese thermal spa. *International Journal of Systematic Bacteriology* **24**, 102 - 112.

Park, S.Y., Yamane, K., Adachi, S., Shiro, Y., Weiss, K.E. and Sligar, S.G. (2000). Crystallization and preliminary X-ray diffraction analysis of a cytochrome P450 (CYP119) from *Sulfolobus solfataricus*. *Acta Crystallographica Section D* **56**, 1173 – 1175.

Poulos, T.L., Finzel, B.C., Gunsalus, I.C., Wagner, G.C., and Kraut, J. (1985). The 2.6-A crystal structure of *Pseudomonas putida* cytochrome P-450. *Journal of Biological Chemistry* **260**, 16122 – 16130.

Proctor, R.H., Brown, D.W., Plattner, R.D. and Desjardins, A.E. (2003). Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in *Gibberella moniliformis*. *Fungal Genetics and Biotechnology* **38**, 237 – 249.

Puchkaev, A.V., Wakagi, T. and Ortiz de Montellano, P.R. (2002). CYP119 plus a *Sulfolobus tokodaii* str. 7 ferredoxin and 2-oxoacid ferredoxin oxidoreductase constitute a high temperature cytochrome P450 catalytic system. *Journal of the American Chemical Society* **124**, 12682 – 12683.

Puchkaev, A.V., Koo, L.S. and Ortiz de Montellano, P.R. (2003). Aromatic stacking as a determinant of the thermal stability of CYP119 from *Sulfolobus solfataricus*. *Archives of Biochemistry and Biophysics* **409**, 52 – 58.

Puchkaev, A.V. and Ortiz de Montellano, P.R. (2005). The *Sulfolobus solfataricus* electron donor partners of thermophilic CYP119: An unusual non-NAD(P)H-dependent cytochrome P450 system. *Archives in Biochemistry and Biophysics* **434**, 169 – 177.

Roberts, G.A., Grogan, G., Greter, A. and Flitsch, S.L. (2002). Identification of a new class of cytochrome P450 from a *Rhodococcus* sp. *Journal of Bacteriology* **184**, 3898 – 3908.

Roberts, G.A., Celik, A., Hunter, D.J., Ost, T.W., White, J.H., Chapman, S.K., Turner, N.J. and Flitsch, S.L. (2003). A self-sufficient cytochrome P450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox center. *Journal of Biological Chemistry* **278**, 48914 – 48920.

Rothschild, L.J. and Mancinelli, R.L. (2001). Life in extreme environments. *Nature* **409**, 1092 – 1101.

Rylott, E.L., Jackson, R.G., Edwards, J., Womack, G.L., Seth-Smith, H.M., Rathbone, D.A., Strand, S.E. and Bruce, N.C. (2006). An explosive-degrading cytochrome P450 activity and its targeted application for the phyto-remediation of RDX. *Nature Biotechnology* **24**, 216 – 219.

Schneider, S., Marles-Wright, J., Sharp, K.H. and Paoli, M. (2007). Diversity and conservation of interactions for binding heme in b-type heme proteins. *Natural Products Report* **24**, 621 – 630.

Seckbach, J. and Oren, A. (2004). Introduction to the extremophiles. In: Origins (J. Seckbach, ed.), pp. 373 – 393. Kluwer Academic Publishers, Dordrecht.

Seo, J.A., Proctor, R.H. and Plattner, R.D. (2001). Characterization of four clustered and co-regulated genes associated with fumonisin biosynthesis in *Fusarium verticilliodes*. *Fungal Genetics and Biotechnology* **34**, 155 – 165.

Serizawa, N. and Matsuoka, T. (1991). A two component-type cytochrome P-450 monooxygenase system in a prokaryote that catalyses hydroxylation of ML-236B to pravastatin, a tissue-selective inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Biochimica et Biophysics Acta* **1084**, 35 – 40.

Shibata, Y., Matsui, K., Kajiwara, T. and Hatanaka, A. (1995). Fatty acid hydroperoxide lyase is a heme protein. *Biochemistry and Biophysics Research Communications* **207**, 438 – 443.

Smith, G.C., Tew, D.G. and Wolf, C.R. (1994). Dissection of NADPH-cytochrome P450 oxidoreductase into distinct functional domains. *Proceedings of the National Academy of Sciences U.S.A.* **91**, 8710 – 8714.

Suzuki T., Iwasaki, T., Uzawa, T., Hara, K., Nemoto, N., Kon, T., Ueki, T., Yamagishi, A. and Oshima, T. (2002). *Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan. *Extremophiles* **6**, 39 – 44.

Takaya, N., Suzuki, S., Kuwazaki, S., Shoun, H., Maruo, F., Yamaguchi, M. and Takeo, K. (1999). Cytochrome P450nor, a novel class of mitochondrial cytochrome P450 involved in nitrate respiration in the fungus *Fusarium oxysporum*. *Archives in Biochemistry and Biophysics* **372**, 340 – 346.

Tian, B. and Hua, Y. (2010). Carotenoid biosynthesis in extremophilic *Deinococcus-Thermus* bacteria. *Trends in Microbiology* **18**, 512 – 520.

Ullrich, V., Zou, M.H. and Bachschmid, M. (2001). New physiological and pathophysiological aspects on the thromboxane A(2)-prostacyclin regulatory system. *Biochimica et Biophysics Acta* **1532**, 1 – 14.

Urlacher, V.B. and Eiben, S. (2006). Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends in Biotechnology* **24**, 324 – 330.

Urlacher, V.B., Lutz-Wahl, S. and Schmid, R.D. (2004). Microbial P450 enzymes in biotechnology. *Applied Microbiology and Biotechnology* **64**, 317 – 325.

Werck-Reichhart, D. and Feyereisen, R. (2000). Cytochrome P450: a success story. *Genome Biology* **1** (REVIEWS3003).

Williams R.A.D., Smith, K.E., Welch, S.G., Micallef, J., Sharp, R.J. (1995). DNA relatedness of *Thermus* strains, description of *Thermus brockianus* sp. nov., and proposal to re-establish *Thermus thermophilus* (Oshima and Imahori). *International Journal of Systematic Bacteriology* **46**, 495 – 499.

Wright, R.L., Harris, K., Solow, B., White, R.H. and Kennelly, P.J. (1996). Cloning of a potential cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *FEBS Letters* **384**, 235 – 239.

Wu, J.T., Wu, L.H. and Knight, J.A. (1986). Stability of NADPH: Effect of Various Factors on the Kinetics of Degradation. *Clinical Chemistry* **32**, 314 – 319.

Wu, K.K. and Liou, J.Y. (2005). Cellular and molecular biology of prostacyclin synthase. *Biochemistry and Biophysics Research Communications* **338**, 45 – 52.

Yano, J.K., Blasco, F., Li, H., Schmid, R., Henne, A. and Poulos, T.L. (2003). Preliminary characterization and crystal structure of a thermostable cytochrome P450 from *Thermus thermophilus*. *The Journal of Biological Chemistry* **278**, 608 – 616.

Yano, J.K., Koo, L.S., Schuller, D.J., Li, H., Ortiz de Montellano, P.R. and Poulos, T.L. (2000). Crystal structure of a thermophilic cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *The Journal of Biological Chemistry* **275**, 31066 – 31092.

Yokoyama, A., Sandmann, G., Hoshino, T., Adachi, K., Sakai, M. and Shizuri, Y. (1995). Thermozeaxanthins, new carotenoid-glycoside-esters from thermophilic eubacterium *Thermus thermophilus*. *Tetrahedron Letters* **36**, 4901 – 4904.

Zhang, Q., Iwasaki, T., Wakagi, T. And Oshima, T. (1996). 2-Oxoacid:ferredoxin oxidoreductase from the thermophilic archaeon, *Sulfolobus* sp. strain 7. *Journal of Biochemistry* **120**, 58 – 599.

Zhang, L., Takaya, N., Kitazume, T., Kondo, T. and Shoun, H. (2001). Purification and cDNA cloning of nitric oxide reductase cytochrome P450nor (CYP55A4) from *Trichosporon cutaneum*. *European Journal of Biochemistry* **268**, 3198 – 3204.

Chapter 2

Cytochrome P450 Monooxygenases from the genus *Thermus*

2.1 Introduction

A distinctive feature of many *Thermus* spp. including *Thermus thermophilus, Thermus* sp. NMX2.A1, *Thermus aquaticus*, *Thermus filiformis*, *Thermus brockianus*, *Thermus oshimai*, *Thermus igniterrae* and *Thermus antranikianii*, is their yellow pigmentation as a result of the carotenoids these bacteria synthesize (Chung *et al*., 2000). However, several colourless *Thermus* strains have also been isolated (Ramaley & Hixson, 1970; Brock & Boylen, 1973; Kristjansson *et al*., 1994; Williams *et al*., 1996; Kieft *et al*., 1999). Non-pigmented strains originate from sources devoid of sunlight (Williams *et al*., 1996) and conversely there seems to be a correlation between yellow pigmented *Thermus* strains and the fact that they were isolated from thermal environments regularly exposed to sunlight. This proposed correlation between pigmentation and sunlight seems plausible since it has been demonstrated by Hoshino *et al*. (1994) and Tabata *et al*. (1994) that carotenoids protect *Thermus* against ultraviolet irradiation.

The *CYP175A1* from *T*. *thermophilus* HB27 is the only *Thermus* CYP450 gene that has been fully characterized after heterologous expression in *Escherichia coli*. The crystal structure was solved to a resolution of 1.8 Å by Yano *et al*. (2003) and Blasco *et al*. (2004) demonstrated that *CYP175A1* adds hydroxyl groups to both β rings of β-carotene to form β,β-carotene-3,3'-diol (zeaxanthin). The native electron transport system of *CYP175A1*, namely a ferredoxin and ferredoxin NAD(P)$^+$ reductase, has also been described and shown to efficiently shuttle electrons to *CYP175A1* to catalyze the hydroxylation of β-carotene at 65°C (Mandai *et al*., 2009).

Genome sequencing has revealed that the *CYP175A1* genes of *T. thermophilus* HB27 and HB8 are located on megaplasmids. Directly upstream from these *CYP175A1* genes are gene clusters that encode the proteins necessary for β-carotene biosynthesis (Tabata *et al*., 1994; Henne *et al*., 2004). The recently shotgun-sequenced and currently

unassembled genome of *Thermus aquaticus* Y51ML23 has also revealed a CYP450 gene, as well as a β-carotene gene cluster very similar to those found on the megaplasmids of the *T. thermophilus* strains. Other similar gene clusters, also located on megaplasmids, have been identified in pigmented bacteria such as *Pantoea agglomerans* (synonym: *Erwinia herbicola*) (Hundle *et al*., 1994) and *R. capsulatus* (Armstrong *et al*., 1989).

In our group *Thermus scotoductus* SA-01 has received much attention after its isolation from an ultra-deep South African gold mine by Kieft and co-workers (1999). This bacterium's ability to reduce heavy metals *e.g*. Co^{3+}, Mn^{4+}, U^{6+} and Cr^{6+} at 65˚C has been well described (Kieft *et al*., 1999; Balkwill *et al*., 2004; Opperman & van Heerden, 2007). Given the fact that CYP450s are such versatile biocatalysts and since only three CYP450s from the *Thermus* genus have to date been identified, our initial aim was to isolate a *CYP175* gene from *T. scotoductus* SA-01. This aim was later amended to include *Thermus* sp. NMX2.A1, a pigmented strain that is phylogenetically closely related to the colourless *T. scotoductus* SA-01.

2.2 Aims

The aims of this study were to:

(i) amplify *CYP175* genes from *T. scotoductus* SA-01 and a closely related strain *Thermus* sp. NMX2.A1,

(ii) to identify the native CYP450 redox partners in *T. scotoductus* SA-01 and

(iii) demonstrate the physiological role of CYP175 in these two strains

2.3 Materials and Methods

2.3.1 Microbiological methods

2.3.1.1 Strains, plasmids, media and growth conditions

Propagation of plasmids in *Escherichia coli* was performed in Luria-Bertani (LB) broth (Sambrook *et al*., 1989) at 37˚C with agitation at 160 r.p.m. Selective pressure was maintained by supplementing the LB broth with antibiotics with the following final concentrations: 100 µg/mL ampicillin, 50 µg/mL streptomycin or 34 µg/mL chloramphenicol. Solid media cultivations were performed by supplementing broth with 15 g/L bacteriological agar and selective pressure was maintained with 60 µg/mL (final concentration) ampicillin, 50 µg/mL (final concentration) streptomycin or 34 µg/mL (final concentration) chloramphenicol.

T. *scotoductus* SA-01 was cultured at 65˚C in TYG broth as described by Opperman and van Heerden (2007). TYG broth contained (per 1 L): 5 g tryptone, 3 g yeast extract and 1 g D-glucose. *Thermus* sp. NMX2.A1 was cultured at 65˚C with agitation at 100 r.p.m. for 24 h in ATCC medium 697 (Kieft *et al*., 1999). ATCC medium 697 contained (per 1 L): 4 g yeast extract, 8 g peptone and 2 g NaCl made up to 1 L with distilled water, pH 7.5. All bacterial strains were cryopreserved at -80˚C in 25% (v/v) glycerol (final concentration). Strains and plasmids used in this study are summarized in Table 2.1.

Table 2.1 Bacterial strains and plasmids used in this study

Strains / Plasmids	Genotype / characteristics	Source / Reference
Thermus scotoductus SA-01	Cream coloured, Isolated 3.2 km below subsurface	Witwatersrand, South Africa Kieft *et al*., (1999) Balkwill *et al*., (2004)
Thermus sp. NMX2.A1	Yellow pigmented, Isolated from hot spring in New Mexico desert	Prof. T.L. Kieft (New Mexico Institute of Mining and Technology, USA); Hudson *et al*., (1986)
Escherichia coli XL-10 Gold	Tet^{r} D(*mcr*A)183 D(*mcr*CB-*hsd*SMR-mrr)173*end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* Hte [F′ *pro*AB *lac*I^{q}ZDM15 *Tn*10 (Tet^{R}) Amy Cam^{R}	Stratagene
Escherichia coli TOP10	F-*mcr*A∆ *(mrr-hsd*RMS-*mcr*BC*)* Φ80*lac*Z ∆M15 ∆*lac*X74 *rec*A1 *ara*D139 ∆*(ara leu) 7697 gal*U *gal*K *rps*L *(*Str^{R}*) end*A1 *nup*G	Invitrogen
Escherichia coli BL21(DE3) (containing pRARE2)[a]	F– *omp*T *hsd*SB(rB–, mB–) *gal dcm* (DE3) Cam^{R}	Invitrogen
pGEM®-T Easy	Cloning vector for blue/white selection, TA-cloning, Amp^{R}	Promega
pET22b(+)	Expression vector, C-terminal 6x His tag, Amp^{R}	Novagen
- Constructs in pET22b(+)	pET22b(+): P450 NMX2.A1 pET22b(+): GDH1 *S*. *solfataricus*	
pCDFDuet-1	Co-expression vector, Dual multiple cloning sites, *aad*A (streptomycin/ spectinomycin resistance)	Novagen
- Constructs in pCDFDuet-1	pCDFDuet-1: FNR *T*. *scotoductus* SA-01 pCDFDuet-1: Fdx *T*. *scotoductus* SA-01 pCDFDuet-1: Fdx + FNR *T*. *scotoductus* SA-01	

[a] This particular *E*. *coli* BL21(DE3) is a lab strain constructed by Dr. J. van Marwijk (University of the Free State). The pRARE2 plasmid was isolated from *E*. *coli* Rosetta-gami™ 2 (Novagen) and encodes for seven rare codon tRNA genes and confers chloramphenicol resistance to *E*. *coli* BL21(DE3).

2.3.2 Recombinant DNA techniques

2.3.2.1 Enzymes, chemicals, kits and other consumables

All DNA modifying enzymes with their respective buffers, PCR reagents and DNA purification kits were, unless stated otherwise, obtained from Bioflux Corporation, Applied Biosystems (ABI), Fermentas, New England Biolabs (NEB) or Roche Molecular Biochemicals.

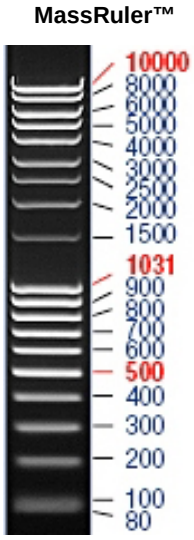
Chemicals used in this study were of an analytical or molecular biology grade and were, unless stated otherwise, obtained from either Merck or Sigma-Aldrich. All oligonucleotides were, unless stated otherwise, obtained from either Inqaba Biotechnical Industries (South Africa) or Integrated DNA Technologies (IDT). Oligonucleotide design and analyses were performed using an online algorithm from IDT with default settings. Oligonucleotides (T7 and SP6) for sequencing purposes were obtained from IDT. All relevant oligonucleotides used in this study are collated in Table 2.2.

Table 2.2 Oligonucleotide primers used in this study

Primer name	5´ – 3´ DNA sequence	Application	Restriction sites/ Comments
P450_F_1	5′-ggc ctc ctc acc gac tgg gg-3′	*Thermus* P450 screen	Anneals to conserved I-helix motif
P450_F_2	5'- gtg gcg ggc cac gag acg -3'	*Thermus* P450 screen	Anneals to conserved K-helix motif
P450_R_1	5′-cgt ctc gtg gcc cgc cac-3′	*Thermus* P450 screen	Inverse sequence of P450_F_2
P450_R_2	5'- ccc cag gca aag cct ctg cc -3'	*Thermus* P450 screen	Anneals to conserved heme binding loop
Lyco_F	5'- aag gac ccc gtg gtg gac ctg -3'	Amplifies *ca*. 1 500 bp amplicon containing *inter alia* a full P450 ORF	Anneals to gene directly upstream from P450
Conserv_R	5'- ggg cga tga ggg gca tga gga -3'	Amplifies *ca*. 1 500 bp amplicon containing *inter alia* a full P450 ORF	Anneals to gene directly downstream from P450
pET22_NdeI_F	5'- cat atg aag cgc ctt tcc ctg agg gag gc -3'	Amplifies P450 ORF from *Thermus* sp. NMX2.A1	*Nde*I
pET22_HindIII_R	5'- aag ctt tca cgc ccg cac ccc -3'	Amplifies P450 ORF from *Thermus* sp. NMX2.A1	*Hin*dIII
pCDFdx_NcoI-F	5'- cca tgg cgc acg tga tct gtg aac cc -3'	Amplifies Fdx ORF from *T. scotoductus* SA-01	*Nco*I
pCDFdx_HindIII_R	5'- aag ctt cta gcc caa gcc ggc cag -3'	Amplifies Fdx ORF from *T. scotoductus* SA-01	*Hin*dIII
pCDFNR_NdeI_F	5'- cat atg gag cac acc gac gtg atc atc att ggt gc-3'	Amplifies FNR ORF from *T. scotoductus* SA-01	*Nde*I
pCDFNR_XhoI_R	5'- ctc gag tca ggc cgg tgc ttt ctc ctc -3'	Amplifies FNR ORF from *T. scotoductus* SA-01	*Xho*I
GDH1_SUL_F_Nde	5'- cat atg aaa gct ata ata gtg aaa ccc cca aac -3'	Amplifies *GDH*1 from *S. solfataricus* P2[a]	*Nde*I
GDH1_SUL_R_Xho	5'- ctc gag tta ttc cca taa tat tct tat ctt gat ttc -3'	Amplifies *GDH*1 from *S. solfataricus* P2[a]	*Xho*I

[a]Oligonucleotide design and PCR amplification of *GDH*1 was performed by Dr. J. van Marwijk. Genomic DNA from *S. solfataricus* P2 was kindly provided by Dr. D.J. Opperman (University of the Free State).

All molecular weight markers were, unless stated otherwise, obtained from Fermentas. The molecular weight marker used in this chapter is shown below:



2.3.2.2 Quantification of nucleic acids

Nucleic acid concentration and purity was determined with a Nanodrop® ND-1000 Spectrophotometer.

2.3.2.3 PCR amplifications

For PCR experiments using the Expand Long Template system, the reaction mixture contained components with the following final concentrations: 1 x Buffer no. 1 (unless otherwise stated), 300 nM forward and reverse oligonucleotides, 350 nM dNTP's, 2 µg/mL BSA, at least 10 ng DNA template and 3.75 U Expand Long Template Polymerase mix. Reactions were made up to 50 µL with PCR grade water. Thermal cycling was performed with an Eppendorf Mastercycler Gradient (Table 2.3)

In most cases PCR analyses was performed by using the whole cells directly from culture or by lysing the cells beforehand and then using the lysate as the source of crude

genomic DNA. When appropriate, genomic DNA was extracted and purified from axenic cultures (Labuschagne & Albertyn, 2007).

Table 2.3 Standard PCR reaction for the Expand Long Template system

Step	Temperature	Time	Number of Cycles
Initial denaturation	94˚C	2 min	1
Denaturation	94˚C	10 sec	
Annealing	55˚C	30 sec	10
Elongation	68˚C	4 - 8 min [a]	
Denaturation	94˚C	10 sec	
Annealing	55˚C	30 sec	
Elongation	68˚C	4 – 8 min + 20 sec for each successive cycle	15
Final elongation	72 ˚C	7 min	1

[a] Elongation time was dependent on the desired amplicon size

2.3.2.3.1 Oligonucleotide design: Towards isolating a CYP450 gene from a *Thermus* sp. NMX2.A1

Oligonucleotide set Lyco_F and Conserv_R (Table 2.2), for the amplification of a possible CYP450 from the yellow pigmented *Thermus* sp. NMX2.A1, were designed based on multiple sequence alignments of the carotenoid gene clusters from *T. thermophilus* strains HB8 and HB27 as well as *T. aquaticus* Y51ML23 (Fig. 2.2). PCR amplification with these primers using whole cells of *Thermus* sp. NMX2.A1 was performed with Expand Long Template Polymerase at an annealing temperature of 59˚C and with an extension time of 90 sec.

2.3.2.3.2 Amplifying the Fdx and FNR from *T. scotoductus* SA-01

The 2.3 Mb genome of *T*. *scotoductus* SA-01 was recently sequenced and annotated (GenBank accession: NC_014974). The annotated open reading frames (ORFs) of Fdx (Locus tag: TSC_c03270) and FNR (Locus tag: TSC_c019740) were used to design oligonucleotides (Table 2.2) for the amplification of these genes. Genomic DNA from *T*. *scotoductus* SA-01 was extracted as described by Labuschagne & Albertyn, (2007) and PCR amplification was performed with Expand Long Template Polymerase with an annealing temperature of 61˚C and extension time of 90 sec.

2.3.2.4 Sequence analyses

Templates for sequencing were purified using the Biospin Plasmid DNA Extraction or Gel Extraction kits (Bioflux) according to the manufacturer's instructions. For double stranded templates 500 ng of DNA was used in each sequencing reaction. Sequencing was performed with an ABI Prism® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit 3.1 (Applied Biosystems, USA) according to the manufacturer's specifications. Universal sequencing primers T7 or SP6 were used at a final concentration of 3.2 pmol in 10 µL reactions that consisted of: 0.5 µL premix, 2 µL 5 x dilution buffer and PCR grade water.

The sequencing PCR cycle comprised of: an initial denaturation of 1 min at 96˚C, followed by 25 successive cycles of denaturation at 96˚C for 10 sec, annealing at 50˚C for 5 sec and elongation at 60˚C for 4 min. The reactions were then cooled to 4˚C. Post reaction clean-up consisted of an EDTA/ethanol precipitation (as recommended by ABI): the sequencing reaction volume was adjusted to 20 µL and 5 µL 125 mM EDTA (pH 8.0) and 60 µL absolute ethanol was added. Reactions were vortexed for 5 sec and left at room temperature for 15 min, after which the reactions were centrifuged at 20 000 x *g* for 10 min at 4˚C. The supernatant was completely removed and 60 µL 70 % (v/v) ethanol was added, followed by centrifugation at 20 000 x *g* at 4˚C for 5 min. The supernatant was completely aspirated and the samples dried under vacuum at 30˚C. Samples were stored in the dark at 4˚C until sequencing analyses could commence.

Nucleotide composition was determined on a 3130xl Genetic Analyzer (ABI). The resulting electropherograms were analyzed and edited using FinchTV 1.4.0 (Geospiza Inc.) and DN*Assist* 3.0. Contigs were assembled using ContigExpress (a component of the Vector NTI suite 9.0.0).

2.3.2.5 Assessment of PCR and restriction digest products

All amplicons and digestion products were electrophoresed and assessed on 0.8 % (w/v) agarose gels (unless stated otherwise) containing 0.6 µg/mL ethidium bromide. Agarose gels were prepared and electrophoresed in 1 x TAE buffer (0.1 M Tris, 0.1 mM glacial acetic acid, pH 8.0 and 0.05 M EDTA) at 6 V/cm for 1 h. Visualizations were done with a GelDoc XR (Bio-Rad Laboratories) and Quantity One 4.6.3 software under short wavelength UV light. For cloning purposes, DNA was visualized with a DarkReader™ transilluminator (Fermentas), excised from agarose gels and purified.

2.3.2.6 Transformation of *E. coli* strains

Subcloning of genes in pGEM®-T Easy (Promega) and propagation of pET22-based constructs were performed in *E. coli* Top 10 cells obtained from Invitrogen. Constructs in pCDFDuet-1 were transformed into *E. coli* XL-10 Gold. For expression studies, confirmed pET22- and pCDF-based constructs were transformed into *E. coli* BL21(DE3)pRARE2. Cells were rendered chemically competent by means of a modified version of the $RbCl_2$ method as originally described by Hanahan, 1983 and used as 50 µL aliquots. Ligations were done according to the manufacturer's directions and transformations were performed as described by Sambrook *et al*., 1989.

After transformation with pGEM®-T Easy cells were plated onto LB plates containing bacteriological agar (15 g/L), supplemented with ampicillin (100 µg/mL), IPTG [isopropylthio-β-D-galactoside (10 mg/mL)] and X-gal [5-bromo-4-chloro-3-indolyl- β-D-galactoside (40 mg/mL)]. Plates were incubated at 37˚C for 15 hours. White clones (presumed positive transformants) were selected and inoculated into 5 mL LB media supplemented with ampicillin (60 µg/mL) and grown for 15 hours at 37˚C with aeration (160 r.p.m).

After transformation with pET22- or pCDF-based constructs, transformants were plated on LB plates containing 15 g/L bacteriological agar and the appropriate antibiotic at the following final concentrations indicated in parentheses: ampicllin (50 µg/mL), streptomycin (50 µg/mL) and chloramphenicol (34 µg/mL). Plates were incubated at 37˚C for 15 hours. Resulting antibiotic resistant colonies (presumptive positive transformants) were selected and inoculated into 5 mL LB media supplemented with appropriate antibiotics at the concentrations as described above and grown for 15 hours at 37˚C with aeration (160 r.p.m). In the case of *E. coli* BL21(DE3)pRARE2 transformations the resulting cultures served as pre-culture for expression studies.

2.3.2.7 Expression in *E. coli* BL21(DE3)pRARE2

Expression of the CYP450 from *Thermus* sp. NMX2.A1 and ferredoxin (Fdx) and ferredoxin $NAD(P)^+$ reductase (FNR) from *T. scotoductus* SA-01, were performed as follows: fresh pre-culture (1 mL) was inoculated into 200 mL LB media with the appropriate antibiotics in 1 L shake flasks and cultured at 37˚C with aeration at 160 r.p.m until OD_{600} = 0.8 – 1.0. Cultures were then induced with 1mM IPTG. The heme precursor, 5-aminolevulinic acid hydrochloride, was added at a concentration of 1 mM to assist in CYP450 biosynthesis as well as 1 mM $FeCl_3 \cdot 6H_2O$. Cultures were then grown at 30˚C with agitation at 120 r.p.m. for 12 h (Momoi *et al*., 2006).

All induced *E. coli* cultures were harvested at 2 400 x *g* for 5 min at 4˚C. Pellets were washed twice with either 10 mL cold 40 mM 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer, pH 7.4 (Momoi *et al*., 2006) or 50 mM Potassium acetate buffer, pH 5.0 (Mandai *et al*., 2009). Wet cell pellets were weighed and resuspended in either HEPES or Potassium acetate buffer to a concentration of 70 g/L containing 1.5 mM 1,4-dithiothreitol (DTT) and Protease inhibitor cocktail set V (EDTA-free). Cell suspensions were disrupted with a One Shot Cell Disruptor (Constant Systems Limited, UK) at 32.5 k.p.s.i. This was followed by centrifugation for 20 min at 12 000 x *g* at 4˚C and the supernatants (cell-free extracts) were immediately used for experiments or kept at 4˚C for future usage.

2.3.2.7.1 CO-difference spectra

Cell-free extracts from *E*. *coli* clones expressing empty pET22b(+) and pET22b(+):NMX CYP450 were assayed in triplicate in microtitre plate strips. One set of six wells containing 190 µL of each cell-free extract or whole cells was exposed to CO gas in an enclosed container for 5 min. The second set was not exposed to CO. Both sets of strips were then reduced with 10 µL of 1.4 M sodium dithionite. Spectra were recorded every 10 min for 40 min at 25˚C from 400 nm – 500 nm (2 nm intervals) with a SpectraMax M2 Microtiter Plate Reader (Molecular Devices). The spectra from the samples without CO treatment were subtracted from the spectra of the CO treated samples to yield difference spectra with a characteristic Soret peak at 450 nm. The $A_{450-490}$ difference was corrected for a path length of 1 cm and the CYP450 concentration was calculated using an extinction coefficient of 91 mM^{-1} cm^{-1} (Omura & Sato, 1964).

2.3.2.7.2 Cytochrome *c* reduction assay

To assess whether the heterologously expressed Fdx and FNR proteins were catalytically active *i*.*e*. capable of shuttling electrons, a cytochrome *c* reduction assay was performed in which NADPH was the electron donor and cytochrome *c* (from equine heart) the final electron acceptor (Mandai *et al*., 2009). Cytochrome *c* reduction activity was calculated from the increase in absorbance at 550 nm (ε = 21.1 $mM^{-1}.cm^{-1}$). The reaction was performed at 45˚C using cell-free extracts containing Fdx and/or FNR. The experiments were performed in 96-well microtiter plates and readings were taken at 15 sec intervals for at least 2 min. Measurements were taken in triplicate. Table 2.4 illustrates the composition of the reactions:

Table 2.4 Components for cytochrome *c* reduction assay using cell-free extracts

Components	Blank	Control (no NADPH)	Experiment
- 1 M Tris-HCl, pH 7.4	10	10	10
- 1 M KCN	0.66	0.66	0.66
- Cell-free extract containing Fdx and/or FNR	2	2	2
- 12.5 mg/mL cytochrome *c*	-	10	10
- 12 mM NADPH	2	-	2
- H_2O	185.34	177.34	175.34
Total volume (in µL)	**200**	**200**	**200**

2.3.2.7.3 Glucose dehydrogenase 1 (GDH1) assay using cell-free extracts

Glucose dehydrogenase 1 (GDH1) from *S*. *solfataricus* P2 regenerates NADPH when given $NADP^+$, glucose (substrate) and $MgCl_2 \cdot 6H_2O$ (Lamble *et al*., 2003). To verity expression and catalytic activity of the heterologously expressed GDH1, glucose dehydrogenase activity was determined spectrophotometrically in 1 mL cuvettes at 37˚C and 65˚C by following the increase in absorbance at 340 nm, corresponding to the reduction of $NADP^+$ over 1 min. D-Glucose and $MgCl_2 \cdot 6H_2O$ solutions were filter sterilized with a 0.20 µm cellulose acetate syringe filter (Gema Medical S.L., Barcelona) and the $NAD(P)^+$ was prepared fresh and kept at 4˚C. Assays were performed in plastic cuvettes with a 1 cm path length using a Cary Bio 3000 spectrophotometer. Standard deviations were calculated from four measurements. Table 2.5 provides the components and the pipetting scheme of the GDH1 assay.

Table 2.5 Components for glucose dehydrogenase assay using GDH1-containing cell-free extracts

Components	Blank	Experiment
- 40mM HEPES buffer, pH 7.4[a]	793.8	773.8
- 100 mM D-Glucose	-	20
- 100 mM $MgCl_2 \cdot 6H_2O$	200	200
- 100 mM $NAD(P)^+$	3	3
- Cell-free extract with GDH1	3.2	3.2
Total volume (in µL)	**1000**	**1000**

[a] Momoi *et al*. (2006) used Potassium phosphate buffer but in this study HEPES was used instead since the $MgCl_2 \cdot 6H_2O$ reacted with the phosphate buffer and caused the magnesium to precipitate.

2.3.2.8 β-carotene hydroxylation experiment

Experiments were done either with whole cells as well with cell-free extract. For the whole cell experiment the CYP450 and redox partners were co-expressed in a single recombinant expression *E*. *coli* strain containing the pET22-construct with the CYP450 gene from *Thermus* sp. NMX2.A1 as well as a single pCDFDuet-1 vector carrying both the Fdx and FNR genes from *T*. *scotoductus* SA-01 (Table 2.1). Expression, harvesting and resuspension were done as described in section 2.3.2.7 but in 50 mM Potassium phosphate buffer (pH 7.4). β-carotene was dissolved in DMSO at a final concentration of 80 µM. The hydroxylation experiment design was based on the biphasic design described by Girhard *et al*. (2009) and comprised the following: 1.56 mL *E*. *coli* whole cell suspension, 400 µL Hexadecane and 40 µL 80 µM β-carotene suspension. Reactions were started by adding 20 µL of 100mM NADPH. Reactions were incubated at 37˚C as well as 65˚C for 12 h after which extractions were performed for TLC and LC/MS analyses.

For the cell-free extract experiment the CYP450, Fdx and FNR were over-expressed individually. Expression, harvesting and resuspension were performed as described in section 2.3.2.7. A 345 µM β-carotene suspension was prepared by mixing β-carotene powder in toluene containing 2.5% (w/v) Tween 20. The suspension was briefly mixed by

vortexing and then sonicated at 85% power for 1 min to create micelles. The toluene was then completely evaporated under a constant stream of N_2 gas and the resulting dry paste was resuspended in an equal volume of 40 mM HEPES buffer, pH 7.4. The β-carotene suspension was used immediately. After the functionality of the components essential for driving the catalysis of β-carotene hydroxylation was confirmed with the above-mentioned assays, the cell-free extracts were added together in equal parts as well as the components needed for NADPH regeneration and the β-carotene suspension. Experiments were performed at both 37˚C and 65˚C at pH 7.4 in HEPES buffer (Momoi *et al*., 2006) as well as at pH 5.0 using potassium acetate buffer (Mandai *et al*., 2009) but without D-Glucose, $MgCl_2 \cdot 6H_2O$ and GDH1. Table 2.6 provides the components and the pipetting scheme of the β-carotene hydroxylation experiments.

Table 2.6 Components of β-carotene hydroxylation experiments at 37˚C and 65˚C

Components	Reaction 1	Reaction 2	Reaction 3	Reaction 4	Reaction 5
- 40mM HEPES buffer, pH 7.4	225	325	325	321.8	-
- 50 mM Potassium acetate buffer, pH 5.0	-	-	-	-	545
- 100 mM D-Glucose	20	20	20	20	-
- 100 mM $MgCl_2 \cdot 6H_2O$	200	200	200	200	-
- 100 mM $NAD(P)^+$	10	10	10[a]	10	10[a]
- 1.36 µM CYP450[c]	100	-	100	100	100
- Fdx[d]	100	100	100	100	100
- FNR[d]	100	100	100	100	100
- GDH1	100	100	-	3.2[b]	-
- 345 µM β-carotene	145	145	145	145	145
Total volume (in µL)	**1000**	**1000**	**1000**	**1000**	**1000**

[a] In these experiments 100 mM NADPH was used and not $NAD(P)^+$

[b] Five units of recombinant GDH1 from *Thermoplasma acidophilum* (Sigma, Product number:G5909)

[c] CYP450 from *Thermus* sp. NMX2.A1 (cell-free extracts)

[d] Ferredoxin and Ferredoxin reductase from *T. scotoductus* SA-01 (cell-free extracts)

Reactions (1 mL) were placed in closed 50 mL Falcon tubes at 37˚C as well as 65˚C for 2 h without shaking. The experiment was repeated and samples incubated at 37˚C for 36 h with gentle agitation at 100 r.p.m. One milliliter samples from the β-carotene

hydroxylation experiments performed at 37˚C as well as 65˚C were extracted twice with an equal volume of ethyl acetate and dried under vacuum at 30˚C.

2.3.2.8.1 TLC analyses and LC/MS analyses

Samples were then either resuspended in 100 µL ethyl acetate for TLC analyses or resuspended in 1 mL acetonitrile for LC/MS analyses. For TLC analyses, 6 µL of each sample was spotted onto TLC plates (TLC Silica gel 60 F_{254} from Merck) together with a β-carotene standard (dissolved in ethyl acetate; final concentration 50 µM). For the separation of possible polar carotenoids *i.e*. β-cryptoxanthin and zeaxanthin, the mobile phase consisted of acetone and hexane (3:1) (Sachindra *et al*., 2005). Plates were inspected visually with UV-light as well as white light for the presence of additional spots. Plates were also left to dry and were then subjected to iodine vapour in an enclosed chamber to detect *inter alia* carotenoids with double bonds.

For LC/MS analyses the method by Momoi *et al*. (2006) was essentially used with the following modifications: LC-MS/MS analysis was done on an AB SCIEX API3200QTRAP triple quadruple ion trap hybrid mass spectrometer with analyte separation performed on an Agilent 1200 SL Series HPLC system (Agilent, Waldbronn, Germany) with binary pump, degasser, column oven and autosampler. Chromatographic separation was achieved on a Phenomenex Luna C18 column (3 µm x 3.0 mm ID x 150 mm) using a mobile phase comprising acetonitrile/methanol/isopropanol (85:10:5) at a flow rate of 0.5 mL/min. Electrospray parameters were as follows: polarity positive, capillary voltage 5500 V, drying gas 50 psi nitrogen, drying gas temperature 500˚C, nebulizer gas pressure 25 p.s.i. The mass spectrometer was operated in selected ion monitoring mode using the respective molecular ions M^+ and MH^+ and the fragment $[M-92]^+$ corresponding to loss of toluene from the molecular ion, *m*/*z* 476.2, 568.2, and 569.2 for zeaxanthin, and *m*/*z* 460.2 552.3, and 553.3 for cryptoxanthin.

2.4. Results

2.4.1 Screening *T. scotoductus* SA-01 for the presence of a CYP450 gene

Our initial strategy for isolating a CYP450 gene from *T. scotoductus* SA-01 was to use PCR together with four oligonucleotide primers (Table 2.2: Oligonucleotide primers P450_F_1, -R_1, -F_2 and -R_2) that were designed based on the conserved CYP450 regions of *CYP175A1* from both strains of *T. thermophilus*. Assuming that *T. scotoductus* SA-01 possessed a CYP450 gene, it was not certain whether *T. scotoductus* SA-01 would have a genomic copy of the gene or a copy located on a plasmid – as is the case for both strains of *T. thermophilus* (Henne *et al*., 2004). To entertain both possibilities we performed PCR using as template whole cells as well as purified, extracted genomic DNA from both *T. scotoductus* SA-01 and *T. thermophilus* HB8. Figure 2.1 illustrates the results when genomic DNA was used as PCR template.

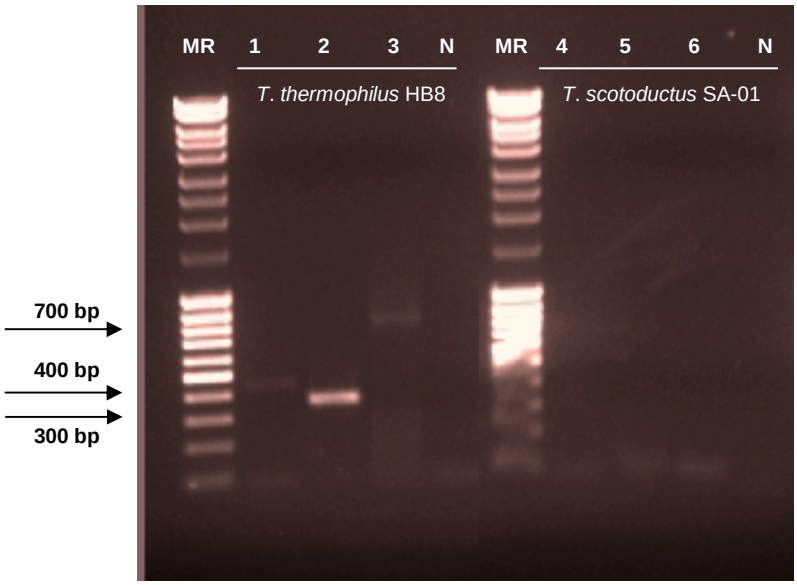


Fig. 2.1 PCR screening for a CYP450 gene in *T. scotoductus* SA-01 using genomic DNA from *T. thermophilus* HB8 as positive control (Lanes **1 – 3**) and *T. scotoductus* SA-01 (Lanes **4 – 6**) as template with various oligonucleotide combinations (see Table 2.2). Lanes: **MR** = 5 µL MassRuler (Fermentas), **1** and **4** = P450_F_1+R_2 (443 bp); **2** and **5** = P450_F_2+R_2 (356 bp); **3** and **6** = P450_F_1+R_2 (782 bp) and **N** = Negative control.

As expected, positive results were obtained when genomic DNA and whole cells (results not shown) from *T*. *thermophilus* HB8 was used as template. No amplicons could be amplified when whole cells or genomic DNA from *T*. *scotoductus* SA-01 was used. We then focused our efforts on *Thermus* sp. NMX2.A1, a yellow pigmented *Thermus* strain which is phylogenetically closely related to the cream coloured *T*. *scotoductus* SA-01.

2.4.2 Amplification of a CYP450 from *Thermus* sp. NMX2.A1 whole cells

Thermus belongs to the *Deinococcus*-*Thermus* phylum and these bacteria often display red or yellow pigmentation due to their ability to synthesize carotenoids (Tian & Hua, 2010). The yellow pigmentation of *Thermus* sp. NMX2.A1 hinted towards the possibility of a CYP450 that forms part of a β-carotene gene cluster as is the case with both strains of *T*. *thermophilus* (Blasco *et al*., 2004; Henne *et al*., 2004).

Multiple sequence alignment of the sections of the three gene clusters from *T*. *thermophilus* HB8 and HB27 as well as *T*. *aquaticus* Y51MC23 containing the CYP450 genes and their immediate adjacent genes, revealed high sequence identity on nucleotide level. The genes adjacent to the 5'-end of the CYP450s are consistently annotated as a photo lyase. However, the genes adjacent to the 3'-ends of the CYP450s bear different names (Fig. 2.2). In spite of the inconsistent gene annotation, all three of these genes adjacent to the 3'-ends are 698 bp in size and share 83% amino acid identity as revealed by a multiple alignment using the ClustalW2 tool (www.ebi.ac.uk/Tools/msa/clustalw2).

Figure 2.3 illustrates a portion of the multiple alignment results from the two genes directly adjacent up- and downstream of *CYP175A1*. The Lyco_F and Conserv_R oligonucleotide set were designed to anneal respectively 51 bp from the translation stop and 94 bp from the translation start on their carotenoid-related genes (indicated by dashed lines). Oligonucleotides P450_F_2 and P450_R_2 (Table 2.2), based on the K-helix and heme binding loop (containing the absolutely conserved **C**X**G** amino acid motif found in all CYP450s) respectively, were also used in the PCR experiment to confirm the presence of a CYP450 gene within the 1500 bp amplicon.

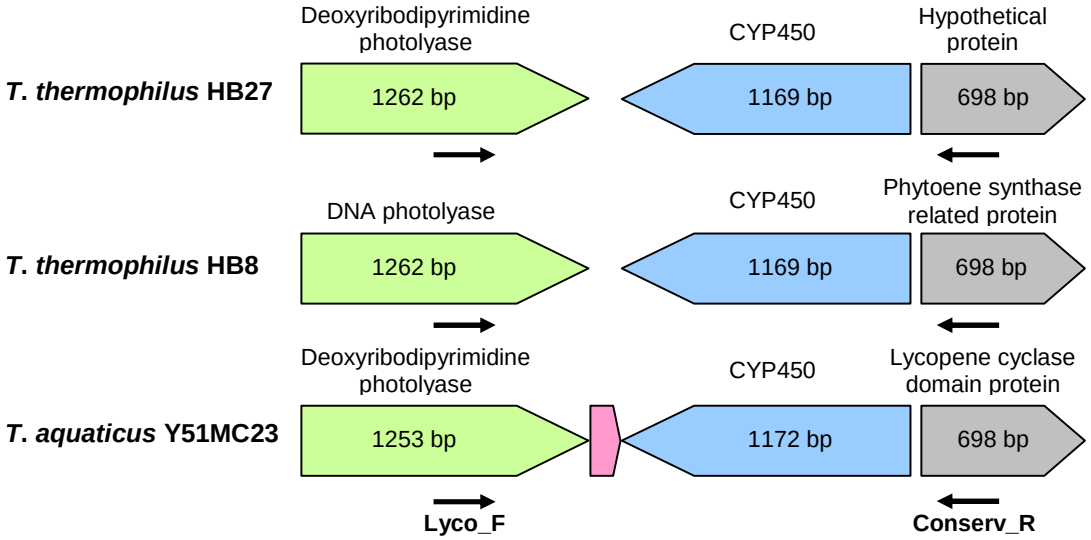


Fig. 2.2 Gene topology of a portion of the β-carotene gene cluster from *T. thermophilus* HB27, *T. thermophilus* HB8 and *T. aquaticus* Y51MC23. The Lyco_F and Conserv_R oligonucleotides are indicated on the figure as black arrows. Gene topology is identical for all three strains except for *T. aquaticus* Y51MC23 which has an additional 179 bp gene encoding a hypothetical protein (indicated in pink).

(A)

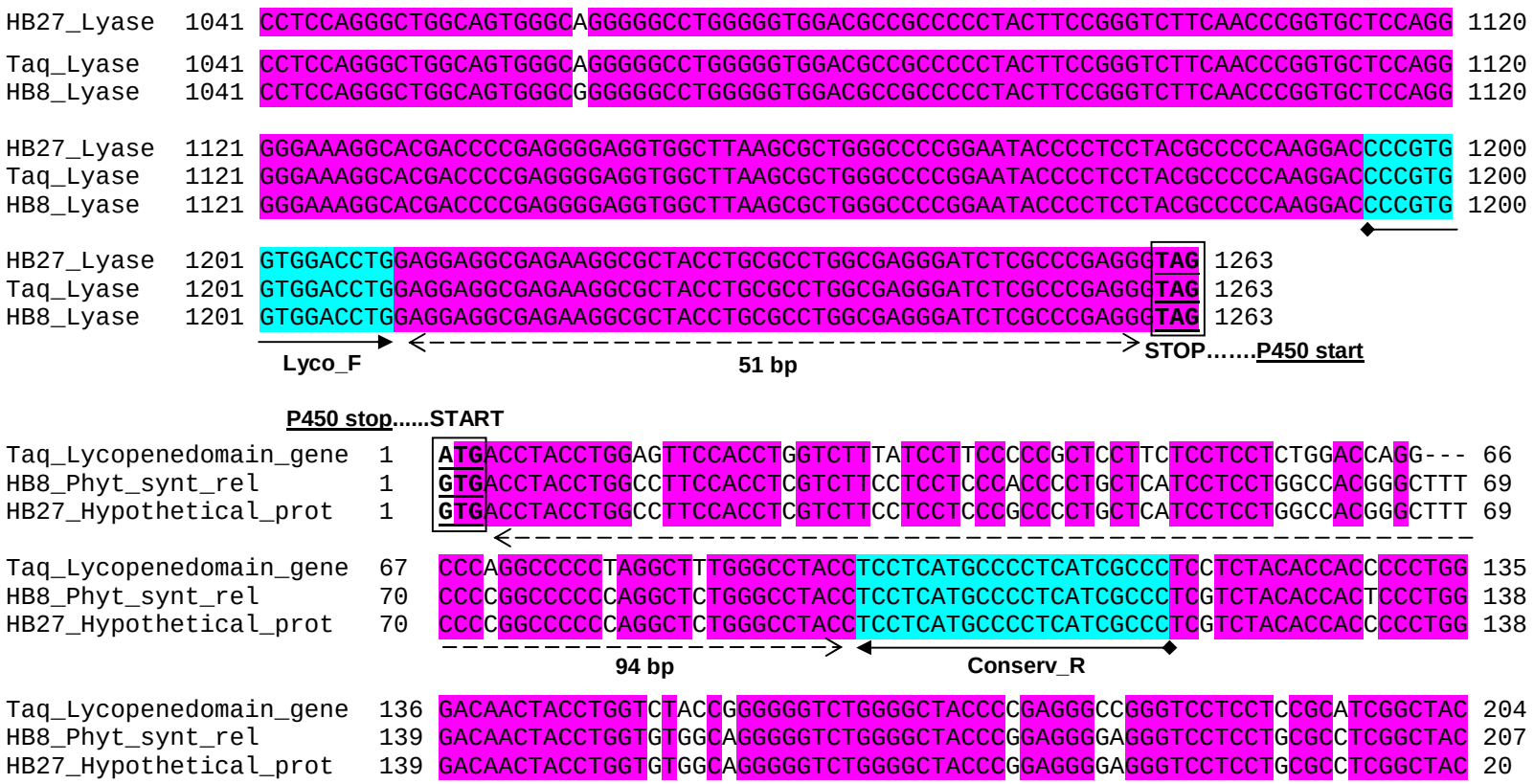


Fig. 2.3 Multiple alignment of genes directly adjacent up- and downstream of the CYP450 gene in strains HB8 and HB27 of *T. thermophilus* as well as *T. aquaticus* Y51ML23. Only a portion of the ORF from each gene is depicted. Oligonucleotides Lyco_F and Conserv_R are indicated in turquoise on the multiple alignment as well as the translation stop and start codons of each gene (boxed, underlined and in boldface). Multiple alignments were performed with DN*Assist* 3.0.

Figure 2.4 (below) illustrates the various amplicons obtained from a PCR using whole cells of *Thermus* sp. NMX2.A1. Oligonucleotide set P450_F_2 + P450_R_2 produced an amplicon of *ca*. 350 bp. The Lyco_F and Conserv_R oligonucleotide pair produced an amplicon that was *ca*. 1 500 bp in size. The 1 500 bp amplicon (Lane 1) was subjected to sequencing analysis.

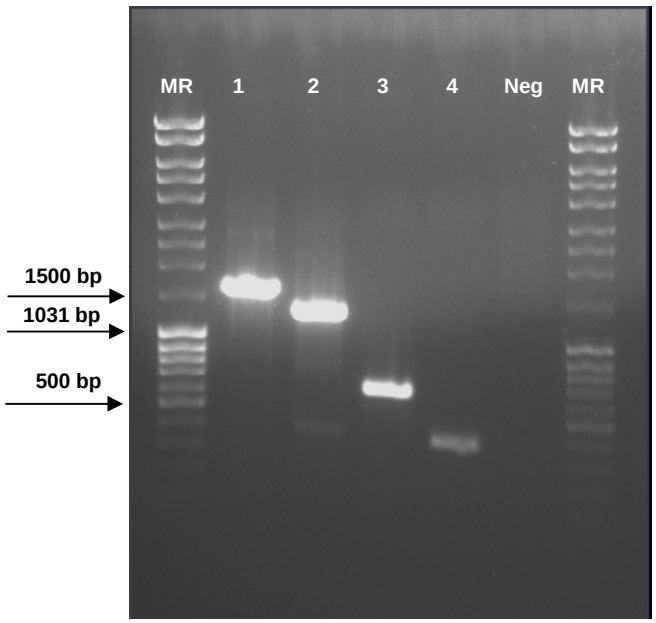


Fig. 2.4 Whole-cell PCR screening for the CYP450 gene in *Thermus* sp. NMX2.A1 with various oligonucleotides. Lanes: **MR** = 5 µL MassRuler (Fermentas) and **1 =** Lyco_F + Conserv_R; **2** = Lyco_F + P450_R_2; **3** = P450-F_2 + Conserv_R and **4** = P450_F_2 + P450_R_2. Expected sizes: **1** = 1545 bp; **2 =** 1275 bp; **3** = 627 bp and **4** = 357 bp and **Neg** = Negative control. Expected amplicon sizes were based on the gene sequences of *T*. *thermophilus* HB27.

Sequence analyses of the amplicon in Lane 1 of Figure 2.4 was performed and revealed a complete ORF of 1 170 bp which translated into a protein with a theoretical molecular mass of 44.14 kDa and isoelectrical point (pI) of 9.79 (as predicted by the pI/Mw tool on the ExPASy Proteomics server).

```
  1 ATGAAGCGCC TTTCCCTGAG GGAGGCCTGG CCCTACCTGA AAGACCTCCA
    M  K  R  L   S  L  R   E  A  W   P  Y  L  K   D  L  Q

 51 GCAAGACCCC CTCGCCGTCC TGCTGGAGTG GGGCCGGGCC CACCCCCGGC
     Q  D  P   L  A  V  L   L  E  W   G  R  A   H  P  R  L

101 TCTTCCTTCC CCTGCCCCGC TTCCCCCTGG CCCTGATCTT TGACCCCGAG
      F  L  P   L  P  R   F  P  L  A   L  I  F   D  P  E

151 GGGGTGGAGG GGGCACTCCT CGCCGAGGGG ACCACCAAGG CCACCTTCCA
    G  V  E  G   A  L  L   A  E  G   T  T  K  A   T  F  Q

201 GTACCGGGCC CTCTCCCGCC TCACGGGGAG GGGCCTCCTC ACCGACTGGG
     Y  R  A   L  S  R  L   T  G  R   G  L  L   T  D  W  G

251 GGAAAAGCTG GAAGGAGGCG CGCAAGGCCC TCAAAGACCC CTTCCTGCCG
      K  S  W   K  E  A   R  K  A  L   K  D  P   F  L  P
301 AAGAGCGTCC GCGGCTACCG GGAGGCCATG GAGGAGGAGG CCCGGGCCTT
    K  S  V  R   G  Y  R   E  A  M   E  E  E  A   R  A  F

351 CTTCGGGGAG TGGCGGGGGG AGGAGCGGGA CCTGGACCAC GAGATGCTCG
     F  G  E   W  R  G  E   E  R  D   L  D  H   E  M  L  A

401 CCCTCTCCCT GCGCCTCCTC GGGCGGGCCC TCTTCGGGGA GCCCCTCTCC
      L  S  L   R  L  L   G  R  A  L   F  G  E   P  L  S

451 CCAAGCCTCG CGGAGCACGC CCTTAAGGCC CTGGACCGGA TCATGGCCCA
    P  S  L  A   E  H  A   L  K  A   L  D  R  I   M  A  Q

501 GACCAGGAGC CCCCTGGCCC TCCTGGACCT CGCCGCCGAA GCCCGCTTCC
     T  R  S   P  L  A  L   L  D  L   A  A  E   A  R  F  R

551 GGAAGGACCG GGGGGCCCTC TACCGCGAGG CGGAAGCCCT CATCGTCCAC
      K  D  R   G  A  L   Y  R  E  A   E  A  L   I  V  H

601 CCGCCCCTCT CCCACCTTCC CCGAGAGCGC GCCCTGAGCG AGGCCGTGAC
    P  P  L  S   H  L  P   R  E  R   A  L  S  E   A  V  T

651 CCTCCTGGTG GCGGGCCACG AGACGGTGGC GAGCGCCCTC ACCTGGTCCT
     L  L  V   A  G  H  E   T  V  A   S  A  L   T  W  S  F
                I-Helix motif
701 TTCTCCTCCT CTCCCACCGC CCGGACTGGC AGAAGCGGGT GGCCGAGAGC
      L  L  L   S  H  R   P  D  W  Q   K  R  V   A  E  S


751 GAGGAGGCGG CCCTCGCCGC CTTCCAGGAG GCCCTGAGGC TCTACCCCCC
    E  E  A  A   L  A  A   F  Q  E   A  L  R  L   Y  P  P

801 CGCCTGGATC CTCACCCGGA GGCTGGGAAG GCCCCTCCTC CTGGGAGAGG
     A  W  I   L  T  R  R   L  G  R   P  L  L   L  G  E  D

851 ACCGGCTCCC CCCGGGCACC ACCCTGGTCC TCTCCCCCTA CGTGACCCAG
      R  L  P   P  G  T   T  L  V  L   S  P  Y   V  T  Q


901 AGGCTCCACT TCCCCGATGG GGAGGCC TTC CGGCCCGAGC GCTTCCTGGA
    R  L  H  F   P  D  G   E  A  F   R  P  E  R   F  L  E
                           K-Helix motif
```

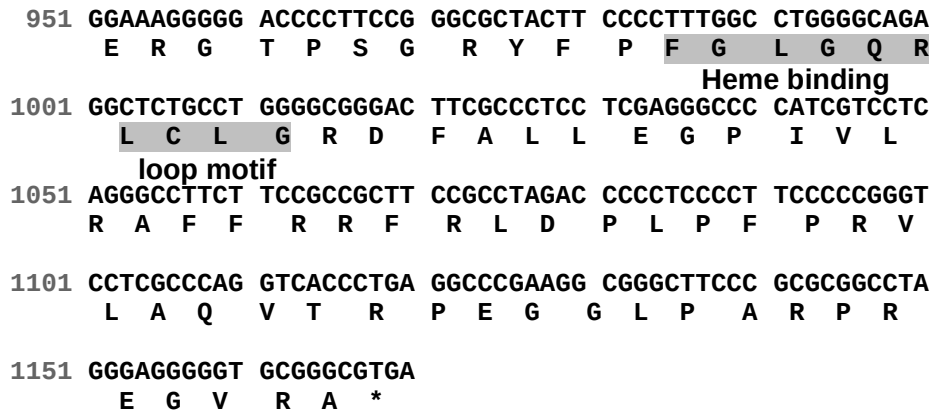


Fig. 2.5 ORF of the CYP450 isolated from *Thermus* sp. NMXA2.A1. Conserved I-Helix, K-Helix and Heme binding loop motifs are indicated in the grey shaded boxes. The overlaid sequence was generated using pDRAW32 version 1.1.109.

Figure 2.5 illustrates the ORF and the corresponding amino acid sequence. All conserved CYP450 amino acid motifs are present (indicated in grey shaded boxes). The gene sequence was deposited onto the NCBI database with the following Genbank accession number: GU220073.

The design of the oligonucleotide set (Lyco_F and Conserv_R, Table 2.2) used to amplify an intact CYP450 from *Thermus* sp. NMX2.A1 was based on β-carotene biosynthesis genes from *T*. *thermophilus* strains HB27 and HB8 as well as *T*. *aquaticus* Y51ML23. It has been confirmed that the CYP450 and β-carotene biosynthesis genes from *T*. *thermophilus* HB27 and HB8 are located on a megaplasmid (Henn *et al*., 2004). However, although a CYP450 and β-carotene biosynthesis genes have been identified from the unassembled genome of *T*. *aquaticus* Y51ML23, the presence of a megaplasmid will have to be confirmed upon completion of the sequencing project in question. The presence of a possible megaplasmid in *Thermus* sp. NMX2.A1 will also have to be confirmed either by sequencing analyses or Pulse Field Gel Electrophoresis (PFGE) (Moreira *et al*., 1995; Moreira & Sá-Correia, 1997). Genomes of other members of the *Deinococcus-Thermus* phylum that have been sequenced include *Deinococcus radiodurans* R1, *Deinococcus geothermalis* DSM 11300, *Meiothermus ruber* DSM 1279 and *Meiothermus silvanus* DSM 9946. Although all of them possess several genes

involved in carotenoid biosynthesis only *M*. *silvanus* DSM 9946 appears to have a β-carotene hydroxylase (Tian & Hua, 2010).

Although the Lyco_F and Conserv_R oligonucleotide set may prove to be useful molecular tools in amplifying compete ORFs of CYP450s from other, unsequenced, yellow pigmented *Thermus* species *e.g*. *Thermus oshimai* and *Thermus igniterrae* or even other yellow pigmented members of the *Deinococcus-Thermus* phylum, caution should be taken not to simply assume that all yellow pigmented members of this phylum will (i) necessarily possess a CYP450 and (ii) possess a CYP450 that forms part of a carotenoid biosynthesis gene cluster.

2.4.3 The CYP450 from *Thermus* sp. NMX2.A1: a putative β-carotene hydroxylase?

Multiple sequence alignment of the CYP450 protein from *Thermus* sp. NMX2.A1 with the CYP450s of *T*. *thermophilus* HB27 and HB8 (Fig. 2.6 A) revealed that the CYP450 from *Thermus* sp. NMX2.A1 had 98% and 97% amino acid identities towards the CYP450s from *T*. *thermophilus* HB27 and HB8 respectively. There are only two unique amino acids found in the CYP450 from *Thermus* sp. NMX2.A1, namely E^{147} and G^{276}, and this CYP450 possesses conserved CYP450 amino acid motifs that are identical to those found in the CYP450s from both strains of *T*. *thermophilus*.

Blasco and co-workers (2004) demonstrated that the CYP450 from *T*. *thermophilus* HB27 (CYP175A1) was a β-carotene hydroxylase and Yano *et al*. (2003), who solved the crystal structure of CYP175A1, identified the α-helical regions important for substrate binding, namely the F-, G- and B'-helices. The B'-helix stretches from T^{65} - S^{72}, the F-helix stretches from S^{150} - S^{170} and the G-helix from D^{176} – I^{198}. In the CYP450s of *Thermus* sp. NMX2.A1 and both strains of *T*. *thermophilus* these helical regions are identical.

Judging from the high amino acid identity between the *Thermus* sp. NMX2.A1 and *T*. *thermophilus* HB27 CYP450s (98%) coupled with the fact that their substrate binding regions are identical (Fig. 2.6 A), it seems plausible to suggest that the CYP450 of *Thermus* sp. NMX2.A1 is a putative β-carotene hydroxylase. Only six amino acid

residues differ between the CYP450s from *T*. *thermophilus* HB27 and *Thermus* sp. NMX2.A1. These amino acids are located on the periphery of the protein (Fig. 2.6 B) – making it unlikely that these amino acids will negatively impact the functionality of the CYP450 from *Thermus* sp. NMX2.A1 CYP450s.

(A)

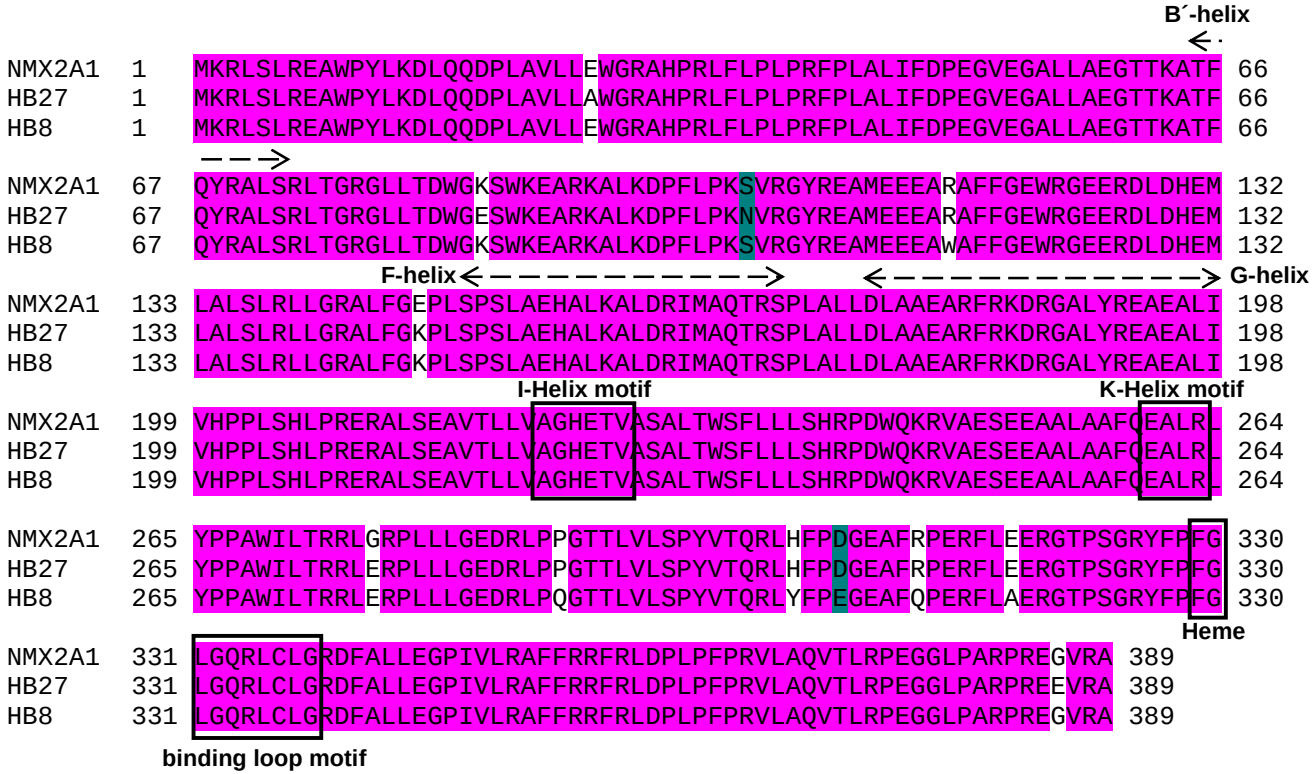


Fig. 2.6 (A) Multiple sequence alignment of CYP450 proteins from *Thermus* sp. NMX2.A1 (**NMX2A1**); *T. thermophilus* HB27 (**HB27**) and *T. thermophilus* HB8 (**HB8**). Conserved CYP450 amino acid motifs are enclosed by rectangles and α-helices important in substrate binding (F, G and B'-helices) are indicated by dashed lines with arrows. Multiple sequence alignments were performed with DNAssist 3.0. Pink colour indicates identical amino acids and turquoise colour indicates similar amino acids.

(B)

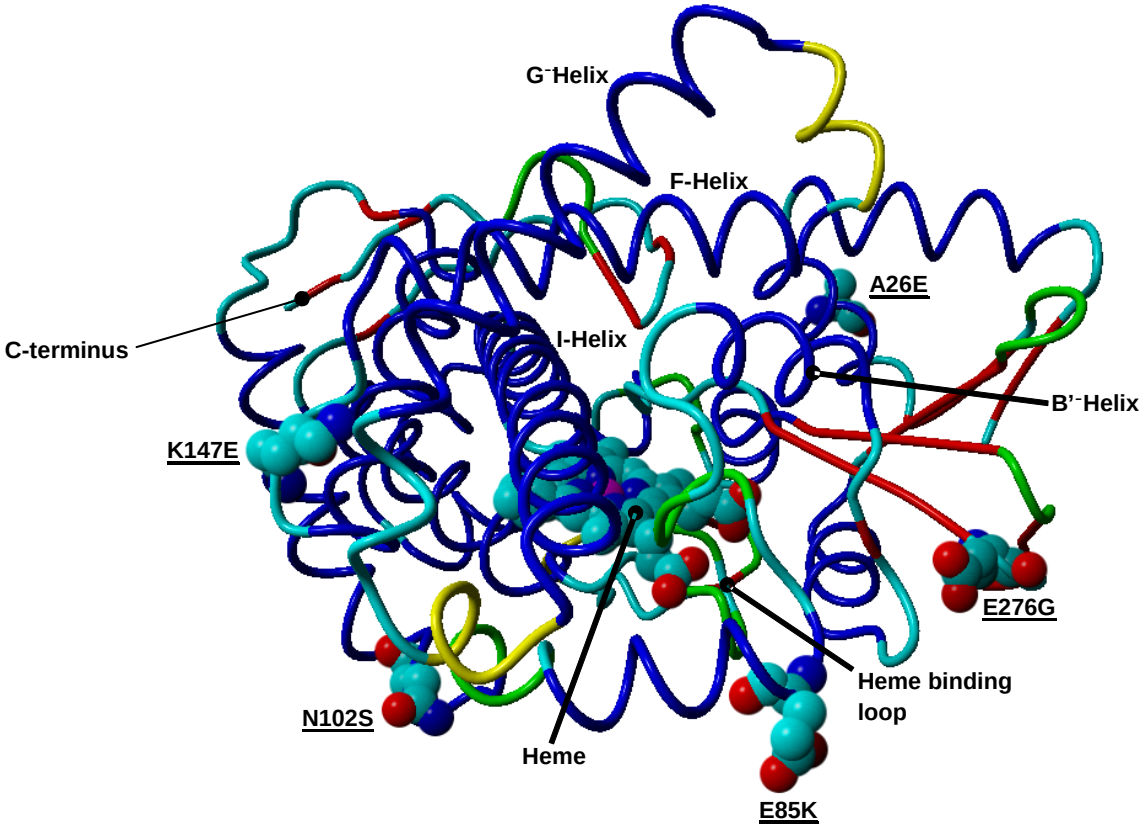


Fig. 2.6 (B) Ribbon structure of CYP175A1 of *T. thermophilus* HB27 edited in YASARA Viewer. Amino acid differences between *T. thermophilus* HB27 and *Thermus* sp. NMX2.A1 are indicated on the structure in single letter amino acid notation (where the first capitol letter represents the amino acid residue of *T. thermophilus* HB27, the number: the position of the amino acid and the second capitol letter the amino acid residue in *Thermus* sp. NMX2.A1 found in the same position as that of *T. thermophilus* HB27. Note: the last four amino acids are missing in this particular structure's C-terminus which includes E386G. The Heme prosthetic group, B'-Helix, I-Helix and G-Helix are also indicated on the structure.

2.4.4 Heterologous expression of the components of the CYP450 electron transfer system

2.4.4.1 Heterologous expression of the CYP450 from *Thermus* sp. NMX2.A1

The newly isolated CYP450 from *Thermus* sp. NMX2.A1 was cloned into pET22b(+) by directional cloning using *Nde*I and *Hin*dIII which resulted in a construct without a 6x Histidine-tag. Momoi *et al*. (2006) and Mandai *et al*. (2009) both used over-expressed CYP175A1 but only the latter authors co-expressed a C-terminal 6x Histidine tag since they intended to use purified CYP175A1 preparations. Both authors reported CYP175A1 activity and thus it would seem as if the presence or absence of a C-terminal 6x Histidine tag has no negative effect on CYP175A1 expression or catalyses. The resulting pET22b(+)-NMX2.A1 CYP450 vector was transformed into an *E*. *coli* BL21(DE3) strain containing an additional pRARE2-LysS plasmid which encodes for 7 rare tRNA codons. The CYP450 gene of *Thermus* sp. NMX2.A1, both the Fdx and FNR genes from *T*. *scotoductus* SA-01 and the GDH1 gene from *S*. *solfataricus* contain rare codons. The corresponding tRNA codons are not found in *E*. *coli* BL21DE3 and consequently this expression strain was transformed with the pRARE2 plasmid to ensure successful heterologous expression of the above-mentioned genes. Induction was performed with IPTG for 12 h. Biomass was harvested and the wet weight concentrations were adjusted to 70 g/L. Cells were disrupted with a French press and the resulting cell-free extracts used to perform CO-difference spectra to calculate the CYP450 content.

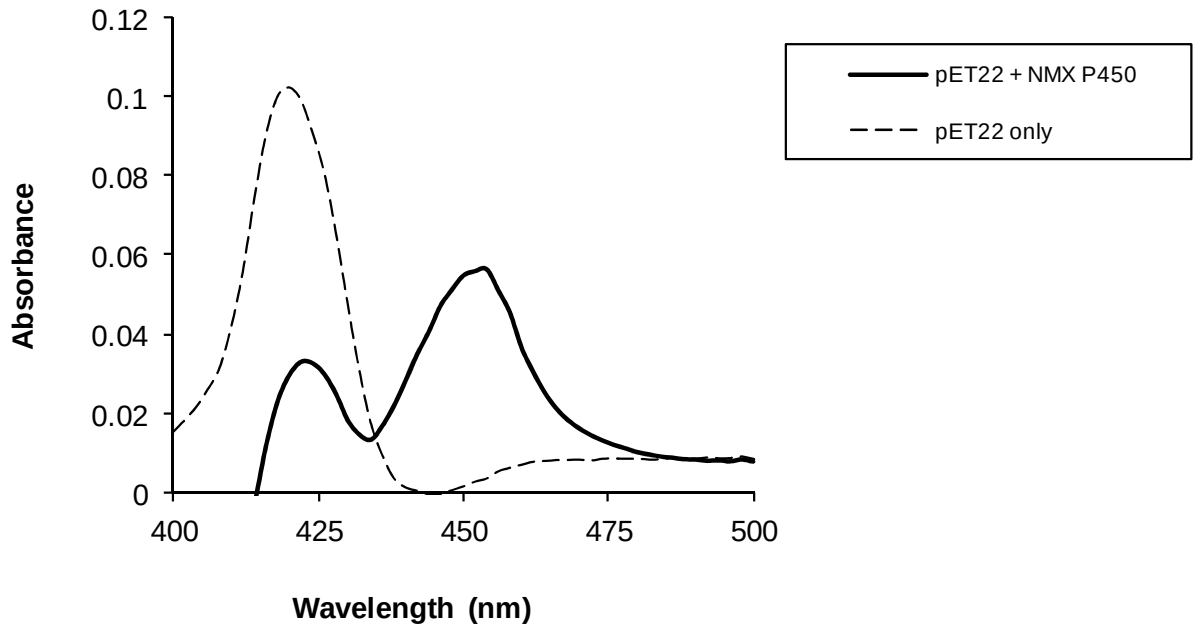


Fig. 2.7 CO-difference spectra of the newly isolated CYP450 from *Thermus* sp. NMX2.A1 using cell-free extracts. Final spectra are a result of subtracting the spectrum of the oxidized CYP450 from that of the sodium dithionite-reduced and CO-bound CYP450. Cell-free extracts containing expressed empty pET22b(+) served as a negative control. Note: this is the average spectra of three separate measurements.

CYP450 induction could be detected 12 h after IPTG addition and based on the CO-difference spectra (Fig. 2.7) obtained from *E*. *coli* cell extracts it was calculated that the content of the properly folded CYP450 was 1.36 µM (± 0.03). Momoi and co-workers (2006) obtained a CYP450 concentration of 3.8 µM in *E*. *coli* cell extract after 44 h of IPTG induction using the pKK233-3 expression vector in *E*. *coli* BL21DE3 Codon Plus.

2.4.4.2 Cloning of Fdx and FNR into pCDFDuet-1

Initially, the FNR in *T*. *thermophilus* HB27 was incorrectly annotated as a thioredoxin reductase but upon closer inspection of the protein sequence (Fig. 2.8 A) it became apparent that the active-site dithiol/disulfide group (protein motif: CXXC), which is required to exchange reducing equivalents with thioredoxin, is not present in this protein (Mandai *et al*., 2009). The FNR identified in the genome sequence of *T*. *scotoductus* SA-

01 (Locus tag: TSC_c019740) aligns well with the FNR from *T*. *thermophilus* HB27 (88% amino acid identity) and possesses the conserved motifs responsible for the binding of FAD (motif: GXGPAG and GXXAXGD) and the binding of NADPH (motif: GXGXXA) (Mandai *et al*., 2009). The redox-active sites are also very similar between the proteins from the two species (Fig. 2.8 A).

The Fdx (79 amino acids) identified in the genome sequence of *T*. *scotoductus* SA-01 (Locus tag: TSC_c03270) and the Fdx from *T*. *thermophilus* HB27 display 97% amino acid identity. The Fdx from *T*. *thermophilus* HB27 is a seven–iron (one [4Fe-4S] and one [3Fe-4S]) cluster ferredoxin with the [3Fe-4S] cluster being the functional component of the ferredoxin (Macedo-Ribeiro *et al*., 2001; Griffin *et al*., 2003). The cysteine residues important in coordinating the iron-sulfur clusters are located at identical positions in Fdx from both HB27 and SA-01. When comparing the Fdx of *T*. *thermophilus* HB27 to that of *T*. *scotoductus* SA-01 there are two amino acid changes: SA-01 has a F^{56} instead of Y^{56} and the C-terminal of the SA-01 Fdx ends with a G^{79} instead of E^{79} (Fig. 2.8 B)

Oligonucleotides for the amplification of *Fdx* and *FNR* were designed based on the sequences from the genome of *T*. *scotoductus* SA-01. The Fdx and FNR genes from *T*. *scotoductus* SA-01 were PCR amplified using the oligonucleotides indicated in Table 2.2 and extracted genomic DNA as template.

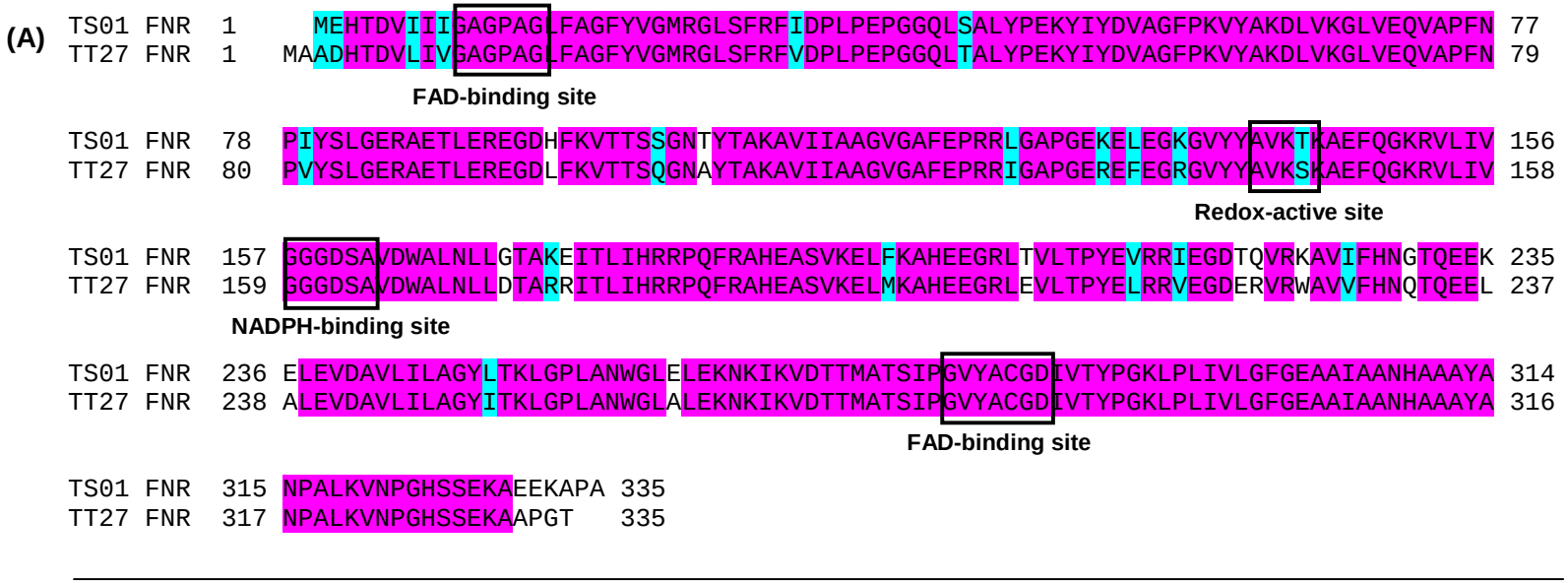


Fig. 2.8 Multiple alignment of FNR proteins **(A)** from *T. scotoductus* SA-01 **(TS01 FNR)** and *T. thermophilus* HB27 **(TT27 FNR).** FAD- and NADPH-binding sites and the redox active site are indicated. Multiple alignment of the Fdx proteins **(B)** of *T. scotoductus* SA-01 (**TS01 Fdx**) and *T. thermophilus* HB27 (**TT27 Fdx**). Cysteines involved in the coordination of the [3Fe-4S] and [4Fe-4S]-clusters are highlighted in yellow and grey respectively. Identical amino acids are highlighted in pink while similar amino acids are highlighted in turquoise. Alignments were performed with DNAssist 3.0.

2.4.4.3 Heterologous expression of Fdx and FNR from *T. scotoductus* SA-01

In order to assess whether the Fdx and FNR from *T. scotoductus* SA-01 were catalytically active *i.e.* whether these proteins could shuttle electrons to oxidized cytochrome *c*, a cytochrome *c* reductase assay was performed. Amplicons of *FNR* and *Fdx* from *T. scotoductus* SA-01 were cloned into the second and first cloning sites of the pCDFDuet-1 expression vector respectively (Fig. 2.9 A and B). *FNR* and *Fdx* were also cloned together into a single pCDFDuet-1 vector (construct not shown).

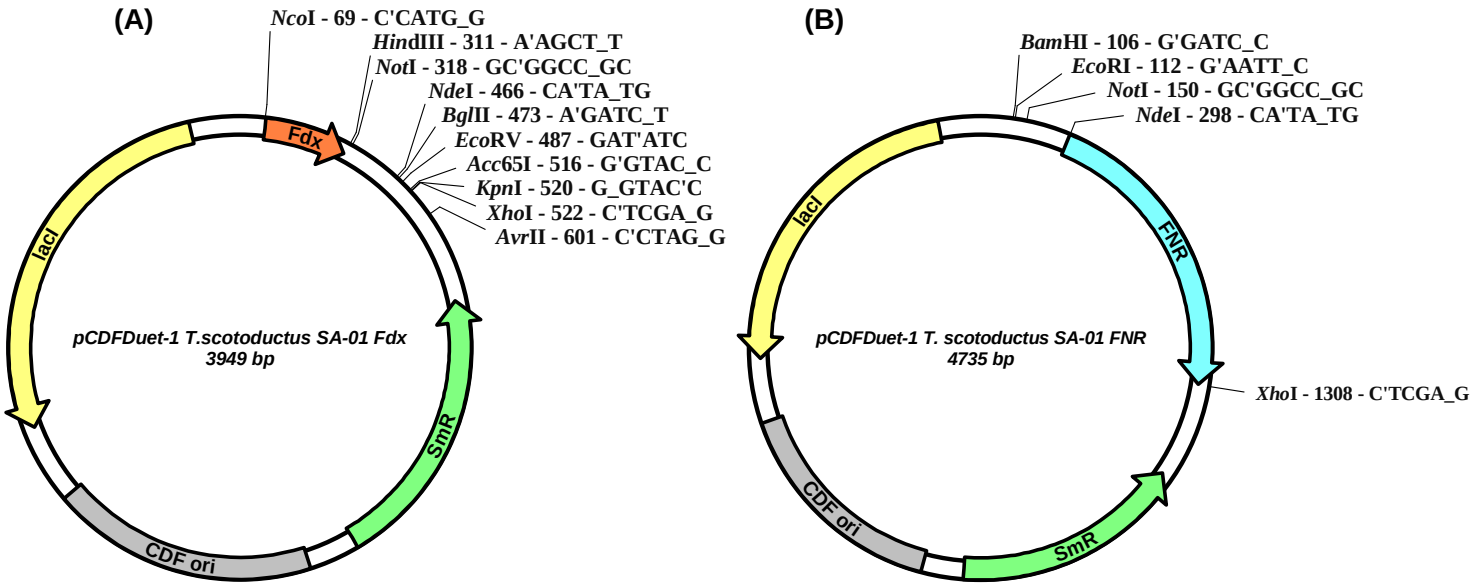


Fig. 2.9 pCDFDuet-1 constructs containing the *Fdx* **(A)** and *FNR* **(B)** genes of *T. scotoductus* SA-01. The *Fdx* gene was directionally cloned into the first multiple cloning site using *Hin*dIII and *Avr*II while the *FNR* gene was cloned into the second multiple cloning site using *Nde*I and *Xho*I.

Cell-free extracts from recombinant *E. coli* strains expressing cloned (i) Fdx, (ii) FNR, (iii) Fdx and FNR and (iv) no redox partners *i.e.* empty expression vector were used in the assay. Figure 2.10 illustrates the results of the reductase assay.

The activities show that the negative control (empty pCDFDuet-1 vector, containing no inserts) displayed almost no activity. As expected, the over-expressed Fdx on its own displayed almost the same order of magnitude of activity as the negative control. Thus, Fdx is unable to shuttle electrons to cytochrome *c* and requires another redox partner to reduce cytochrome *c*.

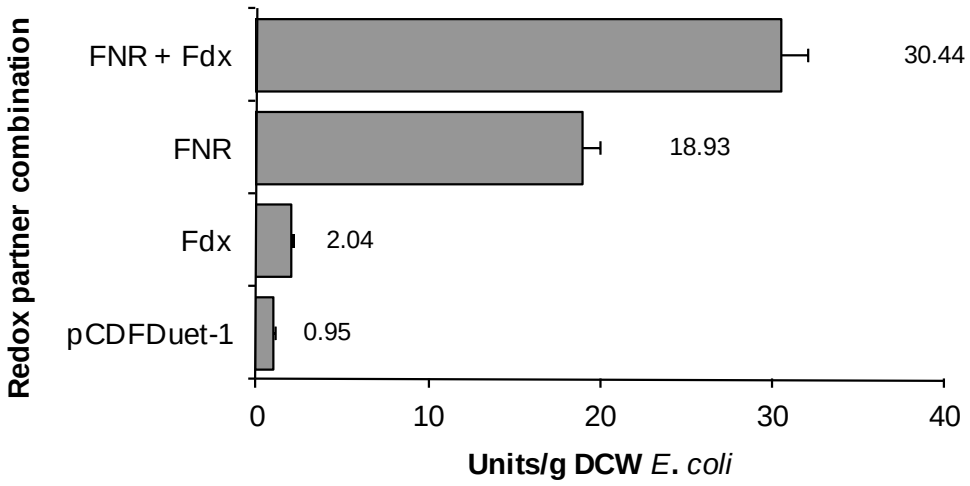


Fig. 2.10 Cytochrome *c* reductase assay using cell-free extracts of recombinant *E. coli* strains over-expressing ferredoxin (Fdx) and ferredoxin-NAD(P)$^+$ reductase (FNR) separately and together as well as the pCDFDuet-1 expression vector with no insert. Assays were performed at 45˚C and activities calculated using ε = 21.1 $mM^{-1}.cm^{-1}$. Activity is expressed as units (U) per gram (g) of dry cell weight (DCW). Unit definition: 1 Unit will catalyze the reduction of 1 µmole cytochrome *c* by NADPH per min in 50 mM Tris-HCl buffer, pH 7.4. Error bars represent triplicate measurements.

Over-expressed FNR expressed alone on the other hand, was able to shuttle electrons to cytochrome *c,* indicating that there is a native *E. coli* Fdx that can interact with the cloned FNR to shuttle electrons to cytochrome c. However, this reaction was enhanced in the presence of Fdx from *T. scotoductus* SA-01. This finding correlates well with the cytochrome *c* reduction results reported by Mandai and co-workers (2009) when purified protein preparations of Fdx and FNR from *T. thermophilus* HB27 were used. Results from our cytochrome *c* reductase assay indicate that the redox partners from *T.*

scotoductus SA-01 are capable of transferring electrons between each other and finally to cytochrome *c*.

2.4.4.3.1 Effect of redox partner co-expression on CYP450 production

Our initial goal was to create a whole-cell system containing the CYP450 from *Thermus* sp. NMX2.A1 together with the redox partners (FNR and Fdx) from *T. scotoductus* SA-01 so that β-carotene could be hydroxylated by whole-cell biotransformation. *E. coli* BL21(DE3)pRARE2-LysS was transformed with pET22- and pCDF vectors containing the CYP450 and Fdx and/or FNR genes to yield strains expressing either: CYP450 only, CYP450 + Fdx, CYP450 + FNR or CYP450 + Fdx + FNR (see Table 2.1 for constructs). CO-difference spectra were performed on whole cells to assess CYP450 production after 12 h and 22 h of IPTG induction (Fig. 2.11).

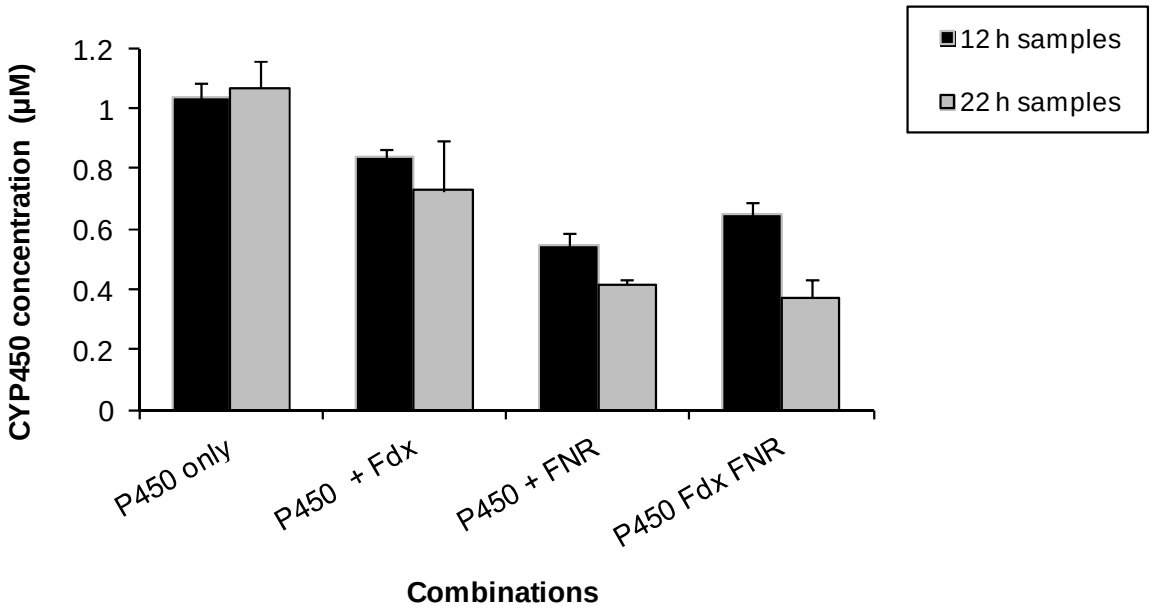


Fig. 2.11 Bar graph summation of the influence of redox partner combinations on CYP450 production when co-expressed after 12 h and 22 h IPTG induction. CO-difference spectra was performed with whole cells. Error bars represent triplicate measurements.

When the CYP450 was expressed on its own in pET22b(+) the CYP450 content was at its highest (1.04 µM and 1.07 µM after 12 h and 22 h of IPTG induction respectively, generating *ca*. 0.5 g/L of biomass [dry weight]). When the CYP450 was co-expressed with the redox partners (in the pCDFDuet-1 vector) the production of CYP450 decreased – especially when the FNR was co-expressed with the CYP450.

2.4.5 Heterologous expression of *GDH1* from *S. solfataricus* P2

Glucose dehydrogenases (GDHs) use glucose as substrate to perform a redox reaction in which glucose is oxidized to gluconate and the co-factor $NAD(P)^+$ is reduced to NADPH. NADPH is the necessary source of electrons that drives the CYP175-system (McLean *et al.*, 2005; Mandai *et al.*, 2009).

GDHs can therefore be used for cofactor regeneration to drive catalysis of the CYP175A1-system from *T*. *thermophilus* HB27 (Mandai *et al*., 2009). Since we had available the amplified gene of GDH1 from *S. solfataricus* we cloned it into pET22 b(+) and evaluated it for use with the CYP175-system. Cell-free extracts containing the over-expressed GDH1 were assayed as essentially described by Lamble *et al*. (2003). Assays were performed at 37˚C and 65˚C in 40 mM HEPES buffer, pH 7.4 and blank reactions contained no glucose. GDH1 activity in cell-free extracts was monitored spectrophotometrically by measuring the increase in absorbance at 340 nm, which corresponds to the reduction of $NAD(P)^+$, for 1 min. Figure 2.12 illustrates the results obtained.

We also tested the commercially available GDH (same protocol was used as with the GDH1 from *S*. *solfataricus*) from *Thermoplasma acidophilum* (available from Sigma, Catalogue number: G5909) at 37˚C as well as 65˚C for usage in our β-carotene hydroxylation experiment. The activity of this dehydrogenase was almost identical at both temperatures making it the better GDH to use in β-carotene hydroxylation experiments.

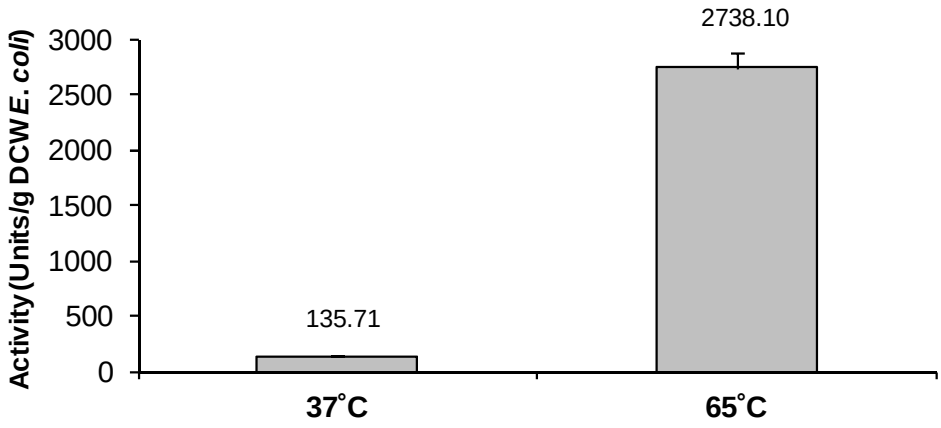


Fig. 2.12 GDH1 assay using *E. coli* cell-free extracts at 37˚C and 65˚C. Assays were performed in 40 mM HEPES, pH 7.4 using glucose as substrate. Unit definition: 1 Unit is the amount of GDH1 needed to reduce 1 µmole of $NAD(P)^+$ to NADPH per min. Error bars are representative of four measurements.

2.4.6 β-Carotene hydroxylation experiments

We initially attempted whole-cell biotransformations at 37˚C as well as 65˚C with an *E. coli* strain co-expressing both redox partners as well as the CYP450 but with no success. We attributed the negative results to low CYP450 levels and possibly very low FNR and/or Fdx expression levels. In addition, the difficulty the substrate (β-carotene) might have in crossing the cell membrane, due to its insolubility and hydrophobic nature, was also an issue. It was decided to rather over-express the CYP450 and the redox partners individually and to work with cell-free extracts - thus eliminating the issues of low CYP450 yields and limited substrate transfer into the cells.

Each component of the CYP450-system (*i.e.* CYP450, FNR and Fdx) was over-expressed individually, the cells broken and the resulting cell-free extract added together in equal volumes (Table 2.6). Hydroxylation experiments were performed at pH 7.4 (Momoi *et al*., 2006) and pH 5.0 (Mandai *et al*., 2009) at 37˚C (Momoi *et al*., 2006) as well as 65˚C (Mandai *et al*., 2009), thus covering the conditions used by Momoi *et al*. (2006) to perform β-carotene hydroxylation experiments with *E. coli* cell extracts

containing over-expressed CYP175A1 and by Mandai and co-workers (2009) to perform hydroxylation of β-carotene with purified CYP175A1. The final β-carotene concentration in our reaction was 50 µM (Momoi *et al*., 2006) *versus* the 20 µM of Mandai *et al*. (2009) and the final CYP450 concentration was 0.14 µM *versus* the 1.9 µM used by Momoi *et al*. (2006) and the 0.03 or 0.4 µM of pure CYP175A1 used by Mandai *et al*. (2009).

TLC analyses revealed no detectable spots correlating to the hydroxylated products of β-carotene namely: β-cryptoxanthin (R_f = 0.40) or zeaxanthin (R_f =0.30) or any other xantophylls for that matter (Hayashi *et al*., 2003; Sachindra *et al*., 2005). Samples that were subjected to LC/MS did not reveal characteristic peaks for β-cryptoxanthin or zeaxanthin. We were only able to identify, with confidence, the peaks corresponding to β-carotene.

2.5 Discussion

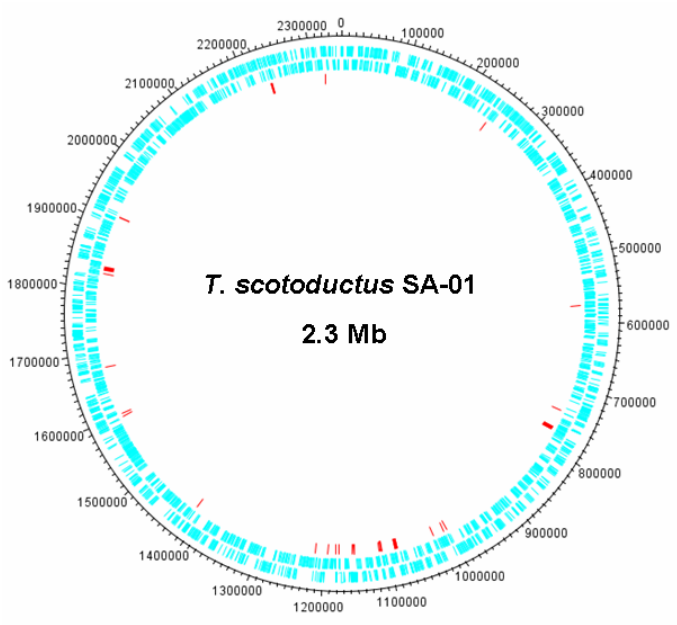
2.5.1 *T. scotoductus* SA-01 does not possess a CYP450

Initial attempts to find a CYP450 in *T. scotoductus* SA-01, which included a PCR strategy using oligonucleotides based on conserved *CYP175A1* regions, using both whole cells as well as extracted genomic DNA as template, did not yield any positive results (section 2.4.1, Fig. 2.1). This was an indication that *T. scotoductus* SA-01 did not possess a CYP450 gene.

Incidentally, during this phase of the study the genome of *T. scotoductus* SA-01 was also being sequenced and annotated. The data from the complete genome unequivocally confirmed the absence of a CYP450 gene in *T. scotoductus* SA-01 (Gounder *et al*., 2011). Not only did the assembled genome reveal the absence of a CYP450 on both the 2.3 Mb chromosome as well as the 8 383 bp pTSC8 plasmid (GenBank accession: NC_014975) of *T. scotoductus* SA-01, but it also revealed that genes similar to those of the *T. thermophilus* HB27 megaplasmid were scattered over portions of the *T. scotoductus* SA-01 chromosome (Fig. 2.13 A).

Although some genes similar to megaplasmid genes from *T. thermophilus* HB27 could be identified in the genome of *T. scotoductus* SA-01, the gene cluster containing the genes necessary for β-carotene biosynthesis could not be identified. The only remnant of this gene cluster in the genome of *T. scotoductus* SA-01 was the NADH-ubiquinone oxidoreductase gene (Fig. 2.13 B). It was then decided to rather search for a CYP450 gene in *Thermus* sp. NMX2.A1 – a yellow pigmented strain that is phylogenetically closely related to the cream coloured *T. scotoductus* SA-01.

(A)



(B)

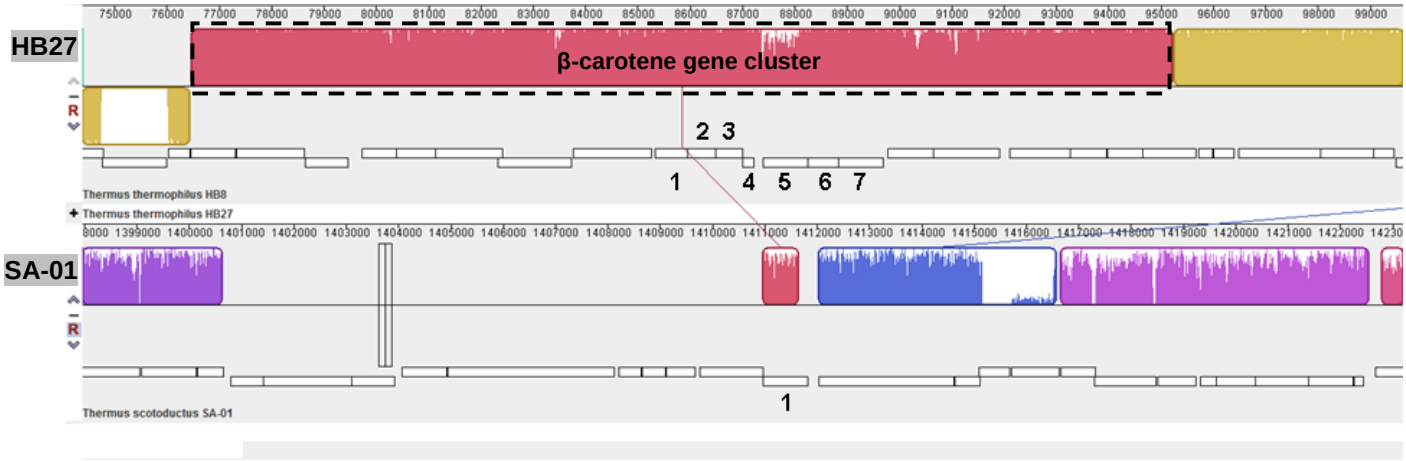


Fig. 2.13 **(A)** Circular diagram of the 2.3 Mb chromosome of *T. scotoductus* SA-01. Chromosomal genes are depicted by turquoise ticks and megaplasmid genes by red ticks. **(B)** Screen shot of a MAUVE alignment showing the β-carotene gene cluster on the megaplasmid in *T. thermophilus* HB27 (red portion) mapped onto the chromosome of *T. scotoductus* SA-01. Gene annotations: **1** = NADH-ubiquinone oxidoreductase; **2** = Regulatory protein; **3** = Probable transcriptional regulator; **4** = Phytoene synthase; **5** = Deoxyribodipyrimidine photolyase; **6** = Cytochrome P450 monooxygenase and **7** = Hypothetical conserved membrane protein (Figures courtesy of Prof. D. Litthauer, University of the Free State).

This close phylogenetic relationship between *Thermus* sp. NMX2.A1 and *T*. *scotoductus* SA-01 was confirmed by DNA/DNA hybridizations and genomic fingerprinting of the BOX A repeat and also correlated with 16S-rRNA gene sequence analyses (Balkwill *et al*., 2004). In addition their morphological-, biochemical- and physiological traits including the ability to reduce various metals are also strikingly similar – thus, reaffirming their close phylogenic relationship (Kieft *et al*., 1999; Balkwill *et al*., 2004). Figure 2.14 illustrates the close phylogenetic relationship between *T*. *scotoductus* SA-01 and *Thermus* sp. NMX2.A1.

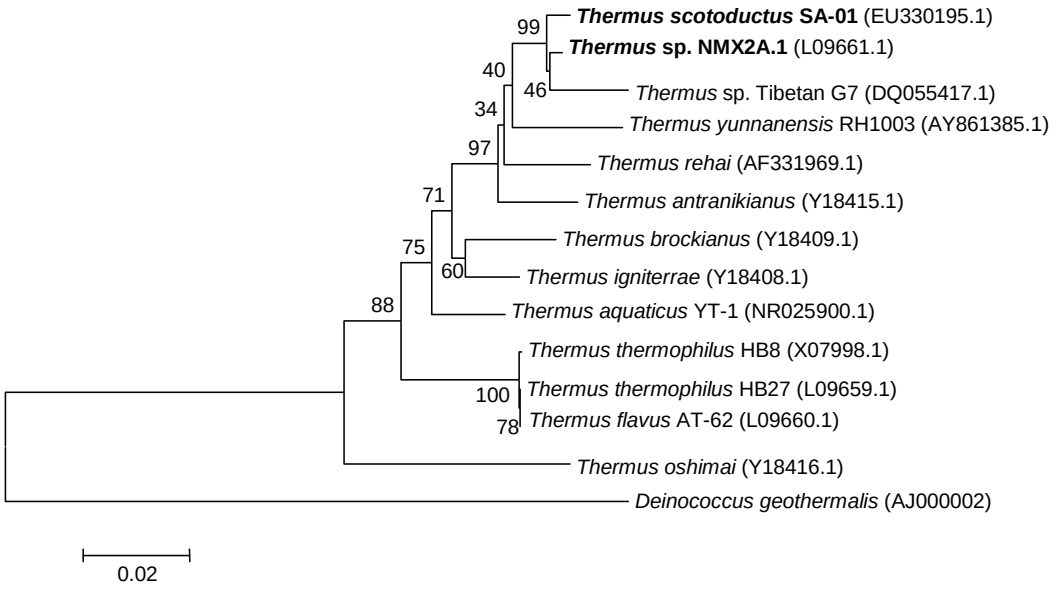


Fig. 2.14 Neighbour-Joining tree of 16S-rRNA gene sequences from various *Thermus* species constructed with MEGA 4.0 software. Multiple alignments were performed with the ClutalW algorithm. The optimal tree was calculated from 1000 replicates and bootstrap values are indicated next to branches. Genbank accession numbers are in parentheses. Evolutionary distances were calculated using the Poisson correction method. *T*. *scotoductus* SA-01 and *Thermus* sp. NMX2.A1 are in boldface.

2.5.2 Pigmentation and carotenoid biosynthesis genes: Indicators of CYP450s in microbes?

We were interested in screening *Thermus* sp. NMX2.A1 for the presence of CYP450 (s) because *Thermus* sp. NMX2.A1 was a yellow pigmented strain which is phylogenetically closely related to *T*. *scotoductus* SA-01. *Thermus* sp. NMX2.A1 was also isolated from a vastly different geographical area than *T*. *scotoductus* SA-01. *T*. *scotoductus* SA-01 was isolated from a gold mine 3.2 km below the subsurface while *Thermus* sp. NMX2.A1 was isolated from a desert hot spring exposed to daily sunlight. *T*. *scotoductus* SA-01 is not pigmented. The yellow coloration of *Thermus* sp. NMX2.A1 could be an indicator of β-carotene and other intracellular accumulated yellow pigments which also hinted at the strong possibility of a megaplasmid containing the β-carotene biosynthesis gene cluster as is the case with *T*. *thermophilus* HB8 and HB27 and other yellow pigmented bacteria *e.g*. *P. agglomerans* (synonym: *E. herbicola*) (Hundle *et al*., 1994) and *Rhodobacter capsulatus* (Armstrong *et al*., 1989).

It is noteworthy to mention that the presence of a β-carotene biosynthesis gene cluster does not necessarily imply that a CYP450 forms part of the gene cluster. For example: *P. agglomerans* and *Pantoea ananatis* (synonym: *Erwinia uredovora*) have carotenoid biosynthesis genes clusters which contain *inter alia* a β-carotene hydroxylase gene but this hydroxylase is not a CYP450 monooxygenase but a hydoxylase that belongs to a superfamily that includes carotene hydroxylases and sterol desaturases (Misawa *et al*., 1990; Hundle *et al*., 1994; Sieiro *et al*., 2003; Momoi *et al*., 2006). In the case of *R*. *capsulatus*, it possesses most of the carotenoid biosynthesis genes like *E*. *herbicola* and *E*. *uredovora* but lacks a β-carotene hydroxylase altogether (Sieiro *et al*., 2003). Thus, it is paramount to bear in mind that although yellow pigmentation and the presence of carotenoid biosynthesis genes may hint at the possibility of a β-carotene hydroxylase gene, the presence and type of hydroxylase should be confirmed by PCR amplification, sequence analyses, cloning and expression to demonstrate activity.

Thermus sp. NMX2.A1 was presumed to have a β-carotene gene cluster due to its yellow pigmentation and we were successful in amplifying a CYP175-gene from this *Thermus* strain since the CYP175-gene did form part of the β-carotene gene cluster.

2.5.3 Proving β-carotene hydroxylase activity: from whole cells to pure proteins

In this study, we were unable to hydroxylate β-carotene when using 50 µM β-carotene suspension and *E*. *coli* cell extracts containing 0.136 µM CYP450 (from *Thermus* sp. NMX2.A1) and FNR and Fdx (both from *T*. *scotoductus* SA-01). Experiments were performed at pH 7.4 and 5.0 at 37˚C as well as 65˚C. There have been a handful of studies in which β-carotene hydroxylation by CYP175A1 was successfully demonstrated (Blasco *et al*., 2004; Momoi *et al*., 2006; Mandai *et al*., 2009). This section provides a brief background of the various approaches that were taken by others to achieve β-carotene hydroxylation. These studies provided the guidelines for the design of our β-carotene hydroxylation experiments.

2.5.3.1 Using *E. coli* whole cells that biosynthesize β-carotene and heterologously express *CYP175A1*

In a study done by Blasco *et al*. (2004), a plasmid (pAC-crtE$_{EU}$-crtB$_{EU}$-I14-Y$_{EU}$) harboring the genes (originating from *E*. *uredovora*) necessary for β-carotene biosynthesis was introduced into *E*. *coli*. Upon induction with IPTG, the *E*. *coli* cultures became yellow due to β-carotene accumulation. Another plasmid (pKK_CYP), carrying the *CYP175A1* gene was then introduced into the abovementioned recombinant *E*. *coli* strain. Upon induction with IPTG the authors reported a slight change in colour in the transformants expressing CYP175A1. Carotenoid pigments were extracted from cell pellets using acetone and hexane and subsequent TLC and HPLC analyses revealed the presence of β-cryptoxanthin, zeaxanthin as well as β -carotene in the strain with the *CYP175A1* gene. Extracted pigments from a negative control, using a pKK233-3 plasmid without any CYP175A1, revealed only β-carotene. All the results concerning extracted pigments were also confirmed with nuclear magnetic resonance (NMR) spectroscopy. The authors also noted that CYP175A1 catalysis was functionally supported by an unknown redox partner from *E*. *coli* – a bacterium that does not possess any endogenous CYP450. It was then suggested that the soluble flavodoxin/flavodoxin reductase (Fpr-Fld) system from *E*. *coli* could possibly replace the CYP450 reductase which supports CYP450 activity.

2.5.3.2 CYP175A1 activity in crude cell-free extracts and with purified proteins

The work performed by Blasco *et al*. (2004) indicated the presence of artificial electron transport proteins from *E*. *coli* that are able to transfer electrons to the heme iron of CYP175A1. In light of this finding, Momoi *et al*. (2006) incubated crude cell-free extracts from *E*. *coli* over-expressing *CYP175A1* using the pKK233-3 expression vector (Yano *et al*., 2003) with 50 µM β-carotene/soybean lecithin suspension and 1 mM NADPH at 37˚C and pH 7.4. HPLC analyses revealed a small peak with a retention time corresponding to β-cryptoxanthin. The authors estimated that 0.013 nmol of β-cryptoxanthin was produced in 30 min at 37˚C when using 0.19 nmol (corresponding to 1.9 µM) CYP175A1. No β-cryptoxanthin was detected when crude cell-free extracts were incubated at 65˚C – an indication that the mesophilic electron transport proteins from *E*. *coli* denatured.

Momoi and co-workers (2006) also used purified CYP175A1 (expressed with the pKK233-3 IPTG induction system) together with two sets of purified redox partners. The first set was the flavodoxin and flavodoxin reductase from *E*. *coli*. In this experiment β-carotene was incubated at 37˚C, pH 7.4 with 1.28 nmol (12.8 µM) of purified CYP175A1, 13 µM flavodoxin and 14 µM flavodoxin reductase. Only very small amounts (concentrations not reported in paper) of β-cryptoxanthin and zeaxanthin were detected. In addition, the production of β-cryptoxanthin and zeaxanthin did not appear to depend on further increasing of CYP175A1 and NADPH (1 mM) concentrations. Finally the authors used an electron transport system from a *Pseudomonas* sp. (putidaredoxin and putidaredoxin reductase) using NADH as electron donor at 37˚C. Both zeaxanthin and β-crypxanthin could be detected by HPLC analyses. The authors also optimized the β-carotene concentration for the CYP175-putidaredoxin/putidaredoxin reductase system. By using 6 µM purified CYP175A1, 20 µM putidaredoxin, 4 µM putidaredoxin reductase and 1 mM NADH, the optimal β-carotene concentration was determined as 50 µM. This β-carotene concentration provided the best rate of β-cryptoxanhin production (*ca*. 0.2 nmol/nmol P450/min) at 37˚C. Time course experiments of β-cryptoxanhin with 50 µM β-carotene and 12 µM CYP175A1 at 37˚C indicated that β-cryptoxanhin production reached a plateau after *ca*. 15 min which corresponded to *ca*. 6 µM of formed β-cryptoxanhin.

The native electron transport proteins from *T*. *thermophilus* HB27 were identified by Mandai *et al*. (2009) as a ferredoxin (Fdx) and ferredoxin-NAD(P)$^+$ reductase (FNR). When purified protein preparations of CYP175A1 and the native electron transport proteins from *T*. *thermophilus* HB27 were used, successful β-carotene hydroxylation was illustrated at 65˚C and pH 5.0 using 20 µM of β-carotene. The turnover rate for this reconstituted system was 12.4 nmol β-cryptoxanhin/nmol P450/min which is *ca*. 54-fold greater than reported by Momoi *et al*. (2006). The study by Mandai *et al*. (2009) finally identified the native redox partners of the CYP175A1-system and this made us confident that the redox partners identified from the genome sequence of *T*. *scotoducutus* SA-01, which were almost identical (Fig. 2.8 A and B), could be reconstituted with the CYP450 from *Thermus* sp. NMX2.A1 to perform β-carotene hydroxylation.

2.5.4 Potential problems with the β-carotene hydroxylation experiments

Our attempts to hydroxylate β-carotene by utilizing crude *E*. *coli* cell-free extracts containing heterologously expressed CYP450 (from *Thermus* sp. NMX2.A1) and Fdx and FNR (both from *T*. *scotoductus* SA-01) were not successful. When Momoi *et al*. (2006) used crude *E*. *coli* extracts containing heterologously expressed CYP175A1 from *T*. *thermophilus* HB27 containing β-carotene and no external surrogate redox partners, a small β-cryptoxanthin peak was detected during HPLC analysis. If we assume that the redox partners from *T*. *scotoductus* SA-01 are not catalytically active and do not reduce the CYP450 from *Thermus* sp. NMX2.A1, we should have, at least, accumulated small amounts of β-cryptoxanthin. This implies that the CYP450 from *Thermus* sp. NMX2.A1 could be the problem and although successful heterologous expression and proper folding was confirmed with CO-difference spectra (Fig. 2.7, section 2.4.4.1) this does not necessarily imply that the protein is catalytically active.

The CYP450 from *Thermus* sp. NMX2.A1 displays 98% amino acid identity to CYP175A1 of *T*. *thermophilus* HB27 with the CYP450 of *Thermus* sp. NMX2.A1 displaying six mutations when compared to CYP175A1 of the HB27 strain (Fig. 2.6 A). Of these six mutations, two are unique when comparing the CYP450 of *Thermus* sp. NMX2.A1 to the CYP175A1 of *T*. *thermophilus* HB8. All six of these mutations are on the periphery of the protein and not in the catalytic site of the protein (Fig. 2.6 B). Therefore

it seems unlikely that these mutations could impair the catalytic function of the CYP450 from *Thermus* sp. NMX2.A1.

Another factor to consider is the final concentration of the CYP450 used. When evaluating the work performed by Momoi *et al*. (2006) the final concentration of CYP175A1 in the cell-free extract was 1.9 µM. The CYP450 in our experiment had a final concentration of 0.136 µM (136 nM). Perhaps the concentration in the reaction was not sufficient to support catalyses? Mandai *et al*. (2009) used as little as 0.03 µM (30 nM) CYP175A1 in their kinetic studies of β-carotene hydroxylation but their system comprised purified CYP175A1, with native purified FNR and Fdx and in addition the assays were performed at 65˚C and pH 5.0. Nonetheless, CYP450 stock concentrations can be improved in the future by inducing more *E*. *coli* culture (*i.e*. increased volume) and by resuspending the biomass in smaller volumes of buffer.

Judging from the cytochrome *c* reductase assay, the redox partners were capable of shuttling electrons from the electron donor NADPH (Fig. 2.10, section 2.4.4.3) between each other and finally to cytochrome *c*. In addition, the FNR and Fdx protein sequences from *T*. *thermophilus* HB27 and *T*. *scotoductus* SA-01 share 88% and 97% amino acid identity respectively. The FNR from *T*. *scotoductus* SA-01 also possesses all the essential binding sites for FAD as well as NADPH. It seems unlikely that the redox partners from *T*. *scotoductus* SA-01 are incompatible with the CYP450 from *Thermus* sp. NMX2.A1 since Momoi *et al*. (2006) managed to demonstrate β-carotene hydroxylation with surrogate electron donors from a *Pseudomonas* sp. However, it has to be taken into consideration that the Fdx, in this system, provides the electrons to the CYP450. The interaction between Fdx and a CYP450 seems to be dependent on electrostatic interactions and weak binding between these proteins can impair catalyses severely (McLean *et al*., 2005; Yang *et al*., 2010).

Although the Fdx from *T*. *scotoductus* SA-01 is 97% identical to the Fdx of *T*. *thermophilus* HB27, there are two amino acid mutations in the protein sequence which could possibly affect the manner in which the Fdx (from *T*.*scotoductus* SA-01) and the CYP450 (from *Thermus* sp. NMX2.A1) interact with each other electrostatically: the polar Y^{56} in the *T*. *thermophilus* HB27 Fdx is replaced by a non-polar F^{56} in the *T*. *scotoductus* SA-01 Fdx and the polar E^{79} (on the C-terminus) by a non-polar G^{79} (Fig. 2.8). In

addition, there are reports that the C-terminal portions of CYP450-associated ferredoxins are important for recognition and electron transfer with their related protein partners (Yang *et al*., 2010). The amino acid differences between the ferredoxins of *T*. *scotoductus* SA-1 and *T*. *thermophilus* HB27 can result in differences in surface charge distributions and could play a significant role in the binding and recognition of the CYP450 from *Thermus* sp. NMX2.A1.

2.6 Concluding remarks

Bacteria from the phylum *Deinococcus*-*Thermus* are categorized as extremophiles since these bacteria display resistance to multiple environmental stresses *e.g*. UV-radiation, oxidation, desiccation and especially high temperatures. These bacteria are usually yellow or red in colour due to their ability to biosynthesize carotenoids like α- and β-carotene and also using these carotenoids as substrates to form other oxygenated pigments (xantophylls). It is, in part, thanks to carotenoids (mostly xantophylls) that these bacteria are in fact able to resist and survive harsh environmental stresses.

There seems to be two themes among β-carotene producing bacteria: (1) the genes responsible for β-carotene biosynthesis form a cluster and are located on a megaplasmid and (2) yellow pigmentation seems to be associated with strains that were isolated from environmental sources regularly exposed to sunlight. These themes were also observed for *T*. *thermophilus* HB27 where β-carotene is hydroxylated by a CYP450 (CYP175A1), which is a β-carotene hydroxylase, to yield the yellow xantophylls β-cryptoxanthin and zeaxanthin. Care should however be taken not to assume that all β-carotene producing bacteria, including species belonging to the *Deinococcus*-*Thermus* phylum, have β-carotene hydroxylase enzymes that are CYP450s. For example: *P. agglomerans* and *P. ananatis* are both β-carotene producing bacteria that have carotenoid biosynthesis gene clusters located on megaplasmids. One of the genes in the cluster is a β-carotene hydroxylase gene but it is not a CYP450 monooxygenase, but a hydroxylase belonging to a superfamily that includes carotene hydroxylases and sterol desaturases. Thus, to identify the type of β-carotene hydroxylase accurately from yellow pigmented bacteria, one has to amplify and sequence the gene. We initially searched for a CYP450 gene in a *T*. *scotoductus* SA-01 strain (not yellow in pigmentation) but were unable to find such a gene. In addition, the genome sequence indicated that *T*.

scotoductus SA-01 possessed megaplasmid genes, but that these genes were scattered throughout the genome. The only remnant of the β-carotene gene cluster found on the genome of *T*. *scotoductus* SA-01 was a single gene encoding a NADH-ubiquinone oxidoreductase. Although *T*. *scotoductus* SA-01 was not yellow or red in pigmentation, several megaplasmid genes were found in the genome of *T*. *scotoductus* SA-01. Could it be that *T*. *scotoductus* SA-01 lost the plasmid since this bacterium was isolated from an environment devoid of sunlight? The genome sequence however assisted us in identifying a Fdx and FNR redox partner set that is almost identical to that of the CYP175A1-system found in *T*. *thermophilus* HB27.

We noticed from literature that a *Thermus* sp. NMX2.A1 strain was phylogenetically closely related to *T*. *scotoductus* SA-01 (based 16S-rRNA sequences) and that this *Thermus* strain was yellow in pigmentation. By using primers based on the adjacent conserved genes in the β-carotene gene clusters of *T. thermophilus* and *T. aquaticus* genome sequences, a CYP450 was isolated from *Thermus* sp. NMX2.A1. This CYP450 displayed 98% amino acid identity to CYP175A1. We could however not confirm that the CYP175A1 homologue is located on a megaplasmid since we used whole-cell PCR to amplify the gene. Since *T*. *scotoductus* SA-01 and *Thermus* sp. NMX2.A1 are closely related we attempted to reconstitute a CYP450-system comprising Fdx and FNR from *T*. *scotoductus* SA-01 and the CYP450 from *Thermus* sp. NMX2.A1.

Cytochrome *c* reductase assays indicated that the Fdx and FNR were capable of shuttling electrons from NADPH to FNR and then from the Fdx to cytochrome *c*. CO-difference spectra of the CYP450 indicated successful expression and proper folding of the *Thermus* sp. NMX2.A1 CYP450. The CYP450, Fdx and FNR, from crude *E*. *coli* cell extracts, were reconstituted in a system containing β-carotene, NADPH and a commercial GDH from *T*. *acidophilum* were used for cofactor regeneration. Hydroxylation experiments were performed at pH 7.4 and pH 5.0 at both 37˚C as well as 65˚C. TLC and LC/MS analyses were inconclusive and we were not able to confidently identify hydroxylated products *i.e*. β-cryptoxanthin and zeaxanthin.

We conclude that the co-factor supply is not a problem since it was provided in excess. We also conclude that the redox partners are probably not the reason for the lack in product formation since the redox partners were capable of shuttling electrons. In

addition, other workers have successfully obtained hydroxylation of β-carotene using CYP175A1 in conjunction with surrogate redox partners from sources other than *Thermus* bacteria. This leaves us with the possibility that the CYP450 from *Thermus* sp. NMX2.A1 might be catalytically inactive. Alternatively, the final concentration of the CYP450 might have been too low to support catalyses. We therefore propose that the following experiments should be performed in future:

2.7 Future research

- Purify the CYP450 from *Thermus* sp. NMX2.A1 and perform binding spectra on the protein using β-ionone as substrate. β-Carotene cannot be used since it strongly absorbs light within the wavelength range of the binding spectra (Momoi *et al*., 2006). This will confirm that the CYP450 is at least capable of binding substrate.
- Once substrate binding in the purified CYP450 has been confirmed, it would also be useful to purify the Fdx as well the FNR and reconstitute the CYP450 system as described by Mandai *et al*. (2009) at both 37˚C as well as 65˚C, pH 5.0.
- The compatibility of the Fdx and FNR from *T*. *scotoductus* SA-01 could also be tested against the CYP175A1 of *T*. *thermophilus* HB27 since β-carotene hydroxylation activity has already been confirmed with this CYP175A1.
- The unique mutations present in the CYP450 of *Thermus* sp. NMX2.A1 can be systematically introduced into the CYP175A1 of *T*. *thermophilus* HB27 to ascertain whether these mutations have an effect on catalysis.
- The stock concentration of the CYP450 in *E*. *coli* extracts can be increased by creating more biomass and resuspending the biomass in a smaller volume of buffer.

Literature cited

Armstrong, G., Alberti, M., Leach, F. and Hearst, J. (1989). Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of *Rhodobacter capsulatus*. *Molecular and General Genetics* **216**, 254 - 268.

Balkwill, D.L., Kieft, T.L., Tsukuda, T., Konstandarithes, H.M., Onstott, T.C., Macnaughton, S., Bownas, J. and Fredrickson, J.K. (2004). Identification of iro-reducing *Thermus* strains as *Thermus scotoductus*. *Extremophiles* **8**, 37 - 44.

Blasco, F., Kauffmann, I. and Schmid, R.D. (2004). CYP175A1 from *Thermus thermophilus* HB27, the first β-carotene hydroxylase of the P450 superfamily. *Applied Microbiology and Biotechnology* **64**, 671 - 674.

Brock, T.D. and Boylen, K.L. (1973). Presence of thermophilic bacteria in laundry and domestic hot-water heater. *Applied Microbiology* **25**, 72 - 76.

Chung, A.P., Rainey, F.A., Valente, M., Nobre, M.F., and da Costa, M.S. (2000). *Thermus igniterrae* sp. nov. and *Thermus antranikianii* sp. nov., two new species from Iceland. *International Journal of Systematic and Evolutionary Microbiology* **50**, 209 - 217.

Girhard, M., Machida, K., Itoh, M., Schmid, R.D., Arisawa, A. and Urlacher, V. (2009). Regioselective biooxidation of (+)-valencene by recombinant *E. coli* expressing CYP109B1 from *Bacillus subtilis* in a two-liquid-phase system. *Microbial Cell Factories* **8**, (page number for citation purpose).

Griffin, S., Higgins, C.L., Soulimane, T. and Wittung-Stafshede, P. (2003). High thermal and chemical stability of *Thermus thermophilus* seven-iron ferredoxin. Linear clusters form at high pH on polypeptide unfolding. *European Journal of Chemistry* **270**, 4736 - 4743.

Gounder, K., Brzuszkiewicz, E., Liesegang, H., Wollherr, A., Daniel, R., Gottschalk, G., Reva, O., Kumwenda, B., Srivastava, M., Bricio, C., Berenguer, J., van Heerden, E. and Litthauer, D. (2011). Sequence of the hyperplastic genome of the naturally competent *Thermus scotoductus* SA-01. *BMC Genomics* **12**, 577 - 590.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**, 557–580.

Hayashi, T., Oka, H., Ito, Y., Goto, T., Ozeki, N., Itakura, Y., Matsumoto, H., Otsuji, Y., Akatsuka, H., Miyazawa, T. and Nagase, H. (2003). Simultaneous analysis of carotenoid colorings in foods by Thin Layer Chromatography. *Journal of Liquid Chromatography and Related Technologies* **26**, 819 - 832.

Henne, A., Brüggemann, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., Johann, A., Lienard, T., Gohl, O., Martinez-Arias, R., Jacobi, C., Starkuviene, V., Schlenczeck, S., Dencker, S., Huber, R., Klenk, H.P., Kramer, W., Merkl, R., Gottschalk, G. and Fritz, H.J. (2004). The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nature Biotechnology* **22**, 547 - 553.

Hoshino, T., Yoshino, Y., Guevarra, E.D., Ishida, S., Hiruta, T., Fujii, R. and Nakahara, T. (1994). Isolation and partial characterization of carotenoid underproducing and overproducing mutants from an extremely thermophilic *Thermus thermophilus* HB27. *Journal of Fermentation and Bioengineering* **7**, 131 - 136.

Hudson, J.A., Morgan, H.W. and Daniel, R.M. (1986). A numerical classification of some *Thermus* isolates. *Journal of General Microbiology* **132**, 531 - 540.

Hundle, B., Alberti, M., Nievelstein, V., Beyer, P.,Kleinig, H., Armstrong, G.A., Burke, D.H. and Hearst, J.E. (1994). Functional assignment of *Erwinia herbicola* Eho10 carotenoid genes expressed in *Escherichia coli*. *Molecular and General Genetics* **245**, 406 - 416.

Kieft TL, Fredrickson JK, Onstott TC, Gorby YA, Kostandarithes HM, Bailey TJ, Kennedy DW, Li SW, Plymale AE, Spadoni CM, Gray MS (1999). Dissimilatory reduction of Fe(III) and other electron acceptors by a *Thermus* isolate. *Applied Environmental Microbiology* **65**, 1214 - 1221.

Kristjansson, J.K., Hjorleifsdottir, S., Marteinsson and V.T., Alfredsson, G.A. (1994). *Thermus scotoductus,* sp. nov., a pigment-producing thermophilic bacterium from hot tap water in Iceland and including *Thermus* sp. X-1. *Systematic and Applied Microbiology* **17**, 44 – 50.

Labuschagne, M. and Albertyn, J. (2007). Cloning of an epoxide hydrolase-encoding gene from *Rhodotorula mucilaginosa* and functional expression in *Yarrowia lipolytica*. *Yeast* **24**, 69 - 78.

Lamble, H.J., Heyer, N.I., Bull, S.D., Hough, D.W. and Danson, M.J. (2003). Metabolic pathway promiscuity in the Archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. *The Journal of Biological Chemistry* **278**, 34066 – 34072.

Macedo-Ribeiro, S.,Martins, B., Pereira, P.J.B., Buse, G., Huber, R.& Soulimane, T. (2001). New insights into the thermostability of bacterial ferredoxins: high-resolution crystal structure of the seven-iron ferredoxin from *Thermus thermophilus*. *Journal of Biological and Inorganic Chemistry* **6**, 663 - 674.

Mandai, T., Fujiwara, S. and Imaoka, S. (2009). A novel electron transport system for thermostable CYP175A1 from *Thermus thermophilus* HB27. *FEBS Journal* **276**, 2416 - 2429.

McLean, K.J., Sabri, M., Marshall, K.R., Lawson, R.J., Lewis, D.G., Clift, D., Balding, P.R., Dunford, A.J., Warman, A.J., McVey, J.P., Quinn, A.-M., Sutcliffe, M.J., Scrutton, N.S. and Munro, A.W. (2005). Biodiversity of cytochrome P450 redox systems. *Biochemical Society Transactions* **33**, 796 - 801.

Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., Harashima, K. (1990). Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of the gene products expressed in *Escherichia coli*. *Journal of Bacteriology* **172**, 6704 - 6712.

Momoi, K., Hofmann, U., Schmid, R.D. and Urlacher, V.B. (2006). Reconstitution of β-carotene hydoxylase activity of thermostable CYP175A1 monooxygenases. *Biochemical and Biophysical Research Communications* **339**, 331 - 336.

Moreira, L.M., da Costa, M.S. and Sá-Correia,I . (1995). Plasmid RFLP profiling and DNA homology in *Thermus* isolated from hot springs of different geographical areas. *Archives in Microbiology* **164**, 7 - 15.

Moreira, L.M. and Sá-Correia, I. (1997). Megaplasmids in *Thermus oshimai* isolates from two widely separated geographical areas: restriction fragment profiling and DNA homology. *Archives in Microbiology* **168**, 473 - 479.

Omura, T. and Sato R. (1964). The carbon monoxide-binding pigment of liver microsomes. I. Evidence of its hemoprotein nature. *Journal of Biological Chemistry* **239**, 2370 - 2378.

Opperman, D.J. and van Heerden, E. (2007). Aerobic Cr(VI) reduction by *Thermus scotoductus* strain SA-01. *Journal of Applied Microbiology* **103**, 1907 - 1913.

Ramaley, R.F. and Hixson, J. (1970). Isolation of non-pigmented, thermophilic bacterium similar to *Thermus aquaticus*. *Journal of Bacteriology* **103**, 527 - 528.

Sachindra, N.M., Bhaskar, N. and Mahendrakar, N.S. (2005). Carotenoids in crabs from marine and fresh waters of India. *LWT - Food Science and Technology* **38**, 221 - 225.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning. A laboratory manual Vol. 1 (2nd ed.). Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Sieiro, C., Poza, M., de Miguel, T. and Villa, T.G. (2003). Genetic basis of microbial carotenogenesis. *International Microbiology* **6**, 11 - 16.

Tabata, K., Ishida, S., Nakahara, T. And Hoshino, T. (1994). A carotenogenic gene cluster exists on a large plasmid in *Thermus thermophilus*. *FEBS Letters* **341**, 251 - 255.

Tian, B. and Hua, Y. (2010). Carotenoid biosynthesis in extremophilic *Deinococcus-Thermus* bacteria. *Trends in Microbiology* **18**, 512 - 520.

Williams, R.A.D., Smith, K.E., Welch, S.G. and Micallef, J. (1996). *Thermus oshimai* sp. nov., isolated from hot springs in Portugal, Iceland, and the azores, and comment on the concept of a limited geographical distribution of *Thermus* species. *International Journal of Systematic Bacteriology* **46**, 403 - 408.

Yang, W., Bell, S.G., Wang, H., Zhou, W., Hoskins, N., Dale, A., Mark Bartlam, M., Wong, L-L. and Rao, Z. (2010). Molecular characterization of a Class I P450 electron transfer system from *Novosphingobium aromaticivorans* DSM12444. *Journal of Biological Chemistry* **285**, 27372 - 27384.

Yano, J.K., Blasco, F., Li, H., Schmid, R.D., Henne, A. and Poulos, T.L. (2003). Preliminary characterization and crystal structure of a thermostable cytochrome P450 from *Thermus thermophilus*. *Journal of Biological Chemistry* **278**, 608 - 616.

Chapter 3

Cytochrome P450 Monooxygenases from Extremely Halophilic Archaea

3.1 Introduction

Cytochrome P450 monooxygenases (CYP450s) from the *Eukarya* and *Bacteria* have enjoyed immense scientific scrutiny since the discovery of CYP450s in rat liver microsomes by Klingenberg in 1958. By 2009 when this project had just started, the CYP450 superfamily comprised according to the CYP450 Homepage of Prof. David Nelson more than 11 500 proteins (Nelson, 2009). Most of these were putative CYP450s identified in the sequenced genomes. More than 90% of these sorted under mammalian, fungal, plant, insect and protist groups. Bacterial CYP450s made up approximately 8% of the CYP450 protein family and those from *Archaea* less than 1%. Current information for the number of sequenced genomes and the number of sequences annotated as CYP450s is summarized in Table 3.1. Thus, bacterial CYP450s currently make up approximately 22% of the CYP450 family while archaeal CYP450s are still only approaching 2%.

The small number of CYP450 sequences from *Archaea* can partly be explained by the fact that a relatively small number of archaeal genomes have been sequenced. This could possibly be ascribed to the fact that most *Archaea* are either difficult to culture or are completely unculturable. However, recent advances in genomics (*e.g*. 454-pyrosequencing-, Illumina- (Solexa) and Helicos technologies) will surely equip us to fully access untapped metagenomes as well as complex genomes (*e.g*. GC-rich genomes) from pure cultures (Delseny *et al*., 2010). This makes the discovery of even more CYP450s very likely and therefore it is plausible to envisage that the archaeal CYP450 number will increase in the near future. However, if we consider the ratio between annotated CYP450s and sequenced genomes it is evident that CYP450s are the least prevalent in *Archaea*.

Table 3.1 Statistics for the distribution of CYP450 sequences (mostly putative from genome sequences) among the three domains of life.

Kingdom	Number[a] of sequenced genomes	Number[b] of annotated CYP450s	$Number_{CYP450s}$ per $Number_{genomes}$
Archaea	85	34 (39[c])	0.4
Bacteria	1638	4667	3.0
Eukarya	303	16689	55
All kingdoms	2026	21390	not applicable

[a] Information available on 21 April 2011 on the genomicBlast page (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi) of the NCBI.

[b] Number of annotated CYP450 sequences used in the compilation of the 3DM CYP450-2011 database (https://fungen.wur.nl/) compiled in April 2011 through a automatic computer generated search of the available protein databases (Swiss Prot, NCBI Non-Redundant Proteins, Patent DB, PDB, Uniprot).

[c] Number of CYP450s from *Archaea* obtained through a BLAST search of the reference protein sequence database of the NCBI on 21 April 2011

CYP450s from *Archaea* have been studied on a very small scale in comparison to CYP450s from the other two kingdoms of life, although 3D structures are available for three archaeal CYP450s. The best studied archaeal CYP450 is the CYP119A1 from the hyperthermophlic acidophile *Sulfolobus acidocaldarius*. CYP119A1 has been cloned, heterologously expressed in *Escherichia coli* and crystallized to a resolution of 1.93 Å. (Wright *et al*., 1996; Yano *et al*., 2000; Nishida & Ortiz de Montellano, 2005). Although the native substrate for CYP119A1 is unknown, hydroxylation of lauric acid has successfully been demonstrated with surrogate as well as the native redox partners (Puchkaev & Ortiz de Montellano, 2005). CYP119A2 from *Sulfolobus tokodaii* strain 7 has also been heterologously expressed and crystallized (Oku *et al*., 2004). The only other archaeal CYP450 that has been studied is CYP231A2 from the thermoacidophile *Picrophilus torridus* and to date only the crystal structure has been studied in depth (Ho *et al*., 2008).

Despite the fact that the three 3D-structures available for archaeal CYP450s are from thermophilic acidophiles, 32 of the 39 CYP450s obtained through a recent (April 2011) BLAST search of the NCBI reference protein sequence database were from halophilic archaea. These 32 CYP450 sequences came from the 13 sequenced genomes of members of the *Halobacteriaceae* family. All the sequenced genomes of halophiles contain at least one CYP450 encoding ORF and the maximum number was nine in the genome of *Haladaptatus paucihalophilus*. Figure 3.1 on the following page illustrates the phylogentic relationship between currently known halophilic archaeal CYP450s as well as other known archaeal CYP450s.

Halophilic archaea sort under the *Halobacteriaceae* family that forms part of the *Euryarchaeota* group in the *Archaea* kingdom. Halophilic archaea flourish and grow optimally in extremely saline environments containing 2 – 5 M NaCl and most grow under aerobic conditions (Oren *et al*., 1997; Coker *et al*., 2007).

The presence of CYP450s in halophilic archaea is quite prevalent when one considers that, on average, each sequenced genome contains 2 CYP450 genes and although not all these CYP450s have been assigned to specific CYP450 protein families, 17 CYP450s have already been assigned to 8 different families. Thus, CYP450s seem to have an important physiological function (s) in halophilic archaea when one considers their prevalence and, based on their different protein families, these CYP450s probably also have different physiological functions in halophilic archaea.

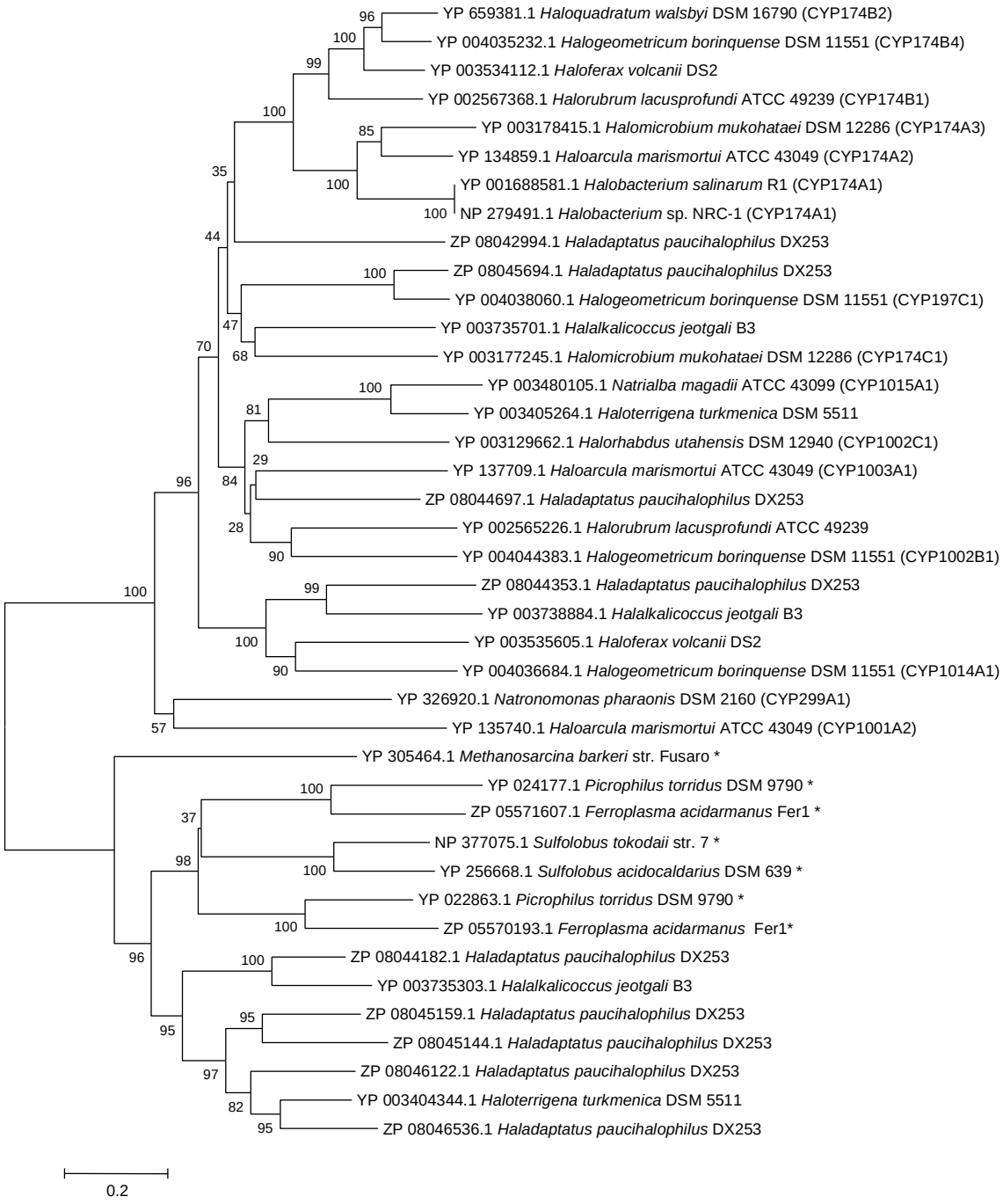


Fig. 3.1 Neighbour-joining tree of all known archaeal CYP450 proteins to date (April 2011) constructed with MEGA 5.04 software. Multiple alignments were performed with the ClustalW algorithm. The optimal tree was calculated from a 1000 replicates and bootstrap values are indicated next to branches. Evolutionary distances were calculated using the Poisson correction method. Haloarchaeal CYP450s that have been assigned to a CYP450 family by Dr. David Nelson are indicated in parentheses. Genera that do not belong to the *Halobacteriaceae* are indicated with asterisks.

From literature one can already speculate about the possible roles that these CYP450s could have in halophilic archaea. There have, for example, been reports of halophilic archaea that are capable of degrading various hydrocarbons (Bertrand *et al*., 1990; Kulichevskaya *et al*., 1991) – a chemical trait that is associated with CYP450s from *e.g*. the asomycetous yeast, *Yarrowia lipolytica* (Fickers *et al*., 2005) and the Gram negative marine bacterium *Alcanivorax borkumensis* (Sabirova *et al*., 2006). One of the outstanding features of most halophilic archaea is their bright red and purple pigmentation which is due to the C_{50} carotenoid bacterioruberin and the membrane protein bacteriorhodopsin – two of the major products of carotenoid metabolism in halophilic archaea (El-Sayed *et al*., 2002, Tarasov *et al*., 2008). In a study by Calo *et al*., (1995), the presence of a hydroxylated carotenoid (trans-astaxanthin) was reported in *Halobacterium salinarum* and *Haloarcula hispanica*. It has been demonstrated in the heterobasidiomycetous yeast *Xanthophyllomyces dendrorhous* (Álvarez *et al*., 2006) and the green algae *Haematococcus pluvialis* (Schoefs *et al*., 2001) that astaxanthin production is dependent on *inter alia*, a CYP450 protein which confers two hydroxyl groups to the molecule.

Bearing in mind the possible physiological functions these CYP450s could have, we decided to elucidate the physiological function(s) of CYP450s from halophilic archaea. To our knowledge this is the first investigation into elucidating the role of CYP450s from halophilic archaea.

The initial aims of this project were therefore to **(i)** isolate an extremely halophilic archaeon from a South African saltern; **(ii)** isolate a CYP450 from the ensuing isolate and **(iii)** characterize the functionality of the CYP450 by heterologous expression in *E. coli.* This was considered feasible since several other genes from various genera of extremely halophilic archaea have been heterologously expressed using *E*. *coli*. When expression in *E. coli* failed, an additional aim became to **(iv)** delete the CYP450 gene in *Halobacterium salinarum* R1 to investigate possible effects on phenotype and study the effect of the deletion on a transriptomic level using microarray. It was also envisaged that the CYP450 deleted strain will be useful for the heterologous expression of CYP450s from other halophiles.

3.2 Materials and Methods

3.2.1 Microbiological methods

3.2.1.1 Isolation of archaeal strains from brine crystals

Brine crystals were collected from a commercial saltpan on the Lemoenkloof farm (GPS coordinates: 28˚ 43' 46'' S 26˚ 03' 18'' E) in February 2009. The crystals were directly inoculated into test tubes containing 5 mL of halophilic growth medium as described by Kamekura *et al*., (1998). Cultures were shaken at a speed of 160 r.p.m. at 37˚C for 2 weeks or until growth was visible.

3.2.1.2 Media and growth conditions

Propagation of plasmids in *Escherichia coli* was performed in Luria-Bertani (LB) broth (Sambrook *et al*., 1989) at 37˚C with agitation at 160 r.p.m. Selective pressure for *E*. *coli* was maintained by supplementing the LB broth with ampicillin, kanamycin or gentamicin to final concentrations of 100 µg/mL, 50 µg/mL and 100 µg/mL respectively. Solid media cultivations were performed by supplementing the growth media with 15 g/L bacteriological agar and selective pressure was maintained with antibiotics at the following final concentrations: 60 µg/mL ampicillin, 50 µg/mL kanamycin or 100 µg/mL gentamicin. *Pseudomonas fluorescens* KOB2Δ1 was cultured in liquid LB medium at 30˚C with agitation at 160 r.p.m. without any antibiotic. Recombinant *P*. *fluoerescens* KOB2Δ1 strains for heterologous expression (containing pCOM8 constructs) were cultured in LB medium containing 100 µg/mL gentamicin at 30˚C.

All extremely halophilic, archaeal strains used in this study were cultured in complete medium described by Oesterhelt and Krippahl (1983) that contained (per 1 L): 20 g $MgSO_4 \cdot 7H_2O$, 3 g tri-sodium citrate, 250 g NaCl, 2 g KCl and 10 g Peptone. Strains were cultured at 40˚C at a shaking speed of 200 r.p.m. Solid media cultivations were performed by supplementing the broth with 15 g/L bacteriological agar and when appropriate, selective pressure for the halophilic archaeal strains was maintained with a

final concentration of 10 µg/mL mevinolin (lovastatin) dissolved in DMSO (final DMSO concentration in medium was 0.1 % v/v). When required plates were also spread with 40 µL of 20 mg/mL X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside; 40 mg/mL dissolved in Dimethyl formamide and diluted with water). Bacterial strains were cryopreserved at -80˚C in 50 % (v/v) glycerol and all halophilic, archaeal strains used in this chapter, at -80˚C in 25 % (v/v) glycerol. Strains used in the study are listed in Table 3.2

Table 3.2 Bacterial and Archaeal strains and plasmids used in this study

Strains / Plasmids	**Genotype / characteristics**	**Source / Reference**
Halobacterium salinarum R1[a]	Wildtype genotype Gas vesicle deficient	Stoeckenius & Kunau (1968) Strahl & Greie (2008)
TOP10 *Escherichia coli*	F-*mcr*A Δ*(mrr-hsd*RMS-*mcr*BC*)* Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ*(ara leu) 7697 gal*U *gal*K *rps*L *(*Str^R*) end*A1 *nup*G	Invitrogen
Escherichia coli BL21 (DE3)	F– *omp*T *hsd*SB(rB–, mB–) *gal dcm* (DE3)	Invitrogen
Pseudomonas fluorescens KOB2Δ1	Δ*alkB* mutant from *P*. *fluorescens* CHAO	Smits *et al*., 2002
pGEM®-T Easy	Blue/White selection, TA-cloning Amp^R	Promega
pET28b (+)	6x His tag (N- and C-terminal) Kan^R	Novagen
pCOM8	P*alkB*, *oriT alkS* regulator, Gen^R	Smits *et al*., 2001
pMKK100[a]	Blue/Red selection, shuttle and suicide vector *bga*H , Amp^R, Mev^R	Koch & Oesterhelt (2005) del Rosario *et al*., (2007)
pNG168	Shuttle vector Mev^R	ATCC (MBA77) DasSarma, (1995)

[a]*Halobacterium salinarum* R1 and the pMKK100 suicide vector were generous gifts from Prof. Dieter Oesterhelt from the Max Planck Institute of Biochemistry, Martinsried, Germany.

3.2.2 Recombinant DNA techniques

3.2.2.1 Enzymes, chemicals, kits and other consumables

All DNA modifying enzymes with their respective buffers, PCR reagents and DNA purification kits were, unless otherwise stated, obtained from Bioflux Corporation, Applied Biosystems (ABI), Fermentas, New England Biolabs (NEB) or Roche Molecular Biochemicals.

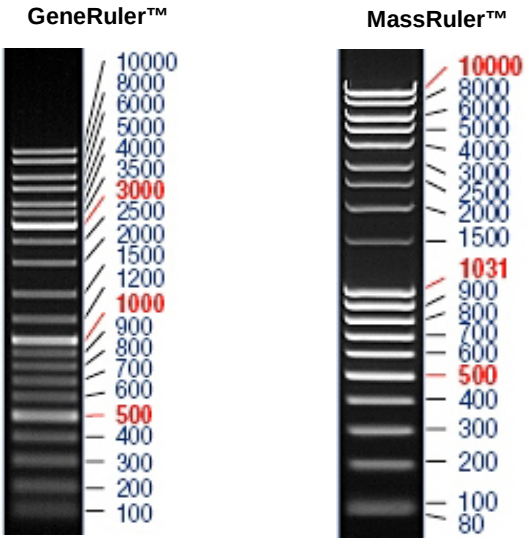
Chemicals were of analytical or molecular biology grade and were, unless otherwise stated, obtained from either Merck or Sigma-Aldrich,. All oligonucleotides were, unless otherwise stated, obtained either from Inqaba Biotechnical Industries Pty (Ltd) (South Africa) or Integrated DNA Technologies (IDT). Oligonucleotide design and analyses were performed using an online algorithm from IDT with default settings (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/). Oligonucleotides (T7 and SP6) for sequencing purposes were obtained from IDT. All relevant oligonucleotides used in this chapter are collated in Table 3.3.

Table 3.3 Oligonucleotide primers used in this chapter

Primer name	5´ – 3´ DNA sequence	Application	Restriction sites/ Comments
23S_F	5'- cta agc tca cct ccc ggc tg -3'	Haloarchaeal genus typing	No restriction sites
23S_R	5'- ccg ata gtg aac aag tag tgt gaa cga acg -3'	Haloarchaeal genus typing	No restriction sites
SP_F1	5'- cag tag gaa cca ggt gta ggt cag c -3'	Haloarchaeal P450 screen	Conserved I-helix motif
SP_R2	5'- ctt ccc gtt cgg tgg tgg -3'	Haloarchaeal P450 screen	Conserved heme binding loop
HM_F1	5'- atg tca aag acg ccg ccc g -3'	ORF of saltpan P450	No restriction sites
HM_R1	5'- tta ccg ttc ctg cac gcg c -3'	ORF of saltpan P450	No restriction sites
HM_F1_Nde	5'- catatg tca aag acg ccg ccc g -3'	Cloning of saltpan P450 into pET28b(+)	*Nde*I
HM_R1_Hind	5'- aagctttta ccg ttc ctg cac gcg c -3'	Cloning of saltpan P450 into pET28b(+)	*Hin*dIII
HM_F1_Nde	5'- catatg tca aag acg ccg ccc g -3'	Cloning of saltpan P450 into pCOM8	*Nde*I
HM_R1_Xma	5'- cctaggtta ccg ttc ctg cac gcg c -3'	Cloning of saltpan P450 into pCOM8	*Xma*JI
US_Hind_F	5'- aag ctt cga agt cgg cgt cct gct c -3'	US-Deletion cassette	*Hin*dIII
Prom_R	5'- ctg cag acg tac gtc tcc atg ggt ccc -3'	US-Deletion cassette	*Pst*I
Term_F	5'- ctg cag aac agg aga tgc gga tgc gg -3'	DS-Deletion cassette	*Pst*I
DS_Bam_R	5'- gga tcc ctg gga cgt cgg cat gag -3'	DS-Deletion cassette	*Bam*HI
421_F	5'- atg cag gat gcc ggc att cc -3'	P450 deletion screen	Gene specific primers[a]
424_R	5'- tca ctc gtc tac gtg gtc ga -3'	P450 deletion screen	Gene specific primers[a]
Int_F	5'- gcg gcc gtt gtg tgg ctg gtt t -3'	Site specific integration	Pair with DS_Bam_R
1464_F	5'- atg acg agc gtc cag aac acc -3'	*Bat* gene amplification	Gene specific primers[a]
1464_R	5'- tca ctc ctc gaa gaa cgc tcc -3'	*Bat* gene amplification	Gene specific primers[a]
1465_F	5'- atg ctc ggt agt gac gtg tgt -3'	*Brp* gene amplification	Gene specific primers[a]
1465_R	5'- tca tgg gac gta cca gat gcc -3'	*Brp* gene amplification	Gene specific primers[a]
1467_F	5'- atg ttg gag tta ttg cca aca -3'	*Bop* gene amplification	Gene specific primers[a]
1467_R	5'- tca gtc gct ggt cgc ggc cgc -3'	*Bop* gene amplification	Gene specific primers[a]

[a]Oligonucleotides were a generous gift from Prof. Shiladitya DasSarma, Department of Microbiology and Immunology, University of Maryland (Baltimore), Maryland, USA.

All molecular weight markers were, unless otherwise stated in the text, obtained from Fermentas. Molecular weight markers are shown below:



3.2.3 Quantification of nucleic acids

Nucleic acid concentrations and purity were determined and assessed with a Nanodrop® ND-1000 Spectrophotometer.

3.2.4 PCR amplification

Polymerase chain reactions (PCR) were performed with *Taq* DNA Polymerase (New England Biolabs) or the Expand Long Template system (Roche Molecular Biochemicals) according to the manufacturer's specifications. For the PCR experiments using *Taq* DNA Polymerase, the reaction mixture (final volume of 50 µL using PCR grade water) contained components with the following final concentrations: 5 µL 10 x ThermoPol buffer, 0.2 µM forward and reverse oligonucleotides (melting temperatures were always 60˚C unless stated otherwise), 0.2 µM dNTP's, at least 10 ng template DNA, 2 µg/mL Bovine Serum Albumin (BSA) and 0.5 µL *Taq* DNA Polymerase (2.5 U). For PCR experiments using the Expand Long Template system, the reaction mixture contained components with the following final concentrations: 5 µL 10 x Buffer no. 1 (unless stated otherwise), 300 nM forward and reverse oligonucleotides, 350 nM dNTP's, 2 µg/mL BSA, at least 10 ng DNA template and 3.75 U Expand Long Template Polymerase mix.

Reactions were made up to 50 µL with PCR grade water. Thermal cycling (Tables 3.4 and 3.5) was performed with an Eppendorf Mastercycler Gradient

Table 3.4 Standard PCR cycling reaction for the *Taq* DNA Polymerase

Step	Temperature	Time	Cycle number
Initial denaturation	94˚C	2 min	1
Denaturation	94˚C	15 sec	
Annealing	55˚C	30 sec	10
Elongation	72˚C	45 sec[a]	
Denaturation	94˚C	15 sec	15
Annealing	55˚C	30 sec	
Elongation	72˚C	45 sec + 5 sec for each successive cycle	
Final elongation	72˚C	7 min	1

Table 3.5 Standard PCR reaction for the Expand Long Template system

Step	Temperature	Time	Cycle number
Initial denaturation	94˚C	2 min	1
Denaturation	94˚C	10 sec	
Annealing	55˚C	30 sec	10
Elongation	68˚C	4 - 8 min[b]	
Denaturation	94˚C	10 sec	
Annealing	55˚C	30 sec	
Elongation	68˚C	4 – 8 min + 20 sec for each successive cycle	15
Final elongation	72˚C	7 min	1

[a] Elongation time was based on 45 sec/1kb.
[b] Elongation time was dependent on the desired amplicon size

3.2.5 Sequence analyses

Templates for sequencing were purified using the Biospin Plasmid DNA Extraction or Gel Extraction kits (Bioflux) according to the manufacturer's instructions. For double stranded templates 500 ng of DNA was used in each sequencing reaction. Sequencing was performed with an ABI Prism® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit 3.1 (Applied Biosystems, USA) according to the manufacturer's specifications. Universal sequencing primers T7 or SP6 were used at a final concentration of 3.2 pmol in a 10 µL reaction that consisted of: 0.5 µL premix, 2 µL 5 x dilution buffer and PCR grade water.

The sequencing PCR cycle comprised of: an initial denaturation of 1 min at 96˚C, followed by 25 successive cycles of denaturation at 96˚C for 10 sec, annealing at 50˚C for 5 sec and elongation at 60˚C for 4 min. The reactions were then cooled to 4˚C. Post reaction clean-up consisted of an EDTA/Ethanol precipitation (as recommended by ABI): the sequencing reaction volume was adjusted to 20 µL and 5 µL 125 mM EDTA (pH 8.0) and 60 µL absolute ethanol was added. Reactions were vortexed for 5 s and left at room temperature for 15 min after which the reactions were centrifuged at 20 000 x *g* for 10 min at 4˚C. The supernatant was completely removed and 60 µL 70 % (v/v) ethanol was added, followed by centrifugation at 20 000 x *g* at 4˚C for 5 min. The supernatant was completely aspirated and the samples dried under vacuum at 30˚C. Samples were stored in the dark at 4˚C until sequencing analyses could commence.

Nucleotide composition was determined on a 3130xl Genetic Analyzer (ABI). The resulting electropherograms were analyzed and edited using FinchTV 1.4.0 (Geospiza Inc.) and DN*Assist* 3.0. Contigs were assembled using ContigExpress (a component of the Vector NTI suite 9.0.0).

3.2.6 Assessment of PCR and restriction digest products

All amplicons and digestion products were electrophoresed and assessed on 0.8 % (w/v) agarose gel (unless stated otherwise) containing 0.6 µg/mL ethidium bromide. Agarose gels were prepared and electrophoresed in 1 x TAE buffer (0.1 M Tris, 0.05 M EDTA and 0.1 mM glacial acetic acid, pH 8.0) at 6 V/cm for 1 h. Visualizations were done with a

GelDoc XR (Bio-Rad Laboratories) and Quantity One 4.6.3 software under short wavelength UV light. For cloning purposes, DNA was visualized with a DarkReader™ transilluminator (Fermentas), excised from agarose gels and purified.

3.2.7 Transformation of *E. coli*, *P. fluorescens* and *H. salinarum* R1

3.2.7.1 *E. coli* Top 10 transformations

Transformations in *E*. *coli* were performed with Top 10 cells obtained from Invitrogen. Cells were rendered chemically competent by means of a modified version of the $RbCl_2$ method as originally described by Hanahan (1983) and used as 50 µL aliquots. Cloning of ligation mixtures were done with pGEM®-T Easy (Promega) according to the manufacturer's specifications. Transformations were performed as described by Sambrook *et al*. (1989) and cells were then plated onto LB plates supplemented with ampicllin, IPTG [isopropylthio-β-D-galactoside (10 mg/mL) and X-gal [5-bromo-4-chloro-3-indolyl- β-D-galactoside (40 mg/mL)]. Plates were incubated at 37˚C for 15 hours. White clones (presumptive positive transformants) were selected and inoculated into 5 mL LB-media supplemented with ampicillin and grown for 15 hours at 37˚C with aeration (160 r.p.m).

3.2.7.2 *E. coli* BL21 (DE3) transformations and induction

Transformations of *E*. *coli* BL21 (DE3) with pET28 b(+), containing the saltpan CYP450, was performed as described above except that selection was done with 50 µg/mL kanamycin and no X-gal was added to the plates. For induction purposes, all strains were grown in 50 mL LB media at 37˚C until OD_{620} = 0.5. Cultures were supplemented with δ-aminolevlinic acid (which is a heme precursor, final concentration 1 mM) and $FeCl_2 \cdot 4H_2O$ (final concentration 0.5 mM). The final IPTG concentration and induction temperature combinations tested were as follows: 1mM IPTG / 35˚C, 0.25 mM / 20˚C, 0.25 mM IPTG / 16˚C and 0.25 mM IPTG / 4˚C (in this experiment cultures were not agitated). Induction was performed for as long 24 h. Cells were harvested at 4000 x *g* for 15 min and the pellets washed three times with 10 mL 20 mM MOPS, pH 7.0. Cell suspensions were sonicated at 60 % power for two cycles consisting of 30 sec followed by 1 min on ice. The *E*. *coli* lysate was centrifuged at 20 000 x *g* for 10 min. Both the

soluble and insoluble fraction (pellet) of the crude protein preparation were loaded onto a 10 % SDS PAGE (Laemmli, 1970) and stained (Fairbanks *et al*., 1971) to evaluate CYP450 expression.

3.2.7.3 *P. fluorescens* KOB2Δ1 transformations

P. *fluorescens* KOB2Δ1 was transformed as described by Højberg *et al*. (1999). In short, *P*. *fluorescens* KOB2Δ1 was cultured in 25 mL LB medium at 35˚C for 5 h (OD_{620} = 0.6 – 0.7). Cells were harvested at 4 000 x *g* for 15 min at 4˚C and resulting pellets were washed twice with 10 mL of 1mM 3-morpholinopropanesulfonic acid (MOPS) buffer containing 15 % (w/v) glycerol, pH 7.0. Washed pellets were resuspended in 200 µL MOPS-glycerol solution and used immediately. *P*. *fluorescens* KOB2Δ1 cell suspension (40 µL) was mixed with *ca*. 0.5 µg plasmid DNA (pCOM8 containing the cloned saltpan CYP450 or pCOM8 only to serve as a negative control) and transformed by electroporation by using the following conditions: 25 µF, 200 Ω and 2.5 kV/cm. One milliliter of pre-warmed (35˚C) SOC medium (containing 20 g/L Bacto tryptone, 5 g/L Yeast extract, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 2.5 mM KCl and 20 mM D-glucose) was added to the cell suspension, mixed by gentle pipetting and incubated at 35˚C for 3 h. The entire mixture was plated onto five LB plates containing 100 µg/mL gentamicin and incubated at 30˚C overnight or until colonies appeared.

3.2.7.3.1 Induction of *P. fluorescens* KOB2Δ1 transformed with the pCOM8 vector containing the saltpan CYP450

P. *fluorescens* KOB2Δ1 is a *alkB* knock-out strain *i.e*. it has no alkane hydroxylase and has thus lost its ability to degrade *n*-alkanes. By transforming the knock-out strain with pCOM8 (containing the saltpan CYP450), one can attempt to prove that the possible function of the saltpan CYP450 is to assist in the oxidation of *n*-alkanes. The pCOM8 vector also contains an *alkB* promoter as well as a positive regulator called *alkS*. *AlkS* is activated by C_7-C_{12} *n*-alkanes, alkenes, and gratuitous inducers such as haloalkanes, ethylacetate, ethylether, and dicyclopropylketone (DCPK) (Grund *et al.*,1975; Smits *et al*., 2001).

To assess the saltpan CYP450's possible ability to oxidize *n*-alkanes, transformed *P*. *fluorescens* KOB2Δ1 (carrying pCOM8 containing the saltpan CYP450 gene) was

streaked onto LB plates containing gentamicin and incubated at 30˚C overnight. Biomass was then streaked onto solid minimal E_2 medium (Smits *et al*., 2001) containing gentamicin as well as δ-aminolevlinic acid and $FeCl_2 \cdot 4H_2O$. C_8 – C_{10} *n*-alkanes were provided as sole carbon source in vapour form by placing 1 mL *n*-alkane in the petri dish lid and incubating the plates in a desicator at 30˚C for 5 days. A *P*. *fluorescens* KOB2Δ1 strain containing the empty pCOM8 vector was also included as a negative control.

To evaluate expression of the CYP450 from the saltpan isolate by SDS-PAGE, the saltpan CYP450-containing *P*. *fluorescens* KOB2Δ1 strain was cultured and induced as described by Smits *et al*., (2001). Strains were cultured on LB plates as described above and inoculated into liquid 100 mL E_2 medium containing 0.2 % (w/v) tri-sodiumcitrate gentamicin. Strains were cultured at 30˚C aerobically until OD_{450} = 0.5 – 1.0 was reached. Cultures were then induced with 0.05 % (v/v) DCPK (final concentration) and δ-aminolevlinic acid and $FeCl_2 \cdot 4H_2O$ were added. Induction with DCPK was performed for 5 h. Cell pellets were harvested at 4 000 x *g* and washed once with 10 mM $MgSO_4$ followed by two wash steps with 50 mM potassium phosphate buffer, pH 7.5. Pellets were finally resuspended in 0.6 mL potassium phosphate buffer and crude protein extract for SDS PAGE was performed as described in section 3.3.7.2.

3.2.7.4 *H. salinarum* R1 transformations

H. *salinarum* R1 cells were transformed as described by Cline *et al*. (1987) with some minor modifications as described by Koch & Oesterhelt (2005). To increase the competence of the cells, cells were firstly cultured in 35 mL complete medium (Oesterhelt and Krippahl, 1983) in 100 mL Erlenmeyer flasks to a cell density of OD_{600} = 0.3 – 0.4. Cells were then diluted 200 fold in 35 mL complete medium and re-cultured 3 times to a cell density of OD_{600} = 0.3 – 0.4. Cells were then finally cultured to a density of OD_{600} = 0.5 – 0.8. Two milliliters of the culture was then harvested at 8 000 x *g* for 2 min at room temperature and the supernatant completely removed. The pellet was gently resuspended in 200 µL spheroplasting solution (SPS) which contained (per 100 mL): 11.7 g NaCl, 0.186 g KCl, 0.606 g Tris base and 15 g sucrose, pH 8.75 adjusted with 10 M NaOH. Cells were converted to spheroplasts by adding 5 µL EDTA in SPS (18.6 g EDTA per 100 mL SPS) to the suspension followed by very gentle mixing and incubation for 10 min at room temperature.

The entire mixture was transferred to a 2 mL microfuge tube and gently mixed with a solution of 1.0 – 1.5 µg plasmid DNA and 1 M NaCl (14 µL in total) to yield a final volume of *ca*. 220 µL. The mixture was incubated for 5 min at room temperature after which 220 µL 60 % (w/v) PEG_{600} (in SPS) was added to the lid of the tube (positioned horizontally). The lid was closed, followed by immediate, vigorous shaking to avoid cell lysis due to excessive high PEG_{600} concentrations. After 15 min, 1.6 mL complete medium was added and the mixture centrifuged at 8 000 x *g* for 2 min at room temperature. The supernatant was completely removed and the spheroplasts gently re-suspended in 2 mL complete medium, followed by centrifugation as described above. The resulting pellet was gently suspended in 600 µL complete medium. Note that the complete medium as well as all the SPS-containing transformation solutions were filter sterilized with a 0.20 µm cellulose acetate syringe filter (Gema Medical S.L., Barcelona).

The 2 mL microfuge tubes were closed and covered with parafilm, placed in a horizontal position and shaken at 250 r.p.m. for 24 h at 37˚C to allow the S-layer of the cells to recover. Fifty microliters were directly plated onto solid medium containing mevinolin and the remainder of the mixture was pelleted as described above, re-suspended in 50 µL complete medium and plated onto mevinolin containing plates. Plates were incubated in an enclosed box (containing a Petri dish lid with water to prevent the plates from drying out) at 40˚C for 5 – 7 days or until colonies appeared.

3.2.8 Construction of a pMKK100 based cassette for CYP450 knock-out

Figure 3.2 provides a summary of the construction of the deletion cassette. Oligonucleotides were designed to amplify *ca*. 1 kb of the upstream (US) and downstream (DS) region adjacent to the CYP450 gene (Fig. 3.2 and Table 3.3). *H. salinarum* R1 was cultured until late stationary phase (OD_{600} = 1.2) in 5 mL complete medium as previously described. Two milliliters of culture was harvested for 1 min at maximum speed and the genomic DNA extracted from the cell pellet with the glass bead method described by Labuschagne and Albertyn (2007). The only modification was that cell suspensions were vortexed for 2 min with a 1 min incubation period on ice after the first minute of vortexing. PCRs using above mentioned primers were performed with Expand Long Template Polymerase to ensure a high degree of fidelity. The upstream and downstream fragements were fused and the fused deletion cassette cloned into the

pMKK100 suicide vector (Fig. 3.2). The resulting construct was transformed into competent *H. salinarum* R1 cells to create a CYP450 knock-out.

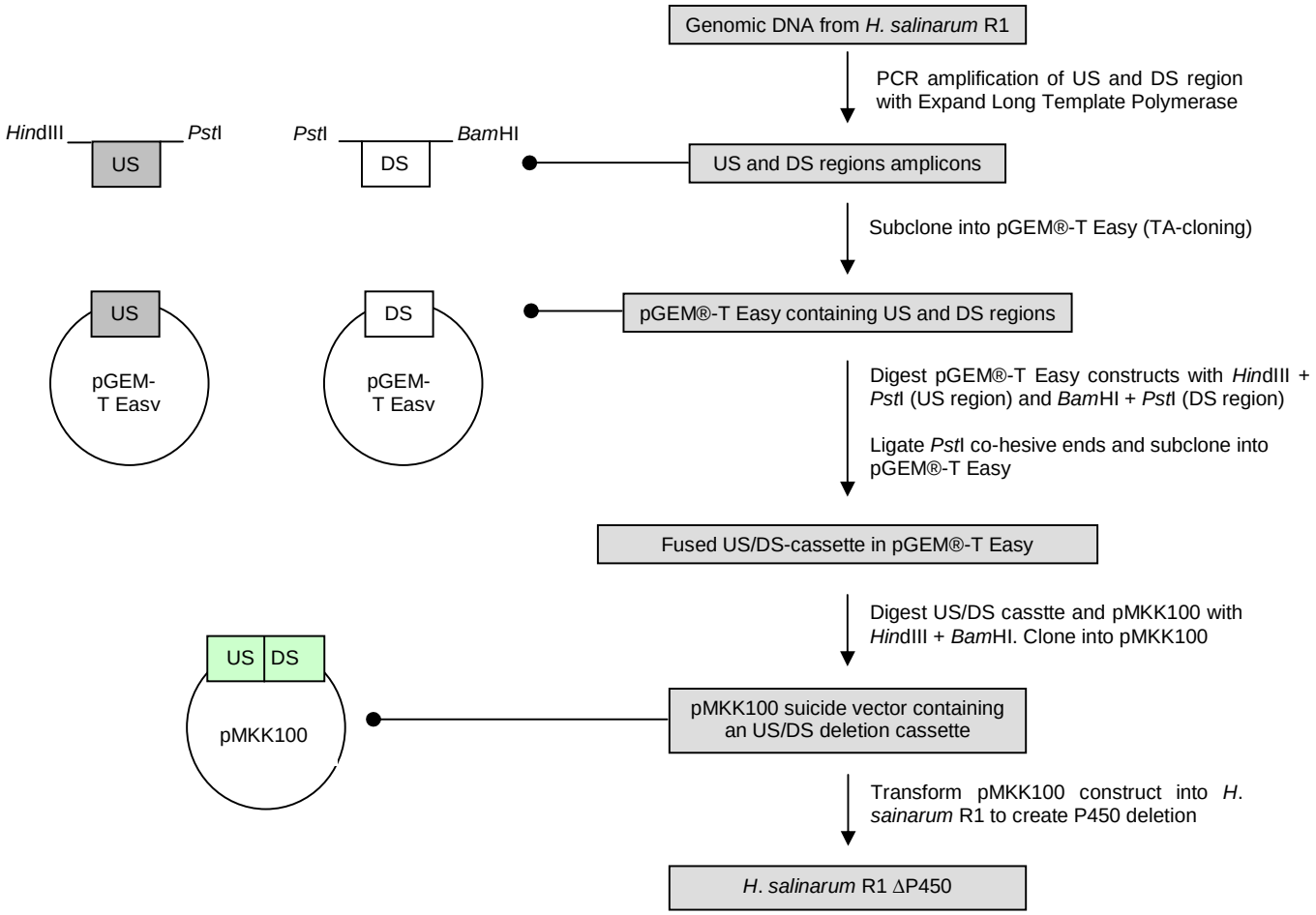


Fig. 3.2 Flow chart of cloning strategy for creating the US/DS-deletion cassette in the suicide vector pMKK100 to generate a CYP450 knock-out in *H. salinarum* R1.

3.2.9 Deleting the single chromosomal copy of CYP450 in *H. salinarum* R1

After transformation and streaking on plates containing mevinolin and X-gal, the plates were incubated at 40˚C for 5 – 7 days or until colonies appeared. Successfully transformed cells *i.e*. cells harboring the integrated pMKK100 plasmid containing the *bgaH* gene (halophilic β-galactosidase) formed bright blue colonies on X-gal containing plates (Patenge *et al*., 2000).

When the single colonies became large enough, four blue colonies were picked and transferred to four test tubes containing 5 mL complete medium without any mevinolin. Cultures were incubated at 40˚C at 100 r.p.m and after the cultures reached an OD_{600} = 0.3 – 0.4, the cultures were diluted 200-fold in 35 mL complete medium in 100 mL Erlenmeyer flasks without any mevinolin. Cells were cultured 3 times to a cell density of OD_{600} = 0.3 – 0.4 and then finally cultured to a density of OD_{600} = 0.5 – 0.8 at a shaking speed of 100 r.p.m. After the final round of culturing (OD_{600} = 0.5 – 0.8), the cell suspensions were diluted 10^{-5} and 10^{-6} fold with complete medium to a final volume of 100 µL. The entire 100 µL was plated out onto complete medium containing X-gal and no mevinolin.

Plates were then incubated at 40˚C for 5 -7 days or until colonies became visible. At least 10 red colonies from each plate were picked for the deletion screening experiment and all blue colonies excluded.

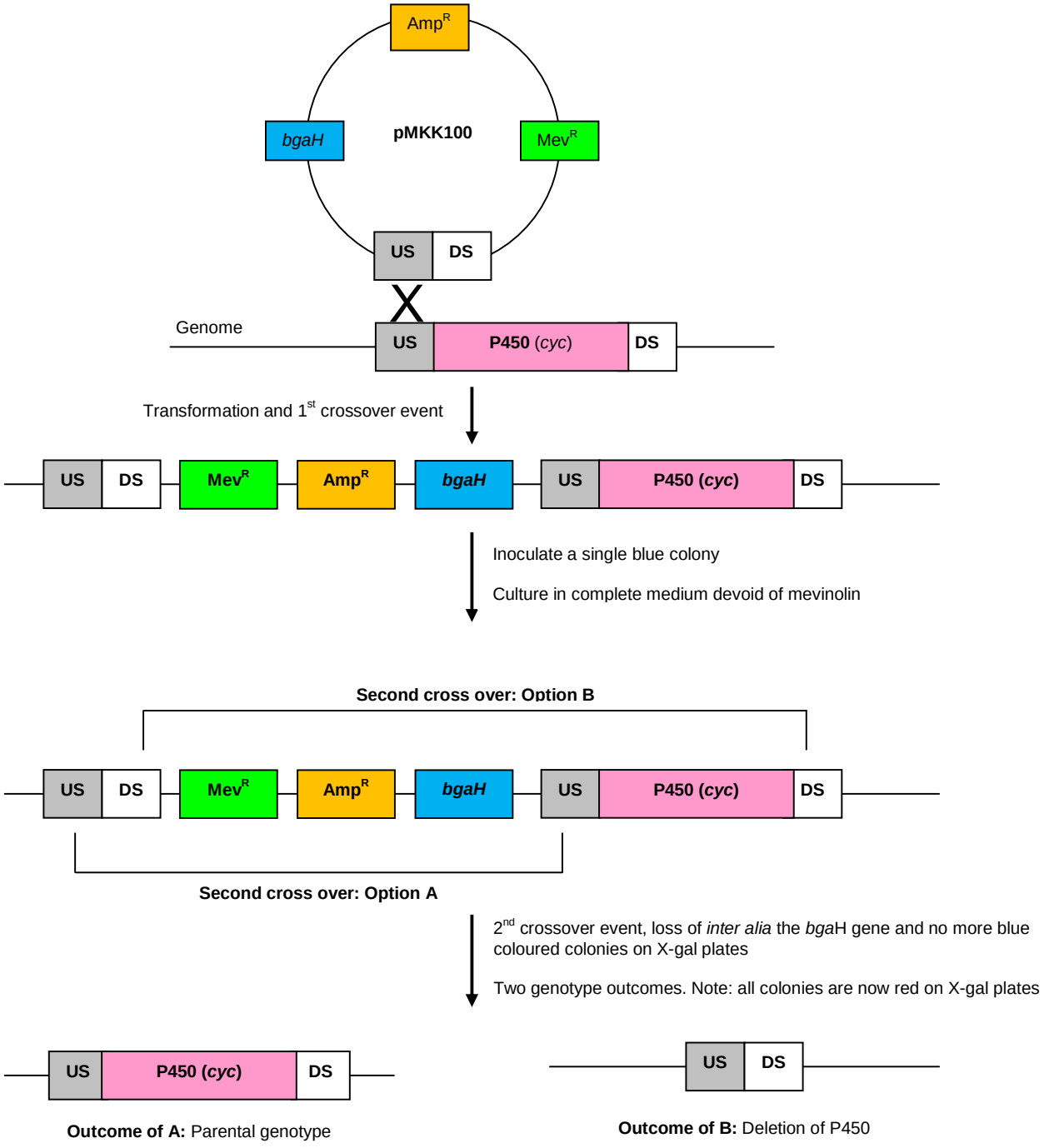


Fig. 3.3 Summary of the Blue/Red selection experiment to delete the CYP450 by utilizing the pMKK100 suicide vector.

3.2.10 Screening for CYP450 deletion mutants with PCR

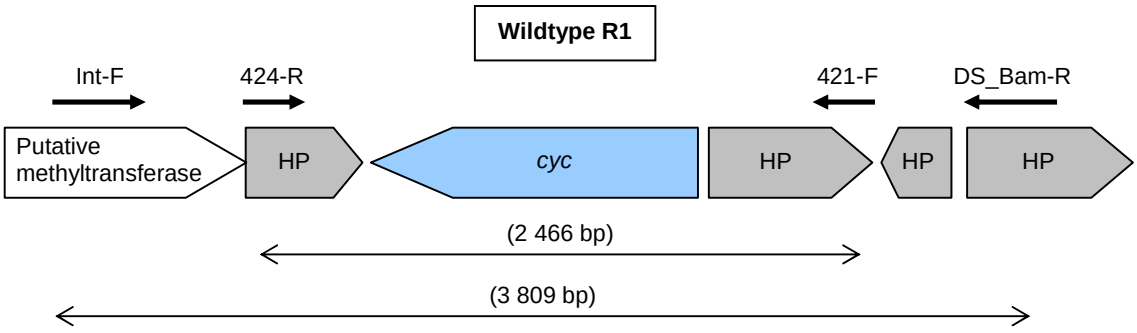
Red colonies were picked and inoculated into 5 mL complete medium without any selective pressure. Cultures were grown at 40˚C until an OD_{600} = 0.4 and their genomic DNA extracted as described by Labuschagne and Albertyn, (2007). To assess (i) successful CYP450 deletion and (ii) confirm the locus of the deletion, two separate PCR screens were performed. The deletion screening PCR was performed with oligonucleotides (421-F and 424-R) based on sequences of genes adjacent to the CYP450 gene. See http://halo4.umbi.umd.edu/cgi-bin/haloweb/nrc1.pl?operation=query&id=334 for gene topologies and annotations of the *H*. *salinarum* NRC-1 and R1 genomes) as well as DasSarma *et al*., 2010.

PCR reactions were performed with 2.5 U *Taq* DNA Polymerase, 0.25 µM forward and reverse oligonucleotides, 1 x ThermoPol buffer, 200 µM dNTPs, 4 % (v/v) Dimethyl sulfoxide (DSMO) and 1 µL extracted genomic DNA (10 ng). Reactions were made up to 50 µL with PCR grade water and all concentrations stated above represent final concentrations. The following cycling profile was used: initial denaturing at 94˚C for 2 min, 34 cycles at 94˚C for 30 sec, annealing at 55˚C for 45 sec followed by a 2 min 30 sec extension at 72˚C. A single cycle of extension was performed for 5 min at 72˚C followed by cooling to 4˚C. An amplicon size of 2 466 bp indicated an intact CYP450 gene whilst an amplicon size of 1266 bp was indicative of a successful CYP450 deletion.

Strains that displayed a CYP450 deletion genotype based on the first round of PCR were then subjected to a second round of PCR to confirm that the deletion occurred at the correct chromosomal locus. This second PCR was performed with an additional oligonucleotide (Int-F), that was designed based on the putative methyltransferase gene sequence – a part of the chromosome that was not genetically manipulated *in vitro*, and the DS-Bam-R oligonucleotide.

The PCR screen with Int-F and DS_Bam-R was performed with 3.75 U of Expand Long Template polymerase. The cycle profile was identical to that described in Table 3.5 except that annealing was performed at 61˚C and elongation was performed for 2 min 30 sec. Figure 3.4 A and B illustrates the positions of the oligonucleotides relative to the either intact or deleted *cyc* (CYP450) gene.

(A)



(B)

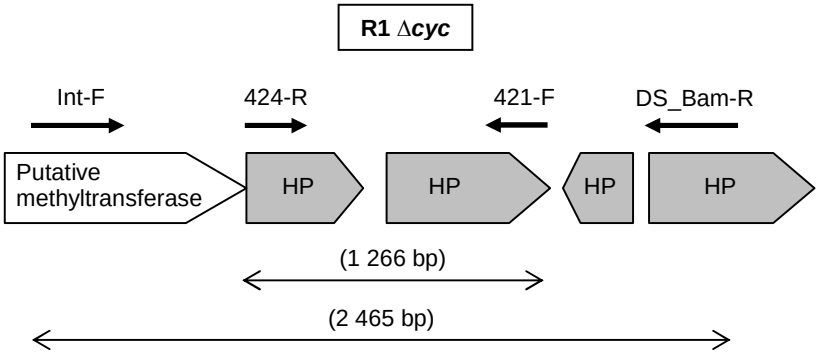


Fig. 3.4 **(A)** Gene topology of wildtype *H. salinarum* R1. **(B)** Gene topology of *H. salinarum* R1 Δ*cyc*. Expected amplicon sizes from utilizing each oligonucleotide pair are indicated in parentheses. Figure legend: **HP** = hypothetical protein and ***cyc*** = cytochrome P450 monooxygenase.

3.2.11 Growth and pigment extraction of wildtype and Δ*CYP174A1* strains of *H. salinarum* R1

Wildytype and Δ*CYP174A1* strains of *H*. *salinarum* R1 were cultured as described in section 3.2.9 in 100 mL Erleynmeyr flasks containing 35 mL complete medium. at 40˚C, shaken at 200 r.p.m. for 96 h. Growth of both the wildtype and deletion strains was followed spectrophotometrically at 600 nm using a SpectraMax M2 (Molecular Devices). Cultures concurrently grown with cultures for the growth experiment were used for wet weight determinations. Five milliliter samples were removed from both sets of cultures after 86 h and 96 h growth and biomass was harvested by centrifugation at 17 000 x *g* and ambient temperature for 10 min and their supernatants decanted. As much of the residual supernatant was removed with absorbent tissue paper before the tubes were weighed.

Pigments were extracted from the biomass used for the wet weight determinations by adding 2 mL ice cold acetone and shaking the tubes at an angle of 45˚ for 1 h at 4˚C. Extractions were performed until pellets appeared white. After each extraction the pellet was collected by centrifuging the mixture at maximum speed for 10 min and the red supernatant removed and transferred to a 15 mL Falcon tube. From the pooled supernatants, 1 mL was transferred to a microfuge tube, dried under vacuum at 30˚C and resuspended in 0.3 mL fresh acetone. The concentrated extracted pigments were subjected to a wavelength scan (200 nm – 750 nm, 2 nm intervals) in 96-well UV microtitre plates using the SpectraMax M2. Fresh acetone served as a blank.

3.2.12 Microarray analyses of the *H. salinarum* R1 Δ*CYP174A1* transcriptome

Wildtype and Δ*CYP174A1* strains of *H*. *salinarum* R1 were cultured in complete medium in triplicate at 40˚C (see section 3.2.9). Samples for total RNA extraction were taken at the late logarithmic- and late stationary phases of growth which corresponded to OD_{600} of *ca*. 0.3 and 1.3 respectively. Total RNA extraction, cDNA synthesis, *Cy*3-dCTP and *Cy*5-dCTP labeling of the cDNA and computational analyses of results was performed as described by Coker *et al*. (2007) except for the following: extracted RNA for each

strain was standardized to 6 µg, labeled cDNA representing the two growth phases were hybridized on a single Agilent slide containing 9 956 replicated genes from *Halobacterium* sp. NRC-1 and hybridizations were performed at 65˚C for 15 h. (*Note: The complete method for the RNA extraction, labeling and hybridization has been condensed for this thesis. Copy right to this method resides with Prof. S. DasSarma, University of Maryland, Baltimore, USA)*.

3.2.13 Purple membrane isolation using a sucrose gradient

Wildytype and Δ*CYP174A1* strains of *H*. *salinarum* R1 were cultured in 1 L complete medium (in 5 L Erlenmeyer flasks) at 40˚C, agitated at 200 r.p.m. until an OD_{600} = 1.2 was reached (this corresponds to late stationary phase). Cells were harvested at 6 000 x *g* for 10 min at 4˚C and the resulting pellets were resuspended in 5 mL basal salts medium (complete medium without peptone). The cell lysates were placed in pre-soaked dialysis tubing with a molecular weight cut off (MWCO) of 10 kDa (SnakeSkin® Pleated Dialysis Tubing, Thermo Scientific) and dialyzed against 5 L pure water at 4˚C for at least 15 h with two water changes (after t = 2 h and t = 4 h and lastly for overnight dialysis).

The cell paste was transferred to a test tube and 50 µL of 10 mg/mL DNaseI was added. The tubes were incubated at 37˚C for 1 h with agitation at 180 r.p.m. until the cell lysate was completely watery and not viscous. A sucrose gradient was prepared with filter sterilized sucrose (dissolved in water) by adding it to the bottom of an ultracentrifuge tube (38.5 mL capacity) in the following order: 11 mL 30 % (w/v) sucrose, 11 ml 40 % (w/v) sucrose, 11 mL 50 % (w/v) sucrose and 2 mL 60 % (w/v) sucrose. The DNaseI digested cell lysates were gently layered onto the gradients and placed in a balanced SW32 Ti rotor. Tubes were spun at 132 000 x *g* for 17 h at 18˚C in a Beckman Coulter Optima™ L-100 Ultracentrifuge (S. DasSarma. 1984. Ph.D. thesis, MIT, Cambridge, USA).

3.2.14 PCR screening for insertion elements in the *bop* cluster

Genomic DNA (as template) from *Halobacterium* sp. NRC-1 as well as oligonucleotides for the amplification of the bacterioopsin (*bop*), bacterioopsin related protein (*brp*) and bacterioopsin gene activator (*bat*) open reading frames (ORFs) were generous gifts from Prof. Shiladitya DasSarma (University of Maryland, Maryland, USA). Genomic DNA was isolated from wildtype and Δ*CYP174A1 H. salinarum* R1 strains as described in section 3.2.8 and PCR amplification was performed as described in section 3.3.10. Oligonucleotide sets used in this section are described in Table 3.3. An increase in the size of amplicons is usually indicative of spontaneous insertions in the ORFs of *bat* and *brp* resulting in *bop* transcription abolishment and consequent cessation of purple membrane production.

3.3 Results

3.3.1 Isolation of a saltpan isolate, identification and cloning of a CYP450 gene

A single pure isolate of an extremely halophilic archaea was isolated from salt crystals from the Lemoenkloof commercial saltern. Amplification of a 16S-rRNA gene from this isolate with a set of universal archaeal oligonucleotides (20bF and 1492R; Rincón *et al*., 2006) and subsequent sequencing and analyses confirmed that the isolate belonged to the *Archaea* kingdom. However, sequence analyses of the particular 16S-rRNA amplicon were not sufficient to distinguish between genera of *Halobacterium* and *Haloarcula*. Consequently, a set of oligonucleotides was designed (Table 3.3) based on a variable region (701 bp in size) of the *rrlA* gene (23S-rRNA) of *Halobacterium* sp. NRC-1 (*rrlA;* NCBI Gene ID: 1448872), and *Haloarcula marismortui* ATTC 43049 (*rrlA-1*; NCBI Gene ID: 3128671) (Grosjean *et al*., 2008). MegaBLAST analyses revealed that the 23S rRNA gene from the saltpan isolate displayed maximum identity of 98% towards the *rrlA-1* gene from *H*. *marismortui* ATCC 43049 and 83% maximum identity to the *rrlA* genes of both *Halobacterium* sp. NRC-1 and *H*. *salinarum* R1. Multiple alignment (using ClustalW2) of the 701 bp 23S-rRNA sequence from the saltpan isolate with the 23S rRNA genes of *Halobacterium* sp. NRC-1, *H*. *salinarum* R1 and *H*. *marismortui* ATCC 43049 supported the MegaBLAST results. These results indicated that the saltpan isolate probably belonged to the *Haloarcula* genus. For the purpose of this study the saltpan isolate will be referred to as *Haloarcula* sp. LK-1.

The genome of *H*. *marismortui* ATCC 43049 (Baliga *et al*., 2004) contains three CYP450s from three distinct families (CYP174, CYP1001 and CYP1003). Oligonucleotide set SP_F1 and SP_R1 (Table 3.3) was designed based on multiple alignments of these three genes with CYP174A1 from *Halobacterium salinarum*. Oligonucleotide SP_F1 was based on the conserved I-helix motif while SP_R1 was based on the conserved heme-binding loop. A PCR using whole cells of the *Haloarcula* sp. LK-1 strain produced a 412 bp amplicon. Subsequent DNA sequencing and analyses revealed that a CYP450 from the CYP174-family had been amplified. Oligonucleotide set HM_F1 and HM_R1 (Table 3.3) based on the sequence of CYP174A2 from *H.*

marismortui (Genbank accession: YP 134859.1) was used to amplify the entire ORF of the CYP174-like gene which was 1 338 bp in length. To our knowledge this is the very first case of an extremely halophilic arhaeal CYP450 isolated by PCR. This gene translated into a 445 amino acid protein. Figure 3.5 shows the DNA sequence as well as the corresponding amino acid sequence of the newly cloned CYP174A2. The gene sequence is available on NCBI with the following Genbank accession number: HM135515. The CYP174A2 from the saltpan isolate displayed 98% amino acid indentity to the CYP174A2 of *H*. *marismortui* ATCC 43049 which indicates that the CYP450 from *Haloarcula* sp. LK-1 is most probably an ortholog of *CYP174A2*.

```
  1 ATGTCAAAGA CGCCGCCCGG ACCGAAGGGC GAACCGTTGT TCGGCAGTAG
    M  S  K  T   P  P  G   P  K  G   E  P  L  F   G  S  S

 51 TCGCACGTAC GCTCGGGACC CGTTCCGGTT CATCTCGGCG CTGGAGCGGG
     R  T  Y   A  R  D  P   F  R  F   I  S  A   L  E  R  A

101 CCTACGGCGA CGTGGCCCGA TTCGACATGG GGCCGATGGA TACGGTCATG
      Y  G  D   V  A  R   F  D  M  G   P  M  D   T  V  M

151 CTCTGTGACC CGACAGCAAT CGAGCGCGTG CTGGTTTCGG AGGCCGACCA
    L  C  D  P   T  A  I   E  R  V   L  V  S  E   A  D  Q

201 GTTCCGCAAA CCCGACTTTC AGGGCGACGC GTTGGGGGAC CTGCTGGGTG
     F  R  K   P  D  F  Q   G  D  A   L  G  D   L  L  G  D

251 ACGGGCTGCT GTTGAGCGAA GGCGAGACCT GGGAGCAGCA GCGAAAGCTC
      G  L  L   L  S  E   G  E  T  W   E  Q  Q   R  K  L

301 GCGAACCCCG CGTTCTCGAT GGCCCGACTG TCGGGGATGG CTGACCGGAT
    A  N  P  A   F  S  M   A  R  L   S  G  M  A   D  R  I

351 CACCGGCCAC GCGAAGGACC GCATCGCCGA CTGGTCCCAC GGCGATGTCA
     T  G  H   A  K  D  R   I  A  D   W  S  H   G  D  V  I

401 TTGACGCCGA GCAGTCGATG ACCCGGGTCA CGCTGGACGT GATTCTGGAC
      D  A  E   Q  S  M   T  R  V  T   L  D  V   I  L  D

451 CTGATGATGG GTGTCGAACT CTCTGAGCAG CGAGTCCAGA CCATCGAGGA
    L  M  M  G   V  E  L   S  E  Q   R  V  Q  T   I  E  E

501 GCAACTGCTG CCGCTGGGTC AACGGTTCGA ACCGGACCCC ATCCGGTTCG
     Q  L  L   P  L  G  Q   R  F  E   P  D  P   I  R  F  A

551 CCATGCCACA GTGGATGCCG ATGCCGGACG ACGCCGAGTT CAACCGCGCT
      M  P  Q   W  M  P   M  P  D  D   A  E  F   N  R  A

601 GTGCGGACGC TTGACGAGGT ACTGGACGAC ATTATCGAGG TCCGTGAGGA
    V  R  T  L   D  E  V   L  D  D   I  I  E  V   R  E  D
```

```
651 CTCGGTCGGG TCCGGTGACG ACGGCCCGAT GGACTTCCTG TCGGTGCTCC
     S  V  G   S  G  D  D   G  P  M   D  F  L   S  V  L  L

701 TGCGTGCCCG CGACGAGGGG AATCAGTCGC CCGAGCAACT GCGCGATGAA
      R  A  R   D  E  G   N  Q  S  P   E  Q  L   R  D  E
                          I-Helix motif
751 ATGATGACGA TGCTGCTTGC GGGCCACGAC ACGACGGCAC TGACGCTGAC
    M  M  T  M   L  L  A   G  H  D   T  T  A  L   T  L  T

801 CTACACCTGG TTCCTACTGT CGGAACACCC CGAAGTCGAA CAGCGAGTCC
     Y  T  W   F  L  L  S   E  H  P   E  V  E   Q  R  V  H

851 ACGAGGAACT GGATGATGTC ATCGGCGACG ACCGGCCGGG GATGGAACAC
      E  E  L   D  D  V   I  G  D  D   R  P  G   M  E  H
                                        K-Helix motif
901 GTCCGCGAGT TGGACTACCT CGAATGGGTG ATTCAGGAAG CGATGCGGCT
    V  R  E  L   D  Y  L   E  W  V   I  Q  E  A   M  R  L
951 CTACCCGCCC GTATACACTA TCTTCCGGGA GCCGACAGAG GATGTGACGC
     Y  P  P   V  Y  T  I   F  R  E   P  T  E   D  V  T  L

1001 TATCGGGATA TGATGTCGAG GCCGGAACGA CGTTGATGGT CCCACAGTGG
      S  G  Y   D  V  E   A  G  T  T   L  M  V   P  Q  W

1051 GGTGTTCACC GCTCCGAGCG GTTCTACGAT GACCCCGAGA CGTTCGACCC
     G  V  H  R   S  E  R   F  Y  D   D  P  E  T   F  D  P

1101 GGAGCGCTGG AAGCCCGAGC GAGCGAACGA GCGGCCCCGG TTCGCCTACT
      E  R  W   K  P  E  R   A  N  E   R  P  R   F  A  Y  F
                Heme binding loop motif
1151 TCCCGTTCGG TGGTGGCCCG CGCCACTGCA TCGGGAAGCA CCTCGCCATG
      P  F  G   G  G  P   R  H  C  I   G  K  H   L  A  M

1201 CTTGAAGCAC AACTCATCAC TGCAACGACG GCCAGCCAGT ACCGACTGGA
     L  E  A  Q   L  I  T   A  T  T   A  S  Q  Y   R  L  E

1251 GTTCCAGGGA GAGACGCCAC TGGAGCTGTT ACCGTCGTTG ACCGCCCATC
      F  Q  G   E  T  P  L   E  L  L   P  S  L   T  A  H  P

1301 CCCGGCAGAA AATGTCCATG CGCGTGCAGG AACGGTAA
      R  Q  K   M  S  M   R  V  Q  E   R
```

Fig. 3.5 Saltpan CYP450 DNA and amino acid sequence. The 1338 bp gene translates into a 445 amino acid protein with a theoretical molecular mass of 50.7 kDa and pI of 4.57 (as predicted by the pI/Mw tool on the ExPASy Proteomics server). Highly conserved CYP450 amino acid motifs are highlighted in grey on the figure.

3.3.2 Attempts at heterologous expression of the *CYP174A2* ortholog

To date no CYP450s from extremely halophilic archaea have been cloned and expressed heterologously. Since the physiological role of these CYP450s were also unknown it was thought best to first express and purify CYP174A2 in order to characterize and possibly elucidate its function. Successful heterologous expression of several genes from *H*. *salinarum* (Ishibashi *et al*., 2001), *Haloferax volcanii* (Connaris *et al*., 1999), *Haloferax mediterranei* (Pire *et al*., 2001) and *Halomonas* sp. #593 (Yonezawa *et al*., 2003) in *E*. *coli* have been reported in the past. These expression studies were encouraging and they suggested that heterologous expression of CYP450s from halophilic archaea in *E*. *coli* might be achievable.

3.3.2.1 Heterologous expression of *CYP174A2* in *E. coli* BL21 (DE3)

The CYP174A2 from *Haloarcula* sp. LK-1 was cloned into the pET28b(+) expression vector and expression was induced with various IPTG concentrations at various temperatures in *E*. *coli* BL21 (DE3) (section 3.2.7.2). Online topology and primary structure analyses prediction software on the ExPASy proteomics server *e.g*. PSORTdb and ProtScale (using Kyte and Doolittle hydrophobicity plots) indicated that the CYP450 protein would be localized in the cytoplasm. Unfortunately, none of the expressed protein could be detected in the soluble fraction. Figure 3.6 illustrates a typical SDS-PAGE obtained, loaded with both soluble as well as the insoluble fractions.

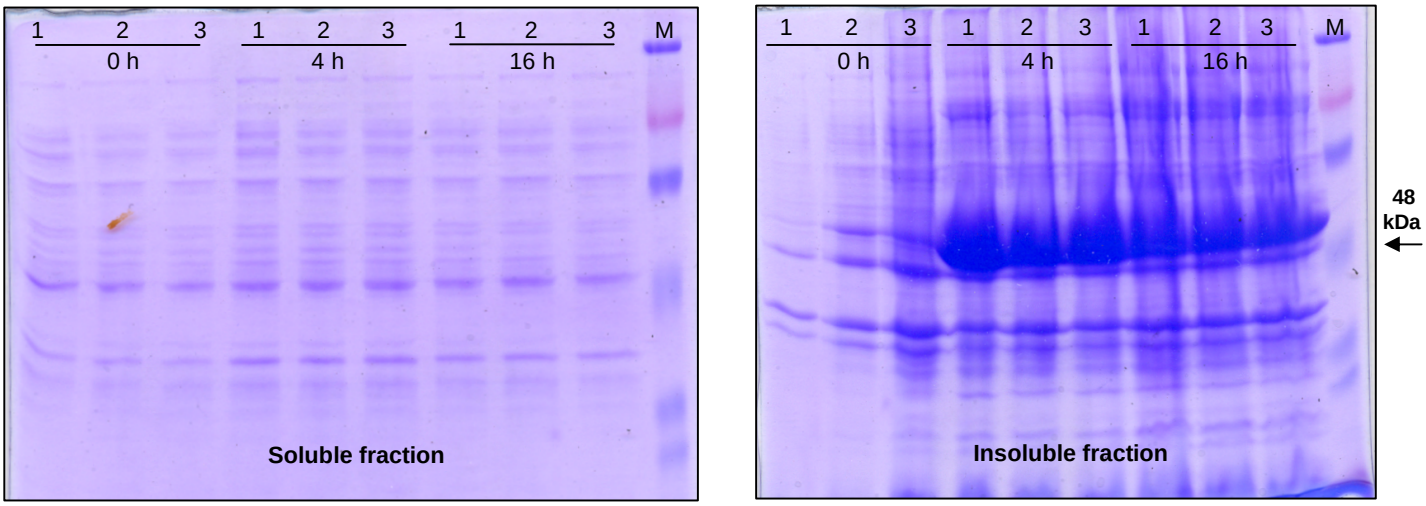


Fig. 3.6 10 % SDS-PAGE loaded with crude protein extract from *E. coli* BL21 (DE3) expressing the CYP450 from *Haloarcula* sp. LK-1 using pET28b (+). Induction was performed with 1 mM IPTG at 35˚C for 16 h. Lanes: **1 - 3** = independent recombinant clones; **M** = 3-Color Prestained Molecular Weight Marker. Samples were taken at 0, 4 and 16 h for analyses. An expressed CYP450 protein with a theoretical molecular mass of 50.7 kDa was expected.

3.3.2.2 Heterologous expression of CYP174A2 in *Pseudomonas fluorescens* KOB2Δ1

Pseudomonas fluorescens KOB2Δ1 has an *alkB* deletion. The *alkB* gene is an alkane hydroxylase and forms part of the *alkBFGHJKL* operon, which encodes proteins involved in n-alkane degradation (van Beilen *et al*., 1994). An *alkB* deletion consequently renders *P. fluorescens* KOB2Δ1 incabable of hydroxylating and in so doing degrading *n*-alkanes. To restore the *n*-alkane degrading phenotype in *P. fluorescens* KOB2Δ1, an alkane-responsive expression vector (pCOM8) was developed that enables the researcher to clone the *alkB* gene or any other similar genes encoding prototeins with a *n*-alkane hydroxylating function (Smits *et al*., 2001; Smits *et al*., 2002).

The pCOM8 system appeared useful since there have been reports of halophilic archaea that are capable of degrading *n*-alkanes (Bertrand *et al*., 1990; Kulichevskaya *et*

al., 1991; Zvyagintseva *et al*., 1995; Raghavan and Furtado, 2000). Raghavan & Furtado (2000) even suggested that since halphilic archaea possess characteristic C_{20} and C_{40} isoprenoid lipids, it was possible that the utilization of hydrocarbons was an inherent trait. However, none of these above-mentioned publications made mention of the fact that a CYP450 could be involved in the degradation of *n*-alkanes. If CYP174A2 was indeed capable of hydroxylating *n*-alkanes, the pCOM8 system in *P*. *fluorescens* KOB2Δ1 offered a means for testing this hypothesis.

A pCOM8:*CYP174A2* construct was created and transformed into *P*. *fluorescens* KOB2Δ1 by electroporation but none of the recombinant strains produced any detectable CYP174A2 (as assessed from SDS-PAGE) upon induction with DCPK and transformants were also not able to grow on *n*-alkanes (results not shown).

3.3.3 Creating a Δ*CYP174A1* in *H. salinarum* R1

To study the CYP174A2 from *Haloarcula* sp. LK-1 by means of heterologous expression and *in vitro* testing became a non-viable option. Instead we opted for creating a CYP450 knock-out mutant in *H*. *salinarum* R1. The rationale for selecting *Halobacterium salinarum* R1 was as follows: Firstly, *H. salinarum* R1 and *Halobacterium* sp. NRC-1 both possess only one CYP450 gene, *i.e*. CYP174A1, while most of the halobacteria with putative CYP450s belonging to the CYP174 family possess several CYP450s from other CYP450-families in their genomes *e.g*. *Haloarcula marismortui* has a CYP174A2, CYP1001A2 and a CYP1003A1. This would surely complicate the CYP450 knock-out strategy. Secondly, *Halobacterium* sp. NRC-1 and *H*. *salinarum* R1 are genetically accessible. Several excellent gene knock-out systems are available for *Halobacterium* sp. NRC-1 (Peck *et al*., 2000) and *H*. *salinarum* R1 (Koch & Oesterhelt, 2005), the genomes of both strains were fully sequenced and annotated (Ng *et al*., 2000; Pfeiffer *et al*., 2008) and finally microarray platforms are available for both strains (Müller & DasSarma, 2005; Coker *et al*., 2007; Twellmeyer *et al*., 2007). It was decided to create the *CYP174A1* deletion with the aid of the pMKK100 suicide vector which confers mevinolin resistance and permits blue/red selection on plates containing X-gal (Koch & Oesterhelt, 2005). It was envisaged that the knock-out strain would be useful for heterologous expression of CYP450s from halophilic *Archaea* and to evaluate the effect

of the *CYP174A1* deletion on phenotype as well as global gene expression with microarray analyses.

3.3.3.1 Creating a *CYP174A1* deletion cassette in pMKK100

One kilobase fragments of US and DS regions (adjacent directly up- and down stream of the *CYP174A1* gene) were amplified from the genome of *H*. *salinarum* R1. The individual amplicons were ligated to each other with *Pst*I cohesive ends to yield a *ca*. 2 kb cassette with *Bam*HI and *Hin*dIII recognition sites engineered on the 5' and 3' ends of the cassette respectively to facilitate unidirectional cloning into the pMKK100 suicide vector. Figure 3.7 shows the liberated US/DS-deletion cassette after endonuclease digestion with *Hin*dIII and *Bam*HI as well as the vector map.

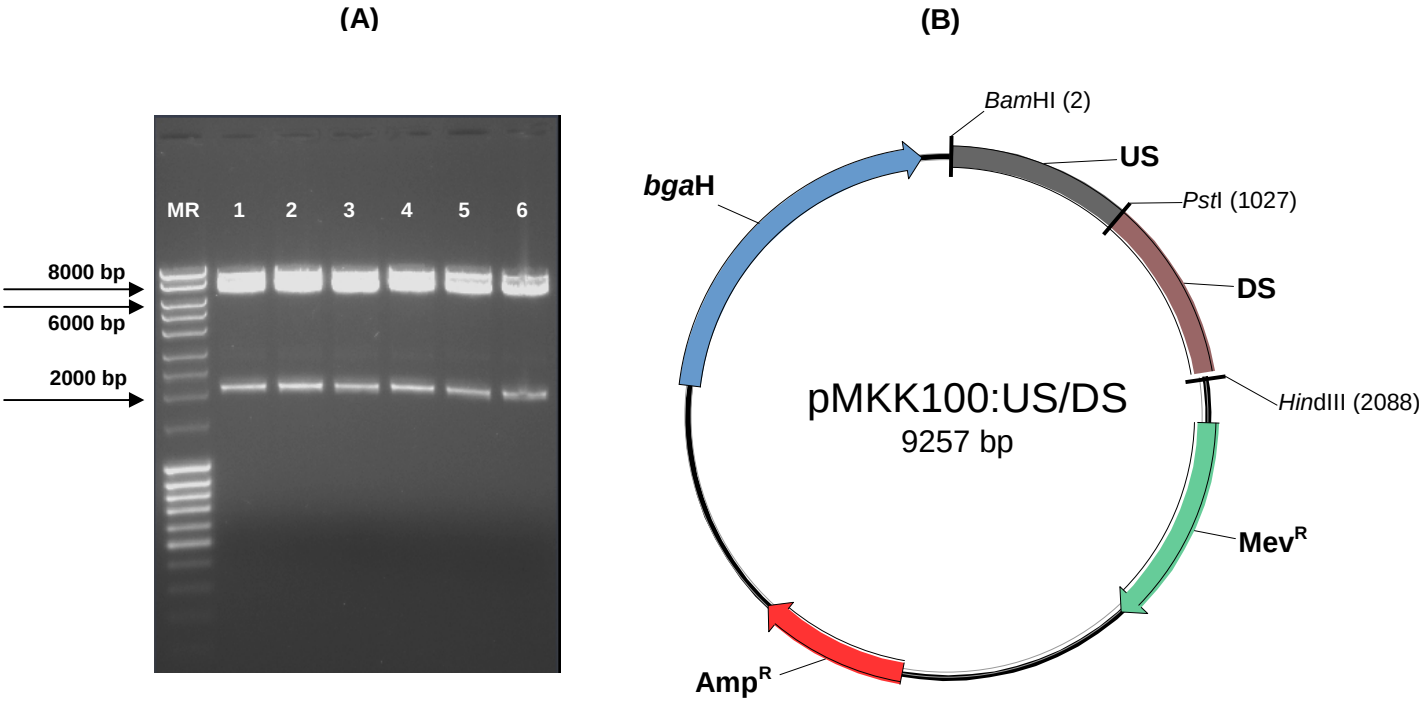


Fig. 3.7 Confirmation that the US/DS casette was inserted into pMKK100. **(A)** Double digestion of pMKK100:US/DS with 5 U *Bam*HI and 10 U *Hin*dIII endonucleases at 37˚C for 1 h. Lanes: **MR** = 5 µL MassRuler (Fermentas) and **1 - 6** = *ca*. 2 kb US/DS deletion cassette liberated from pMKK100 (remaining fragment *ca.* 7.2 kb) . **(B)** Vector map of pMKK100:US/DS containing ampicillin (Amp^R)- and mevinolin (Mev^R) resistance markers as well as the halophilic β-galactosidase gene (*bgaH*). Transcription of the Mev^R and *bgaH* genes is driven by their native promoters.

3.3.3.2 Identifying Δ*CYP174A1* strains by PCR screening

Successful transformation of *H. salinarum* R1 with pMKK100:US/DS and subsequent plating onto X-gal and mevinolin containing medium yielded blue colonies. Blue colonies were cultured in complete medium without mevinolin several times to induce a cross-over event (see Fig. 3.3). After the crossover event and plating of cultures, colonies did not display blue coloration (due to the loss of the suicide vector harboring the *bgaH* gene) but instead remained red. In total, five blue colonies were selected from two independent deletion experiments (one from a first experiment and four from a second experiment) and their subsequent red pigmented progeny inoculated into complete

medium. Genomic DNA was extracted from these cultures and used as templates for a PCR screening with oligonucleotides 421_F and 424_R (Table 3.3) to ascertain whether the *CYP174A1* gene was successfully deleted (see section 3.2.10 Figures 3.4 A and B). Figure 3.8 illustrates typical results from PCR screening for *CYP174A1*-deletion

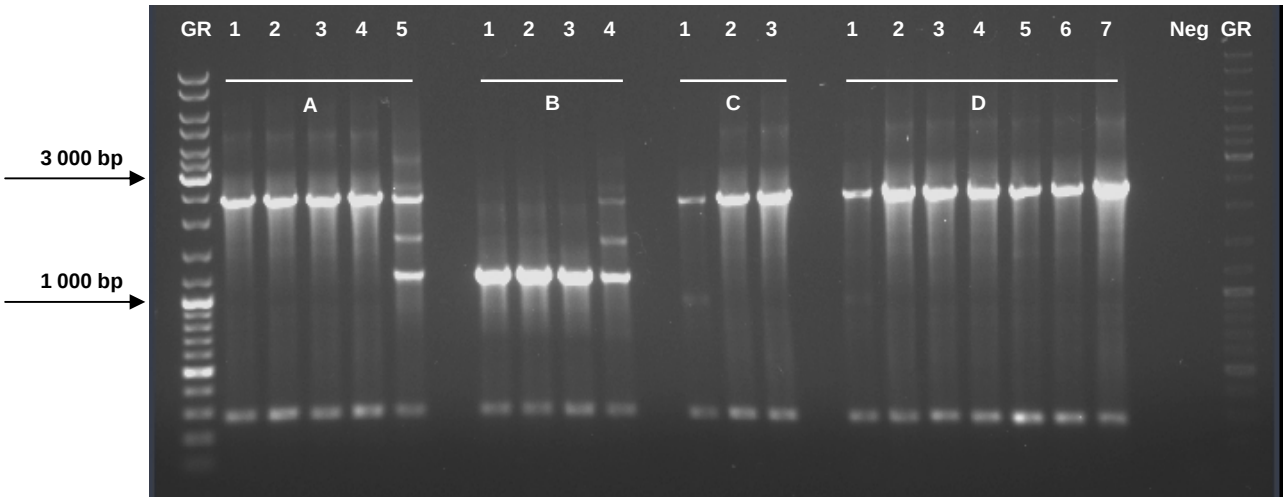


Fig. 3.8 PCR screening results to identify clones that are deficient of the *CYP174A1* gene. **A - D** represents red progeny from four blue colonies. PCR screened red clones from **A**, **C** and **D** displayed the wildtype genotype (2 446 bp amplicon expected) and red clones from **B** displayed the deletion genotype (1 266 bp amplicon expected). Lanes: **GR** = 5 µL GeneRuler (Fermentas) and **Neg** = negative control.

3.3.3.3 Growth experiments and pigment extraction of the Δ*CYP174A1* and wildtype strains

The clones that displayed the Δ*CYP174A1* genotype (Fig. 3.8) as well as the wildtype strain were cultured in complete medium (section 3.2.1.2) at 40˚C with agitation at 200 r.p.m. for 96 h to see if a growth related phenotype could be observed due to the *CYP174A1* deletion. Growth was monitored spectrophotometrically at OD_{600} (Fig. 3.9 A). Samples were also taken for wet weight determinations (Fig. 3.9 B). Cultures were also streaked onto solid media and cultured at 40˚C for five days. We noticed a difference in pigmentation in the liquid cultures (Fig. 3.9 C) after *ca*. 40 h of growth. Differences in

pigmentation between the deletion and wildtype strain on solid medium became evident after five days (Fig. 3.9 D).

The difference in pigmentation between the wildtype and Δ*CYP174A1* strains, prompted us to extract the acetone soluble pigments from the biomass (El-Sayed *et al*., 2002) taken at 86 h and 96 h. The acetone extracted red pigments were subjected to a UV/Vis wavelength scan (200 nm – 750 nm) and produced a characteristic so-called 'three finger' shape that is typical of C_{50} carotenoids (Fig. 3.10 A).

All the deletion strains always generated more biomass when compared to the wildtype strain (Fig. 3.9 A and B). Thus when the acetone extracted pigments were analyzed for both strains it was initially thought that the increased pigment content in the deletion strain (Fig. 3.10 A) was a function of biomass. However, after normalizing the biomass (wet weight) of both strains and re-extracting the pigments with acetone, the same trend was observed with regards to pigment content in both the wildtype and deletion strains.

Although the pigment content in the deletion strain was always higher as compared to the wildtype strain, the amount of pigment extracted from each individual deletion strain was not always consistent between experiments.

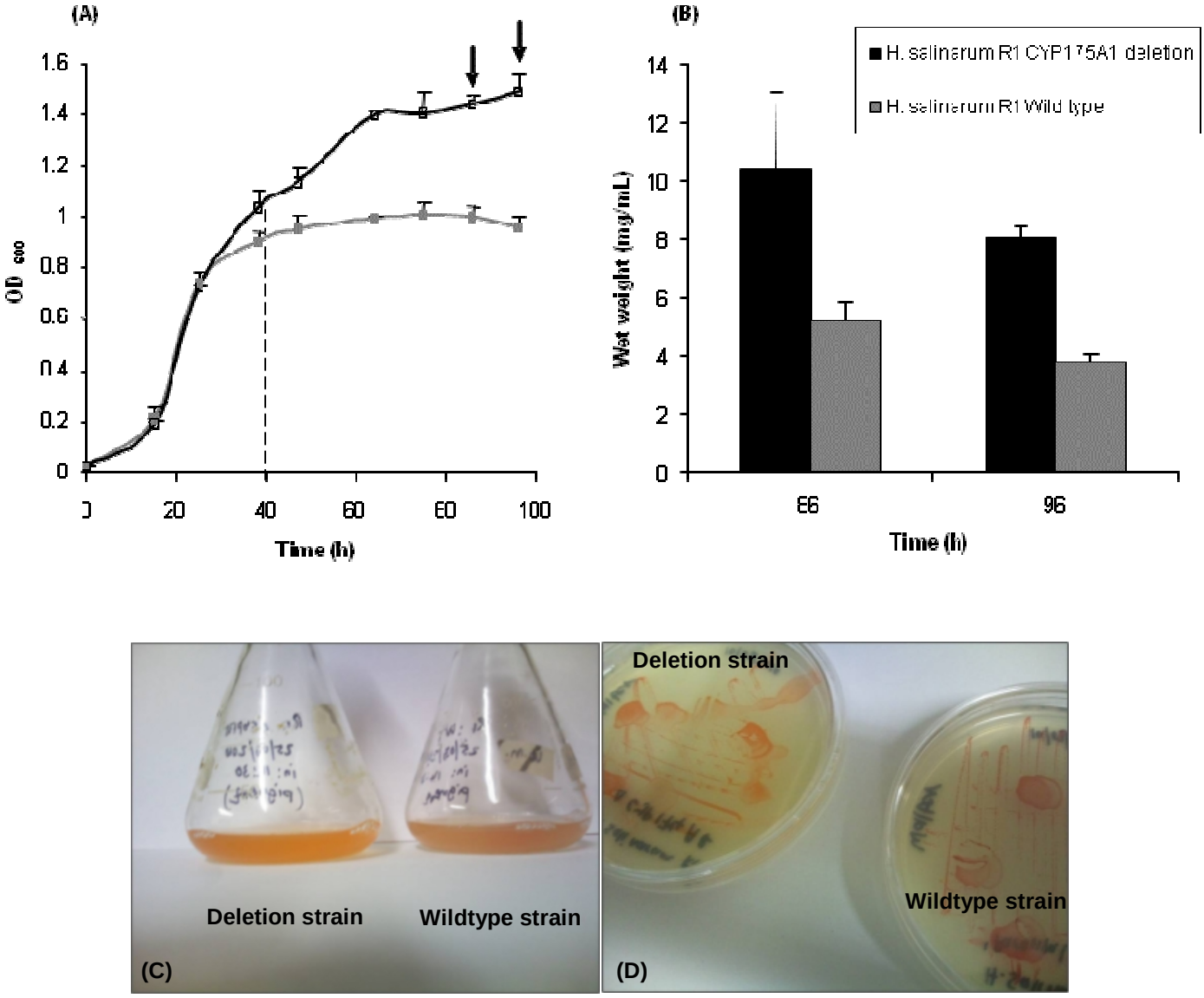


Fig. 3.9 **(A)** Growth curves of wildtype and Δ*CYP174A1* strains of *H. salinarum* R1. Change in pigmentation was observed at *ca*. 40 h (dashed line). **(B)** Wet weight determinations of cultures taken at 86 h and 96 h (indicated by arrows on growth curve). Wildtype and Δ*CYP174A1* data are depicted in grey and black respectively. Growth of the wildtype and Δ*CYP174A1* strains of *H. salinarum* R1 on **(C)** liquid culture and **(D)** solid medium. Strains were cultured aerobically at 40˚C for at least 5 days. Error bars are representative of three measurements.

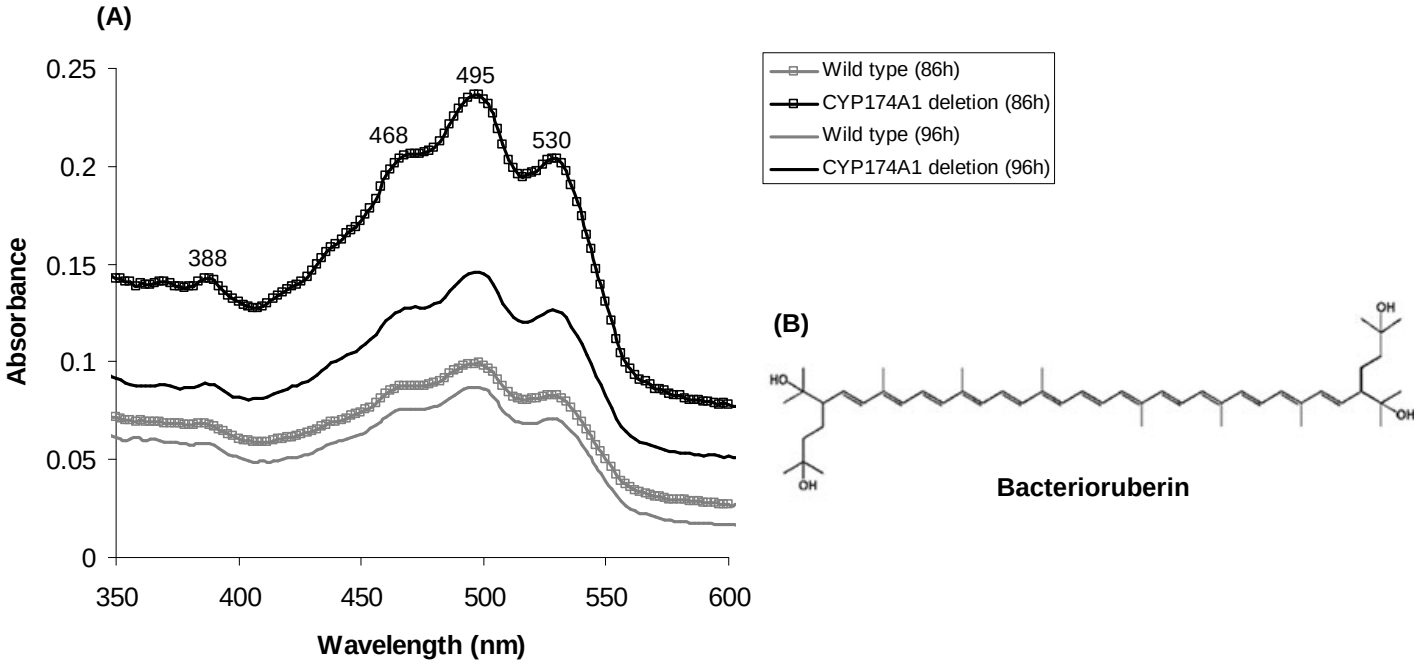


Fig. 3.10 **(A)** Wavelength scan of acetone soluble pigments extracted from cell pellets harvested at 86 h and 96 h of growth from wildtype and Δ*CYP174A1* strains of *H*. *salinarum* R1. Absorption maxima of the pigment in acetone are indicated on the spectra. **(B)** Chemical structure of bacterioruberin.

The spectra of the acetone extracted pigments (Fig. 3.10A) were typical spectra for bacterioruberin-like carotenoids (Fig. 3.10 B) (D'Souza *et al*., 1997; Fang *et al*., 2010). These results indicated that the deletion of *CYP174A1* apparenly had an effect on carotenoid metabolism in *H*. *salinarum* R1 causing accumulation of these pigments. This gave a clue as to which metabolic pathways might be affected and which genes to focus on when microarray experiments were concluded.

3.3.4 Transcriptomic analyses of *H. salinarum* R1 Δ*CYP174A1*

A microarray experiment was performed to assess the impact of the Δ*CYP174A1* on the transcriptome of *H*. *salinarum* R1. It was performed with cDNA synthesized from *H*. *salinarum* R1 wildtype and Δ*CYP174A1* strains. This cDNA was hybridized with DNA originating from *Halobacterium* sp. NRC-1. Some considerations need to be taken into account when working with genetic material from strains R1 and NRC-1.

Firstly, the chromosomes of both strains are almost identical in size – with the chromosome of *Halobacterium* sp. NRC-1 being slightly larger (13.2 kb). Comparison of the two chromosomes revealed complete co-linearity and nearly identical DNA sequences (Ng *et al*., 2000; Pfeiffer *et al*., 2008). Aside from differences related to insertional elements (IS elements), there are only 12 differences on the chromosome of *H*. *salinarum* R1 when compared to *Halobacterium* sp. NRC-1: four point mutations five single-base frameshifts and three insertion/deletion events.

However, *Halobacterium* sp. NRC-1 harbours two plasmids namely: pNRC100 and pNRC200, while *H*. *salinarum* R1 has four plasmids: pHS1, pHS2, pHS3 and pHS4. More than 350 kb of plasmid sequence among the two strains can be matched and are virtually identical at DNA sequence level. The plasmids also contain additional sequences that cannot be matched and is restricted to 4.5 kb in strain NRC-1 and to 210 kb in strain R1. Although a large portion of the plasmid sequences are virtually identical among the two strains there is, overall, highly different plasmid architecture. This variable plasmid architecture is a result of short co-linear regions with frequent colinearity breakpoints. Sequence duplications are also detectable in the plasmids. In the NRC-1 strain a large portion of the pNRC100 plasmid is duplicated in the pNRC200 plasmid and in the R1 strain pHS1, pHS2 and pHS4 plasmids are all related to each other due to large-scale duplications (Figure 3.11).

Since total RNA was extracted from *H*. *salinarum* R1 and reverse transcribed to yield cDNA that was hybridized with DNA from *Halobacterium* sp. NRC-1, the above-mentioned variation in plasmid region architecture will have consequences for the microarray analyses. Most notably, the genes located in regions P, S, V and W, found exclusively on plasmids pHS2 and pHS3 from *H*. *salinarum* R1, will for example not hybridize during the microarray process. If we assume that genes in these regions were significantly regulated, these genes would consequently not be detected during post microarray analyses. This implies that an incomplete picture will be obtained of the *H*. *salinarum* R1 transcriptome.

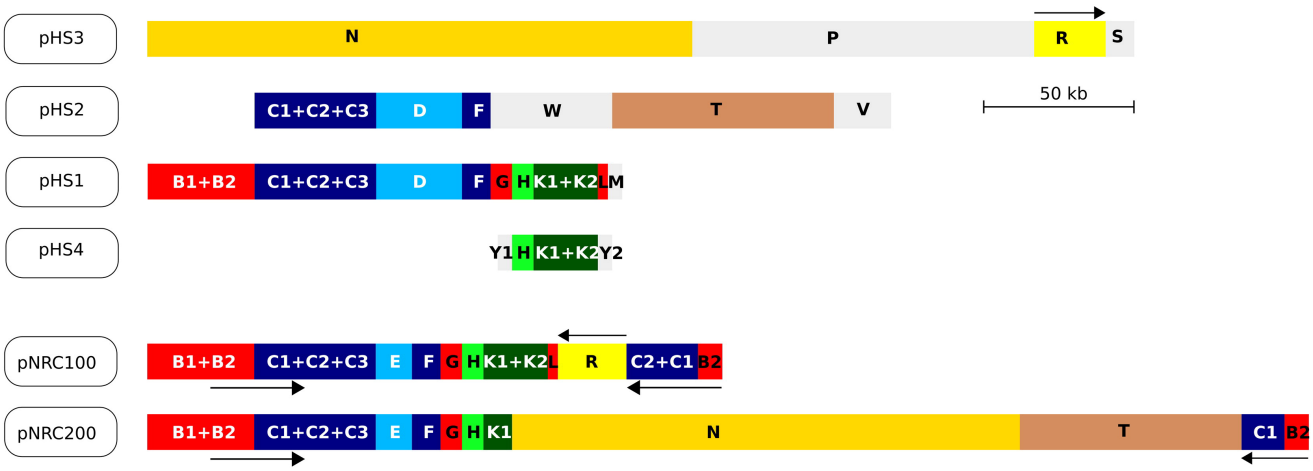


Fig. 3.11 Comparison of plasmid architecture between *Halobacterium* sp. NRC-1 (pNRC100 and pNRC200) and *H. salinarum* R1 (pHS1, pHS2, pHS3 and pHS4). Plasmid regions are depicted as coloured linear bars. Regions unique to strain R1 are indicated in grey (P, S, V and W). Colinearity between regions C and F is interrupted by strain specific alternative sequences D (19.3 kb) and E (4.5 kb). Duplicated regions that are inverted are indicated by arrows (Taken from: Pfeiffer *et al*., 2008.

Although this microarray study was performed with DNA from two non-identical strains, this is not the first study to have done so. In a study by Baliga *et al*. (2002), the authors compared the transcriptomes of *Halobacterium* sp. NRC-1 and *Halobacterium* sp. S9 as well as two mutants decending from the S9 strain in response to oxygen and light. Twellmeyer *et al.* (2008) evaluated the transcriptome of *H. salinarum* R1 also in response to light and oxygen and made mention of the fact that although Baliga and co-workers (2002) used a different approach and different strains, that a reasonable overlap of gene regulation in both the studies were visible.

Wildtype and Δ*CYP174A1* strains of *H. salinarum* R1 were cultured in triplicate in complete medium (section 3.3.1.2) at 40˚C and samples for total RNA extraction were taken at the late logarithmic- and stationary phases of growth (T1 and T2 in Fig. 3.12).

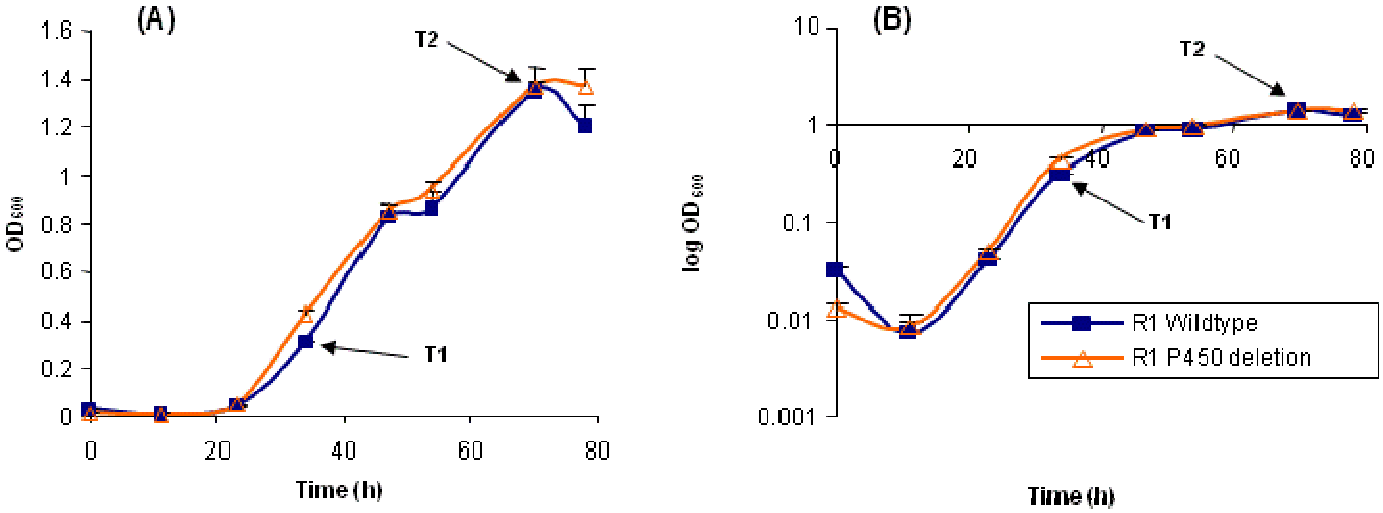


Fig. 3.12 Growth of *H. salinarum* R1 strains at 40˚C. Samples for total RNA extraction were taken at Time 1 **(T1)** and Time 2 **(T2)** which corresponds to the late logarithmic and stationary phases of growth respectively. Growth is represented on an arithmetic **(A)** as well as a logarithmic scale **(B)**. Error bars are representative of triplicate experiments.

The amount of total RNA extracted from each of the cultures at the various time points was normalized and cDNA synthesized. cDNA from the wildtype strain was labeled with the *Cy*-5 probe (red) whereas cDNA from the deletion strain was labeled with the *Cy*-3 probe (green). The fluorescent-labeled cDNA from the wildtype and the deletion strain was mixed in a 1:1 ratio and hybridized to an Agilent microarray slide containing 9 956 replicated genes from *Halobacterium* sp. NRC-1. Processed data from the two colour arrays for samples from the logarithmic and stationary phases of growth were plotted as $\log_2(x)$ ratio *vs*. gene number (Fig. 3.13). Genes that displayed a $\log_2(x)$ ratio of $-0.5 < x < 0.5$ (less than 1.5 fold change) were considered not to be differentially expressed. Genes that displayed $\log_2(x)$ ratios greater than +0.5 and smaller than -0.5 were tabulated (Tables A.1 – A.4) in the following gene groups: chromosomal genes, RNA genes, genes located on the pNCR100 plasmid and genes located on the pNRC200 plasmid. The tabulated data is available in Annexture A.

The microarray analyses of *H. salinarum* R1 Δ*CYP174A1* revealed that 487 and 255 genes were significantly regulated during the logarithmic and stationary phases of growth respectively. Gene names and corresponding predicted functions were assigned

by using the HaloWeb server and the annotated genome of *Halobacterium* sp. NRC-1 as reference (DasSarma *et al*., 2010). Almost 51% of the genes with significantly changed expression had no predicted function assigned to them. Interestingly this percentage is almost identical to what was reported in the study by Twellmeyer *et al*. (2008). Although microarray is a powerful tool to evaluate the regulatory dynamics of transcriptomes, our complete understanding of *Halobacterium* transcriptomes is lacking due to the high percentage of genes with unknown physiological function.

The microarray analyses confirmed that the *cyc* gene (*CYP174A1*) deletion was successful since its transcript levels were consistently reduced during both phases of growth with, the $\log_2(x)$ values almost indentical (logarithmic phase: -2.001 ± 0.656 and stationary phase: -2.263 ± 0.519). The deletion was accomplished by using a deletion cassette comprising *ca*. 1 kb sequences flanking *CYP174A1* (Fig. 3.2). To rule out the possibility that we disturbed genes within the flanking regions, which could have consequences on the transcriptomic profile, we compared the expression levels of six genes that are close neighbours of the *cyc* gene, to their regulation in wildtype *Halbacterium* sp. NRC-1 from a study conducted by Facciotti *et al*. (2010) (Table 3.6). Facciotti *et al*. (2010) found that the *cyc* gene as well as one other close neighbour (vng419) was upregulated during the transition from log to stationary phase, while three of the six close neighbours were down regulated. If deletion of the *cyc* gene had compromised the integrity of these genes one would have expected that wildtype genes upregulated during transition from log to stationary phase would have lower transcript levels in stationary phase cultures of the mutant while wildtype genes downregulated during transition from log to stationary phase would have lower transcript levels in log phase cultures of the mutant when compared to the wildtype strain.

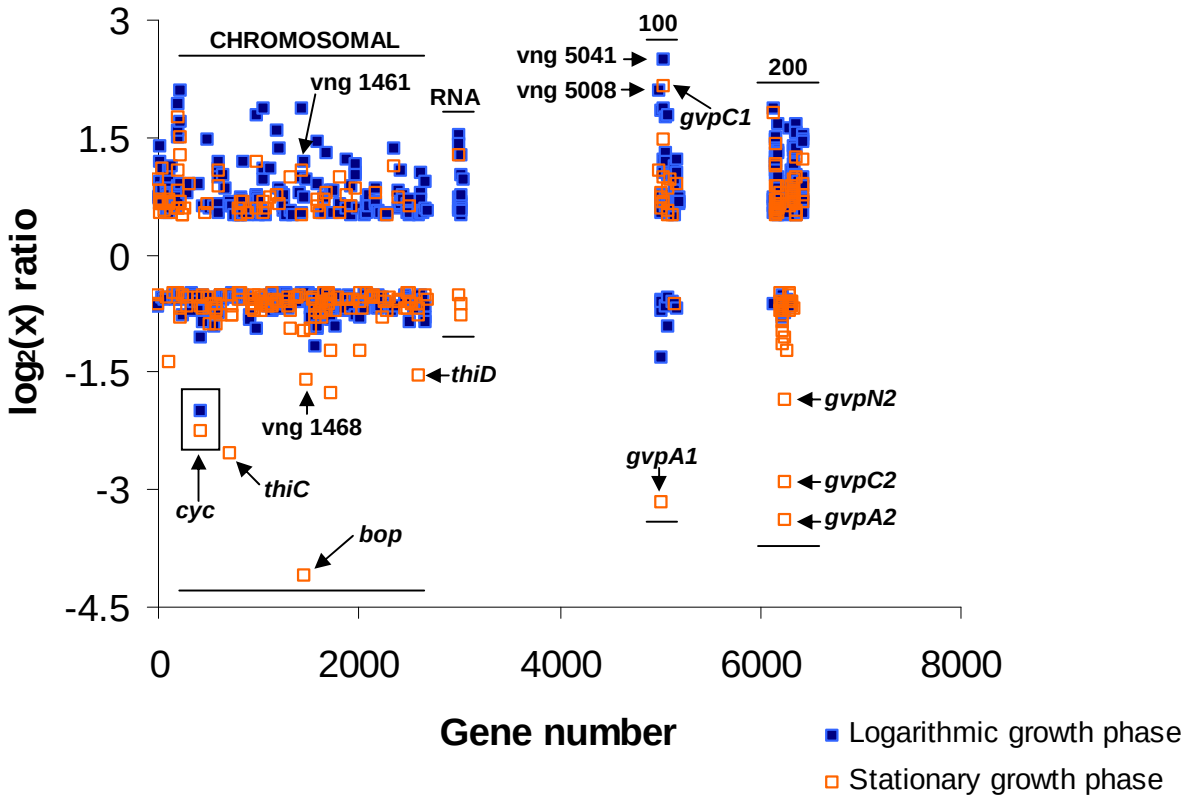


Fig. 3.13 Scatter plot comparing DNA microarrays of *Halobacterium* sp. NRC-1 hybridized with cDNA from the wildtype and CYP450 deletion strain of *Halobacterium salinarum* R1 cultures grown at 40˚C and sampled at late logarithmic and stationary phase. $Log_2(x)$ values of the *Cy*5/*Cy*3 ratio for each gene is plotted against the gene number. Gene numbers (from a NRC-1 point of view) correspond to: chromosomal genes (1 – 2679), RNA genes (3000 – 3051), genes on plasmids pNRC100 (5000 – 5256) and pNRC200 (6000 – 6487). Genes that displayed a $\log_2(x)$ ratio of $-0.5 < x < 0.5$ were considered not to be differentially expressed.

Table 3.6 Comparison of microarray data for *cyc* loci of *Halobacterium* strains. The data for *Halobacterium* sp. NRC-1 is for the transition from exponential to stationary phase (Facciotti *et al*., 2010), while the data from the current study is for the comparison of the wildtype and Δ*CYP174A1* strains of *H. salinarum* R1 in stationary phase.

***cyc* locus (NRC-1 Gene ID)[c]**	**Protein function**	**Fold change[a] from exponential to stationary phase (NRC-1 wildtype strain)**	**Fold change[b] between Δ*CYP174A1* and wildtype R1 strains in stationary phase**
vng419 (OE1626)	Unknown	+47.80	+1.01 (0.480)[d]
vng420 (OE1628)	Unknown	Not reported	+1.09 (0.453)
vng421 (OE1629)	Unknown	-0.05	+1.18 (0.336)
cyc	Cytochrome P450 monooxygenase	+110.04	-5.10 (8.00E-06)
nacR (OE1633)	NAC alpha-BTF3-like transcriptional regulator	+0.05	-1.14 (0.480)
pimT2 (OE1636)	L-isoaspartyl-protein carboxyl methyltransferase	Not reported	-1.24 (0.281)
tfs2 (*rpoM1*)	Transcription elognation factor TFIIS	+0.02	+1.01 (0.290)

[a] Data taken from the supplementary section of Facciotti *et al*. (2010) – 'Fold change' in this case is calculated as the ratio between the average non-logged ratio for the last four samples (replicates included) taken in stationary phase to the average of the first four samples taken in exponential phase. +/- = incease or decrease in fold change respectively.

[b] Represents linear fold change. +/- = incease or decrease in fold change respectively.

[c] *H*. *salinarum* R1 gene ID in parentheses.

[d] p-values of fold change of current study.

None of the genes flanking the *CYP174A1* were differentially expressed in the mutant since fold changes were either > -1.5 or < +1.5. This was regarded as a positive indication that none of the ORFs or promoter regions of the genes flanking the *cyc* gene were severely altered on nucleotide level by the deletion.

When evaluating differential expression of genes in the Δ*CYP174A1* strain, the chromosomal *bop* gene (Figure 3.13 and Annexure 1, Table A.1, Gene ID: vng 1467) was the most striking, due to its severely decreased transcript levels (19 fold change) during the stationary phase of growth ($\log_2(x) = -4.106 \pm 0.602$). Transcript levels of this gene were even more drastically reduced than that of the *cyc* gene (5-fold change). It has early on already been reported that the *bop* gene which encodes bacterioopsin (BO) (DasSarma, 1989; Peck *et al*., 2001), is induced in wildtype *Halobacterium* during the stationary phase of growth (Yang & DasSarma, 1990; Shand & Betlach, 1991). BO and retinal form bacteriorhodopsin (BR) which forms a specialized, two dimentional crystalline lattice in the membranes referred to as purple membrane (PM) (DasSarma, 1989). The biosynthesis and regulation of PM is controlled by a cluster of genes referred to as the *bop*-regulon (Peck *et al*., 2001; Tarasov *et al*., 2008). In a study by Facciotti *et al.* (2010), using *Halobacterium* sp. NRC-1, gene regulation during the transition from exponential to stationary phase was studied with microarray. We compared the differential expression of the *bop*-regulon genes in the Δ*CYP174A1* strain to regulation of the bop-regulon from wildtype *Halobacterium* sp. NRC-1 as reported by Facciotti *et al.* (2010) (Table 3.7).

Table 3.7 Comparison of microarray data for bop-regulon of *Halobacterium* strains. The data for *Halobacterium* sp. NRC-1 is for the transition from exponential to stationary phase (Facciotti *et al*., 2010), while the data from the current study is for the comparison of the wildtype and Δ*CYP174A1* strains of *H. salinarum* R1 in stationary phase

bop-regulon gene (NRC-1 Gene ID)[c]	**Protein function**	**Fold change[a] from exponential to stationary phase (NRC-1 wildtype strain)**	**Fold change[b] between Δ*CYP174A* and wildtype R1 strains in stationary phase**
crtB1	Phytoene synthase	+146.86	-1.03 (0.472)[d]
vng1459 (OE3095F)	Unknown	+146.14	-1.99 (0.002)
vng1461 (OE3097R)	Unknown	Not reported	+1.28 (0.264)
cdc48a (*aaa2*)	Cell division cycle protein	Not reported	+1.01 (0.460)
blp	Bacterioopsin-linked protein	+135.48	-1.24 (0.237)
bat	Bacterioopsin activator	+131.26	+1.01 (0.356)
brp	Bacteriorhodopsin-related protein	+165.23	-1.43 (0.249)
brz	Bacteriorhodopsin-regulating zinc finger protein	Not reported	+1.11 (0.445)
bop	Bacteriorhodopsin precursor	+122.23	-19.10 (8.73E-20)
vng1468 (OE3107)	Unknown	+112.64	-3.11 (1.58E-06)

[a] Data taken from the supplementary section of Facciotti *et al*. (2010) – 'Fold change' in this case is calculated as the ratio between the average non-logged ratio for the last four samples (replicates included) taken in stationary phase to the average of the first four samples taken in exponential phase.+/- = incease or decrease in fold change respectively.

[b] Represents linear fold change. +/- = incease or decrease in fold change respectively.

[c] *H*. *salinarum* R1 gene ID in parentheses.

[d] p-values of fold change of current study.

When examining the gene regulation data from the wildtype *Halobacterium* sp. NRC-1 strain in Table 3.7, it can be seen that seven of the ten *bop*-regulon genes (exceptions vng1461, *cdc48a* and *brz*) were induced in the transition from logarithmic to stationary phase. Interestingly two genes (vng1459 and vng1468) together with *bop*, had reduced transcript levels in the stationary phase cultures of the Δ*CYP174A1* strain. All the other genes comprising the bop-regulon were not differentially expressed.

The *thiC* gene is the second chromosomal gene of the Δ*CYP174A1* strain that displayed significantly lower trascript levels during the stationary phase (6 fold change) when compared to the wildtype strain. Its fold change was also larger than that of the *cyc* gene (5 fold change). The *thiD* gene also had 3 fold lower transcript levels in the mutant strain during stationary phase. Both these genes encode proteins involved in thiamine biosynthesis (Fig 3.14). The thiC protein catalyses the conversion of aminoimidazole ribotide (AIR) to 4-amino-2-methyl-5-phosphomethylpyrimidine (HMP-P) while thiD is involved in adding phosphate groups to 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) as well as HMP-P (Rodionov *et al*., 2002). Facciotti *et al.* (2010) found that *thiC* and *thiD* were both transiently expressed during exponential growth, but not significantly regulated during the transition from exponential to stationary phase. In their study the *thiL* (Monophosphate kinase) gene was down regulated during the transition from exponential to stationary phase The thiL protein is responsible for the synthesis of thiamine diphosphate *via* a single phosphorylation of thiamine phosphate (Rodionov *et al*., 2002).

There are 256 genes from the pNRC100 plasmid on the microarrays and 487 from the pNRC200 plasmid. Fifty-six genes associated with the 'pNRC100' plasmid were differentially expressed in the Δ*CYP174A1* strain. Of these 33 have unknown function and 7 code for genes involved in gas vesicle formation. From a 'pNRC200' point of view, 101 genes were differentially expressed of which 62 genes have unknown function and 12 genes code for gas vesicle formation. These gas vesicle genes included some of the genes that displayed the most obvious differential expression in the stationary phase cultures of the Δ*CYP174A1* strain with transcript levels 11 to 4 fold reduced for gvpA2, gvpA1, gvpC2 and gvpN2. Surprisingly transcript levels for the gvpC1 gene were 3 fold higher.

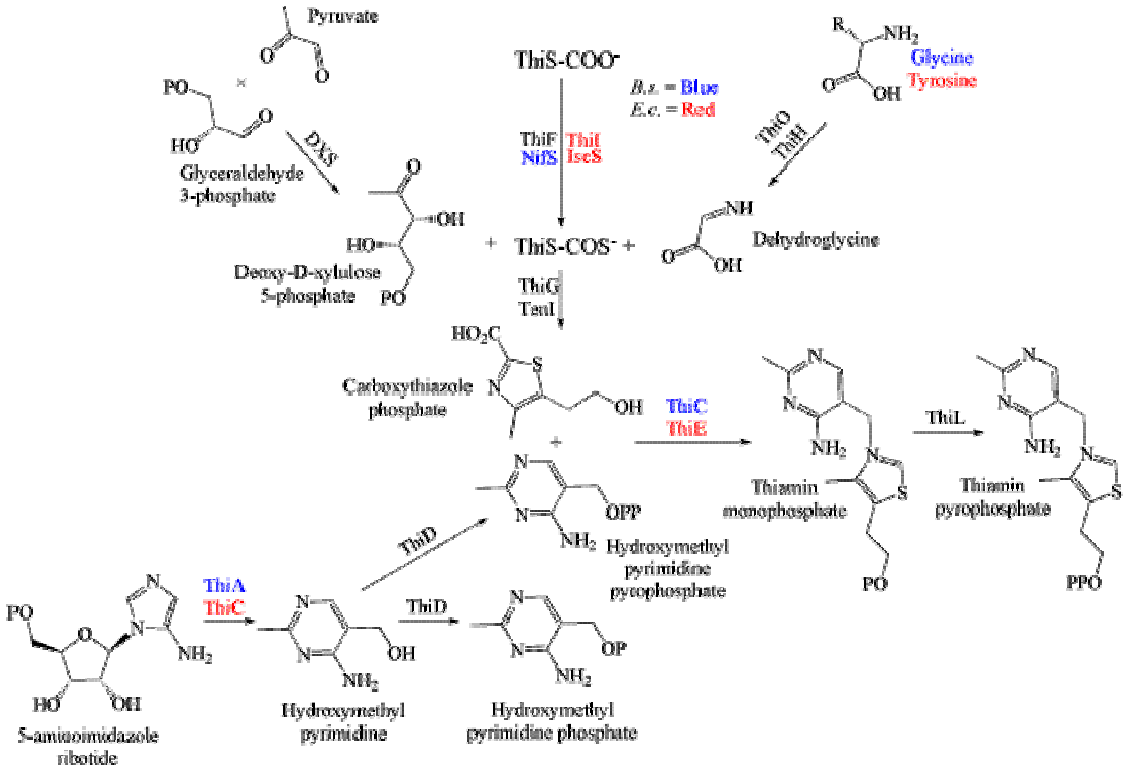


Fig. 3.14 Physiological roles of *thiC* and *thiD* in thiamine biosynthesis in *Halobacterium*. Relevant abbreviations are defined in the text. (Jurgenson *et al*., 2009)

With *Halobacterium* sp. NRC-1 Facciotti *et al.* (2010) observed upregulation of the gas vesicle genes during the transition from exponential to stationary phase, as observed for the *bop*-regulon. *Halobacterium* sp. NRC-1 is known to produce gas vesicles – a striking characteristic of this strain (DasSarma, 1989). In our case the relatively strong differential expression of the genes involved in gas vesicle formation was unexpected for *H*. *salinarum* R1 and is difficult to interpret, since from a phenotypic point of view, *H*. *salinarum* R1 is gas vesicle deficient due to a 1.3 kb insertion element (ISH3) located 15 bp before the transcription start site of the *gvpA* gene (DasSarma *et al*., 1988; DasSarma, 1989). However, Twellmeyer *et al*. (2007) also previously observed in microarray experiments, using *H*. *salinarum* R1, the unexpected regulation of these genes albeit during mid-exponential phase and under phototrophic conditions. The gas

vesicle genes that were induced in the above mentioned study were: *gvpL1*, *gvpK1* and *gvpI1* (on the pHS1 plasmid) whereas *gvpI2* and *gvpC2* (on the chromosome) were repressed. Thus, the unexpected differential expression of these gas vesicle genes in *H*. *salinarum* R1 is not without precedent.

3.3.5 Purple membrane isolation by means of a sucrose gradient

It was decided to corroborate the microarray results, which indicated reduced or no expression of the *bop* gene involved in BR and thus PM biosynthesis, with a sucrose gradient to separate the PM from the red pigments (Danon *et al*., 1977). Three *CYP174A1*-deletion mutants were evaluated for purple membrane synthesis and compared with the wildtype *H*. *salinarum* R1 strain by utilizing a sucrose gradient which separates red pigments from the purple membrane. All strains were cultured at 40˚C until an OD_{600} = 1.2 was reached – this corresponds to the late stationary phase of growth when *bop* is usually up-regulated. After dialysis and DNaseI treatment, cell pastes were layered onto a sucrose gradient and subjected to ultra-centrifugation. Figure 3.15 illustrates typical sucrose gradient results observed when comparing purple membrane synthesis of the wildtype strain *versus* that of the deletion strain. The wildtype strain displayed an intact PM while all three the knock-out strains displayed a complete absence of PM. The biosynthesis of PM is a tightly regulated system (*bop* regulon) that is governed mainly by oxygen levels and light intensity (Baliga *et al*., 2001; Peck *et al*., 2001; Tarasov *et al*., 2008). Based on the microarray- and sucrose gradient results, it was concluded that *CYP174A1* probably plays a role in PM biosynthesis in *Halobacterium*.

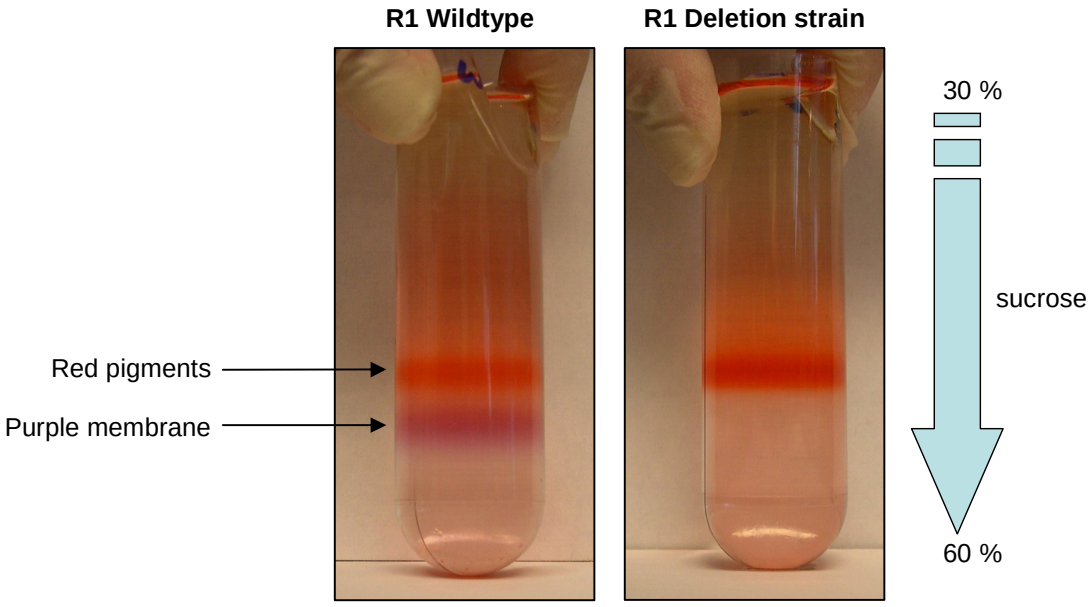


Fig. 3.15 Sucrose gradients of 5 mL dialyzed and DNaseI-treated cell lystates from wildtype and Δ*CYP174A1 H*. *salinarum* R1 strains ultracentrifuged at 132 000 x *g* for 17 h at 18˚C.

3.3.6 Evaluating the genetic integrity of the *bop* gene cluster

The *bop* gene encodes an apoprotein bacterioopsin that covalently binds to a chromophore retinal in a 1:1 fashion to yield bacteriorhodopsin (BR) that forms a two-dimensional crystalline lattice in the membrane (Baliga *et al*., 2001; Tarasov *et al*., 2008). The *bop* gene cluster consists of *bat*, *bop*, *brp* and *blp* genes of which *bat* and *brp* play significant roles in *bop* regulation and consequently BR biosynthesis (Tarasov *et al*., 2008). Abolishment of *bop* expression, and consequently PM synthesis, is mostly attributed to spontaneous inserstions either in *bop* itself (DasSarma, 1989) or in the *brp* gene (Betlach *et al*., 1984). Several insertion sequences (IS) have been identified in *bop* from various purple membrane deficient *Halobacterium* strains. IS information is summarized in Table 3.8.

Table 3.8 IS in the *bop* gene in purple membrane deficient *Halobacterium* strains

IS designation	Size of IS	Reference
ISH 1	1 118 bp	DasSarma *et al*., 1983
ISH 2	520 bp	DasSarma *et al*., 1983
ISH 23	1 000 bp	Pfeiffer *et al*., 1984
ISH 26	1 700 bp	Ebert *et al*., 1987
ISH 27	1 400 bp	Pfeifer *et al*., 1985
ISH 28	1 000 bp	Pfeifer *et al*., 1985
ISH S1	1 400 bp	Ovchinnikov *et al*., 1984

As a rule, ISH 1 integrates into *bop* at a single site and ISH 2 at several sites within or upstream of the *bop* gene (DasSarma *et al*., 1983). All the above-mentioned IS give rise to single mutations and results in the abolishment of purple membrane synthesis. Some *Halobacterium* IS-generated-mutants display pleiotropic loss of pigment synthesis (purple membrane and loss of carotenoids). This is usually the result of ISH 2 insertions together with IS elements, 0.5 – 3 kb in size, that integrate upstream of the structural *bop* gene (Pfeifer *et al*., 1985; Leong *et al*., 1988). This is not the case with the *H*. *salinarum* R1 ∆*CYP174A1* mutant since the red pigments are still present as judged from the sucrose gradients and spectrophotometric scans.

To assess the presence of IS in the *bop* and *brp* genes, the entire ORF of both genes were PCR amplified. In addition the *bat* gene was also amplified to assess its gene size. The sizes of IS-free *bop*, *brp* and *bat* genes in *H*. *salinarum* R1 are 786 bp, 1 104 bp and 2 022 bp respectively. Figure 3.16 shows the results of the PCR screen for spontaneous insertions into the *bat*, *brp* and *bop* genes of wildtype *H*. *salinarum* R1, a *H*. *salinarum* R1 *CYP174A1*-deletion strain and *Halobacterium* sp. NRC-1 (positive control). All amplicons were of the correct size and it was concluded that IS-elements were not responsible for the repression of *bop* and resulting loss of PM.

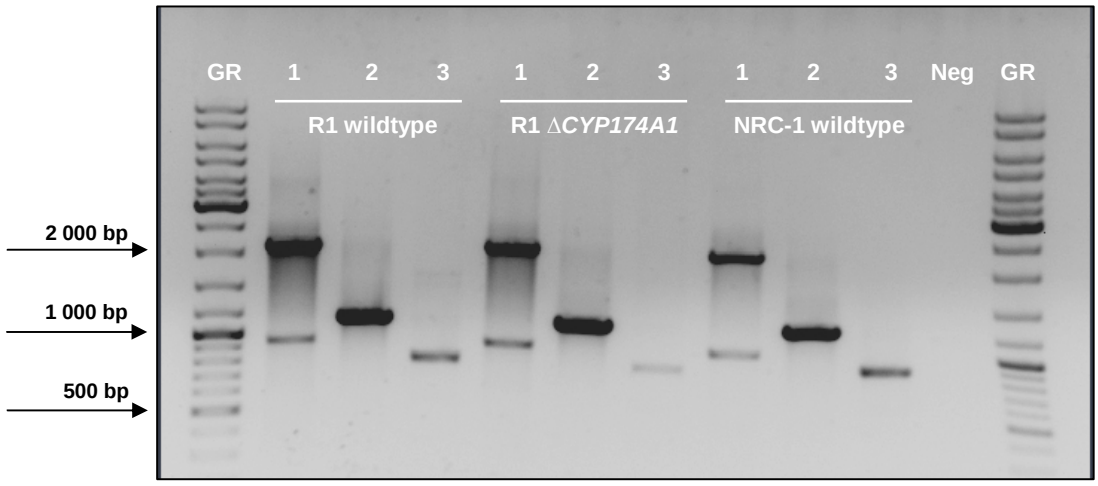


Fig. 3.16 PCR amplification of the *bat*, *brp* and *bop* genes to screen for the presence of spontaneous insertions. Expected sizes of amplicons in wildtype as well as the deletion strains of *H. salinarum* R1: *bat* = 2 022 bp, *brp* = 1 104 bp, *bop* = 786 bp. For *Halobacterium* sp. NRC-1: *bat* = 2 025 bp, *brp* = 1 080 bp, *bop* = 789 bp. Lanes: **GR** = 5 µL GeneRuler (Fermentas); **1** = *bat*; **2** = *brp* and **3** = *bop* and **Neg** = Negative control.

3.4 Discussion

3.4.1 CYP450s are relatively abundant in halophilic archaea

To date, our knowledge of CYP450s from extremely halophlic archaea is limited to nucleic acid sequence data as a result of several whole genome sequencing projects. None of these CYP450s have however been cloned, expressed or characterized. When evaluating the number of all the known archaeal CYP450s, it would seem as if CYP450s in halophilic archaea are relatively abundant. All the sequenced genomes of halophiles contain at least one CYP450 encoding ORF and the maximum number recorded is nine in the genome of *Haladaptatus paucihalophilus*. Of the 32 known halophilic archaeal CYP450s, 17 have been formerly assigned to eight different CYP450 protein families (Fig. 3.1). Currently, the CYP174-family appears to be the most frequently occurring family in halophilic archaea since it is present in 8 different genera including *Halobacterium* and *Haloarcula*. The abundance of CYP450s in halophilic archaea implies that CYP450s must have some essential function in all halophiles and it would appear as if CYP450s belonging to the CYP174-family have a common function in these organisms.

3.4.2 Heterologous expression of the *Haloarcula* sp. LK-1 *CYP174A2* in bacterial hosts

An extremely halophilic arhaeon was isolated from salt crystals taken from Lemoenkloof commercial saltpan outside Bloemfontein (South Africa). 23S-rRNA sequence analysis placed this isolate in the the genus *Haloarcula* (section 3.3.1) and consequentely the isolate was designated as *Haloarcula* sp. LK-1. By using whole cells of *Haloarcula* sp. LK-1 and oligonucleotides based on the CYP450 sequences of *Halobacterium* sp. NRC-1, *H*. *salinarum* R1 and *H*. *marismortui*, a complete CYP450 gene was amplified. The amino acid sequence of the *Haloarcula* sp. LK-1 CYP450 had 98 % identity to the CYP174A2 protein from *H*. *marismortui.* Alignment of the saltpan CYP450 with the CYP174A2 of *H*. *marismortui* is illustrated in Figure 3.17.

```
Saltpan P450  1   MSKTPPGPKGEPLFGSSRTYARDPFRFISALERAYGDVARFDMGPMDTVMLCDPTAIERVLVSEADQFRKPDFQGDA 77
Hm CYP174A2   1   MSKTPPGPKGEPLFGSSRTYARDPFRFISALERAYGDVARFDMGPMDTVMLCDPTAIERVLVSEADRFRKPDFQGDA 77

Saltpan P450  78  LGDLLGDGLLLSEGETWEQQRKLANPAFSMARLSGMADRITGHAKDRIADWSHGDVIDAEQSMTRVTLDVILDLMMG 154
Hm CYP174A2   78  LGDLLGDGLLLSEGETWEQQRKLANPAFSMARLSGMADRITGHAEDRIADWSHGDVIDAEQSMTRVTLDVILDLMMG 154

Saltpan P450 155  VELSEQRVQTIEEQLLPLGQRFEPDPIRFAMPQWMPMPDDAEFNRAVRTLDEVLDDIIEVREDSVGSGDDGPMDFLS 231
Hm CYP174A2  155  VELSEQRVQTIEEQLLPLGQRFEPDPIRFAMPQWMPMPDDAEFNRAVRTLDEVLDDIIEVREDSLGTDEDGPMDFLS 231

Saltpan P450 232  VLLRARDEGNQSPEQLRDEMMTMLLAGHDTTALTLTYTWFLLSEHPEVEQRVHEELDDVIGDDRPGMEHVRELDYLE 308
Hm CYP174A2  232  VLLRARDDGNQSPEQLRDEMMTMLLAGHDTTALTLTYTWFLLSEHPEVEQRVHEELDDVIGDDRPGMEHVRELDYLE 308

Saltpan P450 309  WVIQEAMRLYPPVYTIFREPTEDVTLSGYDVEAGTTLMVPQWGVHRSERFYDDPETFDPERWKPERANERPRFAYFP 385
Hm CYP174A2  309  WVIQEAMRLYPPVYTIFREPTEDVTLSGYEVEAGTTLMVPQWGVHRSERFYDDPETFDPERWKPERASERPRFAYFP 385

Saltpan P450 386  FGGGPRHCIGKHLAMLEAQLITATTASQYRLEFQGETPLELLPSLTAHPRQKMSMRVQER 445
Hm CYP174A2  386  FGGGPRHCIGKHLAMLEAQLITATTASQYRLEFQGETPLELLPSLTAHPRQKMSMRVQER 44
```

Fig. 3.17 Multiple alignment of the salpan CYP450 protein with the CYP174A2 protein from *H. marismortui* ATCC43049 (Hm CYP174A2). Similar amino acids are highlighted in turquoise while identical amino acids are highlighted in pink. Alignments were performed using DNAssist 2.2. Information regarding the theoretical pI and molecular weight of the *Haloarcula* sp. LK-1 (saltpan isolate) CYP450 is provided in Fig. 3.5.

We attempted to study this CYP450 from *Haloarcula* sp. LK-1 by heterologous expression in *E*. *coli* but were unable to achieve this since all the expressed protein localized to the insoluble fraction (section 3.3.2.1 and 3.3.2.2). In addition to heterologous expression in *E*. *coli*, we also attempted to perform heterologous expression in *P*. *fluorescenc*e, but could not detect any expressed CYP450 protein in either the soluble or insoluble fractions.

Although some groups have successfully expressed genes from *H*. *salinarum* (Ishibashi *et al*., 2001), *Haloferax volcanii* (Connaris *et al*., 1999), *Haloferax mediterranei* (Pire *et al*., 2001) and *Halomonas* sp. #593 (Yonezawa *et al*., 2003) in *E.coli*, other groups also experienced the same problems as we did. Most notably are the studies by Connaris *et al*. (1999); Ishibashi *et al*. (2001); Pire *et al*. (2001) and Yonezawa *et al*. (2003). In all these studies the expressed halophilic proteins were insoluble as well as inactive. This was probably due to the fact that most halophilic enzymes from extreme halophilic archaea require high salt concentrations for their activity and stability. Most of these enzymes could regain their activity by refolding them in the presence of salts *in vitro* (Yonezawa *et al*., 2003). We doubted that this will work for heme-containing enzymes such as CYP450s.

An avenue that might be worth exploring is the possibility of using the hsp70 (DnaK) chaperone which is found in *H*. *marismortui* as well as in both strains of *Halobacterium salinarum* (Macario and Conway de Macario, 1999). By co-expressing chaperones with the target gene, the halophilic proteins are more likely to fold correctly and thus not form aggregates. In a study conducted by Ahn *et al*. (2004) the heterologous expression of human CYP1A2 in *E*. *coli* was greatly enhanced when the CYP450 gene was co-expressed with the human hsp70 chaperone. A similar trend was reported by Stephens *et al*. (2011) when the GTP cyclohydrolase I gene from *Plasmodium falciparum* was expressed in *E*. *coli* in conjunction with the *P*. *falciparum* hsp70 protein. Thus by using the strategy of native chaperones from the halophilic host, the degree of correct folding and solubility of CYP450s from extremely halophilic archaea might be markedly improved.

3.4.3 Elucidating the physiological role of CYP174As by creating a *CYP174A1* knock-out in *H. salinarum* R1 and doing microarray analysis

The CYP450 we isolated is almost identical to the CYP174A2 from *H*. *marismortui* (Fig. 3.17) and can thus be grouped into the CYP174 protein family. The CYP174 famlily currently consists of subfamilies A, B and C (Fig. 3.1) which each respectively have four, three and one CYP450 genes assigned to them. The CYP450 from *Haloarcula* sp. LK-1 sorts under the CYP174A protein family which comprises CYP174A1 from *Halobacterium* sp. NRC-1 and also *H*. *salinarum* R1, CYP174A2 from *H*. *marismortui* and CYP174A3 from *Halomicrobium mukohataei*. Based on the family grouping it seems plausible that CYP450 proteins belonging to the CYP174A-family will have similar physiological functions and since we were experiencing difficulties with heterologous expression of the *Haloarcula* sp. LK-1 CYP450 and consequently had no idea of what the physiological function was, we decided to adopt another strategy: namely to delete the *CYP174A1* in *H*. *salinarum* R1 and study the effects of this CYP450 deletion on the transcriptome of *H*. *salinarum* R1 using microarray. Deleting the CYP450 genes from the other strains containing members of the CYP174A-family was not practical since *H*. *marismortui* has two additional CYP450s (*CYP1001A2* and *CYP1003A1*) and *H*. *mukohataei* has one additional CYP450 (*CYP174C1*). *H*. *salinarum* R1 and *Halobacterium* sp. NRC-1 both possess a single CYP450 (*CYP174A1*) which would make the deletion process much simpler and in addition avoid any possible physiological influences from other native CYP450s as could be the case with *H*. *marismortui* and *H*. *mukohataei*.

A *CYP174A1* knock-out was achieved in *H*. *salinarum* R1 with the aid of a pMKK100 suicide vector. When evaluating the growth of the wildtype and Δ*CYP174A1* strains (Fig. 3.9 A) it was observed that both strains grew the same during logarithmic phase. However, during stationary phase the Δ*CYP174A1* strain consistently produced more biomass than the wildtype strain. In addition, a difference in pigmentation was also observed (Fig. 3.9 C and D).

To gain some understanding of the effect of the *CYP174A1* knock-out in *H*. *salinarum* R1, the transcriptome was evaluated by doing microarray analysis on samples taken

during logarithmic- and stationary phases of growth. Microarray results indicated that there were some differences in transcript levels, albeit small, during logarithmic phase when comparing the two strains. However, more pronounced differences were evident during the stationary phase of growth, most notably the suppression of the *bop* gene involved in purple membrane (PM) biosynthesis. A sucrose gradient confirmed the microarray result since PM could be isolated from the wildtype strain but not from the Δ*CYP174A1* strain (Fig. 3.15). The production of red pigments was however still intact in both strains. The absence of PM is usually attributed to insertion elements (ISH), but this possiblity was ruled out by a PCR screening (section 3.3.6, Fig. 3.16).

3.4.4 How might a *CYP174* deletion affect the *bop* regulon

BR is the key component of the PM in *Halobacterium* and is strongly induced under microaerobic conditions during stationary phase (Peck *et al*., 2001; Tarasov *et al*., 2008). The biosynthesis of BR is tightly regulated by a cluster of genes refered to as the *bop* regulon (Fig. 3.18) which comprises *crtB1*, *blp*, *bat*, *brp*, *brz* and *bop*. A summary of the functioning *of the bop* regulon is given in Figure 3.18. In addition to gene regulation, negative feedback loops also exist: substrate inhibition by retinal, inhibition by bacterioopsin or activation by BR (Sumper & Herrmann, 1976).

The *crtB1* gene codes for phytoene synthase, which is a key enzyme in the biosynthesis of C_{40} carotenoids *e.g*. lycopene and β-carotene and consequently also retinal (Baliga *et al*., 2001). Downstream from the *crtB1* gene is the *blp* (*b*op-*l*inked *p*rotein) gene that codes for a protein of unknown function. *Blp* is co-regulated with the *bop* gene when low oxygen tension is experienced by the cells (Gropp & Betlach, 1994).

The *brp* (*b*acterioopsin-*r*elated *p*rotein) and *bat* (*b*acterioopsin *a*ctivator of *t*ranscription) genes form a transcription unit and their termination and initiation codons overlap (Peck *et al*., 2001). Expression of *bat* is inhibited by oxygen while expression of *brp* is enhanced by light (Shand & Betlach, 1991). The *bat* gene encodes a *trans*-acting factor that induces *bop* at low oxygen tension, which naturally occurs in the stationary phase. *Bat* contains a GAF domain, a C-terminal DNA-binding helix-turn-helix motif and a PAS/PAC (oxygen sensing) region (Baliga *et al*., 2001).

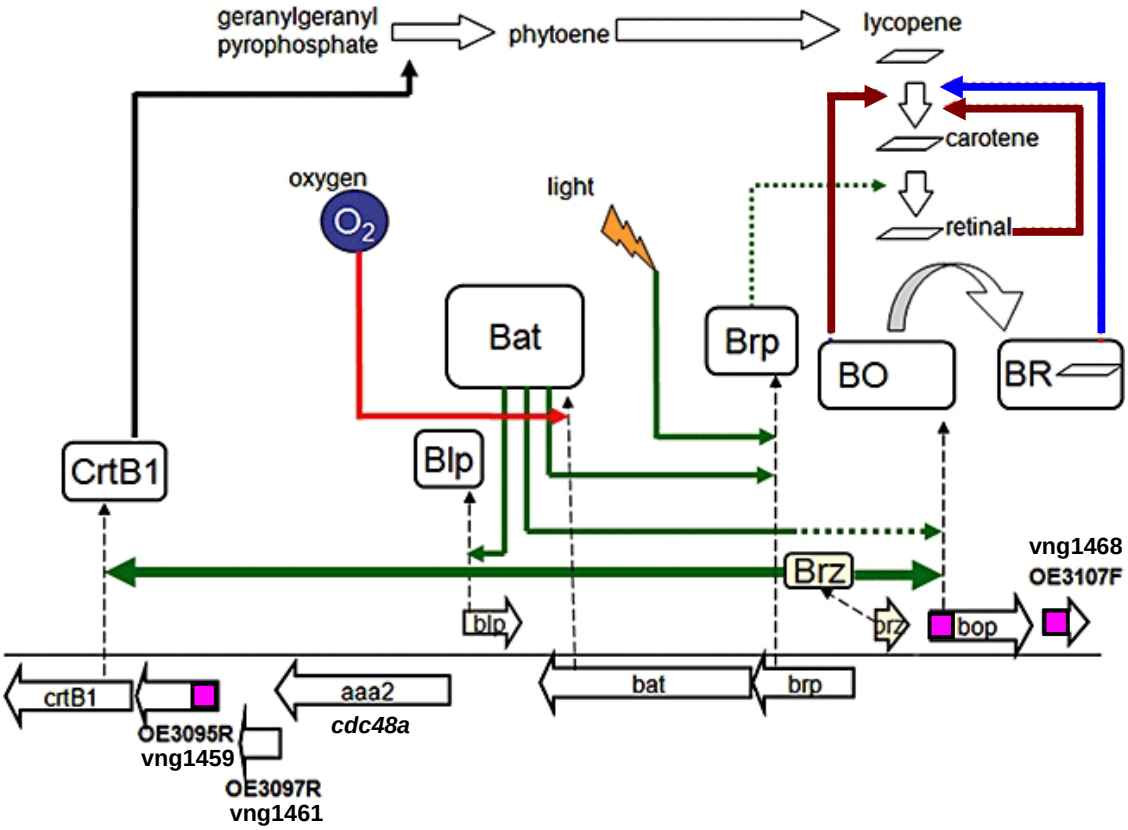


Fig. 3.18 The *bop* gene regulation network. Genes are depicted as arrows and their protein products in boxes. Gene names from *Halobacterium* sp. NRC-1 are indicated either above or below the corresponding *H. salinarum* R1 gene names. Gene activation is indicated by green arrows and gene inhibition by a red arrow. Gene activation by *brz* is indicated by a thick green arrow. Activation of carotenoid conversion to retinal by *brp* is indicated by a green dotted arrow. Substrate inhibition by retinal and inhibition by BO are indicated by brown arrows and activation by BR by a blue arrow. Genes that were differentially expressed in this study contain a pink box (Adapted from: Tarasov *et al.*, 2008).

An in-frame *brp* deletion led to β-cartotene accumulation and a decrease in retinal, and consequently a decrease in BR levels (Peck *et al*., 2001). Deletion of a *brp*-homologue called *blh*, resulted in normal BO, retinal, BR and β-carotene levels. However, a Δ*blh*Δ*brp*-strain displayed no formation of BR or retinal and a concomitant 5.3 fold increase in β-carotene when compared to the wildtype. In addition, BO levels remained almost the same as in the wildtype cells. Brp and blh are most probably 15,15′-β-carotene dioxygenases, since they are approximately the same size and share conserved motives with a proven 15,15′-β-carotene dioxygenase from an uncultured marine bacterium (Kim *et al*., 2009). Since β-carotene is also converted to retinal by mammalian 15,15′-β-carotene monooxygenases (Fig. 3.19 B) *via* a monooxygenase mechanism, one might have proposed such a function for CYP174s. However, since deletion of *brp* and *blh* completely abolishes formation of retinal and BR this is highly unlikely.

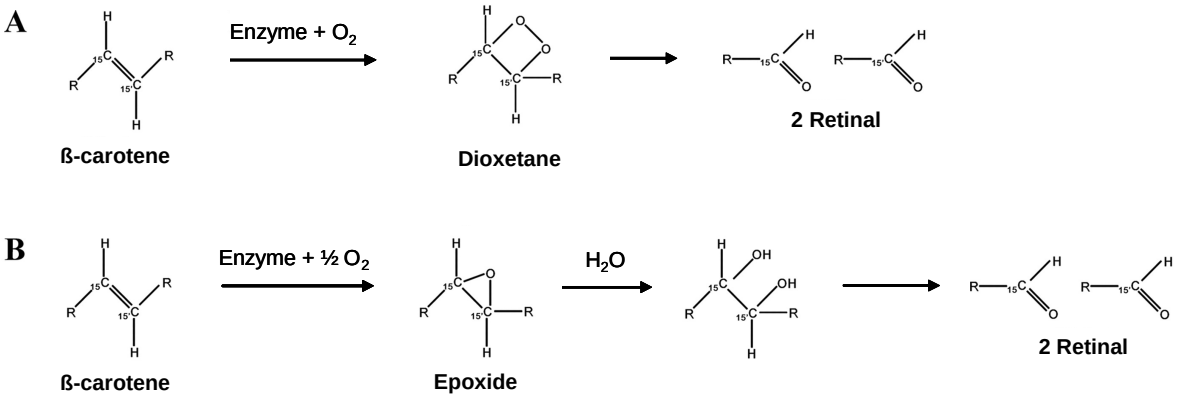


Fig. 3.19 Cleavage of β-carotene by 15,15′-β-carotene dioxygenase **(A)** as well as monooxygenases **(B)** to yield two molecules of retinal. Note: full chemical structures are not shown here (Adapted from: Kim *et al*., 2009).

The final known component of the *bop* regulon is a gene located in the *bop-brp* intergenic region called *brz* (*b*acteriorhodopsin-*r*egulating *z*inc finger protein) which codes for a 60 amino acid, zinc finger motif-containing protein. Although *bat* activates *brp*-, *bop*- and most probably *blp* expression, *bop* as well as *crtB1* are also activated by

brz. *Brz* however, does not affect *bat* expression, which could be indicative that *bop* activation by *bat* is not direct but instead mediated by *brz* (Tarasov *et al*., 2008).

The *brz* protein controls regulation of *bop* and *crtB1* but not *bat* (Tarasov *et al*., 2008). Deletion of the *brz* gene as well as site-directed mutagenized versions of *brz* in *H*. *salinarum* R1 resulted in reduced transcript levels of *bop*, *crtB1*, vng1459/OE3095 (a gene with unknown function adjacent to *crtB1*) as well as vng1468/OE3107 (a gene of unknown function directly adjacent to *bop*). Interestingly enough, the pairing of vng1468/OE3107 with *bop* is also found in *Haloarcula marismortui* and *Haloquadratum walsbyi* – both these halophiles also contain BO. This conserved gene pairing in these three genera may be indicative of a functional association (Tarasov *et al*., 2008). In addition, orthologues of the vng1459/OE3095 gene upstream of *crtB1* have also been reported in *H*. *marismortui* and *H*. *walsbyi* and all three also contain a zinc-finger motif very similar to Brz in *Halobacterium*. Whether the vng1459/OE3095 gene encodes a protein with a regulatory function is still unknown. It is noteworthy to mention here that *H*. *marismortui* and *H*. *walsbyi* both contain CYP450s belonging to the CYP174 family (Fig. 3.1).

Microarray results from this study indicated that only *bop*, vng1459/OE3095 and vng1468/OE3107 had significantly reduced expression levels in stationary phase cultures of the *CYP174A1* deletion strain. Thus three of the four genes with reduced transcript levels in the *brz* deletion strain also displayed reduced expression levels in the CYP174A1 deletion strain, although expression of *brz* was not affected and PCR had shown that there was no IS introduced into *brz*. It thus appears as if a hydroxylated product produced by CYP174A1 might be involved in yet another pathway that regulates *bop*. Tarasov *et al*. (2008) already suggested that vng1459/OE3095 which also contains a zinc finger motif, might be involved together with *brz* in a hierarchical regulatory network

3.4.5 Possible effects of lack of BO and BR on retinal regulation and pigmentation

The BO apoprotein forms part of a negative feedback loop that causes inhibition of lycopene cyclase (*crtY*) which converts lycopene into β-carotene (Sumper & Herrmann, 1976; Peck *et al*., 2002). This inhibition would result in increased lycopene levels, a decrease in β-carotene levels and ultimately a cessation in retinal and consequent BR formation. Since *bop* is down-regulated in this study, the assumption can be made that lycopene cyclase will not be inhibited (at least not by mature BO) and that β-carotene and retinal will be synthesized. However, BR is absent from the *H. salinarum* R1 Δ*CYP174A1* strain which implies that retinal would have to accumulate in the cells. Accumulation of retinal would result in *crtY* inhibition which would again cause a drop in β-carotene levels and accumulation of lycopene.

Lycopene is the branch point for the synthesis of functional pigments in *Halobacterium*, since lycopene can feed the retinal pathway (that will give rise to BR and then PM) or the bacterioruberin (red pigments) pathway (El-Sayed *et al*., 2002). It seems plausible to suggest that excess lycopene is shifted towards bacterioruberin synthesis. This might explain the difference in pigmentation which was observed in stationary phase cultures of the wildtype and the deletion strains.

3.4.6 Speculating on the possible function of *CYP174A1*

Results from the microarray and sucrose gradient experiments point towards the possibility that *CYP174A1* plays a role in BO regulation and, as a consequence, also the biosynthesis of PM in *H. salinarum* R1. CYP174A1 metabolites could possibly act as signaling molecules as is the case with *e.g*. CYP4A1 metabolites of arachidonic acid that act as intracellular signaling molecules in vascular tissue (Harder *et al*., 1997). Some CYP450 monooxygenases have also been implicated in bacterial quorum sensing pathways *e.g*. in *Rhodococcus erythropolis* the CYP450 hydroxylates the quorum sensing molecules of other bacteria so that these molecules may not be recognized as quorum sensing signals anymore by their own bacterial population (Uroz *et al.*, 2005).

Another possibility could be that CYP174A1 is involved in astaxanthin production – a ketocarotenoid with potent antioxidant and singlet oxygen-quenching activities (Makino *et al*., 2008). It has been reported by Calo *et al*. (1995) that some species of extremely halophilic archaea like *Haloarcula hispanica* and *H. salinarum* contain trans-astaxanthin. In *H. salinarum* about 11% of the pigment (on a pigment per weight basis) was *trans*-astaxanthin, while 24% was 3-hydroxy-echinenone (Calo *et al*., 1995). Astanxanthin synthesis, which can occur *via* 3-hydroxy-echinenone, requires four enzymatic steps that involve the addition of two keto and two hydroxyl moieties at the 4,4' and 3,3'-positions of the β-ionone rings of β-carotene respectively (Fig. 3.20) (Martín *et al*., 2008).

Thus, β-carotene ketolase and β-carotene hydroxylase activity is required. A survey of both genomes of *H. salinarum* R1 and *Halobacterium* sp. NRC-1 did not reveal any ketolase encoding genes nor did BLAST analyses using known ketolase protein sequences *e.g*. *crtO* and *crtW*. We did however, come across an interesting fact regarding the hydroxylase needed for astaxanthin production: in the yeast *X. dendrorhous*, a known astaxanthin producing organism, the *crtS* gene was identified as a key role player in astaxanthin production and it was also found that *crtS* is a CYP450 hydroxylase (Álvarez *et al*., 2006). In addition, a CYP450 hydroxylase has also been implicated in astaxanthin production in the green algae *H. pluvialis* (Schoefs *et al*., 2001). Given the presence of *trans*-astaxanthin in *H. salinarum* (Calo *et al*., 1995) and that CYP450s have been implicated in astaxanthin production, we therefore suggest that CYP174A1 could be involved in astaxanthin production in *H. salinarum*. According to Martín *et al*. (2008) there had been a hypothesis that the CYP450 encoded by *crtS* might have both ketolase and hydroxylase activity. Although this hypothesis was later disproved, it might even be possible that CYP174s might have such bifunctional activity. It is then also possible that astaxanthin (or perhaps another hydroxylated carotenoid produced by CYP174A1) might be the chemical messenger that regulates BO synthesis.

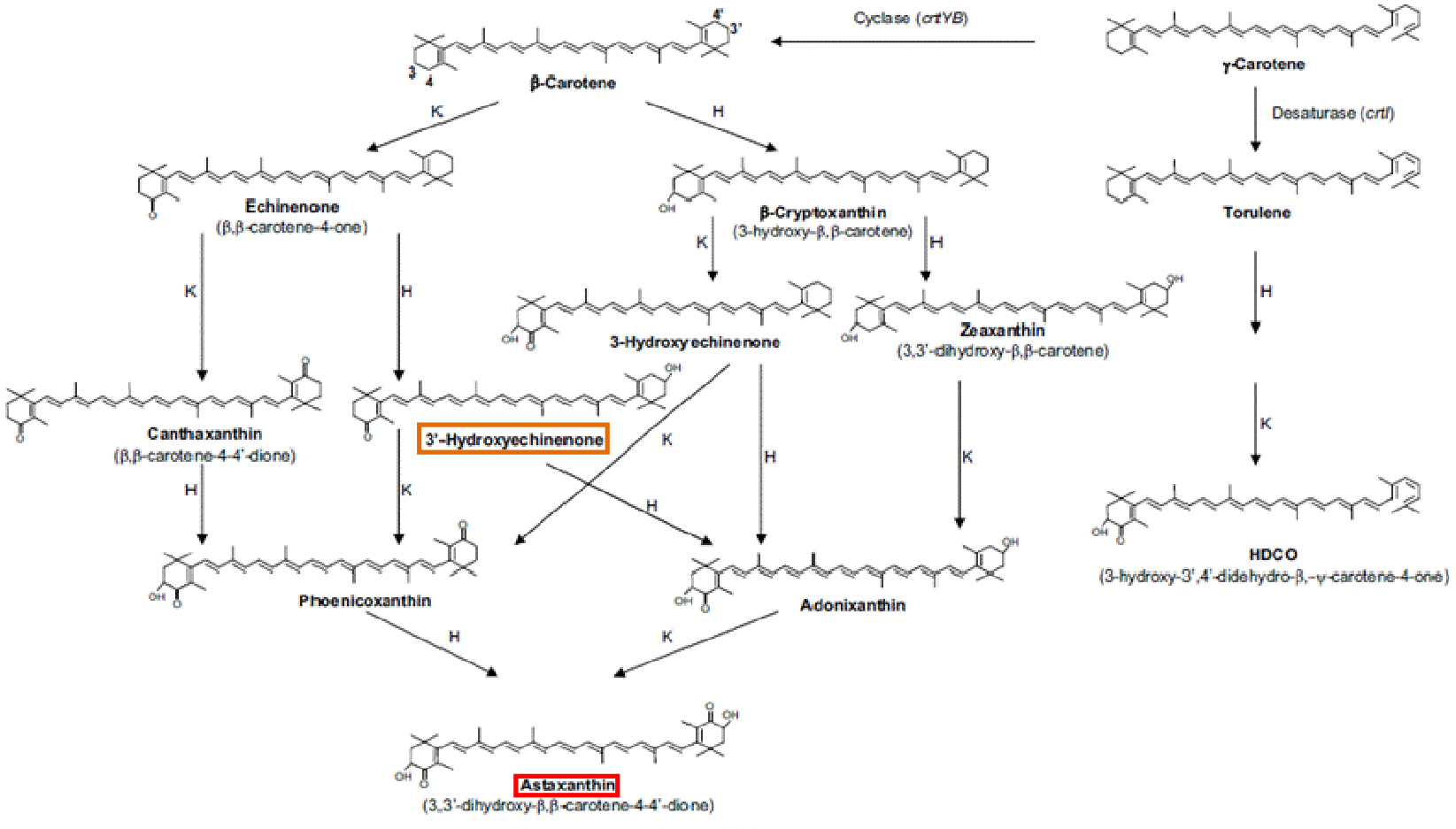


Fig. 3.20 Astanxanthin synthesis with the aid of a ketolase (K) and a hydroxylase (H). The ketocarotenoids, 3-Hydroxyechinenone and astanxanthin, identified in *H. salinarum* are indicated with rectangles (Martín *et al*., 2008).

3.5 Future research

- Heterologous expression of CYP450s from halophilic archaea in *E*. *coli* appears to be a non-viable option since the expressed protein misfolds and localizes in the insoluble fraction. Using *E*. *coli* strains that co-express haloarchaeal chaperons such as hsp70 could solve this problem.
- Creating a *CYP174A1* deletion in *Halobacterium* sp. NRC-1 and repeat the microarray experiment to ensure a complete picture of the transcriptome.
- Replacing the deleted *CYP174A1* either with *CYP174A1* or with the *CYP174A2* homologue from *Haloarcula* sp. LK-1 to confirm their roles in bacterioopsin- and purple membrane biosynthesis. This should ideally be performed with a vector such as pNP22 containing a constitutive promoter *e*.*g*. *Fdx*-promoter.
- Measure β-carotene and retinal levels in both the wildtype and deletion strains with the aid of High Performance Liquid Chromotography (HPLC) and Mass Spectrometry (MS).
- Perform Thin Layer Chromotography (TLC) and/or LC-MS on pigments extracted from *the* Δ*CYP174A1* and wildtype *H*. *salinarum* R1 strains to compare pigment levels and especially to confirm the possible absence of astaxanthin and perhaps 3-hydroxy-echinenone (Asker *et al*., 2002) in the deletion strain.
- Ultimately the Δ*CYP174A1* strain created in this study can used to clone and express the remainder of the halophlic archaeal CYP450s of which the physiological roles remain unknown, especially the CYP450s belonging to the CYP174A-family.

Literature cited

Ahn, T., Yang, S. and Yun, C-H. (2004). Enhanced expression of human cytochrome P450 1A2 by co-expression with human molecular chaperone Hsp70. *Toxicology Letters* **153**, 267 – 272.

Álvarez, V., Rodríguez-Sáiz, M., de la Fuente, J.L., Gudiña, E.J. Godio, R.P., Martín, J.F. and Barredo, J-L. (2006). The *crtS* gene of *Xanthophyllomyces dendrorhous* encodes a novel cytochrome-P450 hydroxylase involved in the conversion of β-carotene into astaxanthin and other xanthophylls. *Fungal Genetics and Biology* **43**, 261 – 272.

Asker, D., Awad, T and Ohta, Y. (2002). Lipids of *Haloferax alexandrinus* strain TM^T: an extremely halophilic canthaxanthin-producing archaeon. Journal of *Bioscience and Bioengineering* **93**, 37 – 43.

Baliga, N.S., Bonneau, R., Facciotti, M.T., Pan, M., Glusman, G., Deutsch, E.W., Shannon, P., Chiu, Y., Weng, R.S., Gan, R.R., Hung, P., Date, S.V., Marcotte, E., Hood, L. and Ng, W.V. (2004). Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Research* **14**, 2221 – 2234.

Baliga, N.S., Kennedy, S.P., Ng, W.V., Hood, L. and DasSarma, S. (2001). Genomic and genetic dissection of an archaeal regulon. *Proceedigns of the National Academy of Sciences USA* **98**, 2521 – 2525.

Baliga, N.S., Pan, M., Goo, Y.A., Yi, E.C., Goodlett, D.R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W.V. and Hood, L. (2002). Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach. *Prodeedings of the National Academy of Sciences USA* **99**, 14913 – 14918.

Bertrand, J.C., Almallah, M., Acquaviva, M. and Mille, G. **(1990)**. Biodegradation of hydrocarbons by an extremely halophilic archaebacterium. *Letters in Applied Microbiology* **11**, 260-263.

Betlach, N.S., Friedman, J., Boyer, H.W. and Pfeiffer, F. (1984). Characterization of a halobacterial gene affecting bacterio-opsin gene expression. *Nucleic Acid Research* **12**, 7949 – 7959.

Calo, P., de Miguel, T., Sieiro, C., Velazquez, J.B. and Villa, T.G. (1995). Ketocarotenoids in halobacteria: 3-hydroxy echinenone and trans-astaxantin. *Journal of Applied Bacteriology* **79**, 282 – 285.

Cline, S.W. and Doolittle, W.F. (1987). Efficient transfection of the archaebacterium *Halobacterium halobium*. *Journal of Bacteriology* **169**, 1341 – 1344.

Coker, J.A., DasSarma, P., Kumar, J., Müller, J.A. and DasSarma, S. (2007). Transcriptional profiling of the model archaeon *Halobacterium* sp. NRC-1: responses to changes in salinity and temperature. *Saline Systems* **3**, 1 – 17.

Connaris, H., Chaudhuri, J.B., Dason, M.J. and Hough, D.W. (1999). Expression, reactivation, and purification of enzymes from *Haloferax volcanii* in *Escherichia coli*. *Biotechnology and Bioengineering* **64**, 38 – 45.

DasSarma, S., Rajbhandary, U.L. and Khorana, H.G. (1983). High-frequency spontaneous mutation in the bacterio-opsin gene in *Halobacterium halobium* is mediated by transposable elements. *Proceedings of the National Academy of Sciences USA* **80**, 2201 – 2205.

DasSarma, S. Ph.D. thesis, **(1984)**. The bacterio-opsin gene in *Halobacterium halobium*: high-frequency inactivation by insertion sequences and structure of the messenger RNA. Massachussetts Institute of Technology, Cambridge, USA.

DasSarma, S., Halladay, J.T., Jones, J.G., Donovan, J.W., Giannasca, P.J. and Tandeau de Marsac, N. (1988). High frequency mutations in a plasmid-encoded gas vesicle gene in *Halobacterium halobium*. *Proceedings of the National Academy of Sciences USA* **85**, 6861 – 6865.

DasSarma, S. (1989). Mechanisms of genetic variability in *Halobacterium halobium*: the purple membrane and gas vesicle mutations. *Canadian Journal of Microbiology* **35**, 65-72.

DasSarma, S. (1995). Natural plasmids and plasmid vectors of halophiles. In: DasSarma S, Fleischmann EM, editor. *Halophiles.* Plainview, NY: Cold Spring Harbor Laboratory Press. pp. 243–250.

DasSarma, P. and DasSarma, S. (2008). On the origin of prokaryotic "species": the taxonomy of halophilic *Archaea*. *Saline Systems* **4**, 1 – 5.

DasSarma, S.L., Capes, M.D., DasSarma, P. and DasSarma, S. (2010). HaloWeb: the haloarchaeal genomes database. *Saline Systems* **6**, 1 – 4.

Danon, A., Brith-Lindner, M. and Caplan, S.R. (1977). Biogenesis of the purple membrane of *Halobacterium halobium*. *Biophysics of Structure and Mechanism* **3**, 1 – 17.

del Rosario, R.C., Staudinger, W.F., Streif, S., Pfeiffer, F., Mendoza, E. and Oesterhelt, D. (2007). Modelling the $CheY^{D10K,Y100W}$ *Halobacterium salinarum* mutant: sensitivity analysis allows choice of parameter to be modified in the phototaxis model. *Institute of Engineering and Technology Systems Biology* **1**, 207 – 221.

Delseny, M., Han, B. and Hsing, Y.I. (2010). High throughput DNA sequencing: The new sequencing revolution. *Plant Science* **179**, 407 – 422.

D'Souza, S.E., Altekar, W., and D'Souza, S.F. (1997). Adaptive response of *Haloferax mediterranei* to low concentrations of NaCl (<20%) in the growth medium. *Archives in Microbiol*ogy **168**, 68–71.

Ebert, K., Hanke, C., Delius, H., Goebel, W. and Pfeiffer, F. (1987). A new insertion element, ISH26, from *Halobacerium halobium*. *Molecular and General Genetics* **206**, 81 – 87.

El-Sayed, W.S.M., Takaichi, S., Saida, H., Kamekura, M., Abu-Shady, M., Seki, H. and Kuwabara, T. (2002). Effects of light and low oxygen tension on pigment biosynthesis in *Halobacterium salinarum*, revealed by a novel method to quantify both retinal and carotenoids. *Plant and Cell Physiology* **43**, 379 – 383.

Facciotti, M.T., Pang, W.L., Lo, F., Whitehead, K., Koide, T., Masumura, K., Pan, M., Kaur, A., Larsen, D.J., Reiss, D.J., Hoang, L., Kalisiak, E., Northen, T., Trauger, S.A., Siuzdak, G. and Baliga, N.S. (2010). Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis. *BMC Systems Biology* **4**, 64 – 82.

Fairbanks, G., T.L. Steck, D.F.H. Wallach. **(1971)**. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* **10**, 2606 -2617.

Fang, C-J., Ku, K-L., Lee, M-H. and Su, N-W. (2010). Influence of nutritive factors on C_{50} carotenoids production by *Haloferax mediterranei* ATCC 33500 with two-stage cultivation. *Bioresource Technology* **101**, 6487 – 6493.

Fickers, P., Benetti, P-H, Waché, Y. Marty, A., Mauersberger, S., Smit, M.S. and Nicaud, J-M. (2005). Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *FEMS Yeast Research* **5**, 527 - 543.

Gropp, F. and Betlach, M.C. (1994). The *bat* gene of *Halobacterium halobium* encodes a trans-acting oxygen inducibility factor. *Biochemistry* **91**, 5475 – 5479.

Grosjean, H., Gaspin, C., Marck, C., Decatur, W.A.and de Crécy-Lagard, V. (2008). RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC Genomics* **9**, 470 – 496.

Grund, A., Shapiro, J., Fennewald, M., Bacha, P., Leahy, J., Markbreiter, K., Nieder, M., and Toepfer, M. (1975). Regulation of alkane oxidation in *Pseudomonas putida*. *J. Bacteriol.* **123**, 546 - 556.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557–580.

Harder, D.R., Lange, A.R., Gebremedhin, D., Birks E.K. and Roman, R.J. (1997). Cytochrome P450 metabolites of arachidonic acid as intracellular signaling molecules in vascular tissue. *Journal Vascular Research* **34**, 237 – 243.

Ho, W.W., Li H., Nishida, C.R., de Montellano, P.R. and Poulos, T.L. (2008). Crystal structure and properties of CYP231A2 from the thermoacidophilic archaeon *Picrophilus torridus*. *Biochemistry* **47**, 2071-2079.

Højberg, O., U. Schnider, H. V. Winteler, J. Sørensen, and D. Haas. (1999). Oxygen-sensing reporter strain of *Pseudomonas fluorescens* for monitoring the distribution of low-oxygen habitats in soil. *Applied and Environmental Microbiology* **65,** 4085 - 4093.

Ishibashi, M., Tokunagaa, H., Hiratsukaa, K., Yonezawaa, Y., Tsurumarua, H., Arakawab, T. and Tokunagaa, M. (2001). NaCl-activated nucleoside diphosphate kinase from extremely halophilic archaeon, *Halobacterium salinarum*, maintains native conformation without salt. *FEBS Letters* **493**, 134 – 138.

Jurgenson, C.T., Ealick, S.E., and Begley, T.P. (2009). Biosynthesis of thiamine pyrophosphate, in EcolSal – *Escherichia coli* and *Salmonella*: cellular and molecular biology (Böck, A., Curtiss III, R., Kaper, J. B., Neidhardt, F. C., Nyström, T., Rudd, K. E., and Squires, C. L., Eds.), ASM Press, Washington, D.C.

Kamekura, M, Seno, Y and Tomioka, H. (1998). Detection and expression of a gene encoding a new bacteriorhodopsin from an extreme halophile strain HT (JCM 9743) which does not possess bacteriorhodopsin activity. *Extremophiles* **2,** 33 – 40.

Kim, Y-S, Kim, N-H, Yeom, S-J, Kim, S-W and Oh, D-K (2009). *In vitro* characterization of a recombinant Blh protein from an uncultured marine bacterium as a β-carotene 15,15′-dioxygenase. *Journal of Biological Chemistry* **284**, 15781 – 15793.

Klingenberg, M. (1958). Pigments of rat liver microsomes. *Archives in Biochemistry and Biophysics* **75**, 376 – 386.

Koch, M.K. and Oesterhelt, D. (2005). MpcT is the transducer for membrane potential changes in *Halobacterium salinarum*. *Molecular Microbiology* **55**, 1681 – 1694.

Kulichevskaya, I.S., Milekhina, E.I., Borzenkov, I.A., Zvyagintseva, I.S. and Belyaev, S.S. (1991). Oxidation of petroleum hydrocarbons by extremely halophilic archaebacteria. *Mikrobiologiya* **60**, 860-866.

Labuschagne, M. and Albertyn, J. (2007). Cloning of an epoxide hydrolaseencoding gene from *Rhodotorula mucilaginosa* and functional expression in *Yarrowia lipolytica. Yeast*. **24**, 69-78.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

Leong, D., Pfeiffer, F., Boyer, H. and Betlach, M. (1988). Characterization of a second gene involved in bacterio-opsin gene expression in a halophilic archaebacterium. *Journal of Bacteriology* **170**, 4903 – 4909.

Macario, A.J.L. and Conway de Macario, E. (1999). The archaeal molecular chaperone machine: peculiarities and paradoxes. *Genetics* **152**, 1277 – 1283.

Makino, T., Harada, H., Ikenaga, H., Matsuda, S., Takaichi, S., Shindo, K., Sandmann, G., Ogata, T. and Misawa, N. (2008). Characterization of cyanobacterial carotenoid ketolase *crtW* and hydroxylase *ctrR* by complementation analysis in *Escherichia coli*. *Plant and Cellular Physiology* **49**, 1867 – 1878.

Martín,J.F., Gudiña, E. and Barredo, J.L. (2008). Conversion of β-carotene into astaxanthin: Two separate enzymes or a bifunctional hydroxylase-ketolase protein? *Microbial Cell Factories* **7**:3 (page numbers not citation purposes).

Müller, J.A. and DasSarma, S. (2005). Genomic analyses of anaerobic respiration in the archaeon *Halobacterium* sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. *Journal of Bacteriology* **187**, 1659 – 1667.

Nelson, D. (2009). The cytochrome P450 homepage. *Human Genomics* **4**, 59 – 65.

Ng, W.V., Kennedy, S.P., Mahairas, C.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Pohlschroder, M., Spudich, J.L., Jung, K.W., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proceedings of National Academy of Sciences USA* **97**, 12176 – 12181.

Nishida, C.R. and Ortiz de Montellano, P.R. (2005). Thermophilic cytochrome P450 enzymes. *Biochemical and Biophysical Research Communications* **338**, 437–445.

Oesterhelt, D. and Krippahl, G. (1983). Phototrophic growth of halobacteria and its use for isolation of photosynthetically deficient mutants. *Annals of Microbiol*, **134B**, 137–150.

Oku, Y., Ohtaki, A., Kamitori, S., Nakamura, N., Yohda, M., Ohno, H. and Kawarabayashi, Y. (2004). Structure and direct electrochemistry of cytochrome P450 from the thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7, *Journal of Inorganic Biochemistry* **98**, 1194–1199.

Oren, A., Ventosa, A. and Grants W.D. (1997). Proposed minimal standards for description of new taxa in the order *Halobacteriales*. *International Journal of Systematic Bacteriology* **47**, 233 – 238.

Ovchinnikov, Y.A., Zozulya, S.A., Zaitseva, F.M., Gurev, S.O., Sverdlov, F.D., Krupenko, M.A. and Aleksandrov, A.A. (1984). The new insertion sequence element of *Halobacterium halobium* localized within the bacterio-opsin gene. *Bioorganicheskaia Khimiia* **10**, 560 – 563.

Patenge, N., Haase, A., Bolhuis, H. and Oesterhelt, D. (2000). The gene for a halophilic β-galactosidase (*bgaH*) of *Haloferax alicantei* as a reporter gene for promoter analyses in *Halobacterium salinarum*. *Molecular Microbiology* **36**, 105 – 113.

Peck, R.F., DasSarma, S. and Krebs, M.P. (2000). Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. *Molecular Microbiology* **35**, 667 – 676.

Peck, R.F., Echavarri-Erasun, C., Johnson, E.A., Ng, V.W., Kenney, S.P., Hood, L., DasSarma, S. and Krebs, M. (2001). *brp* and *blh* are required for synthesis of the retinal cofactor of bacteriorhdopsin in *Halobacterium salinarum*. *The Journal of Biological Chemistry* **276**, 5739 – 5744.

Peck, R.F., Johnson, E.A. and Krebs, M. (2002). Identification of a lycopene β - cyclase required for bacteriorhodopsin biogenesis in the archaeon *Halobacteium salinarum*. *Journal of Bacteriology* **184**, 2889 – 2897.

Pfeiffer, F., Friedman, J., Boyer, H.W. and Betlach, M. (1984). Characterization of insertions affecting the expression of the bacterio-opsin gene in *Halobacterium halobium*. *Nucleic Acids Research* **12**, 2489 – 2497.

Pfeiffer, F. and Betlach, M. (1985). Genome organization in *Halobacterium halobium*: a 70 kb island of more (AT) rich DNA in the chromosome. *Molecular and General Genetics* **198**, 449 – 455.

Pfeiffer, F., Schuster, S.C., Broicher, A., Falb, M., Palm, P., Rodewald, K., Ruepp, A., Soppa, J., Tittor, J., D. Oesterhelt, D. (2008). Evolution in the laboratory: The genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1 *Genomics* **91**, 335 – 346.

Pire, C., Esclapez, J., Ferrer, J. And Bonete, M-J. (2001). Heterologous overexpression of glucose dehydrogenase from the halophilic archaeon *Haloferax mediterranei*, an enzyme of the medium chain dehydrogenase/reductase family. *FEMS Microbiology Letters* **200**, 221 – 227.

Puchkaev, A.V. and Ortiz de Montellano, P.R. (2005). The *Sulfolobus solfataricus* electron donor partners of thermophilic CYP119. An unusual non-NAD(P)H-dependent cytochrome P450 system, *Archives of Biochemistry and Biophysics* **434**, 169–177.

Raghavan, T.M. & Furtado, I. (2000). Tolerance of an estuarine halophilic archaebacaterium to crude oil and constituent hydrocarbons. *Bulletin of Environmental Contamination and Toxicology* **65**, 725-731.

Rincón, B., Raposo, F., Borja, R., González, J.M., Portillo, M.C. and Saiz-Jiménez, C. (2006). Performance and microbial communities of a continious stirred tank anaerobic reactor treating two-phases olive mill solid wastes at low organic loading rates. *Journal of Biotechnology* **121**, 534 – 543.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. and Gelfand, M.S. (2002). Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *Journal of Biological Chemistry* **277**, 48949 - 48959.

Sabirova, J., Ferrer, M., Regenhardt, D., Timmis, K.N. and Golyshin, P.N. (2006). Proteomic insights into metabolic adaptions in *Alcanivorax borkumensis* induced by alkane utilization. *Journal of Bacteriology* **188**, 3763 – 3773.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning – A laboratory manual, 2nd ed., Edited by C. North, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Schoefs, B., Rmiki, N-E, Rachadi, J. and Lemoine Y. (2001). Astaxanthin accumulation in *Haematococcus* requires a cytochrome P450 hydroxylase and an active synthesis of fatty acids. *FEBS Letters* **500**, 125 – 128.

Shand, R.F. and Betlach, M.C. (1991). Expression of the *bop* gene cluster of Halobacterium halobium is induced by low oxygen tension and by light. *Journal of Bacteriology* **173**, 4692 – 4699.

Smits, T.H.M., Seeger, M.A., Witholt, B. and van Beilen, J.B. (2001). New alkane-responsive expression vectors for *Escherichia coli* and *Pseudomonas*. *Plasmid* **46**, 16-24.

Smits, T.H.M., Balada, S.B., Witholt, B. and van Beilen, J.B. (2002). Functional analysis of alkane hydroxylases from gram-negative and gram-positive bacteria. *Journal of Bacteriology* **184**, 1733 – 1742.

Stephens, L.L., Shonhai, A., and Blatch, G.L. (2011). Co-expression of the *Plasmodium falciparum* molecular chaperone, PfHsp70, improves the heterologous production of the antimalarial drug target GTP cyclohydrolase I, PfGCHI. *Protein Expression and Purification* **77**, 159 – 165.

Stoeckenius, W and Kunau, W.H. (1968). Further characterization of particulate fractions from lysed cell envelopes of *Halobacterium halobium* and isolation of gas vacuole membranes. *Journal of Cell Biology* **38**, 337 – 357.

Strahl, H. and Greie, J.C. (2008). The extremely halophilic archaeon *Halobacterium salinarum* R1 responds to potassium limitation by expression of the K^+-transporting KdpFABC P-type ATPase and by a decrease in intracellular K^+. *Extremophiles* **12**, 741 – 752.

Sumper, M. and Herrmann, G. (1976). Biosynthesis of purple membrane: control of retinal synthesis by bacterio-opsin. *FEBS Letters* **71**, 333 – 336.

Tarasov, V.Y., Besir, H., Schwaiger, R., Klee, K., Furtwängler, K., Pfeiffer, F. and Oesterhelt, D. (2008). A small protein from the *bop-brp* intergenic region of *Halobacterium salinarum* contains a zinc finger motif and regulates *bop* and *crtB1* transcription. *Molecular Microbiology* **67**, 772 – 780.

Twellmeyer, J., Wende, A., Wolfertz, J., Pfeiffer, F., Panhuysen, M., Zaigler, A., Soppa, J., Welzl, G. and Oesterhelt, D. (2007). Microarray analysis in the archaeon *Halobacterium salinarum* R1. *PloS ONE* **10**, 1 – 9.

Uroz, S., Chhabra, S.R., Cámara, M., Williams, P., Oger, P. and Dessaux, Y. (2005). N-Acylhomoserine lactone quorum-sensing molecules are modified and degraded by *Rhodococcus erythropolis* W2 by both amidolytic and novel oxidoreductase activities. *Microbiology* **151**, 3313 – 3322.

van Beilen, J.B., Kingma, J. and Witholt, B. (1994). Substrate specificity of the alkane hydroxylase of *Pseudomonas oleovorans* GPo1. *Enzyme and Microbial Technology* **16,** 904 – 911.

von Lintig, J. and Vogt, K. (2000). Filling the gap in vitamin A research. Molecular identification of an enzyme cleaving β-carotene to retinal. *Journal of Biological Chemistry* **275**, 388 – 400.

Werck-Reichhart, D. and Feyereisen, R. (2000). Cytochrome P450: a success story. *Genome Biology* **1**, 3003.1 – 3003.9.

Wright, R.L., Harris, K., Solow, B., White, R.H. and Kennelly, P.J. (1996). Cloning of a potential cytochrome P450 from the archaeon *Sulfolobus solfataricus*. *FEBS Letters* **384**, 235–239.

Yang, C.F. and DasSarma, S. (1990). Transcriptional induction of purple membrane and gas vesicle synthesis in the archaebacterium *Halobacterium halobium* is blocked by a DNA gyrase inhibitor. *Journal of Bacteriology* **172**, 4118 – 4121.

Yano, J.K., Koo, L.S., Schuller, D.J., Li, H., Ortiz de Montellano, P.R. and Poulos, T.L. (2000). Crystal structure of a thermophilic cytochrome P450 from the archaeon *Sulfolobus solfataricus*. Journal of Biological Chemistry **275**, 31086–31092.

Yonezawa, Y., Tokunaga, H., Ishibashi, M., Taura, S. and Tokunaga, M. (2003). Cloning, expression, and efficient purification in *Escherichia coli* of a halophilic nucleoside diphosphate kinase from the moderate halophile *Halomonas* sp. #593. *Protein Expression and Purification* **27**, 128–133.

Zvyagintseva, I.S., Belyaev, S.S., Borzenkov, I.A., Kostrikina, N.A., Milekhina, E.I. and Ivanov, M. (1995). Halophilic archaebacterium from the kalamkass oil field. *Mikrobiologya* **64**, 83-87.

Chapter 4

Concluding remarks

At a glance the genera *Thermus* and *Halobacterium* do not have much in common. The contrasts are quite stark – members of *Thermus* are thermophilic bacteria while members of *Halobacterium* are extremely halophilic archaea. In spite of these differences on phylogenetic and phenotypic level, there is one trait that these genera share and which has been a common theme throughout this study: production of isoprenoids and specifically carotenoid pigments.

Yellow pigments in *Thermus* (specifically β-cryptoxanthin and zeaxanthin) and red pigments from *Halobacterium* (bacterioruberin and astaxanthin) both play a part in protecting the cells of these organisms against oxidative stress and preventing DNA damage due to UV-radiation. It is also interesting to note that the abovementioned oxygenated pigments (or xantophylls) all require substrates that are downstream products of the isoprenoid biosynthesis pathway: namely lycopene and β-carotene. β-cryptoxanthin, zeaxanthin and astaxanthin are synthesized from β-carotene which in turn is synthesized from lycopene while bacterioruberin is directly synthesized from lycopene.

Cytochrome P450 monooxygenases (CYP450s) have been implicated in the biosynthesis of pigments *e.g*. astaxanthin production in the heterobasidiomycetous yeast *Xanthophyllomyces dendrorhous* (Álvarez *et al*., 2006) and the green algae *Haematococcus pluvialis* (Schoefs *et al*., 2001), and β-cryptoxanthin and zeaxanthin production by CYP175A1 from *Thermus thermophilus* HB27 (Blasco *et al*., 2004). Also the CYP175A from *Thermus* sp. NMX2.A1, which we investigated in Chapter 2, is strongly implicated in pigment synthesis, since the closely related non-pigmented *Thermus scotoductus* SA-01 lacks a CYP450 (CYP175).

The presence of a hydroxylated carotenoids including trans-astaxanthin reported in *inter alia Halobacterium salinarum* by Calo *et al*. (1995), have been demonstrated in many different halophilic archaea, although no genes encoding typical β-carotene hydroxylases or ketolases have been identified in the sequenced genomes of the *Halobacteriaceae*. CYP450s in halophilic archaea must have an important physiological role when one considers that all sequenced genomes of halophilic archaea contain at least one CYP450 and with one of them, *Haladaptatus paucihalophilus,* having nine different CYP450s in its

genome. Based on the effect that the CYP174A1 deletion in *Halobacterium salinarum* R1 had on its pigment synthesis – a noticeable increase in bacterioruberin levels and abolishment of purple membrane synthesis, it seems very likely that CYP174A1 plays a role in pigment metabolism in *H*. *salinarum* R1 and possibly other halophilic archaea having CYP450s from the same CYP450-protein family. It might even be possible that CYP174A1 is a bifunctional enzyme with both hydroxylase and ketolase activity that can alone convert β-carotene into astaxanthin. Such a bifunctional activity had at one stage been proposed for the β-carotene hydroxylating CYP450 from *Xanthophyllomyces dendrorhous* (Martín *et al*., 2008).

The fact that isoprenoid and carotenoid biosynthetic pathways are ancient pathways that occur in all three domains of life (Umeno *et al*., 2005; Lombard & Moreira, 2011), makes the possible involvement of CYP450s from extremophiles in the hydroxylation of these ancient metabolites particularly intriguing. Even more intriguing is the possibility that such hydroxylated carotenoids (or isoprenoids) might have a gene regulation (signaling) function in the halophilic archaea, similar to the signaling function of hydroxylated fatty acid or steroid hormones in eukaryotes. CYP450s play a very important role in creating a very large diversity of these chemical messengers in eukaryotes.

Literature cited

Álvarez, V., Rodríguez-Sáiz, M., de la Fuente, J.L., Gudiña, E.J. Godio, R.P., Martín, J.F. and Barredo, J-L. (2006). The *crtS* gene of *Xanthophyllomyces dendrorhous* encodes a novel cytochrome-P450 hydroxylase involved in the conversion of β-carotene into astaxanthin and other xanthophylls. *Fungal Genetics and Biology* **43**, 261 – 272.

Blasco, F., Kauffmann, I. and Schmid, R.D. (2004). CYP175A1 from *Thermus thermophilus* HB27, the first β-carotene hydroxylase of the P450 superfamily. *Applied Microbiology and Biotechnology* **64**, 671 - 674.

Calo, P., de Miguel, T., Sieiro, C., Velazquez, J.B. and Villa, T.G. (1995). Ketocarotenoids in halobacteria: 3-hydroxy echinenone and trans-astaxantin. *Journal of Applied Bacteriology* **79**, 282 – 285.

Lombard, J. and Moreira, D. (2011). Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Molecular Biology and Evolution* **28**, 87 – 99.

Martín,J.F., Gudiña, E. and Barredo, J.L. (2008). Conversion of β-carotene into astaxanthin: Two separate enzymes or a bifunctional hydroxylase-ketolase protein? *Microbial Cell Factories* **7**:3 (page numbers not citation purposes).

Schoefs, B., Rmiki, N-E, Rachadi, J. and Lemoine Y. (2001). Astaxanthin accumulation in *Haematococcus* requires a cytochrome P450 hydroxylase and an active synthesis of fatty acids. *FEBS Letters* **500**, 125 – 128.

Umeno, D., Tobias, A.V. and Arnold, F.H. (2005). Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiology and Molecular Biology Reviews* **69**, 51 – 78.

Summary / Opsomming

Only five CYP450s from extremophiles have been studied, while genome sequence information indicate that CYP450s are prevalent in members of the bacterial phylum *Deinococcus-Thermus* as well as the archaeal family *Halobacteriaceae* that belong to the phyulm *Euryarchaeota.* A property shared by these phylogenetically distant extremophiles is the production of carotenoid pigments. It became the purpose of this study to use genome sequence information to clone and study new CYP450s from the genera *Thermus* and *Halobacterium* and to explore the role of these CYP450s in pigment production.

The non-pigmented thermophilic bacterium *Thermus scotoductus* SA-01 was screened by PCR for the presence of a cytochrome P450 monooxygenase (CYP450). No CYP450 could be found and subsequent genome sequencing confirmed this finding. However, a CYP450 gene (*CYP175A*) was isolated from the closely related yellow pigmented strain *Thermus* sp. NMX2.A1 using oligonucleotides based on the DNA sequence of the β-carotene gene cluster from three *Thermus* strains.

The genome sequence of *T*. *scotoductus* SA-01, revealed a ferredoxin (*Fdx*) and ferredoxin reductase (*FNR*) that were almost identical to those of *Thermus thermophilus* HB27. In *T*. *thermophilus* HB27 the Fdx and FNR are the native redox partners for CYP175A1, a β-carotene hydroxylase. After heterologous expression in *Escherichia coli*, we attempted to hydroxylate β-carotene with the CYP450 from *Thermus* sp. NMX2.A1 and the redox partners of *T*. *scotoductus* SA-01 using cell free extracts, but no products were detected.

Thirty two CYP450s have been identified in the sequenced genomes of thirteen extremely halophilic archaea. Initial attempts to clone and heterologously express a *CYP174A2*-homologue from a *Haloarcula* LK-1 strain in *E. coli* and *Pseudomonas fluorescens* were unsuccessful. In order to study the physiological role of CYP450s in halophilic archaea and to create a strain that can be used for heterologous expression of CYP450s from halophiles *CYP174A1* was deleted from *H. salinarum* R1. *CYP174A1* is the only CYP450 in *H*. *salinarum* R1 and *H*. *salinarum* R1 is a genetically tractable strain.

Upon culturing the wildtype and deletion strains, a difference in red pigmentation of stationary phase cultures was observed; implying that CYP174A1 might play a role in

carotenoid synthesis. Microarray analyses revealed that the *bop* gene, which codes for bacterioopsin (BO) was severely repressed in stationary phase cultures of the deletion strain and sucrose gradient experiments showed a consequent loss of purple membrane (PM) in the deletion strain. The classical causes of *bop* repression *e.g*. insertion elements in the *bop* open reading frame as well as in the *brz* gene was ruled out by PCR screening. In addition to *bop* repression, the neighboring vng1459 and vng1468 genes (both part of the *bop*-regulon) were also down regulated, but the genes normally involved in regulation of the *bop* gene were not affected. Currently the functions of vng1459 and vng1468 are unknown.

Retinal, together with BO, is a key component of bacteriorhodopsin (BR) and essential for PM synthesis. Retinal is formed by the central cleavage of β-carotene which can be achieved by monooxygenases or dioxygenases.The Blh and Brp proteins in *H. bacterium salinarum* are very closely related to a confirmed bacterial 15,15′-β-carotene dioxygenase and studies have shown that deletion of both *brp* and *blh* results in complete abolishment of retinal and BR. It is therefore unlikely that CYP174A1 plays a role in retinal biosynthesis. Another possible function for CYP174A1 might be the hydroxylation of β-carotene, since it is known that *H. salinarum* strains produce hydroxylated carotenoids such as *trans*-astaxanthin, but no genes encoding typical β-carotene hydroxylases or ketolases have been identified in the genomes of *H. salinarum* strains. This will imply that hydroxylated carotenoids play a role in the regulation of *bop*.

Keywords: *CYP175A1*, ferredoxin, ferredoxin reductase, β-carotene, *CYP174A1*, deletion, microarray, purple membrane.

Tot op hede is nog net vyf CYP450s vanuit ekstremofiele bestudeer ondanks die feit dat genoomvolgordebepaling data aandui dat CYP450s volop teenwoordig is in lede van die bakteriese filum *Deinococcus-Thermus* sowel as die archaea familie *Halobacteriaceae* wat behoort aan die filum *Euryarchaeota*. Hierdie twee ver vewante filogenetiese ekstremofiele het wel een fenotipe in gemeen naamlik: die sintese van karoteen pigmente. Die doelwit van hierdie studie was om nuwe CYP450s van die *Thermus* en *Halobacterium* genera te kloneer en bestudeer en sodoende hulle rol in pigment metabolism te ondersoek.

Die pigmentlose, termofiliese bakterium *Thermus scotoductus* SA-01 was ge-evalueer vir die teenwoordigheid van ‘n sitokroom P450 mono-oksigenase (CYP450) deur middel van PKR analise. Geen CYP450 kon gevind word nie en die daaropvolgende genoomvolgordebepaling het hierdie bevinding bevestig. Daar was egter 'n CYP450 geen (*CYP175A*) geïsoleer vanuit ‘n geneties verwante geel gepigmenteerde *Thermus* sp. NMX2.A1 stam m.b.v. priem stukke wat se DNS-volgorde gebaseer was op dié van ‘n β-karoteen geen groepering uit drie verskillende *Thermus* stamme. Vanaf die genoom-volgorde van *T*. *scotoductus* SA-01, was daar 'n ferredoksien (*Fdx*) en ferredoksien reduktase (*FNR*) geïdentifiseer wat byna identies was aan die *Fdx* en *FNR* van *Thermus thermophilus* HB27. In *T*. *thermophilus* HB27 is Fdx en FNR die wilde tipe redoks-paar vir CYP175A1 ('n β-karoteen hydroksilase). Die CYP450 van *Thermus* sp. NMX2.A1 tesame met die redoks-paar van *T*. *scotoductus* SA-01 was heteroloog uitgedruk in *Escherichia coli* en met behulp van hierdie drie komponente was daar gepoog om β-karoteen te hidroksileer m.b.v. sel-vrye ekstrakte. Hierdie eksperimente was egter onsuksesvol.

Twee-en-dertig CYP450s van dertien ekstreme halofiliese archaea is al geïdentifiseer danksy inligting verkry vanuit heel genoom DNS-volgordebepaling projekte. Tot op hede is nog geen van hierdie CYP450s bestudeer nie. Aanvanklike pogings om ‘n *CYP174A2* geen, vanuit 'n *Haloarcula* LK-1 stam in *E*. *coli* en *Pseudomonas fluorescens* heteroloog uit te druk, was onsuksesvol aangesien die uitgedrukte proteïen verkeerdelik gevou en in die onoplosbare fraksie gelokaliseer het. Pogings om die onoplosbare proteïen te denatureer en hervou, was ook onsuksesvol.

Daar was besluit om eerder 'n CYP450 (*CYP174A1*) delesie in *H*. *salinarum* R1 te maak en dan die delesie stam se transkriptoom te bestudeer m.b.v. ‘microarray’ analise om sodoende die fisiologiese rol van CYP174A1 in halofiliese archaea te bepaal. *CYP174A1* is die enigste CYP450-geen in *H*. *salinarum* en is geneties maklik manipuleerbaar .

Tydens groei eksperimente van die wilde tipe en delesie stamme was daar 'n noemenswaardige verskil in rooi pigmentasie waargeneem tydens die stasionêre fase van groei. Hierdie fenotipe was ‘n aanduiding dat die *CYP174A1*-delesie dalk karoteen metabolimse in *Halobacterium* beïnvloed het. ‘Microarray’ ontleding het getoon dat die *bop* geen, wat kodeer vir bakterio-opsien (BO), erg onderdruk was in die delesie stam en daaropvolgende sukrose-gradiënt eksperimente het hierdie resultaat bevestig aangesien daar ‘n verlies van pers membraan (PM) was in die delesie stam, maar nie in die wilde tipe stam nie. Die algemene oorsake van *bop* onderdrukking, bv. insersie-elemente in die *bop* oop lees raam sowel as in die *brz* geen was uitgeskakel d.m.v. PKR analise. Tesame met *bop*, was twee na-burige gene naamlik vng1459 en vng1468 ook af-waarts gereguleer. Die funksie van vng1459 en vng1468 is onbekend. Gene wat deel vorm van die ‘*bop*-regulon’ wat gewoontlik gereguleer word tydens die stasionêre fase van groei was ongeaffekteerd in die delesie stam.

Retinal tesame met BO, is 'n belangrike komponent van bakteriorodopsien (BR) en gevolglik, PM sintese. Retinal word gevorm deur die sentrale splyting van β-karoteen en die hierdie tipe splyting kan bewerkstellig word deur 'n mono-oksigenase sowel as ‘n di-oksigenase ensiem. In *H. salinarum* R1 is daar 'n Blh proteïen wat baie nou verwant is aan ‘n bakteriële 15,15′-β-karoteen di-oksigenase sowel as 'n Brp proteïen met 'n soortgelyke funksie. Studies het getoon dat die delesie van *blh* geen invloed het op instra-sellulêre retinal vlakke nie. Delesie van *brp* veroorsaak 'n afname in retinal vlakke sowel as BR vlakke, máár die delesie van beide *brp* sowel as *blh* veroorsaak ‘n algehele verlies van retinal en BR en gevolglik PM. Dit is dus onwaarskynlik dat CYP174A1 'n rol speel in retinal biosintese. ’n Ander moontlike funksie van CYP174A1 is die hidroksilering van β-karoteen aangesien dit bekend is dat *H. salinarum* stamme gehidrosileerde pigmente sintetiseer soos bv. *trans*-astaxantin. Tot op hede is die gene wat kodeer vir tipiese β-karoteen hidroksilases of ketolases nog nie in die genome van *H. salinarum* stamme gevind nie. Hierdie dui op ‘n moontlikheid dat hierdie gehidroksileerde karotene ‘n rol kan speel in die regulering van *bop*.

Sleutelwoorde: *CYP175A1*, ferredoksien, ferredoksien reduktase, β-karoteen, *CYP174A1*, delesie, ‘microarray’, pers membraan.

Annexure A – Microarray data (Chapter 3)

Table A.1 Differentially expressed genes on chromosome of *H. salinarum* R1 during exponential and stationary growth

Gene ID	Gene name (NRC-1) [a]	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1	vng1	No prediction available	-1.598	-0.657	0.232	-1.596	-0.642	0.309
2	*yvrO*	Amino acid ABC transporter.ATP-binding protein	-1.603	-0.664	0.221	-1.473	-0.526	0.305
3	vng3	No prediction available	-1.578	-0.624	0.314	*	*	*
11	*hsx*	Probable sugar transferase	2.020	0.724	1.042	2.101	0.951	0.609
18	vng18	No prediction available	1.855	0.769	0.602	*	*	*
19	vng19	No prediction available	1.918	0.929	0.179	1.765	0.802	0.225
22	vng22	No prediction available	1.756	0.707	0.558	*	*	*
24	vng24	No prediction available	1.590	0.652	0.218	*	*	*
27	vng27	No prediction available	2.100	0.777	0.943	1.604	0.643	0.332
28	vng28	No prediction available	2.504	1.178	0.698	1.717	0.710	0.457
29	vng29	No prediction available	3.434	1.385	1.103	1.491	0.533	0.356
30	vng30	No prediction available	1.929	0.798	0.647	*	*	*
38	vng38	No prediction available	1.710	0.679	0.529	1.751	0.639	0.734

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
47	*graD6*	Glucose-1-phosphate thymidylyltransferase	2.230	1.098	0.409	2.242	1.104	0.395
54	vng54	No prediction available	2.196	0.835	0.815	*	*	*
56	vng56	No prediction available	-1.575	-0.553	0.663	*	*	*
59	vng59	No prediction available	-1.530	-0.591	0.251	*	*	*
61	vng61	No prediction available	*	*	*	1.548	0.549	0.493
89	*pimT1*	L-isoaspartyl protein carboxyl methyltransferase	*	*	*	-1.525	-0.554	0.398
91	vng91	No prediction available	1.923	0.881	0.397	*	*	*
97	*hsp2*	Putative heat shock protein	1.450	0.523	0.196	*	*	*
98	*rimK*	Ribosomal protein S6 modification protein	2.251	0.935	0.932	*	*	*
115	*yusZ1*	Oxidoreductase	*	*	*	-2.706	-1.369	0.434
117	vng117	No prediction available	2.550	0.765	1.121	1.643	0.646	0.440
118	vng118	No prediction available	1.484	0.536	0.307	*	*	*
119	vng119	No prediction available	2.024	0.993	0.268	1.586	0.656	0.163

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
120	vng120	No prediction available	2.166	1.060	0.386	1.696	0.700	0.388
140	vng140	No prediction available	2.954	1.133	1.013	1.483	0.531	0.333
141	vng141	No prediction available	1.830	0.573	0.876	*	*	*
143	vng143	No prediction available	1.620	0.593	0.564	*	*	*
152	*prrC*	Regulatory protein	-1.473	-0.525	0.306	*	*	*
153	*mbl*	MreB-like protein	*	*	*	-1.444	-0.503	0.277
157	*oxlT*	Oxalate/formate antiporter	-1.593	-0.557	0.583	*	*	*
172	*mutS1b*	Mismatch repair protein	-1.528	-0.584	0.283	*	*	*
174	*cat1*	Cationic amino acid transporter	-1.461	-0.534	0.190	*	*	*
176	*act1*	Hpa2-like histone acetyltransferase	1.855	0.743	0.627	*	*	*
180	*hop*	Halorhodopsin	1.594	0.606	0.429	*	*	*
192	*ftsZ2*	Cell division protein	*	*	*	-1.472	-0.495	0.449
196	vng196	No prediction available	1.769	0.760	0.412	1.545	0.579	0.363

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
207	vng207	No prediction available	4.537	1.924	0.857	3.835	1.752	0.766
208	vng208	No prediction available	4.103	1.483	1.194	1.888	0.799	0.625
209	*xcd1*	Integrase/recombinase	2.072	0.871	0.716	1.717	0.573	0.749
210	vng210	No prediction available	1.622	0.683	0.210	*	*	*
212	vng212	No prediction available	4.034	1.590	1.132	3.362	1.080	1.269
214	vng214	No prediction available	2.457	0.903	1.003	1.992	0.683	0.956
216	vng216	No prediction available	5.526	2.101	1.154	3.291	1.493	0.897
217	vng217	No prediction available	4.076	1.694	1.006	3.020	1.254	1.054
222	vng222	No prediction available	-1.543	-0.584	0.334	-1.824	-0.820	0.381
226	*htrA*	Serine proteinase	-1.447	-0.508	0.265	*	*	*
227	vng227	No prediction available	-1.462	-0.520	0.285	-1.661	-0.692	0.360
237	*rpoP*	DNA-directed RNA polymerase subunit P	*	*	*	1.429	0.500	0.198
248	vng248	No prediction available	-1.734	-0.772	0.253	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
252	vng252	No prediction available	*	*	*	1.452	0.490	0.380
254	*tfbG*	Transcription initiation factor IIB	*	*	*	-1.475	-0.524	0.331
261	vng261	No prediction available	*	*	*	1.512	0.577	0.246
267	vng267	No prediction available	-1.478	-0.539	0.267	*	*	*
282	vng282	No prediction available	*	*	*	-1.644	-0.671	0.355
283	vng283	No prediction available	*	*	*	-1.505	-0.560	0.295
297	vng297	No prediction available	1.839	0.794	0.513	*	*	*
308	*trpA*	Tryptophan synthase alpha chain	-1.483	-0.509	0.419	*	*	*
312	vng312	No prediction available	1.873	0.888	0.223	1.918	0.911	0.283
327	*gadD*	Glutamate decarboxylase	-1.436	-0.503	0.233	*	*	*
330	*ppsA*	Phosphoenolpyruvate synthase	-1.655	-0.707	0.234	*	*	*
351	vng351	No prediction available	-1.552	-0.606	0.293	*	*	*
361	vng361	No prediction available	-1.441	-0.512	0.211	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
401	*epf2*	mRNA 3'-end processing factor homolog	-1.558	-0.605	0.316	*	*	*
402	vng402	No prediction available	1.870	0.884	0.234	*	*	*
414	*purH*	Phosphoribosylaminoimidazole-succinocarboxamide formyltransferase	*	*	*	-1.508	-0.563	0.285
422	*cyc*	Cytochrome P450 monooxygenase	-4.432	-2.001	0.656	-5.103	-2.263	0.519
430	vng430	No prediction available	-2.115	-1.057	0.267	-1.606	-0.654	0.288
437	*ydaJ*	Putative aminohydrolase	1.733	0.607	0.753	*	*	*
439	vng439	No prediction available	-1.630	-0.689	0.211	*	*	*
451	*phoU*	Transcriptional regulator	-1.469	-0.533	0.251	*	*	*
453	*pstA2*	Phosphate ABC transporter permease	-1.840	-0.856	0.262	-1.644	-0.696	0.253
455	*pstC2*	Phosphate ABC transporter permease	-1.694	-0.723	0.318	*	*	*
457	*phoX*	Phosphate ABC transporter periplasmic phosphate-binding	*	*	*	-1.478	-0.511	0.406
467	*yafB*	Aldehyde reductase	*	*	*	1.477	0.518	0.358
491	*dnaK*	Heat shock protein	*	*	*	-1.503	-0.556	0.296

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
501	*vfmO1*	Multidrug resistance protein homolog	3.164	1.457	0.725	1.561	0.629	0.194
505	vng505	No prediction available	-1.735	-0.720	0.4224	*	*	*
507	vng507	No prediction available	-1.493	-0.550	0.287	*	*	*
509	vng509	No prediction available	1.573	0.594	0.374	*	*	*
520	vng520	No prediction available	1.569	0.634	0.215	*	*	*
525	*yurZ*	ABC transporter.membrane component	*	*	*	-1.877	-0.880	0.281
527	vng527	No prediction available	*	*	*	-1.588	-0.643	0.265
546	vng546	Prediction not available	-1.687	-0.706	0.371	*	*	*
557	vng557	Prediction not available	1.715	0.614	0.707	*	*	*
562	*pnh2*	Putative NADH-ubiquinone oxidoreductase subunit chain L	-1.989	-0.925	0.438	*	*	*
583	*cybB2*	Cytochrome *b* subunit of the bc complex	*	*	*	-1.645	-0.693	0.265
584	*cybA*	Rieske Fe-S protein	*	*	*	-1.901	-0.898	0.285
585	vng585	No prediction available	*	*	*	-1.693	-0.733	0.267

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
586	*cybD*	Membrane subunit of the bc complex	*	*	*	-1.622	-0.638	0.413
594	vng594	No prediction available	-1.695	-0.742	0.236	*	*	*
601	vng601	No prediction available	2.156	1.092	0.210	2.213	1.072	0.406
610	*hhoA*	4-hydroxybenzoate octaprenyltransferase	2.201	0.991	0.634	*	*	*
611	vng611	No prediction available	2.745	1.183	0.914	1.960	0.861	0.575
612	vng612	No prediction available	2.745	1.183	0.914	*	*	*
613	vng613	No prediction available	1.869	0.654	0.849	*	*	*
620	*edp*	Proteinase IV homolog	*	*	*	-1.463	-0.526	0.258
630	*ribE*	Riboflavin synthase beta subunit	*	*	*	-1.468	-0.536	0.223
636	*ndhG1*	NADH dehydrogenase/oxidoreductase	-1.457	-0.518	0.267	-1.759	-0.768	0.355
643	*nolC*	NADH dehydrogenase/oxidoreductase-like protein	-1.481	-0.546	0.247	*	*	*
647	*nuoM*	F420H2:quinone oxidoreductase chain M	-1.516	-0.571	0.284	*	*	*
654	*act3*	Hpa2-like histone acetyltransferase	2.283	1.015	0.666	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
662	*coxC*	Cytochrome *c* oxidase subunit III	-1.454	-0.513	0.271	*	*	*
677	vng677	No prediction available	1.893	0.848	0.465	*	*	*
692	vng692	No prediction available	-1.486	-0.562	0.166	*	*	*
713	vng713	No prediction available	1.628	0.658	0.338	*	*	*
715	*thiC*	Thiamine biosynthesis protein	1.616	0.624	0.426	-6.217	-2.532	0.557
717	vng717	No prediction available	*	*	*	-1.462	-0.530	0.228
741	vng741	No prediction available	-1.424	-0.495	0.203	*	*	*
746	vng746	No prediction available	*	*	*	-1.638	-0.658	0.380
748	*prkA*	Kinase anchor protein	*	*	*	-1.747	-0.777	0.269
749	*prk*	Protein kinase	1.644	0.620	0.553	-1.771	-0.783	0.340
757	*tfeA*	Transcription initiation factor IIE alpha subunit	-1.485	-0.542	0.281	*	*	*
768	vng768	No prediction available	1.455	0.499	0.349	*	*	*
789	vng789	No prediction available	-1.466	-0.539	0.187	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
796	*cgs*	Cystathionine gamma-synthase	*	*	*	1.516	0.573	0.283
812	*htr18*	Htr18 transducer	1.559	0.610	0.288	-1.508	-0.566	0.270
813	*potD*	Spermidine/putrescine-binding protein	*	*	*	-1.441	-0.505	0.255
816	*chi*	Chitinase	1.577	0.502	0.612	*	*	*
822	vng822	No prediction available	-1.529	-0.587	0.275	*	*	*
824	*gdb*	Molybdopterin-guanine dinucleotide biosynthesis protein A	*	*	*	-1.472	-0.503	0.388
826	*dmsR*	Dimethylsulfoxide reductase operon activator	1.505	0.575	0.211	*	*	*
829	*dmsA*	Dimethylsulfoxide reductase	1.704	0.736	0.321	*	*	*
836	vng836	No prediction available	2.086	0.691	0.944	1.666	0.656	0.480
837	vng837	No prediction available	1.430	0.502	0.203	1.424	0.490	0.237
840	vng840	No prediction available	1.810	0.706	0.622	*	*	*
846	vng846	No prediction available	1.628	0.541	0.725	*	*	*
847	vng847	No prediction available	-1.456	-0.519	0.256	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
861	vng861	No prediction available	2.386	1.191	0.440	1.597	0.657	0.239
862	*hisF*	Imidazoleglycerol-phosphate synthase	-1.488	-0.551	0.256	*	*	*
867	*asnA*	Asparagine synthetase	-1.495	-0.556	0.259	*	*	*
885	*top6B*	DNA topoisomerase VI subunit B	-1.488	-0.559	0.201	*	*	*
905	*pmu2*	Phosphomannomutase	-1.482	-0.548	0.230	*	*	*
907	vng907	No prediction available	1.477	0.512	0.352	1.459	0.528	0.217
918	vng918	No prediction available	1.507	0.582	0.168	*	*	*
921	*potA1*	Spermidine/putrescine ABC transporter ATP-binding	-1.763	-0.778	0.335	*	*	*
923	*sfuB*	Iron transporter-like protein	-1.814	-0.840	0.232	*	*	*
937	*gap*	Glyceraldehyde-3-phosphate dehydrogenase	*	*	*	-1.659	-0.697	0.311
949	*gspE3*	Type II secretion system protein	-1.519	-0.590	0.190	*	*	*
950	*fapH*	Flagella-related protein H	*	*	*	-1.525	-0.572	0.318
970	*cheC1*	Chemotaxis protein	-1.525	-0.579	0.297	-1.607	-0.638	0.361

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
971	*cheA*	Chemotaxis protein	*	*	*	-1.500	-0.513	0.441
973	*cheB*	Chemotaxis protein	*	*	*	-1.679	-0.716	0.295
974	*cheY*	Chemotaxis protein	*	*	*	-1.685	-0.720	0.307
979	vng979	No prediction available	2.063	0.586	0.988	*	*	*
985	vng985	No prediction available	2.094	0.751	0.900	*	*	*
989	*xcd*	Integrase/recombinase	1.799	0.768	0.476	1.534	0.574	0.348
990	vng990	No prediction available	4.111	1.773	0.872	2.526	1.188	0.651
993	vng993	No prediction available	1.752	0.736	0.440	*	*	*
994	vng994	No prediction available	1.624	0.684	0.218	*	*	*
996	*boa4*	Bacterio-opsin activator-like protein	-2.044	-0.946	0.502	*	*	*
1007	vng1007	No prediction available	1.649	0.710	0.187	*	*	*
1008	*flaA1a*	Flagellin A1 precursor	1.455	0.496	0.360	*	*	*
1018	*adh3*	Alcohol dehydrogenase	*	*	*	-1.490	-0.524	0.376

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1021	vng1021	No prediction available	*	*	*	-1.596	-0.616	0.419
1023	vng1023	No prediction available	*	*	*	-1.682	-0.716	0.305
1024	vng1024	No prediction available	1.470	0.529	0.281	*	*	*
1025	vng1025	No prediction available	*	*	*	-1.479	-0.537	0.284
1026	vng1026	No prediction available	*	*	*	-1.612	-0.638	0.375
1038	vng1038	No prediction available	1.484	0.557	0.186	*	*	*
1041	vng1041	No prediction available	1.661	0.557	0.641	*	*	*
1058	vng1058	No prediction available	*	*	*	1.777	0.663	0.789
1060	vng1060	No prediction available	2.379	0.967	0.950	*	*	*
1063	vng1063	No prediction available	2.346	1.110	0.596	1.558	0.531	0.602
1064	vng1064	No prediction available	3.808	1.855	0.431	*	*	*
1065	vng1065	No prediction available	1.868	0.497	0.994	1.755	0.573	0.791
1074	*ykfB2*	Chloromuconate cycloisomerase	-1.577	-0.619	0.333	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1075	*menA*	Menaquinone biosynthesis	-1.556	-0.589	0.379	*	*	*
1079	*menB*	Dihydroxynaphthoic acid synthase	-1.516	-0.575	0.273	*	*	*
1083	*menF*	Isochorismate synthase	-1.512	-0.551	0.362	*	*	*
1093	vng1093	No prediction available	*	*	*	-1.621	-0.661	0.321
1094	vng1094	No prediction available	-1.486	-0.553	0.228	*	*	*
1100	vng1100	No prediction available	-1.613	-0.667	0.260	*	*	*
1115	vng1115	No prediction available	2.158	1.092	0.222	1.650	0.719	0.098
1117	vng1117	No prediction available	-1.544	-0.588	0.336	*	*	*
1128	*korA*	Putative 2-ketoglutarate ferredoxin oxidoreductase (alpha)	-1.466	-0.527	0.265	*	*	*
1137	*rpl18e*	50S ribosomal protein L18E	-1.445	-0.504	0.275	-1.462	-0.500	0.372
1138	*rpl13p*	50S ribosomal protein L13P	*	*	*	-1.626	-0.640	0.418
1139	*rps9p*	30S ribosomal protein S9P	-1.455	-0.528	0.188	*	*	*
1141	*rpoK*	DNA-directed RNA polymerase subunit K	*	*	*	-1.655	-0.673	0.371

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1150	*idsA*	Geranylgeranyl diphosphate synthase	-1.484	-0.539	0.297	*	*	*
1153	*gltS*	Glutamyl-tRNA synthetase	-1.676	-0.707	0.326	*	*	*
1164	vng1164	No prediction available	-1.548	-0.601	0.287	*	*	*
1165	*ksgA*	Dimethyladenosine transferase	-1.497	-0.543	0.327	*	*	*
1175	*phoR*	PhoR protein homolog GAF.HisKA.HATPase	*	*	*	-1.561	-0.593	0.365
1178	vng1178	No prediction available	-1.633	-0.681	0.271	-1.545	-0.603	0.269
1185	*pqqE*	Coenzyme PQQ synthesis protein	3.074	1.576	0.356	1.728	0.754	0.316
1200	vng1200	No prediction available	2.292	0.841	1.031	1.689	0.637	0.605
1202	vng1202	No prediction available	-1.510	-0.547	0.365	*	*	*
1211	*hutI*	Imidazolone-5-propionate hydrolase	3.380	1.355	1.036	*	*	*
1215	*pai1*	Sporulation regulator homolog	-1.651	-0.701	0.256	*	*	*
1216	*pgk*	3-phosphoglycerate kinase	-1.516	-0.578	0.255	*	*	*
1227	vng1227	No prediction available	-1.466	-0.534	0.225	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1234	*hik2*	Sensory histidine protein kinase	-1.572	-0.635	0.221	*	*	*
1238	vng1238	No prediction available	1.785	0.783	0.396	*	*	*
1249	vng1249	No prediction available	-1.464	-0.517	0.321	*	*	*
1250	vng1250	No prediction available	-1.425	-0.495	0.215	*	*	*
1253	*rfa7*	Replication factor A related protein - rfa32	-1.579	-0.641	0.232	*	*	*
1261	vng1261	No prediction available	1.497	0.558	0.268	*	*	*
1263	vng1263	No prediction available	1.689	0.745	0.181	*	*	*
1264	vng1264	No prediction available	1.546	0.609	0.243	*	*	*
1273	*moaC*	Molybdenum cofactor biosynthesis protein	1.445	0.507	0.266	*	*	*
1279	vng1279	No prediction available	-1.457	-0.510	0.302	*	*	*
1283	vng1283	No prediction available	-1.604	-0.647	0.316	*	*	*
1284	*trkH1*	TRK potassium uptake system protein	-1.429	-0.504	0.172	*	*	*
1296	vng1296	No prediction available	-1.462	-0.527	0.247	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1300	vng1300	No prediction available	-1.684	-0.703	0.393	*	*	*
1302	vng1302	No prediction available	*	*	*	-1.597	-0.653	0.259
1306	*sdhA*	Succinate dehydrogenase subunit A	*	*	*	-1.488	-0.539	0.306
1308	*sdhB*	Succinate dehydrogenase subunit B	*	*	*	-1.528	-0.599	0.191
1314	vng1314	No prediction available	*	*	*	-1.973	-0.960	0.240
1315	vng1315	No prediction available	*	*	*	-1.667	-0.710	0.283
1317	vng1317	No prediction available	1.650	0.547	0.620	*	*	*
1329	vng1329	No prediction available	*	*	*	2.028	0.983	0.300
1370	*hemU*	Iron (III) ABC transporter permease	-1.588	-0.615	0.453	-1.515	-0.545	0.373
1379	*suhB*	Inositol monophosphatase family	-1.543	-0.610	0.216	*	*	*
1383	*rad3a*	Helicase	-1.625	-0.613	0.537	*	*	*
1390	*tfl*	Transcription initiation factor IIB -like homolog	*	*	*	-1.498	-0.550	0.305
1416	*folD*	Methylenetetrahydrofolate dehydrogenase	*	*	*	-1.476	-0.523	0.337

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1417	vng1417	No prediction available	1.853	0.787	0.544	*	*	*
1423	vng1423	No prediction available	1.523	0.517	0.509	*	*	*
1428	*htlA*	Htr-like protein	2.191	0.983	0.609	*	*	*
1429	vng1429	No prediction available	*	*	*	1.450	0.512	0.266
1432	*dys*	Deoxyhypusine synthase	-1.486	-0.552	0.235	*	*	*
1440	vng1440	No prediction available	3.788	1.854	0.399	2.116	1.073	0.151
1442	*htr12*	Htr12 transducer	-1.432	-0.496	0.256	*	*	*
1454	vng1454	No prediction available	-1.451	-0.500	0.332	*	*	*
1459	Vng1459	No prediction available	*	*	*	-1.985	-0.963	0.276
1461	vng1461	No prediction available	2.892	1.176	0.844	*	*	*
1465	*brp*	Bacteriorhodopsin related protein	2.381	0.730	1.388	*	*	*
1467	*bop*	Bacterio-opsin	*	*	*	-19.101	-4.106	0.602
1468	vng1468	No prediction available	2.208	0.963	0.686	-3.111	-1.614	0.267

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1497	vng1497	No prediction available	-1.436	-0.497	0.267	*	*	*
1498	*celM*	Endoglucanase	-1.556	-0.611	0.278	*	*	*
1501	*helB*	DNA helicase	-1.518	-0.575	0.273	*	*	*
1523	*htr8*	Htr8 transducer	-1.635	-0.687	0.251	*	*	*
1530	vng1530	No prediction available	-1.804	-0.812	0.325	-2.005	-0.960	0.355
1550	*cbiT*	Cobalamin biosynthesis	*	*	*	-1.500	-0.538	0.364
1554	*cbiG*	Cobalamin biosynthesis	-1.573	-0.627	0.279	-1.528	-0.586	0.268
1557	*cbiH*	Cobalamin biosynthesis	-1.615	-0.652	0.327	*	*	*
1558	vng1558	No prediction available	-1.716	-0.765	0.197	*	*	*
1559	vng1559	No prediction available	-1.629	-0.661	0.355	*	*	*
1561	vng1561	No prediction available	-1.599	-0.648	0.293	*	*	*
1562	vng1562	No prediction available	-1.451	-0.513	0.256	*	*	*
1566	*cobN*	Cobalamin biosynthesis protein	-1.827	-0.840	0.292	-1.777	-0.809	0.248

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1567	*cbiC*	Precorrin isomerase	-1.889	-0.901	0.221	-1.712	-0.728	0.364
1572	vng1572	No prediction available	-1.709	-0.754	0.235	*	*	*
1573	*cbiA*	Cobyrinic acid a,c-diamide synthase	-1.499	-0.525	0.414	*	*	*
1574	*cobI*	Cobalamin adenosyltransferase	-1.796	-0.807	0.322	*	*	*
1576	*cbiP*	Cobyric acid synthase	-2.272	-1.162	0.257	*	*	*
1577	vng1577	No prediction available	-1.580	-0.641	0.235	*	*	*
1578	vng1578	No prediction available	-1.515	-0.545	0.399	-1.581	-0.643	0.219
1580	vng1580	No prediction available	-2.064	-0.950	0.511	*	*	*
1581	vng1581	No prediction available	-1.959	-0.919	0.383	*	*	*
1589	vng1589	No prediction available	1.922	0.905	0.330	1.563	0.624	0.243
1590	vng1590	No prediction available	3.129	1.432	0.745	1.716	0.685	0.506
1592	*cysT2*	Sulfate transport system permease protein	-1.449	-0.508	0.277	*	*	*
1603	*gcvP1*	Glycine dehydrogenase subunit 1	*	*	*	-1.698	-0.756	0.149

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1613	vng1613	No prediction available	*	*	*	1.486	0.530	0.390
1626	vng1626	No prediction available	1.488	0.542	0.292	*	*	*
1630	vng1630	No prediction available	*	*	*	-1.542	-0.606	0.232
1644	*nrdj*	Ribonucleoside reductase classs II	*	*	*	-1.809	-0.795	0.415
1645	vng1645	No prediction available	1.620	0.542	0.617	*	*	*
1650	vng1650	No prediction available	1.560	0.563	0.476	*	*	*
1651	vng1651	No prediction available	1.658	0.715	0.204	*	*	*
1653	vng1653	No prediction available	*	*	*	1.632	0.690	0.222
1654	*cdc48E*	Cell division cycle protein	1.688	0.538	0.731	*	*	*
1667	*cdc48c*	Cell division cycle protein	-1.432	-0.497	0.245	-1.439	-0.501	0.259
1675	vng1675	No prediction available	2.485	1.305	0.152	1.735	0.789	0.129
1678	vng1678	No prediction available	*	*	*	-1.694	-0.724	0.324
1689	*rpl3p*	50S ribosomal protein L13P	-1.554	-0.607	0.285	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1690	*rpl4e*	50S ribosomal protein L4E	-1.740	-0.783	0.213	*	*	*
1691	*rpl23p*	50S ribosomal protein L23P	-1.595	-0.665	0.159	*	*	*
1692	*rpl2p*	50S ribosomal protein L2P	-1.730	-0.765	0.281	-1.521	-0.574	0.288
1693	*rps19p*	30S ribosomal protein S19P	-1.540	-0.613	0.174	*	*	*
1695	*rpl22p*	50S ribosomal protein L22P	-1.636	-0.690	0.232	*	*	*
1697	*rps3p*	30S ribosomal protein S3P	-1.448	-0.508	0.268	*	*	*
1698	*rpl29p*	50S ribosomal protein L29P	-1.666	-0.714	0.250	-1.446	-0.495	0.320
1699	vng1699	No prediction available	-1.536	-0.594	0.268	*	*	*
1700	*rps17p*	30S ribosomal protein S17P	-1.532	-0.605	0.166	-1.445	-0.509	0.252
1702	*rpl24p*	50S ribosomal protein L24P	-1.739	-0.772	0.277	-1.519	-0.571	0.301
1703	*rps4e*	30S ribosomal protein S4E	-1.573	-0.639	0.198	*	*	*
1705	*rpl5p*	50S ribosomal protein L5P	*	*	*	-1.460	-0.510	0.317
1707	*rps8p*	30S ribosomal protein S8P	-1.625	-0.680	0.249	-1.508	-0.572	0.232

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1709	*rpl6p*	50S ribosomal protein L6P	-1.667	-0.716	0.245	*	*	*
1711	*rpl32e*	50S ribosomal protein L32E	-1.486	-0.536	0.315	*	*	*
1713	*rpl19e*	50S ribosomal protein L19E	-1.454	-0.518	0.247	-1.606	-0.637	0.356
1714	*rpl18p*	50S ribosomal protein L18P	-1.590	-0.637	0.298	*	*	*
1715	*rps5p*	30S ribosomal protein S5P	-1.802	-0.824	0.267	-1.532	-0.587	0.282
1716	*rpl30p*	50S ribosomal protein L30P	-1.645	-0.690	0.284	*	*	*
1718	*rpl15p*	50S ribosomal protein L15P	-1.557	-0.602	0.318	*	*	*
1720	vng1720	No prediction available	-1.597	-0.605	0.505	*	*	*
1733	*htr17*	Htr17 transducer	1.897	0.810	0.581	-2.499	-1.243	0.490
1734	vng1734	No prediction available	1.795	0.811	0.308	-3.476	-1.779	0.231
1748	vng1748	No prediction available	1.571	0.533	0.604	*	*	*
1758	vng1758	No prediction available	1.841	0.801	0.426	*	*	*
1770	*upk*	Predicted protein kinase	-1.492	-0.545	0.298	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1772	*pgp*	Phosphoglycolate phosphatase	-1.475	-0.519	0.347	*	*	*
1773	*cad*	Pterin-4a-carbinolamine dehydratase	*	*	*	-1.667	-0.723	0.202
1774	*hemA*	Glutamyl-tRNA reductase	-1.586	-0.651	0.205	*	*	*
1775	vng1775	No prediction available	-1.930	-0.911	0.329	*	*	*
1793	*sip*	Putative stress-induced protein	*	*	*	-1.472	-0.537	0.242
1806	vng1806	No prediction available	1.470	0.540	0.218	*	*	*
1811	*eye*	Erythromycin esterase-like	1.788	0.642	0.702	*	*	*
1815	*carA*	Carbamoyl-phosphate synthase small subunit	*	*	*	2.017	0.993	0.237
1816	*trh3*	Transcription regulator	*	*	*	1.554	0.530	0.536
1829	*guaAb*	GMP synthase subunit B	-1.583	-0.651	0.184	*	*	*
1849	vng1849	No prediction available	-1.713	-0.731	0.365	*	*	*
1867	*potC*	Spermidine/putrescine ABC transporter permease	-1.633	-0.652	0.399	*	*	*
1868	*potB*	Spermidine/putrescine ABC transporter permease	-1.700	-0.679	0.489	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1882	*nadA*	Quinolinate synthetase	1.608	0.591	0.512	*	*	*
1883	*nadB*	L-aspartate oxidase	*	*	*	-1.540	-0.569	0.394
1887	*Gpm*	Phosphoglycerate mutase	-1.512	-0.579	0.223	*	*	*
1890	vng1890	No prediction available	2.709	1.197	0.861	1.796	0.765	0.502
1895	vng1895	No prediction available	-1.460	-0.520	0.277	*	*	*
1905	vng1905	No prediction available	-1.467	-0.540	0.194	*	*	*
1912	*trpD2*	Phosphoribosyl transferase	*	*	*	-1.520	-0.566	0.320
1920	vng1920	No prediction available	1.543	0.610	0.210	*	*	*
1921	*crcB2*	Possible chromosomal condensation protein	-1.613	-0.666	0.256	*	*	*
1924	*trkA6*	TRK potassium uptake system protein	1.424	0.495	0.204	*	*	*
1925	vng1925	No prediction available	*	*	*	-1.626	-0.685	0.216
1927	vng1927	No prediction available	2.054	0.856	0.742	1.604	0.615	0.444
1944	*purS*	Phosphoribosylformylglycinamidine synthase	*	*	*	-1.523	-0.578	0.282

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
1969	*gpdA2*	Glycerol-3-phosphate dehydrogenase chain A	-1.782	-0.773	0.405	*	*	*
1972	*gpdC*	Glycerol-3-phosphate dehydrogenase chain C	-1.557	-0.581	0.444	*	*	*
1973	vng1973	No prediction available	2.154	1.029	0.471	*	*	*
1976	vng1976	No prediction available	3.016	1.161	1.119	*	*	*
1977	vng1977	No prediction available	2.056	1.003	0.329	1.906	0.853	0.481
1993	vng1993	No prediction available	1.706	0.573	0.810	*	*	*
2005	*hiss*	Histidyl-tRNA synthetase	-1.521	-0.586	0.235	*	*	*
2011	*thiL*	Thiamine monophosphate kinase	-1.802	-0.815	0.312	-1.573	-0.619	0.314
2014	vng2014	No prediction available	*	*	*	-2.435	-1.243	0.343
2017	*lysS*	Lysyl-tRNA synthetase	-1.606	-0.667	0.214	*	*	*
2027	vng2027	No prediction available	*	*	*	-1.443	-0.502	0.273
2042	vng2042	No prediction available	-1.647	-0.704	0.213	*	*	*
2045	*Gcp*	O-sialoglycoprotein endopeptidase homolog	-1.601	-0.631	0.374	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
2049	vng2049	No prediction available	-1.620	-0.668	0.283	-1.677	-0.701	0.353
2051	*rpoE*	DNA-directed RNA polymerase subunit E	-1.634	-0.679	0.286	*	*	*
2054	vng2054	No prediction available	-1.525	-0.588	0.248	*	*	*
2056	*eif2g*	Translation initiation factor eIF-2 subunit gamma	-1.533	-0.593	0.255	*	*	*
2065	*Dgs*	Pantothenate metabolism flavoprotein	-1.436	-0.508	0.206	*	*	*
2071	*lfl2*	Long-chain fatty-acid-CoA ligase	1.459	0.498	0.372	*	*	*
2084	*phnE*	Transport protein	-1.444	-0.514	0.215	*	*	*
2089	vng2089	No prediction available	1.672	0.657	0.509	*	*	*
2091	vng2091	No prediction available	*	*	*	-1.454	-0.495	0.361
2094	*aclR7*	Transcription regulator	1.444	0.514	0.217	*	*	*
2096	*cctB*	Thermosome subunit beta	1.722	0.752	0.313	*	*	*
2108	*thrC3*	Threonine synthase	1.685	0.610	0.724	*	*	*
2112	*xrlR4*	Xre like HTH	-1.513	-0.568	0.294	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
2139	*atpA*	H^+-transporting ATP synthase subunit A	-1.510	-0.568	0.273	-1.535	-0.552	0.414
2140	*atpF*	H+-transporting ATP synthase subunit F	-1.510	-0.569	0.271	*	*	*
2141	*atpC*	H^+-transporting ATP synthase subunit C	-1.490	-0.550	0.273	-1.452	-0.514	0.263
2144	*atpI*	H^+-transporting ATP synthase subunit I	*	*	*	-1.485	-0.541	0.295
2160	*rfa3*	Replication factor A related protein	*	*	*	-1.521	-0.565	0.324
2175	vng2175	No prediction available	-1.497	-0.553	0.285	*	*	*
2178	*phlR*	PhiH1 repressor homolog E-6	1.745	0.676	0.644	1.884	0.793	0.592
2179	vng2179	No prediction available	2.030	0.833	0.691	*	*	*
2182	vng2182	No prediction available	1.690	0.650	0.514	*	*	*
2190	*ileS*	Isoleucyl-tRNA synthetase	-1.634	-0.694	0.204	*	*	*
2193	*cbaA*	ba3-type cytochrome oxidase subunit I	1.482	0.555	0.190	*	*	*
2205	vng2205	No prediction available	-1.460	-0.516	0.297	*	*	*
2206	*Pmu1*	Phosphomannomutase	-1.578	-0.618	0.352	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
2210	*endA*	tRNA intron endonuclease	-1.593	-0.653	0.232	*	*	*
2214	*dinF*	DNA damage-inducible protein	1.609	0.541	0.622	*	*	*
2219	*Dsa*	Dihydrolipoamide S-acetyltransferase	*	*	*	-1.515	-0.564	0.319
2223	*leuS*	Leucine-tRNA synthetase	-1.594	-0.651	0.249	*	*	*
2246	vng2246	No prediction available	-1.494	-0.556	0.255	-1.801	-0.814	0.320
2259	vng2259	No prediction available	*	*	*	-1.540	-0.599	0.265
2273	vng2273	No prediction available	1.441	0.502	0.277	*	*	*
2280	*rfcA*	Replication factor C small subunit	*	*	*	-1.681	-0.729	0.233
2294	*hisA*	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	1.605	0.525	0.706	1.477	0.506	0.417
2307	vng2307	No prediction available	*	*	*	-1.530	-0.561	0.397
2313	vng2313	No prediction available	-1.566	-0.625	0.251	*	*	*
2330	*Hem3*	Porphobilinogen deaminase	-1.690	-0.689	0.437	*	*	*
2333	*recJ*	Single stranded DNA specific exonuclease	-1.677	-0.729	0.221	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
2346	*dppC2*	Dipeptide ABC transporter permease	-1.596	-0.653	0.248	*	*	*
2353	vng2353	No prediction available	2.613	1.358	0.281	2.200	1.124	0.197
2392	vng2392	No prediction available	*	*	*	-1.576	-0.616	0.335
2393	*tssA*	Probable thiosulfate sulfurtransferase	*	*	*	-1.512	-0.572	0.265
2397	*cysA*	Sulfate transport system ATP-binding protein	2.306	0.822	1.077	1.803	0.714	0.648
2399	vng2399	No prediction available	1.754	0.575	0.929	*	*	*
2412	vng2412	No prediction available	1.543	0.610	0.218	*	*	*
2421	*Hal*	O-acetyl homoserine	1.574	0.641	0.198	*	*	*
2423	*serB*	phosphoserine phosphatase	2.658	1.073	1.015	*	*	*
2431	vng2431	No prediction available	1.580	0.643	0.224	*	*	*
2433	vng2433	No prediction available	2.073	0.742	0.849	*	*	*
2445	*arlR15*	transcriptional regulator	1.861	0.750	0.632	*	*	*
2451	vng2451	No prediction available	1.653	0.709	0.219	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
2458	vng2458	No prediction available	-1.461	-0.526	0.251	*	*	*
2465	vng2465	No prediction available	*	*	*	-1.651	-0.680	0.349
2468	vng2468	No prediction available	-1.484	-0.553	0.219	*	*	*
2475	vng2475	No prediction available	-1.525	-0.566	0.340	*	*	*
2482	*pstB1*	Phosphate ABC transporter ATP-binding	1.765	0.640	0.764	*	*	*
2483	*pstA1*	Phosphate ABC transporter permease	1.744	0.515	0.860	*	*	*
2484	*pstC1*	Phosphate transporter permease	*	*	*	-1.452	-0.499	0.331
2486	*yqgG*	Phosphate ABC transporter binding	*	*	*	-1.612	-0.679	0.169
2501	vng2501	No prediction available	2.013	0.705	1.015	1.728	0.615	0.728
2507	*pyrD*	Dihydroorotate dehydrogenase	1.658	0.502	0.946	*	*	*
2508	*Nhp*	Nonhistone chromosomal protein homolog	1.465	0.531	0.242	*	*	*
2515	vng2515	No prediction available	-1.844	-0.855	0.279	*	*	*
2516	vng2516	No prediction available	-1.547	-0.603	0.271	-1.432	-0.497	0.242

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
2520	vng2520	No prediction available	*	*	*	-1.528	-0.602	0.173
2527	*dppD*	Dipeptide ABC transporter ATP-binding	*	*	*	-1.607	-0.634	0.373
2563	vng2563	No prediction available	-1.667	-0.695	0.347	*	*	*
2574	*Can*	Aconitase	-1.514	-0.569	0.290	*	*	*
2576	vng2576	No prediction available	-1.602	-0.647	0.311	-1.615	-0.642	0.379
2591	vng2591	No prediction available	-1.493	-0.562	0.224	*	*	*
2594	vng2594	No prediction available	*	*	*	-1.766	-0.786	0.309
2603	vng2603	No prediction available	1.552	0.618	0.208	*	*	*
2604	*thi1*	Thiamine biosynthetic enzyme	*	*	*	-1.812	-0.706	0.649
2606	*thiD*	Hydroxymethylpyrimidine phosphate kinase	1.459	0.496	0.383	-2.942	-1.533	0.254
2612	*Rli*	RNase L inhibitor homolog	-1.463	-0.525	0.259	*	*	*
2614	*arlR16*	Transcriptional regulator	1.764	0.772	0.377	*	*	*
2616	*Cxp*	Probable carboxypeptidase	1.638	0.686	0.277	*	*	*

Table A.1 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
2619	vng2619	No prediction available	*	*	*	-1.528	-0.602	0.173
2622	vng2622	No prediction available	-1.565	-0.620	0.269	*	*	*
2626	vng2626	No prediction available	2.102	1.053	0.233	*	*	*
2637	vng2637	No prediction available	-1.537	-0.586	0.310	*	*	*
2653	vng2653	No prediction available	1.860	0.522	0.895	*	*	*
2662	*rpoC*	DNA-directed RNA polymerase subunit C	-1.751	-0.792	0.209	*	*	*
2665	*rpoB*	DNA-directed RNA polymerase subunit B	-1.829	-0.849	0.249	*	*	*
2666	*rpoB*	DNA-directed RNA polymerase subunit B	-1.632	-0.665	0.349	*	*	*
2668	*rpoH*	DNA-directed RNA polymerase subunit H	*	*	*	-1.478	-0.530	0.309
2669	*imp7*	Putative integral membrane protein	1.630	0.598	0.598	*	*	*
2673	vng2673	No prediction available	2.057	0.930	0.522	*	*	*
2678	vng2678	No prediction available	1.491	0.549	0.293	*	*	*
2679	*Csg*	Cell surface glycoprotein	*	*	*	-1.530	-0.589	0.259

Table A.2 Differentially expressed RNA genes of *H. salinarum* R1 during exponential and stationary growth

Gene ID	Gene name (NRC-1) [a]	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
3001	*rrlA*	23S rRNA	3.058	1.531	0.487	2.620	1.259	0.611
3003	*Rrs*	16S rRNA	1.883	0.720	0.690	*	*	*
3004	*Rrt*	7S RNA	1.524	0.588	0.244	*	*	*
3005	*trn1*	Phe-tRNA-GAA	1.508	0.556	0.322	*	*	*
3007	*trn11*	Ala-tRNA-GGC	3.014	1.403	0.671	*	*	*
3014	*trn18*	Leu-tRNA-CAG	*	*	*	-1.454	-0.513	0.276
3016	*trn2*	Val-tRNA-TAC	2.450	1.280	0.189	*	*	*
3018	*trn21*	Glu-tRNA-TTC	1.436	0.505	0.220	*	*	*
3019	*trn22*	Glu-tRNA-CTC	2.058	1.014	0.284	-1.585	-0.632	0.301
3025	*trn28*	Val-tRNA-GAC	1.611	0.668	0.239	*	*	*
3026	*trn29*	Gly-tRNA-GCC	1.577	0.643	0.207	*	*	*
3028	*trn30*	Pro-tRNA-GGG	1.505	0.586	0.104	*	*	*

Table A.2 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
3031	*trn33*	Pro-tRNA-CGG	*	*	*	-1.760	-0.774	0.343
3032	*trn34*	Gly-tRNA-CCC	1.707	0.751	0.247	*	*	*
3035	*trn37*	His-tRNA-GTG	1.483	0.553	0.214	*	*	*
3051	*trn9*	Met-tRNA-initiation	1.948	0.942	0.245	*	*	*

Table A.3 Differentially expressed genes on pNRC100 plasmid of *H. salinarum* R1 during exponential and stationary growth

Gene ID	Gene name (NRC-1) [a]	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
5008	vng5008	No prediction available	4.450	2.097	0.403	2.151	1.060	0.338
5010	*sojA*	Spo0A activation inhibitor	1.463	0.536	0.190	*	*	*
5011	vng5011	No prediction available	-2.553	-1.313	0.334	*	*	*
5017	vng5017	No prediction available	-1.509	-0.567	0.276	*	*	*
5020	*gvpL1*	GvpL protein cluster A	-1.683	-0.726	0.266	*	*	*
5021	*gvpK1*	GvpK protein cluster A	-1.554	-0.610	0.273	*	*	*
5022	*gvpJ1*	GvpJ protein cluster A	*	*	*	1.532	0.581	0.318
5028	*gvpE1*	GvpE protein cluster A	1.597	0.660	0.213	1.760	0.791	0.273
5029	*gvpD1*	GvpD protein cluster A	1.709	0.762	0.179	1.667	0.727	0.173
5030	*gvpA1*	GvpA protein cluster A	3.875	1.835	0.527	-9.189	-3.175	0.260
5032	*gvpC1*	GvpC protein cluster A	3.952	1.854	0.567	3.066	1.461	0.594
5035	*sojB*	Spo0A activation inhibitor	2.493	1.201	0.579	1.665	0.682	0.399

Table A.3 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
5041	vng5041	No prediction available	6.244	2.483	0.671	4.583	2.153	0.352
5044	*tnp2*	Transposase 11 family	*	*	*	1.986	0.972	0.233
5052	*tbpB*	Transcription initiation factor IID	2.664	1.043	0.936	*	*	*
5054	vng5054	No prediction available	2.812	1.306	0.675	*	*	*
5057	*cydB*	Cytochrome *d* oxidase chain II	1.557	0.618	0.242	*	*	*
5064	vng5064	No prediction available	-1.613	-0.658	0.307	*	*	*
5066	*phoT1*	Sodium dependent phosphate transporter	3.922	1.761	0.655	*	*	*
5069	vng5069	No prediction available	1.729	0.667	0.560	*	*	*
5071	*yfmO2*	Multidrug resistance protein homolog	1.716	0.719	0.384	*	*	*
5073	vng5073	No prediction available	2.489	1.194	0.612	*	*	*
5075	vng5075	No prediction available	1.608	0.672	0.192	*	*	*
5076	*trxA1*	Thioredoxin	-1.988	-0.929	0.464	*	*	*
5081	vng5081	No prediction available	2.105	1.029	0.339	*	*	*

Table A.3 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
5082	vng5082	No prediction available	-1.476	-0.544	0.217	*	*	*
5083	vng5083	No prediction available	1.644	0.696	0.246	*	*	*
5093	vng5093	No prediction available	1.492	0.566	0.175	1.490	0.553	0.266
5094	*orc9*	Orc / cell division control protein 6	*	*	*	1.442	0.502	0.268
5096	vng5096	No prediction available	1.628	0.695	0.153	*	*	*
5097	vng5097	No prediction available	3.498	1.780	0.272	1.802	0.839	0.172
5098	*yobE*	General secretion pathway protein homolog	2.085	1.007	0.409	*	*	*
5101	vng5101	No prediction available	1.728	0.769	0.252	*	*	*
5106	vng5106	No prediction available	1.734	0.760	0.316	*	*	*
5109	vng5109	No prediction available	*	*	*	1.937	0.941	0.192
5116	vng5116	No prediction available	1.604	0.653	0.268	*	*	*
5136	*repI*	Replication protein	2.153	1.025	0.486	1.428	0.499	0.208
5141	*hepA*	ATP-dependent RNA helicase	-1.556	-0.618	0.241	*	*	*

Table A.3 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
5145	vng5145	No prediction available	*	*	*	-1.639	-0.626	0.479
5147	*repJ*	Replication protein	3.276	1.114	1.277	*	*	*
5149	vng5149	No prediction available	-1.573	-0.627	0.279	*	*	*
5150	vng5150	No prediction available	2.906	0.948	1.229	*	*	*
5154	vng5154	No prediction available	1.567	0.508	0.752	*	*	*
5157	vng5157	No prediction available	1.544	0.600	0.264	*	*	*
5160	vng5160	No prediction available	1.938	0.825	0.700	*	*	*
5165	*Int*	No entry	1.692	0.653	0.558	*	*	*
5168	vng5168	No prediction available	1.657	0.692	0.329	*	*	*
5169	vng5169	No prediction available	-1.716	-0.684	0.511	*	*	*
5170	vng5170	No prediction available	1.928	0.911	0.300	*	*	*
5174	vng5174	No prediction available	1.861	0.757	0.603	*	*	*
5181	*arsD*	Arsenic resistance repressor	2.126	1.026	0.421	*	*	*

Table A.3 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
5185	vng5185	No prediction available	2.673	1.218	0.739	*	*	*
5197	vng5197	No prediction available	1.860	0.651	0.948	*	*	*
5198	vng5198	No prediction available	1.669	0.725	0.205	*	*	*
5199	vng5199	No prediction available	1.672	0.676	0.436	*	*	*
5201	vng5201	No prediction available	1.602	0.669	0.170	*	*	*

Table A.4 Differentially expressed genes on pNRC200 plasmid of *H. salinarum* R1 during exponential and stationary growth

Gene ID	Gene name (NRC-1) [a]	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6141	vng6141	No prediction available	1.676	0.721	0.274	*	*	*
6144	*trsE*	Transfer complex protein	1.673	0.626	0.739	*	*	*
6146	vng6146	No prediction available	1.655	0.711	0.212	*	*	*
6147	vng6147	No prediction available	-1.602	-0.644	0.322	*	*	*
6148	vng6148	No prediction available	2.131	0.529	1.077	*	*	*
6149	vng6149	No prediction available	4.490	1.858	0.921	4.195	1.817	0.859
6152	vng6152	No prediction available	2.151	1.013	0.501	*	*	*
6156	vng6156	No prediction available	2.369	1.233	0.176	1.441	0.507	0.239
6157	vng6157	No prediction available	3.144	1.463	0.700	2.403	1.145	0.608
6158	vng6158	No prediction available	2.776	1.293	0.748	*	*	*
6160	vng6160	No prediction available	2.938	1.183	1.032	2.321	0.881	0.966
6161	vng6161	No prediction available	*	*	*	1.631	0.686	0.242

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6164	*orc2*	Orc / cell division control protein 6	3.093	1.501	0.594	2.293	1.114	0.448
6165	vng6165	No prediction available	2.249	1.160	0.162	2.699	1.414	0.224
6168	vng6168	No prediction available	1.552	0.533	0.475	*	*	*
6170	vng6170	No prediction available	1.765	0.567	0.836	2.211	0.553	1.091
6173	*srl1*	Smc and rad50 like ATPase	2.323	1.067	0.643	1.574	0.587	0.445
6178	*kdpC*	Potassium-transporting ATPase C chain	1.965	0.821	0.692	1.530	0.572	0.361
6180	vng6180	No prediction available	3.186	1.332	0.971	1.571	0.571	0.474
6184	*cat4*	Cationic amino acid transporter	1.452	0.5007	0.321	*	*	*
6185	vng6185	No prediction available	1.883	0.882	0.308	1.470	0.540	0.211
6187	*orc3*	Orc / cell division control protein 6	2.694	1.299	0.607	1.707	0.687	0.452
6188	vng6188	No prediction available	1.925	0.896	0.384	1.609	0.616	0.465
6189	vng6189	No prediction available	4.553	1.667	1.258	2.086	0.868	0.652

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6194	vng6194	No prediction available	1.510	0.586	0.161	*	*	*
6195	vng6195	No prediction available	1.603	0.599	0.476	*	*	*
6196	*phoT2*	Sodium dependent phosphate transporter	2.187	0.980	0.623	*	*	*
6197	vng6197	No prediction available	2.014	0.770	0.819	1.602	0.636	0.381
6199	*cdc48d*	Cell division cycle protein	*	*	*	-1.707	-0.725	0.371
6201	*Hsp5*	Heat shock protein	1.650	0.670	0.392	*	*	*
6205	vng6205	No prediction available	1.648	0.712	0.149	1.731	0.777	0.210
6206	vng6206	No prediction available	*	*	*	-1.588	-0.619	0.376
6208	*imp5*	Putative transmembrane protein	*	*	*	-1.449	-0.505	0.298
6224	vng6224	No prediction available	1.636	0.592	0.560	*	*	*
6229	*gvpL2*	GvpL protein cluster B	-1.495	-0.515	0.421	-1.784	-0.785	0.374
6230	*gvpK2*	GvpK protein cluster B	-1.867	-0.814	0.505	-2.275	-1.146	0.327

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6232	*gvpJ2*	GvpJ protein cluster B	-1.600	-0.633	0.386	-1.640	-0.700	0.200
6235	*gvpH2*	GvpH protein cluster B	*	*	*	-1.902	-0.873	0.383
6236	*gvpG2*	GvpG protein cluster B	-1.525	-0.588	0.240	-1.983	-0.964	0.256
6237	*gvpF2*	GvpF protein cluster B	*	*	*	-1.625	-0.677	0.250
6239	*gvpE2*	GvpE protein cluster B	*	*	*	-1.643	-0.664	0.389
6240	*gvpD2*	GvpD protein cluster B	*	*	*	-1.705	-0.721	0.361
6241	*gvpA2*	GvpA protein cluster B	1.456	0.514	0.290	-10.802	-3.395	0.325
6242	*gvpC2*	GvpC protein cluster B	*	*	*	-8.079	-2.915	0.543
6244	*gvpN2*	GvpN protein cluster B	*	*	*	-3.865	-1.866	0.459
6246	*gvpO2*	GvpO protein cluster B	*	*	*	-2.154	-1.060	0.343
6261	*yocR*	Sodium-dependent transporter	-1.723	-0.762	0.259	*	*	*
6262	*zurM*	ABC transporter permease protein	2.012	0.585	1.143	*	*	*

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6266	vng6266	No prediction available	2.040	0.558	1.056	-2.416	-1.219	0.387
6272	*orc5*	Orc / cell division control protein 6	-1.527	-0.588	0.252	*	*	*
6281	*ugpC*	sn-glycerol-3-phosphate transport system ATP-binding	1.457	0.527	0.215	*	*	*
6287	*arlR19*	Transcriptional regulator	1.995	0.845	0.632	*	*	*
6290	vng6290	No prediction available	3.485	1.612	0.744	1.565	0.605	0.338
6291	vng6291	No prediction available	2.079	0.831	0.771	*	*	*
6294	*perA*	Peroxidase / Catalase	*	*	*	-1.670	-0.723	0.218
6301	*aph*	Alkaline phosphatase	-1.680	-0.651	0.566	-1.466	-0.501	0.359
6305	vng6305	No prediction available	*	*	*	-1.485	-0.521	0.379
6306	vng6306	No prediction available	*	*	*	-1.564	-0.607	0.331
6315	*arcB*	Ornithine carbamoyltransferase	*	*	*	-1.611	-0.658	0.287
6316	*arcC*	Carbamate kinase	*	*	*	-1.583	-0.620	0.338

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6317	*arcA*	Arginine deiminase	*	*	*	-1.576	-0.622	0.317
6327	vng6327	No prediction available	2.951	1.065	1.246	2.102	0.866	0.791
6329	vng6329	No prediction available	2.386	0.913	0.899	*	*	*
6334	vng6334	No prediction available	1.828	0.820	0.364	1.613	0.676	0.197
6339	vng6339	No prediction available	2.387	1.004	0.791	*	*	*
6341	vng6341	No prediction available	*	*	*	-1.633	-0.679	0.284
6343	vng6343	No prediction available	2.845	1.107	1.072	2.540	0.946	0.970
6344	vng6344	No prediction available	2.726	1.334	0.598	1.794	0.792	0.386
6345	vng6345	No prediction available	*	*	*	1.687	0.737	0.222
6346	vng6346	No prediction available	2.573	1.153	0.759	1.800	0.691	0.664
6348	vng6348	No prediction available	2.469	0.995	0.913	2.239	0.687	1.132
6349	vng6349	No prediction available	2.415	0.983	0.857	1.813	0.799	0.406

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6351	*tfbC*	Transcription initiation factor IIB	3.196	1.372	0.900	1.960	0.754	0.801
6353	vng6353	No prediction available	2.764	1.339	0.602	2.555	1.230	0.554
6354	*comA*	Competence-like protein	2.123	0.965	0.578	*	*	*
6355	vng6355	No prediction available	2.375	0.867	0.991	*	*	*
6357	vng6357	No prediction available	1.633	0.504	0.695	1.633	0.500	0.707
6359	vng6359	No prediction available	2.873	0.998	1.125	1.569	0.544	0.522
6361	*tnp2*	Putative transposase	2.129	0.935	0.676	*	*	*
6362	*polB2*	DNA polymerase B2	2.191	0.974	0.677	1.615	0.506	0.747
6365	vng6365	No prediction available	3.206	1.660	0.234	1.992	0.989	0.117
6366	vng6366	No prediction available	3.857	1.556	1.138	*	*	*
6368	vng6368	No prediction available	1.511	0.580	0.212	*	*	*
6373	*phrH*	PhiH1 repressor homolog	1.596	0.657	0.225	*	*	*

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6375	vng6375	No prediction available	1.591	0.564	0.569	*	*	*
6378	vng6378	No prediction available	2.089	0.933	0.607	*	*	*
6383	*lctP*	L-lactate permease	1.956	0.805	0.702	*	*	*
6391	vng6391	No prediction available	1.692	0.745	0.204	*	*	*
6403	*rfa6*	Replication factor A related protein - rfa32	2.027	0.815	0.757	*	*	*
6404	vng6404	No prediction available	2.285	0.734	1.092	*	*	*
6408	*phzF*	Phenazine biosynthetic protein	1.651	0.700	0.262	*	*	*
6416	vng6416	No prediction available	1.717	0.539	0.766	*	*	*
6418	vng6418	No prediction available	1.607	0.620	0.415	*	*	*
6421	vng6421	No prediction available	2.963	1.528	0.338	*	*	*
6429	vng6429	No prediction available	3.474	1.471	0.963	2.104	0.884	0.733
6430	vng6430	No prediction available	1.503	0.516	0.594	*	*	*

Table A.4 continues…

Gene ID	Gene name (NRC-1)	Predicted function	Late exponential phase			Stationary phase		
			Fold change value	$Log_2(x)$ Ratio	Standard deviation of $log_2(x)$	Fold change value	$Log_2(x)$ ratio	Standard deviation of $log_2(x)$
6432	vng6432	No prediction available	3.256	1.434	0.877	1.957	0.752	0.780
6434	vng6434	No prediction available	2.237	0.994	0.686	*	*	*
6437	vng6437	No prediction available	1.917	0.676	0.790	*	*	*
6439	vng6439	No prediction available	1.988	0.702	0.945	1.798	0.670	0.719
6441	vng6441	No prediction available	2.010	0.883	0.617	*	*	*

Note:

[a] Gene names and regulons are from a *Halobacterium* sp. NRC-1 perspective.

- Asterisks (*) indicate no differential regulation in the *Halobacterium salinarum* R1 Δ*CYP174A1* strain.