ENGINEERING OF ANILINE DIOXYGENASE FOR BIOREMEDIATION AND INDUSTRIAL APPLICATIONS

ANG EE LUI

NATIONAL UNIVERSITY OF SINGAPORE

&

UNIVERSITY OF ILLINOIS AT URBANA CHAMPAIGN

2007

ENGINEERING OF ANILINE DIOXYGENASE FOR BIOREMEDIATION AND INDUSTRIAL APPLICATIONS

ANG EE LUI

B. Eng (Hons.), National University of Singapore

A THESIS SUBMITTED FOR

THE DEGREE OF PHILOSOPHY IN ENGINEERING

DEPARTMENT OF CHEMICAL AND BIOMOLECULAR ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

&

UNIVERSITY OF ILLINOIS AT URBANA CHAMPAIGN

2007

Acknowledgements

My heartfelt thanks to my advisors, Associate Professor Jeffrey Obbard and Associate Professor Huimin Zhao for the guidance, inspiration, support and patience they have given me throughout my PhD. I am eternally grateful to them for doing everything possible (and more) to help me along this journey.

I would like to thank my prelim committee members, Dr Richard Braatz, Dr Nick Sahinidis, and Dr Chris Rao from UIUC, as well as Dr Lanry Yung from NUS, for their advice on my project.

I would like to thank my friends in the Zhao lab for all their help in my work, especially Zhilei for her guidance when I first started in the lab. It was hard work but thank you guys so much for making the late nights in the lab so lively and not so lonely!

To my friends at UIUC – Mike, Ty, Karina, Nate, Mo, Jon, Zeng Yi, Wenjuan, Karu, Jungkul, Olga, Ryan Woodyer and Ryan Sullivan, Sheryl, Lily, Charlotte, Jing, Kim Seng, Christian, Rob, Neel, Josh, Halong, Alice, Esther, Eng Kiat, the Singapore Students' Association, the International Football Club, and everyone else – thanks for the great time and making me feel so at home in Urbana Champaign.

I would like to thank Jeff's group for their hospitality and welcoming me into the group as though I was there all along when I returned to Singapore. Also, I never would have finished this work if not for the kindness and generosity of Dr Choe, Haibin and Nian Rui. Thank you very much for putting me up in your lab in NUS.

Special thanks to my mother, Tan Kim Lian, who toiled all these years to give me this opportunity and was always there to give me support. It definitely took a lot of courage and sense of adventure for me to take up this program and I would like to thank my brother, Heng Ung, for inspiring me with both these qualities.

To the most special person in my life, Xing Yi, thank you for being beside me through the best and worst times, always giving me advice and more importantly, loving me. Throughout this journey, just like during the Chicago Marathon in 2004, there is no one else I would rather have beside me. I cherish the memories we created together, but look forward even more to our exciting life ahead with you by my side.

Lastly, I would like to thank God for always being there for me and giving me strength and hope during difficult times.

Contents

Summary

The aniline dioxygenase (AtdA) is a multi-component enzyme that has potential uses in bioremediation of aromatic amines and biorefining processes such as the denitrogenation of carbazole. However, the lack of characterization of the enzyme has limited its development as a practical biocatalyst. The overall objective of this project was to first determine the substrate specificity of AtdA, and then probe for the molecular determinants of its substrate specificity as well as its activity. Using the insights gained from the characterization studies, biomolecular engineering techniques were then used to improve the activity of AtdA as well as to expand its substrate range for application in bioremediation and industrial applications.

The first part of the dissertation presents the development of the tools required for the engineering of AtdA. An expression system, in which both the expression level and the activity of the AtdA enzyme were improved over the original plasmid construct, was established. The liquid phase Gibbs' reagent screening method, which was sensitive and efficient enough to allow for screening of the large genetic libraries generated, was then developed.

A gene deletion assay on AtdA was used to narrow the target subunit to AtdA3. Subsequently, saturation mutagenesis of the active site residues of subunit AtdA3, identified using a homology model, enhanced the promiscuity of AtdA to accept the substrate 2-isoppropylaniline (2IPA), which was not accepted by the wild type enzyme. A

single V205A mutation was found to be responsible for creating the enhanced substrate range of the mutant 1-K31. However, the expanded substrate range of the 1-K31 came at the expense of its activity for aniline (AN) and 2,4-dimethylaniline (24DMA). This is the first study on the molecular determinants for substrate specificity of a five subunit Rieske-dioxygenase, AtdA, and it was shown that the α-subunit of the enzyme (AtdA3) indeed plays a part in controlling the substrate specificity and activity of the enzyme. Using knowledge gained from these findings, saturation and random mutagenesis was then employed to enhance the activity of 1-K31.

Another round of saturation mutagenesis on active site residues with 1-K31 as parent followed by random mutagenesis using error-prone polymerase chain reaction (epPCR) yielded the mutant 3-R21. Whole cell activity assay revealed that the activity of 3-R21 for AN, 24DMA and 2IPA were 27.7, 9.8, and 2.2 nmol/min/mg protein respectively. The activities of 3-R21 for AN, 24DMA and 2IPA were improved by 8.9, 98.0, and 2.0-fold respectively over its parent 1-K31. In particular, the activity of the final mutant 3-R21 was improved by 3.5-fold over the WT AtdA enzyme, while the AN activity was restored to the WT level. Overall, mutant 3-R21 had three mutations – V205A (carried over from the 1-K31 parent), I248L (from the second round of active site residue saturation mutagenesis) and S404C (from epPCR).

This study improved the understanding of the structural determinants of the substrate specificity of AtdA, and enhanced the substrate range and activity of AtdA, making it a better enzyme for bioremediation. The 3-R21 mutant created also serves as a useful platform in the stepwise evolution strategy to engineer AtdA for carbazole denitrogenation application.

List of Tables

List of Figures

Nomenclature

1NDO	Crystal structure of napthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816-4
1ULJ	Crystal structure of biphenyl dioxygenase from *Rhodococcus* sp. strain RHA1
1WQL	Crystal structure of cumene dioxygenase from *Pseudomonas fluorescens* IP01
24DMA	2,4-Dimethylaniline
2ABPD	2-Aminobiphenyl-2,3-diol
2EA	2-Ethylaniline
2IPA	2-Isopropylaniline
2MA	2-Methylaniline
2SBA	2-Sec-butylaniline
2TBA	2-Tert-butylaniline
34DMA	3,4-Dimethylaniline
3IPC	3-Isopropylcatechol
3MC	3-Methylcatechol
4MC	4-Methylcatechol
AN	Aniline
AtdA	Aniline dioxygenase
CarA	Carbazole-1,9a-dioxygenase
CB	Chlorobiphenyl
CBA	(Chloro)benzoate

DMF	Dimethylformamide
epPCR	Error-prone polymerase chain reaction
GAT	glutamine amidotransferase
GS	Glutamine synthetase
IPTG	Isopropyl-β-D-thiogalactopyranoside
MBTH	*N*-methylbenzothiazolinone-2-hydrazone
NDO	Naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816-4
PAH	Polycyclic aromatic hydrocarbons
PCB	Polychlorinated biphenyls
PCR	Polymerase chain reaction
POP	Persistent organic pollutants
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
WT	Wild type

Chapter 1
Introduction

1.1 Background and motivation

Aniline and its derivatives are widely used as intermediates in the pharmaceutical and azo-dye manufacturing industries (Radomski, 1979; Grayson, Eckroth et al., 1984), and can be released to the environment through effluent streams from these industries (Rai, Bhattacharyya et al., 2005). These compounds are highly toxic and there have been numerous reports on their carcinogenic effects (Weisburger, Russfield et al., 1978; Nohmi, Miyata et al., 1983; Shardonofsky and Krishnan, 1997; Przybojewska, 1999; Markowitz and Levin, 2004; Bomhard and Herbold, 2005). Biodegradation is the main route for removing aromatic amine pollutants from the natural environment (Lyons, Katz et al., 1984), with the hydroxylation of the aromatic ring constituting the first step of biodegradation (Bugg and Winfield, 1998). However, to date, there have not been any reports of isolated enzymes responsible for the degradation of some classes of aromatic amine such as the xylidine. Thus, an enzyme with an ability to hydroxylate a wide range of aniline homologues would be a practical and valuable biocatalyst for the remediation of harmful aromatic amine contaminants.

Aniline dioxygenase, AtdA, is a multi-component enzyme isolated from *Acinetobacter* sp. strain YAA, which carries out simultaneous deamination and oxygenation of aniline and *o*-toluidine to catechol and 3-methylcatechol, respectively (Takeo, Fujii et al., 1998a; Takeo, Fujii et al., 1998b). AtdA is encoded by five separate

genes (*atdA1-A5*), which constitute four putative components: AtdA1 which is a glutamine synthetase-like protein; AtdA2 which is a glutamine amidotransferase-like protein; AtdA3 and AtdA4 which resemble the large (α) and small (β) subunits of the terminal class dioxygenase; as well as AtdA5 which is a reductase component (Takeo, Fujii et al., 1998a).

Studies have shown that the substrate specificity of various dioxygenases, such as the naphthalene, biphenyl and 2,4-dinitrotoluene dioxygenases, are determined by their terminal α–subunits (Tan and Cheong, 1994; Parales, Parales et al., 1998; Parales, Emig et al., 1998). From these findings, various directed evolution and saturation mutagenesis studies on the terminal α–subunits have successfully altered the substrate specificity of these dioxygenases (Sakamoto, Joern et al., 2001; Barriault, Plante et al., 2002; Barriault and Sylvestre, 2004; Keenan, Leungsakul et al., 2004; Keenan, Leungsakul et al., 2005; Leungsakul, Keenan et al., 2005). Results from these studies indicate the likelihood that AtdA3 controls the substrate specificity of aniline dioxygenase.

However, unlike the dioxygenases in the above mentioned works, which only require the α, β, and reductase subunits to carry out the benzene ring hydroxylation reactions, AtdA requires all five subunits to be present to undertake aniline hydroxylating activity (Fujii, Takeo et al., 1997). To date, it has not been reported which of the five subunits controls the substrate specificity of aniline dioxygenase. The lack of characterization of the structural determinant of the substrate specificity of AtdA limits its development as a biocatalyst for industrial applications. Hence, elucidation of the molecular determinants of the substrate specificity of AtdA is first required before engineering of the enzyme to expand its substrate range.

In addition to bioremediation applications, AtdA can be potentially applied to industrial applications such as biorefining. With the depletion of crude oil reserves, middle and heavy petroleum feedstocks, which contain high levels of nitrogen impurities, are becoming more important as precursors for lighter feedstocks. The combustion of nitrogen compounds in fuels, results in formation of nitrogen oxides (NO_x), which consequently contribute to acid rain and air pollution. Nitrogen compounds undesirable in refining processes as they are strong inhibitors of the hydrotreatment processes (Nagai and Kabe, 1983; Girgis and Gates, 1991; Laredo, Montesinos et al., 2004), and causes gum formation in fuel in storage (Dinneen and Bickel, 1951; Ford, Holmes et al., 1981). Hence, treatment of heavy feedstocks to remove nitrogen contaminants is necessary to meet increasingly stringent environmental emission regulations as well as to maximize the efficiency of refinery processes.

Biological denitrogenation, which is the use of microorganisms to denitrogenate feedstocks, has advantages over industrial methods as it can be applied at ambient temperature and pressure, resulting in lower energetic costs. Most research on microbial denitrogenation has concentrated on the removal of non-basic nitrogen compounds as they represent the majority of total nitrogen present and are harder to remove (Benedik, Gibbs et al., 1998). One of the main components of the non-basic nitrogen compounds is carbazole (Mushrush, Beal et al., 1999; Laredo, Leyva et al., 2002), which has been used as a model non-basic compound in many previous microbial degradation studies (Grosser, Warshawsky et al., 1991; Ouchiyama, Zhang et al., 1993; Kobayashi, Kurane et al., 1995; Kirimura, Nakagawa et al., 1999; Schneider, Grosser et al., 2000; Kilbane, Daram et al., 2002).

Most of the microbial carbazole degradation pathways discovered use the same *meta*-cleavage degradation pathway as that of the c*arABC* operon isolated from *Pseudomonas* sp strain CA10 (Sato, Ouchiyama et al., 1997) (Figure 1.1). However, microbial denitrogenation process is economically unfeasible as the precious fuel value of carbazole is lost when carbazole is converted to biomass via the tricarboxylic acid (TCA) cycle (Benedik, Gibbs et al., 1998). To date, there is no enzymatic pathway that can denitrogenate carbazole and at the same time preserve its fuel value.

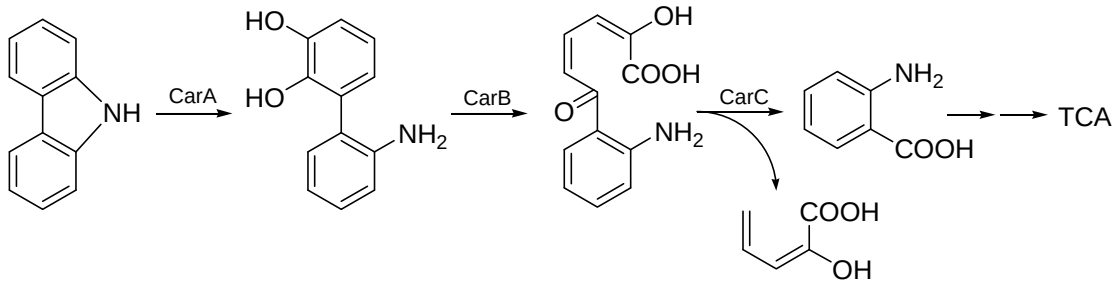


Figure 1.1 Common microbial carbazole degradation pathway.

As AtdA is capable of removing the amine group from aniline without a loss of carbon content, it has the potential to be applied to the denitrogenation of carbazole. This denitrogenation pathway can be achieved via the combination of carbazole-1,9a-dioxygenase (CarA) and a genetically engineered AtdA (Figure 1.2).

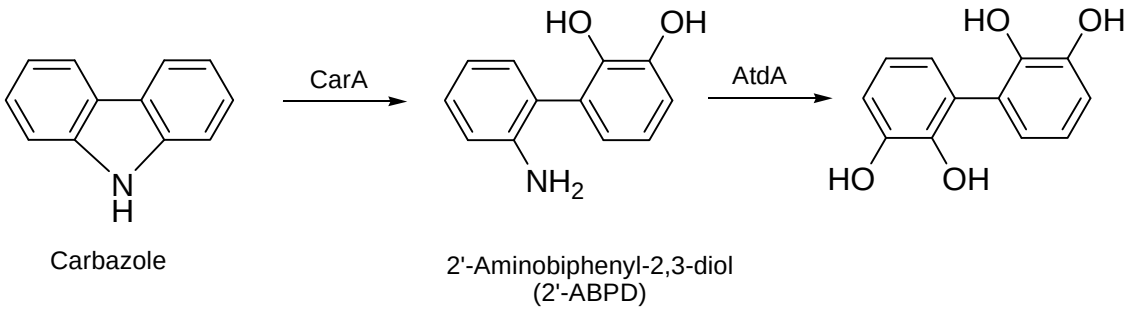


Figure 1.2 Proposed carbazole denitrogenation pathway.

1.2 Objectives

The overall objective of this project was to first characterize the AtdA enzyme by determining the molecular determinants of its substrate specificity as well as its activity. Using the insights gained from the characterization studies, further engineering work was conducted to improve the activity of AtdA as well as expand its substrate range for application in bioremediation and industrial applications. The specific objectives and scope of the project are as follows:

1. To set up a bacterial host-plasmid system that functionally expresses AtdA with high activity. High activity of the enzyme will facilitate in the development of a sensitive screening or selection system for engineering of the AtdA enzyme.

2. To develop an efficient and a sensitive screening or selection system to identify AtdA mutants with improved or novel activity.

3. To identify and probe the residues determining the activity as well as the substrate specificity of the aniline dioxygenase using molecular modeling and saturation

mutagenesis of the substrate binding pocket residues in AtdA3. The structure-function relationship elucidated from this work can be applied to the engineering of AtdA to widen its utility as a biocatalyst.

4. To improve the activity and widen the substrate range of AtdA using further rounds of saturation mutagenesis on active site residues and directed evolution. As higher activity against aromatic amines would make AtdA a more efficient catalyst, it is desirable to improve the activity of the enzyme. Furthermore, as industrial effluents contain a mixture of aromatic amine contaminants rather than just a single compound, it is desirable to widen the substrate specificity of the enzyme to make it a more generic catalyst for breaking down these pollutants. Directed evolution was used to identify residues that are further away from the active site of the enzyme yet have profound effects on its activity and substrate specificity.

In summary, the AtdA enzyme has potential uses in bioremediation of aromatic amines and biorefining such as the denitrogenation of carbazole. However, the lack of characterization of the enzyme has limited its development as a practical biocatalyst. The primary goal of this project was to characterize this enzyme, improve its activity, and widen its substrate specificity, thereby increasing its usefulness in the bioremediation and industrial applications.

1.3 References

Barriault, D., M. M. Plante and M. Sylvestre (2002). "Family shuffling of a targeted *bphA* region to engineer biphenyl dioxygenase." J Bacteriol **184** (14): 3794-800.

Barriault, D. and M. Sylvestre (2004). "Evolution of the biphenyl dioxygenase BphA from *Burkholderia xenovorans* LB400 by random mutagenesis of multiple sites in region III." J Biol Chem **279** (46): 47480-8.

Benedik, M. J., P. R. Gibbs, R. R. Riddle and R. C. Willson (1998). "Microbial denitrogenation of fossil fuels." Trends in Biotechnology **16** (9): 390-395.

Bomhard, E. M. and B. A. Herbold (2005). "Genotoxic activities of aniline and its metabolites and their relationship to the carcinogenicity of aniline in the spleen of rats." Crit Rev Toxicol **35** (10): 783-835.

Bugg, T. D. H. and C. J. Winfield (1998). "Enzymatic cleavage of aromatic rings: mechanistic aspects of the catechol dioxygenases and later enzymes of bacterial oxidative cleavage pathways." Nat Prod Rep **15**: 513-530.

Dinneen, G. U. and W. D. Bickel (1951). "Gum Formation in Shale-Oil Naphtha." Industrial & Engineering Chemistry Research **43** (7): 1604 - 1607.

Ford, C. D., S. A. Holmes, L. F. Thompson and D. R. Latham (1981). "Separation of nitrogen compound types from hydrotreated shale oil products by adsorption chromatography on basic and neutral alumina." Analytical Chemistry **53** (6): 831-836.

Fujii, T., M. Takeo and Y. Maeda (1997). "Plasmid-encoded genes specifying aniline oxidation from *Acinetobacter* sp strain YAA." Microbiology **143**: 93-99.

Girgis, M. J. and B. C. Gates (1991). "Reactivities, Reaction Networks, and Kinetics in High-Pressure Catalytic Hydroprocessing." Industrial & Engineering Chemistry Research **30** (9): 2021-2058.

Grayson, M., D. Eckroth, H. F. Mark, D. F.Othmer, C. G. Overberger and G. T. Seaborg (1984). Kirk-Othmer Encyclopedia of Chemical Technology, Vol. 2, Third Edition. New York, John Wiley and Sons**:** 309-375.

Grosser, R. J., D. Warshawsky and J. R. Vestal (1991). "Indigenous And Enhanced Mineralization Of Pyrene, Benzo[A]Pyrene, And Carbazole In Soils." Applied And Environmental Microbiology **57** (12): 3462-3469.

Keenan, B. G., T. Leungsakul, B. F. Smets, M. A. Mori, D. E. Henderson and T. K. Wood (2005). "Protein engineering of the archetypal nitroarene dioxygenase of *Ralstonia* sp. strain U2 for activity on aminonitrotoluenes and dinitrotoluenes through alpha-subunit residues leucine 225, phenylalanine 350, and glycine 407." J Bacteriol **187** (10): 3302-10.

Keenan, B. G., T. Leungsakul, B. F. Smets and T. K. Wood (2004). "Saturation mutagenesis of *Burkholderia cepacia* R34 2,4-dinitrotoluene dioxygenase at DntAc valine 350 for synthesizing nitrohydroquinone, methylhydroquinone, and methoxyhydroquinone." Appl Environ Microbiol **70** (6): 3222-31.

Kilbane, J. J., A. Daram, J. Abbasian and K. J. Kayser (2002). "Isolation and characterization of *Sphingomonas* sp GTIN11 capable of carbazole metabolism in petroleum." Biochemical And Biophysical Research Communications **297** (2): 242-248.

Kirimura, K., H. Nakagawa, K. Tsuji, K. Matsuda, R. Kurane and S. Usami (1999). "Selective and continuous degradation of carbazole contained in petroleum oil by resting cells of *Sphingomonas* sp CDH-7." Bioscience Biotechnology And Biochemistry **63** (9): 1563-1568.

Kobayashi, T., R. Kurane, K. Nakajima, Y. Nakamura, K. Kirimura and S. Usami (1995). "Isolation Of Bacteria Degrading Carbazole Under Microaerobic Conditions, Ie Nitrogen Gas Substituted Conditions." Bioscience Biotechnology And Biochemistry **59** (5): 932-933.

Laredo, G. C., S. Leyva, R. Alvarez, M. T. Mares, J. Castillo and J. L. Cano (2002). "Nitrogen compounds characterization in atmospheric gas oil and light cycle oil from a blend of Mexican crudes." Fuel **81** (10): 1341-1350.

Laredo, G. C., A. Montesinos and J. A. De los Reyes (2004). "Inhibition effects observed between dibenzothiophene and carbazole during the hydrotreating process." Applied Catalysis a-General **265** (2): 171-183.

Leungsakul, T., B. G. Keenan, H. Yin, B. F. Smets and T. K. Wood (2005). "Saturation mutagenesis of 2,4-DNT dioxygenase of *Burkholderia* sp. strain DNT for enhanced dinitrotoluene degradation." Biotechnol Bioeng **92** (4): 416-26.

Lyons, C. D., S. Katz and R. Bartha (1984). "Mechanisms and pathways of aniline elimination from aquatic environments." Appl Environ Microbiol **48** (3): 491-6.

Markowitz, S. B. and K. Levin (2004). "Continued Epidemic of Bladder Cancer in Workers Exposed to *Ortho*-Toluidine in a Chemical Factory." J Occup Environ Med **46** (2): 154-160.

Mushrush, G. W., E. J. Beal, D. R. Hardy and J. M. Hughes (1999). "Nitrogen compound distribution in middle distillate fuels derived from petroleum, oil shale, and tar sand sources." Fuel Processing Technology **61** (3): 197-210.

Nagai, M. and T. Kabe (1983). "Selectivity of molybdenum catalyst in hydrodesulfurization, hydrodenitrogenation, and hydrodeoxygenation: Effect of additives on dibenzothiophene hydrodesulfurization." Journal of Catalysis **81** (2): 440-449.

Nohmi, T., R. Miyata, K. Yoshikawa, M. Nakadate and M. Ishidate, Jr. (1983). "Metabolic activation of 2,4-xylidine and its mutagenic metabolite." Biochem Pharmacol **32** (4): 735-8.

Ouchiyama, N., Y. Zhang, T. Omori and T. Kodama (1993). "Biodegradation of Carbazole by *Pseudomonas* Sp CA06 and CA10." Bioscience Biotechnology and Biochemistry **57** (3): 455-460.

Parales, J. V., R. E. Parales, S. M. Resnick and D. T. Gibson (1998). "Enzyme specificity of 2-nitrotoluene 2,3-dioxygenase from *Pseudomonas* sp. strain JS42 is determined by the C-terminal region of the alpha subunit of the oxygenase component." J Bacteriol **180** (5): 1194-9.

Parales, R. E., M. D. Emig, N. A. Lynch and D. T. Gibson (1998). "Substrate specificities of hybrid naphthalene and 2,4-dinitrotoluene dioxygenase enzyme systems." J Bacteriol **180** (9): 2337-44.

Przybojewska, B. (1999). "Assessment of aniline derivatives-induced DNA damage in the liver cells of B6C3F1 mice using the alkaline single cell gel electrophoresis ('comet') assay." Cancer Lett **147** (1-2): 1-4.

Radomski, J. L. (1979). "The primary aromatic amines: their biological properties and structure-activity relationships." Annu Rev Pharmacol Toxicol **19**: 129-57.

Rai, H. S., M. S. Bhattacharyya, J. Singh, T. K. Bansal, P. Vats and U. C. Banerjee (2005). "Removal of Dyes from the Effluent of Textile and Dyestuff Manufacturing Industry: A Review of Emerging Techniques With Reference to Biological Treatment." Crit Rev Environ Sci Technol **35** (3): 219-238.

Sakamoto, T., J. M. Joern, A. Arisawa and F. H. Arnold (2001). "Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate." Appl Environ Microbiol **67** (9): 3882-7.

Sato, S. I., N. Ouchiyama, T. Kimura, H. Nojiri, H. Yamane and T. Omori (1997). "Cloning of genes involved in carbazole degradation of *Pseudomonas* sp. strain CA10: nucleotide sequences of genes and characterization of *meta*-cleavage enzymes and hydrolase." J Bacteriol **179** (15): 4841-9.

Schneider, J., R. J. Grosser, K. Jayasimhulu, W. L. Xue, B. Kinkle and D. Warshawsky (2000). "Biodegradation of carbazole by *Ralstonia* sp RJGII.123 isolated from a hydrocarbon contaminated soil." Canadian Journal Of Microbiology **46** (3): 269-277.

Shardonofsky, S. and K. Krishnan (1997). "Characterization of methemoglobinemia induced by 3,5-xylidine in rats." Journal of Toxicology and Environmental Health **50** (6): 595-604.

Takeo, M., T. Fujii and Y. Maeda (1998a). "Sequence analysis of the genes encoding a multicomponent dioxygenase involved in oxidation of aniline and *o*-toluidine in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (1): 17-24.

Takeo, M., T. Fujii, K. Takenaka and Y. Maeda (1998b). "Cloning and sequencing of a gene cluster for the *meta*-cleavage pathway of aniline degradation in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (5): 514-517.

Tan, H. M. and C. M. Cheong (1994). "Substitution of the ISP alpha subunit of biphenyl dioxygenase from *Pseudomonas* results in a modification of the enzyme activity." Biochem Biophys Res Commun **204** (2): 912-7.

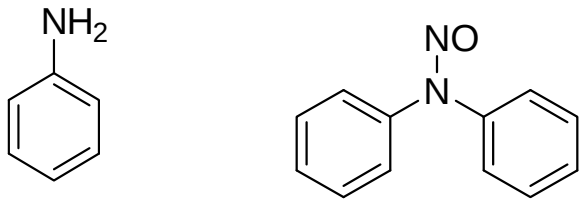
Weisburger, E. K., A. B. Russfield, F. Homburger, J. H. Weisburger, E. Boger, C. G. V. Dongen and K. C. Chu (1978). "Testing of twenty-one environmental aromatic amines or derivatives for long-term toxicity or carcinogenicity." Journal of Environmental Pathology and Toxicology **2** (2): 325-356.

Chapter 2
Literature review

2.1 Aromatic amines

Aromatic amines are compounds with one or more aromatic rings in which at least one of the hydrogen atoms has been replaced by an amine (NH_2) group. The simplest aromatic amine is aniline, and this compound may be substituted at other positions of the aromatic ring to form monocyclic aromatic amines such as *o*-toluidine (2-methylaniline), *m*-toluidine, *p*-toluidine, and xylidines (dimethylanilines). Aromatic amines can also be highly complex molecules with conjugated aromatic or heterocyclic structures with multiple substitutions (Figure 2.1) (Pinheiro, Touraud et al., 2004). The sheer number of aromatic amine structures makes it impossible to review all of them in one study. Hence, this review will only focus on the monocylic aromatic amines and how they are degraded in the environment.

The sources, environmental fates, and toxicity of these compounds, as well as methods of removing them from the environment will be reviewed in this chapter. Recent development in the application of biomolecular engineering to bioremediation will then be presented to demonstrate the potential of applying biomolecular engineering tools to enhance the bioremediation process.



1 **2**

Figure 2.1 Chemical structure of the simplest form of aromatic amine, aniline (1), and a more complex aromatic amine, *N*-Nitrosodiphenylamine (2).

2.2 Sources of aromatic amines in the environment

Aromatic amines are commonly used as intermediates in the varnish, perfume, dye, pharmaceutical and pesticide manufacturing industries (Grayson, Eckroth et al., 1984). Major sources of aromatic amines released into the natural environment come from the discharges of textile, dye manufacturing (Michaels and Lewis, 1985; Michaels and Lewis, 1986; Schnell, Bak et al., 1989; Essington, 1994; Rai, Bhattacharyya et al., 2005), or from coal gasification and shale oil extraction processes (Zachara, Felice et al., 1984).

Aromatic amines can also be produced by the degradation of azo dyes by microorganisms (Chung and Stevens, 1993). Azo dyes undergo reductive cleavage under anaerobic conditions to produce aromatic amines (Keck, Klein et al., 1997; Stolz, 2004). In these reactions, microorganisms enzymatically produce reduced mediator compounds (e.g. flavins or quinones) which in turn reduce the azo group in a purely chemical reaction to form amines. The amines that are formed in the course of these reactions may then be degraded aerobically.

2.3 Environmental fates of aromatic amines

Aniline in solution adsorbs strongly to colloidal organic matter, which effectively increases its solubility and movement into ground water. It is also moderately adsorbed to organic material in the soil, dependent upon the soil pH (pKa of 4.596) (Howard, 1989). Aniline has a vapor pressure of 0.67 mm Hg at 25 ºC and it will slowly volatilize from soil and surface water and is subject to biodegradation. Although rapidly degraded in the atmosphere, aniline can be deposited in the soil by wet and dry deposition, and by adsorption on aerosol particles (US EPA, 1985).

The fate of aniline, a representative of aromatic amine pollutants, was comprehensively evaluated using polluted pond water as a model environment (Lyons, Katz et al., 1984). The study found that biodegradation was the major route of aniline removal from aquatic environments, with evaporation, and binding to humic components playing minor roles. The major metabolic product from aniline biodegradation was catechol, formed from the oxidative deamination of aniline. Oxidation of the aniline to other minor products, phenylhydroxylamine, nitrosobenzene, or nitrobenzene, has also been reported (Kaufman, Plimmer et al., 1972; Kaufman, Plimmer et al., 1973). These minor products have been found to undergo subsequent dimerization and polymerization reactions to form azo (Bartha and Pramer, 1970; Zepp, Baughman et al., 1981), azoxy (Kaufman, Plimmer et al., 1972; Kaufman, Plimmer et al., 1973), and phenoxazine (Briggs and Walker, 1973) products. Xylidines from rocket fuel contaminated soils have been found to undergo biodegradation by microorganisms, but the metabolites were not identified (Rozkov, Vassiljeva et al., 1999).

In addition to microbial oxidation, aromatic amines in soil have been found to undergo acylation as well. 4'-chloroacetanilide were isolated as metabolites of 4-chloroaniline (Kaufman, Plimmer et al., 1973), while 3,4-dichloroformylanilide was formed from soil samples spiked with 3,4-dichloroaniline (Kearney and Plimmer, 1972). It is believed that acylation may serve as a microbial detoxification mechanism by competing with azobenzene formation in utilizing the aniline formed by metabolism of substituted urea herbicides (Tweedy, Loeppky et al., 1970).

2.4 Toxicity of aromatic amines

Earliest concerns of aromatic amine toxicity arose in the late nineteenth century, when workers in dye manufacturing industries were diagnosed with urinary bladder cancer (Weisburger, 1997). There is evidence for aniline to induce chromosome aberrations in rats *in vivo* conditions, but this is limited to high, toxic dose levels (Bomhard and Herbold, 2005). This effect could be mediated by the quantitatively major metabolites of aniline, *p*-aminophenol and *p*-hydroxyacetanilide. In addition to its possible carcinogenic effects, aniline has been reported to produce methemoglobin from hemoglobin, rendering red blood cells incapable of carrying oxygen (Kearney, Manoguerra et al., 1984). Decreased hemoglobin, erythrocyte count, and coagulative factors were reported in an occupational study on workers with chronic exposure to 1.3 to 2.75mg/m^3 (0.19-0.39 mg/kg/day) aniline for 3 to 5 years (US EPA, 1994a). An increase in methemoglobin was also reported on reexamination of these workers after one year.

On the other hand, aniline derivatives have been found to be more carcinogenic. There is strong epidemiological evidence that *ortho*-toluidine (*o*-toluidine) causes bladder

cancer in humans (Sellers and Markowitz, 1992). Further evidence of human bladder carcinogenicity of *o*-toluidine was provided by a study of workers in a chemical factory showing a high incidence of bladder cancer (Markowitz and Levin, 2004). 2,4-xylidine, showed marked toxic effect on the liver of rats (Magnusson, Bodin et al., 1971), and it is known as a reductive product of the azo dye, Ponceau R, which is tumorigenic in rats and mice (Ikeda, Horiuchi et al., 1966; Ikeda, Horiuchi et al., 1968). In turn, 2,4-dimethylphenylhydroxylamine, which is the metabolite of 2,4-xylidine, proved to be potent direct mutagen for *S. typhimurium* TA100 (Nohmi, Miyata et al., 1983). 2,4-xylidine also elicited positive DNA repair responses with rat hepatocytes, demonstrating its genotoxicity, or carcinogenic potential (Yoshimi, Sugie et al., 1988). This finding was further substantiated when a single intraperitoneal injection of 2,4-xylidine at a dose of 100 mg/kg body weight to mice resulted in an increased number of liver cell with damaged DNA (Przybojewska, 1999).

2.5 Methods of aromatic amine removal

2.5.1 Chemical methods

Various chemical methods of aromatic amine degradation have been reported, all of which were based on the oxidation of these compounds. 2,4,6-Triphenylpyrylium is a photocatalyst which works through the generation of free pyrylium radicals in the presence of light (Miranda and Garcia, 1994). The use of pyrylium-containing zeolites improved the stability of pyrylium during the oxidative degradation of 2,4-xylidine by photosensitization (Amat, Arques et al., 2004). The 2,4-xylidine degradation rate in an annular photochemical reactor was found to follow first order kinetics initially with a rate

constant of about 2.2 min^{-1}. However, the degradation did not go to completion and stopped after 20 min, with about 45 % of the 2,4-xylidine oxidized.

Ozonation of aromatic amines, whereby ozone inserts oxygen into the aromatic ring to break the C-C double bond, is another method of degradation. Aromatic amines, such as aniline, decomposed to several by-products during ozonation (e.g., nitrobenzene and azobenzene, acetic and formic acid) before mineralization (Beltran-Heredia, Torregrosa et al., 2001; Sauleda and Brillas, 2001; Sarasa, Cortes et al., 2002).

The Fenton reagent uses a mixture of hydrogen peroxide and ferrous salt to generate hydroxyl radicals for the oxidation of organic compounds (Fenton, 1894). Due to its powerful oxidizing ability, the Fenton reagent has attracted attention in wastewater treatment (Barbeni, Minero et al., 1987; Lipczynska-Kochany, 1991; Sedlak and Andren, 1991). Using the light enhanced Fenton reaction, which was carried out in an annular photochemical reactor, 200 mg/l of 3,4-xylidine was completely degraded in 25 min (Oliveros, Legrini et al., 1997).

The main drawback of these chemical methods is that they have to be carried out in a reactor. This necessitates the removal of the polluted soil or groundwater from the site, which can be costly and damage the surrounding environment. On the other hand, bioremediation has distinct advantages over physicochemical remediation methods as it can be more cost-effective and achieve the complete degradation of organic pollutants without collateral destruction of the site material or its indigenous flora and fauna (Timmis and Pieper, 1999).

2.5.2 Biodegradation of aromatic amines

Several microorganisms capable of degrading aniline and its simple methylated analogues have been isolated. From these strains, the gene clusters encoding for the enzymes responsible for the degradation of aromatic amines have been cloned and sequenced. All the gene clusters had similar nucleotide sequence and arrangement as the gene cluster of *Pseudomonas Putida* UCC22 (Fukumori and Saint, 1997). The degradation of aniline by these strains occurs through the meta-cleavage pathway and consists of two main steps (Figure 2.2). The first step involves the dihydroxylation of the aromatic ring by a Rieske non-heme iron dioxygenase to produce a catechol. The catechol is the further degraded via the cleavage of the dihydroxylated aromatic ring.

The *Pseudomonas Putida* UCC22 strain harboring a catabolic plasmid pTDN1 was able to metabolize aniline, *m*-toluidine and *p*-toluidine (McClure and Venables, 1986). Five genes, *tdnQTA1A2B,* were found to encode for proteins involved in aromatic amine degradation (Fukumori and Saint, 1997). TdnQ shows about 30 % homology to glutamine synthetases (GS) from *Salmonella typhimurium* (Yamashita, Almassy et al., 1989), while TdnT is similar to the glutamine amidotransferase (GATs) domain in GMP synthetase (Tesmer, Klem et al., 1996). TdnA1 and A2 are similar the large and small subunits of terminal Rieske dioxygenases (Wackett, 2002) respectively, while TdnB is a reductase component. It was also found that TdnT was not essential for aniline degradation (Fukumori and Saint, 1997).

Acinetobacter sp. strain YAA was able to use aniline and *o*-toluidine as the sole carbon and energy source (Fujii, Takeo et al., 1997). The five genes responsible for

aniline degradation ability were *atdA1, A2, A3, A4,* and *A5* (Takeo, Fujii et al., 1998a; Takeo, Fujii et al., 1998b) .

Delftia tsuruhatensis AD9 was isolated as an aniline-degrading bacterium from the soil surrounding a textile dyeing plant (Liang, Takeo et al., 2005). Strain AD9 was also able to utilize *m*-toluidine and *p*-toluidine as a sole source of carbon, but not *o*-toluidine, 4-chloroaniline, 2-chloroaniline, 2,4-xylidine, 3,4-dichloroaniline or 2,4-dichloroaniline. The gene cluster *tadQTA1A2B,* which encodes for a multi-component aniline dioxygenase, was found to be responsible for aniline oxidation activity. The presence of other gene clusters encoding for meta-cleavage enzymes for catechol degradation (*tadD1C1D2C2EFGIJKL*) suggests that strain AD9 degrades aniline via catechol through a *meta*-cleavage pathway by the chromosome-encoded *tad* gene cluster (Figure 2.2).

Delftia acidovorans strain 7N, which was capable of degrading aniline via the *meta*-cleavage pathway, was isolated from activated sludge samples (Urata, Uchida et al., 2004). The gene cluster with eight open reading frames (ORF_{7N}A to H) encoded for the genes responsible for aniline degradation. ORF_{7N}A to E constitutes the multi-component aniline oxygenase while ORF_{7N}F, G and H encodes for a putative LysR-type regulator, a small ferredoxin-like protein, and a catechol-2,3-dioxygenase respectively. The catechol-2,3-dioxygenase was reported to accept catechol, 3-methylcatechol, 4-methylcatechol, 4-chlorocatechol, and to a much smaller extent, 2,3-dihydroxybiphenyl, as substrates but the substrate range of the aniline dioxygenase enzyme was not determined.

The *tdnQTA1A2B* gene cluster from *Frateuria* sp. ANA-18 encodes for a multi-component aniline dioxygenase (Murakami, Hayashi et al., 2003). Deletion of *tdnA1A2* or *tdnQ* genes resulted in loss of aniline oxidation activity.

The acquisition of biodegradative capabilities by native microorganisms at contaminated sites through evolutionary processes such as random mutation occur at a slow rate, particularly when multiple biodegradation traits are required – as is the case with sites co-contaminated with more than one aromatic amine. In this context, accelerating these evolutionary processes via biomolecular engineering has become an increasingly attractive bioremediation strategy. The following section reviews the application of biomolecular engineering to the field of bioremediation.

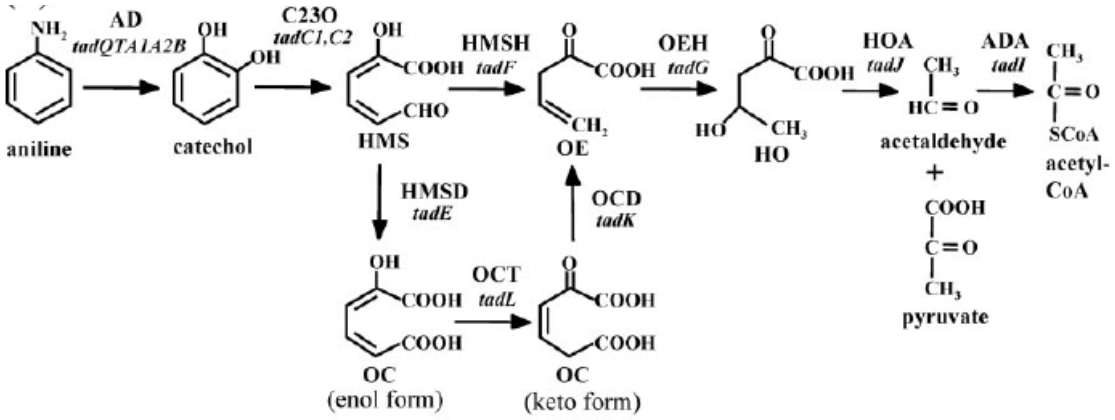


Figure 2.2 The meta-cleavage pathway of aniline by *Delftia tsuruhatensis* AD9. Figure adapted from Liang et al. 2005.

2.6 Biomolecular engineering in bioremediation

The objective of this section is to highlight and evaluate the recent developments in biomolecular engineering for enhancing the bioremediation capability of microorganisms. However, there has been no report on biomolecular engineering applied

to the biodegradation of aromatic amines to date. Hence, this section focuses on two major classes of persistent organic pollutants (POPs), i.e. polycyclic aromatic hydrocarbons (PAH), and polychlorinated biphenyls (PCB). Like aromatic amines, these pollutants are toxic but amendable to microbial degradation, particularly via the *meta*-cleavage pathway starting with the dihyroxylation of their aromatic rings by Rieske dioxygenases.

The exciting and rapidly developing area of biomolecular engineering holds potential opportunities for rapid advancement in bioremediation technology and offers the prospect of degrading some of the most recalcitrant and toxic xenobiotic POPs at large in the modern global environment.

2.6.1 Tools for biomolecular engineering

Biomolecular engineering is a relatively new field of research to engineer biomolecules, such as proteins and nucleic acids, and biomolecular processes to achieve desired biomolecular functions. This field can be classified into five main areas, namely: (1) bioinformatics, (2) protein chemistry and engineering, (3) recombinant techniques, (4) metabolic pathway engineering, and (5) bioprocess engineering (Ryu and Nam, 2000). From these areas, two different, yet complementary strategies have been developed to genetically engineer enzymes or microorganisms for bioremediation applications: rational design and directed evolution (Figure 2.3).

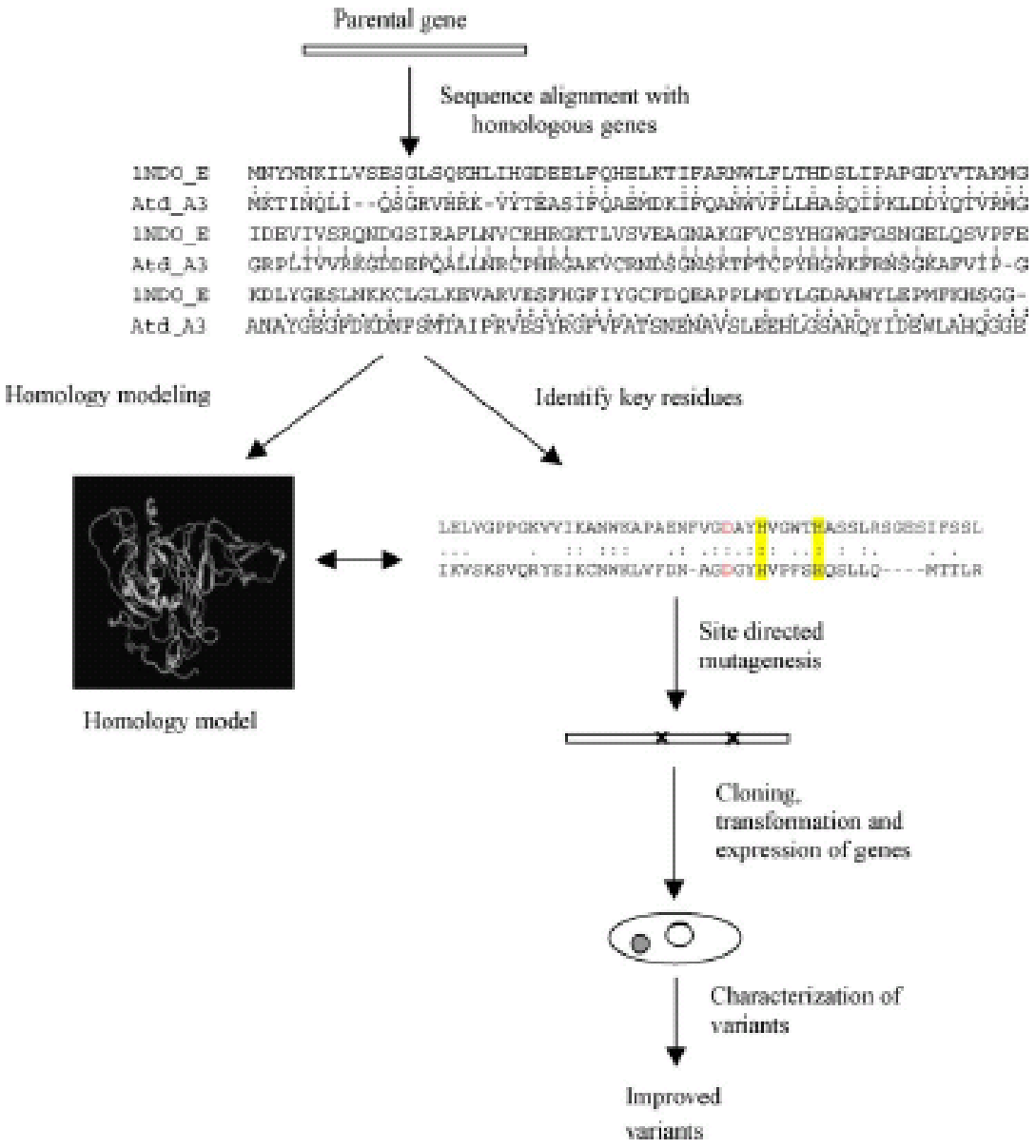


Figure 2.3 (A) Generalized scheme of the rational design process.
The parental gene is first aligned against other genes with high homology or similar function. If crystal structures of homologous proteins are available, a homology model can be built and it can be used to identify key residues. Key residues can also be identified using sequence consensus. Site directed mutagenesis is then carried at these residues and the variants are screened to isolate improved variants.

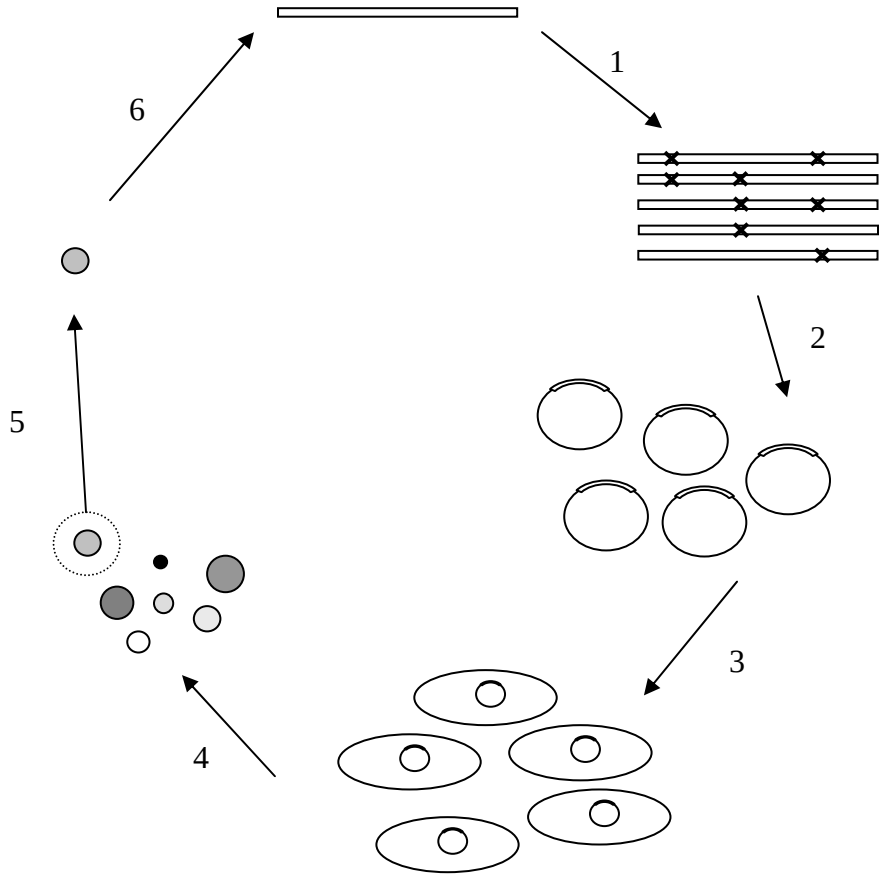


Figure 2.3 (B) Generalized scheme of the directed evolution process.
A library of genetic variants with random mutations is created from the parental gene (1). The variants are then cloned into vectors (2) and transformed into suitable hosts such as *E. coli* or yeast (3). Protein expression is then induced (4) and the library is screened or for improved phenotype (5). The gene from the variant exhibiting the desired phenotype is isolated (6) and used as the parental gene for the next round of evolution (1). The process is repeated until the desired phenotype is attained or no further improvements can be made.

The rational design approach for bioremediation typically involves either the construction of a single microorganism in which desirable biodegradation pathways or enzymes from different organisms are brought together to perform specific reactions using recombinant DNA technology (whole cell level); or the engineering of enzymes with desired features using site-directed mutagenesis (protein level) (Figure 2.3A). Enzymes are delicately folded proteins where even small changes in the amino acid sequence can disrupt its configuration and catalytic properties. Moreover, it is near impossible to predict the impact of a modification in a single trait of the enzyme on other properties. Thus, to successfully modify an enzyme using rational design, a huge amount of information relating to the structural, mechanistic and dynamics of the protein is required. This places an enormous demand on manpower and laboratory resources. Nonetheless, the potential of rational design is rapidly expanding with the recent advances in enabling technologies such as X-ray crystallography and bioinformatics.

In contrast to rational protein engineering methods, directed evolution of an enzyme does not require the structure of the enzyme to be known, and the method can identify mutations that influence enzyme activity through subtle long-range interactions (Figure 2.3B). Directed evolution is a powerful tool for protein engineering that has steadily gained popularity over the years, as evidenced by the growing number of publications (Figure 2.4).

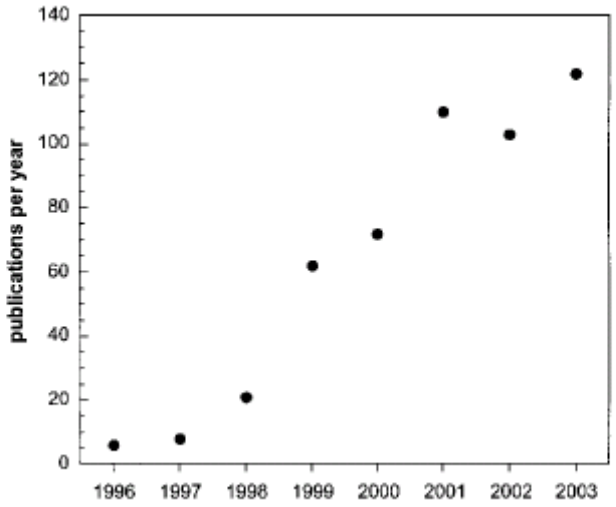


Figure 2.4 Number of publications per year in the area of directed evolution of proteins. The number is derived from a search of the ISI web-of-science database for "directed evolution" while excluding articles, the central theme of which is not the directed evolution of proteins. About half of these papers deal with the generation of new protein variants by library screening or selection; a third describe methodologies for generating genetic diversity, screening, and selection; and the remaining are primarily review papers. Figure reproduced from (Tawfik, 2004).

Directed evolution mimics a simple algorithm that nature has successfully used over eons of time i.e. genetic diversification coupled with natural selection pressure (Arnold, 1998; Schmidt-Dannert, 2001). However, directed evolution, unlike natural evolution, has a specific goal that is controlled empirically and can collapse the process into a matter of months, or even weeks.

In essence, directed evolution involves the creation of a diverse library of gene variants through random mutagenesis, such as error prone PCR or gene recombination techniques, such as the *in vitro* staggered extension process (StEP) recombination (Zhao, Giver et al., 1998) and *in vivo* DNA shuffling (Butler and Alcalde, 2003), followed by selection, or screening, to obtain the enzymes or pathways with the desired characteristics. The process is iterative where the selected or screened enzymes are

subjected to further rounds of random mutagenesis or gene recombination to produce a new generation of enzyme or biochemical pathway variants in a microorganism. For the bioremediation of organic pollutants, this means the generation of a genetically capable organism or enzyme for the complete biodegradation of the compound of interest. Methods of library creation has been reviewed in detail elsewhere (Farinas, Bulter et al., 2001; Zhao, Chockalingam et al., 2002).

It should be noted that the various methods of generating diverse genetic library are by no means mutually exclusive. DNA shuffling, combined with point mutation, can greatly accelerate the process of evolution. This method has been used in optimizing the biocatalyst in production of small molecule pharmaceuticals, viruses and vaccines (Patten, Howard et al., 1997). Random mutagenesis has been successfully combined with saturation mutagenesis, combinatorial multiple cassette mutagenesis (CMCM), DNA shuffling, as well as site directed mutagenesis to improve enzymes in industrial biocatalytical applications, for example, to generate enantioselective lipase enzymes for use in synthetic organic chemistry (Reetz, 2002).

Developing a method to generate a diverse genetic library is only part of the process in optimizing the biocatalyst. The optimization will only be successful if there is an efficient method of searching and identifying the improved proteins out of the vast number of variants generated (Zhao and Arnold, 1997). In selection methods, the organisms producing the variants are subjected to selective pressures and only host organisms possessing the required trait of properties will be able to survive. The disadvantage of this method is that the cells might be able to find alternative ways to survive the selective pressure, rendering the method ineffective. Another disadvantage of

selection is that most functions of the enzymes of interest in cannot be directly linked to their ability to survive and hence, no definitive selective pressure can be applied to the variants. In this case, the mutants must be screened one by one.

Screening differs from selection in sense that the variants are examined separately. Methods of screening include visual screens, digital imaging spectroscopy and conventional assays like ELISA. Most current screening methods are labor intensive, hence limiting their application to small scale operations. Automation and miniaturization will bring about the ability to screen larger libraries and increase the chances of identifying novel catalytic activities.

2.6.2 Naphthalene dioxygenase engineering

Polycyclic aromatic hydrocarbons (PAH) are aromatic compounds made up of two or more fused benzene rings. PAHs are recalcitrant and can persist in the environment for long periods, but are conducive to biodegradation by certain enzymes of bacteria and fungi (Hammel, Kalyanaraman et al., 1986). Most enzymatic degradation of PAHs starts with the oxidation of the aromatic rings. The naphthalene dioxygease (NDO) from *Pseudomonas* sp strain NCIB 9816-4 catalyzes the first step of the aerobic degradation of naphthalene, which is a common form of PAH, by hydroxylating the aromatic ring of naphthalene to form a homochiral (+)-cis-(1R,2S)-dihydroxy-1,2-dihydronaphthtalene (cis-naphthalene dihydrodiol) (Jeffrey, Yeh et al., 1975). The NDO consists of three subunits, namely, the terminal dioxygenase component, which carries out the dihydroxylation reaction, the ferredoxin component, which is involved in electron transfer to the terminal dioxygenase from the last component, the reductase component.

The terminal dioxygenase component is made up of the α and β subunits, in a $\alpha_3\beta_3$ conformation (Karlsson, Parales et al., 2003). Due to its ability to catalyze a wide range of aromatic compounds (Boyd and Sheldrake, 1998) and the availability of its terminal dioxygenase crystal structure (Karlsson, Parales et al., 2003), NDO is the subject of many biomolecular engineering studies, particularly by rational design methods.

Site directed mutagenesis was carried out on nine active site residues of the α subunit, which were identified using the NDO crystal structure (Parales, Lee et al., 2000). The F352V and F352L mutation altered the stereochemistry of naphthalene conversion. The wild type (WT) enzyme produces > 99 % (+)-1R,2S cis-naphthalene dihydrodiol, while F352V and F352L produced 92 and 96 % (+)-1R,2S cis-naphthalene dihydrodiol, respectively. When biphenyl was used as a substrate, the WT produced biphenyl-2,3-hydrodiol as the major product (87 %) and biphenyl-3,4-hydrodiol as the minor product (13 %). The F352V mutation dramatically changed the regiospecificity of the enzyme, with the mutant producing 96 % biphenyl-3,4-hydrodiol and 4 % biphenyl-2,3-hydrodiol. When phenanthrene was used as substrate, the major product of the WT enzyme was phenanthrene-3,4-hydrodiol (90 %) and the minor product was phenanthrene-1,2-hydrodiol. On the other hand, the F352V mutant produced phenanthrene-1,2-hydrodiol as the major product (83 %). The F352L mutant produced a small amount of phenanthrene-9,10-hydrodiol (5 %), which was not detected as a product of the WT enzyme.

Subsequently, it was found that mutation of phenylalanine at residue 352 to glycine, alanine, isoleucine, and threonine also altered the stereoselectivity of NDO (Parales, Resnick et al., 2000). In particular, the F352T mutant formed the opposite enantiomer of biphenyl cis-3,4-dyhydrodiol [60 % (-)-(3S,4R)] to that formed by the WT [>98 % (+)-

(3R,4S)]. These studies were the first evidence that the residue 352 in NDO plays an important role in the stereoselectivity of NDO. Subsequently equivalent residues of F352 in other dioxygenases were also found to have profound effect on those dioxygenases (Pollmann, Wray et al., 2003; Rui, Kwon et al., 2004; Keenan, Leungsakul et al., 2005; Ju and Parales, 2006; Lee, Ang et al., 2006).

The F352 residue in WT NDO lies in the deepest region of the active-site cavity, next to the mononuclear iron (Karlsson, Parales et al., 2003). The crystal structure of the F352V mutant (Ferraro, Okerlund et al., 2006) revealed that the effects observed from the mutation were mainly due to steric effects. The F352V mutant had a significantly larger active-site cavity than the WT due to the decreased bulk of the residue 352. The increased volume of the active-site allows phenanthrene to bind in a new orientation, thus allowing a different set of carbon atoms to be positioned near the mononuclear iron for hydroxylation.

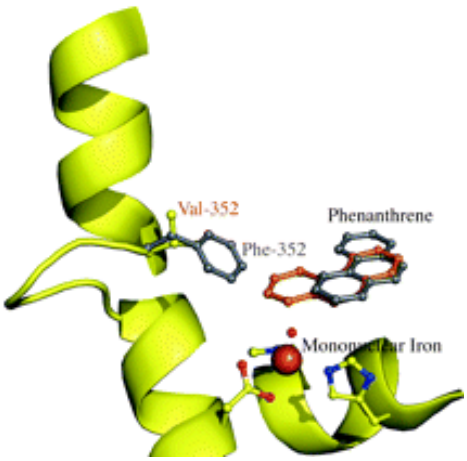


Figure 2.5 Overlay of the crystal structures of wild-type NDO and mutant F352V in complex with phenanthrene. Phenanthrene is shown in the orientations observed in the wild type (gray balls and sticks) and Phe-352-Val (orange balls and sticks). Residue 352, phenanthrene (gray balls and sticks), and valine (yellow balls and sticks) are also shown in the cartoon model. The mononuclear iron (large brown sphere), iron-coordinating residues (yellow balls and sticks), and water (small red sphere) are also shown. Figure reproduced from (Ferraro, Okerlund et al., 2006).

2.6.3 Biphenyl dioxygenase engineering

Polychlorinated biphenyls (PCB) are a class of chemicals consisting of 209 member compounds, collectively known as congeners. These compounds differ by their degree of chlorination and the position of the chlorinated sites. Since the 1930s, PCBs have had a wide range of applications from an extender in insecticide to an insulator in transformer production. Due to their low solubility in water, PCBs tend to dissolve in organic solvents and fats and move up the food chain via bioaccumulation, finally entering the human body via the ingestion pathway. It has been reported that PCBs affect brain development and exert significant effects on short-term memory, as well as planning and attention skills in young children (Winneke, Walkowiak et al., 2002). Although PCB production and use was phased out in many countries in the mid-1980s due to concerns over potential toxicity and adverse effects on wildlife, the compound are still ubiquitous throughout the global environment and its biota because of its resistance to degradation (Alcock, Behnisch et al., 1998).

PCBs can be degraded by microorganisms via a *meta*-cleavage pathway to yield tricarboxylic acid cycle intermediate and (chloro)benzoate (CBA). The initial step in the aerobic biodegradation of PCBs is the dioxygenation of PCB congeners by the biphenyl dioxygenase enzyme. In this step, the enzyme catalyzes the incorporation of two hydroxyl groups into the aromatic ring of a PCB congener, which increases the reactivity of the PCBs, and make them more susceptible to enzymatic ring fission reactions (Bruhlmann and Chen, 1999). Biphenyl dioxygenase is a multicomponent enzyme consisting of a terminal dioxygenase (made up of a large α and a small β subunit), ferredoxin, and ferredoxin reductase encoded by the *bph* operon (Erickson and Mondello,

1992). The substrate recognition of the enzyme is controlled by the large α subunit encoded by the *bphA* gene. Although the *bphA* gene is similar between bacteria species, the substrate specificity of the biphenyl dioxygnease enzymes can differ greatly. For example, although the *Burkholderia cepacia* strain LB400 and *Pseudomonas pseudoalcaligenes* strain KF707 both show near identical sequences in their *bph* operons, the biphenyl dioxygenase enzyme from LB400 preferentially deoxygenates *ortho*-substituted PCBs while that from KF707 preferentially deoxygenates *para*-substituted PCBs (Erickson and Mondello, 1993; Gibson, Cruden et al., 1993).

By targeting a fragment of *bphA* gene near that is critical for enzyme specificity (~489-1532 bp) and using DNA shuffling techniques to recombine particular gene fragments from *Burkholderia cepacia* strain LB400, *Comamonas testosteroni* B-365 and *Rhodococcus globerulus* P6, Barriault et al. were able to obtain variants with superior degradation capabilities for PCBs (Barriault, Plante et al., 2002). The hybrid BphA, II-9, was able to oxygenate 2,6-dichlorobiphenyl, which is a very persistent PCB congener, by up to 58 % after 18 hours. Both parental BphA enzymes of II-9 (*Burkholderia cepacia* strain LB400 and *Comamonas testosteroni* B-365) could only oxygenate the same PCB congener by less than 10 %. II-9 also showed marked improvement in activity towards 3,3'-, 4,4'-, 2,3,4'- and 2,3,4'- congeners relative to the primary enzymes.

Using a rational design approach, Suenaga et al. (2002) developed a three-dimensional model of the KF707 biphenyl dioxygenase enzyme, BphA1, in *Pseudomonas pseudoalcaligenes* based on crystallographic analyses of the naphthalene dioxygenase enzyme from *Pseudomonas* sp. strain NCIB 9816-4 (Suenaga, Watanabe et al., 2002). From the model, key positions near the active site of the enzyme were chosen

for site-directed mutagenesis. The resulting mutants showed altered regio-specificities for various PCB congeners compared to the wild-type enzyme. The mutants Ile335Phe, Thr376Asn and Phe377Leu were able to degrade 2,5,2',5'-tetrachlorobiphenyl, a PCB congener that is not degradable by the wild-type biphenyl dioxygenase.

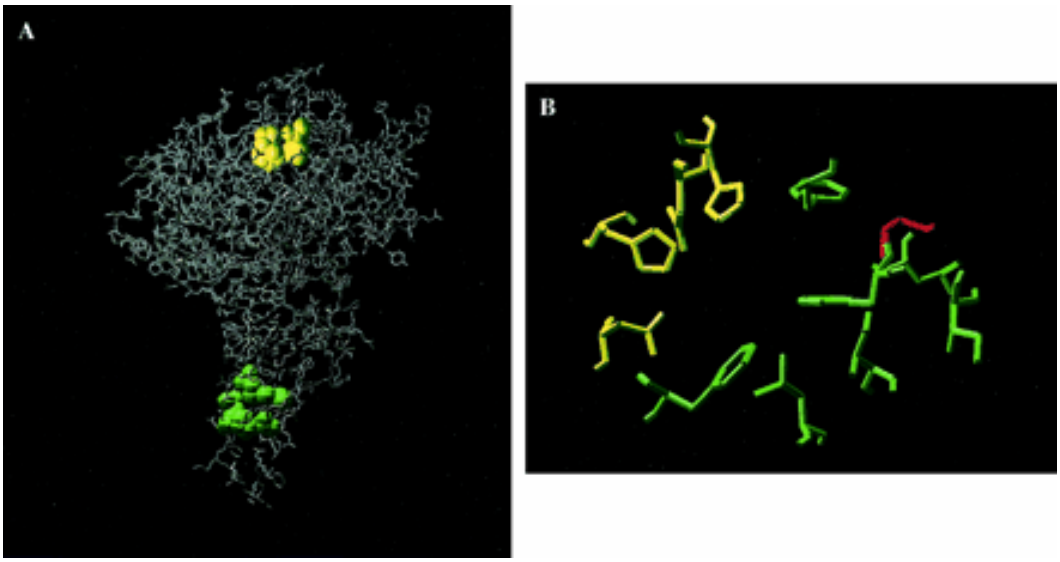


Figure 2.6 Proposed structure of BphA1 based on crystallographic analyses of the naphthalene dioxygenase (A) and the proposed structure near the active site in BphA1 (B). (A) The amino acids surrounding the active iron site are shown in yellow, and those coordinated by the Rieske cluster are shown in green. (B) The yellow amino acids coordinate the active iron site and are involved in electron transfer. The amino acids in green and red are targets for site-directed mutagenesis. Figure reproduced from (Suenaga, Watanabe et al., 2002).

Although the oxidative enzymes encoded by the *bph* gene operon confer upon microorganisms the ability to degrade PCBs, the metabolite from the degradation, CBA, cannot be further degraded by these enzymes. This is a potential problem when engineering aniline dioxygenases to accept novel substrates as well.

The plasmid pE43 contains oxygenolytic *ortho*-dechlorination *ohb* gene, originally found in *Pseudomonas aeruginosa*, whereas the plasmid pPC3 carries the hydrolytic *para*-dechlorination *fcb* gene from *Arthrobacter globiformis.* By individually

transforming these recombinant plasmids into PCB-cometabolizing *Comamonas testosteroni* VP44, two recombinant variants, VP(pPE43) and VP(pPC3), capable of using *ortho*- and *para*-chlorobiphenyls (CBs) as sole carbon sources have been obtained (Hrywna, Tsoi et al., 1999). The parental strain, VP44, grew only on low concentrations of 2- and 4-CBs and accumulated stoichiometric amounts of the corresponding CBA while the recombinant variants were able to grow on and dechlorinate 2- and 4-CBs by up to 95 %. However, complications may arise when applied to higher chlorinated PCBs, especially those chlorinated on both aromatic rings, as no dehalogenation of chloropentadiene, which is a metabolic product of PCB chlorinated on both aromatic rings, has been documented (Brenner, Arensdorf et al., 1994). Natural evolutionary processes may give rise to microorganisms that can dehalogenate chloropentadiene in due time but the process may take several years or even decades. Biomolecular engineering can potentially be applied to shorten this process of developing a novel enzyme that can carry out this reaction.

2.6.4 More engineering on dioxygenases

There are numerous examples of biomolecular engineering work on the biphenyl dioxygenase and other dioxygenases using different methods of creating library diversity such as error prone PCR, DNA shuffling and saturation mutagenesis. Results from some of these studies are summarized on Table 2.1. Due to the stereospecificity of some aryl dioxygenases, they are useful as biocatalysts for the manufacture of a range of fine chemicals (Boyd and Sheldrake, 1998). Hence, some of the dioxygenases listed in Table 2.1 were engineered for industrial rather than bioremediation purposes. An interesting

example listed is the multigene DNA shuffling of the toluene dioxygenase operon (*todC1C2BA*) and the tetrachlorobenzene dioxygenase operon (*tecA1A2A3A4*) (Newman, Garcia et al., 2004). This is the first directed evolution work on a multi enzyme complex in which all components were subjected to DNA shuffling to produce chimeras with crossovers across the whole operon instead of just one gene.

2.7 Conclusion

Human activities have brought about widespread pollution to the natural environment. A number of organic pollutants, such as polycyclic aromatic hydrocarbons, and polychlorinated biphenyls are very resistant to degradation and represent an ongoing toxicological threat to both wildlife and human beings. Over recent years, a growing number of potential hazards linked to the ubiquitous presence of organic pollutants in the environment have been reported. Bioremediation is an attractive alternative to the remediation of these pollutants at a contaminated site based on more traditional physico-chemical techniques as it can be more cost effective and can selectively degrade the pollutants without damaging the site or its flora and fauna. However, despite being hailed as a panacea to the safe and effective remediation of contaminated environmental media, to date, bioremediation technologies have had limited application due to the challenges of substrate and environmental variability, as well as the limited biodegradative potential and viability of naturally occurring microorganisms. In particular, the engineering and environmental release of genetically manipulated microorganisms has run into both technical and public obstacles, leading to severe constraints for their effective use in the field. Now, with the recent advent of rapidly advancing technology in directed evolution

techniques, the prospect of short-circuiting the process of natural evolution to degrade environmental xenobiotic pollutants has been created. This has opened exciting new vistas for enhancing bioremediation programs, and avoids many of the pitfalls associated with direct manipulation of an organism's genome.

Table 2.1 Summary of biomolecular engineering studies on aryl dioxygenases.

Enzyme	Mutation	Effect	Reference
TecA Tetrachlorobenzene DO from *Ralstonia* sp. PS12	L272W	Increased activity for 2,3-, 2,5-, 2,6-, 3,4-, and 3,5-dichlorotoluene.	(Pollmann, Wray et al., 2003)
TodC1 Toluene DO from *Pseudomonas putida* F1	M220A	Introduced dehalogenating activity to toluene dioxygenase.	(Beil, Mason et al., 1998)
	L306	Increased activity for indene, altered product distribution.	(Zhang, Stewart et al., 2000)
	Stop451T (lengthened TodC1 by 38 aa)	Increased activity for picoline and tolunene.	(Sakamoto, Joern et al., 2001)
TecA1A2A3A4 and TodC1C2BA (Multigene shuffling)	--	Increased activity for *p*-xylene by 4.4 fold.	(Newman, Garcia et al., 2004)
NbzA Nitrobenzene DO from *Comamonas* sp JS765	N258V	Altered hydroxylation site on substrate.	(Ju and Parales, 2006)
	F293Q	Increased activity for 2,6-dinitrotoluene.	
	I350F	Altered hydroxylation site on substrate.	
BphA Biphenyl DO from *Burkholderia xenovorans* LB400	P334T	Introduced 4,4'-, and 3,4,4'-chlorobiphenyl activity.	(Zielinski, Kahl et al., 2006)
	F384L	Introduced 3,5,4'-chlorobiphenyl activity.	
	M231A/ F378A/ F384A	Altered substrate regiospecificity by factors of >7.	(Zielinski, Kahl et al., 2003)
BphA1 Biphenyl DO from *Pseudomonas pseudoalcaligenes* KF707	T376N	Introduced 3,4-dioxygnase activity for 2,5,4'- and 2,5,2',5'-chlorobiphenyl.	(Kumamaru, Suenaga et al., 1998; Suenaga, Nishi et al., 1999)
	T376V	Improved activity.	(Suenaga, Goto et al., 2001)
	I335F	Altered regiospecificity for 2,2'-, and 2,5,2'-chlorobiphenyl	(Suenaga, Watanabe et al., 2002)
	F377A	Altered regiospecificity for 2,2'-, 2,5,2'-, and 2,5,4'-chlorobiphenyl, introduced activity for 2,5,2',5'-chlorobiphenyl	
	F377L	Altered regiospecificity for 2,2'-dichlorobiphenyl	
	T376N	Altered regiospecificity for 2,2'-dichlorobiphenyl	
PrnD Aminopyrrolnitrin oxygenase from *Pseudomonas fluorescens* Pf-5	F312S	Altered substrate selectivity for aminopyrrolnitrin over para-aminobenzyl amine by > 300 fold.	(Lee, Ang et al., 2006)
	L277V	Improved catalytic efficiency for aminopyrrolnitrin by 3.5 fold.	

2.8 References

Alcock, R. E., P. A. Behnisch, K. C. Jones and H. Hagenmaier (1998). "Dioxin-like PCBs in the environment-human exposure and the significance of sources." Chemosphere **37** (8): 1457-72.

Amat, A. M., A. Arques, S. H. Bossmann, A. M. Braun, S. Göb, M. A. Miranda and E. Oliveros (2004). "Oxidative degradation of 2,4-xylidine by photosensitization with 2,4,6-triphenylpyrylium: homogeneous and heterogeneous catalysis." Chemosphere **57** (9): 1123-1130.

Arnold, F. H. (1998). "Design by Directed Evolution." Accounts of Chemical Research **31** (3): 125-131.

Barbeni, M., C. Minero, E. Pelizzetti, E. Borgarello and N. Serpone (1987). "Chemical degradation of chlorophenols with Fenton's reagent." Chemosphere **16** (10-12): 2225-2237.

Barriault, D., M. M. Plante and M. Sylvestre (2002). "Family shuffling of a targeted *bphA* region to engineer biphenyl dioxygenase." Journal of Bacteriology **184** (14): 3794-3800.

Bartha, R. and D. Pramer (1970). "Metabolism of acylanilide herbicides." Advances in Applied Microbiology **13**: 317-341.

Beil, S., J. R. Mason, K. N. Timmis and D. H. Pieper (1998). "Identification of chlorobenzene dioxygenase sequence elements involved in dechlorination of 1,2,4,5-tetrachlorobenzene." J Bacteriol **180** (21): 5520-8.

Beltran-Heredia, J., J. Torregrosa, J. R. Dominguez and J. A. Peres (2001). "Kinetics of the reaction between ozone and phenolic acids present in agro-industrial wastewaters." Water Res **35** (4): 1077-85.

Bomhard, E. M. and B. A. Herbold (2005). "Genotoxic activities of aniline and its metabolites and their relationship to the carcinogenicity of aniline in the spleen of rats." Crit Rev Toxicol **35** (10): 783-835.

Boyd, D. R. and G. N. Sheldrake (1998). "The Dioxygenase-catalysed formation of vicinal *cis*-diols." Natural product reports **15** (3): 309-324.

Brenner, V., J. J. Arensdorf and D. D. Focht (1994). "Genetic construction of PCB degraders." Biodegradation **5** (3-4): 359-77.

Briggs, G. G. and N. Walker (1973). "Microbial metabolism of 4-chloroaniline." Soil Biology and Biochemistry **5** (5): 695-697.

Bruhlmann, F. and W. Chen (1999). "Tuning biphenyl dioxygenase for extended substrate specificity." Biotechnology and Bioengineering **63** (5): 544-551.

Butler, T. and M. Alcalde (2003). Preparing libraries in *Saccharomyces cerevisiae*. Directed evolution vol 1. Methods in molecular biology series. F. H. Arnold and G. Georgiou. Totowa, N.J., Humana Press. **1**.

Chung, K.-T. and S. E. Stevens (1993). "Degradation of azo dyes by environmental microorganisms and helminths." Environmental toxicology and chemistry **12** (11): 2121-2132.

Erickson, B. D. and F. J. Mondello (1992). "Nucleotide Sequencing and Transcriptional Mapping of the Genes Encoding Biphenyl Dioxygenase, a Multicomponent Polychlorinated-Biphenyl-Degrading Enzyme in *Pseudomonas* Strain-LB400." Journal of Bacteriology **174** (9): 2903-2912.

Erickson, B. D. and F. J. Mondello (1993). "Enhanced Biodegradation of Polychlorinated-Biphenyls after Site-Directed Mutagenesis of a Biphenyl Dioxygenase Gene." Applied and Environmental Microbiology **59** (11): 3858-3862.

Essington, M. E. (1994). "Adsorption of aniline and toluidines on Montmorillonite." Soil Science **158** (3): 181-188.

Farinas, E. T., T. Bulter and F. H. Arnold (2001). "Directed enzyme evolution." Curr Opin Biotechnol **12** (6): 545-51.

Fenton, H. J. H. (1894). "Oxidation of tartaric acid in presence of iron." Journal of the Chemical Society **65**: 899-910.

Ferraro, D. J., A. L. Okerlund, J. C. Mowers and S. Ramaswamy (2006). "Structural basis for regioselectivity and stereoselectivity of product formation by naphthalene 1,2-dioxygenase." J Bacteriol **188** (19): 6986-94.

Fujii, T., M. Takeo and Y. Maeda (1997). "Plasmid-encoded genes specifying aniline oxidation from *Acinetobacter* sp strain YAA." Microbiology **143**: 93-99.

Fukumori, F. and C. P. Saint (1997). "Nucleotide sequences and regulational analysis of genes involved in conversion of aniline to catechol in *Pseudomonas putida* UCC22(pTDN1)." J Bacteriol **179** (2): 399-408.

Gibson, D. T., D. L. Cruden, J. D. Haddock, G. J. Zylstra and J. M. Brand (1993). "Oxidation of Polychlorinated-Biphenyls by *Pseudomonas* Sp Strain LB400 and *Pseudomonas-Pseudoalcaligenes* KF707." Journal of Bacteriology **175** (14): 4561-4564.

Grayson, M., D. Eckroth, H. F. Mark, D. F.Othmer, C. G. Overberger and G. T. Seaborg (1984). Kirk-Othmer Encyclopedia of Chemical Technology, Vol. 2, Third Edition. New York, John Wiley and Sons**:** 309-375.

Hammel, K. E., B. Kalyanaraman and T. K. Kirk (1986). "Oxidation of Polycyclic Aromatic-Hydrocarbons and Dibenzo P -Dioxins by Phanerochaete-Chrysosporium Ligninase." Journal of Biological Chemistry **261** (36): 6948-6952.

Howard, P. H. (1989). Handbook of Environmental Fate and Exposure Data. Vol 1. Chelsea, MI, Lewis Publishers.

Hrywna, Y., T. V. Tsoi, O. V. Maltseva, J. F. Quensen and J. M. Tiedje (1999). "Construction and characterization of two recombinant bacteria that grow on *ortho*- and *para*-substituted chlorobiphenyls." Applied and Environmental Microbiology **65** (5): 2163-2169.

Ikeda, Y., S. Horiuchi, T. Furuya and Y. Omori (1966). "Chronic toxicity of Ponceau MX in the rat." Food Cosmet Toxicol **4** (5): 485-92.

Ikeda, Y., S. Horiuchi, K. Kobayashi, T. Furuja and K. Kohgo (1968). "Carcinogenicity of Ponceau MX in the mouse." Food Cosmet Toxicol **6** (5): 591-8.

Jeffrey, A. M., H. J. Yeh, D. M. Jerina, T. R. Patel, J. F. Davey and D. T. Gibson (1975). "Initial reactions in the oxidation of naphthalene by *Pseudomonas putida*." Biochemistry **14** (3): 575-84.

Ju, K. S. and R. E. Parales (2006). "Control of substrate specificity by active-site residues in nitrobenzene dioxygenase." Appl Environ Microbiol **72** (3): 1817-24.

Karlsson, A., J. V. Parales, R. E. Parales, D. T. Gibson, H. Eklund and S. Ramaswamy (2003). "Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron." Science **299** (5609): 1039-42.

Kaufman, D. D., J. R. Plimmer, J. Iwan and U. I. Klingebiel (1972). "3,3',4,4'-Tetrachloroazoxybenzene from 3,4-dichloroaniline in microbial culture." Journal of Agricultural and Food Chemistry **20** (4): 916-919.

Kaufman, D. D., J. R. Plimmer and U. I. Klingebiel (1973). "Microbial Oxidation of 4-Chloroaniline." Journal of Agricultural and Food Chemistry **21** (1): 127-132.

Kearney, P. C. and J. R. Plimmer (1972). "Metabolism of 3,4-Dichloroaniline in Soils." Journal of Agricultural and Food Chemistry **20** (3): 127-132.

Kearney, T. E., A. S. Manoguerra and J. V. Dunford, Jr. (1984). "Chemically induced methemoglobinemia from aniline poisoning." West J Med **140** (2): 282-6.

Keck, A., J. Klein, M. Kudlich, A. Stolz, H. J. Knackmuss and R. Mattes (1997). "Reduction of azo dyes by redox mediators originating in the naphthalenesulfonic acid degradation pathway of *Sphingomonas* sp. strain BN6." Appl Environ Microbiol **63** (9): 3684-3690.

Keenan, B. G., T. Leungsakul, B. F. Smets, M. A. Mori, D. E. Henderson and T. K. Wood (2005). "Protein engineering of the archetypal nitroarene dioxygenase of *Ralstonia* sp. strain U2 for activity on aminonitrotoluenes and dinitrotoluenes through alpha-subunit residues leucine 225, phenylalanine 350, and glycine 407." J Bacteriol **187** (10): 3302-10.

Kumamaru, T., H. Suenaga, M. Mitsuoka, T. Watanabe and K. Furukawa (1998). "Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase." Nat Biotechnol **16** (7): 663-6.

Lee, J. K., E. L. Ang and H. Zhao (2006). "Probing the substrate specificity of aminopyrrolnitrin oxygenase (PrnD) by mutational analysis." J Bacteriol **188** (17): 6179-83.

Liang, Q., M. Takeo, M. Chen, W. Zhang, Y. Xu and M. Lin (2005). "Chromosome-encoded gene cluster for the metabolic pathway that converts aniline to TCA-cycle intermediates in *Delftia tsuruhatensis* AD9." Microbiology **151** (Pt 10): 3435-46.

Lipczynska-Kochany, E. (1991). "Degradation of aqueous nitrophenols and nitrobenzene by means of the Fenton reaction." Chemosphere **22** (5-6): 529-536.

Lyons, C. D., S. Katz and R. Bartha (1984). "Mechanisms and pathways of aniline elimination from aquatic environments." Appl Environ Microbiol **48** (3): 491-6.

Magnusson, G., N. O. Bodin and E. Hansson (1971). "Hepatic changes in dogs and rats induced by xylidine isomers." Acta Pathol Microbiol Scand [A] **79** (6): 639-48.

Markowitz, S. B. and K. Levin (2004). "Continued Epidemic of Bladder Cancer in Workers Exposed to *Ortho*-Toluidine in a Chemical Factory." J Occup Environ Med **46** (2): 154-160.

McClure, N. C. and W. A. Venables (1986). "Adaptation of *Pseudomonas putida* mt-2 to growth on aromatic amines." Journal of General Microbiology **132** (8): 2209-2218.

Michaels, G. B. and D. L. Lewis (1985). "Sorption and toxicity of azo and triphenylmethane dyes to aquatic microbial populations." Environmental toxicology and chemistry **4**: 45-50.

Michaels, G. B. and D. L. Lewis (1986). "Microbial Transformation Rates of Azo and Triphenylmethane Dyes." Environmental Toxicology and Chemistry **5** (2): 161-166.

Miranda, M. A. and H. Garcia (1994). "2,4,6-Triphenylpyrylium Tetrafluoroborate as an Electron-Transfer Photosensitizer." Chemical Reviews **94** (4): 1063-1089.

Murakami, S., T. Hayashi, T. Maeda, S. Takenaka and K. Aoki (2003). "Cloning and functional analysis of aniline dioxygenase gene cluster, from *Frateuria* species ANA-18, that metabolizes aniline via an *ortho*-cleavage pathway of catechol." Biosci Biotechnol Biochem **67** (11): 2351-8.

Newman, L. M., H. Garcia, T. Hudlicky and S. A. Selifonov (2004). "Directed evolution of the dioxygenase complex for the synthesis of furanone flavor compounds." Tetrahedron **60** (3): 729-734.

Nohmi, T., R. Miyata, K. Yoshikawa, M. Nakadate and M. Ishidate, Jr. (1983). "Metabolic activation of 2,4-xylidine and its mutagenic metabolite." Biochem Pharmacol **32** (4): 735-8.

Oliveros, E., O. Legrini, M. Hohl, T. Müller and A. M. Braun (1997). "Industrial waste water treatment: large scale development of a light-enhanced Fenton reaction." Chemical Engineering and Processing **36** (5): 397-405.

Parales, R. E., K. Lee, S. M. Resnick, H. Jiang, D. J. Lessner and D. T. Gibson (2000). "Substrate specificity of naphthalene dioxygenase: effect of specific amino acids at the active site of the enzyme." J Bacteriol **182** (6): 1641-9.

Parales, R. E., S. M. Resnick, C. L. Yu, D. R. Boyd, N. D. Sharma and D. T. Gibson (2000). "Regioselectivity and enantioselectivity of naphthalene dioxygenase during arene *cis*-dihydroxylation: control by phenylalanine 352 in the alpha subunit." J Bacteriol **182** (19): 5495-504.

Patten, P. A., R. J. Howard and W. P. Stemmer (1997). "Applications of DNA shuffling to pharmaceuticals and vaccines." Curr Opin Biotechnol **8** (6): 724-33.

Pinheiro, H. M., E. Touraud and O. Thomas (2004). "Aromatic amines from azo dye reduction: status review with emphasis on direct UV spectrophotometric detection in textile industry wastewaters." Dyes and Pigments **61**: 121-139.

Pollmann, K., V. Wray, H. J. Hecht and D. H. Pieper (2003). "Rational engineering of the regioselectivity of TecA tetrachlorobenzene dioxygenase for the transformation of chlorinated toluenes." Microbiology **149** (Pt 4): 903-13.

Przybojewska, B. (1999). "Assessment of aniline derivatives-induced DNA damage in the liver cells of B6C3F1 mice using the alkaline single cell gel electrophoresis ('comet') assay." Cancer Lett **147** (1-2): 1-4.

Rai, H. S., M. S. Bhattacharyya, J. Singh, T. K. Bansal, P. Vats and U. C. Banerjee (2005). "Removal of Dyes from the Effluent of Textile and Dyestuff Manufacturing

Industry: A Review of Emerging Techniques With Reference to Biological Treatment." Crit Rev Environ Sci Technol **35** (3): 219-238.

Reetz, M. T. (2002). "Directed evolution of selective enzymes and hybrid catalysts." Tetrahedron **58** (32): 6595-6602.

Rozkov, A., I. Vassiljeva, M. Kurvet, A. Kahru, S. Preis, A. Kharchenko, M. Krichevskaya, M. Liiv, A. Käärd and R. Vilu (1999). "Laboratory study of bioremediation of rocket fuel-polluted groundwater." Water Research **33** (5): 1303-1313.

Rui, L., Y. M. Kwon, A. Fishman, K. F. Reardon and T. K. Wood (2004). "Saturation mutagenesis of toluene *ortho*-monooxygenase of *Burkholderia cepacia* G4 for Enhanced 1-naphthol synthesis and chloroform degradation." Appl Environ Microbiol **70** (6): 3246-52.

Ryu, D. D. Y. and D. H. Nam (2000). "Recent progress in biomolecular engineering." Biotechnology Progress **16** (1): 2-16.

Sakamoto, T., J. M. Joern, A. Arisawa and F. H. Arnold (2001). "Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate." Appl Environ Microbiol **67** (9): 3882-7.

Sarasa, J., S. Cortes, P. Ormad, R. Gracia and J. L. Ovelleiro (2002). "Study of the aromatic by-products formed from ozonation of anilines in aqueous solution." Water Res **36** (12): 3035-44.

Sauleda, R. and E. Brillas (2001). "Mineralization of aniline and 4-chlorophenol in acidic solution by ozonation catalyzed with Fe^{2+} and UVA light." Applied Catalysis B: Environmental **29** (2): 135-145.

Schmidt-Dannert, C. (2001). "Directed evolution of single proteins, metabolic pathways, and viruses." Biochemistry **40** (44): 13125-36.

Schnell, S., F. Bak and N. Pfennig (1989). "Anaerobic degradation of aniline and dihydroxybenzenes by newly isolated sulfate-reducing bacteria and description of *Desulfobacterium anilini*." Archives of Microbiology **152** (6): 556-563.

Sedlak, D. L. and A. W. Andren (1991). "Oxidation of chlorobenzene with Fenton's reagent." Environmental Science & Technology **25** (4): 777-782.

Sellers, C. and S. Markowitz (1992). "Reevaluating the carcinogenicity of *ortho*-toluidine: A new conclusion and its implications." Regulatory Toxicology and Pharmacology **16** (3): 301-317.

Stolz, A. (2004). "Basic and applied aspects in the microbial degradation of azo dyes." Applied Microbiology and Biotechnology **56** (1-2): 69-80.

Suenaga, H., M. Goto and K. Furukawa (2001). "Emergence of multifunctional oxygenase activities by random priming recombination." J Biol Chem **276** (25): 22500-6.

Suenaga, H., A. Nishi, T. Watanabe, M. Sakai and K. Furukawa (1999). "Engineering a hybrid pseudomonad to acquire 3,4-dioxygenase activity for polychlorinated biphenyls." J Biosci Bioeng **87** (4): 430-5.

Suenaga, H., T. Watanabe, M. Sato, Ngadiman and K. Furukawa (2002). "Alteration of regiospecificity in biphenyl dioxygenase by active-site engineering." Journal of Bacteriology **184** (13): 3682-3688.

Takeo, M., T. Fujii and Y. Maeda (1998a). "Sequence analysis of the genes encoding a multicomponent dioxygenase involved in oxidation of aniline and *o*-toluidine in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (1): 17-24.

Takeo, M., T. Fujii, K. Takenaka and Y. Maeda (1998b). "Cloning and sequencing of a gene cluster for the *meta*-cleavage pathway of aniline degradation in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (5): 514-517.

Tawfik, D. S. (2004). "Book Review for: Directed enzyme evolution: Screening and selection methods. Methods in molecular biology Vol. 230 edited by Frances H. Arnold and George Georgiou. 2003.: Totowa, NJ: Humana Press. 383 pages. Directed evolution library creation: Methods and protocols. Methods in molecular biology, Vol. 231, edited by Frances H. Arnold and George Georgiou. 2003. Totowa, NJ: Humana Press. 224 pages. Directed molecular evolution of proteins: Or how to improve enzymes for biocatalysis, edited by Susanne Brackmann and Kay Johnsson. 2002. Weinheim, Germany: Wiley-VCH. 357 pages." Protein Science **13**: 2836-2837.

Tesmer, J. J., T. J. Klem, M. L. Deras, V. J. Davisson and J. L. Smith (1996). "The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families." Nat Struct Biol **3** (1): 74-86.

Timmis, K. N. and D. H. Pieper (1999). "Bacteria designed for bioremediation." Trends in Biotechnology **17** (5): 201-204.

Tweedy, B. G., C. Loeppky and J. A. Ross (1970). "Metobromuron: Acetylation of the Aniline Moiety as a Detoxification Mechanism." Science **168** (3930): 482-483.

Urata, M., E. Uchida, H. Nojiri, T. Omori, R. Obo, N. Miyaura and N. Ouchiyama (2004). "Genes involved in aniline degradation by *Delftia acidovorans* strain 7N and its distribution in the natural environment." Biosci Biotechnol Biochem **68** (12): 2457-65.

US EPA (1985). "Health and Environmental Effects Profile for Aniline." United States Environmental Protection Agency (US EPA), Office of Solid Waste and Emergency Response

US EPA (1994a). "Integrated Risk Information System (IRIS) Online. Coversheet for Aniline. " United States Environmental Protection Agency (US EPA), Office of Health and Environmental Assessment

Wackett, L. P. (2002). "Mechanism and applications of Rieske non-heme iron dioxygenases." Enzyme And Microbial Technology **31** (5): 577-587.

Weisburger, J. H. (1997). "A perspective on the history and significance of carcinogenic and mutagenic N-substituted aryl compounds in human health." Mutation Research **376**: 261-266.

Winneke, G., J. Walkowiak and H. Lilienthal (2002). "PCB-induced neurodevelopmental toxicity in human infants and its potential mediation by endocrine dysfunction." Toxicology **181-182**: 161-5.

Yamashita, M. M., R. J. Almassy, C. A. Janson, D. Cascio and D. Eisenberg (1989). "Refined atomic model of glutamine synthetase at 3.5 A resolution." J Biol Chem **264** (30): 17681-90.

Yoshimi, N., S. Sugie, H. Iwata, K. Niwa, H. Mori, C. Hashida and H. Shimizu (1988). "The genotoxicity of a variety of aniline derivatives in a DNA repair test with primary cultured rat hepatocytes." Mutat Res **206** (2): 183-91.

Zachara, J. M., L. J. Felice and J. K. Sauer (1984). "Sorption of aniline on selected Alfisols from the eastern coal region." Soil Science **138** (3): 209-19.

Zepp, R. G., G. L. Baughman and P. F. Schlotzhauer (1981). "Comparison of photochemical behavior of various humic substances in water: I. Sunlight induced reactions of aquatic pollutants photosensitized by humic substances." Chemosphere **10** (1): 109-117.

Zhang, N., B. G. Stewart, J. C. Moore, R. L. Greasham, D. K. Robinson, B. C. Buckland and C. Lee (2000). "Directed Evolution of Toluene Dioxygenase from *Pseudomonas putida* for Improved Selectivity Toward *cis*-Indandiol during Indene Bioconversion." Metabolic Engineering **2** (4): 339-348.

Zhao, H. and F. H. Arnold (1997). "Combinatorial protein design: strategies for screening protein libraries." Curr Opin Struct Biol **7** (4): 480-5.

Zhao, H., K. Chockalingam and Z. Chen (2002). "Directed evolution of enzymes and pathways for industrial biocatalysis." Curr Opin Biotechnol **13** (2): 104-10.

Zhao, H. M., L. Giver, Z. X. Shao, J. A. Affholter and F. H. Arnold (1998). "Molecular evolution by staggered extension process (StEP) in vitro recombination." Nature Biotechnology **16** (3): 258-261.

Zielinski, M., S. Kahl, H. J. Hecht and B. Hofer (2003). "Pinpointing biphenyl dioxygenase residues that are crucial for substrate interaction." Journal Of Bacteriology **185** (23): 6976-6980.

Zielinski, M., S. Kahl, C. Standfuss-Gabisch, B. Camara, M. Seeger and B. Hofer (2006). "Generation of novel-substrate-accepting biphenyl dioxygenases through segmental random mutagenesis and identification of residues involved in enzyme specificity." Appl Environ Microbiol **72** (3): 2191-9.

Chapter 3
Functional expression of aniline and carbazole dioxygenases

3.1 Introduction

In order to engineer AtdA to accept a larger range of substrates and improve its activity by directed evolution, the enzyme has to be functionally expressed. Moreover, improving the expression level of the protein in a host cell can increase the overall activity of the cell, consequently allowing for a more sensitive screening and/or selection method to be developed for the directed evolution process.

Gene fragments containing the *atdA* operon was first isolated and cloned in the pUC19 cloning vector (Takeo, Fujii et al., 1998a) to give the pAS91 and pAS93 vectors. Although the a*tdA* genes in these plasmids were shown to be functionally expressed by *E. coli* hosts, the expression of the AtdA enzyme was not optimized because the pUC19 vector is a cloning vector. In this chapter, only the overall activity of the enzyme will be targeted for improvement. Different promoters and the ribosomal binding site (RBS) sequences upstream of the gene, which can directly influence the expression of a gene, can be used to improve expression and consequently overall activity.

The *atdA* operon was cloned into the pAS91 and pAS93 plasmids from the source micro-organism, *Acinetobacter* sp. strain YAA, without any modifications to the gene sequence. The overlapping gene sequences between the subunits make the current gene structure unsuitable for engineering work as the subunit of interest

cannot be easily excised and ligated back into the plasmid without disrupting the gene of the subunits upstream and downstream of it. Thus there is a need to separate the overlapping genes before engineering work can be done on the subunit of interest.

This chapter describes the construction of the plasmid used for the engineering of AtdA, which includes the use of different vectors and separation of the overlapping gene sequences, as well as the verification of the overall activity of the AtdA in *E. coli*. The functional expression of the CarA enzyme in *E. coli* and the preparation of 2'-aminobiphenyl-2,3-diol (2ABPD) are also described in this chapter.

3.2 Materials and methods

3.2.1 Materials

Aniline, 2,4-dimethylaniline (24DMA), 3,4-dimethylaniline (34DMA), 2-methylaniline (2MA), 2-ethylaniline (2-EA), 2-isopropylaniline (2IPA), 2-secbutylaniline (2SBA), 2-tertbutylaniline (2TBA), catechol, isopropyl-β-D-thiogalactopyranoside (IPTG), dimethylformamide (DMF), ampicillin, and all other chemicals were purchased from Sigma (St Louis, MO) unless otherwise stated. 3-isopropylcatechol (3IPC) was purchased from Chem Service (West Chester, PA). Gibbs' reagent was purchased from MP Biomedicals (Solon, OH). Quikchange XL Site Directed Mutagenesis kit and *Pfu Turbo* DNA polymerase were purchased from Stratagene (La Jolla, CA). Primers were purchased from Integrated DNA Technologies (Coralville, IA) and 1st Base (Singapore). PCR-grade deoxynucleotide triphosphates were obtained from Roche Applied Sciences (Indianapolis, IN). All DNA modifying enzymes were purchased from New England Biolabs (Beverly, MA).

E. coli JM109 and BL21 (DE3) were purchased from Novagen (Madison, WI), while chemical competent *E. coli* DH5α was purchased from the Cell Media Facility

at the University of Illinois (Urbana, IL). The plasmid pTrc99A was from Amersham Pharmacia (Piscataway, NJ). The plasmids pAS91 and pAS93, both containing the aniline dioxygenase gene cluster, were kindly provided by Dr. Masahiro Takeo from the Dept of Applied Chemistry, Himeiji Institute of Technology, Hyogo, Japan. The pUCARA plasmid used in this work was kindly provided by Dr Hideaki Nojiri from the Biotechnology Research Center, The University of Tokyo (Japan).

3.2.2 pTrcA-2 plasmid construction

The plasmid pAS91 was digested with SalI restriction enzyme and the 9kb DNA fragment containing the *atdA* gene cluster was amplified was gel purified using a QIAEX II gel purification kit from Qiagen (Valencia, CA). The pTrc99A plasmid was digested with SalI restriction enzyme and then dephosphrylated using shrimp alkaline phosphatase. The digested vector was then gel purified using the QIAEX II gel purification kit. The 9 kb fragment from pAS91 was ligated to the digested pTrc99A using T4 DNA ligase. The resulting plasmid was pTrcA-2.

3.2.3 pTA1-1plasmid construction

From the plasmid pAS91, the gene segment containing *atdA1A2* was amplified using the primers pTrcA1 F and pTrcA2 RII, and the *atdA3* gene was amplified using pTrcA3 FII and pTrcA3 RII, while the gene segment containing *atdA4A5* was amplified using the primers pTrcA4 FII and pTrcA5 RII. The sequences of primers used are given in Table 3.1. The PCR products were gel purified using a QIAEX II gel purification kit from Qiagen (Valencia, CA) and treated with the restriction enzyme DpnI to remove any residual methylated template from the products. Overlap extension PCR was used to join the three fragments together. The overlap extension

PCR reaction mix consists of 85 ng of *atdA1A2*, 50 ng of *atdA3*, 60ng of *atdA4A5*, 2 μl of 10x *Pfu* buffer, 2 μl of 10x dNTP (mixture of dATP, dTTP, dGTP, and dCTP, each at a concentration of 100 mM), 2 U of *Pfu Turbo* DNA polymerase, and water to make up 20 μl. The PCR program consists of 94 °C for 2 min, 10 cycles of 94 °C for 1 min, 55 °C for 1.5 min, 72 °C for 6 min, and a final extension time of 10 min at 72 °C. The reconstituted *atdA* operon was gel purified, digested with SalI restriction enzyme and ligated into pTrc99A using T4 DNA ligase. The resulting plasmid was pTA1-1.

Table 3.1. Primers used in the cloning of the *atdA1-A5* gene. Underlined bases represent the respective restriction sites.

Primer	Sequence
pTrcA1 F	5'-TAGTTAGATTAAGTCGACGATAGACAAAGGAATGTCCCATGAG-3'
pTrcA2 RII	5'-TCATGGTCTGTTCCTGAATTCTTATACATCACCCACCAAGTTCTG-3'
pTrcA3 FII	5'-GAATTCAGGAACAGACCATGAAAACCATAAATCAACTAATTCAGT-3'
pTrcA3 RII	5'-ATTATTCATGGCGTACCTCAACCGGTTTAAACCTCAGACTGATTC-3'
pTrcA4 FII	5'-ACCGGTTGAGGTACGCCATGAATAATAATAAAGATCTTAGTATAA -3'
pTrcA5 RII	5'-CTGACACCTTACCTTGTCGACTTATTGGTCAAAGGATATTTC-3'

3.2.4 pTA2-3 plasmid construction

Quikchange XL Site Directed Mutagenesis kit was used for this mutation, according to the PCR and transformation protocol recommended in the manual. The primers used to remove the EcoRI site in the pTA1-1 multiple cloning site were MCS QC F (5'-AGGAAACAGACCATGCCATTCGAGCTCGGTACCCGG-3') and MCS QC R (5'-CCGGGTACCGAGCTCGAATGGCATGGTCTGTTTCCT-3'). The primers used to remove the EcoRI site in *atdA2* were A2 QC F (5'-GGTGGGGCCGTATGCAGTCGGTATCCAAGGACATCCAGAG-3') and A2QC R (5'-CTCTGGATGTCCTTGGATACCGACTGCATACGGCCCCACC-3'). The

primers used to introduce EcoRI site flanking the 5' end of *atdA3* were A3 EcoRI F (5'-CTTGGTGGGTGATGTATAAGAATTCAGGAACAGACCATGAAAACC-3') and A3 EcoRI R (5'-GGTTTTCATGGTCTGTTCCTGAATTCTTATACATCACCC ACCAAG-3'). The resulting plasmid was pTA2-3.

3.2.5 Sample preparation for SDS-PAGE analysis

Overnight cell cultures in LB with 100 mg/l of ampicillin were inoculated into fresh LB with ampicillin (100 mg/l) at a volume ration of 1: 100 and incubated in a 37 ºC shaker at a speed of 250 rpm. When optical density (OD) reaches 0.5, the cultures were induced with 1 mM IPTG and incubated in a 30 ºC shaker at a speed of 250 rpm for 3 hr.

After induction, the OD of the cells was measured and the cells were then centrifuged at 6,000 × *g* in a Beckman J2-21M Induction Drive Centrifuge at 4 ºC. The supernatant is discarded and the cell pellet was resuspended in 50 mM Tris-HCl (pH 7.5). The final OD of the cells was adjusted to about 15. Then the cells were disrupted using a single pass through the Constant Systems Cell Disruptor (Warwick, UK) at 20.3 kpsi. The disrupted cells were centrifuged at 16,000 × *g* in a benchtop centrifuge for 5 min. The supernatant was separated from the cell pellet and then stored at -20 ºC until ready for SDS PAGE analysis.

3.2.6 Check for AtdA activity

The cells were grown in LB and when their OD reached 0.5, the cells were induced with 1 mM IPTG at 30 ºC in a shaker at 250 rpm for 2.5 hr. The cells were then washed once with M9 minimal media (Miller, 1972) and resuspended in the M9 minimal media to a final OD of 0.75. Subsequently, the substrates, 2-methylaniline

(2MA) and 2-ethylaniline (2EA), were added separately to the resuspended cells to a final concentration of 2 mM. The cultures were then incubated in at 30 ºC in a shaker at 250 rpm for one day. The colors of the media are monitored.

3.2.7 CarA resting cell assay

The 2xYT media used in carbazole dioxygenase, CarA, resting cell assay was prepared as described by (Sambrook, Fritsch et al., 1989). The minimal media (MM) used assay had the following composition: 800 ml Millipore water, 1.32 g $Na_2HPO_4.7H_2O$, 0.64 g KH_2PO_4, 0.012 g $FeSO_4.7H_2O$, 0.012 g $CaCl_2.2H_2O$, 0.0059 g $MgSO_4$, 0.02 g yeast extract. The initial pH was adjusted to 7.0. The media was then filter sterilized.

A colony of *E. coli* JM109 or BL21 harboring the pUCARA plasmid was grown overnight in 2 ml 2xYT with 100 μg/l ampicillin at 37 ºC under constant shaking. The overnight culture was then inoculated into 200 ml 2xYT with 100 μg/l ampicillin and incubated in a 37 ºC shaker until the OD at 550 nm reaches about 0.8. At that point, protein expression is induced by adding IPTG to a final concentration of 1 mM. The culture is then incubated in a 30 ºC shaker for 6 to 8 hrs. 1 ml of sample was taken at different times during induction for SDS-PAGE analysis. After the induction period, the cells were spun down and washed with ice cold MM twice. The cells were resuspended in MM with a final volume of 20 ml. Carbazole was added to a final concentrations of 100 mg/l or 1 g/l. DMSO was added to a final concentration of 3 % vol/vol.

At stipulated times, samples of the resting cell assay were taken and acetonitrile (ACN) was added at a 1.33 volume ratio to dissolve any insoluble carbazole. The

mixture is vortexed for 1 min and spun down at 400 rpm for 10 min at 4 ºC in an Eppendorf 5810R centrifuge. The supernatant was collected and analyzed by HPLC.

3.2.8 Identification of carbazole and 2ABPD

The identification and quantitation of carbazole and 2'-aminobiphenyl-2,3-diol (2ABPD) was done using an Agilent 1100 series HPLC with the following solvents: H_2O with 1 % acetic acid (solvent C) and acetonitrile with 1 % acetic acid (solvent D). The column used is the SB-C8 reverse phase column. The method used is 10 % to 70 % solvent D in 20 min under a flow rate of 1.5 ml/min. Carbazole and 2ABPD were detected using a wavelength of 260 nm.

3.3 Cloning of atdA operon into expression vector: pTrcA-2

The aniline dioxygenase (AtdA) is a multi-component enzyme isolated from *Acinetobacter* sp. strain YAA. It consists of five separate subunits: AtdA1 which is a glutamine synthetase-like protein, AtdA2 which is a glutamine amidotransferase-like protein, AtdA3 and AtdA4 which resemble the large (α) and small (β) subunits of the terminal class dioxygenase respectively, as well as AtdA5 which is a reductase component (Takeo, Fujii et al., 1998a). The *atdA* gene was originally cloned into the pUC19 vector (pAS91 and pAS93) (Takeo, Fujii et al., 1998a; Takeo, Fujii et al., 1998b).

Although the AtdA enzyme was functionally expressed by *E. coli* hosts harboring the pAS91 and pAS93 plasmids, the expression of the AtdA enzyme was not optimized because the pUC19 vector is a cloning vector. To improve the expression level of the operon, the gene fragment from the pAS91 plasmid encoding the entire *atdA* operon was excised using the SalI restriction enzyme and ligated into

the pTrc99A expression vector using the same restriction sites, yielding the resulting vector pTrcA-2.

3.3.1 SDS-PAGE analysis of AtdA expression by pTrcA-2

The pTrcA-2 vector was transformed into *E. coli* BL21 (DE3) and the expression level of AtdA in pTrcA-2 was compared to that in pAS91, as shown in Figure 3.1. From the figure, it can be seen that the expression level of AtdA subunits is higher in pTA2-3 than pAS91. With pTrcA-2, all the subunits of the AtdA operon can be seen except AtdA5 (~37.2 kDa), the last subunit in the operon.

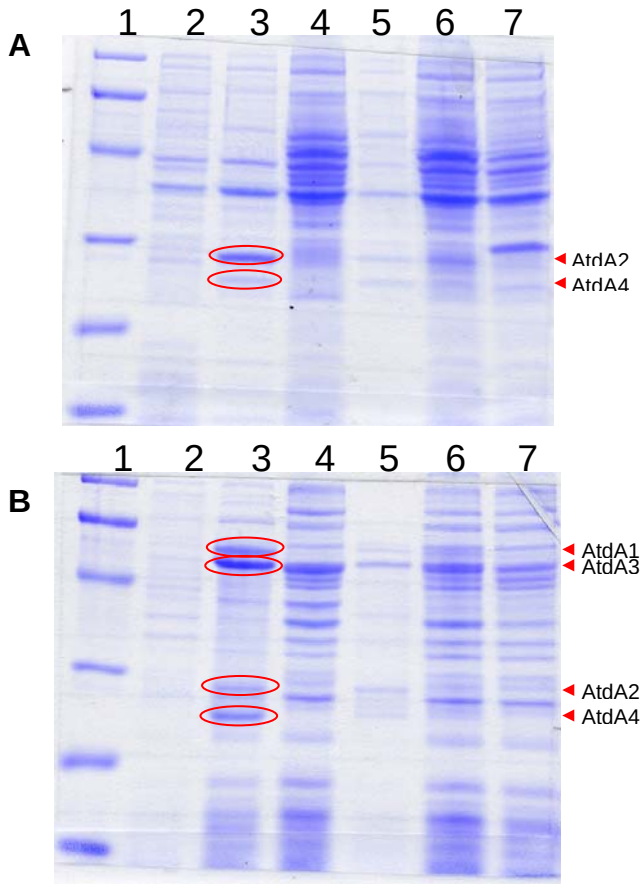


Figure 3.1 SDS-PAGE analysis of AtdA expression by pTrcA-2 and pAS91 in *E. coli* BL21. (A) soluble fraction (B) total fraction.
Lane: (1) Protein marker (from top to bottom: 97.4, 66.2, 45.0, 31.0, 21.5, and 14.4 kDa), (2) pTrcA-2 before induction, (3) pTrcA-2 after 4 hr induction with 1 mM IPTG, (4) pTrc99A (negative control), (5) pAS93 before induction, (6) pAS93 after 4 hr induction with 1 mM IPTG, (7) pUC19 (negative control).
Red circles indicate the bands of the expressed AtdA subunits.

3.4 Introduction of restriction sites flanking AtdA3: pTA2-3

The AtdA subunit targeted for characterization and engineering studies is AtdA3 as it is the α-subunit of a Rieske terminal dioxygenase, which was shown to control the substrate specificity of various dioxygenase enzymes (Tan and Cheong, 1994; Parales, Parales et al., 1998; Parales, Emig et al., 1998; Wackett, 2002). Although the expression level of aniline dioxygenase in pTrcA-2 was increased compared to the original plasmid, pAS91, the construct was still not suitable for characterization and engineering works because the target gene, *atdA3*, overlaps with *atdA2* and *atdA4*. As shown in Figure 3.2A, the start codon of *atdA3* overlaps with the stop codon of *atdA2* while the end of *atdA3* overlaps with *atdA4* between the ribosomal binding site (RBS) and the start codon. Introduction of restriction sites directly to the flanks of *atdA3* will disrupt sequence of *atdA2* and *atdA4*. As such, there was a need to separate the sequences of these three genes when introducing the restriction sites to flank *atdA3*. Overlap extension PCR (Higuchi, Krummel et al., 1988) was used to introduce AvrII and AgeI restriction sites before and after the *atdA3* gene respectively. The *atdA* operon was then cloned into pTrc99A using SalI restriction sites to form the plasmid pTA1-1. The non-coding sequence after *atdA5* in the SalI gene fragment from pTrcA-2 was not amplified in the PCR.

Subsequently, a silent mutation was used to remove the EcoRI restriction site, GAATTC, found on *atdA2*. This EcoRI site was mutated to GTATCC, while the AvrII restriction site at the 5'-end of *atdA3* was mutated to that of EcoRI, which is a more efficient restriction enzyme than AvrII. SDS-PAGE analysis was then used to check the expression of the resulting plasmid, pTA2-3. The complete sequence of pTA2-3 is provided in Appendix A.

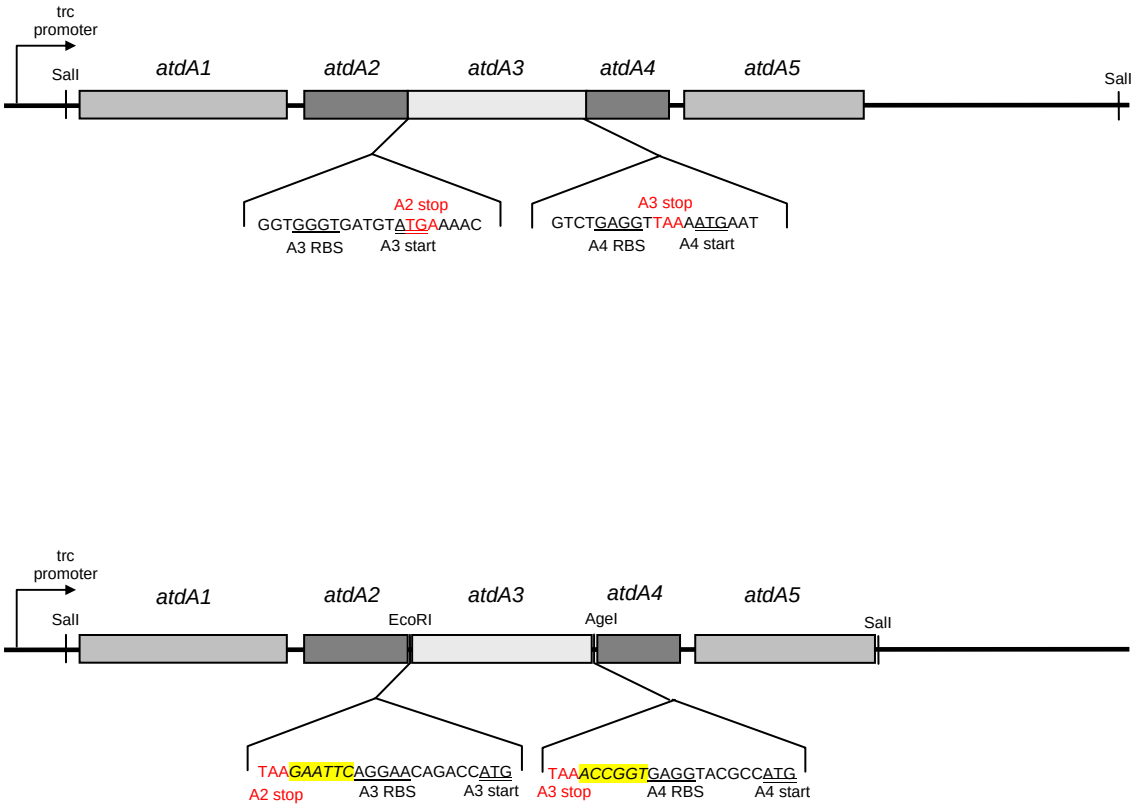


Figure 3.2 Plasmid construct of (a) pTrcA-2, and (b) pTA2-3. Sequences at the 5'- and 3'- ends of *atdA3* are shown in each figure. Red sequences represent stop codons, underlined sequences represent the RBS of each gene, double underlined sequences are the start codons, while sequences italicized and highlighted in yellow represents the restriction sites introduced.

3.4.1 SDS-PAGE analysis of AtdA expression by pTA2-3

The expression level of AtdA from the vector pTA2-3 was then compared to that of pTrcA-2. SDS-PAGE analysis of *E. coli* JM109 with pTA2-3 and pTrcA-2 showed that the expression level of *AtdA* in pTA2-3 is similar to that of pTrcA-2 (Figure 3.3).

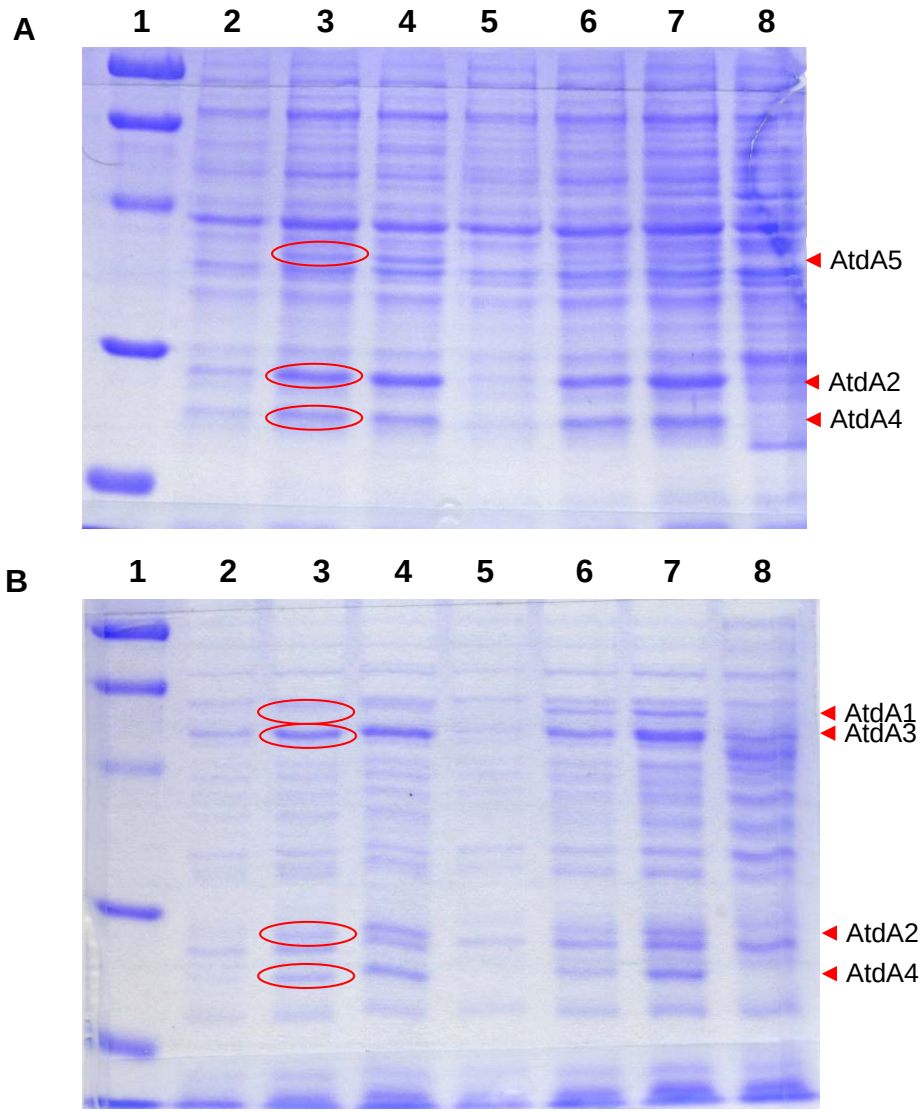


Figure 3.3 SDS-PAGE analysis of AtdA expression by pTA2-3 and pTrcA-2 in *E. coli* JM109 induced with 1 mM IPTG. (A) soluble fraction (B) total fraction.
Lane: (1) Protein marker (from top to bottom: 97.4, 66.2, 45.0, 31.0, 21.5, and 14.4 kDa), (2) pTA2-3 before induction, (3) pTA2-3 after 2 hr induction, (4) pTA2-3 after 4 hr induction, (5) pTrcA-2 before induction, (6) pTrcA-2 after 2 hr induction, (7) pTrcA-2 after 4 hr induction, (8) pTrc99A (negative control).
Red circles indicate the bands of the expressed AtdA subunits.

3.5 AtdA activity of various plasmid constructs

The aniline dioxygenating activities of *E. coli* JM109 cells separately transformed with the original plasmid construct, pAS91, and the new plasmid constructs, pTrcA-2 and pTA2-3, were compared to verify if the higher AtdA expression levels with the new plasmid constructs led to higher activity.

Figure 3.4 shows the cultures after one day of incubation with 2MA and 2EA. The AtdA enzyme converts 2MA and 2EA into their corresponding catechol products, which in turn undergoes autooxidation to give colored compounds. Hence, the higher the intensity of the color formed, the higher the cell activity. As can be observed from the figure, cultures of cells harboring the pTrcA-2 and pTA2-3 plasmids are darker than the cells with pAS91, with the difference more pronounced for 2EA than 2MA. The more pronounced difference in color intensity between pAS91 and pTrcA-2 for 2EA compared to 2MA suggests a slower conversion rate by the enzyme for the larger substrate, 2EA, hence the greater difference between more active cells and less active ones. The difference in color intensity between pTrcA-2 and pTA2-3, on the other hand, was very slight.

These results show that the new plasmid constructs have improved the expression level as well as the overall activity of the AtdA enzyme in *E. coli*. Thus, the plasmid, pTA2-3, was used for the engineering of the AtdA enzyme.

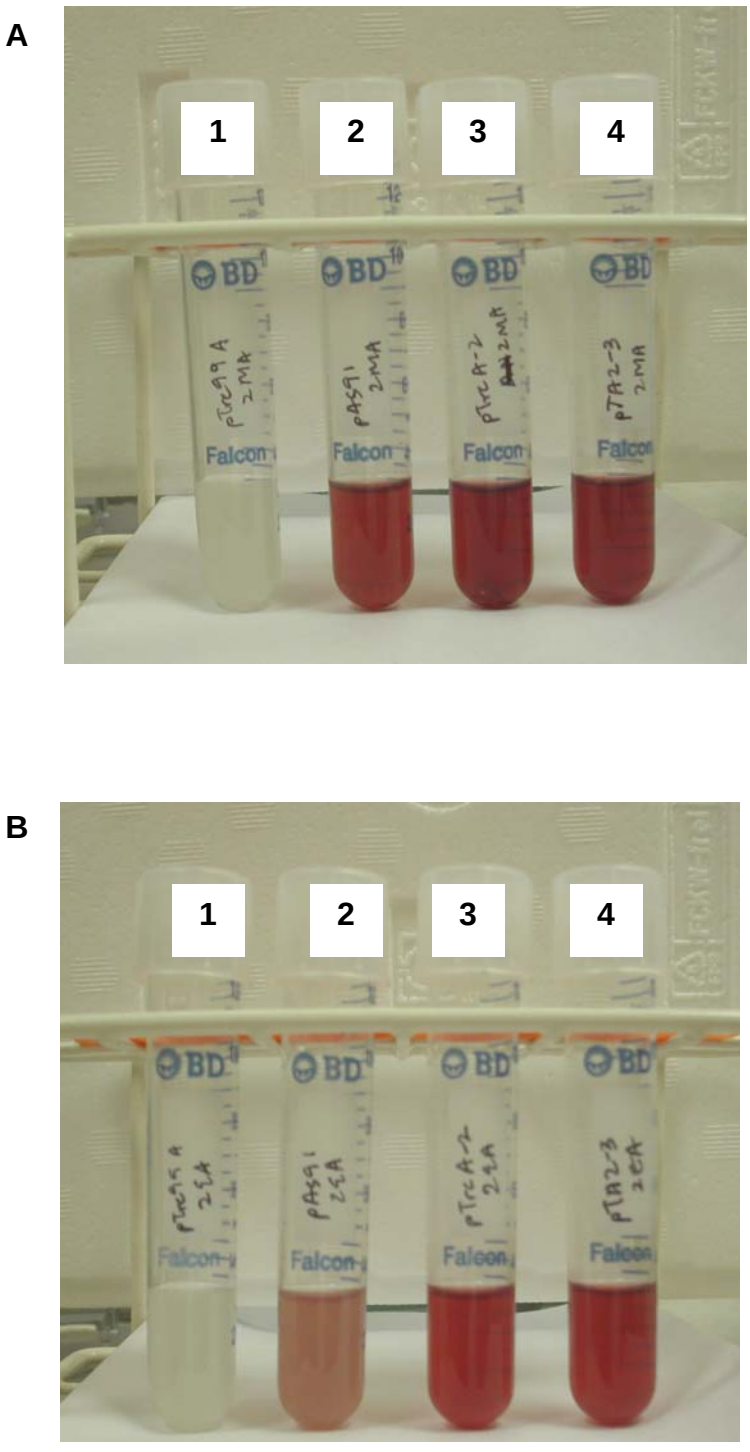


Figure 3.4 *E. coli* cells expressing AtdA from different plasmid constructs after 1 day incubation with 2MA (A), and 2EA (B). Tube 1: pTrc99A (negative control); Tube 2: pAS91 (original vector); Tube 3: pTrcA-2; Tube 4: pTA2-3.

3.6 Functional expression of carbazole-1,9a-dioxygenase

The first step in the proposed carbazole denitrogenation pathway is the angular dioxygenation of carbazole by the carbazole-1,9a-dioxygenase enzyme, CarA, to form 2ABPD (Figure 3.5). The *carA* operon was isolated from *Pseudomonas resinovorans st* CA10 and was cloned into the pUC19 plasmid by Sato et al. in 1997 to give the pUCARA plasmid (Sato, Ouchiyama et al., 1997). CarA is a multi-component enzyme consisting of 3 subunits, CarAa, CarAc and CarAd, with molecular weights of 43, 36 and 11 kDa respectively (Sato, Nam et al., 1997). Subunit CarAa is the α-subunit of a Rieske-type terminal dioxygenase, while CarAc showed 58.0, 57.1, and 57.1 % homologies with the ferredoxin components of chlorobenzene dioxygenase (Werlen, Kohler et al., 1996), biphenyl dioxygenase (Masai, Yamada et al., 1995), and toluene dioxygenase (Zylstra and Gibson, 1989), respectively. CarAd is a reductase component of the dioxygenase, showing 49.3% homology with that of naphthalene dioxygenase (Kurkela, Lehvaslaiho et al., 1988).

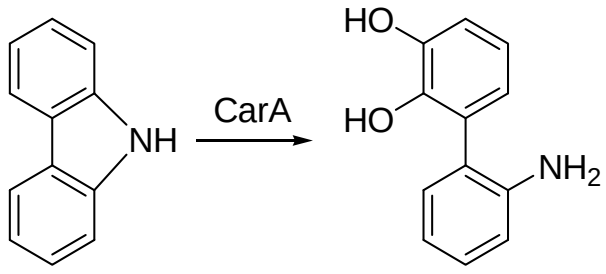


Figure 3.5 Angular dioxygenation of carbazole by CarA enzyme from *Pseudomonas resinovorans st* CA10 to form the product, 2ABPD.

The pUCARA plasmid was transformed into *E. coli* BL21 (DE3) and the expression of the CarA enzyme was analyzed using SDS-PAGE (Figure 3.6). The CarAa and CarAc subunits were expressed, but the CarAd subunit could not be identified as its molecular weight was too low and may therefore have migrated together with protein fragments in the gel front. To further verify if the CarA enzyme was functionally expressed, *E. coli* JM109 cells expressing the enzyme were tested for carbazole dioxygenation activity using the resting cell assay.

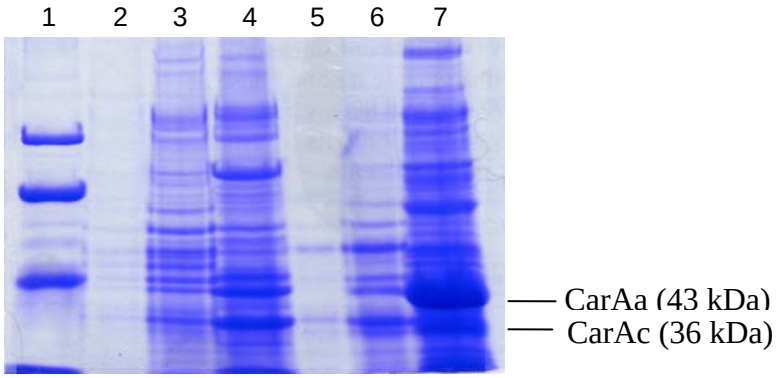


Figure 3.6 SDS-PAGE analysis of *E. coli* BL21 (DE3) with pUCARA plasmids. Lane: (1) marker (from top to bottom: 97.4, 66.2, 45.0, and 31.0 kDa), (2) soluble fraction from negative control, (3) soluble fraction of *E. coli* BL21 with pUCARA, (4) insoluble fraction from negative control, (5) insoluble fraction from *E. coli* BL21 with pUCARA.

From the HPLC analysis of the resting cell assay, it was found that the peak of carbazole, with a retention time of 15.5 min, decreases with time when incubated with *E. coli* JM109 expressing the CarA enzyme, while the area of the product peak, with a retention time of 5.0 min, increased. The identity of the product was then checked using liquid chromatography - mass spectrometry (LC-MS). The compound had an *m/z* of 202 ($M+H^+$), which corresponds to the molecular weight of 2ABPD (201 g/mol). Thus, CarA was functionally expressed in *E. coli* JM109 harboring pUCARA.

3.7 Preparation of 2'-aminobiphenyl-2,3-diol

For directed evolution of the AtdA enzyme, the intermediate of the proposed carbazole denitrogenation pathway, 2ABPD, is required. As 2ABPD is not commercially available, it has to be produced via carbazole dioxygenation with the CarA enzyme. 2ABPD is unstable and undergoes autooxidation if left in the aqueous medium over a period of time. Thus, it needs to be extracted into an organic solvent quickly to prevent its degradation. However, stopping the reaction too early may result in a low product yield, but a longer reaction time for product formation may lead to loss through autooxidation. To find the optimal incubation time and temperature for the production of 2ABPD, a resting cell assay was carried out at three temperatures i.e. 20 ºC, 30 ºC and 37 ºC. The concentration of carbazole, as well as the level of 2ABPD, after incubation with pUCARA transformed *E. coli* JM109, was monitored using HPLC and results are shown in Figure 3.7.

Carbazole was consumed fastest at 37 ºC, with the substrate being completely converted after 7.5 hrs. The highest amount of 2ABPD was also produced at this temperature, where the compound remained stable for up to 22 hours. Hence the reaction should be run at 37 ºC for about 12 hours for optimal yield of 2ABPD.

To produce sufficient 2ABPD for subsequent experiments, the resting cell assay was scaled up to 500 ml with a carbazole concentration of 500 mg/l. However, it was found that the results in the small scale assay were not reproducible when the reaction was scaled up. At larger reaction volumes, the amount of carbazole converted was approximately the same regardless of the starting carbazole concentration (200-500 mg/l, see Figure 3.8). As the assay was carried in minimal media, this trend is more likely caused by exhaustion of cofactors or nutrients than substrate inhibition. When the resting cell assay was carried out in 2xYT medium (Miller, 1972) with 300 mg/l

of carbazole, up to 99.4 % conversion was achieved in 11 hrs, confirming that the media limits the conversion. The 2ABPD produced was then purified using silica gel chromatography, dissolved in dimethylsulfoxide, and stored at -20 ºC for future use (Figure 3.9).

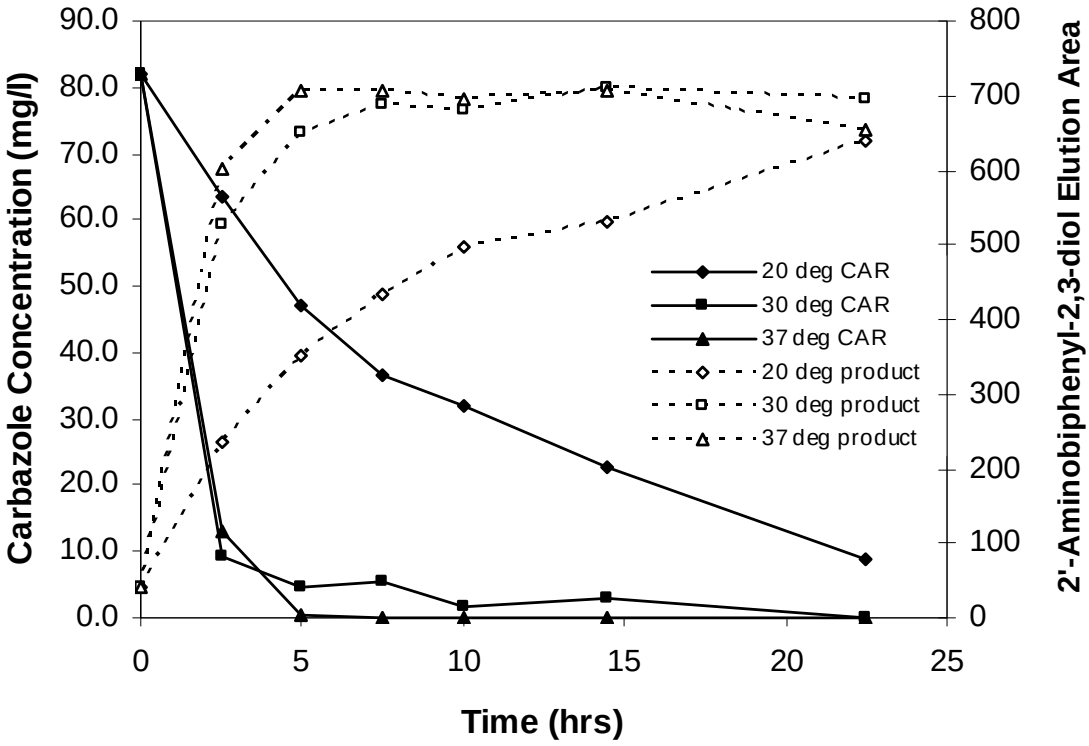


Figure 3.7 Carbazole and 2'-aminobiphenyl-2,3-diol (2ABPD) level with time in the resting cell assay.

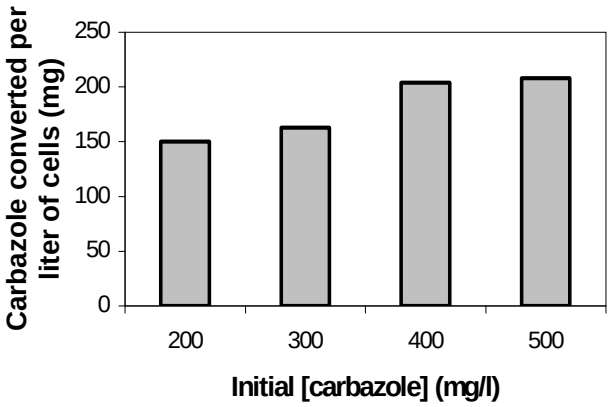


Figure 3.8 Amount of carbazole converted per liter of cells after 16 hr in minimal media.

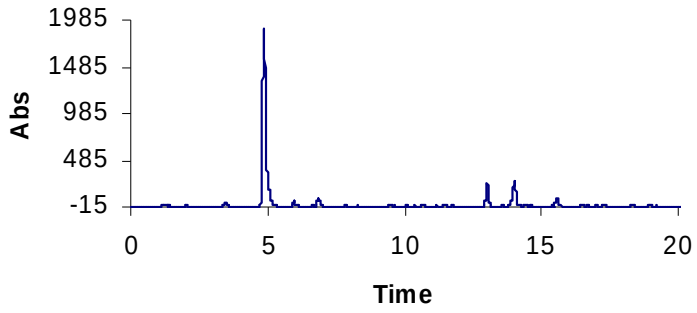


Figure 3.9 HPLC chromatograph of isolated 2ABPD purified by silica gel chromatography. 2ABPD elutes at 5 min.

3.8 Summary

The AtdA and CarA enzymes have been functionally expressed in *E. coli* JM109. The *atdA* gene cluster was cloned into an expression vector pTrc99A and the gene sequence overlap between *atdA2*, *atdA3*, and *atdA4* was removed to facilitate future engineering of *atdA3*. The plasmid to be used for future work is pTA2-3. SDS-PAGE analysis and activity assays have shown that both the expression level as well as the activity of the AtdA enzyme were improved with the new plasmid construct. Furthermore, the activity assay suggested that the activity of the enzyme may be affected by the size of the aniline side-chain. However, further studies with a more extensive range of substrates need to be conducted to verify this observation.

The conditions for conversion of carbazole to 2ABPD by *E. coli* JM109 expressing CarA were also explored. It was found that, over the temperature range tested, the enzyme was most active at 37 ºC. In 2xYT medium, up to 99.4 % of carbazole, at a concentration of 300 mg/l, can be converted in 11 hrs. 2ABPD was also purified for future engineering of AtdA3.

3.9 References

Higuchi, R., B. Krummel and R. K. Saiki (1988). "A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions." Nucleic Acids Res **16** (15): 7351-67.

Kurkela, S., H. Lehvaslaiho, E. T. Palva and T. H. Teeri (1988). "Cloning, Nucleotide-Sequence And Characterization Of Genes Encoding Naphthalene Dioxygenase Of *Pseudomonas-Putida* Strain NCIB9816." Gene **73** (2): 355-362.

Masai, E., A. Yamada, J. M. Healy, T. Hatta, K. Kimbara, M. Fukuda and K. Yano (1995). "Characterization Of Biphenyl Catabolic Genes Of Gram-Positive Polychlorinated Biphenyl Degrader *Rhodococcus* Sp Strain RHA1." Applied And Environmental Microbiology **61** (6): 2079-2085.

Miller, J. H. (1972). Experiments in molecular genetics. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Parales, J. V., R. E. Parales, S. M. Resnick and D. T. Gibson (1998). "Enzyme specificity of 2-nitrotoluene 2,3-dioxygenase from *Pseudomonas* sp. strain JS42 is determined by the C-terminal region of the alpha subunit of the oxygenase component." J Bacteriol **180** (5): 1194-9.

Parales, R. E., M. D. Emig, N. A. Lynch and D. T. Gibson (1998). "Substrate specificities of hybrid naphthalene and 2,4-dinitrotoluene dioxygenase enzyme systems." J Bacteriol **180** (9): 2337-44.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor, N. Y, Cold Spring Harbor Laboratory.

Sato, S. I., J. W. Nam, K. Kasuga, H. Nojiri, H. Yamane and T. Omori (1997). "Identification and characterization of genes encoding carbazole 1,9a-dioxygenase in *Pseudomonas* sp. strain CA10." J Bacteriol **179** (15): 4850-8.

Sato, S. I., N. Ouchiyama, T. Kimura, H. Nojiri, H. Yamane and T. Omori (1997). "Cloning of genes involved in carbazole degradation of *Pseudomonas* sp. strain CA10: nucleotide sequences of genes and characterization of *meta*-cleavage enzymes and hydrolase." J Bacteriol **179** (15): 4841-9.

Takeo, M., T. Fujii and Y. Maeda (1998a). "Sequence analysis of the genes encoding a multicomponent dioxygenase involved in oxidation of aniline and *o*-toluidine in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (1): 17-24.

Takeo, M., T. Fujii, K. Takenaka and Y. Maeda (1998b). "Cloning and sequencing of a gene cluster for the *meta*-cleavage pathway of aniline degradation in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (5): 514-517.

Tan, H. M. and C. M. Cheong (1994). "Substitution of the ISP alpha subunit of biphenyl dioxygenase from *Pseudomonas* results in a modification of the enzyme activity." Biochem Biophys Res Commun **204** (2): 912-7.

Wackett, L. P. (2002). "Mechanism and applications of Rieske non-heme iron dioxygenases." Enzyme And Microbial Technology **31** (5): 577-587.

Werlen, C., H. P. Kohler and J. R. van der Meer (1996). "The broad substrate chlorobenzene dioxygenase and *cis*-chlorobenzene dihydrodiol dehydrogenase of *Pseudomonas* sp. strain P51 are linked evolutionarily to the enzymes for benzene and toluene degradation." J Biol Chem **271** (8): 4009-16.

Zylstra, G. J. and D. T. Gibson (1989). "Toluene degradation by *Pseudomonas putida* F1. Nucleotide sequence of the *todC1C2BADE* genes and their expression in *Escherichia coli*." J Biol Chem **264** (25): 14940-6.

Chapter 4
Screening and selection methods for AtdA

4.1 Introduction

The overall goal of this project was to use directed evolution to engineer the aniline dioxygenase (AtdA) into an enhanced biocatalyst with both improved activity and widened substrate specificity. The two key steps in the successful implementation of directed evolution to improve an enzyme are the generation of a diverse genetic library and screening or selection to identify the desired clones. In directed evolution, the typical library size is much larger than the number of protein variants that can be screened. Moreover, methods for generating a diverse genetic library are generic, while the screens or selection methods for activity are specific for each enzyme. These factors make the availability of a high-throughput screen or selection the bottleneck for most directed evolution endeavors (Aharoni, Griffiths et al., 2005). According to the first law of directed evolution, "you get what you screen for" (You and Arnold, 1996; Schmidt-Dannert and Arnold, 1999). Hence the development of a screening or selection system is very important for the engineering of the aniline dioxygenase enzyme.

Both screening and selection methods can be used for the identification and isolation of improved variants from a library of variants. However, there are differences between the two methods and each has its own advantages and drawbacks. Screening involves the probing of each individual clone in the library for the desired phenotype. In this case, all mutants in the library are analyzed, providing a quantitative measurement of

the each variant. This method is thorough and useful for identifying variants with novel but low activity. However, the thorough probing of the library comes at the price of throughput as all mutants have to be analyzed. The size of the library that can be screen will depend on the availability of automated screening technologies. Typically, a library with a population of about ~10^3-10^4 can be analyzed by each screen.

On the other hand, selection is a process whereby the survival or death of a cell in a genetic library is linked to its ability to display the desired phenotype. The positive selection method, where only variants with desired phenotypes will survive the selection, allows a much larger number of variants (~10^7) to be analyzed at a single time. The major limitation to the size of the library that can be analyzed lies in the transformation efficiency of the host organism, which is 10^6-10^7 for *E. coli.* However, it is not always easy to link survival to a desired property and the host organisms can sometimes adapt to the selection pressure through genetic drift. Another drawback of this method is that it can be difficult to quantify the desired property of variants. Moreover, this method may not be able to identify mutants with novel but low activity as the selection pressure may be too strict for cells displaying low activity to survive.

This chapter describes the development of an efficient system for identifying aniline dioxygenase mutants with improved or novel activity. Both the screening and selection were developed and evaluated for this purpose. For the screening method, various colorimetric assays to detect the substrate, or products were tested and evaluated. For the selection method, the survival of clones was linked to their ability to utilize aniline homologue substrates as the sole source of nitrogen.

4.2 Materials and methods

4.2.1 Chemicals

3-isopropylcatechol (3IPC) was purchased from Chem Service (West Chester, PA). Gibbs' reagent was purchased from MP Biomedicals (Solon, OH). All other chemicals were purchased from Sigma Aldrich (St Louis, MO).

4.2.2 Indophenol blue assay

Phenol solution was prepared by adding 95 % v/v ethanol to 0.11 ml of liquified phenol (≥ 89 %) to make up a volume of 1 ml. Alkali citrate solution was prepared by mixing 2 g of trisodium citrate, and 0.10 g of sodium hydroxide with 10 ml of distilled water. The oxidizing solution was prepared by mixing 1 ml of alkali citrate with 0.25 ml of 5 % sodium hypochlorite. The oxidizing solution was prepared fresh daily, as required.

To 100 µl of ammonium sulfate solution, 4 µl of phenol solution, 4 µl of 5 g/l sodium nitroprusside (stable when stored in the dark for up to 1 month) and 10 µl oxidizing solution were added. The color was left to develop in subdued light for 1 hr and absorbance at 640 nm was measured.

4.2.3 MBTH assay

N-methylbenzothiazolinone-2-hydrazone (MBTH) solution was prepared by dissolving 1 g of MBTH in 1 l of distilled water. Ammonium cerium sulfate solution was prepared by dissolving 2 g of ammonium cerium (IV) sulfate dehydrate in 1 l of distilled water acidified with 4 ml sulfuric acid. To 0.5 ml of sample, 0.5 ml each of the MBTH

and ammonium cerium sulfate solutions was added. The absorbance at 564 nm was then measured.

4.2.4 van Urk reagent assay

The van Urk solution was prepared by adding 1 g of van Urk reagent (1 % p-dimethylaminobenzaldehyde in N hydrochloric acid) and 10 ml of concentrated HCl to 40 ml of distilled water. The solution was stored in the dark and prepared fresh weekly. To 1 ml of sample, 10 µl of van Urk solution was added. After incubation at 30 ºC for 30 min, absorbance at 435 nm was measured.

4.2.5 Gibbs' reagent assay

The pH of an 80 µl sample was adjusted by adding appropriate volumes of 1 M Tris HCl and 0.1M HCl. Then, 10 µl of 0.32 % (w/v) of Gibbs' reagent in ethanol was added to the sample and absorbance at 560 nm was measured.

4.2.6 Gibbs' reagent solid phase screen

E. coli JM109/pTrcA-2 was first grown overnight at 37 ºC on LB plates with 100 mg/l ampicillin and 80 µl of 50 mM IPTG spread on top. The cells were then transferred by a nylon membrane to a piece of filter paper soaked with 2 mM 2-ethylaniline (2EA) and incubated at 30 ºC for 1 hr. The membrane was then transferred to another filter paper soaked with 0.25 % w/v Gibbs' reagent and incubated at 30 ºC for 30 min.

4.2.7 Autooxidation screen

E. coli JM109 cells harboring pTA2-3 plasmid were streaked onto nylon membranes on LB agar plates, which contained 100 mg/l ampicillin and 0.5 mM IPTG, and incubated overnight at 37 ºC such that protein expression was induced as the colonies were forming. The membranes were then transferred to M9 minimal media agar plates containing 0.5 mM IPTG and 5 mM of the desired substrates and incubated at 30 ºC for up to 24 hr. The plates were inspected regularly to monitor the autooxidation of catechol products.

4.2.8 Selection

In the liquid phase selection method, overnight cultures of *E. coli* JM109 strain harboring pAS93 plasmid were first washed with M63 minimal media (Miller, 1972) lacking ammonium sulfate, and inoculated into the same M63 media supplemented with 100 mg/l of ampicillin. IPTG and nitrogen sources, in the form of ammonium sulfate, aniline, or 2EA were added in desired amounts. The cultures were then incubated in a 30 ºC shaker at 250 rpm. The cell densities of the cultures were monitored over the course of 72 hrs.

In the solid phase selection method, overnight cultures of *E. coli* JM109 strain harboring pAS93 plasmid were first washed with M63 minimal media lacking ammonium sulfate and then inoculated onto an M63 minimal media agar plate with no ammonium sulfate added, and supplemented with 100 mg/l ampicillin, 1 mM IPTG, and 4 mM 2EA. The plates were then incubated at 30 ºC and monitored for colony growth.

4.3 Screening

Screening involves the probing of each individual clone in the library for the desired phenotype. In this case, the desired phenotype is improved activity for substrates already accepted by the wild type aniline dioxygenase (WT), and novel activity for substrates not accepted by the WT. There are three possible targets to detect or measure the activity of the enzyme in the aniline dihydroxylation reaction, i.e. the substrate, the catechol product, and the ammonium ion released (Figure 4.1). Screening methods in the liquid phase as well as on agar plates were tested to develop a suitable system.

The desired properties of the method are:

(i) Screen for the property of interest directly, if possible;

(ii) Be sensitive enough to identify improved mutants from background signals;

(iii) Be efficient enough to allow for screening of the large genetic libraries generated;

(iv) Be general enough to be applied to screen for activity on aniline and its various homologue substrates.

Three substrates, aniline, 2-methylaniline (2MA), and 2-ethylaniline (2EA), which were accepted by the AtdA enzyme (discussed in the next chapter), as well as 2-isopropylaniline (2IPA), or their corresponding catechols were used for the evaluation of the screening methods.

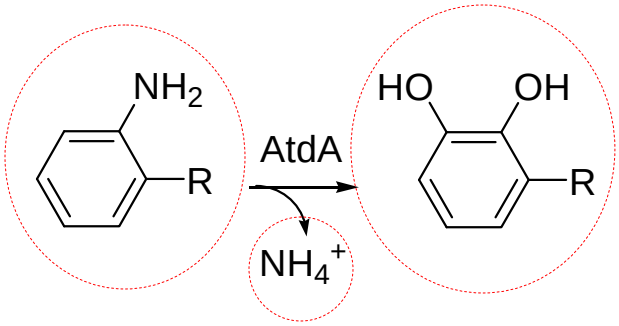


Figure 4.1 Dioxygenation reaction of aniline and its homologues by AtdA. Possible targets for activity measurement are circled. R = hydrogen, methyl, or ethyl group.

The 96-well microplate reader provides an automated platform for the rapid screening of a library based on absorbance. As such, several methods of colorimetric screening were explored. To measure the ammonium ion released, the indophenol blue assay, while MBTH and the van Urk reagent were used to measure the concentration of the substrate. Gibbs' reagent was used to measure the catechol concentration.

4.3.1 Indophenol blue

Since ammonium ions are released in the hydroxylation reaction regardless of the substrate, it is a good target for measuring activity. The indophenol blue assay is a common method for detecting ammonium ions in a variety of samples, from wastewater (Clesceri, Greenberg et al., 1999) to blood plasma (Horn and Squire, 1966). It is based on the reaction of ammonia, hypochlorite, and phenol, catalyzed by sodium nitroprusside, to produce an intensely blue compound, indophenol. The assay used in this work was adopted from the Standard Methods for the Examination of Water and Wastewater by the American Public Health Association (Clesceri, Greenberg et al., 1999).

The standard curve for the calibration of ammonium concentration was prepared using ammonium sulfate as the standard (Figure 4.2). The absorbance increases linearly with the ammonium concentration up to 0.2 mM, after which the intensity of the color exceeds the measurement limit of the spectrometer. Although this method is very sensitive, the aniline substrate was also found to react with the reagents to produce the same blue color, interfering with the measurement of the ammonium ions. The lack of discernible color difference between the substrate and the product makes it impossible to measure concentration of either species. Moreover, this method was laborious as a number of reagents have to be added to the sample during the assay. Hence this method was not pursued further.

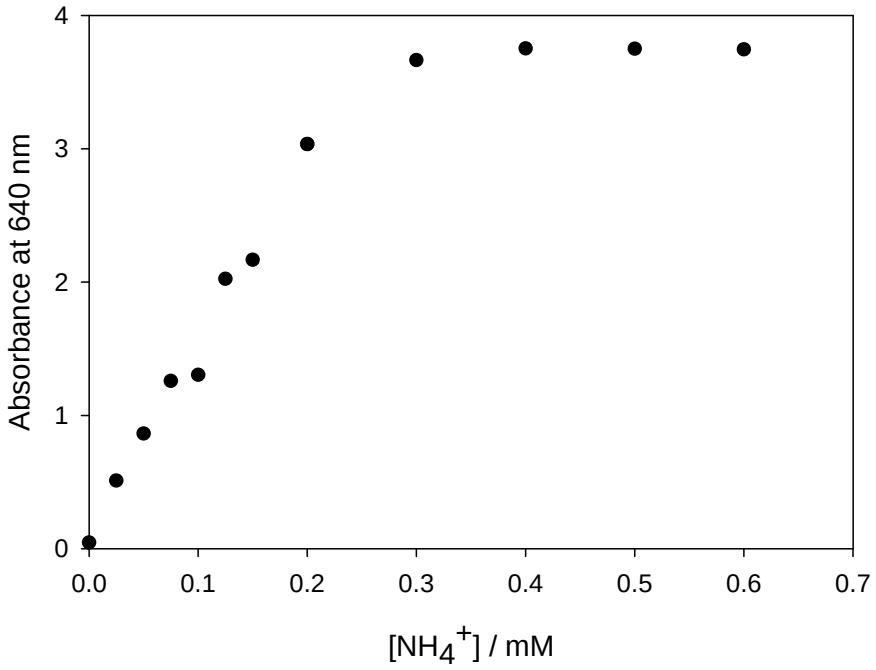


Figure 4.2 Calibration curve of ammonium concentration using the indophenol blue assay.

4.3.2 MBTH reagent

The MBTH assay (Gasparic, Svobodova et al., 1977; Pospisilova, Svobodova et al., 1990; Pospisilova, Polasek et al., 1998) was used to detect aniline and its *ortho*-substituted homologues. In acidic reaction conditions in the presence of an oxidizing agent, MBTH first forms an electrophilic intermediate (Figure 4.3A), which is the active coupling species. The intermediate then undergoes electrophilic substitution with aniline to form the colored compound (Figure 4.3B). The compound is magenta in color and has a strong absorbance at 564 nm (Figure 4.4).

However, it was found that MBTH reacted with catechol as well to give a purple colored compound (with lower intensity than aniline) that overlaps with the absorbance spectrum of aniline, thereby interfering with the accurate measurement of the aniline concentration (Figure 4.4). As the absorbance of the MBTH-catechol product is lower than that of aniline, there is still a possibility that the assay can be applied as an endpoint assay to qualitatively identify active mutants.

To further investigate the effect of catechol interference in the MBTH assay, *E. coli* JM109 cells expressing functional aniline dioxygenase from the pAS93 plasmid were incubated with 1 mM of aniline and, at various times, the supernatant was collected and analyzed with MBTH. As aniline is converted to catechol, the initial absorbance at 564 nm decreased, but the rate of absorbance decrease slowed after 3 hrs as a result of catechol built up. In fact, after 15 hrs, the absorbance increased, indicating that the interference of catechol on the assay was significant (Figure 4.5). Hence this method was not pursued further.

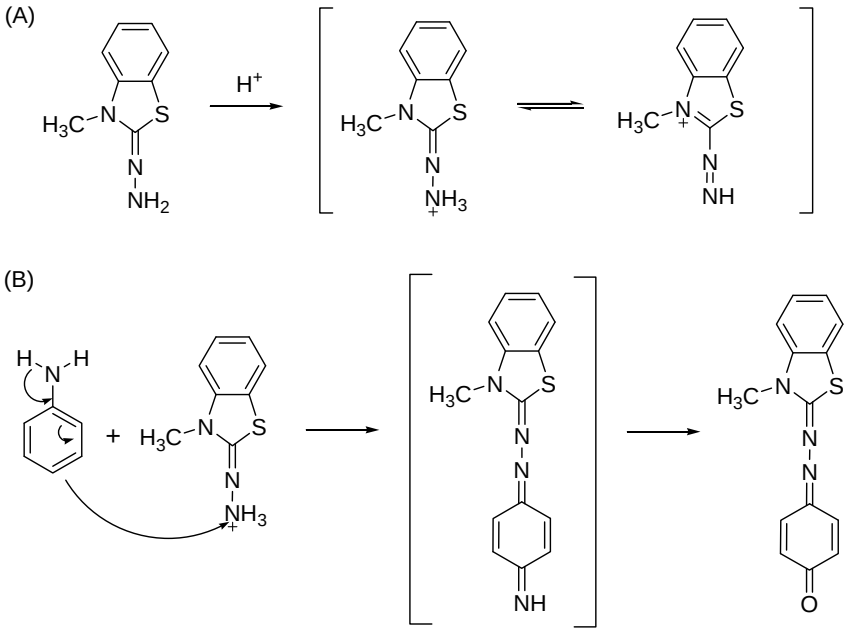


Figure 4.3 (A) Formation of the active coupling intermediate of MBTH. (B) Electrophilic substitution of the intermediate by aniline to form the colored compound.

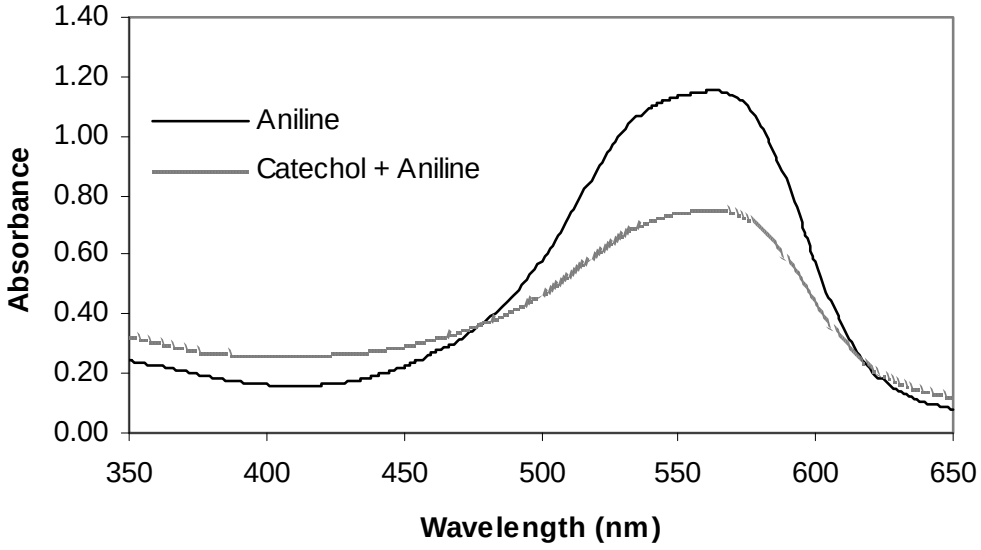


Figure 4.4 Absorbance spectrum of MBTH assay with aniline and a mixture of catechol and aniline.

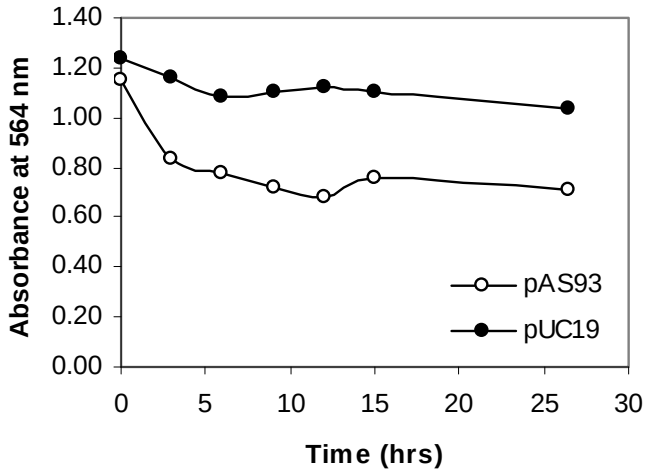


Figure 4.5 Absorbance of the supernatant collected from *E. coli* JM109 cultures with the plasmid pAS93 and expressing AtdA incubated with 1 mM aniline. MBTH was used as the detecting reagent. *E. coli* JM109 with the empty vector pUC19 was used as negative control.

4.3.3 van Urk reagent

The van Urk reagent is used for the detection of organic compounds in thin layer chromatography (TLC) (Ehmann, 1977). When added to aniline, a yellow colored compound was formed with an absorbance peak at 435 nm. The calibration curve of van Urk reagent with two substrates, aniline and 2EA, are shown in Figure 4.6. The molar absorbance of the colored product of the van Urk reagent with 2EA was found to be low. To further investigate the efficacy of the van Urk assay, *E. coli* BL21 (DE3) with the pAS93 plasmid expressing functional AtdA was incubated with 2EA and the cultures were tested with van Urk reagent at various times (Figure 4.7). It was found that the catechol product also reacted with the reagent to give a faint yellowish–orange color. Coupled with the weak absorbance of 2EA, this interference of the catechol product with the assay reagent makes it difficult to detect significant difference in the substrate level between the positive and negative controls.

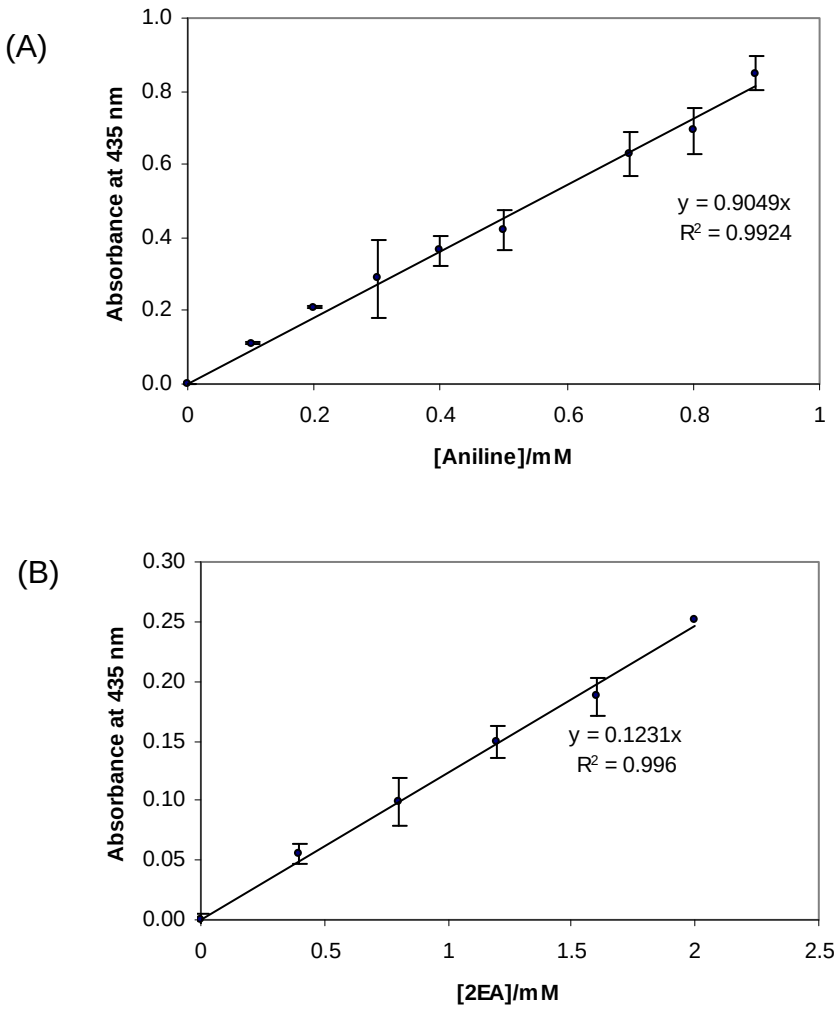


Figure 4.6 Calibration curves of (A) aniline, and (B) 2EA using van Urk's reagent.

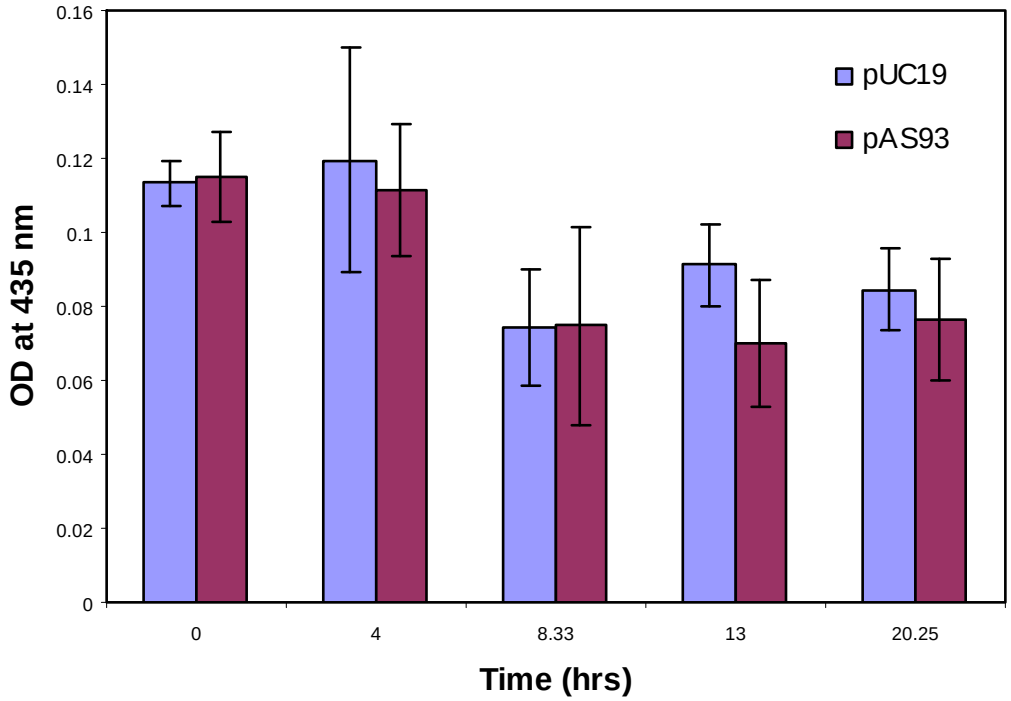


Figure 4.7 Absorbance of *E. coli* JM109 cultures with the plasmid pAS93 and expressing AtdA incubated with 1 mM 2EA after adding van Urk reagent. *E. coli* JM109 with the empty vector pUC19 was used as a negative control.

4.3.4 Gibbs' reagent assay

The Gibbs' reagent (2,6-dichloroquinone-4-chloroimide) is a common reagent used for the detection of phenols (Gibbs, 1927). The reaction scheme of the Gibbs' reagent is shown in Figure 4.8 (Quintana, Didion et al., 1997). According to the reaction scheme, the *para* position of the phenol is involved in the electrophilic substitution reaction with the Gibbs' reagent. Hence phenols substituted with poor leaving groups at the *para* position of the aromatic ring will not react with Gibbs' reagent.

Catechol and 3-isopropylcatechol (3IPC) reacted with Gibbs' reagent to give a pink colored compound with absorbance peak at 560 nm (Figure 4.9). Gibbs (1927) reported that the optimal pH range for the reaction of Gibbs' reagent with phenol is from 8 to 10. However, it was found that at the pH of 8.2, the colored compound formed by the

reaction of Gibbs' reagent with catechol was highly unstable and precipitated after a few minutes of formation. Thus, the stability of the colored compound was titrated against the pH of the reaction media (Figure 4.10). As the pH was reduced, the stability of the colored compound was improved but the rate of color formation was slowed down. At pH 5.8, the colored product was stable up to 5 min after the maximum color intensity was reached. Despite reaching the maximum absorbance at a slower rate, the intensity of the color produced from catechol remained approximately constant at lower pH.

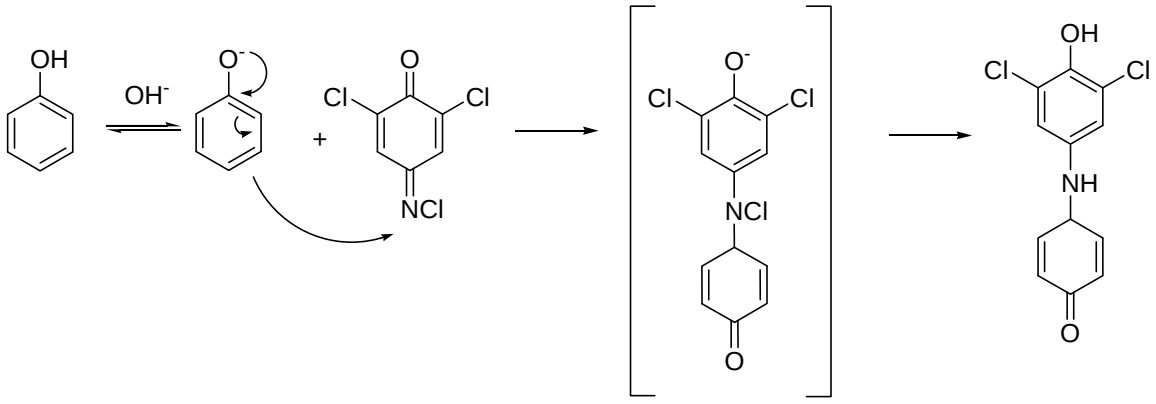


Figure 4.8 Schematic of the reaction between Gibbs' reagent and phenol.

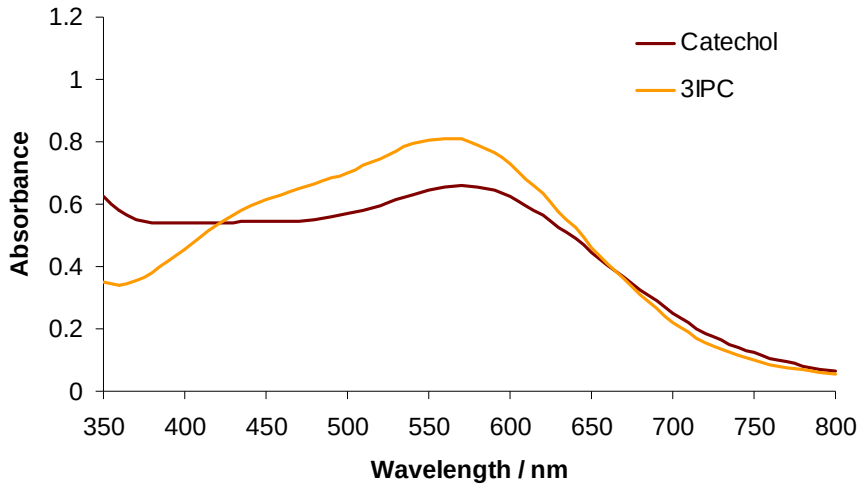


Figure 4.9 Absorbance spectrum of Gibbs' reagent with 1mM of various catechols.

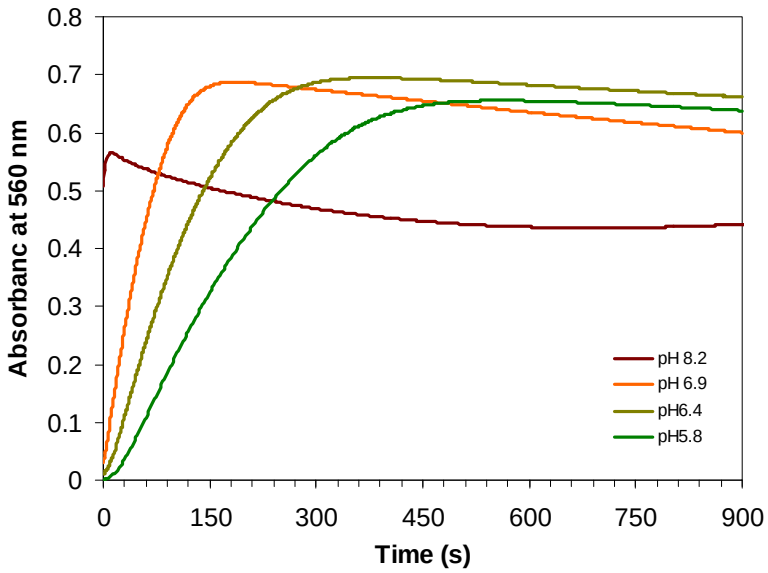


Figure 4.10 Absorbance of the Gibbs' reagent-catechol reaction with time. The decline in absorbance after the maximum point is caused by precipitation of the colored compound. 1mM of catehcol was used.

When the Gibbs' reagent was added to aniline and 2IPA, blue colored compounds with maximum absorbance from 660 to 680 nm were formed (Figure 4.11). However, the rate of color formation was much slower compared to that of catechol (Figure 4.12). Fifteen minutes after the addition of Gibbs' reagent, the absorbance of aniline was more than 10-fold lower than that of catechol. The low interference of aniline in this case prompted further investigation of the Gibbs' reagent's efficacy as a screening method.

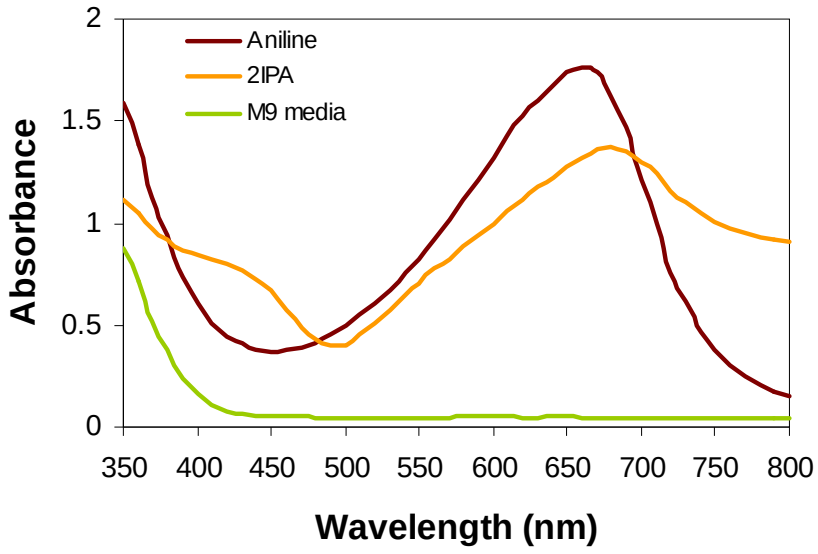


Figure 4.11 Absorbance spectrum of Gibbs' reagent with 1mM aniline and its homologues. M9 media serves as blank.

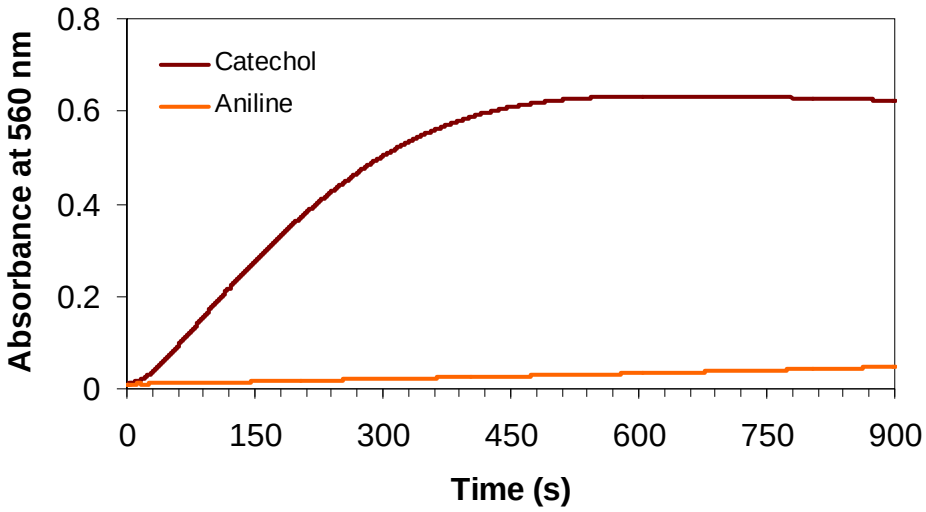


Figure 4.12 Rate of color formation of 1mM of aniline and catechol when reacted with Gibbs' reagent at pH 5.8.

In an actual screening library, the substrate and the product mostly exist as a mixture in the media. Hence the efficacy of Gibbs' reagent to detect catechol in such a mixture was investigated. Different amounts of aniline and catechol were mixed together, with the total concentration of both compounds adding up to 1 mM, and Gibbs' reagent was added to these mixtures at pH 5.8. The absorbance of the colored products formed with time was then monitored (Figure 4.13). In an aniline and catechol mixture, both compounds react with Gibbs' reagent to form their respective colored compounds at a faster rate. In the first 3 minutes of the reaction, only the pink color produced by catechol was observed and the absorbance was approximately proportional to the concentration of the catechol present (relative extinction coefficient ~ 0.87 $mM^{-1}cm^{-1}$).

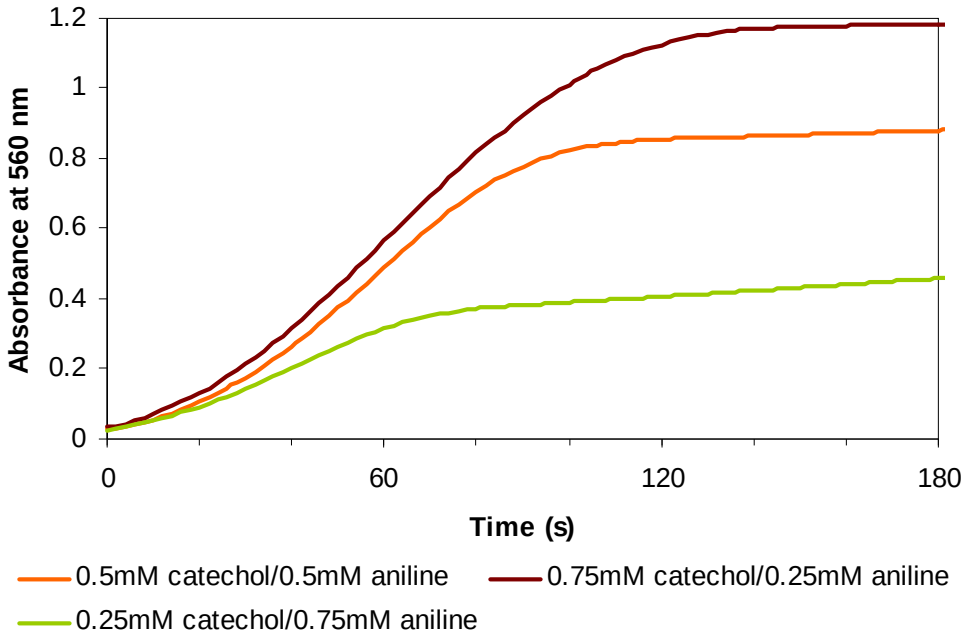


Figure 4.13 Absorbance of the products of Gibbs' reagent and catechol-aniline mixtures with time.

The calibration curve of the Gibbs' reagent with catechol was prepared by adding aniline to make up the total concentration of the samples to 1 mM (Figure 4.14). The same was done for the 2IPA and 3IPC mixture. The absorbance of the colored product from catechol in the catechol-aniline mixture increases with the concentration of catechol, making it a suitable method of detection. The absorbance of 3IPC is not strong and plateaus off after 0.75 mM, but this assay is still applicable as a screening method since the AtdA enzyme does not have any activity on 2IPA.

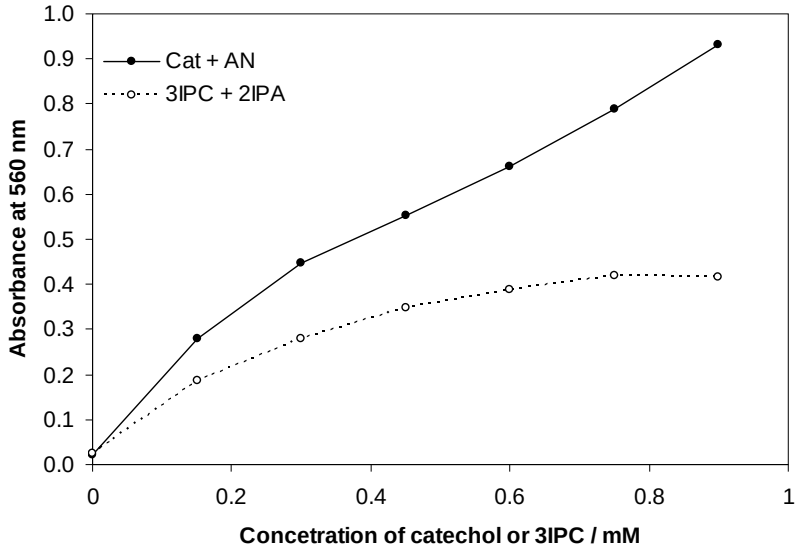


Figure 4.14 Absorbance of colored products from the reaction of Gibbs' reagent with aniline-catechol and 2IPA-3IPC mixtures. The total concentration of components is always 1 mM. Cat represents catechol, and AN represents aniline.

4.3.5 Solid phase screening

To overcome the problem faced in liquid phase screening, a screening method using Gibbs' reagent was carried out in solid media (agar plate). Induced colonies of *E. coli* JM109 harboring pTrcA-2 plasmids were membrane transferred to a piece of filter paper with 2EA and incubated at 30 ºC. The membrane was then transferred onto another piece of filter paper soaked with Gibbs' reagent. The membrane turned blue after incubation on the Gibbs' reagent due to the presence of 2EA. The area around JM109 cells carrying the empty vector remained blue whereas the area surrounding JM109/pTrcA-2 turned to a purplish pink color due to the presence of the hydroxylated product (Figure 4.15). The products formed did not mix with the substrate because of the diffusion limitation on the solid phase. As such, active mutants can easily be identified via a visual screen of the

plate. This also allows for a high throughput screen without the laborious step of colony picking.

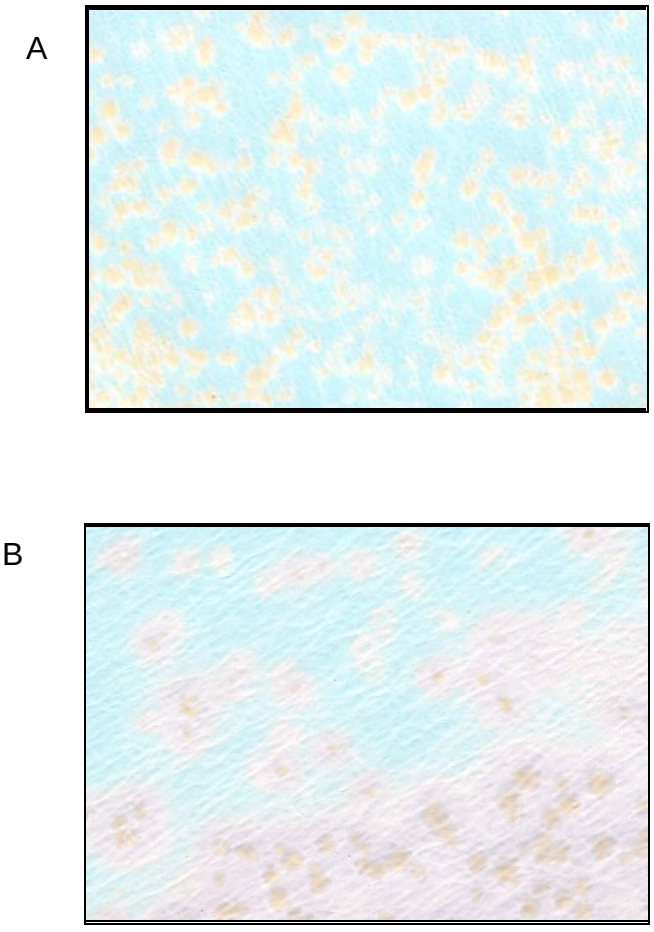


Figure 4.15 Nylon membranes with (A) *E. coli* JM109/pTrc99A (negative control) colonies; (B) *E. coli* JM109/pTrcA-2 colonies after incubation on Gibbs' reagent plate.

The sensitivity of the Gibbs' reagent screen was verified by its ability to identify *E. coli* JM109/pTrcA-2 from a mixture of JM109/pTrcA-2 and JM109/pTrc99A (empty vector). Separate overnight cultures of JM109/pTrcA-2 and JM109/pTrc99A were first diluted 10^4 fold and then plated onto an LB plate at a ratio of 1:10. After overnight incubation at 37 ºC, the colonies were membrane transferred onto an LB plate with IPTG for 8 hrs, and then incubated with 2EA for another 3 hrs. As shown in the Figure 4.16, the active clones can be identified from those with empty vectors by the pink halo formed around the colonies through the reaction of Gibbs' reagent with the phenolic products from 2EA dioxygenation. This assay is sensitive enough to identify active colonies among a population of ~500 colonies per plate.

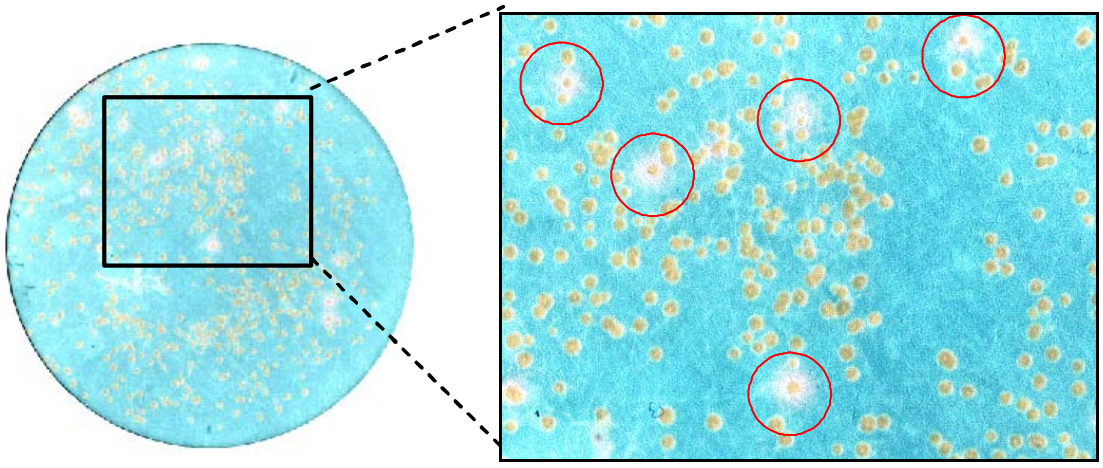


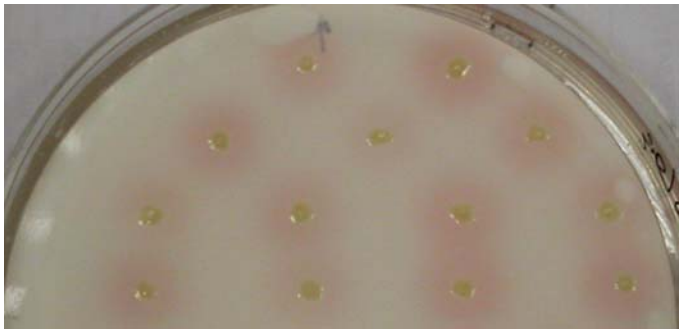
Figure 4.16 *E. coli* JM109 with pTrcA-2 and JM109 with pTrc99A (empty vector) on a nylon membrane at a ratio of 1:10 after incubation with 2EA and Gibbs' reagent screen. Colonies containing the active AtdA enzyme can be identified by the pink halos around them.

The activity of the enzyme is indicated by the radius of the product halo around a colony. However, the size of plated colonies from a transformation reaction varies and this property can affect the amount of substrate converted. For accurate measurement of activity, the radius of the product halo should be normalized to the size of the colony. This would require the use of a high resolution camera and imaging software to process the results (Joern, Sakamoto et al., 2001). As such, this method is more suitable for the detection of novel activity rather than for improved activity. As the goal of the project is to widen the substrate specificity as well as to improve the activity of the enzyme, the liquid phase method was preferred.

4.3.6 Autooxidation

Catechols are unstable in the presence of light and spontaneously polymerize to form colored compounds (Kunz and Chapman, 1981; Fujii, Takeo et al., 1997; Barriault, Plante et al., 2002; Meyer, Schmid et al., 2002), thus allowing the identification of phenolic products around active clones on an agar plate. Using this method, the characteristic colored ring around the colonies appeared only after 5 hrs of incubation with 2-methylaniline and 2-ethylaniline (Figure 4.17). This method is similar to the Gibbs' reagent solid phase screen and has the same limitations. As such, the Gibbs' reagent liquid phase screen was still preferred.

A



B

Figure 4.17 *E.coli* JM109 expressing AtdA after 5 hrs of incubation on M9 minimal media plates supplemented with (A) 2MA, and (B) 2EA. The aromatic amine substrates were converted to their respective catechols, which autooxidize to form the pink halos observed around the colonies.

4.4 Selection

Selection is a process whereby the survival of a cell in a genetic library is linked to its ability to display the desired phenotype. The dihydroxylation of aniline to catechol releases the amine group on the aromatic ring as ammonium, which can be used as a selection criterion for active mutants. The selection pressure in this case is the generation of ammonium for survival. The mutant library would be plated on a M63 minimal media agar with the target aromatic amine substrate as the sole source of nitrogen. Hence, only cells with AtdA enzyme active against the target substrate will be able to generate nitrogen for survival and grow into a colony, from which the plasmids can be rescued and sequenced. The effects of various factors, such as minimal amount of ammonium required for growth, IPTG and substrate concentrations, were first quantified in the liquid minimal media before the method was tested on agar plates.

4.4.1 Effect of IPTG

The cells have to be expressing AtdA when inoculated into the selection media in order to survive. However, AtdA expression, which is induced by IPTG, may be a metabolic burden for the cells, retarding or even preventing growth. Hence the IPTG concentration may need to be optimized.

E. coli JM109 cells harboring pAS93 plasmid were inoculated into M63 minimal media supplemented with 1 mM of ammonium sulfate. The IPTG was added to separate cultures to final concentrations of 0.1 mM and 1 mM respectively. The positive control had no IPTG added, while the negative control had neither IPTG nor ammonium sulfate added. The cell densities of the cultures were monitored over the course of 72 hrs (Figure

4.18). It was found that the cell growth was not affected by IPTG concentration as the cultures with 0.1 mM and 1 mM IPTG grew at the same rate and extent as the positive control.

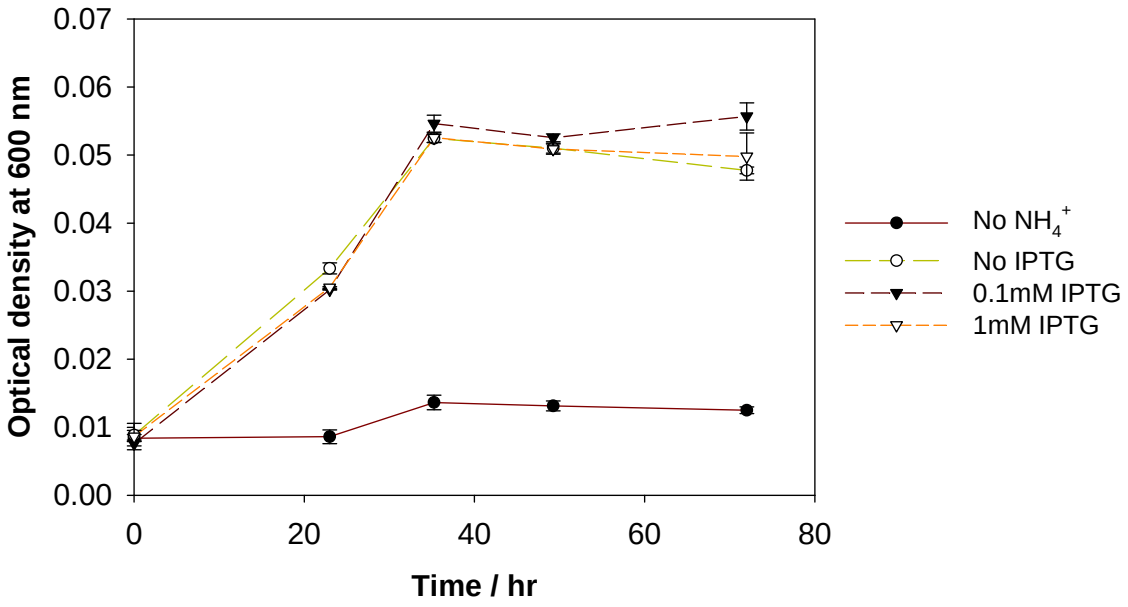


Figure 4.18 Effect of IPTG concentration on the growth of *E. coli* JM109 with pAS93 plasmid.

4.4.2 Minimal ammonium concentration

The minimum amount of ammonium required for growth determines the amount of substrate that needs to be added to the selection media. Overnight cultures of *E. coli* JM109 harboring pAS93 plasmid were inoculated into separate M63 minimal media supplemented with varying concentrations of ammonium sulfate and the cell density was noted after 49 hrs (Figure 4.19). The minimum ammonium concentration required for cell growth was 0.1 mM. As such, the concentration of the aromatic amine used in the selection media need to be higher than 0.1 mM to ensure sufficient nitrogen source for cell growth.

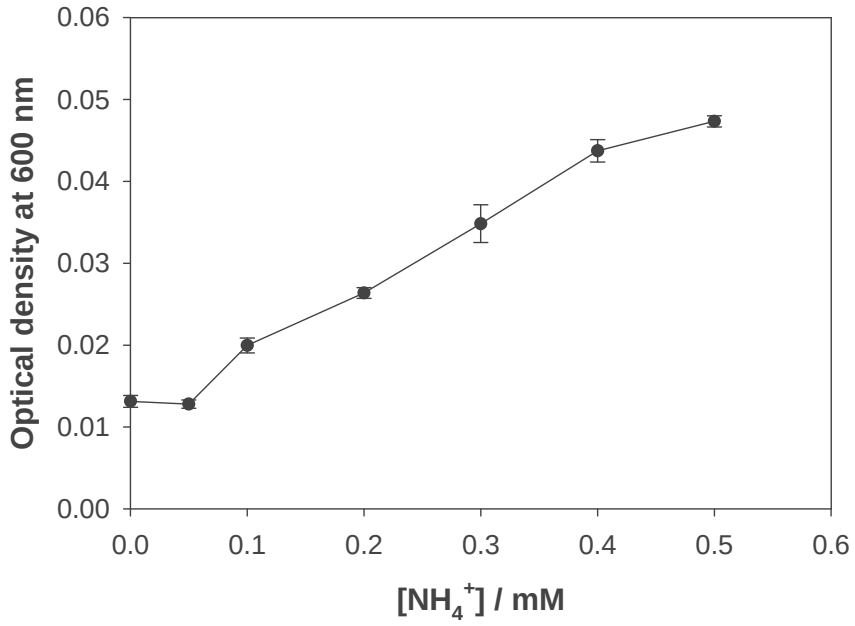


Figure 4.19 Effect of ammonium concentration on the growth of *E. coli* JM109 with pAS93 plasmid.

The efficacy of aromatic amines as nitrogen source was next investigated. Overnight cultures of *E. coli* JM109 harboring pAS93 plasmid were inoculated into separate M63 minimal media supplemented with 0.5 mM of aniline or 2EA as substrates and growth was monitored through the optical density of the cells over a period of 140 hrs (Figure 4.20). The same cells were inoculated into minimal media with 0.5 mM of ammonium sulfate as positive control, while no nitrogen source was provided in the negative control.

Using aniline as the nitrogen source, the cell culture reached the stationary phase at about the same time as the positive control. Its slightly lower cell density may be because as nitrogen source, aniline was not as readily available as free ammonium. The growth on

2EA was slower than on aniline, with the culture reaching saturation after 83 hr. The maximum cell density reached on 2EA was slightly lower than that reached on aniline.

These results have shown that induction of protein expression in cells by IPTG did not cause a significant metabolic burden that resulted in a slow down of cell growth. The minimal concentration of nitrogen required to support growth in a selection media (M63 minimal media) was 0.1 mM. Cells were able to survive and grow in the selection media using 0.5 mM of aniline or 2EA as nitrogen source, with the culture growing faster using the substrate that AtdA is more active. After the efficacy of using aromatic amines as selection markers for enzyme activity was established in the liquid media, the selection method was tested on the solid phase.

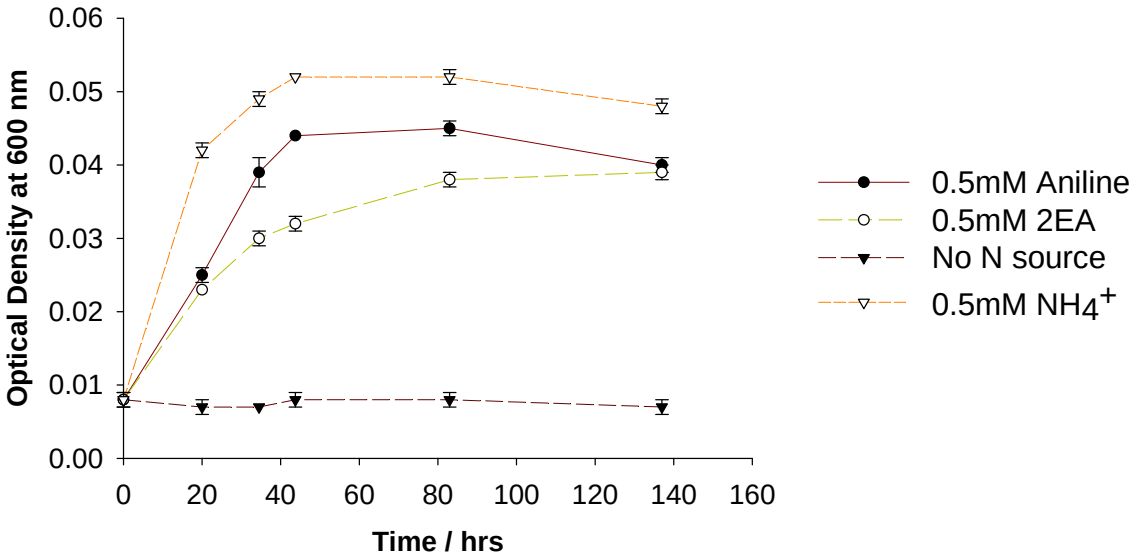


Figure 4.20 The growth curves of *E. coli* JM109 with pAS93 in M63 minimal media using aniline or 2EA as the sole source of nitrogen.

Cultures of *E. coli* JM109 expressing AtdA were washed twice with M63 minimal media lacking nitrogen and plated onto an M63 minimal media plates supplemented with 4 mM 2EA to ensure sufficient nitrogen source for the cells. The plates were then incubated at 30 ºC for a few days. After 2 days, the plates turned pink, indicating the conversion of 2EA to its catechol product, and small colonies grew on them. Randomly selected colonies from the plates were inoculated into LB media overnight. The overnight cultures were then inoculated into liquid M63 minimal medium supplemented with 4 mM 2EA. All the cultures tested were able to dihydroxylated 2EA, as evidenced by the formation of colored autooxidation products after 24 hours.

The advantage of the selection method is that it is much more efficient than the screening method for the identification of mutants with novel activity because a large pool of mutants can be plated onto a single agar plate, and only colonies with the ability to release ammonium from the target substrate, which is the desired phenotype, will grow. The limitation of this method is that it is not very effective at identifying mutants with improved activity. Although it can be argued that the more active mutants will be able to utilize the target substrate faster and grow into larger colonies, the colonies that are formed on minimal media plates are typically small, even when using non-limiting amounts of ammonium sulfate as nitrogen source, and it is difficult to measure differences in the size of the small colonies. Another concern is that novel activity may have been created in a mutant, but the activity of the mutant may not be high enough for the clone to generate enough ammonium for growth, and hence is unable to survive the selection pressure.

4.5 Summary

In summary, several screening and selection methods were investigated in this chapter. The evaluation of the screening and selection methods is summarized in Table 4.1. As the objective of the project was to both improve the existing activity and create novel activity of the aniline dioxygenase, the liquid phase Gibbs' reagent screening method was preferred as it fulfilled both goals, despite being a lower throughtput method than the solid phase method.

Table 4.1 Evaluation of screening and selection methods.

Method	Basis	Advantage	Disadvantage
Indophenol blue	▪ Colorimetric screening ▪ Detection of ammonia ▪ Forms intense blue color at 640 nm	▪ Very sensitive	▪ Reacts with aniline as well to give the same color ▪ Laborious as a lot of reagents need to added separately
MBTH reagent	▪ Colorimetric screening. ▪ Detection of aromatic amines. ▪ Forms magenta color at 564 nm.	▪ Sensitive. ▪ Colored compound shows strong absorbance.	▪ Reacts with catehcols as well, which interfere with absorbance profile
van Urk reagent	▪ Colorimetric screening ▪ Detection of aromatic amines. ▪ Forms yellow color at 435 nm.	▪ Reasonable absorbance with aniline	▪ Weak absorbance with 2EA. ▪ Reacts with catechol as well to interfere with absorbance profile.
Gibbs' reagent (liquid phase)	▪ Colorimetric screening ▪ Detection of catechols	▪ Sensitive. ▪ Quantitative.	▪ Reacts with aromatic amines as well but color formation rate can be controlled with pH. ▪ Not as high throughput as solid phase method
Gibbs' reagent (solid phase)	▪ Colorimetric screening ▪ Detection of catechols	▪ Sensitive ▪ High throughput ▪ No interference of aromatic amines	▪ Not quantitative.
Autooxidation	▪ Colorimetric screening ▪ Detection of catechols	▪ Sensitive ▪ High throughput ▪ No interference of aromatic amines.	▪ Not quantitative ▪ Slow as catechols require time to autooxidize
Selection	▪ Use of aromatic amine as sole nitrogen source	▪ Very high throughput	▪ Not quantitative ▪ Mutants with novel but low activity may not be identified.

4.6 References

Aharoni, A., A. D. Griffiths and D. S. Tawfik (2005). "High-throughput screens and selections of enzyme-encoding genes." Curr Opin Chem Biol **9** (2): 210-6.

Barriault, D., M. M. Plante and M. Sylvestre (2002). "Family shuffling of a targeted *bphA* region to engineer biphenyl dioxygenase." J Bacteriol **184** (14): 3794-800.

Clesceri, L. S., A. E. Greenberg and A. D. Eaton (1999). Standard Methods for the Examination of Water and Wastewater. Washington, D.C., American Public Health Association, the American Water Works Association and the Water Environment Federation.

Ehmann, A. (1977). "The van urk-Salkowski reagent--a sensitive and specific chromogenic reagent for silica gel thin-layer chromatographic detection and identification of indole derivatives." J Chromatogr **132** (2): 267-76.

Fujii, T., M. Takeo and Y. Maeda (1997). "Plasmid-encoded genes specifying aniline oxidation from *Acinetobacter* sp strain YAA." Microbiology **143**: 93-99.

Gasparic, J., D. Svobodova and M. Pospisilova (1977). "Investigation of the Colour Reaction of Phenols With the MBTH Reagent." Mikrochimica Acta **1** (3-4): 241-250.

Gibbs, H. D. (1927). "Phenol Tests. III. The Indophenol Test." Journal of Biological Chemistry **72**: 649-664.

Horn, D. B. and C. R. Squire (1966). "The estimation of ammonia using the indophenol blue reaction." Clin Chim Acta **14** (2): 185-94.

Joern, J. M., T. Sakamoto, A. A. Arisawa and F. H. Arnold (2001). "A versatile high throughput screen for dioxygenase activity using solid-phase digital imaging." Journal of Biomolecular Screening **6** (4): 219-223.

Kunz, D. A. and P. J. Chapman (1981). "Catabolism of pseudocumene and 3-ethyltoluene by *Pseudomonas putida* (arvilla) mt-2: evidence for new functions of the TOL (pWWO) plasmid." J Bacteriol **146** (1): 179-91.

Meyer, A., A. Schmid, M. Held, A. H. Westphal, M. Rothlisberger, H. P. Kohler, W. J. van Berkel and B. Witholt (2002). "Changing the substrate reactivity of 2-hydroxybiphenyl 3-monooxygenase from *Pseudomonas azelaica* HBP1 by directed evolution." J Biol Chem **277** (7): 5575-82.

Miller, J. H. (1972). Experiments in molecular genetics. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Pospisilova, M., M. Polasek and D. Svobodova (1998). "Spectrophotometric study of reactions of substituted phenols with MBTH in alkaline medium: The effect of phenol structure on the formation of analytically useful coloured products." Mikrochimica Acta **129** (3-4): 201-208.

Pospisilova, M., D. Svobodova, J. Gasparic and M. Machacek (1990). "Investigation Of The Color-Reaction Of Phenols With MBTH. 2. Properties Of The Isolated Products Of The Reaction With Phenol, 2,6-Dimethylphenol And 4-Methylphenol." Mikrochimica Acta **3** (1-3): 117-128.

Quintana, M. G., C. Didion and H. Dalton (1997). "Colorimetric method for a rapid detection of oxygenated aromatic biotransformation products." Biotechnology Techniques **11** (8): 585-587.

Schmidt-Dannert, C. and F. H. Arnold (1999). "Directed evolution of industrial enzymes." Trends Biotechnol **17** (4): 135-6.

You, L. and F. H. Arnold (1996). "Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide." Protein Eng **9** (1): 77-83.

Chapter 5
Substrate specificity of AtdA

5.1 Introduction

In this chapter, the substrate specificity and the critical subunits of aniline dioxygenase (AtdA) were examined in order to facilitate in the engineering of AtdA.

Acinetobacter sp strain YAA, from which the aniline dioxygenase (AtdA) enzyme was isolated, was able to utilize aniline and 2-methylaniline as sole carbon source but not 3-methylaniline, 4-methylaniline, chloroanilines, aminophenols or diaminobenzenes (Fujii, Takeo et al., 1997). However, the substrate specificity of AtdA for other aromatic amines such as *ortho*-substituted anilines with larger sidechains, and xylidines were not examined. Furthermore, the substrate specificity of AtdA when expressed in *E. coli* may be different from that as the native host, *Acinetobacter* sp strain YAA. Thus the substrate range of AtdA, expressed in *E. coli*, has to be characterized to determine the target substrate to be used in AtdA engineering. To achieve this goal, the AtdA enzyme will be tested for activity against a variety of aniline homologue substrates substituted at the *ortho* position of the aromatic, with progressively larger alkyl sidechains, which work as steric probes to elucidate the substrate binding pocket size and topography of AtdA.

The AtdA enzyme has five subunits, but not all subunits are critical to its activity or substrate specificity. To narrow the potential library to be screened, the subunits critical for dioxygenase activity will be determined. This is done using a gene deletion study whereby each gene of interest is individually deleted from the *atdA* operon and the

resultant construct is tested for activity. This would facilitate the characterization of the enzyme's substrate specificity and activity determinants, as well as enable directed evolution to be focused on the critical subunits instead of the entire operon.

This chapter analyzes the substrate specificity of the wild type AtdA, and identifies the critical subunits for its activity. Both of these findings are useful for the characterization of the enzyme and will serve as reference points for the further characterization and engineering of AtdA described in the subsequent chapters.

5.2 Materials and methods

5.2.1 Materials

Aniline, 2,4-dimethylaniline (24DMA), 3,4-dimethylaniline (34DMA), 2-methylaniline (2MA), 2-ethylaniline (2EA), 2-isopropylaniline (2IPA), 2-secbutylaniline (2SBA), 2-tertbutylaniline (2TBA), catechol, isopropyl-β-D-thiogalactopyranoside (IPTG), dimethylformamide (DMF), ampicillin, chloramphenicol and all other chemicals were purchased from Sigma (St Louis, MO) unless otherwise stated. 3-isopropylcatechol (3IPC) was purchased from Chem Service (West Chester, PA). Gibb's reagent was purchased from MP Biomedicals (Solon, OH). Quikchange XL Site Directed Mutagenesis kit and *Pfu Turbo* DNA polymerase were purchased from Stratagene (La Jolla, CA). *Taq* DNA polymerase was purchased from Promega (Madison, WI). Primers were purchased from Integrated DNA Technologies (Coralville, IA) and 1st Base (Singapore). PCR-grade deoxynucleotide triphosphates (dNTPs) were obtained from Roche Applied Sciences (Indianapolis, IN). All DNA modifying enzymes were purchased from New England Biolabs (Beverly, MA). All DNA gel purifications were

carried out using the QIAEX II gel purification kit from Qiagen (Valencia, CA). All plasmid isolations were carried out using the QIAprep Miniprep kit from Qiagen.

E. coli JM109 and BL21 (DE3) were purchased from Novagen (Madison, WI), while chemical competent *E. coli* DH5α was purchased from the Cell Media Facility at the University of Illinois (Urbana, IL). The plasmid pTrc99A was from Amersham Pharmacia (Piscataway, NJ). The pACYCDuet-1 and pETDuet-1 plasmids were from Novagen. The plasmids pAS91 and pAS93, both containing the aniline dioxygenase gene cluster, were kindly provided by Dr. Masahiro Takeo from the Dept of Applied Chemistry, Himeiji Institute of Technology, Hyogo, Japan.

5.2.2 Substrate specificity assay

E. coli JM109 expressing AtdA was inoculated into 5 ml LB with ampicillin (100 mg/l) and grown overnight in a 37 °C shaker at 250 rpm. Subsequently, 0.3 ml of the overnight culture was inoculated into 3 ml of M9 minimal media (Sambrook, Fritsch et al., 1989) with 100 mg/l ampicillin and 1 mM IPTG, and incubated in a 30 °C shaker for 4 hr at 250 rpm to induce protein expression. Aniline or its analogue substrates were then added to each tube to a final concentration of 1 mM, and the culture was incubated for 24 hrs in a 30 °C shaker at 250 rpm. The cultures were then observed for formation of colored autooxidation products.

5.2.3 Construction of plasmids for gene deletion assay

The pTA2-3 plasmid was used as the template for all polymerase chain (PCR) reactions to amplify the *atdA* genes. The *atdA1* gene was amplified using the primers A1_EcoRI_F (5'-GCCGTCCTTGAATTCGATGAGTGAGAAATTAGATTT-3') and A1_SalI_R (5'- GCATTGTTACGTCGACTCACAGTAAGTTGAAGTATT-3'). The *atdA2* gene was amplified using the primers A2_FseI_F (5'-ATTATTAAAGGCCGGCCACATGTCTAAACGCTTTGC-3') and A2_AvrII_R (5'-TCATGGTCTGTTCCTCCTAGGTTATACATCACCCACCAAGTTCTG-3'). The *atdA3* gene was amplified using the primers A3_EcoRI_F (5'-CTCAGGTGGGAATTCGATGAAAACCATAAATCAACT-3') and A3_SalI_R (5'-CTCAAGATGGTCGACTTAAACCTCAGACTGATTCT-3'). The *atdA4A5* gene was amplified using the primers A4_FseI_F (5'-GAACATTAAAGGCCGGCCACATGAATAATAATAAAGATCT-3') and A5_AvrII_R (5'-GTTAGTGATGCCTAGGTTATTGGTCGAAGGATATTT-3'). The restriction site encoded in each primer sequence is underlined.

The PCR reaction mix for each gene consists of 150 ng of the pTA2-3 template, 50 pmol of the forward and reverse primers each, 10 μl of 10x *Taq* polymerase buffer, 6 μl of 25 mM $MgCl_2$, 10 μl of 10x dNTP (mixture of dATP, dTTP, dGTP, and dCTP, each at a concentration of 100 mM), 1.25 U of *Taq* DNA polymerase and *Pfu Turbo* DNA polymerase respectively, and water to make up 100 μl. The PCR program consists of 94 °C for 3 min; 25 cycles of 94 °C for 45 s, 50 °C for 45 s, 72 °C for 2 min; and a final extension time of 7 min at 72 °C. The PCR products were then gel purified. The *atdA1* and *atdA3* PCR products were digested with EcoRI and SalI, while the *atdA2* and *atdA4A5* PCR products were digested with FseI and AvrII.

To construct the plasmid pACYC A1, the pACYCDuet-1 plasmid was first digested with EcoRI and SalI, gel purified, and ligated with the digested *atdA1* PCR product. To construct pACYC A2, the pACYCDuet-1 plasmid was digested with FseI and AvrII, gel purified, and ligated with the digested *atdA2* PCR product. To construct pACYC A1A2, the pACYC A2 plasmid was digested with EcoRI and SalI, gel purified, and ligated with the digested *atdA1* PCR product. To construct the plasmid pET A4A5, the pETDuet-1 plasmid was first digested with FseI and AvrII, gel purified, and ligated with the digested *atdA4A5* PCR product. To construct the plasmid pET A3A4A5, the pETA4A5 plasmid was first digested with EcoI and SalI, gel purified, and ligated with the digested *atdA3* PCR product. All ligations were carried out overnight at 16 ºC using the T4 DNA ligase. The salts from the ligation reactions were then removed by precipitating the ligated DNA with n-butanol (Thomas, 1994). The ligation reactions were then transformed into *E. coli* BL21 (DE3) by electroporation.

5.2.4 Gene deletion studies

E. coli BL21 (DE3) colonies harboring the various plasmid constructs were picked into separate culture tubes with 3 ml of LB media containing 100 mg/l ampicillin and 35 mg/l chloramphenicol and were grown overnight in a 37 ºC shaker at 250 rpm. 50 µl of each of the overnight cultures were inoculated into 5 ml of LB media with the same antibiotics composition and grown in a 37 ºC shaker at 250 rpm. At an optical density (OD_{600}) of ~0.5 to 0.6, IPTG was added to each culture to a final concentration of 1 mM and the cultures were then incubated for 3 hrs in a 30 ºC shaker at 250 rpm.

The cultures were harvested by centrifuging at 6000 × g for 10 min. The supernatant was discarded and the cell pellets were gently resuspended with 5 ml of M9 minimal media with 100 mg/l ampicillin, 35 mg/l chloramphenicol and 1 mM IPTG. 2MA was then added to each culture to a final concentration of 2 mM and the cultures were incubated in a 30 ºC shaker at 250 rpm for 24 hrs. The cultures were constantly monitored for the formation of autooxidation products.

5.2.5 Whole cell activity assay

An overnight LB culture of JM109 with WT or mutant plasmid was inoculated into 150 ml LB to an OD_{600} of 0.02 and incubated in a 37 °C shaker at 250 rpm. When the OD_{600} reached 0.50 to 0.55, IPTG was added to a final concentration of 1 mM. The culture was then incubated in a 30 °C shaker at 250 rpm for 3 hr. The induced culture was then centrifuged at 4000 × g for 10 min. The supernatant was discarded and the cell pellet was resuspended with 150 ml of modified M9 buffer (M9 minimal media with 0.1% glucose). The resuspended cells were centrifuged using the same conditions again. The supernatant was discarded and the cell pellet was resuspended with modified M9 buffer to a final OD of about 10. Then 5 ml of the resuspended cells were aliquoted into a 50 ml centrifuge tube and 5 µl of 1 M substrate, dissolved in *N*,*N*-dimethylformamide, was added to a final concentration of 1 mM. The cells were then incubated at 30 °C with 250 rpm shaking. 0.5 ml samples were drawn at various time points. The samples were centrifuged at 16,000 × g in a benchtop centrifuge for 3 min and the supernatant was stored in -20 °C until ready for analysis.

The substrate and products were separated and quantified using high–pressure liquid chromatography (HPLC) with a 250 × 4.60 mm Synergi 4μ Polar-RP 80A column from Phenomenex (Torrance, CA). All HPLC methods used were isocratic with a flowrate of 1 ml/min. Aniline was analyzed using 90% potassium phosphate (pH 7.0) and 10% acetonitrile as mobile phase. 2IPA was analyzed using 60% potassium phosphate (pH 7.0) and 40% acetonitrile as mobile phase. 24DMA was analyzed using 70% potassium phosphate (pH 7.0) and 30% acetonitrile as mobile phase.

For each culture, 1 ml of the resuspended cells was centrifuged at 6,000 × g in a benchtop centrifuge for 3 min and the supernatant was discarded. The cell pellet was resuspended in 50 mM Tris-HCl (pH 7.5) and disrupted using a single pass through the Constant Systems Cell Disruptor (Warwick, UK) at 20.3 kpsi. The disrupted cells were centrifuged at 16,000 × g in a benchtop centrifuge for 5 min and the supernatant was assayed for protein concentration using the BCA Protein Assay kit from Pierce (Rockford, IL). The whole cell activity was calculated by normalizing the initial rate of substrate conversion or product formation to the protein concentration.

5.3 Substrate specificity of AtdA

As the substrate range of the AtdA enzyme had not been extensively characterized, it was necessary to determine this property before probing the molecular determinants of the enzyme's substrate specificity. To determine the substrate specificity of the WT AtdA enzyme, *E. coli* JM109 expressing the WT enzyme was incubated individually with a series of *ortho*-substituted anilines with progressively larger alkyl side chains, namely, aniline, 2-methylaniline (2MA), 2-ethylaniline (2EA), 2-isopropylaniline (2IPA), 2-

secbutylaniline (2SBA), and 2-tertbutylaniline (2TBA), as well as two xylidine substrates, 2,4-dimethylaniline (24DMA) and 3,4-dimethylaniline (34DMA), as shown in Figure 5.1. Dihydroxylation of a particular substrate by the enzyme produces its corresponding catechol, which undergoes autooxidation to form colored compounds, indicating activity for that substrate (Figure 5.2A) (Kunz and Chapman, 1981; Fujii, Takeo et al., 1997; Barriault, Plante et al., 2002; Meyer, Schmid et al., 2002).

Of the *ortho*-substituted substrates, the WT AtdA3 showed activity for aniline, 2MA, and 2EA (Figure 5.2B). However, the enzyme was inactive against substrates with an ortho side chain larger than an ethyl group (2IPA, 2SBA, and 2TBA). Since 2EA and 2IPA differ only by a single methyl group on the ortho side chain, the substrate specificity of the enzyme may be controlled by steric hindrance to the ortho side chain along the substrate channel or in the substrate binding pocket. For the xylidine substrates, 24DMA was accepted as a substrate but a change of the position of one of the methyl groups from *ortho* (24DMA) to *meta* (34DMA) renders the substrate unacceptable to the enzyme. This may indicate that the steric limitation of the enzyme's binding pocket lies in the area between the *ortho* and *para* position of the aromatic substrate.

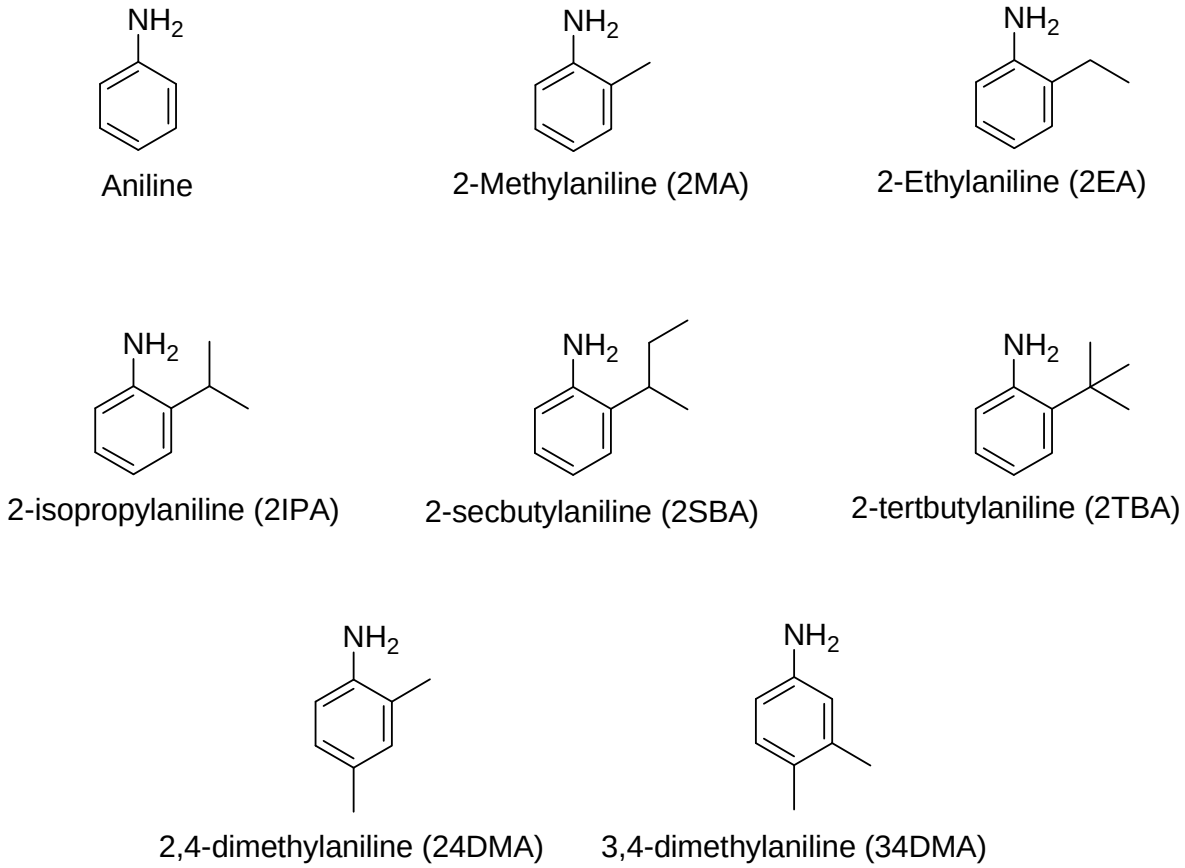
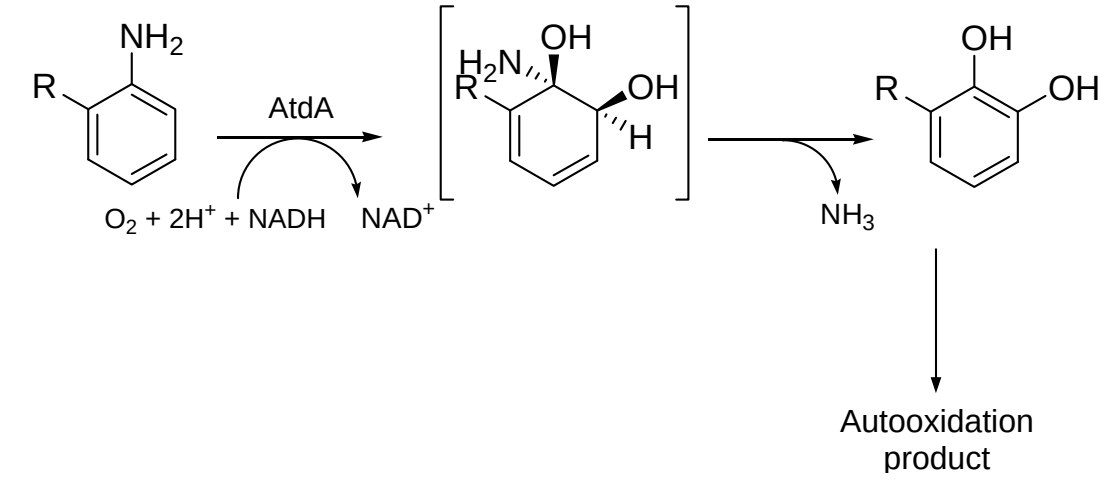


Figure 5.1 Aniline, its *ortho*-substituted homologues (first two rows), as well as xylidine substrates (third row) used to determine the substrate specificity of AtdA.

(A)



(B)

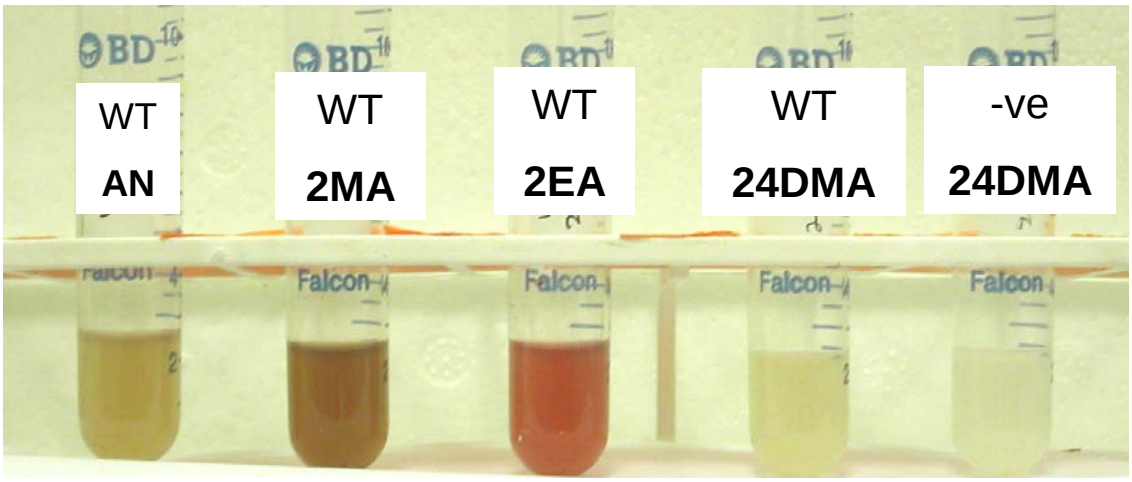


Figure 5.2 (A) Production of autooxidation products from aromatic amines by AtdA; (B) *E. coli* JM109 expressing WT AtdA3 after overnight exposure to aniline and other substrates. The negative control (-ve) is *E. coli* JM109 harboring the empty vector, pTrc99A.

5.4 Effect of methyl sidechain position on enzyme activity

Following the finding that AtdA acted on 24DMA but not 34DMA, further investigation was carried out to elucidate the effects of the position of the alkyl side chain on the aromatic ring of aniline on the enzyme activity. To achieve this aim, whole cell activity assay was carried out using JM109 cells expressing AtdA, with 1 mM of 2-methylaniline (2MA), 3-methylaniline (3MA), and 4-methylaniline (4MA) as substrates respectively. The assay was carried out as described in Materials and Methods (Section 5.2.5). As presented in Figure 5.3, the activity on the substrate decreases as the distance between the amine and methyl groups increases. Amongst the methylaniline isomers, 2MA was the best substrate, and based on the initial rate measurement, it was converted at 19.5 nmol/min.mg protein, which was 5.9 and 23.9 fold faster than 3MA (3.3 nmol/min.mg protein) and 4MA (0.8 nmol/min.mg protein) respectively.

The dihydroxylation of 3MA can have two possible products, 3-methylcatechol (3MC) or 4-methylcatechol (4MC), as shown in Figure 5.4. Interestingly, the only product from the dihydroxylation of 3MA by the WT enzyme was 4MC, as determined from the retention time of the products in HPLC analyses (Figure 5.5). The elution time of the 3MC authentic standard was 10.6 min while that of 4MC was 9.4 min. 2MA was not observed in Figure 5.6A as it was completely converted to 3MC after 120 min.

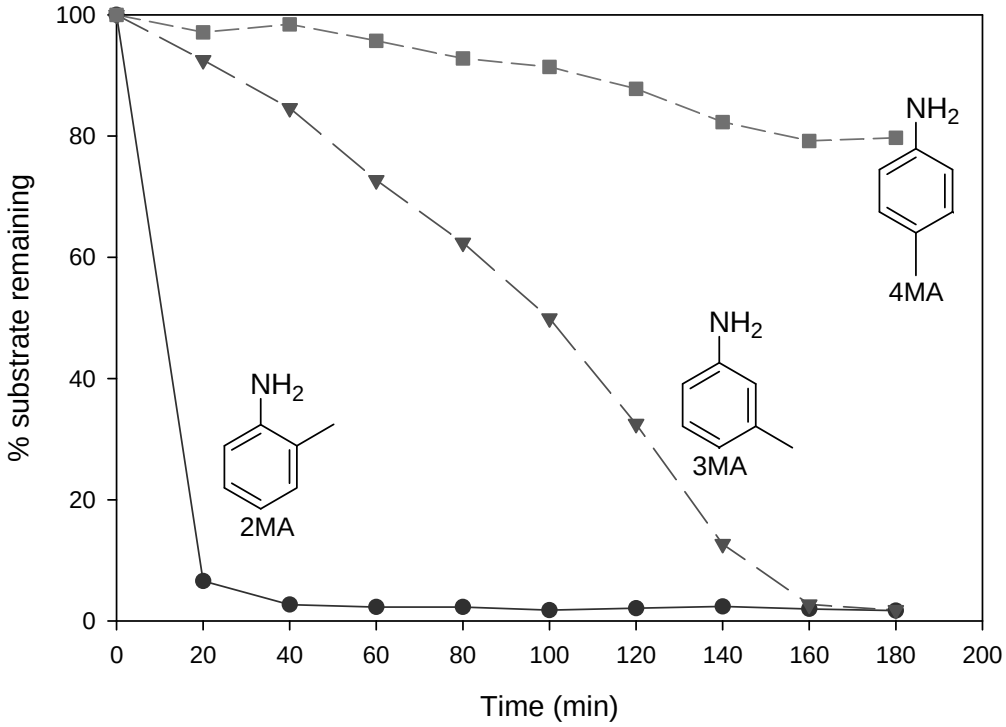


Figure 5.3 Percentage of methylaniline substrate remaining with time in whole cell assay. Starting concentrations of all three substrates were 1 mM.

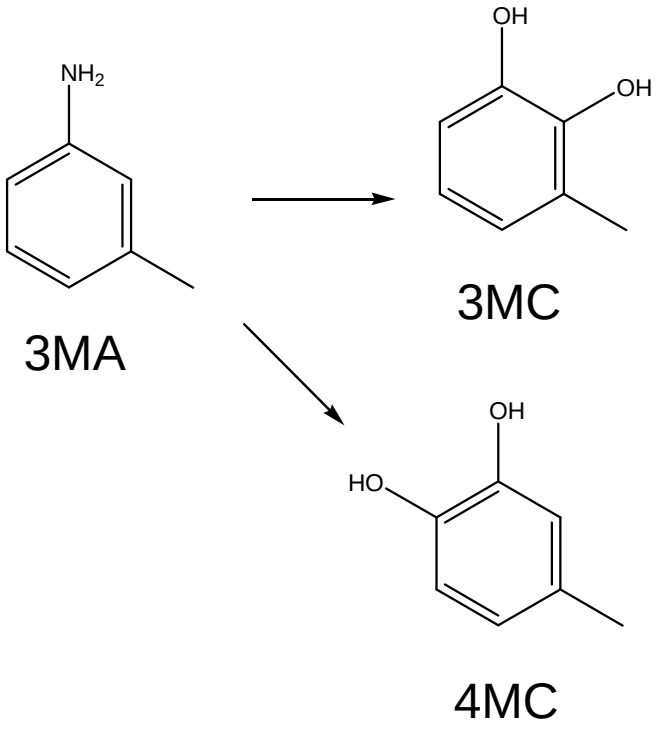


Figure 5.4 Possible products from the dihydroxylation of 3MA by AtdA.

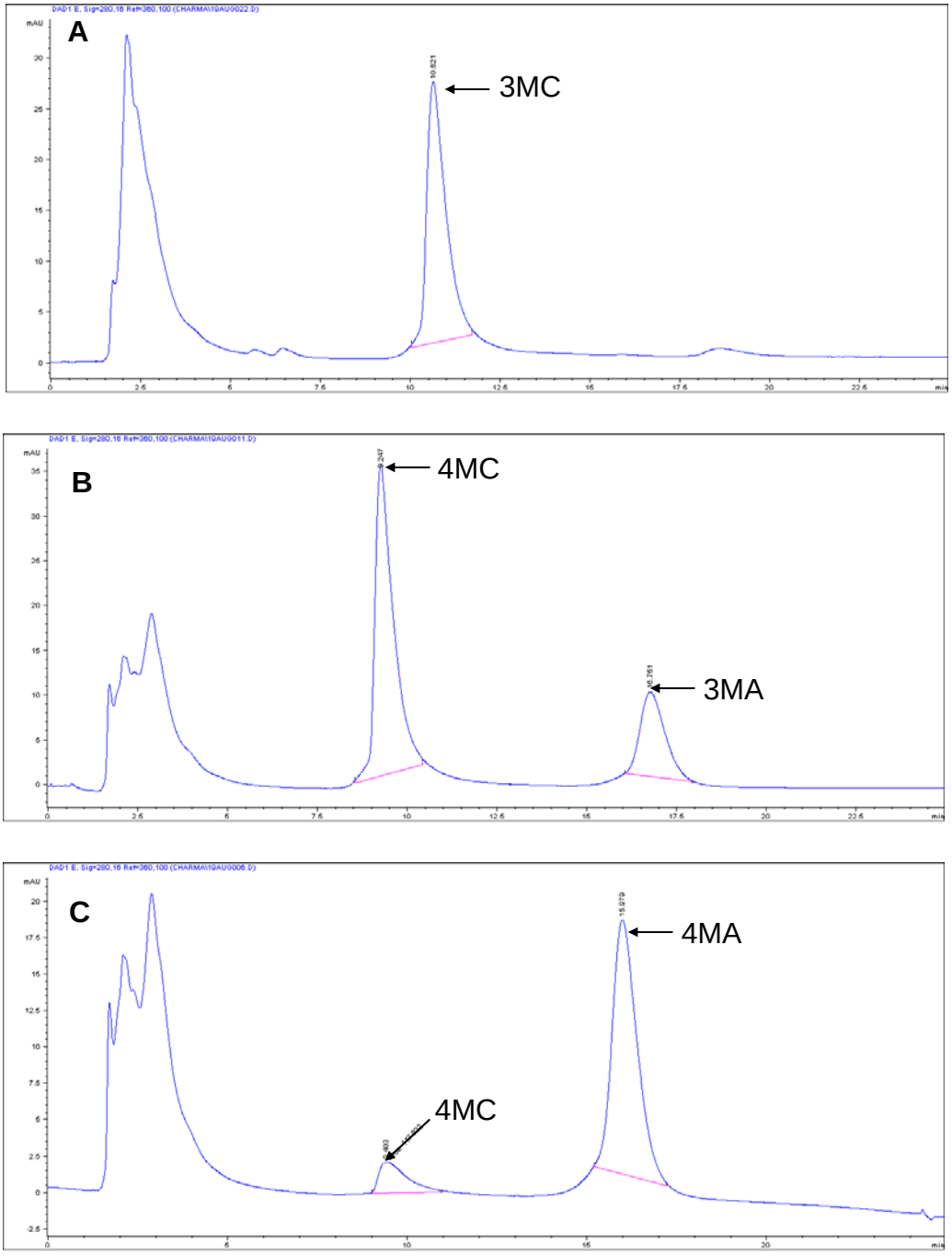


Figure 5.5 HPLC chromatograms of *E. coli* JM109 cells expressing AtdA after 120 min of incubation with (A) 2MA, (B) 3MA, and (C) 4MA.

Since AtdA accepted the *ortho*-substituted aromatic amines better than substrates substituted at other positions on the aromatic ring, it may be easier to create novel activity for *ortho*-substituted substrates not accepted by the WT. Thus, 2IPA, and 2SBA were chosen as target substrates to probe for residues controlling the substrate specificity of the enzyme, while aniline and 24DMA were chosen as target substrates to probe for residues determining the activity of AtdA3. Before probing for molecular determinants of AtdA's substrate specificity, the target subunit has to be identified.

5.5 Gene deletion studies

A gene deletion assay was carried out to identify the subunits critical for AtdA activity. The first three subunits of *atdA* gene cluster, *atdA1*, *A2* and *A3*, were targeted. The *atdA4* gene is homologous to the β-subunit of a terminal Rieske dioxygenase. It was not targeted because the α–subunit of the Rieske dioxygenase is generally regarded as the main contributor of substrate specificity (Beil, Mason et al., 1998; Parales, Parales et al., 1998; Wackett, 2002). The *atdA5* gene encodes for a reductase, which is involved in cofactor regeneration in the dihydroxylation reaction and not in the direct binding of the substrate. Hence it was not targeted in the gene deletion assay as well.

The Duet plasmid system was employed in the gene deletion assay. The *atdA1 and A2* genes were cloned individually and together in the pACYCDuet-1 vector to give the pACYC A1, pACYC A2, and pACYC A1A2 vectors respectively, while *atdA4A5* was cloned into the pETDuet-1 vector with and without *atdA3* to give pET A4A5 and pET A3A4A5 respectively (Figure 5.6). The plasmid sequences are found in Appendix A. The plasmids were transformed into *E. coli* BL21 (DE3) according to Table 5.1 for the assay.

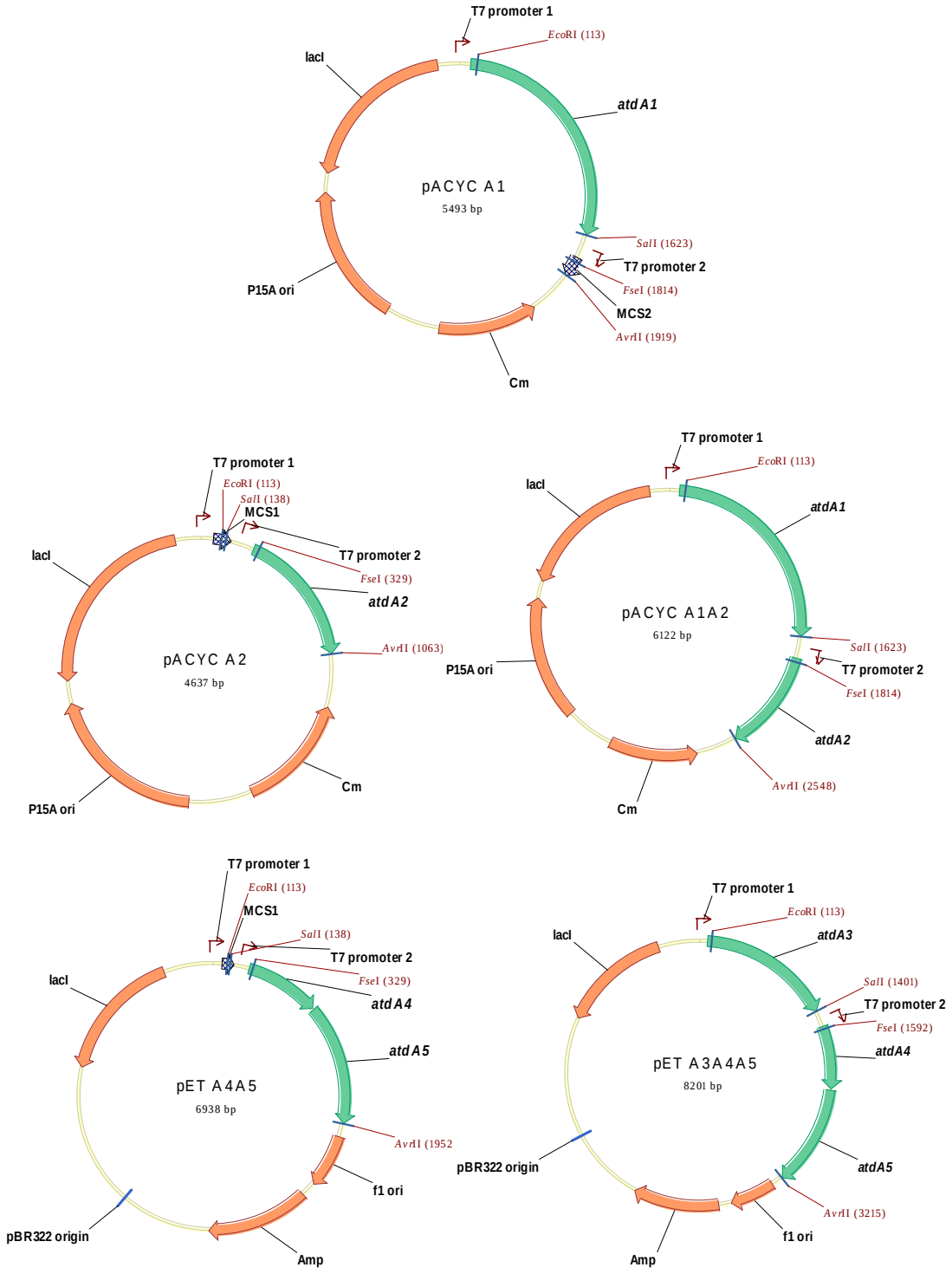


Figure 5.6 Vector maps of plasmids used in the gene deletion assay. MCS represents multiple cloning sites.

Table 5.1 Plasmids used for each deletion construct.

Gene deleted	Plasmids used
atdA1	pACYC A2 & pET A3A4A5
atdA2	pACYC A1 & pET A3A4A5
atdA3	pACYC A1A2 & pET A4A5
Control (no deletion)	pACYC A1A2 & pET A3A4A5

E. coli BL21 (DE3) transformed with the various plasmid constructs shown in Table 5.1 were grown in LB, induced with IPTG, and then washed after overnight with M9 minimal media. 2MA was then added to each culture to a concentration of 2 mM to test for dioxygenase activity. The cultures were incubated in a 30 ºC shaking incubator for 24 hr. 2MA was chosen as the substrate because the product, 3MC, autooxidized the quickest and gave the strongest color amongst the various substrates tested in the earlier substrate specificity assay (Section 5.3).

When the *atdA1* and *A3* genes were removed, no activity for 2MA was detected (Figure 5.7). On the other hand, the construct with *atdA2* removed displayed 2MA activity. These results suggest that AtdA1 and A3 are critical for the activity of the enzyme and hence the substrate specificity and activity may be controlled by one or both of these subunits.

A literature search revealed that the substrate specificity of various dioxygenases, such as the naphthalene, biphenyl and 2,4-dinitrotoluene dioxygenases, are determined by their terminal α–subunits (Tan and Cheong, 1994; Parales, Parales et al., 1998; Parales, Emig et al., 1998). From these findings, various directed evolution and saturation

mutagenesis studies on the terminal α–subunits have successfully altered the substrate specificity of these dioxygenases (Kumamaru, Suenaga et al., 1998; Parales, Lee et al., 2000; Parales, Resnick et al., 2000; Sakamoto, Joern et al., 2001; Barriault, Plante et al., 2002; Barriault and Sylvestre, 2004; Keenan, Leungsakul et al., 2004; Keenan, Leungsakul et al., 2005; Leungsakul, Keenan et al., 2005). These results indicate the likelihood that AtdA3 controls the substrate specificity of aniline dioxygenase. Hence the AtdA3 subunit was targeted for saturation mutagenesis studies to probe for the molecular determinants of the enzyme's substrate specificity and activity. It should be noted that this assay was intended to aid in determining which AtdA subunit would be studied first, and the possibility that the other subunits may play a part in the substrate specificity and activity should not be ruled out.

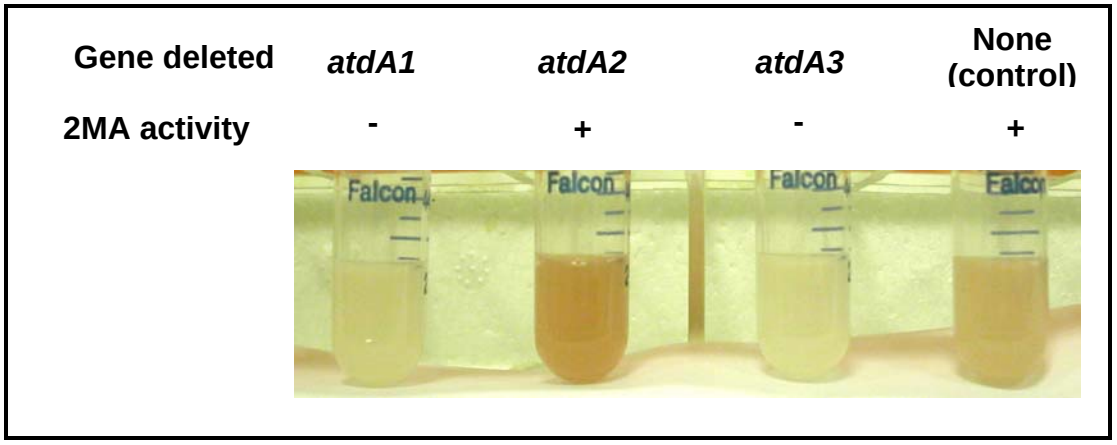


Figure 5.7 *E. coli* BL21 (DE3) with different AtdA deletion constructs after incubation with 2MA for 24 hr.

5.6 Summary

Characterizing the substrate specificity of AtdA and determining the subunits critical for dioxygenase activity are prerequisite to engineering the enzyme to enhance its substrate range and activity. In this chapter, the activity of AtdA on a range of *ortho*-substituted aniline substrates was tested. The effect of methylaniline's alkyl sidechain ring on enzyme activity was also tested. Lastly, a gene deletion assay was used to ascertain which subunits are critical for dioxygenase activity in AtdA.

The AtdA enzyme was inactive against substrates with an *ortho*-sidechain larger than an ethyl group (2IPA, 2SBA, and 2TBA), suggesting that the enzyme substrate specificity may be controlled by steric hindrance to the *ortho*-sidechain along the substrate channel or in the substrate binding pocket. The distance between the amine and methyl groups in methylaniline has an adverse effect on AtdA activity. For methylanilines, 2MA was the best substrate for AtdA, being converted 5.9 and 23.9 fold faster than 3MA and 4MA respectively. Interestingly, the only product from the dihydroxylation of 3MA by the AtdA enzyme was 4MC, while another possible product, 3MC, was not produced. This shows the high regiospecificity of AtdA, and hence the enzyme may be a useful in the production of aromatic diols. The gene deletion assay results showed that AtdA1 and A3 are critical for the activity of the enzyme and hence the substrate specificity and activity may be controlled by one or both of these subunits. AtdA3 was chosen as the subject of investigation in the following chapters as equivalent subunits of AtdA3 in other dioxygenases have shown to play a part in substrate specificity and activity of the enzymes.

From the findings of this chapter, the AtdA3 subunit was targeted for saturation mutagenesis studies to probe for the molecular determinants of the enzyme's substrate specificity and activity. *Ortho*-substituted anilines would be used as target substrates as AtdA was most active against these isomers.

5.7 References

Barriault, D., M. M. Plante and M. Sylvestre (2002). "Family shuffling of a targeted *bphA* region to engineer biphenyl dioxygenase." J Bacteriol **184** (14): 3794-800.

Barriault, D. and M. Sylvestre (2004). "Evolution of the biphenyl dioxygenase BphA from *Burkholderia xenovorans* LB400 by random mutagenesis of multiple sites in region III." J Biol Chem **279** (46): 47480-8.

Beil, S., J. R. Mason, K. N. Timmis and D. H. Pieper (1998). "Identification of chlorobenzene dioxygenase sequence elements involved in dechlorination of 1,2,4,5-tetrachlorobenzene." J Bacteriol **180** (21): 5520-8.

Fujii, T., M. Takeo and Y. Maeda (1997). "Plasmid-encoded genes specifying aniline oxidation from *Acinetobacter* sp strain YAA." Microbiology **143**: 93-99.

Keenan, B. G., T. Leungsakul, B. F. Smets, M. A. Mori, D. E. Henderson and T. K. Wood (2005). "Protein engineering of the archetypal nitroarene dioxygenase of *Ralstonia* sp. strain U2 for activity on aminonitrotoluenes and dinitrotoluenes through alpha-subunit residues leucine 225, phenylalanine 350, and glycine 407." J Bacteriol **187** (10): 3302-10.

Keenan, B. G., T. Leungsakul, B. F. Smets and T. K. Wood (2004). "Saturation mutagenesis of *Burkholderia cepacia* R34 2,4-dinitrotoluene dioxygenase at DntAc valine 350 for synthesizing nitrohydroquinone, methylhydroquinone, and methoxyhydroquinone." Appl Environ Microbiol **70** (6): 3222-31.

Kumamaru, T., H. Suenaga, M. Mitsuoka, T. Watanabe and K. Furukawa (1998). "Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase." Nat Biotechnol **16** (7): 663-6.

Kunz, D. A. and P. J. Chapman (1981). "Catabolism of pseudocumene and 3-ethyltoluene by *Pseudomonas putida* (arvilla) mt-2: evidence for new functions of the TOL (pWWO) plasmid." J Bacteriol **146** (1): 179-91.

Leungsakul, T., B. G. Keenan, H. Yin, B. F. Smets and T. K. Wood (2005). "Saturation mutagenesis of 2,4-DNT dioxygenase of *Burkholderia* sp. strain DNT for enhanced dinitrotoluene degradation." Biotechnol Bioeng **92** (4): 416-26.

Meyer, A., A. Schmid, M. Held, A. H. Westphal, M. Rothlisberger, H. P. Kohler, W. J. van Berkel and B. Witholt (2002). "Changing the substrate reactivity of 2-hydroxybiphenyl 3-monooxygenase from *Pseudomonas azelaica* HBP1 by directed evolution." J Biol Chem **277** (7): 5575-82.

Parales, J. V., R. E. Parales, S. M. Resnick and D. T. Gibson (1998). "Enzyme specificity of 2-nitrotoluene 2,3-dioxygenase from *Pseudomonas* sp. strain JS42 is determined by

the C-terminal region of the alpha subunit of the oxygenase component." J Bacteriol **180** (5): 1194-9.

Parales, R. E., M. D. Emig, N. A. Lynch and D. T. Gibson (1998). "Substrate specificities of hybrid naphthalene and 2,4-dinitrotoluene dioxygenase enzyme systems." J Bacteriol **180** (9): 2337-44.

Parales, R. E., K. Lee, S. M. Resnick, H. Jiang, D. J. Lessner and D. T. Gibson (2000). "Substrate specificity of naphthalene dioxygenase: effect of specific amino acids at the active site of the enzyme." J Bacteriol **182** (6): 1641-9.

Parales, R. E., S. M. Resnick, C. L. Yu, D. R. Boyd, N. D. Sharma and D. T. Gibson (2000). "Regioselectivity and enantioselectivity of naphthalene dioxygenase during arene *cis*-dihydroxylation: control by phenylalanine 352 in the alpha subunit." J Bacteriol **182** (19): 5495-504.

Sakamoto, T., J. M. Joern, A. Arisawa and F. H. Arnold (2001). "Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate." Appl Environ Microbiol **67** (9): 3882-7.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor, N. Y, Cold Spring Harbor Laboratory.

Tan, H. M. and C. M. Cheong (1994). "Substitution of the ISP alpha subunit of biphenyl dioxygenase from *Pseudomonas* results in a modification of the enzyme activity." Biochem Biophys Res Commun **204** (2): 912-7.

Thomas, M. R. (1994). "Simple, effective cleanup of DNA ligation reactions prior to electro-transformation of *E. coli*." Biotechniques **16** (6): 988-90.

Wackett, L. P. (2002). "Mechanism and applications of Rieske non-heme iron dioxygenases." Enzyme And Microbial Technology **31** (5): 577-587.

Chapter 6
Probing the molecular determinants of AtdA substrate specificity

6.1 Introduction

Aniline and its derivatives are widely used as intermediates in the pharmaceutical and azo-dye manufacturing industries (Radomski, 1979; Grayson, Eckroth et al., 1984), and may be released to the environment through effluent streams from these industries (Rai, Bhattacharyya et al., 2005). These compounds are highly toxic and there have been numerous reports on their carcinogenic effects (Weisburger, Russfield et al., 1978; Nohmi, Miyata et al., 1983; Shardonofsky and Krishnan, 1997; Przybojewska, 1999; Markowitz and Levin, 2004; Bomhard and Herbold, 2005). Biodegradation is the main route of removing aromatic amine pollutants from the natural environment (Lyons, Katz et al., 1984), with the hydroxylation of the aromatic ring often constituting the first step of biodegradation (Bugg and Winfield, 1998). Thus, an enzyme with ability to hydroxylate a wide range of aniline homologues would be a practical and valuable biocatalyst for the remediation of harmful aromatic amine contaminants.

The aniline dioxygenase (AtdA) is a multi-component enzyme isolated from *Acinetobacter* sp. strain YAA, which carries out the simultaneous deamination and oxygenation of aniline and *o*-toluidine to produce catechol and 3-methylcatechol, respectively (Takeo, Fujii et al., 1998a; Takeo, Fujii et al., 1998b). The aniline dioxygenase is encoded by five genes (*atdA1-A5*) that constitute four putative

components: AtdA1 which is a glutamine synthetase-like protein, AtdA2 which is a glutamine amidotransferase-like protein, AtdA3 and AtdA4 which resemble the large (α) and small (β) subunits of the terminal class dioxygenase, as well as AtdA5 which is a reductase component (Takeo, Fujii et al., 1998a). The putative reaction pathway of the AtdA enzyme is shown in Figure 6.1. It should be noted that the role of each component in the figure is speculative as there has been no detailed characterization of the function of each component in AtdA, or other closely related aniline dioxygenases, such as that from *Pseudomonas putida* UCC22 (pTDN1) (Fukumori and Saint, 1997). The lack of characterization of the structural determinant of substrate specificity of the AtdA enzyme has thus limited its development as a biocatalyst for the bioremediation of a wide range of aromatic amines.

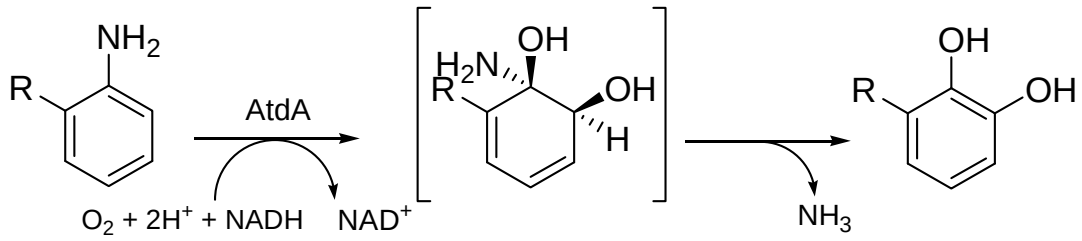


Figure 6.1 Putative aniline dioxygenation pathway of AtdA. Oxygen atoms are incorporated by AtdA into the 1 and 2 positions of the aniline aromatic ring to form a diol, and the amino group then leaves the ring spontaneously, or with the aid of AtdA1 and AtdA2, as suggested by Takeo et al., 1998 (Takeo, Fujii et al., 1998a).

The objective of this chapter is to identify and probe the residues determining the activity as well as the substrate specificity of the aniline dioxygenase using molecular modeling and saturation mutagenesis of the substrate binding pocket residues in AtdA3. The structure-function relationship elucidated from this work can

potentially be applied to the further engineering of AtdA to widen its utility as a biocatalyst. A homology model was built using the crystal structures of naphthalene dioxygenase, 1NDO (Kauppi, Lee et al., 1998), biphenyl dioxygenase, 1ULJ (Furusawa, Nagarajan et al., 2004), and cumene dioxygenase, 1WQL (Dong, Fushinobu et al., 2005), as templates. Fourteen residues within 4.5 Å of the substrate, forming the substrate binding pocket, were identified for saturation mutagenesis studies. Saturation mutagenesis of the substrate binding pocket residues widened the substrate specificity of AtdA to accept 2-isopropylaniline (2IPA), for which the wild type (WT) enzyme has no activity. The activity of AtdA for aniline and 2,4-dimethylaniline (24DMA) was also improved by 1.7- and 2.1-fold respectively.

This is the first known study on the molecular determinants of the substrate specificity of a four-component dioxygenase, AtdA, and it has shown that the α subunit of the terminal dioxygenase (AtdA3) indeed plays a part in the substrate specificity of AtdA. Results from this work will have important implications on the engineering of the aniline dioxygenase for the deamination of aromatic amines, for bioremediation and other industrial applications.

6.2 Materials and methods

6.2.1 Materials

Aniline, 2,4-dimethylaniline (24DMA), 2-isopropylaniline (2IPA), 2-secbutylaniline (2SBA), catechol, isopropyl-β-D-thiogalactopyranoside (IPTG), dimethylformamide (DMF), ampicillin and all other chemicals were purchased from Sigma (St Louis, MO) unless otherwise stated. 3-isopropylcatechol (3IPC) was purchased from Chem Service (West Chester, PA). Gibbs' reagent was purchased from MP Biomedicals (Solon, OH). Quikchange XL Site Directed Mutagenesis kit

and *Pfu Turbo* DNA polymerase were purchased from Stratagene (La Jolla, CA). Primers were purchased from Integrated DNA Technologies (Coralville, IA) and 1st Base (Singapore). All DNA gel purifications were carried out using the QIAEX II gel purification kit from Qiagen (Valencia, CA). All plasmid isolations were carried out using the QIAprep Miniprep kit from Qiagen. *E. coli* JM109 was purchased from Novagen (Madison, WI). The plasmid pTrc99A was from Amersham Pharmacia (Piscataway, NJ).

6.2.2 Homology modeling

A homology model of AtdA3 was constructed using the Insight II software (Insight II, version 2000; Accelrys Inc., San Diego, CA). The crystal structure of the α-subunit of naphthalene dioxygenase, 1NDO (Kauppi, Lee et al., 1998), biphenyl dioxygenase, 1ULJ (Furusawa, Nagarajan et al., 2004), and cumene dioxygenase 1WQL (Dong, Fushinobu et al., 2005), were used as templates. The sequence of AtdA3 was aligned with that of 1NDO, 1ULJ, and 1WQL using CLUSTALW (http://workbench.sdsc.edu/) and was adjusted to ensure that critical residues, such as the catalytic iron coordinating the facial triad of AtdA3 (H204, H209, and D356), were aligned with critical residues of NDO (H208, H213, and D362). Gaps in regions of secondary structures were avoided when the sequences were aligned. Three loop optimization models were generated for each model constructed with Insight II. All the models were checked with the PROSTAT and Profiles-3D functions in Insight II. The model with the highest overall score was chosen. The substrates were docked in the homology models of the WT AtdA3 and the mutants, V205A and I248L, using the Molecular Operating Environment (MOE) software (Chemical Computing Group Inc., Montreal, Canada). Mutations were introduced into the AtdA3 model using the

Rotamer Explorer function, and the rotamer with the lowest free energy was chosen. Each docking run consisted of 25 independent docks with six iteration cycles, and a random start was used to generate substrate positions within the docking box. From the results, the substrate orientation which gave the lowest interaction energy was chosen for another round of docking. A non-random start was used in this case. This process was repeated two times or until there was no significant decrease in the interaction energy of the substrate. The Conolly surface of the substrate binding pocket was generated using the Molecular Surface Function in MOE.

6.2.3 Saturation mutagenesis

Saturation mutagenesis library at each binding pocket residue was created using the Quikchange XL Site Directed Mutagenesis kit, with the plasmid pTA2-3 as the template. The primers listed in Table 6.1, together with their complements, were used in the saturation mutagenesis PCR. The PCR and transformation protocol recommended in the manual were used. Transformants were plated on LB agar plates containing 100 mg/l ampicillin and incubated overnight in 37 °C.

Table 6.1 Primers used in saturation mutagenesis. Underlined bases represent the randomized codon, where N = G, C, A or T and S = G or C

Residue	Sequence of forward primer used
N198	5'- GAA GCT GGT GTT TGA TNN SGC TGG TGA TGG CTA C -3'
G202	5'- GTG TTT GAT AAT GCT GGT GAT NNS TAC CAT GTC CCT TTT TCC -3'
V205	5'- GTG ATG GCT ACC ATN NSC CTT TTT CCC ATC AG -3'
L213	5'- CCT TTT TCC CAT CAG TCA TTG NNS CAA ATG ACT ACG CTT C -3'
I248	5'- GGT CAC TCA GTG NNS GAT CAG CGC -3'
Q250	5'- CAC TCA GTG ATA GAT NNS CGC CCG GAA ATG CAT -3'
K256	5'- CGC CCG GAA ATG CAT NNS GAG TCA GGG TGG GAT -3'
E257	5'- CCG GAA ATG CAT AAA NNS TCA GGG TGG GAT CAG -3'
W260	5'- GAG TCA GGG NNS GAT CAG CAG C -3'
A293	5'- GAG CGA GCG GTT GGT NNS GGA ATG AAT CTT AAT -3'
G294	5'- GCG AGC GGT TGG TGC TNN SAT GAA TCT TAA TAT T -3'
N296	5'- GTT GGT GCT GGA ATG NNS CTT AAT ATT TTT CCA AAC -3'
L304	5'- ATT TTT CCA AAC TTA NNS TTA ATT GGC AAC CAA ATA C -3'
F348	5'- CGC ACT CAG GAG GAT NNS CCA ATT ATG GGT GAG -3'

6.2.4 Screening method

The screening method was adapted from (Sakamoto, Joern et al., 2001) with modifications. Each colony of a library was picked into 200 µl of LB containing ampicillin (100 mg/l) in separate wells of a 96-well microplate. 186 clones were picked for each target residue, with three WT clones included as positive controls in each plate. The plates were incubated overnight at 37 °C with shaking at 250 rpm. 10 µl of the overnight culture was inoculated into new wells containing 90 µl of M9 minimal media supplemented with 5 µM $FeSO_4$, 100 mg/l ampicillin and 1 mM IPTG. Five replicates of each plate were made. The plates were incubated in 30 °C with 250 rpm shaking for 4 hr. Then, 100 µl of M9 media with 5 µM $FeSO_4$, 100 mg/l ampicillin, 1 mM IPTG and 2 mM substrate was added into each well of a plate. A different substrate was added to each plate. The substrates were aniline, 24DMA, 2IPA, and 2SBA. The plates were then incubated in 30 °C with 250 rpm shaking for 45 min for aniline and 4 hr for the other substrates. Optical density at 595 nm was measured after incubation. For aniline, 2IPA and 2SBA, 140 µl of 1M Tris-HCl (pH 5.8) was first added to each well followed by 10 µl of 0.32% (weight/volume) Gibbs' reagent in ethanol and absorbance at wavelength 560 nm was measured. For 24DMA, 10 µl of 0.32% Gibbs' reagent was added directly and absorbance at wavelength 620 nm was measured after 5min. The activity of each mutant, as indicated by the absorbance at 560 nm or 620 nm, was then normalized to its cell density (OD_{595}). Positive mutants from each screen were subjected to a second screen carried out in larger volumes, using culture tubes instead of 96 well microplates.

6.2.5 Whole cell activity assay

An overnight LB culture of JM109 with WT or mutant plasmid was inoculated into 150 ml LB to an OD_{600} of 0.02 and incubated in a 37 °C shaker at 250 rpm. When the OD_{600} reached 0.50 to 0.55, IPTG was added to a final concentration of 1 mM. The culture was then incubated in a 30 °C shaker at 250 rpm for 3 hr. The induced culture was then centrifuged at 4000 × g for 10 min. The supernatant was discarded and the cell pellet was resuspended with 150 ml of modified M9 buffer (M9 minimal media with 0.1% glucose). The resuspended cells were centrifuged using the same conditions again. The supernatant was discarded and the cell pellet was resuspended with modified M9 buffer to a final OD of about 10. Then 5 ml of the resuspended cells were aliquoted into a 50 ml centrifuge tube and 5 µl of 1 M substrate, dissolved in N,N-dimethylformamide, was added to a final concentration of 1 mM. The cells were then incubated at 30 °C with 250 rpm shaking. 0.5 ml samples were drawn at various time points. The samples were centrifuged at 16,000 × g in a benchtop centrifuge for 3 min and the supernatant was stored in -20 °C until ready for analysis.

The substrate and products were separated and quantified using high–pressure liquid chromatography (HPLC) with a 250 × 4.60 mm Synergi 4µ Polar-RP 80A column from Phenomenex (Torrance, CA). All HPLC methods used were isocratic with a flowrate of 1 ml/min. Aniline was analyzed using 90% potassium phosphate (pH 7.0) and 10% acetonitrile as mobile phase. 2IPA was analyzed using 60% potassium phosphate (pH 7.0) and 40% acetonitrile as mobile phase. 24DMA was analyzed using 70% potassium phosphate (pH 7.0) and 30% acetonitrile as mobile phase.

For each culture, 1 ml of the resuspended cells was centrifuged at 6,000 × g in a benchtop centrifuge for 3 min and the supernatant was discarded. The cell pellet was

resuspended in 50 mM Tris-HCl (pH 7.5) and disrupted using a single pass through the Constant Systems Cell Disruptor (Warwick, UK) at 20.3 kpsi. The disrupted cells were centrifuged at 16,000 × g in a benchtop centrifuge for 5 min and the supernatant was assayed for protein concentration using the BCA Protein Assay kit from Pierce (Rockford, IL). The whole cell activity was calculated by normalizing the initial rate of substrate conversion or product formation to the protein concentration.

6.2.6 Identification of products

E. coli JM109 cells with WT or mutant plasmid were grown, induced, washed and resuspended in modified M9 as described in the whole cell activity assay (Section 6.2.4). Substrate was added to a final concentration of 1 mM to 40 ml of the resuspended cells and the resting cell culture was incubated at 30 °C for 3 hr in a shaking incubator at 250 rpm. The culture was then centrifuged at 6000 × g for 10 min and the supernatant was extracted with ethyl acetate. The ethyl acetate was then evaporated with a rotary evaporator under vacuum at 40 °C and the residue was dissolved in 5 ml methanol. The sample was then analyzed by liquid chromatography-mass spectrometry (LC-MS) with an Agilent series 1100 HPLC (Agilent Technologies, Palo Alto, CA) coupled to an Applied Biosystems 4000 Q-Trap mass spectrometer. Separation was achieved with the 250 x 4.60 mm Synergi 4μ Polar-RP 80A column from Phenomenex. Isocratic methods with flowrate of 0.4 ml/min were used for all analyses. Aniline conversion product was analyzed using 60 % 20 mM ammonium acetate (pH 5.4) and 40 % acetonitrile as mobile phase. 2IPA conversion product was analyzed using 50 % 20 mM ammonium acetate (pH 5.4) and 50 % acetonitrile as mobile phase. 24DMA conversion product was analyzed using 40 % 20 mM ammonium acetate (pH 5.4) and 60 % acetonitrile as mobile phase. Negative

electrospray ionization mode with declustering potential and collision energy of -70 eV and -20 eV was employed, respectively.

For the ^{1}H-NMR analysis of the product of 24DMA conversion, the above assay was repeated using 200 ml of resuspended cells. After the extraction and evaporation of ethyl acetate, the sample was dissolved in a mixture of 95 % chloroform and 5 % methanol. The 24DMA dihydroxylation product was then purified using silica gel chromatography, with a mixture of 95 % chloroform and 5 % methanol as the mobile phase. The fraction containing the product was collected and dried with a rotary evaporator under vacuum at 40 °C. The sample was dissolved in $CDCl_3$ and analyzed by 500 MHz ^{1}H-NMR (Bruker AMX500) using tetramethylsilane as internal standard.

6.2.7 Sample preparation for SDS-PAGE analysis

Overnight cell cultures in LB with 100mg/l of ampicillin were inoculated into fresh LB with ampicillin (100 mg/l) at a volume ration of 1: 100 and incubated in a 37 ºC shaker at a speed of 250 rpm. When optical density (OD) reaches 0.5, the cultures were induced with 1mM IPTG and incubated in a 30 ºC shaker at a speed of 250 rpm for 3 hr. After induction, the OD of the cells was measured and the cells were then centrifuged at 6,000 × g in a Beckman J2-21M Induction Drive Centrifuge at 4 ºC. The supernatant is discarded and the cell pellet was resuspended in 50 mM Tris-HCl (pH 7.5). The final OD of the cells was adjusted to about 15. Then the cells were disrupted using a single pass through the Constant Systems Cell Disruptor (Warwick, UK) at 20.3 kpsi. The disrupted cells were centrifuged at 16,000 × g in a benchtop centrifuge for 5 min. The supernatant was separated from the cell pellet and stored at -20 ºC until ready for SDS PAGE analysis.

6.3 Identification of substrate binding pocket residues

A homology model of AtdA3 was built based on the crystal structures of naphthalene dioxygenase (Kauppi, Lee et al., 1998), biphenyl dioxygenase (Furusawa, Nagarajan et al., 2004) and cumene dioxygenase (Dong, Fushinobu et al., 2005). All three templates had more than 25% sequence identity with AtdA3.

To identify the substrate binding pocket of AtdA3, the largest substrate accepted by the WT AtdA, 2EA, was docked into the AtdA3 homology model. The approximate initial position of the substrate was determined based on the possible binding sites identified by the Site Finder function in MOE, as well as the relative position of indole in the crystal structure of naphthalene dioxygenase (Protein Data Base code 1O7N).

Eighteen residues within van der Waal's contact distance (4.5Å) of the substrate were identified as substrate binding pocket residues (Figure 6.2). These residues are N198, D201, G202, H204, V205, H209, L213, I248, Q250, K256, E257, W260, A293, G294, N296, L304, F348, and D356.

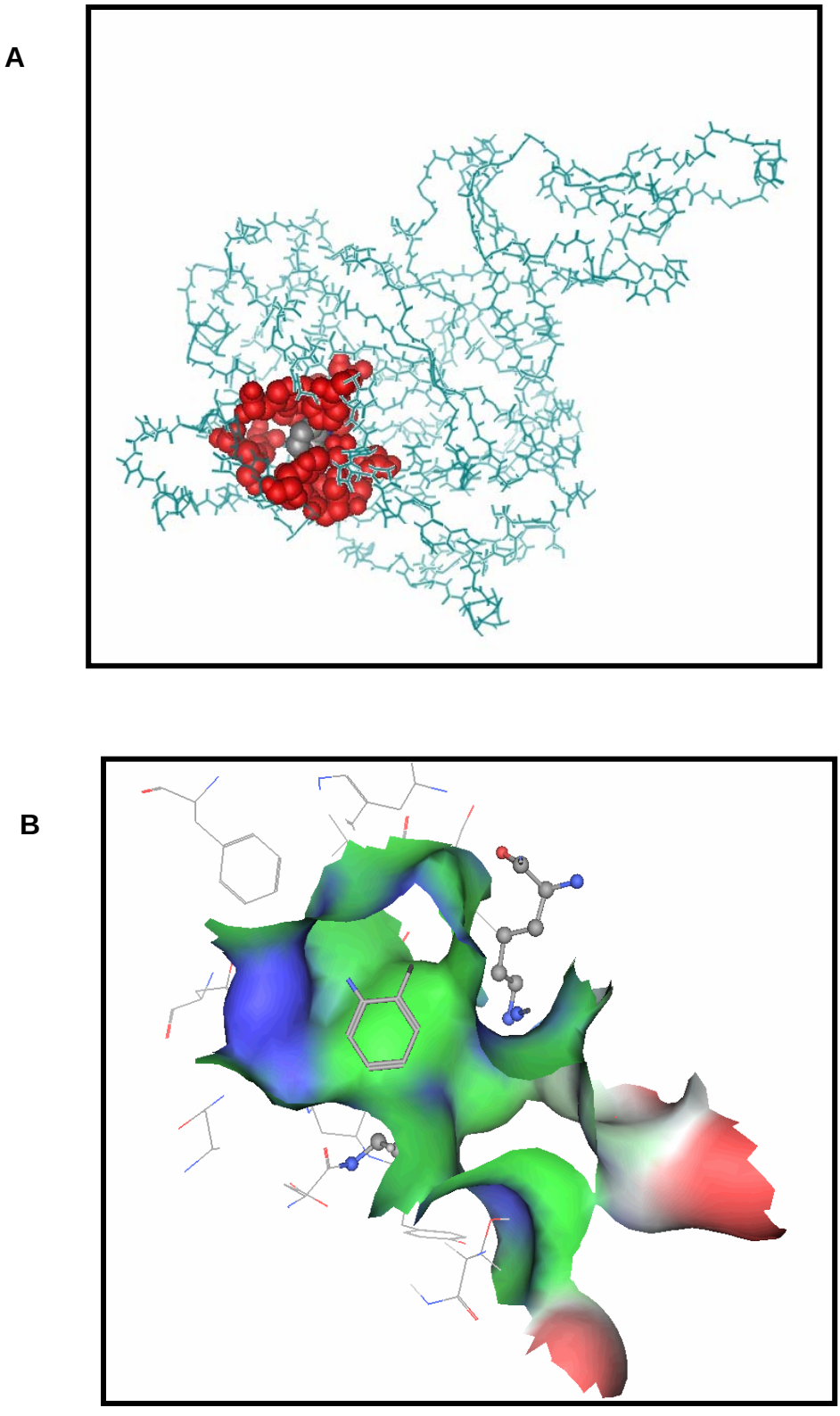


Figure 6.2 (A) Homology model of AtdA3, with 2EA (displayed in grey) docked to the active site. Residues in red form the binding pocket; (B) Close up of AtdA3 binding pocket and the substrate channel with 2EA in bound in the pocket.

6.4 Saturation mutagenesis

From the sequence alignment of AtdA3 with NDO, BphDO and CumDO, residues H204, H209, and D356 correspond to the catalytic facial triad that coordinates the mononuclear iron in the active site (H208, H213, and D362 of NDO (Kauppi, Lee et al., 1998)), while D201 corresponds to D205 of NDO, which plays a critical role in electron transfer between the Rieske [2Fe-2S] center of one α–subunit and mononuclear iron in the adjacent α subunit (Parales, Parales et al., 1999). Hence, these four critical residues were not subjected to saturation mutagenesis. The remaining fourteen sites were mutagenized individually using the NNS codon (where N denotes A, T, G, or C, while S denotes G, or C), resulting in 32 possible codon combinations for each site encoding all possible 20 amino acids. 186 clones were screened in two 96 well microplates per site, ensuring comprehensive coverage of all possible 19 mutations at each site, with three WT clones as control in each plate. Random clones were sequenced to ensure that the corresponding codons were successfully randomized.

Each library was screened using the Gibbs' reagent screening method adapted from (Sakamoto, Joern et al., 2001) with modifications as elaborated in Materials and Methods (Section 6.2.3). Mutants were selected based on improved activity for substrates which were accepted by the WT (aniline and 24DMA), or novel activity for the substrates 2IPA and 2BA. The screening scheme for 2IPA is shown in Figure 6.10. Schemes for the other substrates were the same except for the substrate added to the M9 media and the time of incubation.

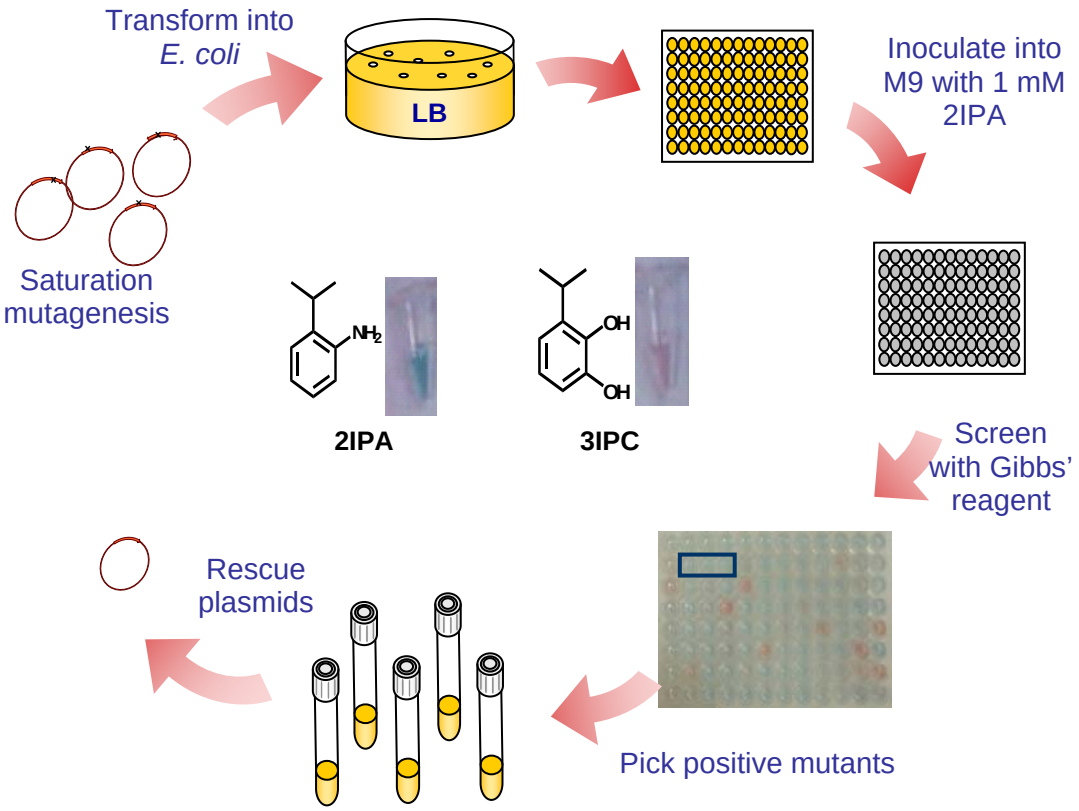


Figure 6.3 Screening scheme for saturation mutagenesis studies using 2IPA as substrate.

6.4.1 V205 Library

From the V205 saturation mutagenesis library, several mutants (1-K31 to 1-K36) with novel activity for 2IPA, a substrate not accepted by the WT enzyme, were found. These mutants were inoculated into LB and grown overnight in a 37 ºC shaker for a secondary screen. The overnight cultures were reinoculated into M9 media with 1mM IPTG and 2 mM 2IPA and incubated in a 30 ºC shaker for 8 hrs and analyzed with Gibbs' reagent. Results are shown in Figure 6.4. Each row in the figure contains the triplicates of the mutant and WT, with their positions randomized to ensure fairness in the test. All the mutants displayed activity for 2IPA in the secondary screen as indicated by the pink colored product formed with Gibbs' reagent while the WT cultures did not display any activity. The plasmids from these mutants were isolated and transformed into *E. coli* JM109. The *E. coli* JM109 cultures transformed

with the plasmids displayed activity for 2IPA as well, confirming that the 2IPA activity was a result of mutations in the AtdA3 gene and not the *E. coli* hosts. Sequencing of these mutants revealed that all had the V205A mutation.

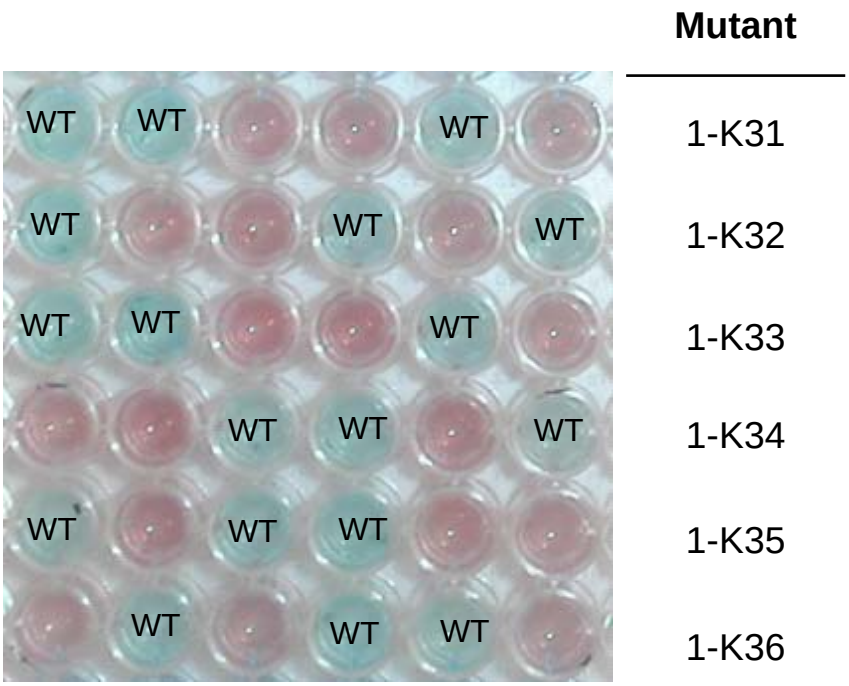


Figure 6.4 Secondary screening result of mutants displaying 2IPA activity. Wells with the mutants turned pink upon the addition of Gibbs' reagent due to the presence of 3IPC. The wells with WT turned blue as only 2IPA was present.

6.4.2 L248 library

The saturation mutagenesis library of I248 yielded two mutants (1-A21 and 1-A22) with improved aniline and 24DMA activity. Both mutants were subjected to the same secondary screen as the mutants from the residue 205 library, and the activities of the mutants for 24DMA, as calculated by the intensity of their color formation with Gibbs' reagent (normalized to their cell density), relative to that of the WT are presented in Figure 6.5. Sequencing revealed that both mutants had the I248L mutation.

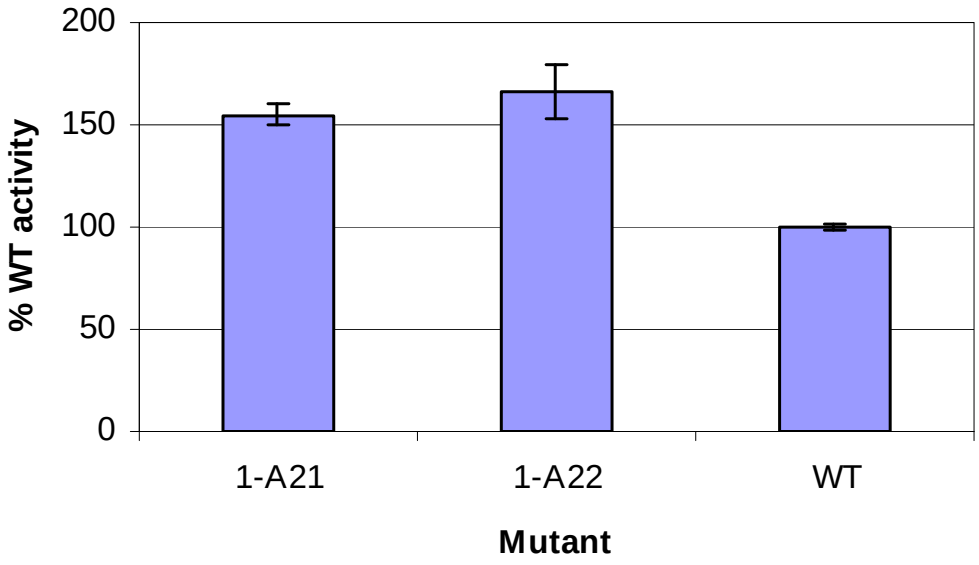


Figure 6.5 Activities of mutants for 24DMA from the saturation mutagenesis library of residue 248 relative to the WT.

6.4.3 F348 library

In studies on various other dioxygenases, the mutagenesis of the residue corresponding to F348 of AtdA3 (F352 of NDO from *Pseudomonas* sp. strain NCIB 9816-4) significantly altered the activity or the substrate specificity of the dioxygenase (Parales, Lee et al., 2000; Parales, Resnick et al., 2000; Pollmann, Wray et al., 2003; Keenan, Leungsakul et al., 2004; Rui, Kwon et al., 2004; Ju and Parales, 2006). However, mutation of residue F348 critically impaired the activity of the enzyme in this case. From the saturation mutagenesis library of residue 348, only five active mutants were found (Table 6.2), three of which had the parent residue, phenylalanine, at position 348. These residues were encoded by codon TTC instead of the parental codon of TTT. The other two active mutants were valine and tryptophan

mutants, neither of which had improved activity on aniline or 24DMA, or novel activity for 2IPA or 2SBA.

Table 6.2 Codon and amino acid identities of active F348 mutants.

Mutant	Codon	Amino Acid
WT	TTT	F
1-C11	**GTC**	**V**
1-C12	TTC	F
1-C13	TTC	F
1-C14	TTC	F
1-C15	**TGG**	**W**

6.5 SDS-PAGE analysis

Expression levels of AtdA in the V205A and I248L mutants were compared to that of the WT enzyme (Figure 6.6). Visual inspection of the SDS-PAGE gel showed no observable difference between the concentrations of the AtdA1 (56.8 kDa), AtdA2 (28.5), AtdA3 (50.3 kDa), AtdA4 (24.0 kDa), and AtdA5 (37.2 kDa) subunits in the mutants compared to their corresponding subunits in the WT. Thus the changes in activity and specificity of the mutants were not a result of altered expression.

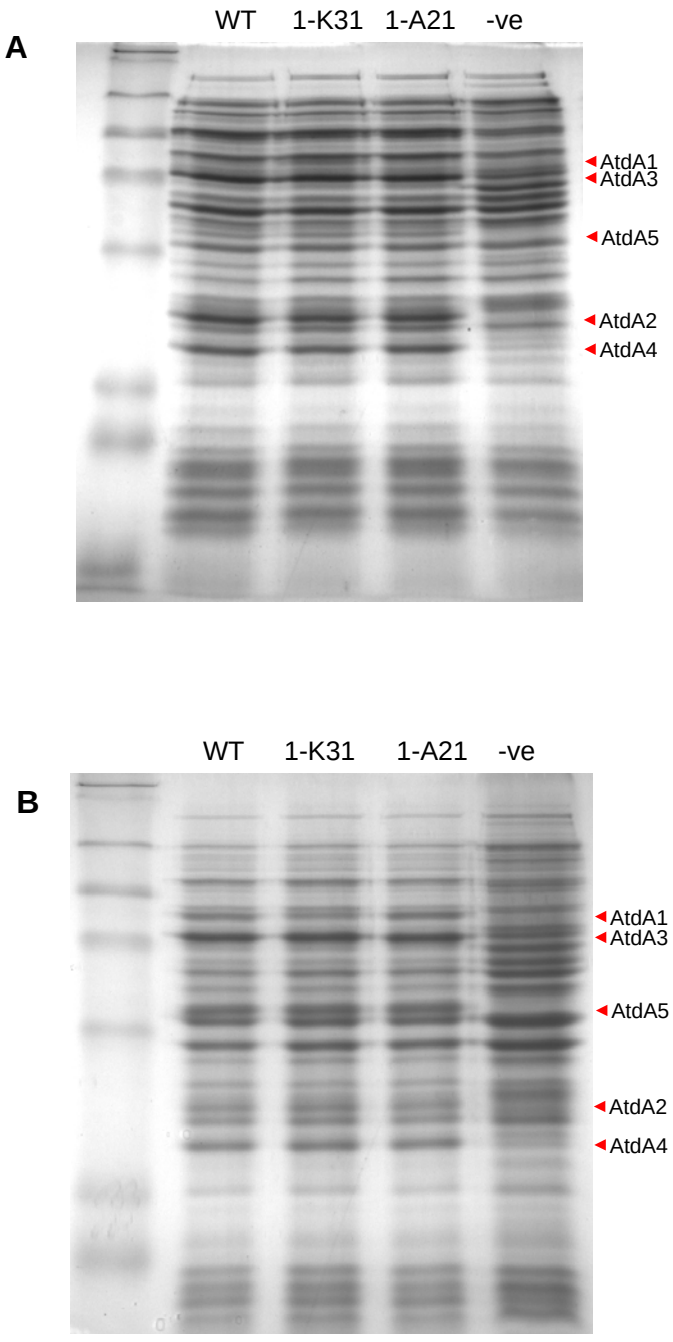


Figure 6.6 SDS-PAGE analysis of soluble fraction (A), and total fraction (B) of *E. coli* JM109 cells expressing AtdA WT and mutants. *E.coli* JM109 cells with pTrc99A plasmids were used as negative control (last lane in both panels).

6.6 Whole cell activity for 2IPA

The positive mutants of each library were characterized using the whole-cell activity assay as described in the Material and Methods section. The V205A mutation introduced a novel activity to the AtdA enzyme, enabling *E. coli* whole cells expressing the mutant to convert 2IPA at the rate of 1.1 nmol/min/mg protein to form 3IPC as the only product (Table 6.3).

The identity of 3IPC was confirmed by comparing its HPLC retention time with that of the authentic standard as well as by coelution with the authentic standard, and LC-MS analysis (*m/z* =151). In contrast, the 2IPA dihydroxylation activity was not detected at all in the WT enzyme or the I248L mutant. The V205A mutation also made the enzyme a better catalyst for the conversion 2IPA, a substrate not accepted by the WT enzyme, than for 24DMA, a substrate accepted by the WT enzyme.

6.7 Whole cell activity for aniline and 24DMA

The rate of catechol formation from aniline by whole cells expressing the I248L mutant is 45.3 nmol/min/mg protein, a 1.7-fold enhancement over the WT enzyme, while that of the V205A mutant was reduced to 3.1 nmol/min/mg protein (Table 6.3). For both these mutants, as well as the WT enzyme, the only product formed was catechol, as confirmed by HPLC coelution with the authentic catechol standard and LC-MS analysis (*m/z* = 109).

The 24DMA conversion rate of mutant I248L was enhanced by 2.1-fold over the WT enzyme to 5.9 nmol/min/mg protein. On the other hand, the 24DMA activity of the V205A mutant was reduced to 0.1 nmol/min/mg protein (Table 6.3). The 24DMA conversion products from I248L, V205A and WT had the same elution time on the

HPLC and all had a molecular ion at *m/z* = 137, corresponding to that of a dimethylcatechol, when analyzed with LC-MS.

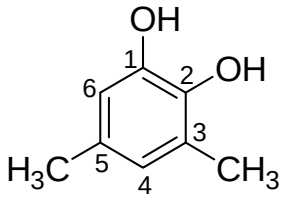
As there was no authentic standard, the product of 24DMA conversion by the WT enzyme was purified and further analyzed using ^{1}H-NMR. As shown in Table 6.4, the two methyl groups were detected at δ 2.20 (s) and δ 2.21 (s), the two aromatic protons at δ 7.26 (s), and the two hydroxyl groups at δ 6.51 (s) and δ 6.54 (s), confirming the product to be 3,5-dimethylcatechol. Thus, the regiospecificity of the enzyme was not altered by the I248L or V205A mutation as the only product from 24DMA conversion was 3,5-dimethylcatechol.

Table 6.3 Conversion rate of 2IPA, aniline and 24DMA by *E. coli* JM109 expressing the WT AtdA enzyme and the V205A and I248L mutants.

AtdA3	2IPA		Aniline		24DMA	
	Rate (nmol/min.mg protein)	Relative rate	Rate (nmol/min.mg protein)	Relative rate	Rate (nmol/min.mg protein)	Relative rate
WT	0	-	26.0 ± 0.20	1.00	2.8 ± 0.1	1.00
V205A	1.1 ± 0.2	∞	3.1 ± 0.10	0.12	0.1 ± 0.02	0.03
I248L	0	-	45.3 ± 7.20	1.74	5.9 ± 0.01	2.10

Table 6.4 500 MHz ^{1}H-NMR data (TMS internal standard) for 24DMA dihydroxylation product.

Protons	ppm
OH-1	6.54
OH-2	6.51
CH_3-3	2.20
CH_3-5	2.21
H-4	7.26
H-6	7.26



6.8 Analysis of mutations and discussion on AtdA1 and A2

Interestingly, the residues V205 and I248 have not been previously reported to influence the substrate specificity of a Rieske dioxygenase. The V205 residue corresponds to V209 in the naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816-4 (NDO) (Karlsson, Parales et al., 2003), V207 of naphthalene dioxygenase form *Ralstonia* sp. strain U2 (NagAc) (Fuenmayor, Wild et al., 1998), A223 of toluene 2,3-dioxygenase (TodC1) (Zylstra and Gibson, 1989), and A234 of biphenyl dioxygenases from *Burkholderia xenovorans* LB400 (BphA LB400)and *Pseudomonas pseudoalcaligenes* KF707 (BphA KF707) (Erickson and Mondello, 1992; Taira, Hirose et al., 1992). The sequence alignment of AtdA3 with the aforementioned dioxygenases is shown in Figure 6.7.

```
NagAc       ------------------MIYEN--LVSEAGLTQKHLIHGDKELFQHEL
NDO         ------------------MNYNNKILVSESGLSQKHLIHGDEELFQHEL
BphA LB400  MSSAIKEVQGAPVKWVTNWTPEAIRGLVDQEKGLLDPRIYADQSLYELEL
BphA_KF707  MSSSIKEVQGAPVKWVTNWTPEAIRGLVDQEKGLLDPRIYADQSLYELEL
TodC1       ----MNQTDTSPIRLRRSWNTSEIEALFDEHAGRIDPRIYTDEDLYQLEL
AtdA3       ------------------MKTINQLIQ--SGRVHRKVYTEASIFQAEM

NagAc       KTIFARNWLFLTHDSLIPSPGDYVTAKMGVDEVIVSRQNDGSVRAFLNVC
NDO         KTIFARNWLFLTHDSLIPAPGDYVTAKMGIDEVIVSRQNDGSIRAFLNVC
BphA LB400  ERVFGRSWLLLGHESHVPETGDFLATYMGEDPVVMVRQKDKSIKVFLNQC
BphA_KF707  ERVFGRSWLLLGHESHVPETGDFLATYMGEDPVVMVRQKDKSIKVFLNQC
TodC1       ERVFARSWLLLGHETQIRKPGDYITTYMGEDPVVVVRQKDASIAVFLNQC
AtdA3       DKIFQANWVFLLHASQIPKLDDYQTVRMGGRPLIVVRKGDDEFQALLNRC

NagAc       RHRGKTLVHAEAGNAKGFVCSYHGWGFGSNGELQSVPFEKELYG-----D
NDO         RHRGKTLVSVEAGNAKGFVCSYHGWGFGSNGELQSVPFEKDLYG-----E
BphA LB400  RHRGMRICRSDAGNAKAFTCSYHGWAYDIAGKLVNVPFEKEAFCDKKEGD
BphA_KF707  RHRGMRICRSDAGNAKAFTCSYHGWAYDIAGKLVNVPFEKEAFCDKKEGD
TodC1       RHRGMRICRADAGNAKAFTCSYHGWAYDTAGNLVNVPYEAESFA------
AtdA3       PHRGAKVCRNDSGNSKTFTCPYHGWKFRNSGKAFVIP-GANAYG-----E

NagAc       TIKKKCLGLKEVPRIESFHGFIYGCFDAEAPTLVDYLGDAAWYLEPIFKH
NDO         SLNKKCLGLKEVARVESFHGFIYGCFDQEAPPLMDYLGDAAWYLEPMFKH
BphA LB400  CGFDKAEWGPLQARVATYKGLVFANWDVQAPDLETYLGDARPYMDVMLDR
BphA_KF707  CGFDKAEWGPLQARVATYKGLVFANWDVQAPDLETYLGDARPYMDVMLDR
TodC1       -CLNKKEWSPLKARVETYKGLIFANWDENAVDLDTYLGEAKFYMDHMLDR
AtdA3       GFDKDNFSMTAIPRVESYRGFVFATSNENAVSLEEHLGSARQYIDEWLAH
```

Figure 6.7 Sequence alignment of AtdA3 with other Rieske dioxygenases.
(Continued on next page.)

```
NagAc        S-GGLELVGPPGKVVIKANWKAPAENFVGDAYHVG-WTHASSLRSGQSIF
NDO          S-GGLELVGPPGKVVIKANWKAPAENFVGDAYHVG-WTHASSLRSGESIF
BphA_LB400   TPAGTVAIGGMQKWVIPCNWKFAAEQFCSDMYHAGTTTHLSGILAG---I
BphA_KF707   TPAGTVAIGGMQKWVIPCNWKFAAEQFCSDMYHAGTTTHLSGILAG---I
TodC1        TEAGTEAIPGVQKWVIPCNWKFAAEQFCSDMYHAGTTSHLSGILAG---L
AtdA3        QGGEIKVSKSVQRYEIKCNWKLVFDN-AGDGYHVP-FSHQSLLQMTTLRY

NagAc        TPLAGNAMLPPEGAGLQMTSKYGSGMGVLWDGYSGVHSADLVPEMMAFG-
NDO         SSLAGNAALPPEGAGLQMTSKYGSGMGVLWDGYSGVHSADLVPELMAFG-
BphA_LB400   PPEMDLSQAQIPTKGNQFRAAWGGHGSGWYVDEPGSLLAVMGPKVTQYWT
BphA_KF707   PPEMDLSQAQIPTKGNQFRAAWGGHGSGWYVDEPGSLLAVMGPKVTQYWT
TodC1        PEDLEMADLAPPTVGKQYRASWGGHGSGFYVGDPNLMLAIMGPKVTSYWT
AtdA3        GGGDIQYFGNADETGMGLYALGNGHS--VIDQRPEMHKESGWDQQRPQPG

NagAc        -GAKQEKLAKEIGDVRARIYRSH-LNCTVFPNNSILTCSGVFKVWNPIDE
NDO          -GAKQERLNKEIGDVRARIYRSH-LNCTVFPNNSMLTCSGVFKVWNPIDA
BphA_LB400   EGPAAELAEQRLGHTGMPVRRMVGQHMTIFPTCSFLPTFNNIRIWHPRGP
BphA_KF707   EGPAAELAEQRLGHT-MPVRRMFGQHMSVFPTCSFLPAINTIRIWHPRGP
TodC1        EGPASEKAAERLGSVERGSKLMV-EHMTVFPTCSFLPGINTVRTWHPRGP
AtdA3        RESYETHVRNNSSQPARDLERAVGAGMNLNIFPNLLLIGNQIQVIDPISV

NagAc        NTTEVWTYAIVEKDMPEDLKRRLADAVQRTFGPAGFWESDDNDNMETESQ
NDO          NTTEVWTYAIVEKDMPEDLKRRLADSVQRTFGPAGFWESDDNDNMETASQ
BphA_LB400   NEIEVWAFTLVDADAPAEIKEEYRRHNIRNFSAGGVFEQDDGENWVEIQK
BphA_KF707   NEIEVWAFTLVDADAPAEIKEEYRRHNIRTFSAGGVFEQDDGENWVEIQK
TodC1        NEVEVWAFTVVDADAPDDIKEEFRRQTLRTFSAGGVFEQDDGENWVEIQH
AtdA3        NETVLHWHATLLAGDNEELN-AIRMRTQEDFPIMG--EVDDVANFESCQE

NagAc        NAKKYQSSNSDLIANLGFGKDVYGDECYPGVVAKSAIGETSYRGFYRAYQ
NDO          NGKKYQSRDSDLLSNLGFGEDVYGDAVYPGVVGKSAIGETSYRGFYRAYQ
BphA_LB400   GLRGYKAKSQPLNAQMGLGRSQTGHPDFPGNVG-YVYAEEAARGMYHHWM
BphA_KF707   GLRGYKAKSQPLNAQMGLGRSQTGHPDFPGNVG-YVYAEEAARGMYHHWM
TodC1        ILRGHKARSRPFNAEMSMDQTVDNDPVYPGRISNNVYSEEAARGLYAHWL
AtdA3        G---LETMPEIEWIDFSRHMNEGENACYQDVIQHKPTSEIHSRHYFDTWL

NagAc        AHISSSNWAEFENTSRNWHTELTKTTDR
NDO          AHVSSSNWAEFEHASSTWHTELTKTTDR
BphA_LB400   RMMSEPSWATLKP--------------
BphA_KF707   RMMSEPSWATLKP--------------
TodC1        RMMTSPDWDALKATR-------------
AtdA3        QLMSAVNKENQSEV--------------
```

Figure 6.7 (continued). Sequence alignment of AtdA3 with other Rieske dioxygenases. Residues highlighted in green represent V205A in AtdA3 and corresponding residues in the other dioxygenases, while residues highlighted in grey are the conserved residues constituting the active site iron binding catalytic facial triad in Rieske dioxygenases.

Based on the homology model of AtdA3, the residue V205 resides in the deepest and narrowest end of the substrate binding pocket and is found next to the facial triad of H204, H209, and D356, which coordinates the catalytic mononuclear iron. From the docking of 2IPA into the V205A mutant binding pocket, it was found that the isopropyl side chain of 2IPA comes within 4.25 Å of the A205 side chain (Figure 6.8A). In contrast, if 2IPA were to assume this position in the binding pocket of the WT enzyme, the side chain of V205 will come within 2.74 Å of the isopropyl side chain of 2IPA (Figure 6.8B). This may result in a steric clash that forces the substrate away from the active site iron, and prevents the substrate from coming into contact with the activated oxygen molecule bound to the catalytic iron, possibly explaining the lack of activity of the WT enzyme towards 2IPA. Removal of the methyl groups from residue 205 via a valine to alanine mutation removes the steric hindrance and allows the approach of 2IPA towards the catalytic iron.

The residue I248 lies at the entrance of the substrate binding pocket of the enzyme, leading to the substrate channel. Mutation from isoleucine to leucine results in a larger entrance to the substrate binding pocket (Figure 6.9A and B). This may allow for easier entry and exit of substrate and product molecules, hence the increase in the activity of the enzyme for all the substrates screened. However, this mutation does not change the specificity of the enzyme.

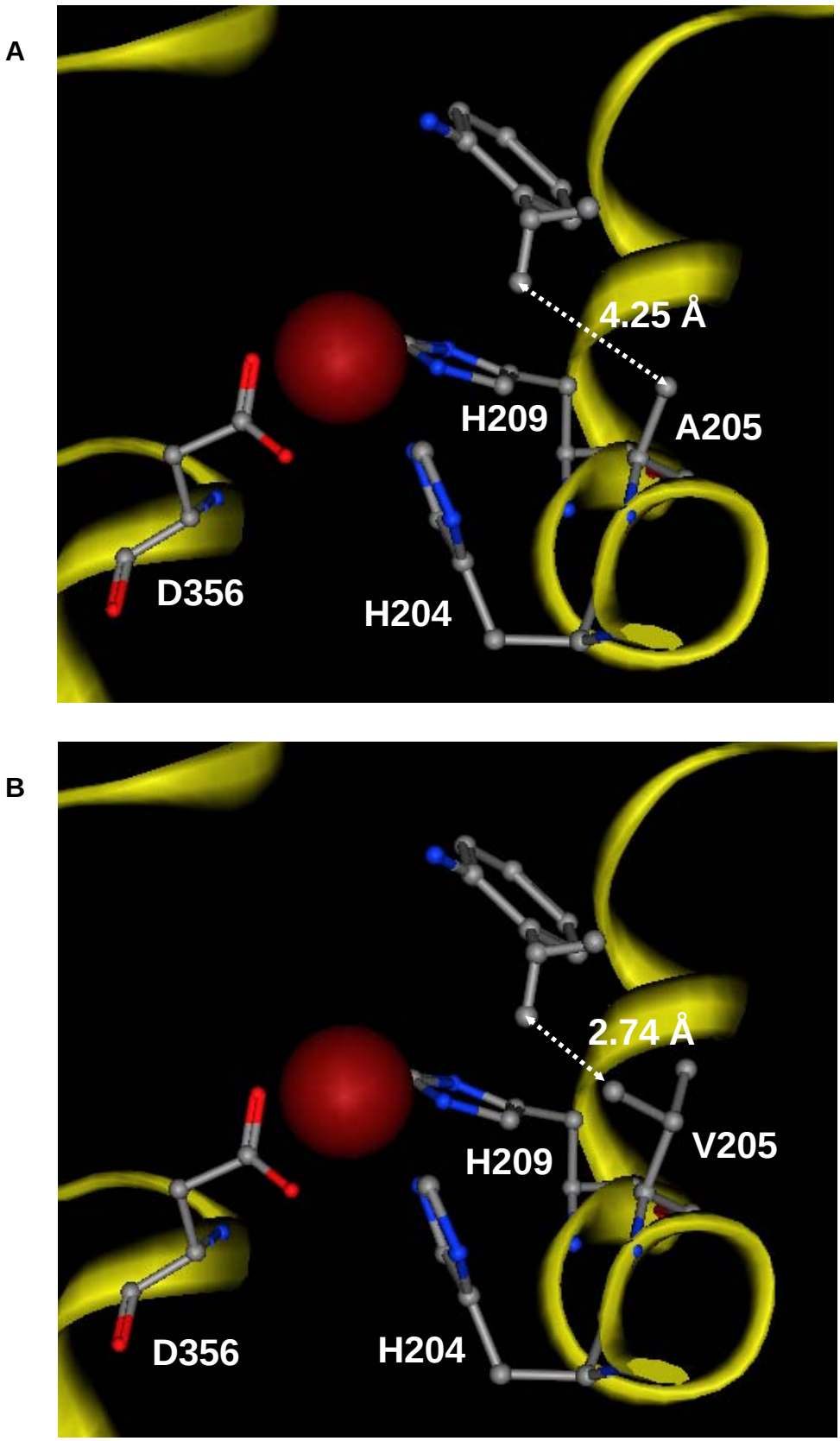


Figure 6.8 The position of the substrate, 2IPA, relative to residue 205 in the substrate binding pocket of (A) mutant V205A and (B) WT AtdA3. Also shown are the mononuclear iron (large brown sphere) and the catalytic facial triad of H204, H209 and D356.

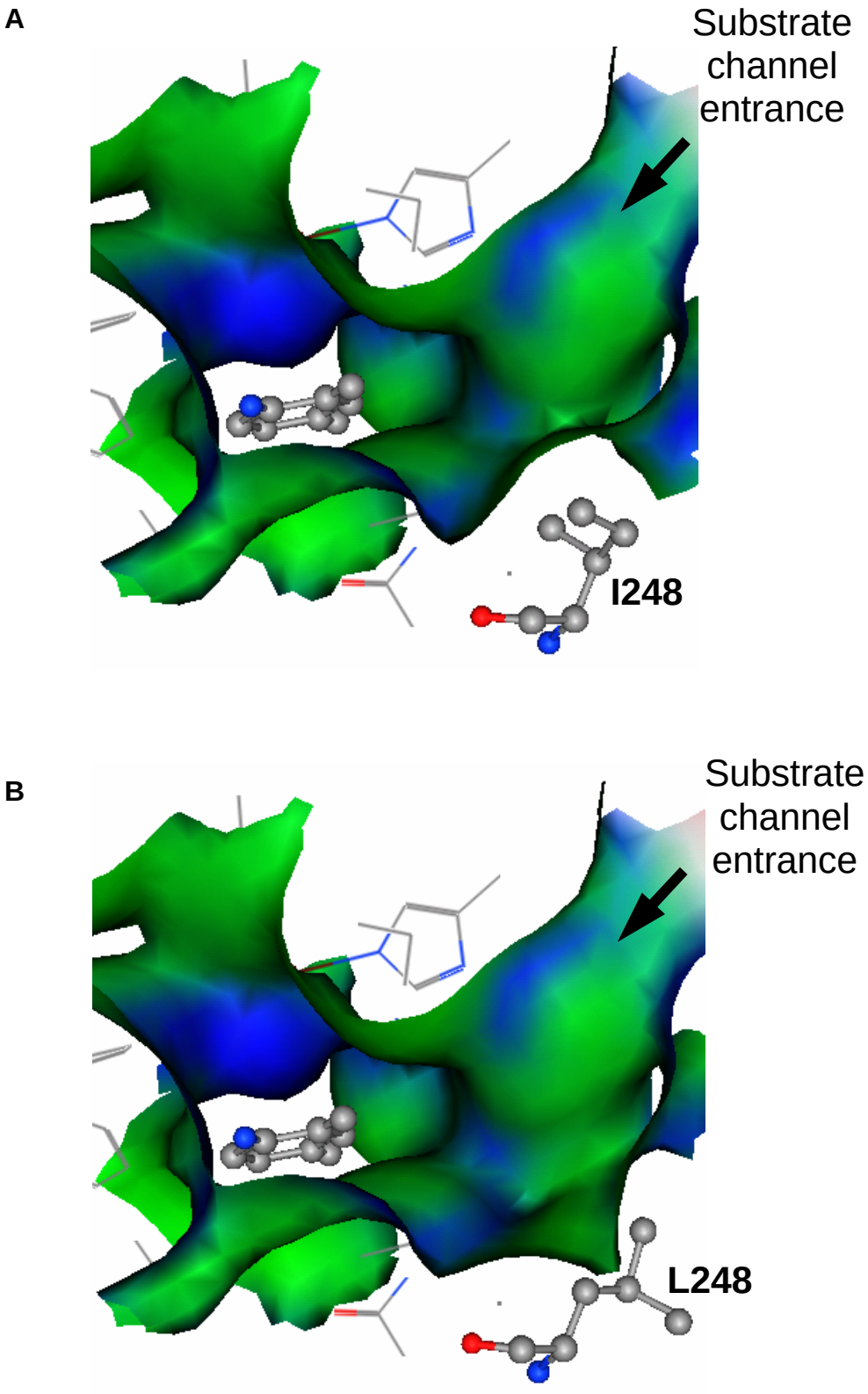


Figure 6.9 Molecular surfaces of the substrate channel leading to the binding pocket of the (A) WT AtdA3 and (B) mutant I248L.

Although it has been shown in this work that AtdA3 controls the substrate specificity of the aniline dioxygenase, it has yet to be explored whether the AtdA1 and AtdA2 components also control the substrate specificity. AtdA1 has a 25.8 % homology to glutamine synthetases from *Salmonella typhimurium* (GS) (Yamashita, Almassy et al., 1989), and the important ATP-binding motif and the tyrosine 426 corresponding to the adenylation site in GS are well conserved. AtdA1 also has a 62.1 % protein sequence identity with TdnQ of the aniline dioxygenase from *Pseudomonas putida* UCC22. It was reported that *E. coli* cells expressing TdnQ had no glutamine synthetase activity (Fukumori and Saint, 1997), suggesting that AtdA1 is unlikely to be involved in the recovery of nitrogen for biosynthesis reactions. AtdA2 exhibits homology to the class-I glutamine amidotransferase (GATs) domain in GMP synthetase (Tesmer, Klem et al., 1996). It has been postulated that since GS and GAT are involved in the addition of an amino group to glutamate and its release from glutamine respectively, AtdA1 and AtdA2 may be involved in the recognition and release of aniline amino groups (Takeo, Fujii et al., 1998a). Hence a similar engineering approach on AtdA1 and AtdA2 may offer useful insights on the substrate specificity and activity of the enzyme as well.

6.9 Summary

This is the first study on the molecular determinants for substrate specificity of a five subunit Rieske-dioxygenase, AtdA. In this study, a homology model was constructed to identify the residues defining the substrate binding pocket of the α-subunit of the aniline dioxygenase, AtdA3, and saturation mutagenesis was applied to these residues to probe the molecular determinants of the activity and specificity of the enzyme. In the study, it has clearly been demonstrated that the substrate

specificity as well as activity of the Rieske dioxygenase, AtdA, can be controlled by the α-subunit of the terminal dioxygenase, AtdA3.

It was found that the mutation V205A had the greatest effect on the substrate specificity on the enzyme as the mutant was able to dihydroxylate 2IPA, a substrate previously unaccepted by the WT enzyme. On the other hand, the I248L mutation enhances the activity of the enzyme for aniline and 24DMA, a carcinogenic pollutant for which no enzyme directly responsible for its biodegradation has been previously identified.

Although the V2025A mutation caused the loss of activity for aniline and 24DMA, the primary goal of this work, which is to probe the molecular determinants of AtdA, was achieved. This finding will facilitate future engineering efforts of the enzyme for both bioremediation and industrial applications using methods such as random mutagenesis and DNA shuffling to explore sequence space further away from the active sites.

6.10 References

Bomhard, E. M. and B. A. Herbold (2005). "Genotoxic activities of aniline and its metabolites and their relationship to the carcinogenicity of aniline in the spleen of rats." Crit Rev Toxicol **35** (10): 783-835.

Bugg, T. D. H. and C. J. Winfield (1998). "Enzymatic cleavage of aromatic rings: mechanistic aspects of the catechol dioxygenases and later enzymes of bacterial oxidative cleavage pathways." Nat Prod Rep **15**: 513-530.

Dong, X., S. Fushinobu, E. Fukuda, T. Terada, S. Nakamura, K. Shimizu, H. Nojiri, T. Omori, H. Shoun and T. Wakagi (2005). "Crystal structure of the terminal oxygenase component of cumene dioxygenase from *Pseudomonas fluorescens* IP01." J Bacteriol **187** (7): 2483-90.

Erickson, B. D. and F. J. Mondello (1992). "Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl dioxygenase, a multicomponent polychlorinated-biphenyl-degrading enzyme in *Pseudomonas* strain LB400." J Bacteriol **174** (9): 2903-12.

Fuenmayor, S. L., M. Wild, A. L. Boyes and P. A. Williams (1998). "A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2." J Bacteriol **180** (9): 2522-30.

Fukumori, F. and C. P. Saint (1997). "Nucleotide sequences and regulational analysis of genes involved in conversion of aniline to catechol in *Pseudomonas putida* UCC22(pTDN1)." J Bacteriol **179** (2): 399-408.

Furusawa, Y., V. Nagarajan, M. Tanokura, E. Masai, M. Fukuda and T. Senda (2004). "Crystal structure of the terminal oxygenase component of biphenyl dioxygenase derived from *Rhodococcus* sp. strain RHA1." J Mol Biol **342** (3): 1041-52.

Grayson, M., D. Eckroth, H. F. Mark, D. F.Othmer, C. G. Overberger and G. T. Seaborg (1984). Kirk-Othmer Encyclopedia of Chemical Technology, Vol. 2, Third Edition. New York, John Wiley and Sons**:** 309-375.

Ju, K. S. and R. E. Parales (2006). "Control of substrate specificity by active-site residues in nitrobenzene dioxygenase." Appl Environ Microbiol **72** (3): 1817-24.

Karlsson, A., J. V. Parales, R. E. Parales, D. T. Gibson, H. Eklund and S. Ramaswamy (2003). "Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron." Science **299** (5609): 1039-42.

Kauppi, B., K. Lee, E. Carredano, R. E. Parales, D. T. Gibson, H. Eklund and S. Ramaswamy (1998). "Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase." Structure **6** (5): 571-586.

Keenan, B. G., T. Leungsakul, B. F. Smets and T. K. Wood (2004). "Saturation mutagenesis of *Burkholderia cepacia* R34 2,4-dinitrotoluene dioxygenase at DntAc valine 350 for synthesizing nitrohydroquinone, methylhydroquinone, and methoxyhydroquinone." Appl Environ Microbiol **70** (6): 3222-31.

Lyons, C. D., S. Katz and R. Bartha (1984). "Mechanisms and pathways of aniline elimination from aquatic environments." Appl Environ Microbiol **48** (3): 491-6.

Markowitz, S. B. and K. Levin (2004). "Continued Epidemic of Bladder Cancer in Workers Exposed to *Ortho*-Toluidine in a Chemical Factory." J Occup Environ Med **46** (2): 154-160.

Nohmi, T., R. Miyata, K. Yoshikawa, M. Nakadate and M. Ishidate, Jr. (1983). "Metabolic activation of 2,4-xylidine and its mutagenic metabolite." Biochem Pharmacol **32** (4): 735-8.

Parales, R. E., K. Lee, S. M. Resnick, H. Jiang, D. J. Lessner and D. T. Gibson (2000). "Substrate specificity of naphthalene dioxygenase: effect of specific amino acids at the active site of the enzyme." J Bacteriol **182** (6): 1641-9.

Parales, R. E., J. V. Parales and D. T. Gibson (1999). "Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity." J Bacteriol **181** (6): 1831-1837.

Parales, R. E., S. M. Resnick, C. L. Yu, D. R. Boyd, N. D. Sharma and D. T. Gibson (2000). "Regioselectivity and enantioselectivity of naphthalene dioxygenase during arene *cis*-dihydroxylation: control by phenylalanine 352 in the alpha subunit." J Bacteriol **182** (19): 5495-504.

Pollmann, K., V. Wray, H. J. Hecht and D. H. Pieper (2003). "Rational engineering of the regioselectivity of TecA tetrachlorobenzene dioxygenase for the transformation of chlorinated toluenes." Microbiology **149** (Pt 4): 903-13.

Przybojewska, B. (1999). "Assessment of aniline derivatives-induced DNA damage in the liver cells of B6C3F1 mice using the alkaline single cell gel electrophoresis ('comet') assay." Cancer Lett **147** (1-2): 1-4.

Radomski, J. L. (1979). "The primary aromatic amines: their biological properties and structure-activity relationships." Annu Rev Pharmacol Toxicol **19**: 129-57.

Rai, H. S., M. S. Bhattacharyya, J. Singh, T. K. Bansal, P. Vats and U. C. Banerjee (2005). "Removal of Dyes from the Effluent of Textile and Dyestuff Manufacturing Industry: A Review of Emerging Techniques With Reference to Biological Treatment." Crit Rev Environ Sci Technol **35** (3): 219-238.

Rui, L., Y. M. Kwon, A. Fishman, K. F. Reardon and T. K. Wood (2004). "Saturation mutagenesis of toluene *ortho*-monooxygenase of *Burkholderia cepacia* G4 for Enhanced 1-naphthol synthesis and chloroform degradation." Appl Environ Microbiol **70** (6): 3246-52.

Sakamoto, T., J. M. Joern, A. Arisawa and F. H. Arnold (2001). "Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate." Appl Environ Microbiol **67** (9): 3882-7.

Shardonofsky, S. and K. Krishnan (1997). "Characterization of methemoglobinemia induced by 3,5-xylidine in rats." Journal of Toxicology and Environmental Health **50** (6): 595-604.

Taira, K., J. Hirose, S. Hayashida and K. Furukawa (1992). "Analysis of *bph* operon from the polychlorinated biphenyl-degrading strain of *Pseudomonas pseudoalcaligenes* KF707." J Biol Chem **267** (7): 4844-53.

Takeo, M., T. Fujii and Y. Maeda (1998a). "Sequence analysis of the genes encoding a multicomponent dioxygenase involved in oxidation of aniline and *o*-toluidine in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (1): 17-24.

Takeo, M., T. Fujii, K. Takenaka and Y. Maeda (1998b). "Cloning and sequencing of a gene cluster for the *meta*-cleavage pathway of aniline degradation in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (5): 514-517.

Tesmer, J. J., T. J. Klem, M. L. Deras, V. J. Davisson and J. L. Smith (1996). "The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families." Nat Struct Biol **3** (1): 74-86.

Weisburger, E. K., A. B. Russfield, F. Homburger, J. H. Weisburger, E. Boger, C. G. V. Dongen and K. C. Chu (1978). "Testing of twenty-one environmental aromatic amines or derivatives for long-term toxicity or carcinogenicity." Journal of Environmental Pathology and Toxicology **2** (2): 325-356.

Yamashita, M. M., R. J. Almassy, C. A. Janson, D. Cascio and D. Eisenberg (1989). "Refined atomic model of glutamine synthetase at 3.5 A resolution." J Biol Chem **264** (30): 17681-90.

Zylstra, G. J. and D. T. Gibson (1989). "Toluene degradation by *Pseudomonas putida* F1. Nucleotide sequence of the *todC1C2BADE* genes and their expression in *Escherichia coli*." J Biol Chem **264** (25): 14940-6.

Chapter 7
Further engineering of AtdA

7.1 Introduction

It was reported in the previous chapter that the AtdA3 subunit of the aniline dioxygenase (AtdA) plays a key role in controlling the substrate range and activity of the enzyme. With a V205A mutation in the AtdA3 subunit, the AtdA enzyme was able to accept a novel substrate 2-isopropylaniline (2IPA). However, the widened substrate range of the enzyme came at the expense of its activity. Compared to the wild type (WT), the V205A mutant had a 8.4- and 28-fold decrease in activity for aniline and 2,4-dimethylaniline (24DMA) respectively. To render AtdA suitable for bioremediation, the activity of the V205A mutant has to be increased. At the same time, a further increase in the substrate range would also be desirable.

In addition to bioremediation, AtdA can potentially be applied to industrial applications such as biorefining due to its ability to denitrogenate aromatic amines. With the depletion of crude oil reserves, middle and heavy petroleum feedstocks are becoming more important as precursors for lighter feedstocks. However, middle and heavy feedstocks often contain high levels of nitrogen impurities, which exist predominantly as heterocyclic aromatic compounds (Katzer and Sivasubramanian, 1979).

The combustion of nitrogen compounds results in the formation of nitrogen oxides (NO_x), which consequently contribute to acid rain and air pollution. Nitrogen compounds are strong inhibitors of the hydrodesulfurization (HDS) reactions during hydrotreatment

processes (Nagai and Kabe, 1983; Girgis and Gates, 1991). In addition, nitrogen compounds such as carbazole exhibited a retarding effect on its own hydrodenitrogenation reaction (Laredo, Montesinos et al., 2004). It was also found that the presence of nitrogen containing compounds leads to instability of fuels in storage by causing gum formation (Dinneen and Bickel, 1951; Ford, Holmes et al., 1981). Hence, treatment of heavy feedstocks to remove nitrogen contaminants is necessary to meet increasingly stringent environmental emission regulations as well as to maximize the efficiency of refinery processes.

Biological denitrogenation, which is the use of microorganisms to denitrogenate feedstocks, has advantages over industrial methods as it can be operated at ambient temperature and pressure, resulting in lower energy costs. In addition, the high selectivity of enzymes involved will not generate undesirable by-products. Most research on microbial denitrogenation has concentrated on the removal of non-basic nitrogen compounds as they represent the majority of total nitrogen and are more challeging to remove (Benedik, Gibbs et al., 1998).

One of the main components of the non-basic nitrogen compounds is carbazole (Mushrush, Beal et al., 1999; Laredo, Leyva et al., 2002), which has been used as a model non-basic compound in many previous microbial degradation studies (Grosser, Warshawsky et al., 1991; Ouchiyama, Zhang et al., 1993; Kobayashi, Kurane et al., 1995; Kirimura, Nakagawa et al., 1999; Schneider, Grosser et al., 2000; Kilbane, Daram et al., 2002).

Most of the microbial carbazole degradation pathways discovered use the same reaction steps, starting with the angular dioxygenation of carbazole to form 2’-

aminobiphenyl-2,3-diol (2'-ABPD), followed by *meta* cleavage of the diol ring to form 2-hydroxy-6-oxo-(2'-aminophenyl)hexa-2,4-dienoic acid. This intermediate is then degraded into 2-hydroxy-4-pentenoate and anthranilic acid, which then enters the tricarboxylic acid (TCA) cycle of the microorganism (Figure 7.1). This pathway results in the loss of precious fuel value of carbazole to biomass, which makes the microbial denitrogenation process economically unfeasible (Benedik, Gibbs et al., 1998). To date, there is no enzymatic pathway that can denitrogenate carbazole and at the same time preserve its fuel value.

As AtdA is capable of removing the amine group from aniline without a loss of carbon content, it has the potential to be applied to the denitrogenation of carbazole. This denitrogenation pathway can be achieved via the combination of carbazole-1,9a-dioxygenase (CarA) and a genetically engineered AtdA (Figure 7.2).

It was reported in Chapter 5 that AtdA only accepted aromatic amines with an *ortho*-substitution no larger than an ethyl group. The disparity in the sidechain size of an ethyl group and that of the 2ABPD, which is a dihydroxyl-benzene, is large. Due to this disparity in substrate size, creating AtdA activity for 2ABPD may require a large number of mutations to achieve success. The introduction of multiple mutations can generate a very large number of variants, making it physically impossible to screen enough mutants to comprehensively cover such a large sequence space (Kuchner and Arnold, 1997).

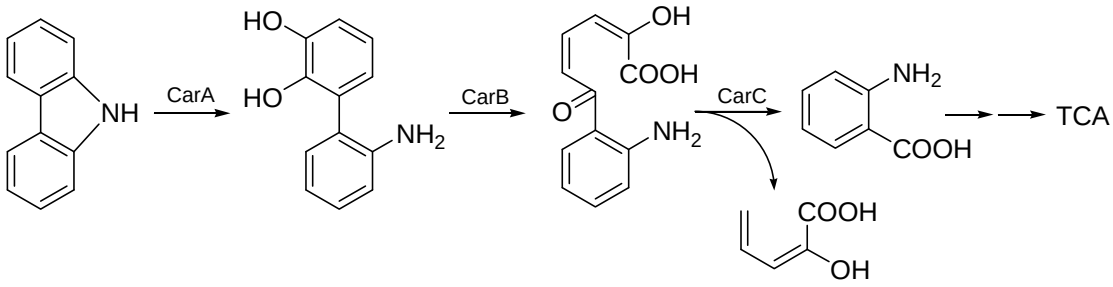
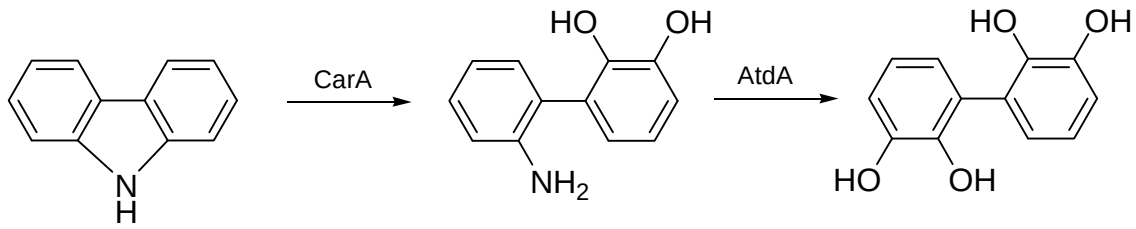


Figure 7.1 Common microbial carbazole degradation pathway.



Carbazole

2'-Aminobiphenyl-2,3-diol
(2'-ABPD)

Figure 7.2 Proposed carbazole denitrogenation pathway.

To overcome this problem, a stepwise evolution approach is proposed, whereby the size of the substrate is progressively enlarged with each round of mutagenesis (Figure 7.3). This method reduces the disparity in the target substrate size for each round, requiring a less drastic change in the level of enzyme promiscuity. Consequently the number of mutations required to achieve the objective of each round may be reduced. The V205A mutant from saturation mutagenesis of AtdA3, which already has activity for 2-isopropylaniline (2IPA) provides a good platform for further engineering AtdA towards 2ABPD activity.

The engineering of AtdA for bioremediation applications and the denitrogenation of carbazole can be carried out simultaneously since both applications share the same goal, which is to widen the enzyme's substrate range to accept larger substrates. Hence, the objective of this chapter is to widen the substrate range for both bioremediation and for carbazole denitrogenation and also to enhance the activity of AtdA for accepted substrates. A recent study of mutations that improved enzyme properties in directed evolution works found that mutations closer to the active site of an enzyme are more effective at creating new catalytic activity (catalytic promiscuity), while mutations both close and distant from the active site can improve activity (Morley and Kazlauskas, 2005). Hence, to further widen the substrate range of the aniline dioxygenase, AtdA, further rounds of saturation mutagenesis on the active site pocket residues was carried out using the V205A mutant as the parent. The best mutant from the saturation mutagenesis studies were then subjected to random mutagenesis using error-prone polymerase chain reaction (epPCR) to further improve its activity and substrate range.

Saturation mutagenesis will first be used to explore the sequence space of the substrate binding pocket residues for beneficial mutations followed by random mutagenesis to access the residues further away from the active site.

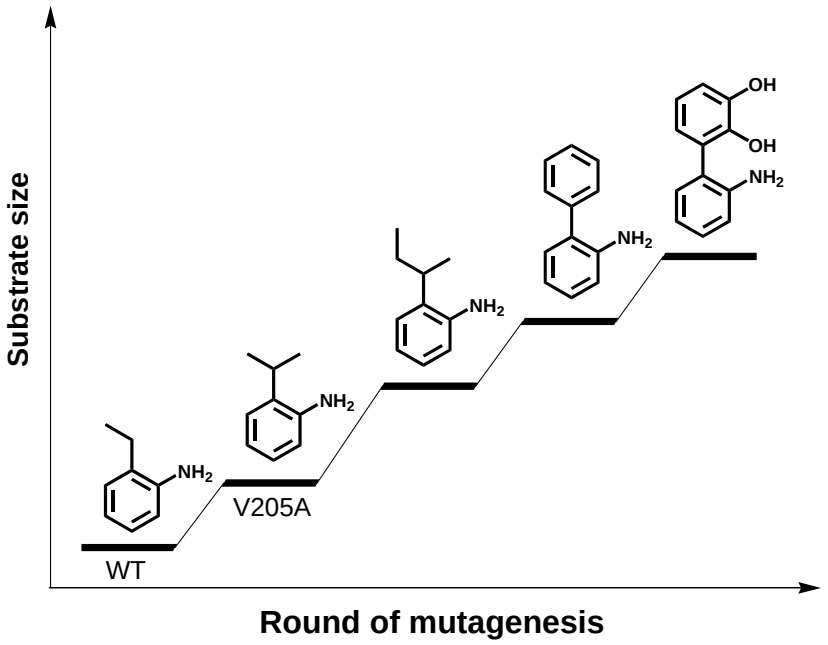


Figure 7.3 The proposed stepwise evolution strategy to engineer 2ABPD activity in AtdA.

7.2 Materials and methods

7.2.1 Materials

The GeneMorph II EZClone Domain Mutagenesis kit used for epPCR library creation was purchased from Stratagene (La Jolla, CA). All other materials were obtained from sources as described in Section 6.2.1.

7.2.2 Saturation mutagenesis

A saturation mutagenesis library at each substrate binding pocket residue was created using the Quikchange XL Site Directed Mutagenesis kit (Strategene), with the mutant 1-K31 and 2-A21 as parental templates for the second and third round respectively. The primers used were the same as those listed in Table 6.1, except for the residues G202 (in the second and third round) and residue Q250 (in the third round). The primers used to create these libraries are listed in Table 7.1. The PCR and transformation protocol recommended in the manual were used. Transformants were plated on LB agar plates containing 100 mg/l ampicillin and incubated overnight in 37 °C.

Table 7.1 Sequences of primers used in saturation mutagenesis which were changed for the second and third round of mutagenesis. Underlined bases represent the randomized codon, where N = G, C, A or T and S = G or C

Residue	Sequence of forward primer used
G202	5'- GTTTGATAATGCTGGTGAT<u>NNS</u>TACCATGCCCCTTTTTCC -3'
Q250	5'- CACTCAGTGCTCGAT<u>NNS</u>CGCCCGGAAATGCAT -3'

7.2.3 Random mutagenesis by error prone PCR

The random mutagenesis library was created using the GeneMorch II EZClone Domain Mutagenesis kit from Stratagene. For the mutant megaprimer synthesis, the primers used were A3 EcoRI F (5'-CTCGGTGGGTGATGTATAGAATTCAGGAACAGACCATGAAAACC-3') and A3 lib R2 (5'-CTTTATTATTATTCATGGCGTACCTCAACCGGT-3'). 700 ng of template was used. 25 amplification cycles was used in the PCR reaction with the melting temperature set at 54 °C. All other components and conditions were as recommended by the manufacturer's manual. The PCR product was gel purified with QIAEX II gel purification kit from Qiagen (Valencia, CA). The EZClone reaction was carried using the protocol recommended by in the manufacturer's manual. Transformants were plated on LB agar plates containing 100 mg/l ampicillin and incubated overnight in 37 °C.

7.2.4 Screening method

The screening method was adapted from (Sakamoto, Joern et al., 2001) with modifications. Each colony of a library was picked into 200 μl of LB containing ampicillin (100 mg/l) in separate wells of a 96-well microplate, with three WT clones included as positive controls in each plate. The plates were incubated overnight at 37 °C with shaking at 250 rpm. 10 μl of the overnight culture was inoculated into new wells containing 90 μl of M9 minimal media supplemented with 100 mg/l ampicillin and 1 mM IPTG. Two replicates of each plate were made. The plates were incubated in 30 °C with 250 rpm shaking for 4 hr. Then, 100 μl of M9 media with 100 mg/l ampicillin, 1 mM IPTG and 2 mM substrate was added into each well of a plate. A different substrate was

added to each plate. The substrates were 24DMA and 2SBA. The plates were then incubated in 30 °C with 250 rpm shaking for 4 hr. Optical density at 600 nm was measured after incubation. 10 μl of 0.32% Gibbs' reagent was added directly to each well and absorbance at wavelength 620 nm was measured after 5min. The activity of each mutant, as indicated by the absorbance at 620 nm, was then normalized to its cell density (OD_{600}). Positive mutants from each screen were subjected to a second screen carried out in larger volumes, using culture tubes instead of 96 well microplates.

7.2.5 Whole cell activity assay

The whole cell activity assay was carried out as described by the protocol in Section 6.2.5.

7.2.6 Sample Preparation for SDS-PAGE Analysis

The sample preparation for SDS-PAGE analysis was carried out as described by the protocol in Section 6.2.7.

7.3 Second round of saturation mutagenesis

A recent study on directed evolution found that mutations closer to the active site of enzymes were more effective at creating novel activity (Morley and Kazlauskas, 2005). Hence the engineering of AtdA would start with saturation mutagenesis of the active site residues. Fourteen active site residues were identified by homology modeling of AtdA3 using the crystal structures of naphthalene, biphenyl, and cumene dioxygenases (Karlsson, Parales et al., 2003; Furusawa, Nagarajan et al., 2004; Dong, Fushinobu et al., 2005) as

templates. These residues were N198, G202, L213, I248, Q250, K256, E257, W260, A293, G294, N296, L304, and F348. From the first round of saturation mutagenesis of these residues in Chapter 6, two mutants, 1-K31 and 1-A21, were identified. The 1-K31 mutant had a V205A mutation, which introduced activity for 2IPA but lowered the overall activity of AtdA for other substrates. The 1-A21 mutant, which had a I248L mutation, increased the AtdA activities for aniline and 24DMA but did not widen the substrate range of AtdA.

Using mutant 1-K31 as the parent, the remaining active site residues of AtdA3 were subjected to a second round of saturation mutagenesis. Mutant 1-K31 was chosen over mutant 1-A21 as the parental template because creating activity for novel substrates was deemed more challenging than improving the activity for substrates already accepted by the WT. 186 clones from each saturation mutagenesis library was screened for novel activity against 2-secbutylaniline (2SBA) and improved activity against 2,4-dimethylaniline (24DMA) using the Gibbs' reagent screening method.

In the second round of mutagenesis, five mutants (2-A21, 2-A22, 2-A23, 2-A24 and 2-A25) with improved activity for 24DMA from the I248 library were isolated. All the mutants were screened a second time using in culture tubes instead of microplates to confirm their improved activity for 24DMA. All the mutants were found to be more active than the parent 1-K31 in the second screen. Sequencing of the mutants revealed all of them to have an isoleucine to leucine mutation at residue 248. The mutant 2-A21, which has the V205A/I248L double mutation, was transformed into *E. coli* JM109 strain for further analysis. However, no active mutants for 2SBA were isolated.

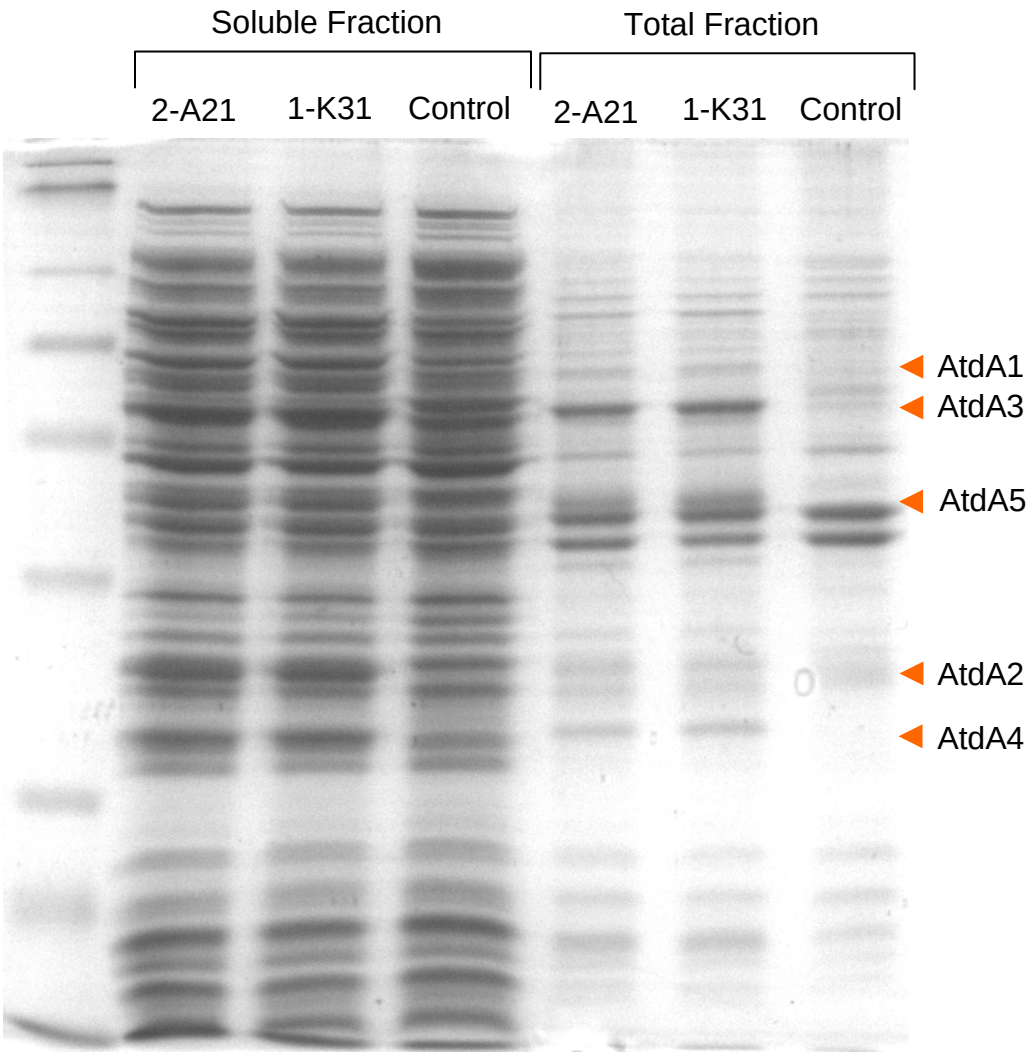


Figure 7.4 SDS-PAGE analysis of the soluble and total fractions of *E. coli* JM109 cells expressing mutant 2-A21 and 1-K31. *E. coli* JM109 cells with the empty vector, pTrc99A, were used as negative controls.

7.4 SDS-PAGE analysis of mutant 2-A21

Expression level of AtdA in the V205A/I248L double mutant (mutant 2-A21) was compared to that of the parent, mutant 1-K31 using SDS-PAGE analysis (Figure 7.4). Visual inspection of the SDS-PAGE gel showed no observable difference between the concentrations of the AtdA1 (56.8 kDa), AtdA2 (28.5), AtdA3 (50.3 kDa), AtdA4 (24.0 kDa), and AtdA5 (37.2 kDa) subunits in the mutant compared to their corresponding subunits in the WT. Thus, the improved activity of the mutant was not a result of altered expression.

7.5 Whole cell activity of mutant 2-A21

The activity of mutant 2-A21, which had the V205A/I248L double mutation, was characterized using the whole cell activity. The substrates used in the assay were aniline (AN), 24DMA, and 2IPA. The results are presented in Table 7.2. The activities of 2-A21 for these three substrates was compared to that of its parent, mutant 1-K31 to gauge the impact of the I248L mutation. The whole cell activities of mutant 1-K31 for the substrates were obtained from the previous chapter (Section 6.7).

The activity of mutant 2-A21 for AN was 9.3 nmol/min/mg protein while that of the parent 1-K31 was 3.1 nmol/min/mg protein. Thus, the introduction of the I248L mutation to mutant 1-K31 resulted in a 3-fold increase in activity for AN. The sole product of AN conversion was catechol, as evidenced by the product's HPLC retention time compared to that of the authentic standard. The I248L mutation in this case resulted in a 17 fold improvement of activity for 24DMA, with the mutant 2-A21 converting 24DMA at a rate of 1.7 nmol/min/mg protein and the parent 1-K31 only converting the same substrate at a

rate of 0.1 nmol/min/mg protein. The mutant 2-A21 converted 24DMA to 3,5-dimethylcatechol (35DMC), as confirmed by coelution with the compound purified as described in the previous chapter (Section 6.2.5). For the substrate 2IPA, mutant 2-A21 displayed an activity of 1.6 nmol/min/mg protein, a 1.5 fold improvement in activity over mutant 1-K31, which had an activity of 1.1 nmol/min/mg protein. 3-Isopropylaniline (3IPC) was the sole product of 2IPA conversion by mutant 2-A21. Again HPLC coelution with authentic 3IPC was the means of identifying the product.

The introduction of I248L mutation had a greater effect on the activity of mutant 1-K31 than it had on the WT, for which the activity for AN and 24DMA was increased by 1.7 and 2.0-fold respectively. From the analysis of the homology model of AtdA3, residue 248 and residue 205 are about 9.65 Å away from each other and do not come into direct contact (Figure 7.5). Hence, the effects of the two mutations are likely to be independent. It was postulated that the I248L mutation improved activity of the enzyme by enlarging the entrance of the substrate binding pocket. The low activity of 1-K31 for AN and 24DMA would mean that the same increase in the ease of substrate entry and product exit from the substrate binding pocket would bring about a greater factor in activity improvement compared to the WT.

Despite the improvement of overall activity of the enzyme in this round of saturation mutagenesis, the AN and 24DMA activity of mutant 2-A21 was still lower than that of the WT, which displayed a rate of 26.0 and 2.8 nmol/min/mg protein for AN and 24DMA respectively. Further engineering of the enzyme would be required to increase the activity of the enzyme.

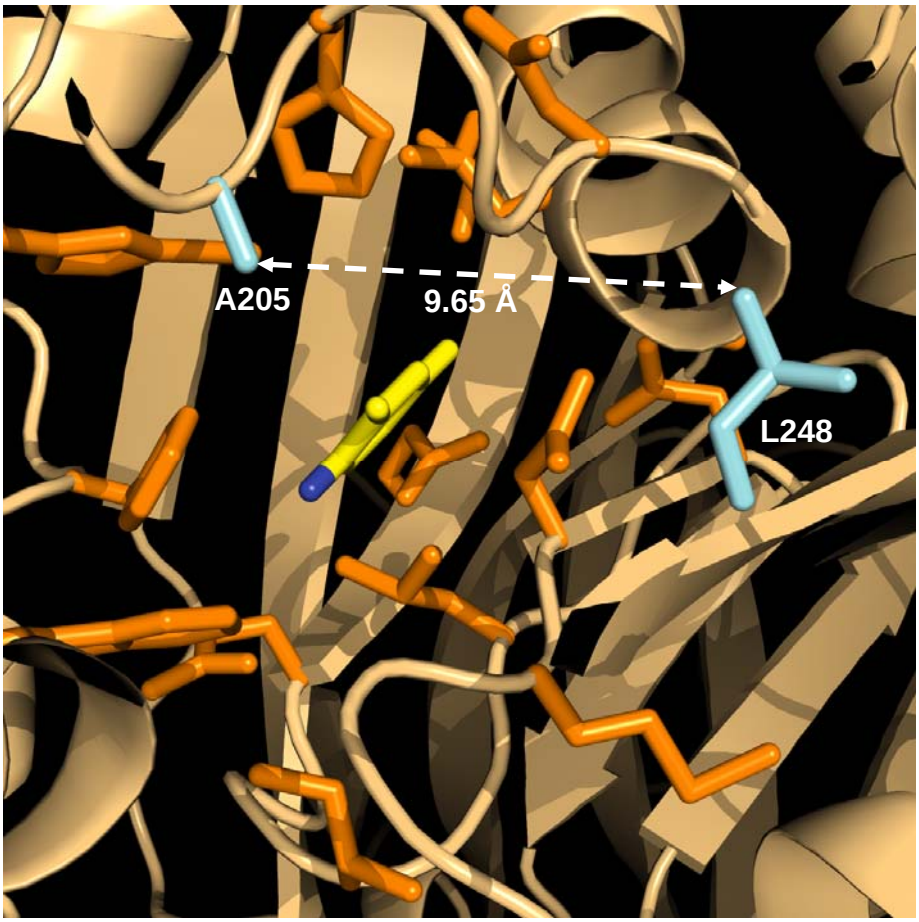


Figure 7.5 Intermolecular distance between L248 and A205 in mutant 2-A21. Residues in orange represent the other substrate binding pocket residues while the molecule in yellow and blue represent 24DMA.

Table 7.2 Conversion rate of aniline, 24DMA, and 2IPA by *E. coli* JM109 expressing the AtdA mutants 1-K31 and 2-A21. Relative rates are calculated with respect to mutant 1-K31, the parent in the second round of active site residue saturation mutagenesis.

Mutant	Mutation	AN		24DMA		2IPA	
		Rate (nmol/min/mg protein)	Relative rate	Rate (nmol/min/mg protein)	Relative rate	Rate (nmol/min/mg protein)	Relative rate
1-K31	V205A	3.1 ± 0.1	1.0	0.1 ± 0.02	1.0	1.1 ± 0.2	1.0
2-A21	V205A/ I248L	9.3 ± 2.4	3.0	1.7 ± 0.1	17.0	1.6 ± 0.5	1.5

7.6 Third round of saturation mutagenesis

Using mutant 2-A21 as the parent, another round of saturation mutagenesis was conducted on the remaining active site residues (N198, G202, L213, Q250, K256, E257, W260, A293, G294, N296, L304, and F348) of AtdA3. The target substrates were again 24DMA and 2SBA. However, no beneficial mutation was found in this round of saturation mutagenesis.

7.7 Directed evolution of AtdA3 by random mutagenesis

As further rounds of saturation mutagenesis of the AtdA3 active site residues did not improve AtdA activity, the focus was shifted to residues further away from the active site of the enzyme. To enhance the activity and substrate range of AtdA3, the mutant 2-A21 was subjected to random mutagenesis using error-prone polymerase chain reaction (epPCR) to generate the mutant library.

7.7.1 Mutation Rate of epPCR library

As epPCR only introduces random point mutations at the DNA level the accessible sequence diversity is limited. On average, about 5.7 different amino acids are accessible by a single-base-pair change in a codon (Kuchner and Arnold, 1997). Mutation of more than one nucleotide in a codon is still possible but the probability is very low as the mutations have to be introduced less than three bases apart. Hence, the number of possible variants of a protein that can be created by introducing *n* mutations in a protein *m* amino acids in length using epPCR is given by ${}^{m}C_{n} \times 5.7^{n}$.

For the AtdA3 subunit, which is 426 amino acids in length, the number of possible mutants that can be created for a given number of amino acid mutations is shown in Table 7.3. For random mutageneis, the library size to screen should be about 10 times the number of possible variants to ensure a 95 % confidence that all possible variants are covered (Kuchner and Arnold, 1997).

Table 7.3 The number of possible variants created by introducing *n* mutations in AtdA3 using epPCR and the number of clones to screen for comprehensive coverage.

n	Number of possible variants	Number of clones to screen
1	2,428	2.4×10^4
2	2.94×10^6	2.94×10^7
3	2.37×10^9	2.37×10^{10}
4	1.43×10^{12}	1.43×10^{13}
5	6.87×10^{14}	6.87×10^{15}

The feasible upper limit for screening using a high throughput assay is about 10^5, which is exceeded when more than one mutation is introduced per *atdA3* gene. Hence, the one amino acid mutation per *atdA3* gene was set as the target mutation rate. This would require approximately 1 to 3 nucleotide mutations per gene on the DNA level.

7.7.2 Random mutagenesis of AtdA3

The epPCR library of the 1.3kb *atdA3* gene was generated using the GeneMorph II EZclone Domain Mutagenesis kit as described in the Materials and Methods section. Sequencing of randomly selected clones from the epPCR library of *atdA3* revealed that the mutation frequency of the library was ~2.7 nucleotide/gene. As there are four possible

transition mutations (purine to purine and pyrimidine to pyrimidine changes) and eight possible transversion mutations (purine to pyrimidine changes and vice versa), an ideal polymerase without mutational bias would have a transition to transversion ratio of 0.5. The transition to transversion ratio of the library in this case was about 0.9, showing a slight bias for transitions mutations over transversion mutations. This deviation from the ideal ratio was a result of the inherent bias of the Mutazyme II DNA polymerase used in DNA amplification (Stratagene, 2006), which is a common drawback of polymerases used in epPCR processes (Wong, Tee et al., 2004) and may limit the actual diversity of the library at the amino acid level.

Approximately 2500 clones were screened for improved activity against 24DMA and novel activity for 2SBA. From the screen, one mutant, 3-R21 exhibited significantly improved 24DMA activity over the parent 2-A21 was isolated. The 3-R21 mutant was then subjected to a secondary screen in a culture to ascertain its activity. Sequencing of the *atdA3* gene revealed a serine to cysteine mutation at residue 404 (S404C mutation). No mutant with 2SBA activity was isolated.

7.8 SDS-PAGE analysis of mutant 3-R21

The expression level of AtdA in *E. coli* XL10 from mutant 3-R21 was compared to that of its parent, 2-A21. The SDS PAGE gel is shown in Figure 7.6. Visual inspection of the SDS-PAGE gel showed that the AtdA1, AtdA3, AtdA4, and AtdA5 subunits in mutant 3-R21 were expressed at a slightly higher level compared to their corresponding subunits in the 2-A21. However, the expression level of AtdA2 is too low in both 3-R21 and 2-A21 for comparison.

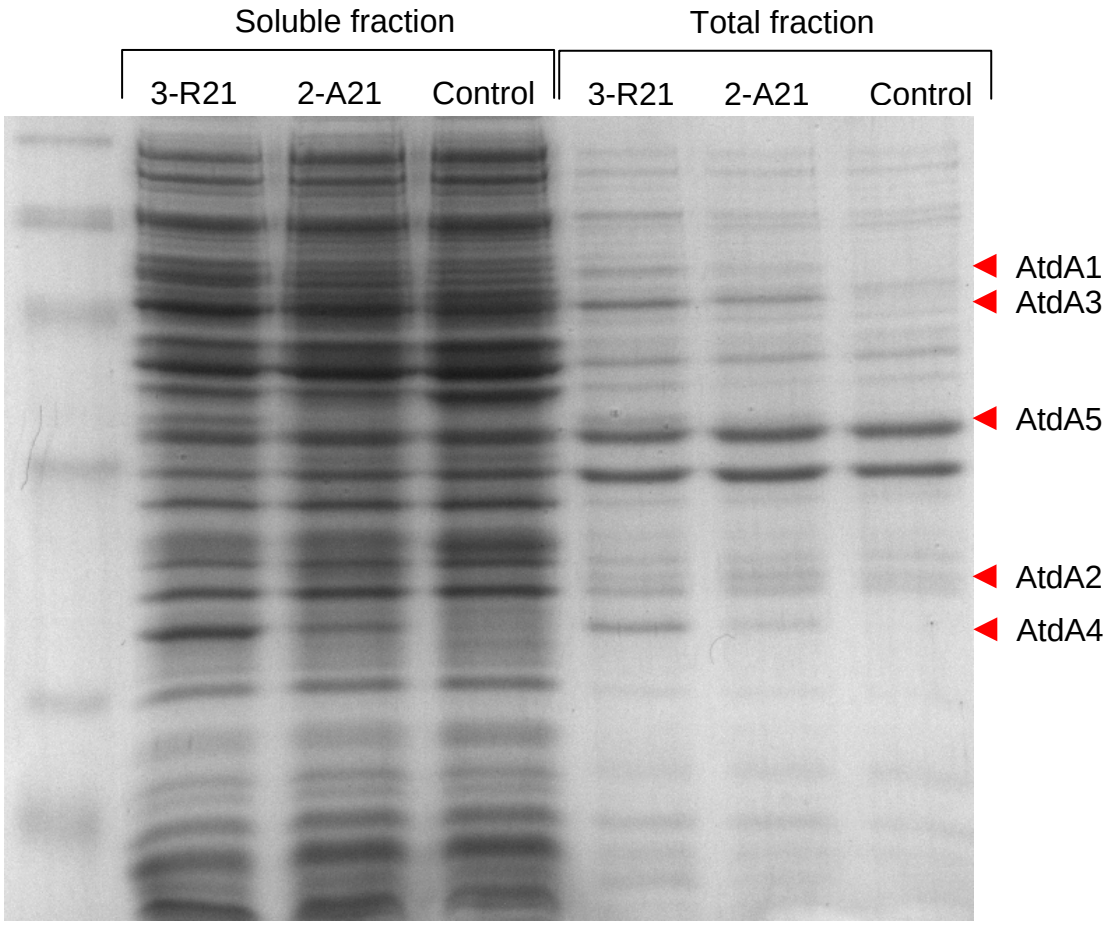


Figure 7.6 SDS-PAGE analysis of the soluble and total fractions of *E. coli* XL10 cells expressing mutant 3-R21 and 2-A21. *E. coli* XL10 with the empty vector, pTrc99A, was used as a negative control.

7.9 Whole cell activity of mutant 3-R21

The activity of mutant 3-R21 for 24DMA was compared to its parent 2-A21 using the whole cell activity assay. 1 mM of 24DMA was added to each resting cell culture and the fraction of 24DMA remaining with time is shown in Figure 7.7. Mutant 3-R21 was able to convert all 24DMA within 90 min while the parent 2-A21 was much slower, with less than 30 % of DMA converted after 210 min.

The 2-A21 mutant had an activity of 1.4 nmol/min/mg protein, calculated based on the initial rate of 24DMA conversion normalized to total protein concentration. This value corroborates with the activity found in the previous activity assay found in Section 7.5 (1.7 nmol/min/mg protein). On the other hand, the 3-R21 mutant displayed an activity of 9.8 nmol/min/mg protein, which is a 7-fold increase in activity compared to 2-A21. The regiospecificity of mutant 3-R21 was unchanged as the only product formed was 35DMC, which was confirmed by HPLC coelution with the standard. The activity of this mutant is higher than that of the 1-A21 mutant found in Chapter 6 (5.9 nmol/min/mg protein).

Similarly, the mutant 3-R21 was tested for its activity on AN and 2IPA. The mutant 3-R21 had activity of 27.7 nmol/min/mg protein. This is a 3-fold increase in activity over mutant 2-A21 (9.3 nmol/min/mg protein) from the previous round. Catechol is the sole product of aniline conversion by 3-R21. For 2IPA, mutant 3-R21 exhibited a conversion rate of 2.2 nmol/min/mg protein, which is a 1.4-fold increase over mutant 2-A21, which displayed an activity of 1.6 nmol/min/mg protein. The product of 2IPA conversion by 3-R21 was 3IPC. The activities of 3-R21 for AN, 24DMA, and 2IPA are tabulated in Table 7.4 together with that of 2-A21.

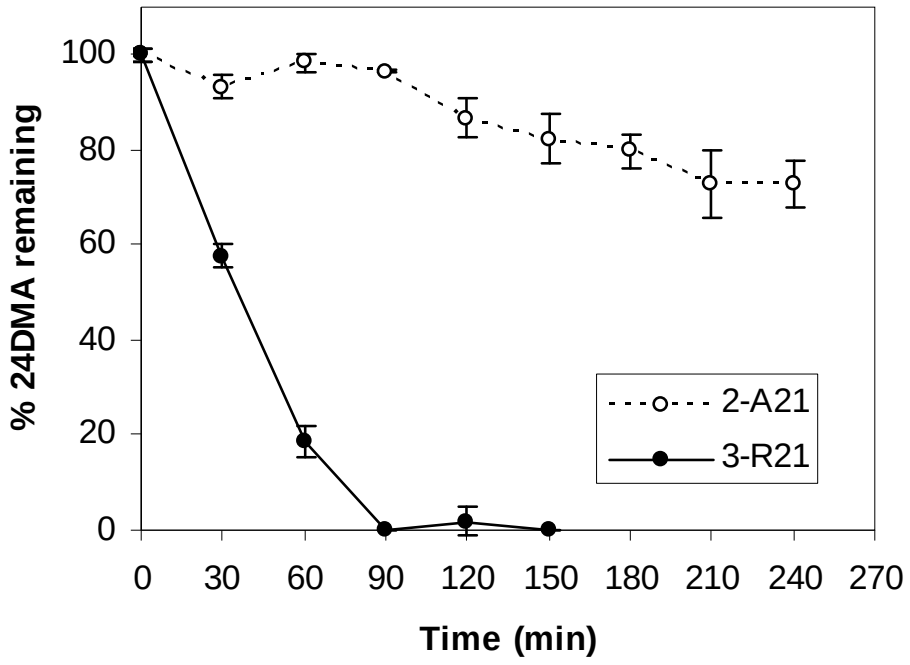


Figure 7.7 Percentage of 24DMA remaining with time when added to resting cell cultures of mutant 3-R21 and its parent 2-A21.

Table 7.4 Conversion rate of aniline, 24DMA, and 2IPA by *E. coli* JM109 expressing the AtdA mutants 2-A21 and 3-R21. Relative rates are calculated with respect to mutant 2-A21, the parent used in epPCR.

Mutant	Mutation	AN		24DMA		2IPA	
		Rate (nmol/min/mg protein)	Relative rate	Rate (nmol/min/mg protein)	Relative rate	Rate (nmol/min/mg protein)	Relative rate
2-A21	V205A/ I248L	9.3 ± 2.4	1.0	1.4 ± 0.1	1.0	1.6 ± 0.5	1.0
3-R21	V205A/ I248L/ S404C	27.7 ± 1.5	3.0	9.8 ± 0.4	7.0	2.2 ± 0.1	1.4

7.10 Structural analysis of mutation

The S404C mutation was mapped onto the homology model of AtdA3 to analyze the molecular basis of its effect. The S404C mutation is located on an α-helix close to the C-terminal of the protein, more than 8 Å away from the substrate binding pocket (Figure 7.8). Residue 404 is not exposed to the surface of the protein and the orientation of the S404 side chain is similar to that of C404 (Figure 7.9). The closest residues to residue 404 in both mutants are I376, H403, and E401. The serine to cysteine mutation shifts the residue 404 side chain slightly towards H403. The slightly nucleophilic C404 residue may be attracted to the positively charged H403, and being more hydrophobic than S404, may have enhanced hydrophobic interactions with I376 and F408. Both of these factors can help to stabilize the α-helix structure near the C-terminal of the protein and help to improve its overall activity.

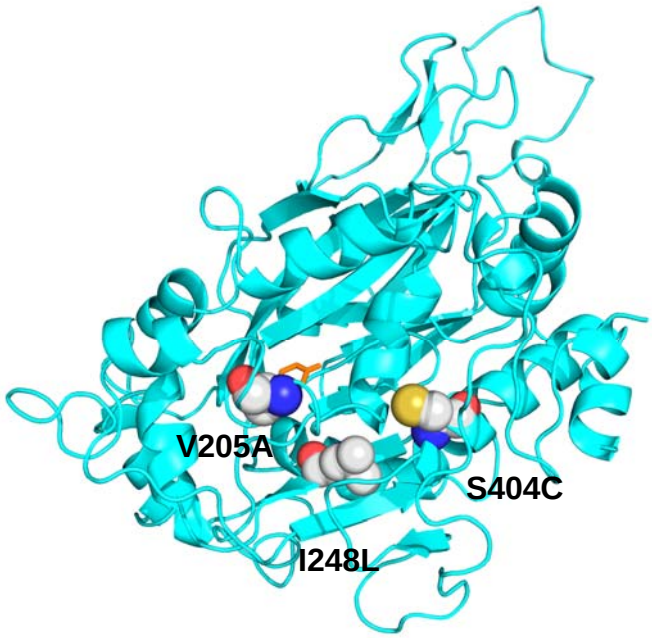


Figure 7.8 Location of V205A, I248L, and S404C mutations in the AtdA3 subunit. The orange molecule represents 24DMA bound in the substrate binding pocket of the subunit.

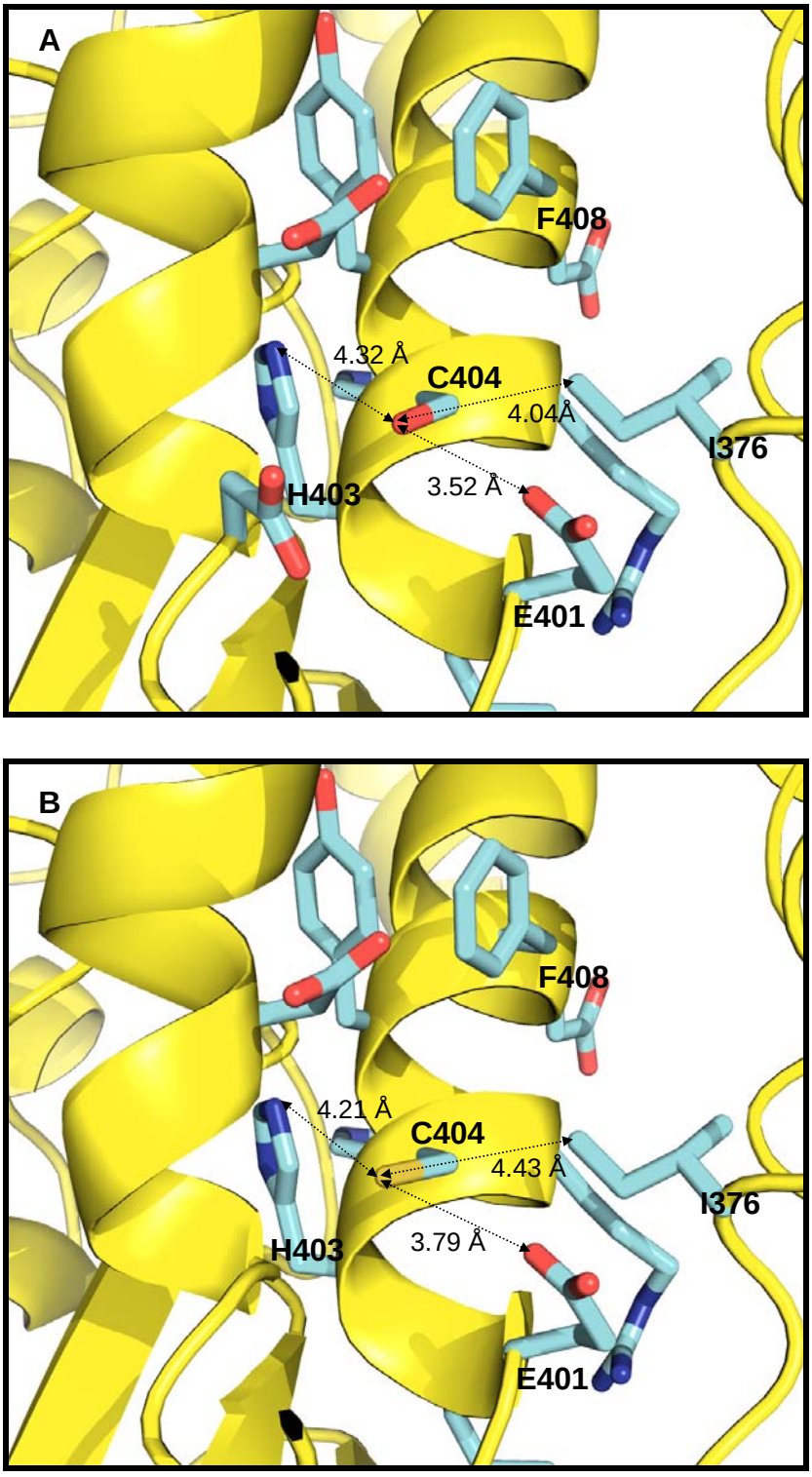


Figure7.9 Residue 404 and its neighboring residues in (A) mutant 2-A21 and (B) mutant 3-R21.

7.11 Summary

The objective of this chapter was to enhance the AtdA enzyme as a biocatalyst for bioremediation as well as potential carbazole denitrogenation applications by widening its substrate range and improving its activity. The mutant V205A (1-K31) isolated from the first round of saturation mutagenesis in Chapter 6, which has an enhanced substrate range, was used as the parent for further rounds of engineering. The target substrate used to screen for improved activity was 24DMA, while 2SBA was used as a target substrate to screen for novel activity. With the second round of saturation mutagenesis on active site residues followed by epPCR, the activities of AtdA for AN, 24DMA and 2IPA were improved over its parent, V205A. In particular, the 24DMA activity of the final mutant, V205A/I248L/S404C (3-R21), was improved by 3.5-fold over the WT AtdA enzyme.

Using a second saturation mutagenesis on the substrate binding pocket of the subunit AtdA3, the mutant V205A/I248L (2-A21), which has improved activity for 24DMA, was isolated. Whole cell activity assay revealed that the activity of the V205A/I248L mutant for AN, 24DMA and 2IPA were increased to 9.3, 1.7, and 1.6 nmol/min/mg protein respectively. The third round of saturation mutagenesis of the substrate binding pocket residues did not yield further improvement in activity or substrate range.

Subsequently, the mutant V205A/I248L was subjected to epPCR and screened for improvement in activity or substrate range using 24DMA and 2SBA respectively. Screening of about 2500 variants yielded a triple mutant, V205A/I248L/S404C (3-R21), with improved 24DMA activity. Whole cell activity assay revealed that the activity of

mutant V205A/I248L/S404C for AN, 24DMA and 2IPA were further increased to 27.7, 9.8, and 2.2 nmol/min/mg protein respectively compared to its parent V205A/I248L.

Overall, the activity of the parent, mutant V205A, for AN, 24DMA, and 2IPA was increased by 8.9, 98.0, and 2.0-fold respectively after one round of saturation mutagenesis followed by epPCR. The activity for AN was restored to the level of the WT AtdA, while the activity of 24DMA was increased by 3.5 fold. The activities of the mutants V205A, V205A/I248L, and V205A/I248L/S404C for AN, 24DMA and 2IPA are presented in Figure 7.10 together with that of the WT for a comparison of the improvements brought about by each round of mutagenesis.

Although the mutant V205A served as a platform for the stepwise evolution strategy to engineer AtdA for the carbazole denitrogenation process, its substrate range could not be increased further with saturation mutagenesis of the substrate binding pocket residues. Random mutagenesis of the AtdA3 subunit did not yield any mutant with a widened substrate range thus far but the library size (2500 variants) screened in this work has not comprehensively covered the sequence space, which requires the screening of about 24000 variants. Screening of more mutants from the epPCR library may yield a mutant with higher activity or even widened substrate range.

The unexplored subunits AtdA1 and A2 may have effects on the enzyme's substrate specificity and activity as well and their sequence space can be probed in a similar way as AtdA3 to widen the substrate range and activity of AtdA. Another possible strategy is to shuffle the *atdA* genes with homologous aniline dioxygenases such as those from *Frateuria sp* ANA-18 (Murakami, Hayashi et al., 2003), *Delftia acidovorans* strain 7N (Urata, Uchida et al., 2004), and *Delftia tsuruhatensis* AD9 (Liang, Takeo et al., 2005).

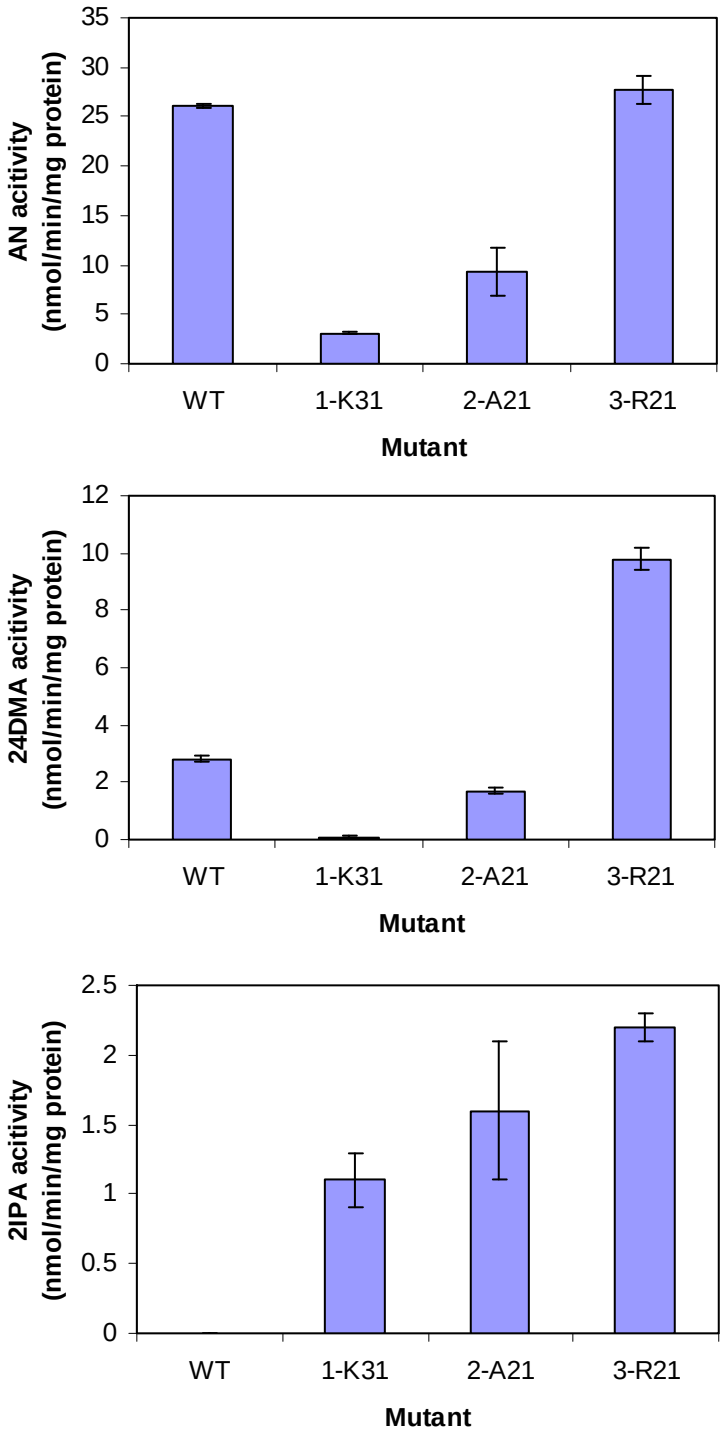


Figure 7.10 Activities of WT, 1-K31 (V205A), 2-A21 (V205A/I248L), and 3-R21 (V205A/I248L/S404C) for (A) AN, (B) 24DMA, and (C) 2IPA

7.12 References

Benedik, M. J., P. R. Gibbs, R. R. Riddle and R. C. Willson (1998). "Microbial denitrogenation of fossil fuels." Trends in Biotechnology **16** (9): 390-395.

Dinneen, G. U. and W. D. Bickel (1951). "Gum Formation in Shale-Oil Naphtha." Industrial & Engineering Chemistry Research **43** (7): 1604 - 1607.

Dong, X., S. Fushinobu, E. Fukuda, T. Terada, S. Nakamura, K. Shimizu, H. Nojiri, T. Omori, H. Shoun and T. Wakagi (2005). "Crystal structure of the terminal oxygenase component of cumene dioxygenase from *Pseudomonas fluorescens* IP01." J Bacteriol **187** (7): 2483-90.

Ford, C. D., S. A. Holmes, L. F. Thompson and D. R. Latham (1981). "Separation of nitrogen compound types from hydrotreated shale oil products by adsorption chromatography on basic and neutral alumina." Analytical Chemistry **53** (6): 831-836.

Furusawa, Y., V. Nagarajan, M. Tanokura, E. Masai, M. Fukuda and T. Senda (2004). "Crystal structure of the terminal oxygenase component of biphenyl dioxygenase derived from *Rhodococcus* sp. strain RHA1." J Mol Biol **342** (3): 1041-52.

Girgis, M. J. and B. C. Gates (1991). "Reactivities, Reaction Networks, and Kinetics in High-Pressure Catalytic Hydroprocessing." Industrial & Engineering Chemistry Research **30** (9): 2021-2058.

Grosser, R. J., D. Warshawsky and J. R. Vestal (1991). "Indigenous And Enhanced Mineralization Of Pyrene, Benzo[A]Pyrene, And Carbazole In Soils." Applied And Environmental Microbiology **57** (12): 3462-3469.

Karlsson, A., J. V. Parales, R. E. Parales, D. T. Gibson, H. Eklund and S. Ramaswamy (2003). "Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron." Science **299** (5609): 1039-42.

Katzer, J. R. and R. Sivasubramanian (1979). "Process and Catalyst Needs for Hydrodenitrogenation." Catalysis Reviews - Science and Engineering **20**: 155-208.

Kilbane, J. J., A. Daram, J. Abbasian and K. J. Kayser (2002). "Isolation and characterization of *Sphingomonas* sp GTIN11 capable of carbazole metabolism in petroleum." Biochemical And Biophysical Research Communications **297** (2): 242-248.

Kirimura, K., H. Nakagawa, K. Tsuji, K. Matsuda, R. Kurane and S. Usami (1999). "Selective and continuous degradation of carbazole contained in petroleum oil by resting cells of *Sphingomonas* sp CDH-7." Bioscience Biotechnology And Biochemistry **63** (9): 1563-1568.

Kobayashi, T., R. Kurane, K. Nakajima, Y. Nakamura, K. Kirimura and S. Usami (1995). "Isolation Of Bacteria Degrading Carbazole Under Microaerobic Conditions, Ie Nitrogen Gas Substituted Conditions." Bioscience Biotechnology And Biochemistry **59** (5): 932-933.

Kuchner, O. and F. H. Arnold (1997). "Directed evolution of enzyme catalysts." Trends Biotechnol **15** (12): 523-30.

Laredo, G. C., S. Leyva, R. Alvarez, M. T. Mares, J. Castillo and J. L. Cano (2002). "Nitrogen compounds characterization in atmospheric gas oil and light cycle oil from a blend of Mexican crudes." Fuel **81** (10): 1341-1350.

Laredo, G. C., A. Montesinos and J. A. De los Reyes (2004). "Inhibition effects observed between dibenzothiophene and carbazole during the hydrotreating process." Applied Catalysis a-General **265** (2): 171-183.

Liang, Q., M. Takeo, M. Chen, W. Zhang, Y. Xu and M. Lin (2005). "Chromosome-encoded gene cluster for the metabolic pathway that converts aniline to TCA-cycle intermediates in *Delftia tsuruhatensis* AD9." Microbiology **151** (Pt 10): 3435-46.

Morley, K. L. and R. J. Kazlauskas (2005). "Improving enzyme properties: when are closer mutations better?" Trends Biotechnol **23** (5): 231-7.

Murakami, S., T. Hayashi, T. Maeda, S. Takenaka and K. Aoki (2003). "Cloning and functional analysis of aniline dioxygenase gene cluster, from *Frateuria* species ANA-18, that metabolizes aniline via an *ortho*-cleavage pathway of catechol." Biosci Biotechnol Biochem **67** (11): 2351-8.

Mushrush, G. W., E. J. Beal, D. R. Hardy and J. M. Hughes (1999). "Nitrogen compound distribution in middle distillate fuels derived from petroleum, oil shale, and tar sand sources." Fuel Processing Technology **61** (3): 197-210.

Nagai, M. and T. Kabe (1983). "Selectivity of molybdenum catalyst in hydrodesulfurization, hydrodenitrogenation, and hydrodeoxygenation: Effect of additives on dibenzothiophene hydrodesulfurization." Journal of Catalysis **81** (2): 440-449.

Ouchiyama, N., Y. Zhang, T. Omori and T. Kodama (1993). "Biodegradation of Carbazole by *Pseudomonas* Sp CA06 and CA10." Bioscience Biotechnology and Biochemistry **57** (3): 455-460.

Sakamoto, T., J. M. Joern, A. Arisawa and F. H. Arnold (2001). "Laboratory evolution of toluene dioxygenase to accept 4-picoline as a substrate." Appl Environ Microbiol **67** (9): 3882-7.

Schneider, J., R. J. Grosser, K. Jayasimhulu, W. L. Xue, B. Kinkle and D. Warshawsky (2000). "Biodegradation of carbazole by *Ralstonia* sp RJGII.123 isolated from a hydrocarbon contaminated soil." Canadian Journal Of Microbiology **46** (3): 269-277.

Stratagene (2006). "GeneMorph II EZClone Domain Mutagenesis Kit Instruction Manual." Stratagene

Urata, M., E. Uchida, H. Nojiri, T. Omori, R. Obo, N. Miyaura and N. Ouchiyama (2004). "Genes involved in aniline degradation by *Delftia acidovorans* strain 7N and its distribution in the natural environment." Biosci Biotechnol Biochem **68** (12): 2457-65.

Wong, T. S., K. L. Tee, B. Hauer and U. Schwaneberg (2004). "Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution." Nucleic Acids Res **32** (3): e26.

Chapter 8
Summary, conclusion and future work

8.1 Summary

The aniline dioxygenase (AtdA) enzyme has potential uses in bioremediation of aromatic amines and biorefining such as the denitrogenation of carbazole. However, the lack of characterization of the enzyme has limited its development as a practical biocatalyst. The overall objective of this investigation was to first characterize the AtdA enzyme by determining its substrate specificity, and then probe for the molecular determinants of its substrate specificity as well as its activity. Using the insights gained from the characterization studies, further engineering work was conducted to improve the activity of AtdA as well as expand its substrate range for use in bioremediation and industrial applications (Figure 8.1).

This is the first study on the molecular determinants for substrate specificity of a five subunit Rieske-dioxygenase, AtdA, and it was shown that the α-subunit of the enzyme (AtdA3) indeed plays a key role in controlling the substrate specificity and activity of the enzyme. In the process, novel activity for 2IPA, a substrate not accepted by the wild type (WT) enzyme, was discovered with an introduction of a single V205A mutation. Using knowledge gained from these findings, the activity of AtdA for 2IPA and 24DMA, a carcinogenic aromatic pollutant, was enhanced by saturation and random mutagenesis.

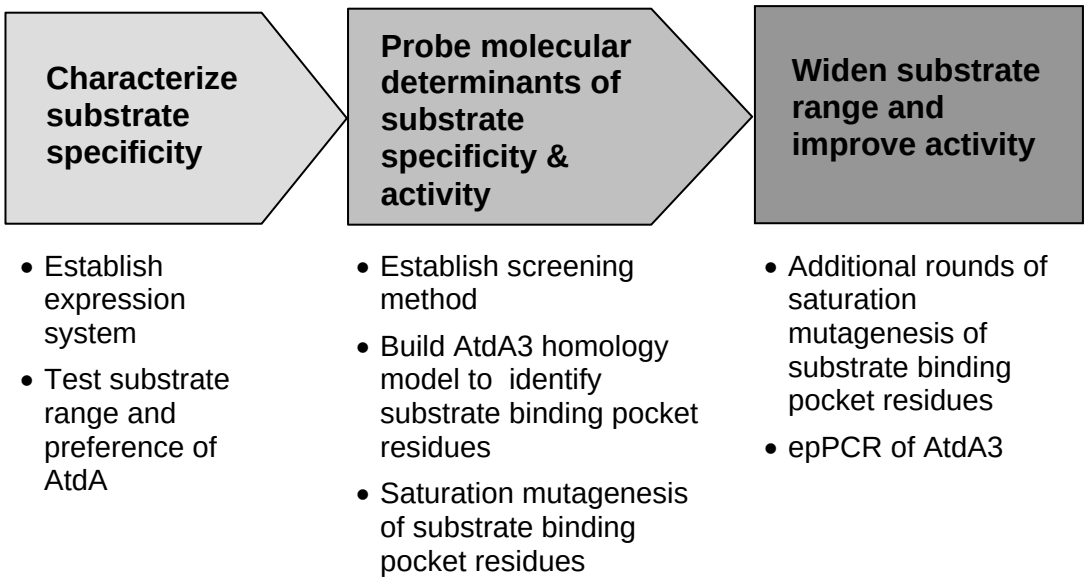


Figure 8.1 Schematic of the project objective and scope.

Chapter 3 reported the cloning of the *atdA* gene cluster into an expression vector pTrc99A and removal of the gene sequence overlaps between *atdA2*, *atdA3*, and *atdA4* to facilitate the characterization of AtdA's substrate specificity and engineering of the enzyme. Both the expression level and the activity of the AtdA enzyme were improved with the new plasmid construct.

As the screens or selections for activity are specific for each enzyme, the availability of a high-throughput screen or selection is the bottleneck for most directed evolution endeavors (Aharoni, Griffiths et al., 2005). The efficacy of several screening and selection methods for detecting AtdA activity was reported. The liquid phase Gibbs' reagent screening method, which detected the presence of catechol products from the hydroxylation of various aromatic amines by AtdA, was found to be sensitive enough to identify improved mutants from background signals and efficient enough to allow for screening of the large genetic libraries generated. Therefore, the method was chosen as

the screening method for AtdA engineering.

Characterizing the substrate specificity of AtdA and determining the subunits critical for dioxygenase activity are prerequisite to engineering the enzyme to enhance its substrate range and activity. Chapter 5 reported the activities of AtdA on a range of *ortho*-substituted aniline substrates, as well as the effect of methylaniline's alkyl sidechain ring on enzyme activity. To narrow down the potential target for AtdA engineering, a gene deletion assay was used to ascertain which subunits are critical for dioxygenase activity in AtdA. The inability of the AtdA enzyme to hydroxylate substrates with an *ortho*-sidechain larger than an ethyl group (2IPA, 2SBA, and 2TBA), provided evidence that the enzyme substrate specificity may be controlled by steric hindrance to the *ortho*-sidechain along the substrate channel or in the substrate binding pocket. For methylanilines, 2MA was the best substrate for AtdA. The gene deletion assay results showed that AtdA1 and A3 are critical for the activity of the enzyme and hence the substrate specificity and activity may be controlled by one or both of these subunits. Since equivalent subunits of AtdA3 (Rieske dioxygenase α-subunit) in other dioxygenases have shown to play a part in substrate specificity and activity of the enzymes (Tan and Cheong, 1994; Parales, Parales et al., 1998; Parales, Emig et al., 1998), the AtdA3 subunit was targeted for saturation mutagenesis studies to probe for the molecular determinants of the enzyme's substrate specificity and activity and o*rtho*-substituted anilines were used as target substrates as AtdA was most active against these isomers.

A homology model was constructed, as reported in Chapter 6, to identify the residues defining the substrate binding pocket of the α-subunit of the aniline

dioxygenase, AtdA3, and applied saturation mutagenesis to these residues to probe the molecular determinants of the activity and specificity of the enzyme. It was clearly demonstrated that the substrate specificity of the AtdA enzyme can be controlled by the AtdA3 subunit. The V205A mutation enables the mutant 1-K31 to dihydroxylate 2IPA, a substrate not accepted by the WT enzyme. It was also found that the I248L mutation enhances the activity of the enzyme for aniline and 24DMA, a carcinogenic pollutant for which no enzyme directly responsible for its biodegradation has been previously identified. Although the V2025A mutation caused the loss of activity for aniline and 24DMA, the primary goal of this chapter, which was to probe the molecular determinants of AtdA, was achieved. This finding will facilitate future engineering efforts of the enzyme for both bioremediation and industrial applications.

The next objective was to enhance the AtdA enzyme as a biocatalyst for bioremediation as well as potential carbazole denitrogenation applications by widening its substrate range and improving its activity. The mutant 1-K31 isolated from the first round of saturation mutagenesis reported in Chapter 6, which has an enhanced substrate range but decreased activity, was used as the parent for further rounds of engineering. A second round of saturation mutagenesis on active site residues followed by epPCR yielded the mutant 3-R21. Overall, this mutant had three mutations – V205A (carried over from the 1-K31 parent), I248L (from the second round of active site residue saturation mutagenesis) and S404C (from epPCR). The activities of 3-R21 for AN, 24DMA and 2IPA were improved by 8.9, 98.0, and 2.0-fold respectively over its parent 1-K31. In particular, the activity of the final mutant 3-R21 was improved by 3.5-fold over the WT AtdA enzyme. Whole cell activity assay revealed that the activity of 3-R21 for

AN, 24DMA and 2IPA were 27.7, 9.8, and 2.2 nmol/min/mg protein respectively. The activity for AN was restored to the level of the WT enzyme while the activity of 24DMA was increased by 3.5 fold relative to the WT.

8.2 Conclusion

In this work, the expression of AtdA was optimized; a high throughput assay for the screening of mutant libraries was developed; the substrate specificity of AtdA was investigated; and a homology model of the AtdA3 subunit was constructed. Using the results of these studies, the AtdA enzyme was put through two rounds of saturation mutagenesis followed by one round of random mutagenesis to give the final mutant 3-R21, which demonstrated enhanced substrate range by accepting 2IPA as a substrate and had a 3-fold improvement in 24DMA activity relative to the WT enzyme.

This study has made strong contributions to the knowledge of the structural determinants that controls the substrate specificity of AtdA. It has also enhanced the substrate range and activity of AtdA, rendering it a potent enzyme for bioremediation aromatic amines. The 3-R21 mutant created is also a useful platform in the stepwise evolution strategy to engineer AtdA for carbazole denitrogenation.

8.3 Future work

Although the mutant 3-R21 served as a platform for the stepwise evolution strategy proposed to engineer AtdA for the carbazole denitrogenation process, its substrate range could not be increased further with saturation mutagenesis of the substrate binding pocket residues. Random mutagenesis of the AtdA3 subunit did not yield any mutant with a

widened substrate range thus far but the library size (2500 variants) screened in this work has not comprehensively covered the sequence space, which requires the screening of about 24000 variants. Screening of more mutants from the epPCR library may yield a mutant with higher activity or even widened substrate range.

AtdA1 has a 25.8 % homology to glutamine synthetases from *Salmonella typhimurium* (GS) (Yamashita, Almassy et al., 1989), and the important motifs of GS are well conserved. AtdA1 also has a 62.1 % protein sequence identity with TdnQ of the aniline dioxygenase from *Pseudomonas putida* UCC22, which had no glutamine synthetase activity when expressed in *E. coli* cells (Fukumori and Saint, 1997). This suggests that AtdA1 is unlikely to be involved in the recovery of nitrogen for biosynthesis reactions. AtdA2 exhibits homology to the class-I glutamine amidotransferase (GATs) domain in GMP synthetase (Tesmer, Klem et al., 1996). It has been postulated that since GS and GAT are involved in the addition of an amino group to glutamate and its release from glutamine respectively, AtdA1 and AtdA2 may be involved in the recognition and release of aniline amino groups (Takeo, Fujii et al., 1998a). Hence a similar engineering approach on the unexplored sequence space of AtdA1 and AtdA2 may offer useful insights on the substrate specificity and activity of the enzyme and improve them as well.

Another possible strategy is to shuffle the *atdA* genes with homologous aniline dioxygenases such as those from *Frateuria sp* ANA-18 (Murakami, Hayashi et al., 2003), *Delftia acidovorans* strain 7N (Urata, Uchida et al., 2004), and *Delftia tsuruhatensis* AD9 (Liang, Takeo et al., 2005).

For a more in depth understanding of the structural determinants of the substrate

specificity and activity of AtdA, an activity assay should be conducted with the purified subunits instead of whole cell assays. This assay would shed more light on the whether each mutation affects the catalytic rate or binding affinity of the enzyme.

In summary, the primary goals of the project, which were to: (1) characterize this enzyme; and (2) improve its activity and widen its substrate specificity thereby increasing its usefulness in the bioremediation and industrial applications, have been achieved. This investigation is the first study on the molecular determinants for substrate specificity of a five subunit Rieske-dioxygenase enzyme, AtdA and it has been clearly demonstrated that AtdA3 plays a key role in determining substrate specificity and activity of the AtdA enzyme. An enhanced biocatalyst for the bioremediation of 24DMA and 2IPA was successfully engineered and the mutant 3-R21 will serve as a valuable platform for future biomolecular engineering of AtdA for the carbazole denitrogenation process via a stepwise evolution strategy.

8.3 References

Aharoni, A., A. D. Griffiths and D. S. Tawfik (2005). "High-throughput screens and selections of enzyme-encoding genes." Curr Opin Chem Biol **9** (2): 210-6.

Fukumori, F. and C. P. Saint (1997). "Nucleotide sequences and regulational analysis of genes involved in conversion of aniline to catechol in *Pseudomonas putida* UCC22(pTDN1)." J Bacteriol **179** (2): 399-408.

Liang, Q., M. Takeo, M. Chen, W. Zhang, Y. Xu and M. Lin (2005). "Chromosome-encoded gene cluster for the metabolic pathway that converts aniline to TCA-cycle intermediates in *Delftia tsuruhatensis* AD9." Microbiology **151** (Pt 10): 3435-46.

Murakami, S., T. Hayashi, T. Maeda, S. Takenaka and K. Aoki (2003). "Cloning and functional analysis of aniline dioxygenase gene cluster, from *Frateuria* species ANA-18, that metabolizes aniline via an *ortho*-cleavage pathway of catechol." Biosci Biotechnol Biochem **67** (11): 2351-8.

Parales, J. V., R. E. Parales, S. M. Resnick and D. T. Gibson (1998). "Enzyme specificity of 2-nitrotoluene 2,3-dioxygenase from *Pseudomonas* sp. strain JS42 is determined by the C-terminal region of the alpha subunit of the oxygenase component." J Bacteriol **180** (5): 1194-9.

Parales, R. E., M. D. Emig, N. A. Lynch and D. T. Gibson (1998). "Substrate specificities of hybrid naphthalene and 2,4-dinitrotoluene dioxygenase enzyme systems." J Bacteriol **180** (9): 2337-44.

Takeo, M., T. Fujii and Y. Maeda (1998a). "Sequence analysis of the genes encoding a multicomponent dioxygenase involved in oxidation of aniline and *o*-toluidine in *Acinetobacter* sp. strain YAA." J Ferment Bioeng **85** (1): 17-24.

Tan, H. M. and C. M. Cheong (1994). "Substitution of the ISP alpha subunit of biphenyl dioxygenase from *Pseudomonas* results in a modification of the enzyme activity." Biochem Biophys Res Commun **204** (2): 912-7.

Tesmer, J. J., T. J. Klem, M. L. Deras, V. J. Davisson and J. L. Smith (1996). "The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families." Nat Struct Biol **3** (1): 74-86.

Urata, M., E. Uchida, H. Nojiri, T. Omori, R. Obo, N. Miyaura and N. Ouchiyama (2004). "Genes involved in aniline degradation by *Delftia acidovorans* strain 7N and its distribution in the natural environment." Biosci Biotechnol Biochem **68** (12): 2457-65.

Yamashita, M. M., R. J. Almassy, C. A. Janson, D. Cascio and D. Eisenberg (1989). "Refined atomic model of glutamine synthetase at 3.5 A resolution." J Biol Chem **264** (30): 17681-90.

Apendix A

Sequences of Plasmid Constructs

A.1 pTA2-3 sequence

FEATURES	Location/Qualifiers
atdA1	327..1826
atdA2	1845..2570
atdA3	2591..3868
atdA4	3889..4482
atdA5	4496..5503
Amp resistance	6050..6908
PBR322 origin	6975..7671
lacI	8259..9338
trc promoter	193..222

```
   1 gtttgacagc ttatcatcga ctgcacggtg caccaatgct tctggcgtca ggcagccatc
  61 ggaagctgtg gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc
 121 gcactcccgt tctggataat gtttttgcg ccgacatcat aacggttctg gcaaatattc
 181 tgaaatgagc tgttgacaat taatcatccg gctcgtataa tgtgtggaat tgtgagcgga
 241 taacaatttc acacaggaaa cagaccatgc cattcgagct cggtacccgg ggatcctcta
 301 gagtcgacta gacaaaggaa tgtcccatga gtgagaaatt agattttata acgaaaaata
 361 atctttggac agataagcag cgagatgcag ccgacaaagt tctcgcagaa attgattctt
 421 tagggcttga gatgattcgg ctctcctggg ctgatcagta tggtctcttg cgaggtaagg
 481 cgctgtcggt agcagccctc aaggcggcat tctcagaagg gtcagaagtt acgatggcac
 541 ctttctcttt caacctagtc agcgaatggg ttttcaaccc atttactgct ggtggtggct
 601 tcggaattga tgagtttgat gagttgggtg gtgtcccgag tgtggtgatg gttccagatc
 661 ctacgacgtt caaggtctta ccttgggcag ataaaaccgg ctggatgctg gcagatctgc
 721 attggaaatc tggtgaacca ttcccattat gtccccgcgg tatcatgaag aaggctgtca
 781 aatcgttaag cgatgaaggt tacttattta aatgcggtat tgagcttgaa tggtacttga
 841 cgaaaattgt tgatcgctca ctttctccag agagtttagg tgcgccaggt gtacagcctg
 901 atgccattca agttcaaccc gtggcgcaag ggtactccta tctacttgaa tatcacttag
 961 atcaggtgga cgatattatg tccaaggttc gaaaaggtct tctcgagctc aatctgcctt
1021 tgcgctcaat agaagatgag ttggcaccaa gccaaatgga aaccacgttt gatgtaatgg
1081 aaggtttaga agcagccgat gcagcgctac ttataaaatc ggccatcaaa caaatatgtt
1141 cacgacatgg ttatcacgca acatttatgt gcaagccggc aattaacggg ttctctgttg
1201 cttcaggctg gcatatgcat caatcactag tggataaaga tacccgaaag aatcttttta
1261 taccctcaga aggggaagtg gtatccccgc taggtcgagc ttatgctggt ggattacttg
1321 caaatggtag tgccgcctcg agtttcacaa caccaactgt gaatgggtat cgaagacgtc
1381 agccgcactc gcttgcacca gaccgaagag cttgggcgaa agaaaacaag gcagcgatgg
1441 tccgtgtaat ttctgcaaca ggcgatccgg ctagccggat cgaaaatcgt attggtgagc
1501 ccggcgccaa cccttattta tatatggcat cacaaattgt ctctgggctt gatggcatta
1561 aaatcaaaag ggatcccggc gggttgcaag gggctcctta tggtgcacaa gtaccaatgc
1621 tgccaacggc tttggctgag gctctggatg ctcttgagca cgattcggag ttgtttagaa
1681 gctgctttgg cgacaccttt attaaatatt ggctgcaatt aagaagatcc gagtgggcaa
1741 gatttctcga tgctgaaggt gctgaggctg ctgagcctac aggtgctgtc acgcagtggg
1801 aacaaaaaga atacttcaac ttactgtgat tttcagaggt caatatgtct aaacgctttg
1861 cattattgtg gtgctctgaa gaagagcgct ttgattatcg agaagaaatg gtaaatgcct
1921 ttaaaactga aaactccgac tgggaagtta taagtgcatt cacagactta aataaaatta
```

```
1981 tcgataatta cgatggcttt gttatcagtg gtagtgagta ctcagttaat gctgataaag
2041 aaaagttttc tggcttattt gaatttattc gagcggtcca taagaaagaa aaaccaattg
2101 ttggcatatg cttcggttgt cagtcccttg ctgtcgcact tggcggagag gtgggtttga
2161 accctagtcg tgagtttagg tttggaactg atgagctcac gtttcaaaat ggacttaaca
2221 aacatgttgg taccagtgaa gagcgagtga ggcttattga aagccatgga gaatgcgtca
2281 ttagacgccc acttggatct acattactcg cacgttctga ttcaactgct gtagaaattt
2341 ttgcggtggg gccgtatgca gtcggtatcc aaggacatcc agagatcagt aaaaaaaccc
2401 tggagcaaga ctttctacgg gttcatctcg aagatggtaa tttgcaagaa gatgaggtac
2461 gcaggtttca tgctgagctg agtggttatc agcctcctca agcgatacgt caattagtga
2521 aagcgactct acacaagcaa attaattttc agaacttggt gggtgatgta taagaattca
2581 ggaacagacc atgaaaacca taaatcaact aattcagtcc gggcgcgtac accgtaaggt
2641 ctatacagaa gcatctattt ttcaagcgga aatggacaaa atatttcaag cgaactgggt
2701 gtttctctta catgcaagcc aaattcctaa acttgatgat tatcaaacgg ttcggatggg
2761 tgggcgacca ctgatcgttg taagaaaagg agatgatgag tttcaggcat tgcttaatcg
2821 ttgtccgcat cgtggggcaa aggtttgtcg aaatgattcg ggtaactcca agacatttac
2881 ttgcccttac catggttgga aattcaggaa ctcaggcaaa gcttttgtta ttcctggtgc
2941 gaatgcctat ggagagggtt tcgataaaga caatttctcc atgacggcaa ttccacgggt
3001 ggaaagctat cgaggttttg tctttgctac cagtaatgag aacgctgttt cgttagagga
3061 gcatcttggt agtgcgcgcc agtatattga tgaatggtta gcccaccagg gtggtgagat
3121 taaagtatcg aagtctgttc aacgttatga aataaaatgt aattggaagc tggtgtttga
3181 taatgctggt gatggctacc atgtcccttt ttcccatcag tcattgctac aaatgactac
3241 gcttcgatat ggcggcgggg atatacagta tttcggtaat gccgatgaga caggtatggg
3301 cctttatgct ttgggtaatg gtcactcagt gatagatcag cgcccggaaa tgcataaaga
3361 gtcagggtgg gatcagcagc gaccacagcc tggccgcgaa agctatgaaa cacacgttcg
3421 taataatagt agccagccag caagagattt agagcgagcg gttggtgctg gaatgaatct
3481 taatattttt ccaaacttac tgttaattgg caaccaaata caggttattg atcctatttc
3541 tgttaatgaa acagttctgc attggcatgc aaccttgctt gctggtgata atgaagagct
3601 caatgcaatt cggatgcgca ctcaggagga ttttccaatt atgggtgagg tggatgacgt
3661 ggccaacttt gaatcatgtc aagaaggact cgagaccatg ccggaaatcg aatggatcga
3721 ttttagcagg catatgaatg aaggagaaaa tgcttgctac caagatgtta tacaacataa
3781 accaacatct gagatccata gtcgccatta ttttgatacc tggctacagc taatgtctgc
3841 agtaaacaaa gagaatcagt ctgaggttta aaccggttga ggtacgccat gaataataat
3901 aaagatctta gtataaaaat tgttgatcct agcgctgtta acctcgctta ttatcaagag
3961 attaagcagt actctgatta tttttggaat ctttctaacc ttggtgagcc tgcgcttgat
4021 cacaagataa atatgtttct gactaaagaa gctcggcttt tggatcagca gtgttttgat
4081 gagtggttaa cactgtttct ggaggatggc tgttactgga ttcctggcag tatgccggcg
4141 gcgtcacctg ccagcgaggc cacatatgag tttcatgata tacgcagact gaaagatcga
4201 attgtaaggc tgcaaacggg ttttgcctat tcgcagatac ctgtttcaaa aactaatcgc
4261 atccttgggg cgccagaagt atgggcagtg ccggggtcaa gcgaggggtt tttggttaga
4321 acaagcttta ttgtgtttga aagccgagat ggcaagtctc aagttttaag tggttggtat
4381 ggttatgtaa ttattaagga taatgatgag ttgaagataa agatgaaaca gataaattta
4441 aacgactgtc tttcgccgca aggcaataat tcatttttc tatagttggt ctaaaatgaa
4501 tacattaaaa tttcgagtta ttgataagat agcggaaacg aaggagtcgt tttcatttgt
4561 tctgaagccg ttggacggtg tcttggctga gcactcccct ggcaagtatt taccaattaa
4621 aatccgaact gaaaaaggtc ttttatttcg gtcttattcc ctatcttcgt cagcttcggc
4681 caatgaggac tttaaaatca cggtaaagag agaaagggga ggcagagggt cgaactggtt
4741 gtgtgacaat gtaaaggttg gcgactttat cgagacactc ccccctgctg gcagtttcca
4801 cccgcagaac tgggatcgag attttgtttt ttttgcagga ggtagtggta taactcctgt
4861 aatatccatt ataaaaacag cgctaaatag acacaaaaat aggattaagt tgttttatgc
4921 taactcctct gaaagttcta taatatttca taaagagtta aaagatctat gtttacaatt
4981 cccggatcgg cttgatatac aattctggtt agacgatgaa aaaggtattc caacctctat
5041 cgcgtttgag caatatattg atgatgctct agaggttgaa tatttttat gtggacctgc
5101 ccccttcatg ggaggtgtag agaatttttt gatcgagtct aaagtcccac ctggactgat
5161 aaccaaggag tcttttgctg ggagtgtttc tgatgataat ggcgatacag ttgaaagctc
5221 agctgaaaaa gatgtcactg taaactttat gctaaatggc attaagaaca gcgttatgtg
5281 ctccgaagat gattttattt taaacgagat aataaaagct ggaattaatg ttcctagttc
5341 gtgctgtgct ggtaattgtg ggtcttgcat gtgtctgctg gtgagtggag atgtaattct
5401 agaaagcaat actgttttgg acgcttctga tgaggaagac ggctggatct tggcgtgtcg
5461 atctaaaccg agatcgaaaa atatagaaat atccttcgac caataagtcg acctgcaggc
5521 atgcaagctt ggctgttttg gcggatgaga gaagattttc agcctgatac agattaaatc
5581 agaacgcaga agcggtctga taaaacagaa tttgcctggc ggcagtagcg cggtggtccc
5641 acctgacccc atgccgaact cagaagtgaa acgccgtagc gccgatggta gtgtggggtc
5701 tccccatgcg agagtaggga actgccaggc atcaaataaa acgaaaggct cagtcgaaag
```

```
5761 actgggcctt tcgttttatc tgttgtttgt cggtgaacgc tctcctgagt aggacaaatc
5821 cgccgggagc ggatttgaac gttgcgaagc aacggcccgg agggtggcgg gcaggacgcc
5881 cgccataaac tgccaggcat caaattaagc agaaggccat cctgacggat ggccttttg
5941 cgtttctaca aactcttttt gtttattttt ctaaatacat tcaaatatgt atccgctcat
6001 gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca
6061 acatttccgt gtcgccctta ttccctttt tgcggcattt tgccttcctg tttttgctca
6121 cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta
6181 catcgaactg gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt
6241 tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc gtgttgacgc
6301 cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc
6361 accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc
6421 cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa
6481 ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga
6541 accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc ctacagcaat
6601 ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca
6661 attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc
6721 ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat
6781 tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag
6841 tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa
6901 gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca
6961 ttttaattt aaaaggatct aggtgaagat ccttttgat aatctcatga ccaaaatccc
7021 ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca aaggatcttc
7081 ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc
7141 agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt
7201 cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag gccaccactt
7261 caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc
7321 tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa
7381 ggcgcagcgg tcgggctgaa cgggggttc gtgcacacag cccagcttgg agcgaacgac
7441 ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc ttcccgaagg
7501 gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga
7561 gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact
7621 tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa
7681 cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc
7741 gttatcccct gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg
7801 ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgcctgat
7861 gcggtatttt ctccttacgc atctgtgcgg tatttcacac cgcatatggt gcactctcag
7921 tacaatctgc tctgatgccg catagttaag ccagtataca ctccgctatc gctacgtgac
7981 tgggtcatgg ctgcgccccg acacccgcca acacccgctg acgcgccctg acgggcttgt
8041 ctgctcccgg catccgctta cagacaagct gtgaccgtct ccgggagctg catgtgtcag
8101 aggttttcac cgtcatcacc gaaacgcgcg aggcagcaga tcaattcgcg cgcgaaggcg
8161 aagcggcatg catttacgtt gacaccatcg aatggtgcaa aacctttcgc ggtatggcat
8221 gatagcgccc ggaagagagt caattcaggg tggtgaatgt gaaaccagta acgttatacg
8281 atgtcgcaga gtatgccggt gtctcttatc agaccgtttc ccgcgtggtg aaccaggcca
8341 gccacgtttc tgcgaaaacg cgggaaaaag tggaagcggc gatggcggag ctgaattaca
8401 ttcccaaccg cgtggcacaa caactggcgg gcaaacagtc gttgctgatt ggcgttgcca
8461 cctccagtct ggccctgcac gcgccgtcgc aaattgtcgc ggcgattaaa tctcgcgccg
8521 atcaactggg tgccagcgtg gtggtgtcga tggtagaacg aagcggcgtc gaagcctgta
8581 aagcggcggt gcacaatctt ctcgcgcaac gcgtcagtgg gctgatcatt aactatccgc
8641 tggatgacca ggatgccatt gctgtggaag ctgcctgcac taatgttccg gcgttatttc
8701 ttgatgtctc tgaccagaca cccatcaaca gtattatttt ctcccatgaa gacggtacgc
8761 gactgggcgt ggagcatctg gtcgcattgg gtcaccagca aatcgcgctg ttagcgggcc
8821 cattaagttc tgtctcggcg cgtctgcgtc tggctggctg gcataaatat ctcactcgca
8881 atcaaattca gccgatagcg gaacgggaag gcgactggag tgccatgtcc ggttttcaac
8941 aaaccatgca aatgctgaat gagggcatcg ttcccactgc gatgctggtt gccaacgatc
9001 agatggcgct gggcgcaatg cgcgccatta ccgagtccgg gctgcgcgtt ggtgcggata
9061 tctcggtagt gggatacgac gataccgaag acagctcatg ttatatcccg ccgttaacca
9121 ccatcaaaca ggattttcgc ctgctggggc aaaccagcgt ggaccgcttg ctgcaactct
9181 ctcagggcca ggcggtgaag ggcaatcagc tgttgcccgt ctcactggtg aaaagaaaaa
9241 ccaccctggc gcccaatacg caaaccgcct ctccccgcgc gttggccgat tcattaatgc
9301 agctggcacg acaggtttcc cgactggaaa gcgggcagtg agcgcaacgc aattaatgtg
9361 agttagcgcg aattgatctg
```

//

A.2 pACYC A1A2 sequence

```
FEATURES             Location/Qualifiers
     lacI            complement(4903..5982)

     Cm resistance   complement(2846..3502)

     P15A origin     3864..4776

     T7 promoter 1   6106..6122

     T7 promoter 2   1699..1715

     atdA1             71..1618

     atdA2           1785..2543

        1 ggggaattgt gagcggataa caattcccct gtagaaataa ttttgtttaa ctttaataag
       61 gagatatacc atgggcagca gccatcacca tcatcaccac agccaggatc cgaattcgat
      121 gagtgagaaa ttagatttta taacgaaaaa taatctttgg acagataagc agcgagatgc
      181 agccgacaaa gttctcgcag aaattgattc tttagggctt gagatgattc ggctctcctg
      241 ggctgatcag tatggtctct tgcgaggtaa ggcgctgtcg gtagcagccc tcaaggcggc
      301 attctcagaa gggtcagaag ttacgatggc acctttctct ttcaacctag tcagcgaatg
      361 ggttttcaac ccatttactg ctggtggtgg cttcggaatt gatgagtttg atgagttggg
      421 tggtgtcccg agtgtggtga tggttccaga tcctacgacg ttcaaggtct taccttgggc
      481 agataaaacc ggctggatgc tggcagatct gcattggaaa tctggtgaac cattcccatt
      541 atgtccccgc ggtatcatga agaaggctgt caaatcgtta agcgatgaag gttacttatt
      601 taaatgcggt attgagcttg aatggtactt gacgaaaatt gttgatcgct cactttctcc
      661 agagagttta ggtgcgccag gtgtacagcc tgatgccatt caagttcaac ccgtggcgca
      721 agggtactcc tatctacttg aatatcactt agatcaggtg gacgatatta tgtccaaggt
      781 tcgaaaaggt cttctcgagc tcaatctgcc tttgcgctca atagaagatg agttggcacc
      841 aagccaaatg gaaaccacgt ttgatgtaat ggaaggttta gaagcagccg atgcagcgct
      901 acttataaaa tcggccatca aacaaatatg ttcacgacat ggttatcacg caacatttat
      961 gtgcaagccg gcaattaacg ggttctctgt tgcttcaggc tggcatatgc atcaatcact
     1021 agtggataaa gatacccgaa agaatctttt tataccctca gaaggggaag tggtatcccc
     1081 gctaggtcga gcttatgctg gtggattact tgcaaatggt agtgccgcct cgagtttcac
     1141 aacaccaact gtgaatgggt atcgaagacg tcagccgcac tcgcttgcac cagaccgaag
     1201 agcttgggcg aaagaaaaca aggcagcgat ggtccgtgta atttctgcaa caggcgatcc
     1261 ggctagccgg atcgaaaatc gtattggtga gcccggcgcc aacccttatt tatatatggc
     1321 atcacaaatt gtctctgggc ttgatggcat taaaatcaaa agggatcccg gcgggttgca
     1381 aggggctcct tatggtgcac aagtaccaat gctgccaacg gctttggctg aggctctgga
     1441 tgctcttgag cacgattcgg agttgtttag aagctgcttt ggcgacacct ttattaaata
     1501 ttggctgcaa ttaagaagat ccgagtgggc aagatttctc gatgctgaag gtgctgaggc
     1561 tgctgagcct acaggtgctg tcacgcagtg ggaacaaaaa gaatacttca acttactgtg
     1621 agtcgacaag cttgcggccg cataatgctt aagtcgaaca gaaagtaatc gtattgtaca
     1681 cggccgcata atcgaaatta atacgactca ctatagggga attgtgagcg gataacaatt
     1741 ccccatctta gtatattagt taagtataag aaggagatat acatatggca gatctcaatt
     1801 ggatatcggc cggccacatg tctaaacgct ttgcattatt gtggtgctct gaagaagagc
     1861 gctttgatta tcgagaagaa atggtaaatg cctttaaaac tgaaaactcc gactgggaag
     1921 ttataagtgc attcacagac ttaaataaaa ttatcgataa ttacgatggc tttgttatca
     1981 gtggtagtga gtactcagtt aatgctgata aagaaaagtt ttctggctta tttgaattta
     2041 ttcgagcggt ccataagaaa gaaaaaccaa ttgttggcat atgcttcggt tgtcagtccc
     2101 ttgctgtcgc acttggcgga gaggtgggtt tgaaccctag tcgtgagttt aggtttggaa
     2161 ctgatgagct cacgtttcaa aatggactta acaaacatgt tggtaccagt gaagagcgag
     2221 tgaggcttat tgaaagccat ggagaatgcg tcattagacg cccacttgga tctacattac
     2281 tcgcacgttc tgattcaact gctgtagaaa tttttgcggt ggggccgtat gcagtcggta
     2341 tccaaggaca tccagagatc agtaaaaaaa ccctggagca agactttcta cgggttcatc
     2401 tcgaagatgg taatttgcaa gaagatgagg tacgcaggtt tcatgctgag ctgagtggtt
     2461 atcagcctcc tcaagcgata cgtcaattag tgaaagcgac tctacacaag caaattaatt
     2521 ttcagaactt ggtgggtgat gtataaccta ggctgctgcc accgctgagc aataactagc
     2581 ataacccctt ggggcctcta aacgggtctt gaggggtttt ttgctgaaac ctcaggcatt
     2641 tgagaagcac acggtcacac tgcttccggt agtcaataaa ccggtaaacc agcaatagac
```

```
2701 ataagcggct atttaacgac cctgccctga accgacgacc gggtcgaatt tgctttcgaa
2761 tttctgccat tcatccgctt attatcactt attcaggcgt agcaccaggc gtttaagggc
2821 accaataact gccttaaaaa aattacgccc cgccctgcca ctcatcgcag tactgttgta
2881 attcattaag cattctgccg acatggaagc catcacagac ggcatgatga acctgaatcg
2941 ccagcggcat cagcaccttg tcgccttgcg tataatattt gcccatagtg aaaacggggg
3001 cgaagaagtt gtccatattg gccacgttta aatcaaaact ggtgaaactc acccagggat
3061 tggctgagac gaaaaacata ttctcaataa acccttttagg gaaataggcc aggttttcac
3121 cgtaacacgc cacatcttgc gaatatatgt gtagaaactg ccggaaatcg tcgtggtatt
3181 cactccagag cgatgaaaac gtttcagttt gctcatggaa aacggtgtaa caagggtgaa
3241 cactatccca tatcaccagc tcaccgtctt tcattgccat acggaactcc ggatgagcat
3301 tcatcaggcg ggcaagaatg tgaataaagg ccggataaaa cttgtgctta tttttcttta
3361 cggtctttaa aaaggccgta atatccagct gaacggtctg gttataggta cattgagcaa
3421 ctgactgaaa tgcctcaaaa tgttctttac gatgccattg ggatatatca acggtggtat
3481 atccagtgat ttttttctcc attttagctt ccttagctcc tgaaaatctc gataactcaa
3541 aaaatacgcc cggtagtgat cttatttcat tatggtgaaa gttggaacct cttacgtgcc
3601 gatcaacgtc tcattttcgc caaaagttgg cccagggctt cccggtatca acagggacac
3661 caggatttat ttattctgcg aagtgatctt ccgtcacagg tatttattcg gcgcaaagtg
3721 cgtcgggtga tgctgccaac ttactgattt agtgtatgat ggtgtttttg aggtgctcca
3781 gtggcttctg tttctatcag ctgtccctcc tgttcagcta ctgacggggt ggtgcgtaac
3841 ggcaaaagca ccgccggaca tcagcgctag cggagtgtat actggcttac tatgttggca
3901 ctgatgaggg tgtcagtgaa gtgcttcatg tggcaggaga aaaaaggctg caccggtgcg
3961 tcagcagaat atgtgataca ggatatattc cgcttcctcg ctcactgact cgctacgctc
4021 ggtcgttcga ctgcggcgag cggaaatggc ttacgaacgg ggcggagatt tcctggaaga
4081 tgccaggaag atacttaaca gggaagtgag agggccgcgg caaagccgtt tttccatagg
4141 ctccgccccc ctgacaagca tcacgaaatc tgacgctcaa atcagtggtg gcgaaacccg
4201 acaggactat aaagatacca ggcgtttccc ctggcggctc cctcgtgcgc tctcctgttc
4261 ctgcctttcg gtttaccggt gtcattccgc tgttatggcc gcgtttgtct cattccacgc
4321 ctgacactca gttccgggta ggcagttcgc tccaagctgg actgtatgca cgaaccccc
4381 gttcagtccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggaaaga
4441 catgcaaaag caccactggc agcagccact ggtaattgat ttagaggagt tagtcttgaa
4501 gtcatgcgcc ggttaaggct aaactgaaag gacaagtttt ggtgactgcg ctcctccaag
4561 ccagttacct cggttcaaag agttggtagc tcagagaacc ttcgaaaaac cgccctgcaa
4621 ggcggttttt tcgttttcag agcaagagat tacgcgcaga ccaaaacgat ctcaagaaga
4681 tcatcttatt aatcagataa aatatttcta gatttcagtg caatttatct cttcaaatgt
4741 agcacctgaa gtcagcccca tacgatataa gttgtaattc tcatgttagt catgccccgc
4801 gcccaccgga aggagctgac tgggttgaag gctctcaagg gcatcggtcg agatcccggt
4861 gcctaatgag tgagctaact tacattaatt gcgttgcgct cactgcccgc tttccagtcg
4921 ggaaacctgt cgtgccagct gcattaatga atcggccaac gcgcggggag aggcggtttg
4981 cgtattgggc gccagggtgg tttttctttt caccagtgag acgggcaaca gctgattgcc
5041 cttcaccgcc tggccctgag agagttgcag caagcggtcc acgctggttt gccccagcag
5101 gcgaaaatcc tgtttgatgg tggttaacgg cgggatataa catgagctgt cttcggtatc
5161 gtcgtatccc actaccgaga tgtccgcacc aacgcgcagc ccggactcgg taatggcgcg
5221 cattgcgccc agcgccatct gatcgttggc aaccagcatc gcagtgggaa cgatgccctc
5281 attcagcatt tgcatggttt gttgaaaacc ggacatggca ctccagtcgc cttcccgttc
5341 cgctatcggc tgaatttgat tgcgagtgag atatttatgc cagccagcca gacgcagacg
5401 cgccgagaca gaacttaatg ggcccgctaa cagcgcgatt tgctggtgac ccaatgcgac
5461 cagatgctcc acgcccagtc gcgtaccgtc ttcatgggag aaaataatac tgttgatggg
5521 tgtctggtca gagacatcaa gaaataacgc cggaacatta gtgcaggcag cttccacagc
5581 aatggcatcc tggtcatcca gcggatagtt aatgatcagc ccactgacgc gttgcgcgag
5641 aagattgtgc accgccgctt tacaggcttc gacgccgctt cgttctacca tcgacaccac
5701 cacgctggca cccagttgat cggcgcgaga tttaatcgcc gcgacaattt gcgacggcgc
5761 gtgcagggcc agactggagg tggcaacgcc aatcagcaac gactgtttgc ccgccagttg
5821 ttgtgccacg cggttgggaa tgtaattcag ctccgccatc gccgcttcca ctttttcccg
5881 cgttttcgca gaaacgtggc tggcctggtt caccacgcgg gaaacggtct gataagagac
5941 accggcatac tctgcgacat cgtataacgt tactggtttc acattcacca ccctgaattg
6001 actctcttcc gggcgctatc atgccatacc gcgaaaggtt ttgcgccatt cgatggtgtc
6061 cgggatctcg acgctctccc ttatgcgact cctgcattag gaaattaata cgactcacta
6121 ta
//
```

A.3 pET A3A4A5 sequence

FEATURES	Location/Qualifiers
lacI	complement(6712..7794)
PBR322 origin	5518..5519
Amp resistance	3900..4757
T7 promoter 1	8185..8201
T7 promoter 2	1477..1493
atdA3	71..1396
atdA4	1563..2189
atdA5	2203..3210

```
   1 ggggaattgt gagcggataa caattcccct ctagaaataa ttttgtttaa ctttaagaag
  61 gagatatacc atgggcagca gccatcacca tcatcaccac agccaggatc cgaattcgat
 121 gaaaaccata aatcaactaa ttcagtccgg gcgcgtacac cgtaaggtct atacagaagc
 181 atctattttt caagcggaaa tggacaaaat atttcaagcg aactgggtgt ttctcttaca
 241 tgcaagccaa attcctaaac ttgatgatta tcaaacggtt cggatgggtg ggcgaccact
 301 gatcgttgta agaaaaggag atgatgagtt tcaggcattg cttaatcgtt gtccgcatcg
 361 tggggcaaag gtttgtcgaa atgattcggg taactccaag acatttactt gcccttacca
 421 tggttggaaa ttcaggaact caggcaaagc ttttgttatt cctggtgcga atgcctatgg
 481 agagggtttc gataaagaca atttctccat gacggcaatt ccacgggtgg aaagctatcg
 541 aggttttgtc tttgctacca gtaatgagaa cgctgtttcg ttagaggagc atcttggtag
 601 tgcgcgccag tatattgatg aatggttagc ccaccagggt ggtgagatta aagtatcgaa
 661 gtctgttcaa cgttatgaaa taaaatgtaa ttggaagctg gtgtttgata atgctggtga
 721 tggctaccat gtccctttt cccatcagtc attgctacaa atgactacgc ttcgatatgg
 781 cggcggggat atacagtatt tcggtaatgc cgatgagaca ggtatgggcc tttatgcttt
 841 gggtaatggt cactcagtga tagatcagcg cccggaaatg cataaagagt cagggtggga
 901 tcagcagcga ccacagcctg gccgcgaaag ctatgaaaca cacgttcgta ataatagtag
 961 ccagccagca agagatttag agcgagcggt tggtgctgga atgaatctta atatttttcc
1021 aaacttactg ttaattggca accaaataca ggttattgat cctatttctg ttaatgaaac
1081 agttctgcat tggcatgcaa ccttgcttgc tggtgataat gaagagctca atgcaattcg
1141 gatgcgcact caggaggatt ttccaattat gggtgaggtg gatgacgtgg ccaactttga
1201 atcatgtcaa gaaggactcg agaccatgcc ggaaatcgaa tggatcgatt ttagcaggca
1261 tatgaatgaa ggagaaaatg cttgctacca agatgttata caacataaac caacatctga
1321 gatccatagt cgccattatt ttgatacctg gctacagcta atgtctgcag taaacaaaga
1381 gaatcagtct gaggtttaag tcgacaagct tgcggccgca taatgcttaa gtcgaacaga
1441 aagtaatcgt attgtacacg gccgcataat cgaaattaat acgactcact ataggggaat
1501 tgtgagcgga taacaattcc ccatcttagt atattagtta agtataagaa ggagatatac
1561 atatggcaga tctcaattgg atatcggccg gccacatgaa taataataaa gatcttagta
1621 taaaaattgt tgatcctagc gctgttaacc tcgcttatta tcaagagatt aagcagtact
1681 ctgattattt ttggaatctt tctaaccttg gtgagcctgc gcttgatcac aagataaata
1741 tgtttctgac taaagaagct cggcttttgg atcagcagtg ttttgatgag tggttaacac
1801 tgtttctgga ggatggctgt tactggattc ctggcagtat gccggcggcg tcacctgcca
1861 gcgaggccac atatgagttt catgatatac gcagactgaa agatcgaatt gtaaggctgc
1921 aaacgggttt tgcctattcg cagatacctg tttcaaaaac taatcgcatc cttggggcgc
1981 cagaagtatg ggcagtgccg gggtcaagcg aggggttttt ggttagaaca agctttattg
2041 tgtttgaaag ccgagatggc aagtctcaag ttttaagtgg ttggtatggt tatgtaatta
2101 ttaaggataa tgatgagttg aagataaaga tgaaacagat aaatttaaac gactgtcttt
2161 cgccgcaagg caataattca ttttttctat agttggtcta aaatgaatac attaaaattt
2221 cgagttattg ataagatagc ggaaacgaag gagtcgtttt catttgttct gaagccgttg
2281 gacggtgtct tggctgagca ctcccctggc aagtatttac caattaaaat ccgaactgaa
2341 aaaggtcttt tatttcggtc ttattcccta tcttcgtcag cttcggccaa tgaggacttt
2401 aaaatcacgg taaagagaga aaggggaggc agagggtcga actggttgtg tgacaatgta
2461 aaggttggcg actttatcga gacactcccc cctgctggca gtttccaccc gcagaactgg
2521 gatcgagatt ttgtttttt tgcaggaggt agtggtataa ctcctgtaat atccattata
```

```
2581 aaaacagcgc taaatagaca caaaaatagg attaagttgt tttatgctaa ctcctctgaa
2641 agttctataa tatttcataa agagttaaaa gatctatgtt tacaattccc ggatcggctt
2701 gatatacaat tctggttaga cgatgaaaaa ggtattccaa cctctatcgc gtttgagcaa
2761 tatattgatg atgctctaga ggttgaatat tttttatgtg gacctgcccc cttcatggga
2821 ggtgtagaga atttttttgat cgagtctaaa gtcccacctg gactgataac caaggagtct
2881 tttgctggga gtgtttctga tgataatggc gatacagttg aaagctcagc tgaaaaagat
2941 gtcactgtaa actttatgct aaatggcatt aagaacagcg ttatgtgctc cgaagatgat
3001 tttattttaa acgagataat aaaagctgga attaatgttc ctagttcgtg ctgtgctggt
3061 aattgtgggt cttgcatgtg tctgctggtg agtggagatg taattctaga aagcaatact
3121 gttttggacg cttctgatga ggaagacggc tggatcttgg cgtgtcgatc taaaccgaga
3181 tcgaaaaata tagaaatatc cttcgaccaa taacctaggc tgctgccacc gctgagcaat
3241 aactagcata acccccttggg gcctctaaac gggtcttgag gggttttttg ctgaaaggag
3301 gaactatatc cggattggcg aatgggacgc gccctgtagc ggcgcattaa gcgcggcggg
3361 tgtggtggtt acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt
3421 cgctttcttc ccttcctttc tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg
3481 ggggctccct ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga
3541 ttagggtgat ggttcacgta gtgggccatc gccctgatag acggtttttc gccctttgac
3601 gttggagtcc acgttcttta atagtggact cttgttccaa actggaacaa cactcaaccc
3661 tatctcggtc tattcttttg atttataagg gattttgccg atttcggcct attggttaaa
3721 aaatgagctg atttaacaaa aatttaacgc gaattttaac aaaatattaa cgtttacaat
3781 ttctggcggc acgatggcat gagattatca aaaaggatct tcacctagat ccttttaaat
3841 taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac
3901 caatgcttaa tcagtgaggc acctatctca gcgatctgtc tatttcgttc atccatagtt
3961 gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt
4021 gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag
4081 ccagccggaa gggccgagcg cagaagtggt cctgcaactt tatccgcctc catccagtct
4141 attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt
4201 gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc
4261 tccggttccc aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt
4321 agctccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg
4381 gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg
4441 actggtgagt actcaaccaa gtcattctga gaatagtgta tgcggcgacc gagttgctct
4501 tgcccggcgt caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc
4561 attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt
4621 tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt
4681 tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg
4741 aaatgttgaa tactcatact cttccttttt caatcatgat tgaagcattt atcagggtta
4801 ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa taggtcatga
4861 ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca
4921 aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac
4981 caccgctacc agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg
5041 taactggctt cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag
5101 gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac
5161 cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt
5221 taccggataa ggcgcagcgg tcgggctgaa cgggggttc gtgcacacag cccagcttgg
5281 agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc
5341 ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc
5401 gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc
5461 acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa
5521 acgccagcaa cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt
5581 tctttcctgc gttatcccct gattctgtgg ataaccgtat taccgccttt gagtgagctg
5641 ataccgctcg ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag
5701 agcgcctgat gcggtatttt ctccttacgc atctgtgcgg tatttcacac cgcatatatg
5761 gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagtata cactccgcta
5821 tcgctacgtg actgggtcat ggctgcgccc cgacacccgc caacacccgc tgacgcgccc
5881 tgacgggctt gtctgctccc ggcatccgct tacagacaag ctgtgaccgt ctccgggagc
5941 tgcatgtgtc agaggttttc accgtcatca ccgaaacgcg cgaggcagct gcggtaaagc
6001 tcatcagcgt ggtcgtgaag cgattcacag atgtctgcct gttcatccgc gtccagctcg
6061 ttgagtttct ccagaagcgt taatgtctgg cttctgataa agcgggccat gttaagggcg
6121 gtttttttcct gtttggtcac tgatgcctcc gtgtaagggg gatttctgtt catgggggta
6181 atgataccga tgaaacgaga gaggatgctc acgatacggg ttactgatga tgaacatgcc
6241 cggttactgg aacgttgtga gggtaaacaa ctggcggtat ggatgcggcg ggaccagaga
6301 aaaatcactc agggtcaatg ccagcgcttc gttaatacag atgtaggtgt tccacagggt
```

```
     6361 agccagcagc atcctgcgat gcagatccgg aacataatgg tgcagggcgc tgacttccgc
     6421 gtttccagac tttacgaaac acggaaaccg aagaccattc atgttgttgc tcaggtcgca
     6481 gacgttttgc agcagcagtc gcttcacgtt cgctcgcgta tcggtgattc attctgctaa
     6541 ccagtaaggc aaccccgcca gcctagccgg gtcctcaacg acaggagcac gatcatgcta
     6601 gtcatgcccc gcgcccaccg gaaggagctg actgggttga aggctctcaa gggcatcggt
     6661 cgagatcccg gtgcctaatg agtgagctaa cttacattaa ttgcgttgcg ctcactgccc
     6721 gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg
     6781 agaggcggtt tgcgtattgg gcgccagggt ggtttttctt ttcaccagtg agacgggcaa
     6841 cagctgattg cccttcaccg cctggccctg agagagttgc agcaagcggt ccacgctggt
     6901 ttgccccagc aggcgaaaat cctgtttgat ggtggttaac ggcgggatat aacatgagct
     6961 gtcttcggta tcgtcgtatc ccactaccga gatgtccgca ccaacgcgca gcccggactc
     7021 ggtaatggcg cgcattgcgc ccagcgccat ctgatcgttg gcaaccagca tcgcagtggg
     7081 aacgatgccc tcattcagca tttgcatggt ttgttgaaaa ccggacatgg cactccagtc
     7141 gccttcccgt tccgctatcg gctgaatttg attgcgagtg agatatttat gccagccagc
     7201 cagacgcaga cgcgccgaga cagaacttaa tgggcccgct aacagcgcga tttgctggtg
     7261 acccaatgcg accagatgct ccacgcccag tcgcgtaccg tcttcatggg agaaaataat
     7321 actgttgatg ggtgtctggt cagagacatc aagaaataac gccggaacat tagtgcaggc
     7381 agcttccaca gcaatggcat cctggtcatc cagcggatag ttaatgatca gcccactgac
     7441 gcgttgcgcg agaagattgt gcaccgccgc tttacaggct tcgacgccgc ttcgttctac
     7501 catcgacacc accacgctgg cacccagttg atcggcgcga gatttaatcg ccgcgacaat
     7561 ttgcgacggc gcgtgcaggg ccagactgga ggtggcaacg ccaatcagca acgactgttt
     7621 gcccgccagt tgttgtgcca cgcggttggg aatgtaattc agctccgcca tcgccgcttc
     7681 cacttttttcc cgcgttttcg cagaaacgtg gctggcctgg ttcaccacgc gggaaacggt
     7741 ctgataagag acaccggcat actctgcgac atcgtataac gttactggtt tcacattcac
     7801 caccctgaat tgactctctt ccgggcgcta tcatgccata ccgcgaaagg ttttgcgcca
     7861 ttcgatggtg tccgggatct cgacgctctc ccttatgcga ctcctgcatt aggaagcagc
     7921 ccagtagtag gttgaggccg ttgagcaccg ccgccgcaag gaatggtgca tgcaaggaga
     7981 tggcgcccaa cagtcccccg gccacggggc ctgccaccat acccacgccg aaacaagcgc
     8041 tcatgagccc gaagtggcga gcccgatctt ccccatcggt gatgtcggcg atataggcgc
     8101 cagcaaccgc acctgtggcg ccggtgatgc cggccacgat gcgtccggcg tagaggatcg
     8161 agatcgatct cgatcccgcg aaattaatac gactcactat a
//
```

A.4 pACYC A1 sequence

FEATURES	Location/Qualifiers
lacI	complement(4274..5353)
Cm resistance	complement(2217..2873)
P15A origin	3235..4147
T7 promoter 1	5477..5493
T7 promoter 2	1699..1715
atdA1	71..1618

```
   1 ggggaattgt gagcggataa caattcccct gtagaaataa ttttgtttaa ctttaataag
  61 gagatatacc atgggcagca gccatcacca tcatcaccac agccaggatc cgaattcgat
 121 gagtgagaaa ttagatttta taacgaaaaa taatctttgg acagataagc agcgagatgc
 181 agccgacaaa gttctcgcag aaattgattc tttagggctt gagatgattc ggctctcctg
 241 ggctgatcag tatggtctct tgcgaggtaa ggcgctgtcg gtagcagccc tcaaggcggc
 301 attctcagaa gggtcagaag ttacgatggc acctttctct ttcaacctag tcagcgaatg
 361 ggttttcaac ccatttactg ctggtggtgg cttcggaatt gatgagtttg atgagttggg
 421 tggtgtcccg agtgtggtga tggttccaga tcctacgacg ttcaaggtct taccttgggc
 481 agataaaacc ggctggatgc tggcagatct gcattggaaa tctggtgaac cattcccatt
 541 atgtccccgc ggtatcatga agaaggctgt caaatcgtta agcgatgaag gttacttatt
 601 taaatgcggt attgagcttg aatggtactt gacgaaaatt gttgatcgct cactttctcc
 661 agagagttta ggtgcgccag gtgtacagcc tgatgccatt caagttcaac ccgtggcgca
 721 agggtactcc tatctacttg aatatcactt agatcaggtg gacgatatta tgtccaaggt
 781 tcgaaaaggt cttctcgagc tcaatctgcc tttgcgctca atagaagatg agttggcacc
 841 aagccaaatg gaaaccacgt ttgatgtaat ggaaggttta gaagcagccg atgcagcgct
 901 acttataaaa tcggccatca aacaaatatg ttcacgacat ggttatcacg caacatttat
 961 gtgcaagccg gcaattaacg ggttctctgt tgcttcaggc tggcatatgc atcaatcact
1021 agtggataaa gatacccgaa agaatctttt tataccctca gaaggggaag tggtatcccc
1081 gctaggtcga gcttatgctg gtggattact tgcaaatggt agtgccgcct cgagtttcac
1141 aacaccaact gtgaatgggt atcgaagacg tcagccgcac tcgcttgcac cagaccgaag
1201 agcttgggcg aaagaaaaca aggcagcgat ggtccgtgta atttctgcaa caggcgatcc
1261 ggctagccgg atcgaaaatc gtattggtga gcccggcgcc aacccttatt tatatatggc
1321 atcacaaatt gtctctgggc ttgatggcat taaaatcaaa agggatcccg gcgggttgca
1381 aggggctcct tatggtgcac aagtaccaat gctgccaacg gctttggctg aggctctgga
1441 tgctcttgag cacgattcgg agttgtttag aagctgcttt ggcgacacct ttattaaata
1501 ttggctgcaa ttaagaagat ccgagtgggc aagatttctc gatgctgaag gtgctgaggc
1561 tgctgagcct acaggtgctg tcacgcagtg ggaacaaaaa gaatacttca acttactgtg
1621 agtcgacaag cttgcggccg cataatgctt aagtcgaaca gaaagtaatc gtattgtaca
1681 cggccgcata atcgaaatta atacgactca ctatagggga attgtgagcg gataacaatt
1741 ccccatctta gtatattagt taagtataag aaggagatat acatatggca gatctcaatt
1801 ggatatcggc cggccacgcg atcgctgacg tcggtaccct cgagtctggt aaagaaaccg
1861 ctgctgcgaa atttgaacgc cagcacatgg actcgtctac tagcgcagct taattaacct
1921 aggctgctgc caccgctgag caataactag cataacccct tggggcctct aaacgggtct
1981 tgaggggttt tttgctgaaa cctcaggcat ttgagaagca cacggtcaca ctgcttccgg
2041 tagtcaataa accggtaaac cagcaataga cataagcggc tatttaacga ccctgccctg
2101 aaccgacgac cgggtcgaat ttgctttcga atttctgcca ttcatccgct tattatcact
2161 tattcaggcg tagcaccagg cgtttaaggg caccaataac tgccttaaaa aaattacgcc
2221 ccgccctgcc actcatcgca gtactgttgt aattcattaa gcattctgcc gacatggaag
2281 ccatcacaga cggcatgatg aacctgaatc gccagcggca tcagcacctt gtcgccttgc
2341 gtataatatt tgcccatagt gaaaacgggg gcgaagaagt tgtccatatt ggccacgttt
2401 aaatcaaaac tggtgaaact cacccaggga ttggctgaga cgaaaaacat attctcaata
2461 aaccctttag ggaaataggc caggttttca ccgtaacacg ccacatcttg cgaatatatg
2521 tgtagaaact gccggaaatc gtcgtggtat tcactccaga gcgatgaaaa cgtttcagtt
2581 tgctcatgga aaacggtgta acaagggtga acactatccc atatcaccag ctcaccgtct
2641 ttcattgcca tacggaactc cggatgagca ttcatcaggc gggcaagaat gtgaataaag
2701 gccggataaa acttgtgctt atttttcttt acggtcttta aaaaggccgt aatatccagc
2761 tgaacggtct ggttataggt acattgagca actgactgaa atgcctcaaa atgttcttta
```

```
2821 cgatgccatt gggatatatc aacggtggta tatccagtga ttttttttctc cattttagct
2881 tccttagctc ctgaaaatct cgataactca aaaaatacgc ccggtagtga tcttatttca
2941 ttatggtgaa agttggaacc tcttacgtgc cgatcaacgt ctcattttcg ccaaaagttg
3001 gcccagggct tcccggtatc aacagggaca ccaggattta tttattctgc gaagtgatct
3061 tccgtcacag gtatttattc ggcgcaaagt gcgtcgggtg atgctgccaa cttactgatt
3121 tagtgtatga tggtgttttt gaggtgctcc agtggcttct gtttctatca gctgtccctc
3181 ctgttcagct actgacgggg tggtgcgtaa cggcaaaagc accgccggac atcagcgcta
3241 gcggagtgta tactggctta ctatgttggc actgatgagg gtgtcagtga agtgcttcat
3301 gtggcaggag aaaaaaggct gcaccggtgc gtcagcagaa tatgtgatac aggatatatt
3361 ccgcttcctc gctcactgac tcgctacgct cggtcgttcg actgcggcga gcggaaatgg
3421 cttacgaacg gggcggagat ttcctggaag atgccaggaa gatacttaac agggaagtga
3481 gagggccgcg gcaaagccgt ttttccatag gctccgcccc cctgacaagc atcacgaaat
3541 ctgacgctca aatcagtggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc
3601 cctggcggct ccctcgtgcg ctctcctgtt cctgcctttc ggtttaccgg tgtcattccg
3661 ctgttatggc cgcgtttgtc tcattccacg cctgacactc agttccgggt aggcagttcg
3721 ctccaagctg gactgtatgc acgaaccccc cgttcagtcc gaccgctgcg ccttatccgg
3781 taactatcgt cttgagtcca acccggaaag acatgcaaaa gcaccactgg cagcagccac
3841 tggtaattga tttagaggag ttagtcttga agtcatgcgc cggttaaggc taaactgaaa
3901 ggacaagttt tggtgactgc gctcctccaa gccagttacc tcggttcaaa gagttggtag
3961 ctcagagaac cttcgaaaaa ccgccctgca aggcggtttt ttcgttttca gagcaagaga
4021 ttacgcgcag accaaaacga tctcaagaag atcatcttat taatcagata aaatatttct
4081 agatttcagt gcaatttatc tcttcaaatg tagcacctga agtcagcccc atacgatata
4141 agttgtaatt ctcatgttag tcatgccccg cgcccaccgg aaggagctga ctgggttgaa
4201 ggctctcaag ggcatcggtc gagatcccgg tgcctaatga gtgagctaac ttacattaat
4261 tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg
4321 aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgccagggtg gtttttcttt
4381 tcaccagtga gacgggcaac agctgattgc ccttcaccgc ctggccctga gagagttgca
4441 gcaagcggtc cacgctggtt tgccccagca ggcgaaaatc ctgtttgatg gtggttaacg
4501 gcgggatata acatgagctg tcttcggtat cgtcgtatcc cactaccgag atgtccgcac
4561 caacgcgcag cccggactcg gtaatggcgc gcattgcgcc cagcgccatc tgatcgttgg
4621 caaccagcat cgcagtggga acgatgccct cattcagcat ttgcatggtt tgttgaaaac
4681 cggacatggc actccagtcg ccttcccgtt ccgctatcgg ctgaatttga ttgcgagtga
4741 gatatttatg ccagccagcc agacgcagac gcgccgagac agaacttaat gggcccgcta
4801 acagcgcgat ttgctggtga cccaatgcga ccagatgctc cacgcccagt cgcgtaccgt
4861 cttcatggga gaaaataata ctgttgatgg gtgtctggtc agagacatca agaaataacg
4921 ccggaacatt agtgcaggca gcttccacag caatggcatc ctggtcatcc agcggatagt
4981 taatgatcag cccactgacg cgttgcgcga gaagattgtg caccgccgct ttacaggctt
5041 cgacgccgct tcgttctacc atcgacacca ccacgctggc acccagttga tcggcgcgag
5101 atttaatcgc cgcgacaatt tgcgacggcg cgtgcagggc cagactggag gtggcaacgc
5161 caatcagcaa cgactgtttg cccgccagtt gttgtgccac gcggttggga atgtaattca
5221 gctccgccat cgccgcttcc acttttccc gcgttttcgc agaaacgtgg ctggcctggt
5281 tcaccacgcg ggaaacggtc tgataagaga caccggcata ctctgcgaca tcgtataacg
5341 ttactggttt cacattcacc accctgaatt gactctcttc cgggcgctat catgccatac
5401 cgcgaaaggt tttgcgccat tcgatggtgt ccgggatctc gacgctctcc cttatgcgac
5461 tcctgcatta ggaaattaat acgactcact ata
//
```

A.5 pACYC A2 sequence

```
FEATURES             Location/Qualifiers
     lacI            complement(3418..4497)

     Cm resistance   complement(1361..2017)

     P15A origin     2379..3291

     T7 promoter 1   4621..4637

     T7 promoter 2    214..230

     atdA2            300..1058

        1 ggggaattgt gagcggataa caattcccct gtagaaataa ttttgtttaa ctttaataag
       61 gagatatacc atgggcagca gccatcacca tcatcaccac agccaggatc cgaattcgag
      121 ctcggcgcgc ctgcaggtcg acaagcttgc ggccgcataa tgcttaagtc gaacagaaag
      181 taatcgtatt gtacacggcc gcataatcga aattaatacg actcactata ggggaattgt
      241 gagcggataa caattcccca tcttagtata ttagttaagt ataagaagga gatatacata
      301 tggcagatct caattggata tcggccggcc acatgtctaa acgctttgca ttattgtggt
      361 gctctgaaga agagcgcttt gattatcgag aagaaatggt aaatgccttt aaaactgaaa
      421 actccgactg ggaagttata agtgcattca cagacttaaa taaaattatc gataattacg
      481 atggctttgt tatcagtggt agtgagtact cagttaatgc tgataaagaa aagttttctg
      541 gcttatttga atttattcga gcggtccata agaaagaaaa accaattgtt ggcatatgct
      601 tcggttgtca gtcccttgct gtcgcacttg gcggagaggt gggtttgaac cctagtcgtg
      661 agtttaggtt tggaactgat gagctcacgt ttcaaaatgg acttaacaaa catgttggta
      721 ccagtgaaga gcgagtgagg cttattgaaa gccatggaga atgcgtcatt agacgcccac
      781 ttggatctac attactcgca cgttctgatt caactgctgt agaaattttt gcggtggggc
      841 cgtatgcagt cggtatccaa ggacatccag agatcagtaa aaaaaccctg gagcaagact
      901 ttctacgggt tcatctcgaa gatggtaatt tgcaagaaga tgaggtacgc aggtttcatg
      961 ctgagctgag tggttatcag cctcctcaag cgatacgtca attagtgaaa gcgactctac
     1021 acaagcaaat taattttcag aacttggtgg gtgatgtata acctaggctg ctgccaccgc
     1081 tgagcaataa ctagcataac cccttggggc ctctaaacgg gtcttgaggg gttttttgct
     1141 gaaacctcag gcatttgaga agcacacggt cacactgctt ccggtagtca ataaaccggt
     1201 aaaccagcaa tagacataag cggctattta acgaccctgc cctgaaccga cgaccgggtc
     1261 gaatttgctt tcgaatttct gccattcatc cgcttattat cacttattca ggcgtagcac
     1321 caggcgttta agggcaccaa taactgcctt aaaaaaatta cgccccgccc tgccactcat
     1381 cgcagtactg ttgtaattca ttaagcattc tgccgacatg gaagccatca cagacggcat
     1441 gatgaacctg aatcgccagc ggcatcagca ccttgtcgcc ttgcgtataa tatttgccca
     1501 tagtgaaaac gggggcgaag aagttgtcca tattggccac gtttaaatca aaactggtga
     1561 aactcaccca gggattggct gagacgaaaa acatattctc aataaaccct ttagggaaat
     1621 aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga
     1681 aatcgtcgtg gtattcactc cagagcgatg aaaacgtttc agtttgctca tggaaaacgg
     1741 tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtctttcatt gccatacgga
     1801 actccggatg agcattcatc aggcgggcaa gaatgtgaat aaaggccgga taaaacttgt
     1861 gcttatttttt ctttacggtc tttaaaaagg ccgtaatatc cagctgaacg gtctggttat
     1921 aggtacattg agcaactgac tgaaatgcct caaaatgttc tttacgatgc cattgggata
     1981 tatcaacggt ggtatatcca gtgatttttt tctccatttt agcttcctta gctcctgaaa
     2041 atctcgataa ctcaaaaaat acgcccggta gtgatcttat ttcattatgg tgaaagttgg
     2101 aacctcttac gtgccgatca acgtctcatt ttcgccaaaa gttggcccag ggcttcccgg
     2161 tatcaacagg gacaccagga tttatttatt ctgcgaagtg atcttccgtc acaggtattt
     2221 attcggcgca aagtgcgtcg ggtgatgctg ccaacttact gatttagtgt atgatggtgt
     2281 ttttgaggtg ctccagtggc ttctgtttct atcagctgtc cctcctgttc agctactgac
     2341 ggggtggtgc gtaacggcaa aagcaccgcc ggacatcagc gctagcggag tgtatactgg
     2401 cttactatgt tggcactgat gagggtgtca gtgaagtgct tcatgtggca ggagaaaaaa
     2461 ggctgcaccg gtgcgtcagc agaatatgtg atacaggata tattccgctt cctcgctcac
     2521 tgactcgcta cgctcggtcg ttcgactgcg gcgagcggaa atggcttacg aacggggcgg
     2581 agatttcctg gaagatgcca ggaagatact taacagggaa gtgagagggc cgcggcaaag
     2641 ccgtttttcc ataggctccg ccccccctgac aagcatcacg aaatctgacg ctcaaatcag
     2701 tggtggcgaa acccgacagg actataaaga taccaggcgt ttcccctggc ggctccctcg
     2761 tgcgctctcc tgttcctgcc tttcggttta ccggtgtcat tccgctgtta tggccgcgtt
```

```
     2821 tgtctcattc cacgcctgac actcagttcc gggtaggcag ttcgctccaa gctggactgt
     2881 atgcacgaac cccccgttca gtccgaccgc tgcgccttat ccggtaacta tcgtcttgag
     2941 tccaacccgg aaagacatgc aaaagcacca ctggcagcag ccactggtaa ttgatttaga
     3001 ggagttagtc ttgaagtcat gcgccggtta aggctaaact gaaaggacaa gttttggtga
     3061 ctgcgctcct ccaagccagt tacctcggtt caaagagttg gtagctcaga gaaccttcga
     3121 aaaaccgccc tgcaaggcgg ttttttcgtt ttcagagcaa gagattacgc gcagaccaaa
     3181 acgatctcaa gaagatcatc ttattaatca gataaaatat ttctagattt cagtgcaatt
     3241 tatctcttca aatgtagcac ctgaagtcag ccccatacga tataagttgt aattctcatg
     3301 ttagtcatgc cccgcgccca ccggaaggag ctgactgggt tgaaggctct caagggcatc
     3361 ggtcgagatc ccggtgccta atgagtgagc taacttacat taattgcgtt gcgctcactg
     3421 cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg
     3481 gggagaggcg gtttgcgtat tgggcgccag ggtggttttt cttttcacca gtgagacggg
     3541 caacagctga ttgcccttca ccgcctggcc ctgagagagt tgcagcaagc ggtccacgct
     3601 ggtttgcccc agcaggcgaa aatcctgttt gatggtggtt aacggcggga tataacatga
     3661 gctgtcttcg gtatcgtcgt atcccactac cgagatgtcc gcaccaacgc gcagcccgga
     3721 ctcggtaatg gcgcgcattg cgcccagcgc catctgatcg ttggcaacca gcatcgcagt
     3781 gggaacgatg ccctcattca gcatttgcat ggtttgttga aaaccggaca tggcactcca
     3841 gtcgccttcc cgttccgcta tcggctgaat ttgattgcga gtgagatatt tatgccagcc
     3901 agccagacgc agacgcgccg agacagaact taatgggccc gctaacagcg cgatttgctg
     3961 gtgacccaat gcgaccagat gctccacgcc cagtcgcgta ccgtcttcat gggagaaaat
     4021 aatactgttg atgggtgtct ggtcagagac atcaagaaat aacgccggaa cattagtgca
     4081 ggcagcttcc acagcaatgg catcctggtc atccagcgga tagttaatga tcagcccact
     4141 gacgcgttgc gcgagaagat tgtgcaccgc cgctttacag gcttcgacgc cgcttcgttc
     4201 taccatcgac accaccacgc tggcacccag ttgatcggcg cgagatttaa tcgccgcgac
     4261 aatttgcgac ggcgcgtgca gggccagact ggaggtggca acgccaatca gcaacgactg
     4321 tttgcccgcc agttgttgtg ccacgcggtt gggaatgtaa ttcagctccg ccatcgccgc
     4381 ttccactttt tcccgcgttt tcgcagaaac gtggctggcc tggttcacca cgcgggaaac
     4441 ggtctgataa gagacaccgg catactctgc gacatcgtat aacgttactg gtttcacatt
     4501 caccaccctg aattgactct cttccgggcg ctatcatgcc ataccgcgaa aggttttgcg
     4561 ccattcgatg gtgtccggga tctcgacgct ctcccttatg cgactcctgc attaggaaat
     4621 taatacgact cactata
//
```

A.6 pET A4A5 sequence

```
FEATURES             Location/Qualifiers
     lacI            complement(5449..6531)

     PBR322 origin   4255..4256

     Amp resistance  2637..3494

     T7 promoter 1   6922..6938

     T7 promoter 2    214..230

     atdA4            300..926

     atdA5            940..1947

        1 ggggaattgt gagcggataa caattcccct ctagaaataa ttttgtttaa ctttaagaag
       61 gagatatacc atgggcagca gccatcacca tcatcaccac agccaggatc cgaattcgag
      121 ctcggcgcgc ctgcaggtcg acaagcttgc ggccgcataa tgcttaagtc gaacagaaag
      181 taatcgtatt gtacacggcc gcataatcga aattaatacg actcactata ggggaattgt
      241 gagcggataa caattcccca tcttagtata ttagttaagt ataagaagga gatatacata
      301 tggcagatct caattggata tcggccggcc acatgaataa taataaagat cttagtataa
      361 aaattgttga tcctagcgct gttaacctcg cttattatca agagattaag cagtactctg
      421 attatttttg gaatctttct aaccttggtg agcctgcgct tgatcacaag ataaatatgt
      481 ttctgactaa agaagctcgg cttttggatc agcagtgttt tgatgagtgg ttaacactgt
      541 ttctggagga tggctgttac tggattcctg gcagtatgcc ggcggcgtca cctgccagcg
      601 aggccacata tgagtttcat gatatacgca gactgaaaga tcgaattgta aggctgcaaa
      661 cgggttttgc ctattcgcag atacctgttt caaaaactaa tcgcatcctt ggggcgccag
      721 aagtatgggc agtgccgggg tcaagcgagg ggtttttggt tagaacaagc tttattgtgt
      781 ttgaaagccg agatggcaag tctcaagttt taagtggttg gtatggttat gtaattatta
      841 aggataatga tgagttgaag ataaagatga aacagataaa tttaaacgac tgtctttcgc
      901 cgcaaggcaa taattcattt tttctatagt tggtctaaaa tgaatacatt aaaatttcga
      961 gttattgata agatagcgga aacgaaggag tcgttttcat ttgttctgaa gccgttggac
     1021 ggtgtcttgg ctgagcactc ccctggcaag tatttaccaa ttaaaatccg aactgaaaaa
     1081 ggtcttttat ttcggtctta ttccctatct tcgtcagctt cggccaatga ggactttaaa
     1141 atcacggtaa agagagaaag gggaggcaga gggtcgaact ggttgtgtga caatgtaaag
     1201 gttggcgact ttatcgagac actcccccct gctggcagtt tccacccgca gaactgggat
     1261 cgagattttg tttttttgc aggaggtagt ggtataactc ctgtaatatc cattataaaa
     1321 acagcgctaa atagacacaa aaataggatt aagttgtttt atgctaactc ctctgaaagt
     1381 tctataatat ttcataaaga gttaaaagat ctatgtttac aattcccgga tcggcttgat
     1441 atacaattct ggttagacga tgaaaaaggt attccaacct ctatcgcgtt tgagcaatat
     1501 attgatgatg ctctagaggt tgaatatttt ttatgtggac ctgccccctt catgggaggt
     1561 gtagagaatt ttttgatcga gtctaaagtc ccacctggac tgataaccaa ggagtctttt
     1621 gctgggagtg tttctgatga taatggcgat acagttgaaa gctcagctga aaaagatgtc
     1681 actgtaaact ttatgctaaa tggcattaag aacagcgtta tgtgctccga agatgatttt
     1741 attttaaacg agataataaa agctggaatt aatgttccta gttcgtgctg tgctggtaat
     1801 tgtgggtctt gcatgtgtct gctggtgagt ggagatgtaa ttctagaaag caatactgtt
     1861 ttggacgctt ctgatgagga agacggctgg atcttggcgt gtcgatctaa accgagatcg
     1921 aaaaatatag aaatatcctt cgaccaataa cctaggctgc tgccaccgct gagcaataac
     1981 tagcataacc ccttggggcc tctaaacggg tcttgagggg ttttttgctg aaaggaggaa
     2041 ctatatccgg attggcgaat gggacgcgcc ctgtagcggc gcattaagcg cggcgggtgt
     2101 ggtggttacg cgcagcgtga ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc
     2161 tttcttccct tcctttctcg ccacgttcgc cggctttccc cgtcaagctc taaatcgggg
     2221 gctcccttta gggttccgat ttagtgcttt acggcacctc gaccccaaaa aacttgatta
     2281 gggtgatggt tcacgtagtg ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt
     2341 ggagtccacg ttctttaata gtggactctt gttccaaact ggaacaacac tcaaccctat
     2401 ctcggtctat tcttttgatt tataagggat tttgccgatt tcggcctatt ggttaaaaaa
     2461 tgagctgatt taacaaaaat ttaacgcgaa ttttaacaaa atattaacgt ttacaatttc
     2521 tggcggcacg atggcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa
     2581 aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa
     2641 tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc
```

```
2701 tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct
2761 gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca
2821 gccggaaggg ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt
2881 aattgttgcc gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt
2941 gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc
3001 ggttcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc
3061 tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt
3121 atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact
3181 ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc
3241 ccggcgtcaa tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt
3301 ggaaaacgtt cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg
3361 atgtaaccca ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct
3421 gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa
3481 tgttgaatac tcatactctt cctttttcaa tcatgattga agcatttatc agggttattg
3541 tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gtcatgacca
3601 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag
3661 gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac
3721 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa
3781 ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc
3841 accacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag
3901 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac
3961 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc
4021 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc
4081 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca
4141 cgagggagct tccagggga aacgcctggt atctttatag tcctgtcggg tttcgccacc
4201 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg
4261 ccagcaacgc ggcctttta cggttcctgg ccttttgctg gccttttgct cacatgttct
4321 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata
4381 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc
4441 gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc atatatggtg
4501 cactctcagt acaatctgct ctgatgccgc atagttaagc cagtatacac tccgctatcg
4561 ctacgtgact gggtcatggc tgcgccccga cacccgccaa cacccgctga cgcgccctga
4621 cgggcttgtc tgctcccggc atccgcttac agacaagctg tgaccgtctc cgggagctgc
4681 atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga ggcagctgcg gtaaagctca
4741 tcagcgtggt cgtgaagcga ttcacagatg tctgcctgtt catccgcgtc cagctcgttg
4801 agtttctcca gaagcgttaa tgtctggctt ctgataaagc gggccatgtt aagggcggtt
4861 ttttcctgtt tggtcactga tgcctccgtg taaggggat ttctgttcat gggggtaatg
4921 ataccgatga aacgagagag gatgctcacg atacgggtta ctgatgatga acatgcccgg
4981 ttactggaac gttgtgaggg taaacaactg gcggtatgga tgcggcggga ccagagaaaa
5041 atcactcagg gtcaatgcca gcgcttcgtt aatacagatg taggtgttcc acagggtagc
5101 cagcagcatc ctgcgatgca gatccggaac ataatggtgc agggcgctga cttccgcgtt
5161 tccagacttt acgaaacacg gaaaccgaag accattcatg ttgttgctca ggtcgcagac
5221 gttttgcagc agcagtcgct tcacgttcgc tcgcgtatcg gtgattcatt ctgctaacca
5281 gtaaggcaac cccgccagcc tagccgggtc ctcaacgaca ggagcacgat catgctagtc
5341 atgccccgcg cccaccggaa ggagctgact gggttgaagg ctctcaaggg catcggtcga
5401 gatcccggtg cctaatgagt gagctaactt acattaattg cgttgcgctc actgcccgct
5461 ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga
5521 ggcggtttgc gtattgggcg ccagggtggt ttttcttttc accagtgaga cgggcaacag
5581 ctgattgccc ttcaccgcct ggccctgaga gagttgcagc aagcggtcca cgctggtttg
5641 ccccagcagg cgaaaatcct gtttgatggt ggttaacggc gggatataac atgagctgtc
5701 ttcggtatcg tcgtatccca ctaccgagat gtccgcacca acgcgcagcc cggactcggt
5761 aatggcgcgc attgcgccca gcgccatctg atcgttggca accagcatcg cagtgggaac
5821 gatgccctca ttcagcattt gcatggtttg ttgaaaaccg gacatggcac tccagtcgcc
5881 ttcccgttcc gctatcggct gaatttgatt gcgagtgaga tatttatgcc agccagccag
5941 acgcagacgc gccgagacag aacttaatgg gcccgctaac agcgcgattt gctggtgacc
6001 caatgcgacc agatgctcca cgcccagtcg cgtaccgtct tcatgggaga aaataatact
6061 gttgatgggt gtctggtcag agacatcaag aaataacgcc ggaacattag tgcaggcagc
6121 ttccacagca atggcatcct ggtcatccag cggatagtta atgatcagcc cactgacgcg
6181 ttgcgcgaga agattgtgca ccgccgcttt acaggcttcg acgccgcttc gttctaccat
6241 cgacaccacc acgctggcac ccagttgatc ggcgcgagat ttaatcgccg cgacaatttg
6301 cgacggcgcg tgcagggcca gactggaggt ggcaacgcca atcagcaacg actgtttgcc
6361 cgccagttgt tgtgccacgc ggttgggaat gtaattcagc tccgccatcg ccgcttccac
6421 tttttcccgc gttttcgcag aaacgtggct ggcctggttc accacgcggg aaacggtctg
```

```
     6481 ataagagaca ccggcatact ctgcgacatc gtataacgtt actggtttca cattcaccac
     6541 cctgaattga ctctcttccg ggcgctatca tgccataccg cgaaaggttt tgcgccattc
     6601 gatggtgtcc gggatctcga cgctctccct tatgcgactc ctgcattagg aagcagccca
     6661 gtagtaggtt gaggccgttg agcaccgccg ccgcaaggaa tggtgcatgc aaggagatgg
     6721 cgcccaacag tcccccggcc acggggcctg ccaccatacc cacgccgaaa caagcgctca
     6781 tgagcccgaa gtggcgagcc cgatcttccc catcggtgat gtcggcgata taggcgccag
     6841 caaccgcacc tgtggcgccg gtgatgccgg ccacgatgcg tccggcgtag aggatcgaga
     6901 tcgatctcga tcccgcgaaa ttaatacgac tcactata
//
```