IN SILICO MODELING AND ANALYSIS FOR IMPROVING DESULFURIZING BACTERIAL STRAINS

SHILPI AGGARWAL

(M.Tech., Indian Institute of Technology, Roorkee, India)

A THESIS SUBMITTED

FOR THE DEGREE OF PHD OF ENGINEERING

DEPARTMENT OF CHEMICAL AND BIOMOLECULAR ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2012

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Shilpi Aggarwal

10 April 2013

ACKNOWLEDGMENTS

I take this chance to express my sincere gratitude and love for my parents (Mr Subash and Mrs Renu) for their undying faith, encouragement, and unconditional love throughout my life. This PhD thesis is dedicated to them as I could have never come so far without their support. It is with immense pleasure and respect I take this opportunity to express my thankfulness to all those who have helped me in shaping my research career and making my stay in Singapore a truly memorable one.

With the term 'PhD', the first person who comes to my mind is my supervisor, Prof I. A. Karimi. I take this chance to thank him for giving me an opportunity to pursue my research career under his able guidance. As a mentor he has made me acquire skills for critical and logical thinking. His directions and guidance has helped me get deeper insights into the subject and made my learning a great experience. Besides, the kindness and patience that he has shown in handling my mistakes during this research work has helped me become a better person. Truly, as I leave his group, I feel a little heavy for I will miss working smoothly under the guidance of such a wonderful supervisor. It was indeed a privilege to have worked with him.

I extend my sincere thanks to Dr. D. Y. Lee for his timely inputs for my research work. His ideas and suggestions have always made me achieve better results and enhanced my knowledge of the subject. I would also like to thank Prof Kilbane from Illinois Institute of Technology, Illinois, Chicago for sharing his knowledge and experience with working on biodesulfurization. His novel ideas and deeper insights about the subject have added new dimensions to my work.

I would also like to thank Prof Y. P. Ting and Dr. Rudiyanto Gunawan for their kind acceptance to be on the panel of examiners during the qualifying exam. I also thank Prof. Li Zhi and Dr.Yan Kun Lin for being my thesis examiners and spending time on evaluating this thesis. I wish to admire and thank all the unknown reviewers of our publications, who gave constructive feedbacks on all our manuscripts and helped us to bring the best out of this research.

I also extend my thanks to my friend, Rajib Saha from Penn State University, Pennsylvania for his regular guidance and discussions about genome scale modeling. His timely support with the subject has greatly helped in enhancing my knowledge and completion of this work. There are no words to thank my dear friends Nidhi, Susheela, Naresh, Manoj, Himani, Mona, Dr. Mukta, Shobha, Tanvi, Vaani, Sadegh, Rahul, and all others, who have helped me sail through this journey with their kind words and encouragement at all times. Special thanks to my labmates (Hanifah, Kunna,Vasanth, Kefeng, Anoop, Rajnish, Karthik, Nishu) and friends at NUS (Sumit, Shivom, Srinath, Vaibhav, Krishna, and more), who made my stay in Singapore and at NUS all the more comfortable and enjoyable.

And most importantly, I thank my elder sister Surabhi, my brother Shivam, my nephew Devom, and my niece Ananya for their love and affection throughout my life. I am always grateful to my grandparents as nothing would have been possible without their blessings. Finally, I thank my cousins and the entire family for always being my source of inspiration. Their love, continued support, and motivation have always been helpful.

TABLE OF CONTENTS

SUMMARY

In the absence of any competitive and economical renewable fuel, fossil fuels continue to be the major source of energy worldwide. The fossil fuels contain various aliphatic, cyclic, and aromatic compounds some of which contain the heteroatom sulfur. As such on combustion, the fossil fuels release various oxides of sulfur (SO_x). The SO_x exert harmful effects on the environment and human health. As such, governments have laid stringent regulations to limit sulfur content in fossil fuels. Thus, desulfurization (removal of sulfur) has become an important step in the pre-processing of fossil fuels. However, hydrodesulfurization, the prevalent desulfurization method is expensive, energy-intensive, and incapable of removing certain recalcitrant sulfur heterocycles such as benzothiophene (BT), dibenzothiophene (DBT), and their derivatives [1]. Biodesulfurization has been identified as a potential economical and energy efficient alternative. It is a process which involves the use of microbial whole cells or enzymes to reduce the sulfur content in fossil fuels. It offers the advantages of being economical, environment-friendly, and specific in action. Several bacterial strains have been studied for their ability to desulfurize various recalcitrant sulfur heterocycles [2]. Of these, the bacteria belonging to the genus *Rhodococcus* and *Gordonia* have attracted much attention owing to their ability to desulfurize the recalcitrant sulfur compounds non-destructively. However, the desulfurization activities exhibited by the wild type desulfurizing strains are too low for any commercial application. Moreover, they exhibit different activities and preferences for the various sulfur heterocycles and therefore, cannot be used for the simultaneous desulfurization of these compounds. Furthermore, the numerous genetic engineering efforts for enhancing the flux through

the desulfurization pathways in these strains have been unsuccessful at giving the desired levels of desulfurization activity.

The inability of the genetically modified strains to give the expected levels of activity can be attributed to a number of reasons. Firstly, although the pathways for desulfurization of sulfur heterocycles have been elucidated, it is unclear how the extracted sulfur atom is incorporated into biomass or released as sulfate [3]. Secondly, the sulfur metabolism in the desulfurizing strains is not well studied. Thirdly, there are several unknown host functions and factors that are likely to control the extent of desulfurization activities in these strains. Finally, most studies have targeted desulfurization pathway for improving the desulfurizing strains but no holistic study of their metabolism has been performed. However, since any cellular phenotype (here, desulfurization) is an emergent behavior of the various metabolic activities occurring within an organism, a holistic systems study of the entire metabolism of desulfurizing strains is critical. Such integrated study of their cellular metabolism can provide better insights into the interactions between the desulfurization pathway and the other cellular activities. Thus, it will further improve our understanding of the cellular factors that may play key roles in determining the level of desulfurization activity of these strains. Finally, such information can help in developing metabolic engineering strategies for improved desulfurizing strains. The absence of any such holistic study has motivated us to study of the metabolic network of the desulfurizing strains using the systems approach.

To this end, we have reconstructed *in silico* stoichiometric model of sulfur metabolism in *R. erythropolis* for the quantitative elucidation of the process by which the sulfur from various sulfur heterocycles gets assimilated into the biomass. Next, we present an *in silico* genome scale metabolic model of *R. erythropolis* that represents

the various biochemical activities and desulfurization as integrated parts of the entire metabolic network. The model has been used to analyze different known phenotypes and provide meaningful insights into the inexplicable experimental observations reported earlier with *R. erythropolis.* Further, the model has been used to design mutants that can give higher desulfurization activities. A detailed analysis with the model has enabled identification of certain host functions that are likely to play key roles in determining the extent of desulfurization activity achievable with *R. erythropolis*. As a continued effort to improve our genome scale model we have expanded it by including additional functionalities and thermodynamic constraints. The expanded model has been explored to design an *in silico* modified strain for bio-refining (i.e., simultaneous biodesulfurization and biodenitrogenation) of the fossil fuels. However, the *in silico R. erythropolis* is unable to utilize BT but DBT and is in accordance with the reports available in literature. Therefore, we have explored another desulfurizing strain, *G. alkanivorans* which is capable of desulfurizing both BT and DBT.

In the next section, we report the reconstruction of an *in silico* genome-scale metabolic model of *G. alkanivorans* for desulfurization of various sulfur heterocycles. The model has been successfully validated and analyzed using the experimental data available in literature. Further, the model analyses give interesting insights into the metabolism of *G. alkanivorans*. Although the *in silico* strain exhibits desulfurization activities for both BT and DBT, it does not utilize them simultaneously. It preferentially utilizes BT. However, an efficient biodesulfurization process should be able to desulfurize various sulfur heterocycles simultaneously at high activity. Therefore, we developed an *in silico* model for the co-culture of *R. erythropolis* and *G. alkanivorans*, to study desulfurization of BT and DBT in a mixture. The *in silico* co-

culture has been used to determine optimal conditions for enhanced desulfurization of the mixture.

NOMENCLATURE

Notations

v_j Reaction flux (mmol /gdcw-h)

α_j Lower bound for reaction flux (mmol /gdcw-h)

β_j Upper bound for reaction flux (mmol /gdcw-h)

S_{ij} Stoichiometric coefficient of metabolite *i* in reaction *j* (dimensionless)

Z Objective function in the optimization problem

c_j Weight associated with the reaction fluxes in objective function (dimensionless)

K_s Substrate constant for substrate *s* (mM)

v_{max} Maximum reaction rate (mmol /gdcw-h)

Subscripts

s Substrate

i Metabolites

j Reactions

ABBREVIATIONS

BT	Benzothiophene
DA	Desulfurizing Activity
DBT	Dibenzothiophene
DEMs	Deadend Metabolites
DFBA	Dynamic Flux Balance Analysis
EPA	Environment Protection Agency
FBA	Flux Balance Analysis
FVA	Flux Variability Analysis
FSA	Flux Sum Analysis
HBP	Hydroxybiphenyl
HDS	Hydrodesulfurization
HSty	o-Hydroxystyrene
NO_x	Oxides of Nitrogen
PASHs	Poly Aromatic Sulfur Heterocycles
SOR	Sulfite Oxidoreductase

SO_x	Oxides of Sulfur
SR	Sulfite Reductase

LIST OF FIGURES

LIST OF TABLES

1 Introduction

The human civilization has largely depended on the utilization of natural resources for its survival and development. Energy is one of the most important components that have played key roles in the endurance and expansion of civilization. Over the decades, the energy requirements have shown ever increasing trends with the increasing population and ever-expanding industrial sector. Fossil fuels are the important natural resources that have been largely exploited for energy over the years. The fossil fuels contribute to nearly 87% of the total energy consumption across the world. Their usage has been instrumental in the improvement of the living standards and has made the modern life possible. In spite of being non-renewable in nature, the fossil fuels continue to be used largely in the absence of any competitive alternate energy source. Indeed, there has been a twenty fold increase in the usage of fossil fuels for energy over the twentieth century [4]. They are used for various transportation systems, manufacturing of several thousands of goods, running different industries, domestic purposes, etc. In most purposes they are put to combustion for the generation of energy. However, the fossil fuels contain several hydrocarbons, polyaromatic heterocycles, etc. and on combustion they lead to the release of various pollutants [2]. The major pollutants include oxides of carbon, oxides of nitrogen (NO_x), oxides of sulfur (SO_x), unburnt hydrocarbons, etc. Of these, SO_x have attracted major attention of the environment and human health protection groups over the past three decades.

1.1 **Need for desulfurizarion**

The SO_x are formed due to the combustion of the sulfur containing compounds present in fossil fuels. They exert numerous harmful effects on the environment and the living beings [2]. They make the air hazy and thus reduce the visibility. The SO_x leads to the production of the acid rain, which has detrimental effects on the aquatic life, human beings, and certain manmade objects such as lime stone sculptures. They lead to the formation of aerosols that owing to their small size (~2.5 μm) can be inhaled by humans and cause chronic respiratory illness. In addition, the SO_x poisons the catalytic convertors installed on engines to combust the unburnt hydrocarbons and capture NO_x that otherwise are a major source of urban pollution. Considering these deleterious effects, the governments have laid stringent regulations to control the release of SO_x [1]. For this, they have put limits on the allowable level of sulfur content in the fossil fuels. As such, the removal/reduction in sulfur content i.e., desulfurization is an important step in the pre-processing of the fossil fuels. The need for desulfurization is further aggravated by the fact that with fast depletion of the lighter crude, the remaining heavier crude is richer in sulfur. The prevalent method used for the desulfurization of fossil fuels is hydrodesulfurization.

1.2 **Hydrodesulfurization**

Hydrodesulfurization is a chemical method used to remove sulfur atoms from the compounds in fossil fuels. In this process, the sulfur in fossil fuels is catalytically reduced to H_2S in the presence of H_2 gas over a suitable catalyst, under conditions of high pressure and temperature [1]. The H_2S gas is then catalytically oxidized to elemental sulfur. Although this method easily treats the aliphatic sulfur compounds

present in the crude, it is currently unable to desulfurize refractory poly aromatic sulfur hetrocycles (PASHs) such as dibenzothiophene (DBT), benzothiophene (BT), and their derivatives. However, in order to meet the regulations for low sulfur fuels, the desulfurization of such recalcitrant sulfur compounds is essential. The removal of sulfur atom from these refractory PASHs by hydrodesulfurization requires more extreme conditions of temperature and pressure, longer residence times, stronger reaction vessels, and facilities that will require a heavy load of capital cost. Moreover, the hydrodesulfurization leads to hydrogenation of olefinic compounds and reduction in their calorific value. In order to increase the calorific value, the HDS stream has to be sent through a fluid catalytic cracking system that further adds to the fuel processing cost. Thus, HDS process becomes expensive and energy intensive [5] for achieving deeper desulfurization. Clearly, there is a need for alternate methods of desulfurization that are economical, less energy intensive, and more efficient in removing the recalcitrant PASHs. In this regard, biodesulfurization is a potential alternative.

1.3 **Biodesulfurization**

Biodesulfurization is a method that uses either microbial whole cells or enzymes as catalysts to remove sulfur atoms from PASHs present in fossil fuels. It is an attractive alternate desulfurization method as it occurs at ambient conditions and is environment friendly. Moreover, since biological catalysts used in biodesulfurization exhibit specificity in their action, they do not bring about unwanted side reactions such as hydrogenation of olefinic compounds. It is more economical and is likely to reduce the operating costs incurred during the desulfurization of fuels substantially [6]. The biodesulfurization process is based on the premise that sulfur is essential for all

microorganisms. As such, they have evolved diverse mechanisms to utilize a wide range of sulfur compounds including the recalcitrant PASHs. Several bacterial strains belonging to *Pseudomonas, Corynebacterium, Rhodococcus, Mycobacterium, Desulfovibrio, Gordonia,* etc. have been characterized for their ability to desulfurize the various PASHs present in fossil fuels [6]. These strains employ different mechanisms or pathways to metabolize the PASHs. Some of them act by breaking the carbon skeleton of the molecule via a destructive pathway while some specifically cleave the C-S bond via a non-destructive pathway leaving the carbon skeleton intact [7]. However, the rupture of the carbon skeleton in PASHs is undesirable as it leads to the loss of calorific value. Therefore, the organisms capable of desulfurizing PASHs via the non-destructive pathway are of utmost importance and interest for the desulfurization of fossil fuels. Most of the desulfurization studies have used DBT as the model sulfur compound to study the desulfurizing pathways. The '4S' pathway is the most commonly studied non-destructive pathway present in bacteria for the desulfurization of DBT. The desulfurizing bacterial strains belonging to genus *Rhodococcus, Gordonia*, and *Corynebacterium*, and *Mycobacterium* possess the '4S' pathway [2]. Of these, the rhodococci and gordonae are well known for their catabolic diversity and hydrophobic cell surface that enable their use as potential candidates for biotransformation and bioremediation of organic compounds present in solvents [8, 9]. As such they are attractive strains for the desulfurization of the organic PASHs present in the fossil fuels consisting of complex mixture of organic compounds.

1.4 *Rhodococcus erythropolis* and *Gordonia alkanivorans*: interesting desulfurizing bacterial strains

R. erythropolis IGTS8 was the first strain to be isolated for its ability to desulfurize PASHs without affecting the carbon skeleton, nearly two decades ago [10]. Since then several rhodococci strains have been characterized for their ability to desulfurize DBT and its derivatives [11-17]. These strains desulfurize DBT and its derivatives via the non-destructive '4S' pathway by specifically cleaving the C-S bond. However, the activities exhibited by these strains are too low for any commercial desulfurization process. A number of attempts have been made to improve the activities of the wild type desulfurizing strains by optimizing the external conditions. In this regard, several works are available in literature which discuss the effects of various nutrients [18, 19], design of suitable media [20-22], optimization of operational conditions [23, 24], etc. on the desulfurizing activities. However, desired improvement in the desulfurizing activities of these strains has not been achieved by changing the external conditions. Numerous efforts have been attempted at enhancing the flux through the desulfurization pathway by performing various manipulations at the genetic level [25-27] in the rhodococci strains. However, desirable desulfurization rates are yet to be attained. Most of these studies have targeted modifications in the '4S' pathway components for improved desulfurization. The inability of these genetically modified rhodococci strains to give the expected levels of activity can be attributed to the presence of the unknown host functions that may exert a control over the extent of desulfurization possible [5]. As such it is likely that the parts of metabolism other than '4S'pathway need to be engineered for an overall improvement in desulfurization. Therefore, it is critical to study the inter-relationships between the various host

functions in rhodococci that may play crucial roles in the functioning of '4S' pathway. Moreover, although the DBT desulfurization mechanisms [7] and cell growth rates on DBT [28-30] have been reported, a quantitative elucidation of the actual process by which sulfur from DBT gets incorporated into the biomass is still unclear [3]. The lack of such holistic studies limits the desired improvement in the rhodococci for use as an efficient desulfurizing biocatalyst. Besides, most of the desulfurizing rhodococci strains are unable to exhibit activities for other thiophenic compounds such as BT and its derivatives. However, since all these compounds are of relevance for reducing the overall sulfur content of fossil fuels, it is critical to study microbes that possess activity for compounds other than DBT. In this regard, several gordoniae strains capable of desulfurizing various PASHs have been isolated and characterized [31-34]. Most of the gordoniae strains exhibit desulfurization for either BT or DBT and their respective derivatives. The desulfurization studies with gordoniae strains have revealed that DBT and BT are desulfurized via distinct pathways. This observation further emphasizes on the necessity of studying bacterial desulfurization of BT besides DBT. It is therefore desirable to have bacterial strains capable of simultaneous desulfurization of BT, DBT, and their derivatives at higher rates. In this regard, *G. alkanivorans* strain has been found to exhibit desulfurization of BT, DBT and other thiophenic compounds [35]. However, it has been seen that *G. alkanivorans* does not utilize BT and DBT simultaneously but desulfurizes BT preferentially. Moreover, it exhibits lower levels of desulfurization activity. Even the genetically modified gordoniae strains are unable to give the desired levels of activity [36]. Most of these works have emphasized on altering/modifying the specific pathways for desulfurization of DBT/BT but have not considered their interactions with the rest of the metabolism. Furthermore, there are

only a few limited studies available in literature on desulfurization with *Gordonia* strains and a lot of research is needed.

It is clear that more efficient desulfurizing strains are required for taking the biodesulfurization process from laboratory scale experiments to refineries level technology. The design of improved desulfurizing strains requires an understanding of their intracellular metabolic architecture. The quantitative elucidation of the interactions between the various metabolic activities and the desulfurizing pathway can help in determining the other host functions that are likely to control the extent of desulfurization possible. For an effective desulfurization process the desulfurizing strain should be able to simultaneously desulfurize BT, DBT, and their derivatives at high levels of activity. However, as mentioned earlier, *R. erythropolis* is the most widely studied strain for DBT desulfurization while *G. alkanivorans* shows utilization of both BT and DBT but preferentially uses BT. Therefore, it is interesting to study the desulfurization characteristics of the two strains and determine the properties common to their metabolic networks which can enable design of pure culture or co-cultures for enhanced desulfurization of the complex mixtures of various PAHSs. However, most desulfurization studies with *R. erythropolis* and *G. alkanivorans* have targeted the 4S pathway solely, and holistic systems studies of their intracellular activities are not available. Since a cell contains several enzymes and metabolites, any of its phenotype (here, desulfurizing activity) results from the coordinated action of the various enzymes and environmental factors [37]. As such it is critical to study the complex interactions among various gene products, environmental factors, and desulfurization activity. However, the complex nature of the metabolic network in a cell makes it difficult to comprehend such interactions and predict the rate or level of desulfurizing activity under any given condition by looking at individual reactions or enzymes.

Systems biology approach enables development of *in silico* models for holistic representation of the various metabolic activities and quantitative elucidation of their interactions. Therefore, a holistic systems biology approach can be used to study the sulfur metabolism and related metabolic activities in *R. erythropolis* and *G.* alkanivorans in an integrated manner.

1.5 *In silico* modeling

Systems study of metabolic network aims at understanding the cellular phenotypic behavior as an emergent property of the integrated metabolic, signaling, and regulatory networks by interpretation and combination of omics data (genomics, transcriptomics, proteomics and metabolomics) that carry global cellular information [38]. Metabolic phenotypes can be defined in terms of flux distributions through a metabolic network. Interpretation and prediction of the metabolic flux distributions require application of mathematical modeling and simulations. Therefore, systems biology involves the *in silico* modeling to represent the complex gene, protein and metabolite interactions across multiple metabolic and regulatory networks [39]. With the availability of whole-genome sequences, it has become possible to reconstruct *in silico* models representing the genome-scale biochemical reaction networks in microorganisms. Although the *in silico* dynamic models give more accurate metabolic and regulatory information, they are currently limited by unavailability of detailed kinetic data. However, the known stoichiometry of reactions based on the genome annotations can be used to construct *in silico* steady state models of cellular metabolism [40]. Such stoichiometric models can be analyzed using linear programming approach to determine possible flux distributions through the metabolism of desulfurizing biocatalysts under defined conditions. Thus, they can provide an estimate for various

phenotypes possible. Moreover, as the genetic manipulations of desulfurizing strains are carried out to enhance flux through the desired '4S' pathway, the *in silico* experiments using such models can provide an estimate of achievable flux.

Fong et al. [41] and Chin et al. [42] have demonstrated the usefulness of *in silico* genome scale stoichiometric model of *Escherichia coli* to successfully predict the effects of genetic manipulations on cellular metabolic characteristics by carrying out *in silico* experiments. Similarly, Saha et al. [43] have developed the maize model that quantitatively represents the metabolic properties of the otherwise complex network of the member of the plant family. Several genome scale models (>50) of various organisms [44-46] have been developed and analyzed to carry out various *in silico* studies for studying their intracellular metabolism and generating testable hypotheses. Numerous useful algorithms and procedures [47-49] have been developed to reconstruct and analyze such *in silico* models of various organisms ranging from the unicellular bacteria to multicellular plants. In a similar way, mathematical modeling can be used to complement the experimental works for designing desulfurizing biocatalysts with improved performance. In turn, the wet lab experiments can be used to generate data for modification of model structure to enable more accurate and validated *in silico* models that can give meaningful interpretations and predictions. Thus, *in silico* models provide useful insights into the cellular systems in an iterative manner.

1.6 **Research objectives**

This research focuses on the systematic study of the two desulfurizing strains, *R. erythropolis* and *G. alkanivorans,* using *in silico* models for an improved understanding of their metabolic networks. This thesis aims at the reconstruction and

analyses of the genome scale models for *R. erythropolis* and *G. alkanivorans* to get insights into their metabolism that can help in addressing the existing questions regarding these desulfurizing strains. For instance, the existing literature has detailed documentations available for the desulfurization pathways and cell growth on DBT. However, it is not clear how the sulfur from DBT actually gets incorporated into the biomass of the above mentioned desulfurizing strains [3]. Furthermore, in spite of continued efforts to improve desulfurization activities of the strains using different genetic engineering strategies, the desired levels are yet to be obtained. In addition, this work also aims at providing explanations for several inexplicable observations reported in literature such as, i) excretion of sulfate as only sulfur product while sulfite is known to be the end product of the desulfurization pathway [3], ii) contrasting reports on the repression of desulfurization in the presence of readily bioavailable sulfur compounds such as sulfate, methionine, cysteine, etc. [50, 51], iii) lack of knowledge of host functions and intracellular factors in the desulfurizing strains that may control the extent of desulfurization activity that they can exhibit, and iv) inability of the strains to exhibit simultaneous desulfurization for a broader range of PASHs.

In this work, we have used a systems-based approach to develop *in silico* models of *R. erythropolis* and *G. alkanivorans* to study their intracellular metabolism and to elucidate the role of other metabolic activities in controlling the flux through the desulfurization pathway. To this end, we first holistically study the sulfur metabolism in *R. erythropolis* to explicate the incorporation of sulfur atom from DBT into the biomass. Next, we reconstruct the *in silico* genome scale metabolic models of *R. erythropolis* and *G. alkanivorans* to study their metabolism and desulfurizing characteristics. These models provide a better understanding of the properties of their metabolic network and the dependence of desulfurization pathway on other metabolic

activities. Further, the *in silico* experiments with these models help in the design of media and modified strains for enhanced desulfurization. In addition, we reconstruct an *in silico* model for the co-culture of *R. erythropolis* and *G. alkanivorans* for simultaneous desulfurization of BT and DBT when present in mixtures. Finally, we give a few testable hypotheses for design of improved strains that can give desired levels of desulfurization activity.

1.7 **Thesis outline**

This thesis consists of nine chapters. Chapter 1 gives a brief introduction to the problems associated with biodesulfurization and the approach that we use to address them. It is followed by Chapter 2 that provides a detailed review of the desulfurization works that are available in literature with an emphasis on studies done with *R. erythropolis* and *G. alkanivorans* strains. We identify the gaps and other inexplicable observations that exist in the currently available work on biodesulfurization which need to be addresses by future works.

Chapter 3 reports a flux-based model of sulfur metabolism reconstructed to understand and analyze sulfur utilization by *R. erythropolis*. The model gives a quantitative elucidation of the process by which the sulfur from DBT gets incorporated into the biomass.

Chapter 4 reports the reconstruction of an *in silico* genome-scale metabolic model for *R. erythropolis* using the available genomic, experimental, and biochemical information. The model has been validated and analyzed using the known phenotypes.

Next, chapter 5 covers our investigation on the possible routes taken by the sulfur from DBT to convert into biomass or other metabolites. With our detailed

analyses of the genome-scale model of *R. erythropolis*, we determine and suggest the host functions that can be manipulated for obtaining enhanced desulfurization activity.

Chapter 6 presents a genome scale metabolic model of *G. alkanivorans*. We have successfully validated our model using the experimental data available in the literature. The analyses with our model provide insights into the roles of various metabolic components for the survival and desulfurization activity of *G. alkanivorans*.

Chapter 7 includes an expanded and improved model of *R. erythropolis*. This model includes additional functionalities and thermodynamic constraints for improved predictability. Further, this expanded model has been used to design and study an *in silico* modified strain of *R. erythropolis* for bio-refining i.e., simultaneous biodesulfurization and biodenitrogenation of fossil fuels.

Chapter 8 gives details on an *in silico* metabolic model of a co-culture of *R. erythropolis* and *G. alkanivorans* for efficient simultaneous desulfurization of BT and DBT in a mixture.

Finally, chapter 9 highlights the major findings and contribution of this research work. It also includes some of the potential research works that can be done as an extension of this work for improved understanding and design of efficient desulfurizing strains.

2 Literature review

2.1 Biological desulfurization of fossil fuels

2.1.1 Need for fossil fuels desulfurization

The advancement of human civilization and industrialization has led to an ever increasing consumption of fossil fuels to meet the energy requirements. The fossil fuels contain various aliphatic, cyclic and aromatic hydrocarbons consisting of various heteroatoms. Thus, the combustion of fossil fuels for energy, leads to the release of various pollutants including unburnt hydrocarbons, and oxides of carbon, nitrogen, and sulfur. The increasing awareness for environmental protection has drawn attention towards the harmful effects of the oxides of sulfur (SO_x) released on combustion of fossil fuels [1, 2, 52, 53]. . The various detrimental effects of SO_x on the environment and human health are stated as follows:

- They make the air hazy or foggy and thus, reduce visibility
- They exert ill-effects on breathing and cause respiratory illness, alterations in pulmonary defences, aggravation of existing cardiovascular disease, and chronic lung diseases
- SO_x tend to combine with the moisture present in air to form sulfate aerosol with extremely small size (~ 2.5μm) which may be easily inhaled and cause respiratory illness

- The acid rain produced from SO_x damages the foliage of trees and agricultural crops. It also leads to the acidification of lakes and streams thereby, adversely affecting the aquatic life
- The acid rain also damages manmade objects such as stone statues, buildings, and monuments

Moreover, SO_x formed on the combustion of fossil fuels, poison the catalytic convertor on the automobile exhaust system which are installed to combust unburnt hydrocarbons that otherwise are a major source of pollution [6]. The control of SO_x emissions is also essential for the new generation engines that are equipped with storage catalysts to capture oxides of nitrogen as the SO_x poisons these catalysts [54]. Considering these harmful effects and environmental threats posed by SO_x, the governments have laid stringent regulations to limit the sulfur levels in the road and non-road fuels. According to Environment Protection Agency (EPA), the allowable fuel sulfur levels in most non-road diesel fuel to be used 2010 onwards was reduced to 15 ppm [1, 55] while the step to 15 ppm sulfur fuel control for locomotive and marine diesel fuel has been implemented since the beginning of 2012 [56]. These regulations have made desulfurization i.e., removal of sulfur, an important step in the pre-processing of the fuels. Another motivation for the desulfurization of fossil fuels arises from the increasing use of fuel cells, which require almost zero-level sulfur fuels [57-59]. The problem of satisfying all these demands for lowering the sulfur content in fossil fuels is further aggravated by the fast depletion of the light crude oil pool which has forced the use of high density crude that usually is richer in sulfur [60]. This is because sulfur tends to concentrate in the larger, denser, and more complex hydrocarbon molecules that are abundant in heavier crude oils. Therefore, the escalating production of crude oil and stringent demands for use of cleaner fuels

require the efficient methods that can reduce the sulfur content of the fossil fuels to the desired level. In a nutshell, there is clear demand for efficient and effective methods for desulfurization of fossil fuels to comply with the government regulations. The prevalent method of desulfurization is hydrodesulfurization.

2.1.2 **Hydrodesulfurization**

Hydrodesulfurization (HDS) is a chemical process which is currently used for the desulfurization of fossil fuels. It involves the reduction of sulfur atom present in compounds of sulfur to H_2S on $CoMo/Al_2O_3$ or $NiMo/Al_2O_3$ catalyst in the presence of H_2 gas, at high pressure (150-250 psi) and high temperature (200-425°C) [1, 61]. The H_2S gas generated in the process is later catalytically oxidized to elemental sulfur. The HDS method can easily remove aliphatic sulfur compounds from the fossil fuels but it is unable to desulfurize certain refractory poly aromatic sulfur heterocycles (PASHs) such as thiophene, dibenzothiophene (DBT), benzothiophene (BT), and their derivatives. Thus, in order to meet the regulations to produce low level sulfur fuels, the removal of such compounds by HDS requires even higher operating temperature and pressure, longer residence time, installation of new units in the existing ones, stronger reaction vessels, and certain additional facilities all of which require a heavy load of capital cost. Moreover, the HDS of PASHs under hydrogen atmosphere leads to the hydrogenation of olefinic compounds and reduction in their calorific value. In order to increase the calorific value, the hydrodesulfurized stream has to be sent through a fluid catalytic cracking system which further adds to the fuel processing costs. It is thus clear that HDS, the current chemical method employed to remove sulfur, is very expensive and requires extreme operating conditions to obtain the desired extent of

desulfurization. Furthermore, like most of the chemical processes, HDS is non-specific in action and besides removing sulfur it also removes:

- The polar compounds of oxygen and nitrogen that add to the lubricating value of the fuel
- The aromatics that are responsible for fuel ignition quality
- The naturally occurring anti-oxidants that contribute to both the physical and thermal stability of the fuel

Thus, these deprivations in the hydrotreated fuel lead to the problems like excessive exhaust temperatures, thermal stress on engine and excessive deposits besides the improper diesel engine operation [62]. Also as the fuels for the ships are stored for prolonged uses, the poor physical and thermal stability are likely to lead to their accelerated degradation. As a result of which, upon combustion they are likely to form large amounts of unburnable, heavy carbon compounds that lead to heavy smoke. Thus, there is a clear demand for inexpensive desulfurization methods that can specifically lower the sulfur content in the fuels without degrading their lubrication properties and stability. As such alternate inexpensive technologies are being studied and developed to produce ultra-low level sulfur fuels. A potential candidate technology that has gained much attention over the past two decades is biodesulfurization.

2.1.3 **Biodesulfurization**

Biodesulfurization is a process that involves the use of the biocatalysts (i.e., whole microbial cells or enzyme extracts) [63] to reduce the amount of sulfur present in fuels. It is based on the realism that sulfur is an integral constituent of several cofactors, amino acids and other precursor metabolites that are essential for the proper metabolic functioning and growth of microorganisms. As such the microorganisms have evolved

diverse metabolic pathways to utilize various complex sulfur compounds including the PASHs. The potential of using biocatalysts for desulfurization arises from their following properties:

- More specific and selective in action
- Environmentally friendly
- Operate at ambient conditions
- Transform a larger number of substrates

Further, their use decreases the energy costs and eliminates the problem of undesirable side reactions and by-products formation [64-68]. Thus, a successful biodesulfurization process will provide the additional benefits of lower capital and operating costs for the removal of sulfur from fossil fuels in a greener way [68].

2.1.3.1 **Types of sulfur compounds**

Sulfur is the most abundant element in petroleum after carbon and hydrogen [69]. Heavy oils found all over the world are viscous and contain 3 to 6% of sulfur [70]. The total sulfur content of the fossil fuels is contributed by the presence of various inorganic as well as organic compounds containing sulfur. The inorganic sulfur compounds include elemental sulfur, hydrogen sulfide, and pyrite that are present either in dissolved or in suspended form. The group of organic sulfur compounds includes both aromatic and aliphatic forms of thiols, sulfides, and heterocycles shown in Fig. 2.1. The inorganic sulfur compounds can be efficiently removed using the physical methods but the removal of sulfur from organic compounds requires an intensive chemical treatment at extreme conditions of temperature and pressure. The reactivity scale for different sulfur compounds at high temperature of about 622 K is given by: non-aromatic sulfides > thiophenes ≈ benzothiophenes >> extended five-

rings thiophenes ≈ six-ring thiophenes > benzonapthothiophenes ≈ compact five-ring thiophenes > C_0/C_1 dibenzothiophenes >> C_{2+} dibenzothiophenes > phenanthrothiophenes [71]. Due to this difference in reactivity some of the organic sulfur compounds such as benzothiophene (BT), dibenzothiophene (DBT) and their derivatives are considered refractory or recalcitrant as they cannot be easily removed from the crude using the conventional HDS process. They are quite stable and require more intensive processes and conditions for their desulfurization. DBT contains a hindered sulfur atom which is removed with difficulty and acts as a simple representative molecule for the various recalcitrant PASHs. Thus, DBT is used as the model compound in most of the biodesulfurization studies.

2.1.3.2 **Biotechnology for desulfurization**

A living cell is mainly made up of the following elements: carbon, hydrogen, oxygen, nitrogen, phosphorus sulfur and certain trace elements. Sulfur accounts for about 0.5-1% of the bacterial dry cell weight [2]. It is present in two of the amino acids (cysteine and methionine), some of cofactors (such as coenzyme A, thymine, acyl carrier protein, etc.), disulfide bonds present in proteins, and certain intermediate metabolites that play important roles in the nutrition, maintenance, and growth of an organism. As such sulfur uptake from different sources is clearly essential for the survival and growth of the microorganisms for which they have evolved diverse biochemical mechanisms.

Several microorganisms, belonging to different genera, have been identified that have the ability to remove sulfur atoms from the BT and DBT molecules [2]. These microorganisms vary in their desulfurizing characteristics such as pathways, physiology, activities, specificities, and stabilities. The bacterial strains that have been

isolated and studied for the purpose of desulfurization belong mainly to the genus *Pseudomonas*, *Rhodococcus*, *Gordonia*, *Mycobacterium*, *Sulfolobus*, *Nocardia*, *Desulfovibrio*, *Desulfotomacaculum*, *Arthobacter*, etc. They differ in the manner that they utilize the sulfur compounds.

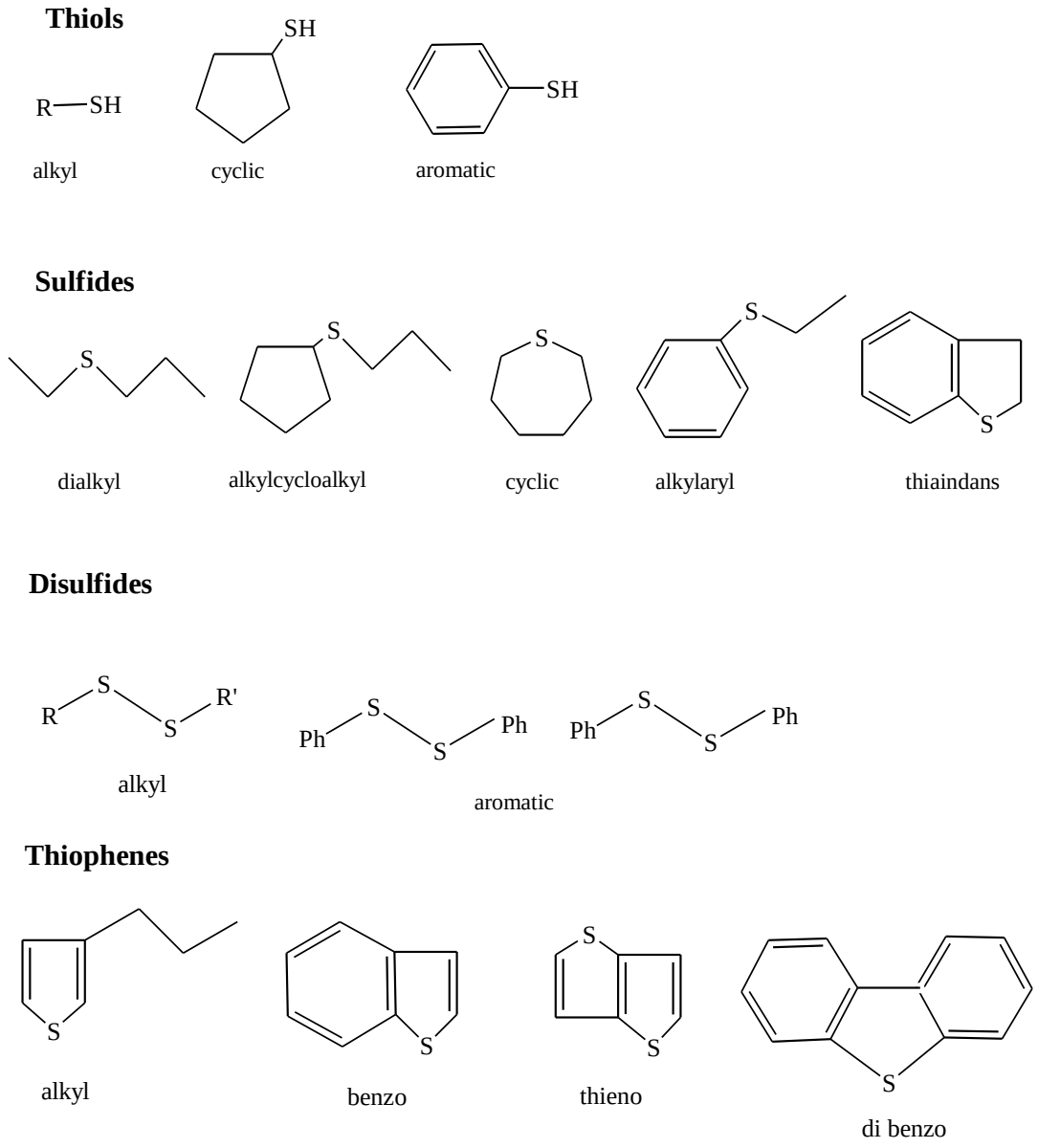


Fig. 2.1 Various organic sulphur compounds found in fossil fuels (adapted from [72] and [2])

2.1.3.3 **Mechanisms of biodesulfurization**

The desulfurization mechanisms in various desulfurizing bacterial strains have been studied using DBT as the model compound. They differ in the pathways/mechanisms that they employ to metabolize DBT. Some of them do it by reducing the sulfur atom while others oxidize it. Some of them break the compound into fragments via a destructive pathway while others do it non-destructively. The two most widely studied biodesulfurization mechanisms are Kodama pathway (destructive) and 4S pathway (non-destructive). The details of these pathways are as follows:

2.1.3.3.1 *Kodama pathway (destructive):*

This pathway was proposed by Kodama et al. [72] in a study to analyze the products of DBT metabolism by the different strains of *Pseudomonas* capable of metabolizing DBT [73]. It involves three main steps: hydroxylation, ring cleavage and hydrolysis as shown in Fig. 2.2. *Pseudomonads* usually exhibit this pathway for DBT metabolism, in which the aromatic rings of DBT are oxidized to water soluble 3-hydroxy-2-formyl-benzothiophene (HFBT) through the oxidative C-C bond cleavage [72-74]. However, the sulfur is not mineralized to inorganic form in this pathway. Although some strains of *Pseudomonas* have also shown biodegradation of HFBT but no sulfur is seen to be released [75]. Therefore, this pathway is more of a hydrocarbon degradation pathway involving C-C bond cleavage while the C-S bond remains intact in one of the fragments. Moreover as the resulting products are water soluble, it results in undesirable loss of calorific value of the fossil fuel.

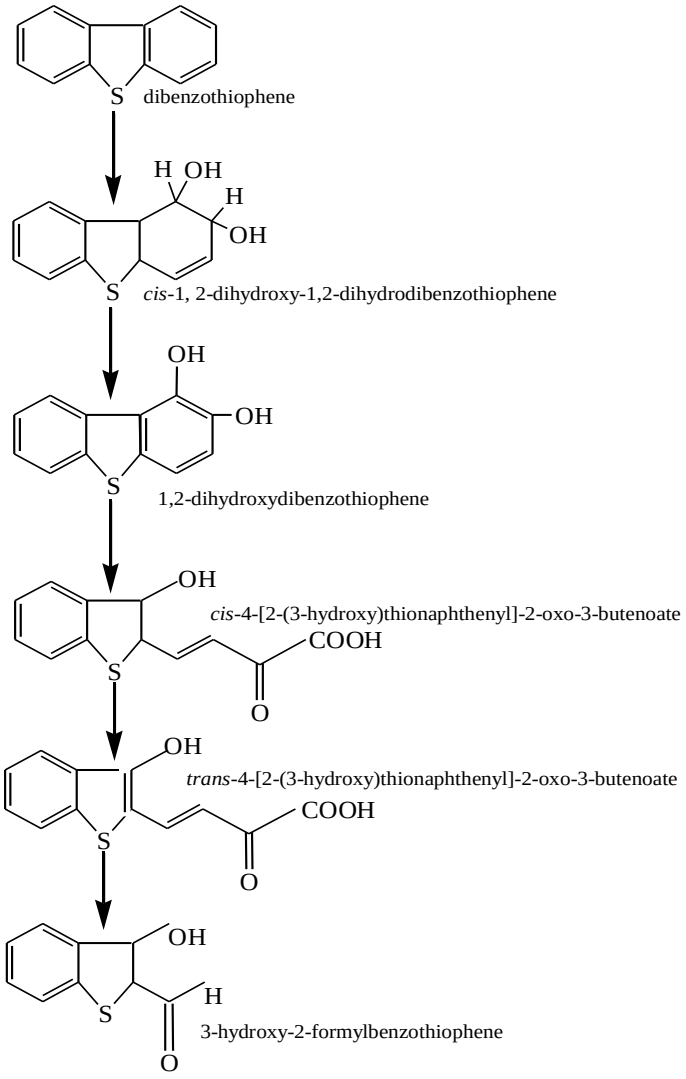


Fig. 2.2 Kodama Pathway

2.1.3.3.2 *4S pathway (non-destructive)*

This is a sulfur specific pathway as it leads to the cleavage of C-S bond present in DBT without affecting the carbon skeleton as shown in Fig. 2.3. Kilbane first proposed the 4S pathway (sulfoxide/ sulfone/ sulfonate/ sulfate) consisting of a set of four thermodynamically favorable reactions [76]. The first two steps are catalyzed by a monooxygenase, DszC, which oxidizes DBT to sulfone via sulfoxide. The sulfone is then oxidized to sulfonate by the action of another monooxygenase, DszA. It has been reported that the activities of DszC and DszA enzymes require the coordinated activity of an FMN:NADH oxidoreductase enzyme, DszD which supplies them with reduced

flavin mononucleotides for their oxidation reactions. The final step involves the action of a desulfinase, DszB which removes the sulfur atom from sulfonate and results in the production of sulfite and 2-hydroxybiphenyl (HBP), the final desulfurized product. Since then several strains of *Rhodococcus* and other genus such as *Corynebacterium*, *Gordona* and *Mycobacterium* have been reported to exhibit similar desulfurization pathway [2, 5].

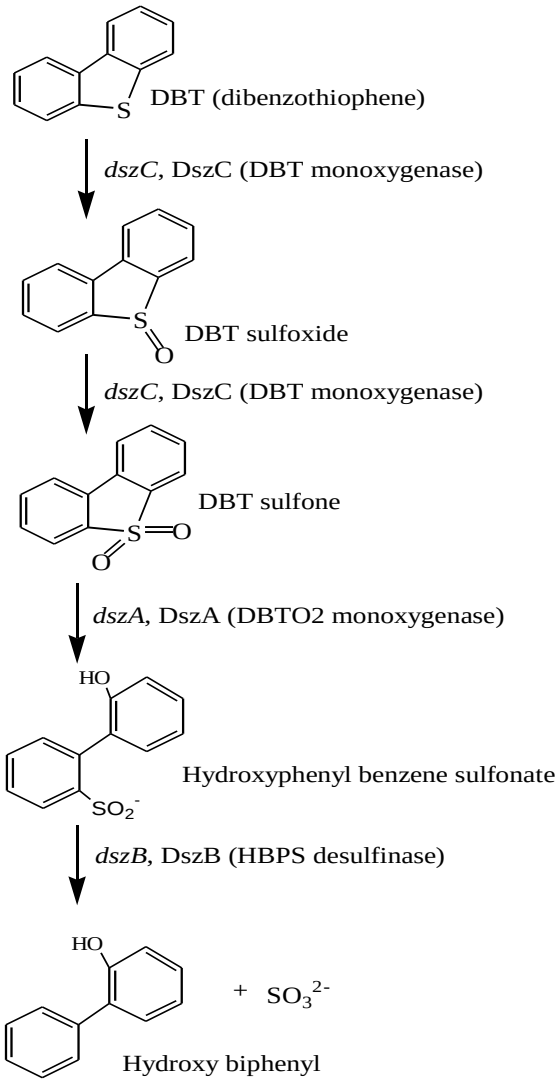


Fig. 2.3 4S Pathway

2.1.3.3.3 *Other mechanisms:*

There are several other mechanisms via which the other organisms utilize DBT. However, they have not been well studied due to lack of complete information. Certain

microorganisms such as *Arthrobacter* [77] and *Brevibacterium* [78] utilize DBT by mineralizing it to CO_2 and sulfite. Various anaerobic strains belonging to *Desulfovibrio, Desulfomactum* and *Thermodesulfobacterium commune* [79] are able to cleave C-S bonds with subsequent reduction of sulfur to hydrogen sulfide under reducing atmosphere. Similarly, some fungi have also been studied for their potential application in biodesulfurization. The versatile *Cunninghamella elegans* [80] and the white rot *Pleurotus ostreatus* [81] have the ability to convert DBT to DBT sulfone but they do not show any further transformation.

Analysis of various DBT metabolic pathways by different microorganisms reveals that the 4S pathway is of utmost interest. It enables desulfurization of DBT, the representative of the most recalcitrant sulfur compounds present in fuels, without rupture of the carbon skeleton. Thus, much research work in biodesulfurization has been focused on microorganisms exhibiting the 4S pathway.

Rhodococcus erythropolis IGTS8 [10] was the first microbial strain isolated with the ability to specifically desulfurize polyaromatic sulfur heterocycles (PASHs) such as dibenzothiophene (DBT) and its derivatives *via* the 4S pathway. Since then several *Rhodococci* strains have been isolated, characterized and genetically modified to obtain improved biocatalysts for specific desulfurization. Moreover, faster growth, simpler developmental cycle [82], environmental persistence, and the remarkable ability of the *Rhodococci* strains to degrade diverse organic compounds and produce surfactants make them highly attractive for biotransformation and bioremediation purposes. Also the presence of mycolic acids facilitates the uptake of hydrophobic compounds and confers hydrophobicity to surface of rhodococci cells. This hydrophobicity enables them to attach to the oil/water interface while growing in aqueous-hydrocarbon system and hence, the uptake of hydrophobic compounds such

as DBT from the oil [6]. Thus, they appear as microbial systems of choice, for biodesulfurization.

2.1.3.4 **Biodesulfurization studies with rhodococci strains**

R. erythropolis IGTS8 was first isolated from a mixed culture soil sample by Kilbane et al. [10]. It was initially recognized and registered as *R. rhodochrous* strain capable of specific C-S bond incision activity. However, later in 1997, the analysis of the nucleotide sequence of 16rRNA from this strain showed it to be *R. erythropolis* IGTS8 (GenBank accession no. AF001265) [2]. Gallagher et al. [83] cultured IGTS8 on various sulfur sources including sulfate, DBT, DBT sulfoxide and sultone. They found similar exponential growth trends with all these sulfur sources however, a richer medium supported faster growth. The realization of specific desulfurization capability of microbial cells encouraged researchers to explore the possibility of developing biodesulfurization as a full-fledged technology. This necessitated the work to gain insights into the molecular mechanism of the 4S pathway.

Considerable research has been conducted on molecular biology of IGTS8 and other similar strains, in order to develop a better understanding of the metabolic machinery involved in the specific sulfur removal. Denome et al. [84] reported that the sulfur oxidation of DBT is controlled by a cluster of three genes. The three desulfurizing (*dsz)* genes, named as *dszA*, *dszB*, and *dszC* [85], are organized on an operon present on a plasmid. They are transcribed in the same direction and are under the control of the same promoter [86]. The plasmid is stable and is not lost on heat treatment. However, when the cells are exposed to UV light for nearly 20 minutes, the desulfurizing activity is lost [84]. It is found that a cluster of three genes (*dsz* ABC) located on the same plasmid is responsible for the specific desulfurization activity in

several *Rhodococcus* sp. strains including IGTS8, X309, X310, If, Ig, and Ih [85, 87]. The location of dsz genes on an endogenous plasmid pointed out to the possible distribution of specific desulfurization activity in other soil microorganisms and studies have been directed on characterizing genes and their encoded proteins [88, 89]. The comparison of *dsz* genes and their encoded proteins against various databases such as GenBank 83, EMBL 39 and Swiss-Prot 28 did not reveal any significant homology [90]. As such it was deduced that they encode enzymes that are particularly employed for desulfurization and are not associated to other commonly known metabolic pathways. The *dsz* genes encode for three enzymes, Dsz A, Dsz B, and Dsz C, of the 4S pathway. These enzymes have been isolated from *R. erythropolis* IGTS8 [7] and *R. erythropolis* D-1 [91] and studied.

The *dszC* gene encodes for Dsz C (DBT monooxygenase) which catalyses the oxidation of DBT to DBT sulfone (DBTO2) in a two-step process with DBT sulfoxide (DBTO) as the intermediate compound [14]. Isotopic labeling studies indicated that the two oxygen atoms are derived from molecular oxygen. The Dsz A enzyme is a monooxygenase encoded by the *dszA* gene. It catalyzes the oxidation of DBTO2 to 2-(2'-hydroxyphenyl) benzene sulfinate (HPBS). The third enzyme of the 4S pathway is a desulfinase, DszB, encoded by *dszB* gene. The desulfinase catalyzes the conversion of HBPS to the final desulfurized product 2- hydroxybiphenyl (HBP). The kinetic analysis of the DBT desulfurization revealed that the rate of desulfurization is limited by the last enzyme in the pathway, which catalyzes the conversion of HBPS to HBP [7]. Further studies have shown that the two monooxygenases require co-ordinated activity of a NADH: FMN oxidoreductase enzyme designated as Dsz D. The *dszD* gene [92, 93] which carries genetic information for oxidoreductase Dsz D is present on the chromosome rather than the plasmid. The monooxygenases essentially require

reducing equivalents for their activity; as a result, their coupling with oxidoreductase becomes critical. Ohshiro et al. [94] searched for non-DBT desulfurizing micro-organisms producing a flavin reductase which could couple more efficiently with DszC than that produced by the DBT-desulfurizing bacterium *R. erythropolis* D-1, and found *Paenibacillus polymyxa* A-1 to be a promising strain. The identification of such coupling reactions may be beneficial in developing an efficient enzymatic desulfurization system and in elucidating the mechanisms of interaction between the monooxygenases, reductases and reduced flavin species.

The identification of the enzymes and genes responsible for specific desulfurization trait further enhanced the interest in determining the conditions under which the desulfurizing rates would be altered or affected. In literature various reports on the effect of the various substrates, products, media compositions, and operational conditions on desulfurizing activities and growth rates have been reported. Omori et al. [16] reported that the cells grown in the presence of sulfate did not exhibit the desulfurizing activity. Moreover, the removal of the by-product sulfate from the medium of desulfurizing strain leads to increase in the desulfurizing activity by nearly 14%. It has been reported in the literature [95] that although sulfate represses the expression of *dsz* genes, it does not exhibit any inhibitory effect on the Dsz enzymes. Such studies point to the fact that, inorganic sulfate is one of the simplest sulfur sources and is preferred for growth by most micro-organisms. However, under the sulfate-limited conditions certain bacteria are able to synthesize enzymes and transport systems involved in scavenging and metabolizing alternative sulfur sources from the environment [96]. The Dsz enzymes probably fall under this category of enzymes.

The repression of desulfurization activity in various bacteria in the presence of readily bioavailable sulfur sources such as sulfate, methionine, cysteine, taurine,

methanesulfonic acid and casamino acids have been observed by various researchers [90, 97, 98]. Li et al. [50] reported that the expression of *dsz* genes was necessary for IGTS8 to carry out desulfurization of DBT and its derivatives. They found that the desulfurizing activity of IGTS8 decreased with the increasing concentration of sulfur containing compounds such as sulfate, methionine, and cysteine in the medium. They also reported that in presence of these sulfur compounds, the transcription of *dsz* genes is repressed probably by the binding of some unidentified repressor protein to the Dsz promoter, which lies upstream of the *dszA* gene.

In the literature, several other *Rhodococci* strains such as SY1, D-1, H2, KA2-5-1, I-19 and SHT87 strains [2, 11, 16, 99] have been isolated and characterized for their ability to specifically desulfurize various PASHs without cleavage of the C-S bonds at ambient conditions. These cells can utilize the DBT as the sole sulphur source and give HBP as the final desulfurized product. The growth of the cells is concomitant with the accumulation of the HBP. It has been reported that the resting cells of various desulfurizing strains exhibit higher rates of desulfurization than the growing cells [11, 15, 99]. However, if microorganisms are to be used as nongrowing biocatalysts, they need to be grown in high densities in a separate process. Furthermore, since the specific activity of desulfurization and cell growth are crucial indicators of the biocatalyst efficiency, knowledge about how to increase the specific activity is very important from the viewpoints of both scientific interest and practical applications. Honda et al. investigated the suitable media composition for obtaining high cell densities of IGTS8 [24]. They found that acetate and sulfate support higher growth as carbon and sulfur sources respectively. However, the desulfurization activity of the cells is repressed in the presence of sulfate but it can be induced after a few hours of incubation of cells on DBT. Further, it was found that the desulfurized product, HBP

exerts toxic effect on the cells. In a similar study with *R. erythropolis* KA2-5-1, Yan et al. [18] observed that ethanol supports a higher cell density and higher desulfurizing activity in comparison to glucose and glycerol. Ohshiro et al. [15, 100, 101] carried out studies on desulfurization of DBT by *R. erythropolis* H-2 in various hydrophobic solvents. They observed that H-2 can convert DBT dissolved in hydrophobic phase to HBP. Relatively long chain alkanes (*e.g.* hexadecane, tetradecane, etc.) support a higher desulfurizing activity of the strain than the shorter alkanes while the hydrocarbons such as *p*-xylene, styrene and toluene are toxic to the strain. The H-2 strain is also able to specifically desulfurize the derivatives of DBT such as 3,4-dibenzoDBT, 2,8-dimethylDBT and 4,6-dimethylDBT. However, most of the *Rhodococci* strains isolated cannot desulfurize benzothiophene (BT). Matsui et al. isolated first *Rhodococcus* sp. strain T09 capable of desulfurizing BT and various alkylated BTs such as isomers of 2-methylbenzothiophene, 2-ehtylbenzothiophene and 2,7-diethylbenzothiophene which are otherwise found resistant to desulfurizing activities of other *Rhodococcus* species but the growth rate is too slow for substituted compounds [102].

Most of these studies had aimed at the desulfurization of DBT, the model sulfur compound, in either an aqueous phase or biphasic media using the hydrocarbons as the model oil. The desulfurizing studies using different derivatives of DBT revealed that the cells preferably desulfurize lesser substituted derivatives such as methyl DBTs, ethyl DBTs than the di-substitutes and tri-substituted DBTs [103] However, in the real scenario the actual material to be desulfurized is petroleum, a complex mixture of hydrocarbons with thousands of different sulfur containing compounds. Therefore, the study of the ability of the microorganisms to desulfurize the crude oil becomes critical. Some of the biodesulfurization studies have been carried out using samples of different

distillation fractions of the crude oil but the desulfurizing activities obtained different are too low [104]. It has been further reported that the microorganisms are intolerant to the crude oil probably due to the toxic effects of the various organic components present in it. Moreover, as most of the microbial cells essentially require water for their existence and growth, the use of an aqueous-oil biphasic medium has to be employed to carry out desulfurization of the oil. In *Rhodococcus*, the Dsz enzymes are soluble and presumably found in the cytoplasm in contrast to the enzymes which carry out metabolism of alkanes and are associated with the cell membrane [6]. The use of a biphasic medium poses mass transfer limitations of the DBT and other PASHs from the oil phase to the microbial cells in the aqueous phase. The production of biosurfactants by the *rhodococci* or the addition of surfactants such as tween 80 [105] helps overcome this problem by formation of emulsions. Although the fine emulsions give increased biodesulfurization yields [106] however, the separation of the hydrocarbon phase from the aqueous phase takes longer and it is difficult to recover the hydrophobic biocatalyst without loss. Moreover, it is observed that the specific desulfurizing activity usually increases with the decreasing oil-water phase ratio [107, 108]. Therefore, the use of biphasic medium for desulfurization studies is undesirable but it is necessitated by the requirement of water by the microbial cells for their growth.

It is known that the enzymes require lesser water than the whole cells for their activity. In addition, the use of isolated enzymes is supposed to be advantageous as it avoids undesirable formation of any by-products. The additional metabolic activities of the whole cell may result in degradation or consumption of the other organic components present in the oil. However, the cell-free extracts were found to exhibit a lower activity, in the order of 0.01 g of DBT removed /(g protein – h), as compared

with 0.4 g/(gdcw- h) using whole cells [107]. Moreover, since the 4S pathway involves a complex enzyme system and requires cofactors generation, the use of purified enzymes is difficult in comparison to the use of whole cells [3]. Also the extraction and purification of enzymes and replenishment of cofactors make the process expensive and complicated. Therefore, use of whole cells for biodesulfurization is preferred over enzymes.

In literature, several studies using immobilized cells have been reported with an objective to overcome the limitation of low oil to water phase ratio and separation. Naito et al. [109] immobilized the cells of KA2-5-1 by entrapping them with different materials such as calcium alginate, agar, photo-cross linkable resin prepolymers (ENT-4000 and ENTP-4000) and urethane prepolymers (PU-3 and PU-6) to carry out biodesulfurization of DBT in *n*-tetradecane as the model compound. It was found that ENT-4000-immobilized cells exhibit highest desulfurization activity than the other cells for as long as 900 hours without any considerable leakage. The desulfurization activity by the immobilized cells is exhibited in the absence of water and the separation is easily achieved. However, they reported that the average desulfurizing rate is lower than that obtained with the free cells in aqueous-oil biphasic medium. They suggested that the probable reason behind the decrease in the reaction rate may be the resistance offered by the support. Ansari et al. [110] studied DBT desulfurization using IGTS8 cells decorated with Fe_3O_4 nanoparticles. The decorated cells showed nearly 56% higher desulfurization rates. They further showed that the nanoparticles increased the permeability of a model black lipid membrane and inferred that the increased permeability of the cell membranes and thereby, enhanced intake of DBT and release of HBP is required for better desulfurization activities. In addition,

the magnetic separation of the decorated cells is easier and it greatly facilitated their recovery and reuse.

The main objective of biodesulfurization research is to develop a commercial biodesulfurization process for petroleum, and it has been estimated that a successful commercial process would require a biocatalyst with desulfurization activity of 1.2 mM DBT/g DCW/h [6] to 3 mM DBT/g DCW/h [111]. However, the desulfurization activity of naturally occurring bacterial cultures is low in comparison to the requirements of a commercial process. The manipulation in the basic genetic makeup of the bacterial system is a key to increase the metabolic flux through the bacteria. Moreover, since the naturally occurring bacteria exhibit desulfurizing activity for a limited number of sulfur compounds, genetically engineered strains may be designed with specifity for a wider range of sulfur compounds. Various studies using different genetic engineering approaches have been reported to design recombinant biocatalysts with improved desulfurization activity [26, 103, 112]. Further improvements in the activity may be achieved by modifying the Dsz enzymes themselves, but this has not been reported so far.

Folsom et al. [103] designed a genetically engineered strain of *R. erythropolis* IGTS8, I-19 that contained multiples copies of *dsz* genes. They observed that the overexpression of key Dsz enzymes increased the specific rates of DBT desulfurization by nearly 25-folds over those for the wild type. The desulfurization of partially hydrodesulfurized middle distillate with I-19 gave a 70% reduction in total sulfur content as compared to 30% reduction obtained for straight-run middle distillate employing wild type strain, ECRD-1. However, the analysis of the untreated and treated sample revealed that DBT and C1-DBT were preferentially attacked, followed by the more substituted DBTs upto C5-DBTs. The difference in reactivity of the

substituted DBT has been attributed to the steric hinderance due to the carbon atom adjacent to the sulfur atom. Hirasawa and colleagues [27] designed recombinant KA2-5-1 strain containing multiple copies of plasmid containing the *dszABC* and *dszD* genes with the same promoter and regulatory region as the wild type. The recombinant strain exhibited four folds higher DBT desulfurizing activity than the parent strain. The treatment of the light gas oil with this recombinant strain also results in a higher decrease in total sulfur content than the parent strain but again the desulfurizing activity is lower than the desired level and the highly alkylated DBTs (> C5) are not desulfurized. However, in the presence of the sulfate the desulfurizing activity is completely repressed in both the parent and the recombinant strain. Therefore, some researchers [26, 97] have designed recombinant strains using promoter deletion and replacement strategies to overcome sulfate repression. Noda et al. [97] used promoter deletion and replacement approach to overcome the sulfate repression in KA2-5-1 cells. They replaced the dsz promoter with a sulfate non-repressible putative promoter, *kap1*. They speculated that the desulfurizing activity of the cells was not affected by up to 1mM sulfate. Furthermore, the desulfurizing activity for DBT as well as light gas oil [113] was two folds greater than the parent strain. However, Franchi et al. [26] observed that the activity of the promoter-replaced strains was also reduced in the presence of sulfate by nearly 50%. Therefore, they postulated sulfate to exert additional regulatory effects on desulfurization. In another study [51] it was reported that the KA2-5-1 mutant, with the interruption in *cbs* (cystathionine β synthase) gene, exhibited desulfurizing activity in the presence of sulfate and methionine.

A successful biodesulfurization process requires a biocatalyst that exhibits higher desulfurization rates for a broader range of sulfur compounds with higher stability. Matsui et al. [114] developed a recombinant strain of BT desulfurizing

Rhodococcus sp. strain T09 by introducing the *dsz* genes operon from KA2-5-1. It was observed that the recombinant strain was capable of utilizing BT, DBT as well their derivatives as the sole sulfur source in different experiments as opposed to the wild type T09. The recombinant plasmid was stable even after 30 generations in 100% of the viable cells and the DBT desulfurizing activity was retained. In their continued study with the recombinant T09 strain, Matsui et al. found that the introduction of *dszD* (NADH:FMN oxidoreductase) additionally gave enhanced DBT uptake rates [115]. However, the analysis of the medium showed that HBP was not produced in stoichiometric ratio with the DBT uptake rate and there was an accumulation of the penultimate product. It is clear that although the addition and expression of *dszD* gene simulates the activity of Dsz A and Dsz C monooxygenases, the activity of the DszB, desulfinase is insufficient to carry out desulfurization completely. However, when an additional copy of *dsz* operon was introduced, the final desulfurization activity increased by almost 3.3 times. In their further studies [116] with the recombinant T09 strain, they reported that the DBT desulfurizing activity was repressed in the presence of inorganic sulfur. But the replacement of the *dsz* promoter with a putative sulfate non-repressible promoter the recombinant strain was able to exhibit desulfurization even in the presence of sulfate, cysteine and methionine. However, the activities have been too low than the expected levels.

As the reaction involving the action of the desulfinase, DszB has been reported to be the rate limiting step, efforts have been reported to increase the desulfurization rate by improving the specific production of DszB. Li et al. [117] investigated the transcriptional characteristics of the native *dsz* operon in IGTS8 by analyzing the quantities of mRNA from the transcription of *dszA*, *dszB*, and *dszC* in the cells of *R. erythropolis* DS-3. The analysis revealed the ratio of mRNAs quantity of *dszA*, *dsz*,

and *dszC* to be 11:3.3:1. However, the Western-blot analysis showed that the expression level of DszB is much lower than DszC probably due to low translation of *dszB* mRNA. Further the gene analysis revealed that the termination codon of *dszA* and initiation codon of *dszB* are overlapping while a 13bp gap exists between *dszB* and *dszC*. They suggested that the translation levels of the Dsz enzymes decrease according to their position on operons due to the polar effects on *dsz* gene transcription in prokaryotes and for DszB in particular due to overlapping. In over to overcome this, they redesigned the *dsz* operon using overlap polymerase chain reaction and introduced the modified operon into *R. erythopolis* 4.1491(dsz^{-}) to produce recombinant cells, *R. erythropolis* DR-2. For comparison they also prepared recombinant cells *R. erythopolis* 4.1491, *R. erythropolis* DR-1 containing the original operon. It was seen that the desulfurization activity obtained with the DR-2 (120 µmol/gdcw-h) is nearly five folds higher than DR-1 (26 µmol/gdcw-h). In another study [118] they further modified the *dsz* operon for achieving still higher rates of DBT desulfurization. In this modification they rearranged the order of the *dsz* genes in the operon to generate *dszBCA* operon and reconstructed the ribosome binding site of *dszB.* The modified operon was inserted into *R. erythropolis* 4.1491(dsz^{-}) to obtain the recombinant strain DRA while the insertion of the original operon gave a desulfurizing recombinant strain, DS3. The western-blot analysis showed that the recombinant strain DRA exhibits higher levels of transcription of *dszB* and *dsz* than BS3. The desulfurizing activity obtained with DRA (320 µmol/gdcw-h) is nearly 12 folds higher than DS3 (26 µmol/gdcw-h).

It is evident from the results of above studies that various strains of *R. erythropolis* are capable of carrying out biodesulfurization by specifically removing sulfur from DBT and its derivatives, commonly found in fossil fuels. However, both the wild desulfurizing rhodococci strains show low levels of activity. Moreover, most

strains do not exhibit desulfurization for BT and its derivatives that otherwise largely contribute to the recalcitrance to hydrodesulfurization process. Nevertheless, the efficient reduction in overall sulfur content of the fossil fuels requires simultaneous desulfurization of the various PASHs at high rates. Therefore, it is crucial to look at strains that naturally exhibit desulfurizing activity for a wider of PASHs. In this regard, *Gordonia* strains have shown interesting results as they show desulfurization for BT, DBT, and their derivatives.

2.1.3.5 **Biodesulfurization studies with gordoniae strains**

The isolation of *Gordonia* sp. strain 213E [119] capable of non-destructively desulfurizing BT has encouraged the exploration of *Gordonia* strains as the potential desulfurizing biocatalysts. Over the last decade several strains of *Gordonia* have been isolated for their ability to desulfurize various PASHs including DBT, BT, and their derivatives [120-125]. Some of these strains [120, 123] have the ability to desulfurize BT and its derivatives but not DBT while others can desulfurize only DBT and its derivatives [121, 125-128]. However, *G. alkanivorans* 1B has the ability to desulfurize thiophene, BT, DBT, and their derivatives [35]. It is therefore, useful to study the gordoniae strains as they can be instrumental in reducing the overall sulfur content in fuels when used alone or in combination with rhodococci strains. The isolation of the different desulfurizing strains of *Gordonia* has been followed by the study of their pathways/mechanisms for the desulfurization of DBT and BT.

The DBT desulfurization activity in the desulfurizing strains of *Gordonia* is conferred by the *dszABC* genes [32, 129-131] present on the chromosomal DNA. These *dsz* genes from the different gordoniae strains share a high similarity between each other. The *dszABC* genes from gordoniae strains are homologous to the

desulfurizing genes present in *R. erythropolis* [129, 130]. The DBT desulfurization takes place in gordoniae via the 4S pathway [121, 125, 127]. Just as in case of the rhodococci strains, the *dszA*, *dszB*, and *dszC* strains encode for the Dsz A, Dsz B, and Dsz C enzymes, respectively. The activities of the enzymes Dsz A and Dsz C depend on the an oxidoreductase enzyme, DszD encoded by *dszD* [130]. However, BT is desulfurized by a pathway that is similar but distinct than the '4S' pathway.

Matsui et al. [123] studied the intermediates of BT desulfurization pathway in *G. rubropertinctus* T08. They reported that BT is desulfurized in a four step process. First, the BT molecule is oxidized to benzothiophene-sultine with the intermediate formation of benzothiophene S-oxide and benzothiophine S,S-dioxide, respectively. The BT sultine is then desulfurized to o-hydroxystyrene by the specific cleavage of the C-S bond. The sulfur atom is released as a sulfite moiety. The identification of the non-destructive pathways for specific desulfurization of DBT and BT molecules in gordoniae strains has fortified the experiments to study different factors and conditions that may influence the desulfurizing activity of gordoniae strains.

There are a few reports in literature on the effects of various medium and environmental conditions on the activity and growth of the desulfurizing gordoniae strains. Several works have studied the effects of pH on the desulfurizing activity of gordinae strains [124, 128, 132]. It has been found that the optimal pH for highest desulfurizing activity is 7 and the activity decreases with a decrease in pH. The optimal temperature for strain *Gordonia* sp. AHV-01 is 30ºC. Studies are also available on the effects of the various nutrients on the desulfurizing activity of the gordoniae strains. It has been found that the carbon source has a prominent effect on the desulfurizing activity with ethanol being the best source [132]. Ethanol is consumed more rapidly than any other carbon source and supports higher growth and desulfurizing rates.

According to Alves et al. [133, 134] there is a significant effect of the presence of various metal ions on the desulfurizing activity. The zinc has been found to enhance both the growth and the desulfurizing activity of *G. alkanivorans* 1B when present in the medium. In addition, there are studies available to determine relative desulfurization of various PASHs which differ in structure and alkyl moieties.

The researchers have studied the desulfurization activity of the gordoniae strains for the mixtures of the various PASHs. Matsui et al. [123] reported that *G. rubropertinctus* sp. T08 can desulfurize BT and its derivatives. However, when they are present in a mixture their assimilation by the bacteria is determined by the position and hydrophobicity of the alkyl group. The derivatives with a shorter side chain (2-methylBT) give higher growth rates than the ones with relatively longer side chains (2-ethylBT). This probably happens due to the steric hinderance effects of the longer side chains. In another report, Alves et al. studied the desulfurization of a mixture of BT and DBT by *G. alkanivornas* 1B [35]. They found that the BT was used as a preferred sulfur source and DBT was consumed only after BT had been exhausted from the medium. They proposed the presence of a non-specific transport system in strain 1B which can uptake BT and DBT but has a higher affinity for BT. However, a suitable desulfurizing biocatalyst would be one that would simultaneously desulfurize the various thiophenic compounds and their derivatives.

The desulfurization of BT as well as DBT by the gordoniae strains is repressed in the presence of the readily bioavailable sulfur sources such as sulfate, sulfite, cysteine, etc. [35, 124, 125, 135]. This poses a problem in the mass production of the gordoniae desulfurizing biocatalysts as an economical process requires the use of a readily bioavailable cheaper source such as sulfate, sulfite, etc. However, as the presence of these sulfur sources represses the desulfurizing activity, the use of DBT

becomes essential for the biocatalyst production. Conversely, the use of DBT as the sulfur source for bulk production of desulfurizing cells is not recommended as it is expensive and water insoluble. Chang et al. [136] proposed a two stage fermentation method for the bulk generation of *G. nitida* CYSK1 for large scale desulfurization processes. During the first stage they grew the cells in a sulfate based medium. The cells harvested from first stage were then treated with DBT in the next stage to induce the desulfurizing activity. This method helped in achieving a cell concentration of 92.6 g/L, which is nearly three times of the concentration (33 g/L) achieved in earlier studies with the desulfurizing strain of *R. erythropolis*. In a similar study, Ghasemali et al. [135] studied the bulk production of *G. alkanivorans* RIPR90A desulfurizing biocatalyst in a two stage process using dimethyl sulfoxide (DMSO) as the sulfur source in the first stage. They found that the cells grown in the presence of DMSO exhibited desulfurizing activity while they did not do so when grown in the presence of readily bioavailable sulfur sources such as sulfate, methionine, and cysteine. In addition, they observed that the desulfurizing activity (1.4 µmol HBP/gdcw-h) of the cells grown on DMSO could further be improved by nearly 4 times (5.11 µmol HBP/gdcw-h) using DBT as an inducer in the second stage. They also found that the highest desulfurizing activity obtained in the two stage fermentation process, using DMSO as the sulfur source in the first stage was nearly 2.8 time than that achieved when sulfate was used. However, in separate reports it has been found that the higher concentrations of DBT (> 1.5 mM) are deleterious to the growth and desulfurizing activity of the cells [132] so it must be supplied within toxic limits when used as an inducer. Furthermore, it is also reported that the end-product of its metabolism i.e, 2-HBP inhibits both the growth and desulfurization at higher concentrations [132, 137,

138]. This implies that the simultaneous removal of 2-HBP from the medium is desirable for continued desulfurization.

The aim of the research on desulfurization is to treat fossil fuels, which contain complex organic compounds and hydrocarbons that are known to be toxic to the cells. However since water is essential for the survival of the cells therefore, in most studies the biphasic medium has been used for desulfurization studies. In biphasic medium, the phase ratio has a considerable effect on the desulfurization activity with lower organic to aqueous phase ratios being more suitable [139]. But a higher organic to aqueous phase ratio is desirable for achieving higher desulfurization. Morever, the use of a biphasic medium is undesirable as it leads to the mass-transfer limitations for the uptake of DBT from the organic phase by the cells present in the aqueous phase. The mass transfer limitations are overcome by the emulsification by bacterial cells [140] or by addition of a suitable surfactant such as Tween-80 [132]. However, the emulsions need to be broken in order to recover the treated oil and separate the biocatalyst which is a challenging job. Choi et al. [141] found that ethanol can be used as a de-emulsifier for an efficient separation of the treated oil, biocatalyst, and water. They also observed that the recovered cells could use the ethanol as a carbon source and their desulfurization activity improved with each recovery due to repeated exposure to DBT. Alternatively, immobilized cells can be used for desulfurization as immobilization offers advantages of treating high oil-water phase ratios and ease of separation of the biocatalyst from the oil. Chang et al. [137] studied the desulfurization by CYSK1 immobilized on celite beads. The celite is chosen as it is inert to the biological attack, stable, and non-toxic in nature. They found that using immobilized cells the biphasic medium with high oil water ratio (~40) could be efficiently desulfurized. In addition, the desulfurization activity and the longevity of the immobilized cells were higher.

However, the highest desulfurization activity achieved with repeated usage (1.81 mg sulfur/gdcw-h) was still lower than the desired value. Moreover, the activity was reduced by nearly 40-50% on storage. These studies have indicated the need for improved desulfurizing biocatalysts and effective ways for their storage.

An efficient way of increasing the flux through the desulfurization pathway is to modify the genetic architecture of the desulfurizing cell. Shavandi et al. [36] designed the recombinant cells of RIPI90A by using *Escherichia coli-Rhodococcus* shuttle vector pRSG43 to efficiently clone and express *dsz* genes under the *lac* promoter. They found that the recombinant cells gave nearly three times higher growth rates and desulfurization activity in shorter fermentation times. However, the highest desulfurization activity achieved is still too low for commercialization. Therefore, there is a clear need for development of efficient gordoniae strains or other desulfurizing biocatalyst with higher activity for a broader range of PASHs.

2.1.4 Need for *in silico* holistic studies of desulfurizing strains

It is clear from the above review that the desulfurization rates exhibited by the wild type desulfurizing strains of *Rhodococci* and *Gordonia* are too low for commercialization. Although isolation of new desulfurizing strains with higher desulfurization activity for a wider number of PASHs is a possible route, the current research efforts mainly focus on the genetic studies to obtain improved strains. However, the genetically engineered strains have also been unable to give desired levels of desulfurization rates. These genetic engineering efforts have largely focused on the alterations within the ‘4S’ pathway and the associated genes and promoter. Nevertheless, since the cellular phenotypes result from the coordinated activities of multiple gene products and environmental factors, it is likely that other unknown host

functions control the desulfurization activity. As such, desulfurization enzymes might be the rate limiting factors in achieving desulfurization activity up to a certain level, but in order to achieve even higher activities alterations in genes and regulatory sequences other than (or in addition to) the *dsz* genes may probably be needed. Therefore, a better understanding of the interactions between the various host factors and the desulfurization pathway is critical. Moreover, as genetic engineering is carried out to increase metabolic flux through an organism by manipulating the basic genetic makeup so it is necessary to determine flux distributions through different pathways rather than a single pathway. Furthermore, there are several inexplicable observations associated with the desulfurizing strains. For instance, it is unclear how the sulfur from DBT gets assimilated into biomass [3]. It is also confusing that while sulfite is the only sulfur containing product obtained during desulfurization of DBT via '4S' pathway, sulfate is seen as the only excreted product. The lack of understanding of the sulfur metabolism in desulfurizing strains is one of the major bottlenecks in enhancing flux through the '4S' pathway. A major reason behind these limitations is the absence of any holistic study of the sulfur metabolism and the rest of the metabolic network in the desulfurizing strains. The biochemical reactions and pathways and their complex interactions largely determine the metabolic fluxes [142] through the various cellular activities (here, desulfurization activity). As such studying the entire metabolism as an integrated network is essential for gaining better insights and understanding into its metabolic machinery of the desulfurizing strains.

However, the extreme complexity of the metabolic networks makes it difficult to understand such intracellular interactions and predict the changes in metabolic fluxes in response to external environment. The *in silico* models of organism may play a role in developing a deeper understanding of the intracellular factors affecting the

desulfurization phenotypes. A systems biology approach can be used to reconstruct and analyze the *in silico* models that represent the metabolic networks of desulfurizing strains.

2.2 **Systems biology for *in silico* metabolic modeling**

A cell is the basic unit of all living organisms. It is a complex multi-component system consisting of several genes, proteins, enzymes, cofactors and various metabolites. All of these components are interrelated and form a complex network. Any cellular function results from the coordinated activity of these components under given environmental conditions. In biotechnology there is a special focus on the cellular metabolism of the organism as it can be exploited for various applications such as biosynthesis of important compounds, bioremediation purposes, and biotransformation to value added products, etc. Therefore, the knowledge of various cellular components and understanding of interconnectivity between them, the control and regulatory mechanisms that determine the possible phenotypes for a cell become critical.

Over the 20th century the biologists have largely followed the reductionists approach to study the cells. Reductionists analyze a large system by breaking it down into pieces and determining the connections between the parts [143]. Similarly the cells are studied by generating information about individual cellular components and their chemical composition, and deducing their biological functions. Such an approach has helped in providing the insight into properties of the individual components in a cell. In fact, this approach has given much success in the field of biochemistry and molecular biology. For instance, the various metabolic activities have been studied and analyzed by studying the various individual biochemical reactions and identification of the enzymes catalyzing them. However, cells exhibit emergent properties that arise

from interactions among their components and with the external factors [144]. Moreover, the dynamic nature of the interactions between these cellular components results in different phenotypes under different external conditions. It is therefore difficult to appropriately interpret the cell behavior by studying only the segregated parts. The reductionist approach has successfully identified most of the cellular components and many of the interactions but, unfortunately, offers no convincing concepts or methods to understand how system properties emerge. The development of high throughput technologies (genomics, transcriptomics, proteomics, and metabolomics) has helped to unravel the detailed molecular composition and complexity of cells and generate data in fundamentally different formats than the previous approaches. It is now generally accepted that the integrative analysis of these datasets and the study of cells as systems has become a critical issue for the future development of biology [145]. However, the *omics* data carry a large amount of complex information which requires tools to analyze and extract knowledge about the whole cell. Systems biology is an integrative science that aims to bridge the individual behavior of biological components with a collective behavior of the system [146]. It is a holistic approach and involves *in silico* modeling and analysis of metabolic pathways, regulatory, and signal transduction networks for understanding cellular behaviors. The term *in silico* modeling refers to the use of computers to formulate and analyze mathematical models that represent the biological systems. Simulation of relevant mathematical model provides a more reliable basis to understand the extremely complex network of the biochemical components because it is often difficult to comprehend and predict the behavior of such complex systems by intuition as our reasoning usually simple casual chains [147].

2.3 *In silico* mathematical models

The *in silico* mathematical models of biochemical networks can help integrate experimental knowledge into a coherent picture of the underlying biochemical mechanisms. Modeling underlines the holistic aspects of the biochemical networks, which may not be clearly obtained if the system is studied in terms of its constituents. In addition, a validated and established model may be used to generate testable hypotheses or perform *in silico* experiments which may otherwise be too complex and time consuming if performed in the wet lab. An *in silico* model is usually a simplified representation of an actual phenomenon and therefore, it is possible to build different models for the same phenomena, depending on the objectives of the model and available measurements [148]. The models vary in the level of details that can be incorporated in them, which in turn determines the accuracy of simulation results and the nature of insights that they can reflect [149]. The metabolic models can be usually grouped into:

a) Kinetic Models : usually based on the stoichiometry and enzyme or microbial kinetics
b) Stoichiometric Models: based on time invariant characteristics of metabolic networks

2.3.1 Kinetic models:

The biological systems are dynamic in nature therefore, in order to understand these systems we require quantitative measurements in time. The term, 'kinetic model' refers to a system of mechanistic differential equations that determine the temporal state of the corresponding system of biochemical reactions. In metabolic networks, such

models enable us to understand and predict the effect of altered enzyme activities or enzyme kinetic properties (e.g., feedback inhibition) on fluxes and levels of metabolites of interest [150]. In principle, a kinetic model of metabolism is constructed by defining stoichiometry and kinetics of all reactions involved in the biological system of interest and setting up the proper mass balances for each metabolite. A kinetic metabolic model is non-linear (ordinary differential equation, logarithmic, power law, linear-logarithmic) and involves a number of parameters. Although it is easy to define the stoichiometry and apply the mass balances, it is usually difficult to specify the kinetics of the biochemical reactions.

Moreover, the kinetic parameters of the reactions are usually based on the information obtained from the kinetic studies that are carried out using enzymes in a test tube (*in vitro*). However, the conditions inside the microorganism (*in vivo*) are very different from *in vitro*, where usually the enzyme and substrate concentration are much lower and higher, respectively. Moreover, various different compounds are present *in vivo* along with the high concentration of total proteins and certain storage compounds (e.g., trehalose, glycogen), all of which might change the (*in vitro*) obtained enzyme kinetic equations and/or parameters. Several reports have indicated that the *in vitro* kinetics do not perform well under *in vivo* conditions [151, 152] and it has been proposed that the enzyme kinetics of metabolic networks must be obtained from *in vivo* studies (using the whole cell) by using, for example, pulse response or (pseudo) steady-state experiments [153]. However, the identification of the kinetic parameters from such *in vivo* experiments poses serious problems.

Furthermore, even though the kinetic models give useful information, it is difficult to use them to model genome-scale networks because of the large number of parameters needed and the computational complexity [37]. In the absence of complete

kinetic information on biochemical reactions, an alternate way is to use the known stoichiometry of these reactions to study the theoretical capabilities and functions of the entire metabolic networks. One such approach involves the reconstruction and analysis of constraint-based genome scale stoichiometric models.

2.3.2 **Stoichiometric models:**

The annotated genome of an organism can be used to easily construct the entire metabolic network of biochemical reactions occurring within that organism and to represent it in the form of a stoichiometric matrix [154]. A stoichiometric matrix *S*, is an $m \times n$ matrix where *m* corresponds to the no. of metabolites and *n* is the no. of fluxes occurring in the organism. The element S_{ij}, of the stoichiometric matrix corresponds to the stoichiometric coefficient of metabolite *i* in the reaction *j*. It is an invariant property of the metabolic network that describes the metabolic architecture and topology of the system [155]. The stoichiometric matrix reflects the interconnectivity of the metabolites and allows us to represent the metabolic network in a mathematical format permitting further analysis to study the system.

The static models built on the known stoichiometric matrix can be analyzed to study the flux through the metabolic network under steady state assumptions and hence, the physiology of the organism. Metabolic fluxes can be seen as a fundamental determinant of the cell physiology because they show the quantitative contributions of various pathways to overall cellular functions. Therefore, a common way of relating cell genotype to phenotype is by analyzing the fluxes in the metabolic network. The static models are based on the assumption of a quasi-steady state according to which the metabolic transients are more rapid than both cellular growth rates and the dynamic changes in the environment of the organism [156]. Thus, under this assumption all

metabolic fluxes leading to the formation and degradation of any metabolite must be balancing, leading to a flux equation:

$$\frac{dx}{dt} = \mathbf{S}.\mathbf{v} - \mathbf{b} = 0 \qquad (2.1)$$

where, S is the *(m×n)*stoichiometric matrix

v is the vector of *n* metabolic reactions

b is the vector containing the net metabolite uptake by the cell

Equation 2.1 represents a typically underdetermined system since the number of fluxes usually exceeds the number of metabolites [142]. Therefore, an infinite number of solutions exist and thus, correspondingly the cell can exhibit a number of phenotypes. The linear optimization can be utilized to obtain a particular solution for the flux distribution. The definition of a suitable objective function Z, is necessary to compute an optimal network state and the resulting flux distribution [157]. The model may be further constrained using the experimentally determined substrate uptake rates and release rates for various products and by-products.

There are several constraint based approaches used to analyze the cellular flux analysis using linear optimization. Flux balance analysis (FBA) is one such approach in which the value of an objective function Z, is maximized/minimized subject to the constraints using experimentally measures uptake/release rates. The general formulation for FBA is

Maximize $\mathbf{Z}$

Subject to $\sum_{j=1}^{n} S_{ij} v_j \quad \forall i \qquad (2.2)$

$$\alpha_j \le v_j \le \beta_j \qquad (2.3)$$

$$\alpha_j \le b_j \le \beta_j \qquad (2.4)$$

Several objective functions such as biomass production, ATP generation, substrate uptake, useful metabolite production, etc., have been used in literature to define various optimal solutions [47, 158].

The ability to generate quantitative hypotheses by analyzing these constraint based stoichiometric models, has given them a wide range of applications in the field of metabolic engineering of various biocatalysts for different purposes such as the large-scale generation of valuable substances and the degradation of pollutants [40]. There have been instances where the reconstructed genome scale models of organisms have provided insights into the intracellular activities of microbial cells [43, 45, 46, 159-162] and provided answers for otherwise inexplicable behaviors of the cells. They are based on the genome annotations of an organism and other relevant biochemical information available in literature and are analyzed to study the cellular systems and provide meaningful insights into the organism for different purposes. Such models provide strategies for metabolic engineering by carrying out various gene deletion and gene addition studies. They enable successful interpretation and prediction of phenotypic behaviour of cells under different conditions and understanding of network.

2.4 **Research focus**

Our thorough literature review shows that a lot of studies have been done with several *R. erythropolis* strains for the DBT desulfurization studies. However, their activity is low and they usually do not exhibit desulfurization of BT. Since BT, DBT, and their derivatives are mainly responsible for recalcitrance to HDS method, it is essential to study strains that can exhibit activity for a wider range of these compounds. To this end, *G. alkanivorans* strains have been studied for their ability to desulfurize several thiophenic compounds. However, they do not exhibit simultaneous desulfurization of

BT and DBT but preferentially utilize BT. Furthermore, the activities exhibited by the wild type desulfurizing strains of rhodococci and gordonia are too low for commercialization. Despite numerous genetic manipulation efforts, desired level of activity is yet to be obtained. The key issues and limitations of the current works are as follows:

- Although the '4S' pathway for DBT metabolism is well elucidated, it is not clear how the sulfur from DBT gets assimilated into biomass
- Sulfur metabolism in desulfurizing strains is not well understood
- It is unclear why and how the sulfite molecule obtained as the end product of '4S' pathway is converted and excreted out as sulphate
- The host functions and factors that may interact with the desulfurizing pathway and control the flux through it, are unknown

For an efficient biodesulfurization process, it is desirable that BT and DBT are desulfurized simultaneously. However, it is usually not the case as either the strains exhibit activity for just one of them or they use one of them preferentially

Due to these bottlenecks in the current research, biodesulfurization process has not moved beyond the laboratory-scale experiments. A major limitation of the current research is the continued use of reductionist approach to study desulfurization. Most studies have revolved around the characterization and modifications of the desulfurizing pathway. However, the desulfurization activity is an emergent phenotype that results from coordination and interactions between the desulfurization pathway and the other metabolic activities. Therefore, it is essential to study all the biochemical activities occurring within the desulfurizing strains as an integrated metabolic network. However, no such holistic study is available in literature. This has motivated us to reconstruct *in silico* models for the metabolic networks of *R. erythropolis* and *G.*

*alkanivoransi*to perform holistic systems study. The course of research taken to answer the above inexplicable or unknown factors is as follows:

- We first develop an *in silico* flux based model of sulfur metabolism in *R. erythropolis*. It presents the network of reactions and pathways involved in the extraction of sulfur atom from DBT and its further processing to synthesize the various sulfur containing precursor metabolites. It thus, gives a quantitative elucidation of the process by which sulfur from DBT gets assimilated into biomass.
- Next, we reconstruct an *in silico* genome scale metabolic model of *R. erythropolis*. It represents the various metabolic activities occurring within *R. erythropolis* as integrated parts of the metabolic network. The model enables the study and analysis of the known experimental phenotypes. Further, it helps in identification of host functions that may exert control on the extent of desulfurizing activity that *R. erythropolis* strain can exhibit.
- Further, we present an *in silico* genome scale metabolic model of *G. alkanivorans* to study desulfurization of BT and DBT. The model has been used to analyse the properties of the metabolic network in *G. alkanivorans.*
- Next, we present an *in silico* model for a co-culture of *G. alkanivorans* and *R. erythropolis* to study desulfurization of a mixture of BT and DBT.

The subsequent chapters cover our works on the reconstruction and analysis of the various *in silico* metabolic models.

3 Flux-based analysis of sulfur metabolism in desulfurizing strains of *Rhodococcus erythropolis*

3.1 Introduction

The increasingly stringent rules for ultra-low-sulfur fuels have inspired efforts to improve existing desulfurization techniques and develop new, efficient, and more economical methods. Biodesulfurization is considered an attractive technique for removal of sulfur from fossil fuels. It involves the action of microbial cells or enzymes to remove the sulfur atom from various PASHs present in fossil fuels. Of the several desulfurizing strains isolated, *R. erythropolis* IGTS8 was the first to be identified for its ability to specifically cleave the C-S bond in PASHs without affecting the C-C bond [10].

Over the last two decades, several strains of *Rhodococcus* [10, 16, 85, 100, 163, 164] have been isolated, characterized, and studied for their desulfurization ability. DBT has been the model compound in most studies and a number of detailed studies on the mechanism of DBT metabolism are available in literature [2, 7, 165]. Several studies have well documented the cell growth rates on DBT [28-30] and have evaluated the effects of the various medium and operational conditions [20, 23] on the desulfurizing activity. However, it is still unclear how the sulfur from DBT gets incorporated into biomass [3]. A quantitative elucidation of the actual process by which sulfur from DBT gets assimilated into various biomass precursors of *R.*

erythropolis has not been reported so far. In order to obtain improved desulfurizing activity using *R. erythropolis* as the biocatalyst, a clear understanding and elucidation of its sulfur metabolism in critical. However, most studies have targeted the 4S pathway solely, and a holistic systems study of intracellular sulfur processing has not been reported. Since the cellular phenotypes are an emergent behavior of complex interactions among various gene products and environmental factors, studying the desulfurization activity using a systems approach is critical. Thus, it is desirable to study the sulfur metabolism and related metabolic activities in a desulfurizing organism using a holistic systems approach.

In this chapter we report the first attempt, to our knowledge, at reconstructing a stoichiometric model for the sulfur metabolism in *R. erythropolis*. It comprises a network of reaction pathways involved in sulfur and central metabolism, and quantitatively describes the assimilation of sulfur from different sources into various biomass precursors. In this we have included the proper assignment of the gene-protein-reaction (GPR) associations for the various metabolic activities associated with the sulfur metabolism in *R. erythropolis*. It successfully predicts two independent cell growth data and several phenotypes reported in the literature such as the effects of sulfate and various carbon sources on biodesulfurization activity. We have successfully used the model to compare the effects of eight carbon sources (citrate, ethanol, fructose, gluconate, glucose, glutamate, glycerol, and lactate) on desulfurizing activity and cell growth.

3.2 **Materials and methods**

3.2.1 **Model construction**

The flux-based models have been widely used to study the metabolic networks of various microorganisms in a holistic manner [45, 47, 166, 167]. Such a model for an organism is built on the known and hypothesized reactions that may take place within the organism [168]based on its genomic, biochemical, and physiological information [169]. In this work, our goal is to develop an *in silico* model that gives a quantitative description of the metabolic and biosynthetic functions of sulfur in *R. erythropolis*. Thus, we limited ourselves largely to the pathways dedicated to the syntheses of sulfur-containing metabolic precursors and their incorporation into biomass. However, we also added select pathways from the central metabolism to elucidate and examine the effects of carbon sources [18] on desulfurization activity and the key role of reducing equivalents [86] in the energy-intensive 4S pathway.

Our basis model used the information on pathways and reactions available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [170] database. We curated the reactions manually and corrected them for carbon and sulfur balances. Further, we included some additional reactions from the literature [86, 161, 162, 171] and MetaCyc [172] to complete the pathways necessary for the biosynthesis and utilization of some key metabolites. For instance, we took the reactions for the 4S pathway from [171], mycothiol biosynthesis from [173], and metabolism of glycerol and glutamate from MetaCyc [172]. Likewise, we adapted the pathways for the biosynthesis of thiamine and biotin from the existing reconstructed metabolic model of a related actinomycete, *Mycobacterium tuberculosis* [161, 162]. Table 3.1 shows the number of reactions taken from each of the above mentioned sources. However, being limited in scope and

pathways, the resulting model could still not synthesize (consume) some substrates (products) such as inositol, pantothenate, etc. that appear in the reactions. Therefore, we assumed an extracellular pool of such metabolites and added transport reactions with unlimited fluxes to simulate their necessary uptake (release).

A biomass equation represents cell growth in a flux-based *in silico* model. It is a synthetic reaction that consumes cell constituents in known constant proportions (derived from cell composition) to form a unit amount of cell biomass. However, as a quantitative analysis of the biomass constituents in *R. erythropolis* is unavailable in the literature, we adapted the biomass equation in our model from the known composition of a related actinomycete, *M. tuberculosis* [161, 162]. We kept only the precursors that either contain sulfur or are involved in the sulfur metabolism, and added other sulfur-containing cofactors such as biotin and thiamin to appropriately reflect the requirements of sulfur and its metabolism. However, we excluded sulfolipids, as they are known to confer pathogenic characteristics to *M. tuberculosis*.

Table 3.1 Number of reactions from different sources for the reduced model

Source	No. of Reactions
KEGG	63
Beste et al. 2007	16
Jamshidi et al. 2007	16
Metacyc	8
Oldfield et al., 1998	4
Rawat et al. 2007	4

For performing the flux balance analysis with the resulting model, we used MetaFluxNet [174].

3.2.2 **Experimental data sources for model validation**

Experimental data are indispensable for validating an *in silico* (computational) model. For this study, we used the experimental data of Izumi et al. [11] and Davoodi-Dehaghani et al. [99] on cell growth and metabolite concentration profiles. Izumi et al. [11] reported that *R. erythopolis* D-1 desulfurized DBT to 2-hydroxybiphenyl (HBP) successfully. They used 500 ml of a glucose-based biosynthetic medium with 0.125 mM DBT as the sole sulfur source at 30 °C to examine the desulfurization activity of growing cells. They measured pH, cell growth, DBT concentration, and HBP concentration at various times during their experiment.

In another study, Davoodi-Dehaghani et al. [99] isolated *R. erythropolis* SHT87. They used growing cells at 30 °C in a 50 ml solution of glycerol containing a synthetic medium with 0.25 mM of DBT as the sole sulfur source. They also measured cell growth, DBT concentration, and HBP concentration at different times over 120 h.

The experimental data from the above two independent studies provided a sound basis for validating our proposed model. We used their cell growth data and DBT/HBP concentration profiles from the exponential phase to compute specific cell growth rates (1/h) and DBT (HBP) uptake (secretion) rates (mmol/gdcw-h).

3.3 **Results and discussion**

3.3.1 **Model for sulfur metabolism in *R. erythropolis***

Our reconstructed model consists of 87 intracellular metabolic reactions, 66 transport reactions, and 196 metabolites related to either sulfur or central metabolism. The sulfur metabolism includes the 4S pathway; the CoA biosynthetic pathway; metabolism of inorganic sulfur, cysteine, and methionine; and biosynthesis of cysteine, methionine,

mycothiol, biotin, and thiamine. The central metabolism includes gluconeogenesis, citric acid cycle, pentose phosphate pathway (PPP), and Embden Meyerhoff Paranas (EMP) pathway for glycolysis. Fig. 3.1 shows a complete picture of the pathways and reactions in our model with full details in the Appendix A.

3.3.2 **Model validation**

We simulated the experiments of Izumi et al. [11] and Davoodi-Dehaghani et al. [99] and compared our predicted cell growth rates with their measured data. As the 4S pathway is aerobic, we assumed unlimited oxygen flux in all of our validation studies and analyses. Sulfur was a limiting substrate in the experiments of Izumi et al. [11] and Davoodi-Dehaghani et al. [99]. We inferred this from the fact that the stationary phase in their experiments was triggered, when DBT concentration went to zero and HBP concentration reached its maximum. Therefore, we allowed unlimited glucose flux for simulating the experiment of Izumi et al. [11] and unlimited glycerol flux for Davoodi-Dehaghani et al. [99].

Then, we fixed the DBT uptake and HBP production rates (mmol/gdcw-h) to be at some values computed from their data, and predicted specific cell growth rates at those values. Fig. 3.2 shows that our growth predictions are in close agreement with the two experimental data. The accuracy of our predictions is confirmed by the argument that the limiting sulfur solely determines the growth.

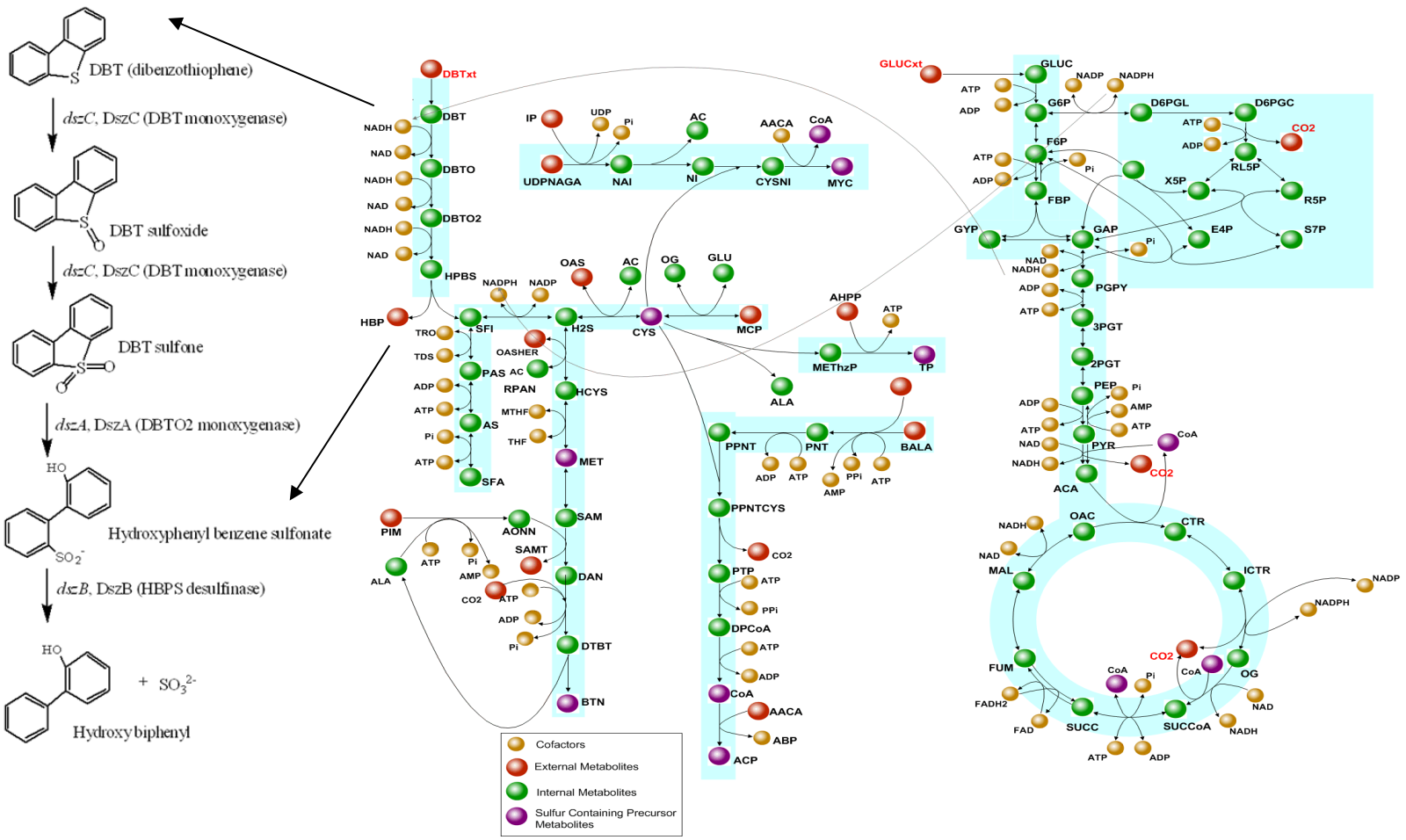


Fig. 3.1 Schematic of metabolic pathways included in the reduced model for sulfur metabolism in R. erythropolis. Detailed 4S pathway shown as the enlarged figure on the left

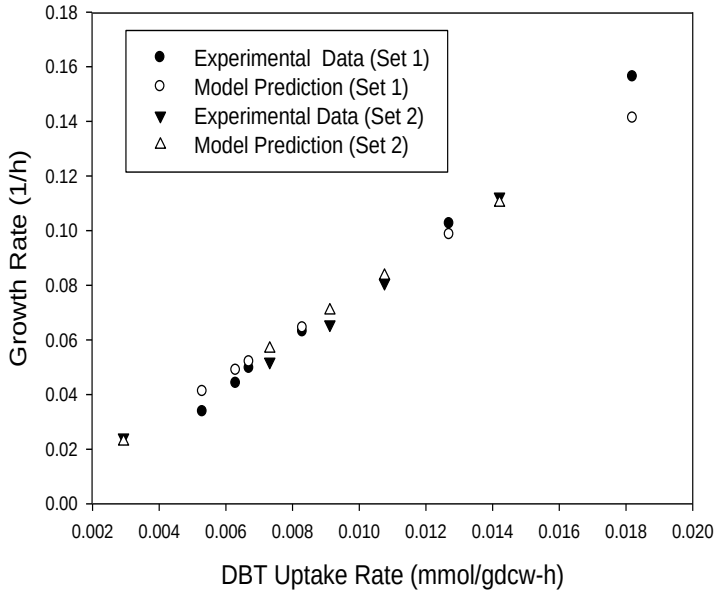


Fig. 3.2 Cell growth rate versus DBT uptake rate. Set 1 corresponds to the data from Izumi et al. [10] and Set 2 corresponds to the data from Davoodi-Dehaghani et al. [99]

3.3.3 **Analysis of sulfur metabolism using alternate sources**

We then studied the utilization of sulfate and DBT as sulfur sources to qualitatively demonstrate the consistency of our model with some literature observations [16, 24] on cell growth and desulfurizing activity.

3.3.3.1 ***In silico* growth on alternate sulfur sources**

In a study on desulfurization by *R. erythropolis* IGTS8 in an acetate-based medium, Honda et al. [24] observed that sulfate promoted higher cell growth than DBT. To study this phenotype, we performed flux balances for two scenarios (Table 3.2) with unlimited acetate uptake. In run 1, we fixed the DBT (sulfate) uptake at 20 (0.0) mg/gdcw-h. In run 2, we fixed sulfate (DBT) at 20 (0.0) mg/gdcw-h. Our model gave a higher cell growth rate (1.29 vs. 0.84 h^{-1}) for sulfate (Run 2) than DBT (Run 1). Then, we fixed the acetate uptake at 20 mg/gdcw-h and studied two more scenarios (Table 2). In run 3, we allowed unlimited (zero) sulfate (DBT) uptake, and did the reverse in run

4. Again, we obtained a higher growth (1.4 vs. 1.06 h^{-1}) for sulfate (Run 3) than DBT (Run 4).

After studying sulfate and DBT separately, we studied them together (Run 5 in Table 2) as well for a fixed acetate uptake of 20 mg/gdcw-h. We fixed the sulfate uptake at 2.16 mg/gdcw-h and allowed unlimited DBT. This sulfate uptake is 10% of its maximum (21.6 mg/gdcw-h) observed in run 3. The model showed a higher growth rate of 1.12 h^{-1} compared to 1.06 h^{-1} obtained previously for run 3 (unlimited DBT, zero sulfate). The DBT uptake was also lower (22.08 vs. 25.76 mg/gdcw-h). This suggests that the organism may grow faster, when it fulfills a part of its sulfur needs via sulfate rather than DBT. In other words, the organism may prefer sulfate, when both DBT and sulfate are present. Since sulfate gives a higher growth rate than DBT, the organism may use DBT, only if sulfate is not present. This clearly confirms the results of Honda et al. [24].

Honda et al. [24] reasoned that the observed lower cell growth with DBT was due to the toxic effect of HBP (its desulfurized product). Since our model does not include such toxic effects, we cannot deny this as a probable explanation. However, we have the following alternate explanation from our study. *R. erythropolis* needs sulfate and sulfide to synthesize its sulfur-containing biomass precursors. If it uses DBT as the sulfur source, then it must use the 4S pathway. 4S converts DBT to sulfite, which is converted to sulfate and sulfide by the sulfur metabolism and then incorporated into the biomass precursors. However, the organism needs 4 moles of NADH per mole of DBT to use DBT in the above manner. In contrast, the organism does not need this extra NADH for metabolizing sulfate. Thus, the organism prefers the energetically less expensive sulfate over DBT for its growth. Although our reduced model does not include all the reactions involving NADH, it is known that NADH is an essential

component for growth. When the organism is forced to use DBT, NADH available for other growth-critical activities inside the cell reduces, thus cell growth reduces.

Table 3.2Flux analyses for the utilization of alternate sulfur sources with acetate as the carbon source and effect of sulfate on desulfurization activity with succinate as carbon source

In silico growth of alternate sulfur sources					
Run	Uptake Inputs			Model Results	
	Acetate	DBT	Sulfate	Growth	Uptake
1	Unlimited	0	20	1.29	Acetate = 18.48
2	Unlimited	20	0	0.84	Acetate = 15.78
3	20	Unlimited	0	1.06	DBT = 25.76
4	20	0	Unlimited	1.40	Sulfate = 21.60
5	20	Unlimited	0.21	1.12	DBT = 22.08
Effect of sulfate on desulfurization activity					
Run	Uptake Inputs			Model Results	
	Succinate	DBT	Sulfate	Desulfurizing Activity	Uptake
6	20	Unlimited	0	0.07	DBT = 12.88
7	20	0	Unlimited	0	Sulfate = 10.80
All uptakes in mg/gdcw-h , growth in h^{-1} and desulfurizing activity in mmol/gdcw-h					

3.3.3.2 **Effect of sulfate on desulfurizing activity**

The reduced DBT uptake in the presence of sulfate prompted us to study the effect of sulfate on desulfurization rates. Several literature studies have reported the effect of sulfate on desulfurization activity. Li et al. [50] reported that although sulfate represses the *dsz* genes, it does not inhibit the activity of desulfurizing enzymes [98]. They observed that the desulfurizing activity increased with decreasing amount of sulfate in the medium. Similarly, Omori et al. [16] also observed enhanced desulfurizing rates

arising from the removal of by-product sulfate from a succinate-based medium. To understand this phenotype using our *in silico* model, we analyzed fluxes for three scenarios (Table 3.2) with a succinate uptake at 20 mg/gdcw-h. In Run 6, we allowed unlimited DBT as the sole sulfur source and obtained the maximum desulfurizing rate of 0.07 mmol/gdcw-h. In Run 7, we allowed unlimited sulfate as the sole sulfur source, and obtained the maximum sulfate uptake of 10.80 mg/gdcw-h. Then, in subsequent runs, we allowed progressively increasing amounts of sulfate (from 0 to 100% of the maximum sulfate uptake of 10.80 mg/gdcw-h from Run 7) with unlimited DBT. From Figure 3.3, we see that the desulfurizing activity clearly decreases with increasing amount of sulfate. Thus, our model successfully explains the observations of Omori et al. [16] and Li et al. [50].

Our earlier comment on energy needs again readily explains this effect. When the desulfurizing enzymes are already present, then the organism is able to take (desulfurize) DBT. However, sulfate promotes higher growth at lower energy, so the organism prefers sulfate consumption over DBT conversion. Only when sulfate is limited, it desulfurizes DBT. In other words, no desulfurization is possible even in the presence of desulfurizing enzymes, if the medium has sufficiently high concentration of sulfate to meet the sulfur needs of *R. erythropolis*. To our knowledge, no previous experimental work has elucidated this phenotype, which our model made possible.

3.3.4 **Effect of carbon source**

Yan et al. [18] studied the relative efficacy of ethanol, glucose, and glycerol as sole carbon sources for the growth and desulfurizing activity of *R. erythropolis*. They reported ethanol to give the highest growth and desulfurizing rates, followed by glucose, and then glycerol. To simulate this phenotype, we considered three separate

scenarios with unlimited DBT and one carbon source. In each scenario, we fixed the uptake of the respective sole carbon source at 20 mg/gdcw-h and used maximum biomass as the cellular objective. Our model gave the highest growth rate of 1.39 h^{-1} and the highest desulfurizing rate of 0.18 mmol HBP/gdcw-h for ethanol. In contrast, the rates were 0.60 h^{-1} and 0.08 mmol HBP/gdcw-h for glucose, and 0.59 h^{-1} and 0.07 mmol HBP/gdcw-h for glycerol. Thus, our model qualitatively confirms the experimental results of Yan et al. [18].

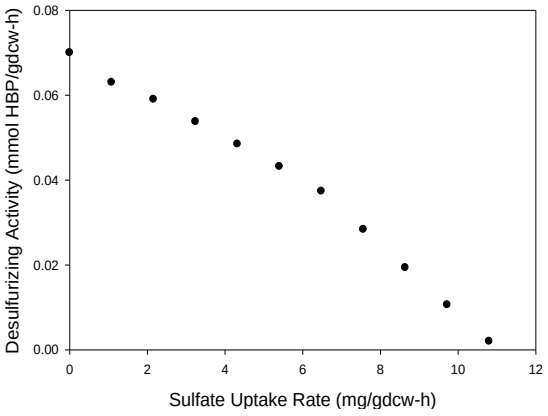


Fig. 3.3 Effect of increasing the sulfate uptake rate on specific desulfurization activity

The extra NADH needed for the 4S pathway again successfully elucidates the above results. The major source of NADH in *R. erythropolis* is the carbon metabolism. Ethanol gives more NADH during this metabolism than glucose and glycerol. The additional NADH enables the cell to increase the flux (or desulfurizing rate) of the 4S pathway, which eventually helps it to increase growth. Extending this, we argue that a carbon source that provides more NADH is likely to enhance both the growth and desulfurizing rates of *R. erythropolis*.

As our model predicted some experimental observations successfully, we examined the suitability of additional carbon sources for desulfurizing activity. We studied citrate, ethanol, fructose, gluconate, glucose, glycerol, glutamate, and lactate as possible sole carbon sources. We computed fluxes for each sole source separately with

an uptake rate of 20 mg/gdcw-h. Figure 3.4 shows the results of our eight simulation runs. The desulfurization and growth rates relative to those of ethanol decrease in the order: ethanol (0.18 mmol HBP/gdcw-h as 100% & 1.39 h^{-1} as 100%) > lactate (67%) > citrate (48%) > glutamate (44%) > glucose = fructose (43%) > glycerol (42%) > gluconate (40%). However, as our model is reduced and has limited scope this prediction is only qualitative in nature. An experimental verification of this prediction is clearly beyond the scope of this work. As a natural goal of any *in silico* model, our intention is simply to offer a new hypothesis that experimental researchers can verify.

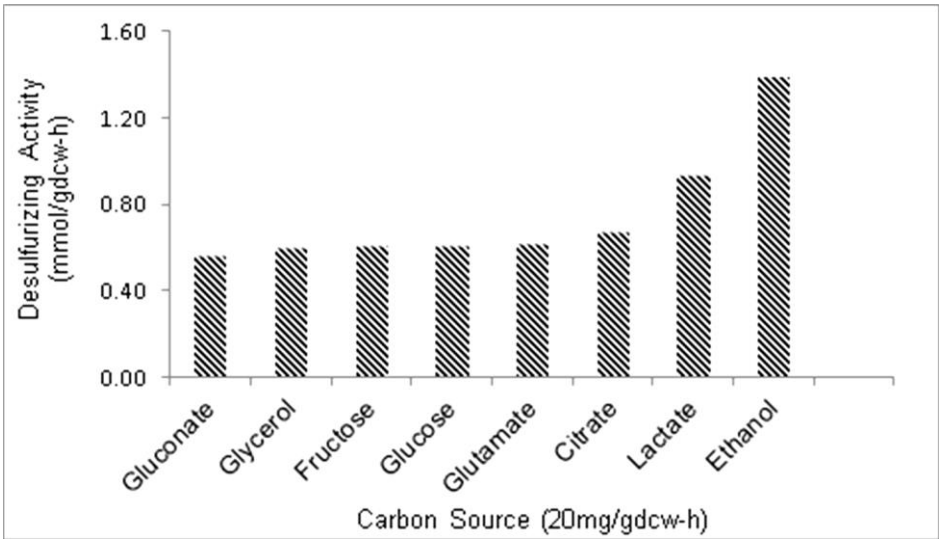


Fig. 3.4 Desulfurizing activities for various carbon sources at 20 mg/gdcw-h

3.4 **Summary**

In this chapter, we have reported the first attempt at reconstructing a flux-based model to analyze sulfur utilization by *R. erythropolis*. It predicts closely the growth rates reported by two independent experimental studies, and gives a clear and comprehensive picture of the pathways that assimilate the sulfur from DBT into biomass. In addition, it successfully elucidates that sulfate promotes higher cell growth than DBT and its presence in the medium reduces DBT desulfurization rates. A study using eight carbon sources suggests that ethanol and lactate yield higher cell growth

and desulfurization rates than citrate, fructose, glucose, gluconate, glutamate, and glycerol. Our analyses show that NADH plays a critical role in desulfurization activity. Any changes in medium design or genetic manipulations, etc., which increase NADH regeneration and supply within the cellular metabolism, are likely to enhance desulfurization activity.

4 Reconstruction of a genome-scale metabolic network of *Rhodococcus erythropolis* for desulfurization studies

4.1 Introduction

The advancement in biotechnology has led to the application of several organisms such as *Pseudomonas*, *Escherichia*, *Rhodococcus*, *Saccharomyces*, etc. have been studied for various biotransformations and bioremediations. Of these, *Rhodococcus* has attracted much attention for its catabolic versatility in degrading or metabolizing recalcitrant and toxic substances. The rhodococci can produce biosurfactants, which makes them potential candidates for bioremediation of soil, enhanced oil recovery, biotransformation for pharmaceuticals, and biotreatments involving organic compounds and/or solvents. The mycolic acids in the rhodococci cell surface confer hydrophobicity and enable them to tolerate and utilize hydrophobic compounds [8, 175]. A key application of *Rhodococcus* is desulfurizing fossil fuels. Many publications have reported on the desulfurization of various organic sulfur compounds and distillation fractions of fuels with wild type and genetically modified strains.

R. erythropolis desulfurizes DBT to 2-hydroxybiphenyl (HBP) via '4S Pathway'. It involves a set of four thermodynamically favorable reactions catalyzed by three enzymes (DszA, Dsz B, Dsz C) encoded by *dszABC* genes [7]. This pathway is of interest as it desulfurizes PASHs (DBT and its derivatives) by specific cleavage of

C-S bond without affecting their carbon skeleton and thus, the calorific value of fuel[2]. However, a biodesulfurization process using naturally occurring bacterial cultures is too slow for commercialization [5]. Despite numerous efforts to increase specific desulfurization activity via various genetic manipulations, further progress is needed to attain desirable desulfurization rates. A holistic study of various intracellular activities in *Rhodococcus erythropolis* is missing in the literature. Since cellular phenotypes are an emergent behavior of complex interactions among various gene products and environmental factors, the study of desulfurization using a systems approach is critical.

The complex interactions among the various metabolic pathways and associated reactions largely determine the metabolic fluxes [142] within an organism and hence its cellular activities and phenotypes. The extreme complexity of the metabolic networks makes it difficult to comprehend and predict the changes in metabolic fluxes in response to external environment. Structured quantitative stoichiometric models [159] based on known reactions can give insights into the possible flux distributions using the technique of Flux Balance Analysis (FBA) [156]. The flux distributions show the interdependence of various reactions and pathways quantitatively, and can help isolate the dominant intracellular factors affecting specific cellular activities. Such a model validated with experimental observations can inspire novel and non-intuitive hypotheses that may lead to further research. Unlike *R. erythropolis*, such models are available for several industrially important organisms such as *E. coli* [176, 177], *P. putida* [178-180], *P. aeruginosa* [181], *S. cerevisiae* [182-184], *P. pastoris* [185, 186], *Z. mobilis* [187, 188], etc. They have led to enhanced understanding and improved strains. In chapter 3, we developed a preliminary model to quantify sulfur metabolism in the desulfurizing strains of *R.*

erythropolis[189]. However, the model captured only a part of the otherwise vast and complex metabolic networks of desulfurizing strains, thus our study was highly limited and lacked full details.

In this chapter, we reconstruct a comprehensive genome-scale metabolic model of *R. erythropolis*. In the process, we identify and propose several possible new genome annotations for *R. erythropolis*. We validate the model using published experimental data on DBT desulfurization. We derive *in silico* minimal media and identify essential genes and reactions for sustaining cell growth and desulfurization activity. We also successfully predict the desulfurization characteristics of some mutant strains reported in the literature, and propose some new mutants based on our model-driven studies to enhance biodesulfurization. We analyze the effect of various carbon and nitrogen sources and their combinations on the rate of desulfurization, and study the suitability of various sulfur sources to obtain higher growth rates. Finally, we compare our model of *R. erythropolis* with that of *E. coli*, the model organism.

4.2 Materials and methods

4.2.1 Reconstruction of genome-scale model

The first step in the reconstruction of a stoichiometric metabolic model is to compile a list of metabolic reactions based on available genome annotations and biochemical literature. Fig. 5 gives an overview of our iterative process for reconstructing a detailed genome-scale model for the metabolic network of *R. erythropolis*. We based our initial model on the ORF (open reading frame) information from the annotated genome sequence of *R. erythropolis* PR4 [190] and associated reaction pathways from the KEGG [170] database. We also modified the reactions appropriately to ensure

elemental balances. However, our initial draft model exhibited several missing, incomplete, and/or isolated pathways. Thus, we improvised in several creative ways for our current incomplete knowledge of the genome sequence of *R. erythropolis*. For instance, we added several reactions based on available biochemical information, Metacyc[172] or literature evidence to complete the otherwise incomplete/broken pathways. We took reaction information on the synthesis of large complex cell wall components from the metabolic models of *Mycobacterium tuberculosis*, a related actinomycete. We also included several reactions to enable the synthesis of some key biomass precursors. Lastly, via a BLASTp analysis with an e-value cutoff of 10^{-5}, we identified candidate genes in *R. erythropolis* PR4, which can confer these missing functionalities to the organism.

A flux-based metabolic model needs an appropriate cellular objective to select a flux distribution that best represents the cellular phenotypes. Most FBA applications assume maximum cell growth as the cellular objective, and employ a linear hypothetical reaction to model cell growth. This reaction computes biomass as a weighted sum of key cellular components (biomass precursors and ATP) contributing to the dry cell weight. ATP quantifies the cell's energy needs for growth and maintenance. While experimentally measured cell composition of *R. erythropolis* is ideal for formulating such a biomass equation, this is not available in the literature. Therefore, we adopted the data on a closely related organism, namely *Mycobacterium tuberculosis* [161, 162]. Such an adaption from related organisms is an established and reliable practice in the literature [159]. However, note that we excluded, from our biomass, components such as sulfolipids that are known to confer pathogenicity to *M. tuberculosis* strains[161]. Also we included certain important sulfur-containing

cofactors such as thiamin and biotin to more appropriately represent the sulfur requirements for cell growth.

4.2.2 **Experimental data**

Experimental data are indispensable for validating an *in silico* model. The experimental data relevant to and sufficient for the purposes of this work is readily available in the published literature, as several studies on desulfurization by several strains of *R. erythropolis* exist. In this study, we use the experimental data from literature [11, 18, 20, 99] to validate our model.

Izumi *et al.*[11] isolated *R. erythopolis* D-1 strain from soil for its ability to specifically desulfurize DBT to 2-hydroxybiphenyl (HBP) successfully. In their study, they used a sulfur-free glucose-based biosynthetic medium containing 5.0 g glucose, 0.5 g KH_2PO_4, 4.0 g K_2HPO_4, 1.0 g NH_4Cl, 0.2 g $MgCl_2 \cdot H_2O$, 0.02 g $CaCl_2$, 0.01 g NaCl, 10 ml metal solution, and 1 ml vitamin mixture in 1000 ml distilled water (pH 7.5). The metal solution contained 0.5 g $FeCl_2 \cdot 4H_20$, 0.5 g $ZnCl_2$, 0.5 g $MnCl_2 \cdot 4H_20$, 0.1g $Na_2MoO_4 \cdot 2H_20$, 0.05 g $CuC1_2$, 0.05 g $Na_2WO_4 \cdot 2H_20$, and 120 mmol HCl in 1,000 ml distilled water. The vitamin mixture comprised 400 mg calcium pantothenate, 200 mg inositol, 400 mg niacin, 400 mg pyridoxine hydrochloride, 200 mg p-aminobenzoic acid, and 0.5 mg cyano-cobalamin in 1000 ml distilled water. They supplemented their glucose-based medium with a sterilized solution of 0.125 mM DBT as the sole sulfur source. They grew the culture at 30 °C in 2-liter flasks containing 500 ml medium with reciprocal shaking (100 strokes per minute). During cultivation, they measured cell growth (via turbidometry at 660 nm) and DBT-HBP concentrations (via gas chromatography) at various times. They observed that medium pH decreased during

desulfurization. To convert their absorbance values (OD_{660}) into dry cell mass, we used the conversion factor of 1 OD_{660} = 0.38 gdcw/L provided in the literature [29].

In a similar study, Davoodi-Dehaghani *et al.*[99] isolated *R. erythropolis* SHT87 capable of utilizing DBT as sole sulfur source. They used growing cells of *R. erythropolis* SHT87 at 30 °C in a 50 ml of a synthetic medium containing 5.0 g glycerol, 0.5 g KH_2PO_4, 4 g K_2HPO_4, 1 g NH_4Cl, 0.2 g $MgCl_2 \cdot 6H_2O$, 0.02 g $CaCl_2$, 0.01 g NaCl, and 10 ml metal solution in 1000 ml deionized water (pH 7.2). The metal solution contained 0.5 g $FeCl_2 \cdot 4H_2O$, 0.5 g $ZnCl_2$, 0.5 g $MnCl_2 \cdot 4H_2O$, 0.1 g $Na_2MoO_4 \cdot 2H_2O$, 0.05 g $CuC1_2$, 0.05 g $Na_2WO_4 \cdot 2H_2O$ and 120 mmol HCl in 1000 ml deionized water. For cultivating their strain, they incubated 250 ml flasks containing 50 ml medium supplemented with 0.25 mM DBT, while shaking at 200 rpm for 120 h. During cultivation, they collected aliquots of the culture to measure cell growth by turbidometric assay at 660 nm and DBT and 2-HBP concentrations by HPLC. The experimental data from these two studies provided a sound basis for validating our proposed model. We used the cell growth and DBT-HBP concentration profiles from the above experimental studies to compute specific cell growth rates (h^{-1}) and DBT (HBP) uptake (production) rates (mmol/gdcw-h).

In addition, we used several other experimental data/observations to predict/explain known results and/or propose new hypotheses for possible experimental verification in future. For instance, we used our final model to predict/explain the experimental data of Yan *et al.* (2000) on the efficacies of carbon sources. They used *R. erythropolis* KA2-5-1 to desulfurize DBT in the presence of three carbon sources (ethanol, glycerol, and glucose). They used the medium previously described by Izumi *et al.*[11] but with different carbon sources. Among ethanol, glycerol, and glucose as sole carbon sources, they found ethanol to support the

highest growth rate and desulfurization activity. As high growth rates and high desulfurizing activities are essential for effective biodesulfurization, they concluded that ethanol would be a better source compared to glucose and glycerol.

We also simulated the experimental conditions of Olmo *et al.*[20] to study the effects of nitrogen sources on the desulfurizing activity and growth of *R. erythropolis*. They used a basal salt medium to study the effect of operational conditions and medium compositions on the desulfurization activity of *R. erythropolis* IGTS8. They reported that the addition of an ammonium salt to a glutamate-based medium enhanced growth and desulfurization activity of IGTS8, although glutamate by itself was sufficient as a dual source of carbon and nitrogen.

4.2.3 **Flux balance analysis**

A flux-based metabolic model assumes that metabolic transients are typically rapid compared to cell growth and environmental changes [191, 192]. Under this assumption, the mass balances on various metabolites give us $S \cdot v = b$, where S is the ($m \times n$) stoichiometric matrix of the reactions, m is the number of metabolites, n is the number of reactions, v is the ($n \times 1$) vector of reaction fluxes, and b is the ($m \times 1$) vector of metabolite fluxes. The fluxes are zero (nonzero) for intracellular (extracellular) metabolites. We can further augment $S \cdot v = b$ by imposing suitable bounds on various fluxes to reflect a given experimental scenario. The resulting model is an underdetermined system [156, 193] with infinite solutions for possible flux distributions. A particular solution can be obtained by optimizing an objective function that appropriately reflects the cellular goal such as biomass maximization discussed before. $Z = \sum_{j=1}^{n} c_j \, v_j$ is the most common objective function [45, 169], where v_j is the flux of reaction j and c_j is a pre-determined constant derived from cell mass

composition. We used MetaFluxNet[174] and GAMS/CPLEX 10.0[194] to solve these linear optimization problems.

4.2.4 **Reaction and gene essentiality analysis**

We used our *in silico* model to ascertain the essentiality of various intracellular reactions. We froze one reaction at a time by setting its flux to zero in the model. Then, we assessed if the remaining *in silico* network of the reactions could produce biomass. Zero (non-zero) biomass growth confirmed that the stopped reaction is essential (non-essential). To assess gene essentiality, we deleted one gene at a time by freezing all its associated reactions and checking if the remaining model could synthesize biomass. If two or more genes catalyzed a particular reaction, then we did not freeze it. Using the same argument as with reaction essentiality, we identified genes whose deletions might be lethal to *R. erythropolis*. Furthermore, since reaction and gene essentiality can also be conditional on medium composition [195], we analyzed reaction/gene essentiality for seven media with varying carbon sources (ethanol, succinate, glutamate, acetate, glucose, fructose and lactate). From the seven sets of essential reactions/genes, we selected the common reactions/genes, which represent the medium-independent or true essential reactions/genes, the deletion of any of which may be lethal for *R. erythropolis*.

4.3 **Results and discussion**

4.3.1 **Reconstructed genome-scale model**

Our proposed metabolic network of *R. erythropolis* comprises 823 reactions (792 intracellular and 31 transport) and 786 metabolites. The 792 intracellular reactions

describe 57 distinct pathways (glycolysis, pentose phosphate pathway, histidine metabolism, pyrimidine metabolism, pantothenate and CoA biosynthesis, etc.) belonging to five major metabolism subsystems (central, amino acids, nucleotide, vitamins and cofactors, and cell wall). Based on the existing genome annotations in KEGG [170], our initial model accounted for 573 ORFs corresponding to 674 reactions associated with 399 genes. However, it could not synthesize several biomass precursors (e.g. methionine, histidine, biotin, etc.) and complex molecules (e.g. micothiol, mycolic acids, and arabinogalactan). It also could not uptake/metabolize certain substrates (acetate, glutamate, DBT, HBP, etc.). Therefore, we inferred and added several reactions from Metacyc[172] and the published literature on *R. erythropolis* or other organisms[8, 86, 171, 173, 196] to provide/complete pathways.

Some examples of our above additions/enhancements are as follows. *R. erythropolis* is known to produce all amino acids[197], so we identified and added from the KEGG database three reactions that enable methionine production in other organisms. In this manner, we added 39 additional reactions to enable the synthesis of various precursor metabolites. For synthesizing mycothiol[173], mycolic acids, and arabinogalactan[198], we took the relevant pathway information from available literature[172, 173, 196]. For cell envelope metabolism, we took the reactions from *M. tuberculosis*[161, 162], as Daffe, McNeil *et al.*[198] have reported the cell wall composition of *R. erythropolis* to be similar to that of *M. tuberculosis*. To justify the uptake/release of compounds such as acetate, glutamate, DBT, HBP, etc., we used the available annotations for their transporters, and/or literature evidence[11, 18]. Our final model accounts for 618 ORFs out of the known 6437 corresponding to 741 reactions associated with 413 genes. Thus, we added 82 extra reactions based on our FBA studies and careful scrutiny of the available literature and experimental information.

Since we added several reactions for which genetic evidence does not exist currently, we also attempted to identify and propose possible ORFs in PR4 that might encode for these missing functions. For this, we performed BLASTp (e-cutoff value 10^{-5}) similarity searches between the translated set of genes associated with these additional reactions in various databases and the genes of *R. erythropolis* PR4. The BLASTp analyses revealed possible annotations for 28 ORFs in *R. erythropolis* (Table 4.1). Table 4.2 shows the overall properties of the final curated version of our genome-scale metabolic model. During our reconstruction process, we have continually updated our model to include any information updates in KEGG. The full model is given in Appendix B. However, the model still has 183 dead-end metabolites with missing links. Connecting these links to other cellular activities may make this model more comprehensive and effective.

4.3.2 **Model validation**

We mimicked the experiments of Izumi *et al.*[11] and Davoodi-Dehaghani *et al.*[99] using our model and compared the predicted cell growth rates with their experimental data to validate our model. Since the concentration profiles for all the components of the sulfur-free rich medium used by Izumi *et al.* [11] and Davoodi-Dehaghani *et al.* [99] are not available, and sulfur was indeed limiting[189] in their experiments, we allowed unlimited fluxes for their uptakes. Our genome-scale model predicted maximum biomass growth rates in close agreement with the experimental values (Fig.4.1). We attribute this error to various model artifacts such as incomplete biomass formulation, maintenance energy needs, missing functionalities, incomplete genome annotation, etc. that arise due to the lack of relevant experimental information for *R. erythropolis*. Clearly, a need exists for measuring complete biomass composition and

other strain-specific parameters such as maintenance energy for *R. erythropolis* for an improved model.

Table 4.1 List of new possible annotations for R. erythropolis

Enzyme	Enzyme Name	ORF	Current Annotation	NCBI accession n	E value
EC 1.11.1.6	catalase	RER_04460	catalase	YP_002763893	3*10^-93
EC 1.18.1.3	ferredoxin reductase	RER_07950	ferredoxin reductase	YP_002764242	3*10^-69
EC 1.2.2.1	formate dehydrogenase	RER_24000	D-3-phosphoglycerate dehydrogenase	YP_002765847	1*10^-31
EC 1.2.3.3	pyruvate oxidase	RER_35280	pyruvate dehydrogenase	YP_002766975	2*10^-62
EC 1.3.1.10	acyl-ACP dehydrogenase	RER_30750	enoyl-(acyl carrier protein) reductase	YP_002766522	1*10^-58
EC 1.4.7.1	glutamate synthase	RER_56700	glutamate synthase	YP_002769117	1*10^-117
EC 1.5.1.20	methylenetetrahydrofolate reductase NAD(P)H	RER_356903	5,10-methylenetetrahydrofolate reductase	YP_002767016	2*10^-134
EC 1.8.2.1	sulfite oxidase	RER_39830	sulfite oxidase	YP_002767430	3*10^-22
EC 2.1.1.79	cyclopropane-fatty-acyl-phospholipid synthase	RER_23240	cyclopropane fatty acid synthase	YP_002765771	2*10-33
EC 2.1.2.11	3-methyl-2-oxobutanoate hydroxymethyltransfe	RER_36510	3-methyl-2-oxobutanoate hydroxymethyltransferase	YP_002767098	1*10^-138
EC 2.3.1.41	beta-ketoacyl-acyl-carrier-protein synthase I	RER_36740	3-oxoacyl-(acyl carrier protein) synthase II	YP_002767121	4*10^-47
EC 2.3.1.85	fatty acid synthase	RER_02210	polyketide synthase	YP_002763668	5*10^-82
EC 2.3.1.86	acyl-AMP ligase, fadD32	RER_02200	long-chain-fatty-acid--CoA ligase	YP_002767774	5*10^-50
EC 2.3.3.8	ATP citrate synthase	RER_44280	succinyl-CoA synthetase subunit alpha	YP_002767875	7*10^-18
EC 2.4.1.-	glycosyltransferase mshA	RER_16170	glycosyltransferase MshA	YP_002765064	1*10^-159
EC 2.4.1.15	alpha,alpha-trehalose-phosphate synthase	RER_09680	alpha,alpha-trehalose-phosphate synthase	YP_002764415	2*10^-99
EC 2.4.2.34	arabinosyl transferases aftA	RER_02330	arabinosyltransferase AftA	YP_002763680	8*10^-168
EC 2.6.1.44	L-alanine-glycine transaminase	RER_46170	phosphoserine aminotransferase	YP_002768064	1*10^-7
EC 3.1.1.23	acylglycerol lipase	RER_53090	monoacylglycerol lipase	YP_002768756	2*10^-82
EC 3.1.1.3	6-phosphogluconolactonase	RER_34630	lipase	YP_002766910	1 *10^-73
EC 3.1.2.2	palmitoyl-CoA hydrolase	RER_29140	acyl-CoA thioesterase	YP_002766361	3*10^-64
EC 3.1.3.12	trehalose-phosphatase	RER_09680	alpha,alpha-trehalose-phosphate synthase	YP_002764415	3*10^-89
EC 5.1.3.2	UDP-glucose 4-epimerase	RER_28000	UDP-glucose 4-epimerase galP	YP_002766247	8*10^-26
EC 6.3.2.7	UDP-N-acetylmuramoylalanyl-D-glutamate--L-ly	RER_35570	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diami	YP_002767004	2*10^-32
EC 1.1.3.21	glycerol-3-phosphate oxidase	RER_46040	glycerol-3-phosphate dehydrogenase	YP_002768051	4*10^-52
EC 1.17.3.2	hypoxanthine oxidase	RER_33910	xanthine dehydrogenase	YP_002766838	1*10^-38
EC 1.8.99.2	phosphoadenosine phosphosulfate reductase	RER_20430	succinate dehydrogenase flavoprotein subunit	YP_002765490	1*10^-11
EC 2.8.1.2	mercaptopyruvate sulfurtransferase	RER_39860	3-mercaptopyruvate sulfurtransferase	YP_002767433	4*10^-62

Table 4.2 Features of the reconstructed genome scale model of *R. erythropolis*

Features	Properties
Reactions in genome scale mode	823
No of ORFs included	618
Gene-associated reactions	741
Non gene-associated reactions	84
Intracellular reactions	792
Transport reactions	31
Metabolites in genome scale model	786
Internal metabolites	755
External metabolites	31

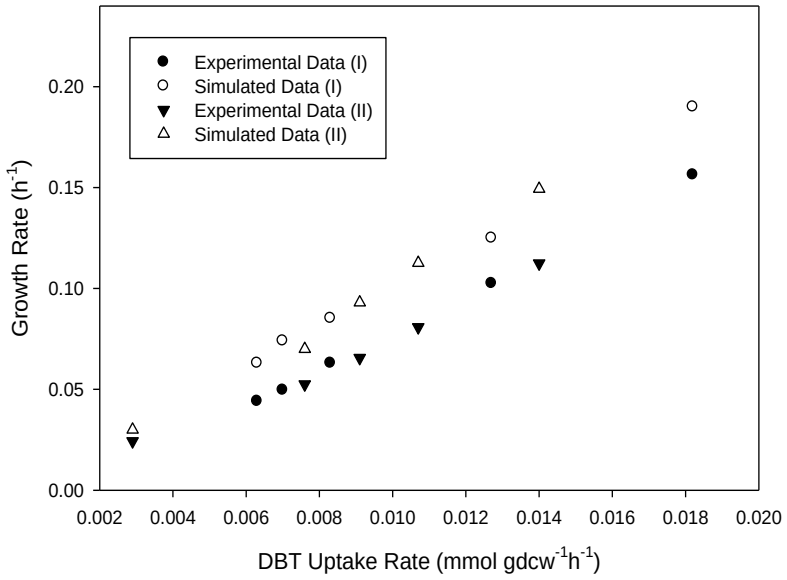


Fig. 4.1 Experimental and simulated growth rates at various DBT uptake rates from Izumi et al.[10] (I) and Davoodi-Dehaghani et al. [99] (II)

4.3.3 **Essential genes and reactions**

Gene/reaction essentiality analyses are instructive in knowing how gene knockouts and redundancy may affect growth and/or metabolism. They can guide strain improvement strategies aimed at reshaping the flux distributions, because one must not knock out any essential gene or reaction. Similarly, one may not want to target a reaction catalyzed by multiple genes, as that would require multiple knock-outs.

Since gene/reaction essentiality may depend on medium composition, we identified essential genes/reactions for seven different media based on different carbon sources (acetate, ethanol, glucose, fructose, lactate, glutamate, glycerol). The seven essential sets contained 272 reactions in common. These reactions are thus essential irrespective of the media. As shown in Fig. 4.2, cell wall biosynthesis, amino acid metabolism, and vitamins and cofactors metabolism have the highest percentages of essential reactions. These are the most vulnerable subsystems in *R. erythropolis*. On the other hand, central metabolic pathways (PPP, TCA, cycle, and Pyruvate metabolism) have several non-essential reactions. This is in agreement with observations on other bacteria [199], and it signifies the ability of a microorganism to survive against several genetic perturbations.

We identified 135 essential genes and 199 uniquely associated essential reactions. While this number is nearly half of 272 (the number of essential reactions), two or more isozymes catalyze approximately 12% of the essential reactions and some genes control multiple reactions. While these reactions are essential at the metabolic level, they are not so at the genetic level. Furthermore, the essential reaction set includes 38 non-gene-associated reactions. Identification of their encoding genes would also increase the number of essential genes.

The essential genes and reactions common to the seven subsets include the components of the 4S pathway for the survival on DBT. However, if the DBT supply was substituted or supplemented by other sulfur sources such as sulfate, cysteine, or methionine, then the expression of these genes was not essential for biomass growth. From this study, it is clear that the expression of the *dsz* genes is essential for a strain to utilize DBT and exhibit desulfurization activity. Moreover, as several sulfur-containing cofactors and building blocks participate in some essential reactions, any

cellular malfunction that prevents it from utilizing DBT when supplied solely, can lead to the loss of desulfurizing activity and cell death.

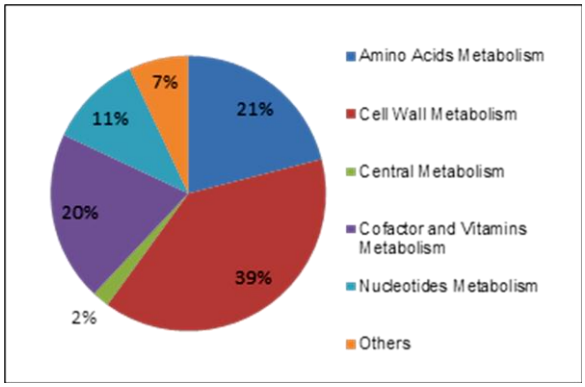


Fig. 4.2 Distribution of various essential reactions among cellular subsystems

4.3.4 **Minimal medium**

Izumi et al.[11] supplemented their glucose-based synthetic medium with a vitamin mixture and a metals solution for growing *R. erythropolis*. However, it would be instructive to identify the minimal media for the same, where an *in silico* model can help. To this end, we simulated growth in the absence of one component at a time from the rich medium. A component must exist in a minimal medium, if its absence results in zero growth. With this, we identified glucose, oxygen, ammonium salt, iron salt, phosphorus source, and DBT as the essential components of a minimal medium. Any of the 12 carbon sources (acetate, citrate, ethanol, fructose, fumarate, formate, gluconate, glycerol, lactate, malate, pyruvate, and succinate) can replace glucose in a minimal medium, and glutamate can replace both glucose and ammonium salt. The vitamins on the other hand are not essential, but they enhance growth. This suggests that the addition of vitamins may be desirable, but not essential, in desulfurization experiments.

4.3.5 **Desulfurization in the presence of cysteine and methionine**

Sulfur is essential for the synthesis of several key metabolites such as cysteine, methionine, coenzyme A, thiamin, etc. that are required for the cellular activities and growth. Therefore, the various *R. erythropolis* strains have evolved diverse metabolic pathways to utilize various sulfur sources such as sulfate, cysteine, methionine, DBT, dimethyl sulfoxide, etc. Although they can utilize DBT as the sole sulfur source, they do not do so in the presence of sulfate, cysteine and methionine [50]. Most available literature has argued that sulfate, cysteine, and methionine repress the *dsz* gene in *R. erythropolis*, which deactivates the conversion of DBT via the 4S pathway.

However, in a study[51] of the wild type and two mutant strains of *R. erythropolis* KA2-5-1, the authors observed no DBT desulfurization activity for the wild type in the presence of either methionine or cysteine, but significant enzymatic activity for the mutants in the presence of methionine. Their gene analysis of the mutants revealed a disruption in the *cbs* gene that encodes for cystathionine β-synthase, an enzyme responsible for converting methionine to cysteine. They suggested that high desulfurizing activity in mutants might be due to reduced amounts of cysteine in cells lacking cystathionine β-synthase activity. Thus, they concluded that methionine does not repress *dsz* genes. On the other hand, no desulfurization activity by the mutants in the presence of cysteine led them to opine that cysteine does repress *dsz* gene.

Our model reproduced the above observations of Tanaka *et al.*[51] successfully. It showed zero DBT uptake (i.e., no desulfurization activity) in the presence of either methionine or cysteine for the wild type strain. When modified suitably (disabled *cbs* gene as per the observation of Tanaka *et al.*[51]) for the mutant, it confirmed significant DBT uptake in the presence of methionine, and zero uptake in

the presence of cysteine. Having reproduced the experimental observations, we now use our model to (1) explain why the mutants uptake DBT in the presence of methionine, (2) propose, contrary to the current opinion, that cysteine does not repress *dsz* gene, (3) devise a strain that uptakes DBT, even in the presence of both methionine and cysteine, and (4) reason why the wild type shows no desulfurization in the presence of cysteine or methionine.

The DBT uptake in the presence of methionine is simply explained by the fact that cysteine is a sulfur-containing amino acid essential for cell growth. The mutants, because of their disfunctional *cbs* gene are unable to produce cystein, use the alternate sulfur-source (DBT) to produce cysteine needed for their growth. However, it is worth noting that the mutant showed nonzero uptake of methionine as well with DBT, even though DBT is supposedly sufficient for all the sulfur needs of *R. erythropolis*. We will explain this later.

To show that cysteine like methionine may not repress *dsz* gene, we simply recall that the mutants used DBT to produce the essential-for-growth cysteine, because they could not use methionine to do the same. Thus, if we can design a strain that is unable to convert cysteine to any sulfur-containing metabolite that is essential for growth, then such a mutant would take DBT even in the presence of cysteine. To this end, we used trial and error to identify genes whose knockouts would prevent the production of an essential sulfur-containing metabolite from cysteine. We found from our model that the simultaneous knockout of *cbs* gene and the gene for cysteine desulfhydrase forces the strain to use DBT in the presence of cysteine. While it should be possible to find some other gene knockouts that would have the same effect, this clearly shows that an appropriately designed strain may uptake DBT even in the presence cysteine, just like what Tanaka *et al.*[51] observed for methionine. In other

words, we suggest that cysteine may not repress *dsz* gene. Interestingly, when we supplied methionine in addition to cysteine, the same strain stopped taking DBT. Again, we explain this later, when we consider desulfurization in the presence of methionine and cysteine.

Having shown desulfurization in the presence of cysteine, it is easy to show the same, even when both cysteine and methionine are present. Using our trial and error approach, we found that the simultaneous removal of genes for cysteine sulfhydrase, cystathionine gamma-lyase, and cystathionine beta-synthase could confer the strain the ability to exhibit desulfurization.

Now, we reason why the wild type does not uptake DBT, when cysteine and/or methionine are present. The explanation is the same as what we proposed in chapter 3 for reduced desulfurizing activity in the presence of sulfate. This concerns the additional energy (4 moles of NADH) required for utilizing every mole of DBT via 4S pathway as a sulfur source. Sulfate, cysteine, and methionine all require lower energy than DBT to produce essential sulfur-containing metabolites. Thus, when any of them is present, the cell prefers them over DBT due to lower energy requirement. However, whenever the cell is unable to utilize them for an essential-for-growth metabolite, then it is forced to metabolize DBT to satisfy its need for that metabolite, thus exhibiting desulfurization activity. The key is that the cell always prefers a bioavailable source, unless it is unable to utilize it as a sole source. The hypothesis of *dsz* gene repression by sulfate, methionine, or cysteine does not seem valid.

4.3.6 **Effects of medium components on desulfurization activity and growth of *R. erythropolis***

4.3.6.1 **Relative effectiveness of carbon sources**

R. erythropolis can metabolize a variety of sugars and other carbon sources [11] to produce energy, redox power, and precursor metabolites necessary for macromolecular biosynthesis. Franchi *et al.*[26] stated that the level and stability of catalytic activity and the cost of carbon source are the key factors affecting the overall cost of a desulfurization process. Therefore, we used our model to study the relative suitability of various carbon sources for cell growth and desulfurization activity. Yan *et al.*[18] studied the relative efficacy of ethanol, glucose, and glycerol as sole carbon sources for the growth and desulfurizing activity of *R. erythropolis*. Simulations using our more detailed and comprehensive model under conditions identical to their experiments successfully predicted the relatively greater effectiveness of ethanol (versus glucose followed by glycerol) for higher growth rate. They also reconfirmed the discussion and elucidation of this based on the additional NADH needed for the 4S pathway.

Since our present model is more comprehensive than the reduced model of presented in chapter 3, we could also study the suitability of additional carbon sources for achieving higher desulfurizing activity. We compared acetate, citrate, ethanol, fructose, fumarate, formate, gluconate, glucose, glycerol, glutamate, lactate, malate, pyruvate, and succinate as sole carbon sources for the growth and desulfurizing activity of *R. erythropolis*. Fig. 4.3 shows the results of our 14 separate simulations for an identical uptake rate of 20 mg/gdcw-h for each sole source. Ethanol still proves to be the most effective carbon source, and the desulfurization and growth rates relative to those of ethanol decrease as: ethanol (2 μmol HBP/gdcw-h as 100% & 0.021 h^{-1} as

100%) > acetate = fructose = glucose = lactate (76.5%) > glycerol (75%) > glutamate (74%) > succinate (72%) > pyruvate (69%) > gluconate (67.8%) > fumarate (62.7%) > citrate (56.6%) > malate (54%) > formate (26.5%). In contrast to our earlier model, the biomass equation in this work includes the stoichiometric coefficients for carbon-based precursors as well. Thus, these results are more comprehensive in that they account for the contribution of carbon source to growth and other metabolic activities besides desulfurization.

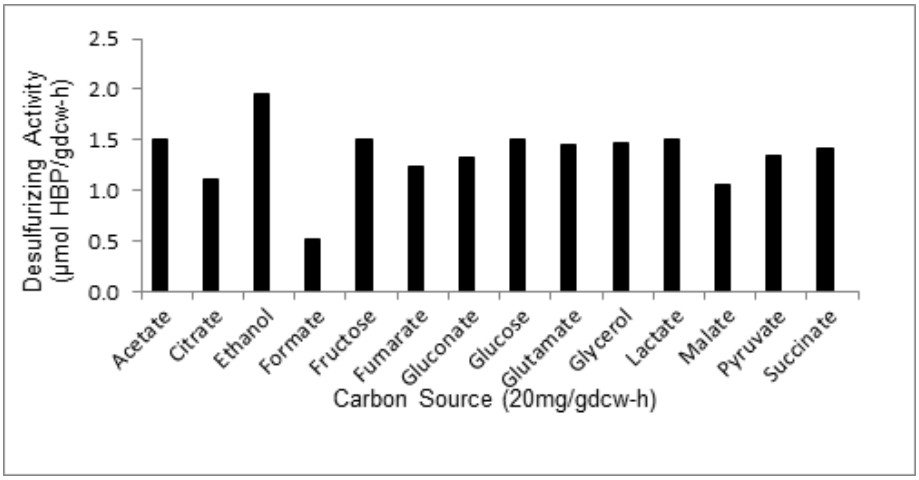


Fig. 4.3 Specific desulfurizing activities for an uptake rate of 20mg/gdcw-h of various carbon sources

4.3.6.2 **Combined effect of carbon and nitrogen sources**

Both carbon and nitrogen sources affect desulfurizing activity. Thus, it is important to study them as a combination rather than just individually. We considered nitrate and ammonium salts as two sole nitrogen sources, and acetate, citrate, ethanol, fructose, formate, fumarate, gluconate, glucose, glycerol, glutamate, lactate, malate, pyruvate, and succinate as 14 sole carbon sources. Thus, we examined 28 combinations of carbon and nitrogen sources. For all 28 simulations, we allowed unlimited supplies of the specific carbon and nitrogen sources, fixed the DBT uptake at 20 mg/gdcw-h to achieve an identical level of desulfurizing activity, and maximized the biomass. The simulations yielded the amounts of carbon and nitrogen sources needed to achieve the

desired desulfurization activity. For any given nitrogen source, the carbon source uptakes naturally vary, as we discussed in the previous section. However, the uptake of the nitrogen source remains constant for all carbon sources except glutamate. We may explain this as follows. The DBT uptake fixes the sulfur supply, and thus the biomass growth. A given biomass growth requires a fixed nitrogen supply, and thus the need for nitrogen is the same for all carbon sources. However, glutamate is unique among the carbon sources, as it can supply both carbon and nitrogen. Thus, when glutamate is the carbon source, the cell needs relatively smaller amount of the "pure" nitrogen source (nitrate 0.25 mg/gcdw-h; ammonia 0.1 mg/gdcw-h) than the amounts (nitrate 0.78 mg/gcdw-h; ammonia 0.21 mg/gdcw-h) required in presence of other carbon sources. Our simulations with glutamate as the sole substrate confirmed the dual role of glutamate, and also the experimental observations by Olmo *et al.*[20] using *R. erythropolis* IGTS8. Furthermore, the fact that the organism does use the "pure" nitrogen source in addition to glutamate suggests that nitrogen in glutamate is limited for growth and desulfurization.

A comparison of the two nitrogen sources reveals that ammonium is a better nitrogen source than nitrate, as it needs relatively lesser carbon source (Fig. 4.4). This may be explained by the fact that the cell must first reduce nitrate into ammonium by using energy in the form of reducing equivalents or NAD(P)H. Thus, nitrate needs more carbon to supply that extra energy. Olmo *et al.*[20] also observed that the addition of 2 g/L ammonium salt to 20 g/L glutamate increased both growth and desulfurization as compared to 20 g/L glutamate alone. To predict this phenotype, we performed two simulations maximizing growth rate with an uptake of 20 mg/gdcw-h glutamate. In the first simulation, we allowed no ammonium. In the second, we allowed up to 2 mg/gdcw-h of ammonium. The desulfurizing activity (1.4 µmol/gdcw-

h) and growth rate (0.015 h^{-1}) were higher for the second run than those (0.9 µmol/gdcw-h; 0.01 h^{-1}) for the first. In other words, as observed by Olmo *et al.*[20], the organism is able to utilize the additional nitrogen in the form of ammonia and increase its growth, given that nitrogen in glutamate is limiting as discussed previously.

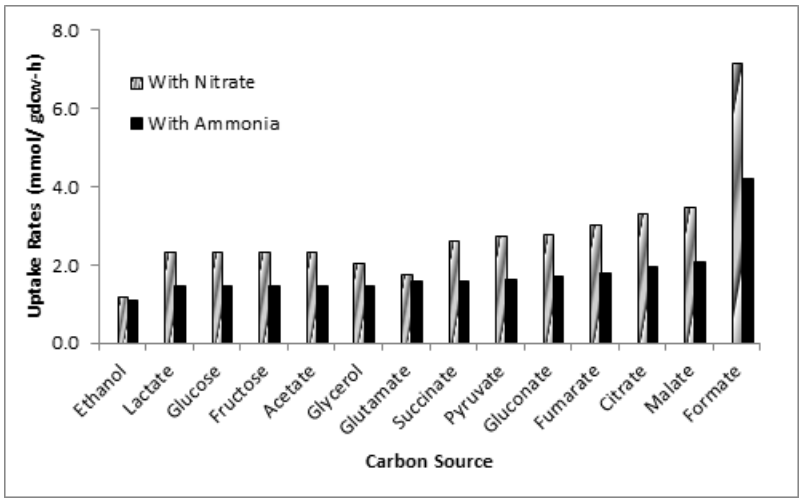


Fig. 4.4 Consumption rates of various carbon sources in the presence of various nitrogen sources for a desulfurizing activity of 20mg/gdcw-h.

For using *R. erythropolis* to desulfurize fossil fuels, one would need to grow this biocatalyst in larger quantities. As with carbon and nitrogen, carbon and sulfur sources also affect its growth rate. We compared 42 combinations of three sulfur sources (cysteine, methionine, and DBT) with the aforementioned 14 carbon sources to identify the combination that would give the highest cell growth rate.

For each of the 42 combinations, we maximized biomass for a fixed supply (20 mg/gdcw-h) of the sulfur source and unlimited supply of the carbon source. We observed that cysteine supported the highest growth rate (1.76 h^{-1}) with all the 14 carbon sources as comapred to other sulfur sources (methionine 1.43 h^{-1}; DBT 1.16 h^{-1}). Furthermore, we observed that of the 14 carbon sources, ethanol supported the highest specific growth rate for all sulfur sources. Therefore, the combination of cysteine and ethanol gave the highest specific growth rate. With cysteine as the sole

sulfur source, the relative order of effectiveness for the 14 carbon sources is ethanol > fructose = glucose = lactate = acetate > glutamate > glycerol > succinate > pyruvate > gluconate > fumarate > citrate > malate > formate.

These studies show that ethanol is the best carbon source of the 14 examined above, irrespective of the nitrogen and sulfur source used. However, the order of relative effectiveness of the remaining 13 carbon sources varied slightly with each combination as evident from (Fig. 4). For instance, glutamate was the second most effective carbon source after ethanol with nitrate as the nitrogen source. However, with ammonia as the nitrogen source, it was the seventh most effective carbon source. This shows that the order of effectiveness of an elemental source for growth or desulfurization may depend on the other components in the medium. We can explain the change in the relative order of glutamate on the basis of its dual-source nature. As discussed before, nitrate has to go through ammonium before it can be assimilated into the cell mass, and this requires additional NADH. While ethanol readily provides this additional NADH[189], extra amounts of the other carbon sources are required for the same. However, since glutamate can supply both carbon and nitrogen, it reduces the need for nitrate, and thus the corresponding additional energy requirement. Thus, a relatively lower amount of glutamate can support the same desulfurizing activity and biomass in the presence of nitrate than other carbon sources (except ethanol).

4.3.7 **Comparison with *E. coli* model**

E. coli is a widely studied and genetically well characterized organism with several industrial applications[177]. However, its wild type is not suitable for desulfurization, as it does not have the *dsz* or any other genes for desulfurization. Since it is relatively easier to reengineer, we can design an *in silico* mutant of E coli, which could show

desulfurization. Then, it would be interesting to compare *R. erythropolis* and that mutant for their desulfurization activities. To this end, we added the *dsz* genes into a literature model (iAF260)[176] of *E. coli* to obtain the model for such a mutant. By comparing this model with that of *R. erythropolis*, we observed several similarities in their central metabolisms and amino acid biosyntheses.

Next, we simulated the models of *R. erythropolis* and the *E. coli* mutant under similar conditions and obtained their desulfurizing activities. We supplied fixed amounts of ethanol (20 mg/gdcw-h) and unlimited supplies of other external metabolites (DBT, oxygen, ammonium salt, etc.). Interestingly, the mutant showed a higher desulfurization rate, but a lower growth rate (4.35 µmol HBP/gdcw-h; 0.017 h^{-1}) than *R. erythropolis* (2 µmol HBP/gdcw-h; 0.021 h^{-1}). In order to explain this phenotype of *E. coli*, we performed another set of simulations. We allowed unlimited supplies of ethanol, DBT, and other nutrients and determined the minimum amount of DBT uptake required for a unit specific growth rate (1 h^{-1}) for both *E. coli* mutant and *R. erythropolis*. The minimum DBT uptake rate was much higher for *E. coli* (2.5 µmol HBP/gdcw-h) than *R. erythropolis* (0.09 µmol HBP/gdcw-h). In addition, the corresponding ethanol uptake (31.57 mmol/gdcw-h) was also much higher for *E. coli* than that for *R. erythropolis* (20.8 mmol/gdcw-h) with a significant amount of CO_2 release. In other words, both carbon (ethanol) and sulfur (DBT) requirements of the mutant are much higher than those for *R. erythropolis*. In the mutant *E. coli*, some ethanol is diverted to CO_2 synthesis. These simulations explain why *E. coli* showed a lower growth but higher desulfurization rate than *R. erythropolis* for a fixed ethanol uptake.

Our above study clearly shows that a genetically modified *E. coli* strain may act as an efficient desulfurizing strain. In fact, it may be possible to promote growth

and consequent desulfurization by limiting the diversion of the carbon flux to other byproducts such as CO_2. Furthermore, this study shows that higher sulfur requirements for growth by an organism may confer it a higher desulfurizing ability. Since the elemental biomass composition of an organism may alter with experimental conditions[200], the identification of conditions that can relatively increase the cellular sulfur content of *R. erythropolis* may lead to its improved desulfurization activity.

4.4 **Summary**

This chapter presents the first genome-scale model for the metabolic network of *R. erythropolis*. It proposes key additional information and new genome annotations to account for missing functionalities in the available genome-scale data on *R. erythropolis*. Our reconstructed model successfully predicts and explains several experimental observations available in the literature on *R. erythropolis*. Our *in silico* studies with appropriate mutants of *R. erythropolis* indicate that cysteine, like methionine, may not repress *dsz* genes. They also show that appropriately designed strains can exhibit desulfurizing activity in the presence of bioavailable sulfur sources such as cysteine and methionine. The gene essentiality analyses highlight that the maintenance of cell membrane's integrity is crucial for a biocatalyst's growth and thus, its functioining. However, some organic compounds including ethanol are known to alter cell membrane's fluidity and functions, and thus exhibit toxicity to microbial cells. Our comparative studies on carbon sources gives alternate sources that could preferentially be used to grow the desulfurizing biocatalyst, as higher concentrations of ethanol may be lethal to it. Our comparative study of *in silico* desulurizing strain of *E. coli* and *R. erythropolis* shows that sulfur requirements for growth play an improtant role in determining the extent of desulfurization.

5 Roles of sulfite oxidoreductase and sulfite reductase in improving desulfurization by *Rhodococcus erythropolis*

5.1 Introduction

Numerous works are available on the characterization of the desulfurizing *Rhodococci* and their activities [11, 13, 85]. A desulfurization activity (DA) of nearly 1.2-3.0 mmol/gdcw-h [5, 6, 111] is desirable for a competitive commercial biodesulfurization process. However, the desulfurization activities of the wild type strains are too low for commercialization. Several studies exist on optimal conditions [23, 24, 189] and media compositions [18-20, 201, 202] for obtaining higher levels of desulfurization. However, sulfur metabolism in general, and the metabolism of sulfur derived from the desulfurization of DBT in particular, are not well understood in *Rhodococcus*. For instance, the 4S pathway is known to produce sulfite, but it is unclear how the cell (1) incorporates sulfite into biomass, and (2) excretes sulfate [3].

A major challenge is to enhance DA, as the activity of wild strains is too low for an industrial process [26]. Genetic engineering has been used to create strains with improved DA. These strains have involved alterations [27, 103, 117, 118] to copy number of *dsz* gene(s), promoters expressing *dsz* genes, *dsz* operon, and genes that supply cofactors needed by Dsz enzymes and have achieved up to 25-fold increase in DA as compared to the parent strains. However, all the above strategies have focused almost exclusively on 4S pathway elements, and they have not achieved desired activity levels. Several investigators have demonstrated that different bacterial hosts

containing identical *dsz* genes exhibit different DA, substrate range, and temperature ranges for the 4S pathway [15, 26]. Clearly, there is a need to look beyond the 4S pathway.

Repression of DA in both the parent and genetically modified strains has been observed in the presence of sulfate. Therefore, some researchers [26, 97] have designed recombinant strains using promoter deletion and replacement strategies to overcome sulfate repression. However, Franchi et al. [26] observed that the activity of the promoter-replaced strains was also reduced in the presence of sulfate by nearly 50%. Therefore, they postulated sulfate to exert additional regulatory effects on desulfurization.

In chapter 4 we presented a genome-scale *in silico* metabolic model of *Rhodococcus erythropolis* and showed it to be useful in determining preferred carbon sources for optimum DA. In this study, we expand the metabolic model of *R. erythropolis*, specifically to better address sulfur metabolism and we present hypotheses regarding the incorporation of sulfur from DBT into biomass, and the key roles of SOR and SR in DA. Based on these insights, we propose strategies for improved desulfurization.

5.2 Materials and methods

5.2.1 Experimental measurements

Our primary source of experimental data is Omori et al. [203]. They measured pH, cell growth, DBT concentration, HBP concentration, and sulfate concentration at various times during their study on DBT desulfurization by *R. erythropolis* SY1 [16, 203]. They used 5 ml of an ethanol-based sulfur-free minimal medium supplemented with 50

µl of 30 mM DBT-ethanol solution at 30 ºC to examine the activity of growing cells. For analyzing and predicting various literature observations qualitatively and quantitatively, we used the genome-scale metabolic model [204] of *R. erythropolis* as the basis. We employed the technique of flux balance analysis [166] to analyze the intracellular flux distributions from our model. As discussed later, we also enhanced our model to include additional functionalities.

5.3 **Results and discussion**

The growth-related sulfur need of *R. erythropolis* is low (about 0.5%-1% of dry cell weight), which naturally constrains its DA. Clearly, if we wish to achieve desulfurization rates beyond this natural limit, then we must either increase this nutritional requirement for sulfur by engineering the environment and/or the genome of *R. erythropolis*, or induce *R. erythropolis* to accumulate/excrete sulfur in some form.

The discussions by Kilbane and LeBorgne [3] suggest a clear need for thoroughly understanding the metabolism of sulfur in the 4S pathway. They make it clear that factors other than the *dsz* genes need to be considered, if optimum desulfurization activity is to be achieved. While sulfur atom in DBT is oxidized to sulfite via the 4S pathway before getting assimilated into biomass [5, 92], the specific enzymes and pathways employed by *Rhodococcus* for the utilization of sulfite derived from DBT desulfurization are not well understood. A major inexplicable phenomenon in desulfurizing strains is the formation and excretion of sulfate when sulfite is the major sulfur metabolite obtained from DBT. Thus, a key issue that must be addressed in an improved metabolic model of *R. erythropolis* is the observation that sulfate is

present extracellularly in desulfurization-competent cultures utilizing DBT as the sole source of sulfur [11, 16, 29, 203].

5.3.1 **Sulfate formation in 4S pathway**

We propose two alternate plausible hypotheses for explaining the sulfate formation from sulfite during DBT desulfurization. The first is based on the presence and toxicity of excess sulfite, and the second is based on enzyme compartmentalization.

5.3.1.1 **Toxicity of excess sulfite**

Sulfite is toxic to microbial cells [205, 206], so the cells would naturally avoid its accumulation. However, being a proven intermediate metabolite in the 4S pathway, its accumulation inside *R. erythropolis*, even if temporary, is highly plausible. The direct incorporation of sulfite derived from DBT into biomass may occur; however, the rate of sulfite utilization may be less than the rate of sulfite formation by the 4S pathway resulting in the intracellular accumulation of toxic levels of sulfite [205, 206]. Assuming that this happens, *R. erythropolis* must have a mechanism to limit the toxic effect of sulfite. Searching for such a mechanism, we learnt that prokaryotes can indeed oxidize sulfite [206-208] to sulfate via two pathways: direct and indirect. SOR catalyzes the former, and a reductase and a sulfurylase effect the latter. To our knowledge, the presence of SOR in *R. erythropolis* has been neither hypothesized nor confirmed in the literature. However, our BLASTp analysis using the amino acid sequence of SOR (E.C. 1.8.2.1) located an ORF with the gene name and locus tag of RER_39830 in the genome of *R. erythropolis* PR4. This can encode for SOR that can convert toxic sulfite to non-toxic sulfate.

Having explained the formation of sulfate in *R. erythropolis*, we now elucidate sulfite accumulation, especially when *R. erythropolis* may have no incentive to produce sulfite beyond its growth-related needs. Consider scenarios in which sulfur is limiting for cell growth. For assimilation into biomass, perhaps sulfite is first converted to sulfide by sulfite reductase (SR) (EC 1.8.1.2). Ploeg et al. [209] reported that SR is encoded by the *cysJI* genes in *Bacillus subtilis*. They also observed that sulfate induces the expression of *cysJI*, and any deletion in *cysJI* blocks the utilization of sulfite or sulfate by the organism. Using the model presented in chapter 4, we studied the role of SR in sulfur metabolism in *R. erythropolis*. For this, we inactivated the SR enzyme by restricting the flux through the associated reaction to be zero. Under this condition *R. Erythropolis* was unable to exhibit any *in silico* growth when sulfite and sulfate were provided solely or in combination as sulfur source(s). This confirmed that SR is essential for utilizing sulfite or sulfate by *R. erythropolis*. We hypothesize that SR is probably not active in *R. erythropolis* initially in the absence of sulfate. Thus, the sulfite obtained from DBT metabolism would accumulate. This intracellular accumulation of toxic sulfite would make *R. erythropolis* activate sulfite oxidoreductase (SOR) to convert it into sulfate. The sulfate would subsequently activate the *cysJ* gene encoding for SR in *R. erythropolis* and the resulting SR would then begin converting both sulfite and sulfate into sulfide for growth. However, for any reason, if any excess sulfite remains, the active SOR would convert it and excrete as sulfate. This is indeed supported by the experimental observations of sulfate as the only excreted sulfur metabolite and in small amounts [11].

5.3.1.2 **Enzyme compartmentalization**

Alternately, it is also possible that the enzymes involved in the sulfite-to-biomass conversion are compartmentalized in the bacterial cell in a location distinct from that of the Dsz enzymes. In this case, even though it makes energetic sense to feed sulfite from the 4S pathway directly into biomass production, the reality is that the production of sulfite by Dsz enzymes creates metabolic stress that is relieved by converting sulfite to sulfate via SOR (as explained previously) so that the sulfur can be routed to the enzyme complex equipped to incorporate it into biomass.

The conversion of all, or some, of the sulfite to sulfate can be understood as a way to avoid intracellular concentrations of sulfite from causing toxicity, but it is not obvious why sulfate formed from sulfite oxidation would be exported outside the cell. Keeping sulfate intracellularly so that it can be incorporated into biomass avoids any energy expenditures associated with the exporting and then re-importing sulfate, yet sulfate is found extracellularly in desulfurization-competent cultures [11, 16, 29, 203]. Moreover, in cases of co-cultures, *R. erythropolis* excretes sulfate that can be readily consumed to meet the sulfur requirements of other bacteria that do not possess *dsz* genes [12]. It is not known if all or only some of the sulfur liberated from the metabolism of DBT by the 4S pathway is released/available as extracellular sulfate, but when co-cultures of desulfurization-competent and desulfurization-deficient bacterial strains are grown with DBT serving as the sole source of sulfur, the population of the desulfurization-deficient culture can exceed the population of the desulfurization-competent culture by ratios as high as 1000-to-1 [12].

This suggests that the entirety of sulfur derived from DBT may be converted from sulfite to sulfate and then sulfate is exported outside the cell before being imported again and incorporated into *R. erythropolis* biomass. The observation by

Omori et al. [203] that adding barium chloride to the growth medium of desulfurization-competent cultures can stimulate the metabolism of the 4S pathway is a further confirmation that sulfite from the 4S pathway is converted into extracellular sulfate before being incorporated into biomass. The explanation why sulfate is found extracellularly may be related to enzymatic compartmentalization, and it certainly merits further investigation.

Another unresolved topic in the metabolism of sulfur by *R. erythropolis* is what enzyme(s) and pathway(s) are used in converting sulfite to sulfate and extracellular sulfate into biomass. Fig. 3.1 shows the pathways for the metabolism of sulfur from DBT to various metabolites. Since sulfite is one step closer to being incorporated into biomass than sulfate, it would seem to be advantageous for the 4S pathway to produce sulfite rather than sulfate. Furthermore, the reduction of sulfite directly into sulfide seems less energy intensive than via sulfate formation. Therefore, an understanding of sulfite-to-sulfate inter-conversions is also essential.

5.3.2 **Sulfite-sulfate inter-conversions in DBT metabolism**

We investigated the possible metabolic routes for the interconversion of sulfite to sulfate and five metabolic pathways were found as shown in Fig. 5 1. Of these routes, the enzymes (marked green) for pathways II, and III have been identified [204] to be present in *R. erythropolis* based on its latest genome annotations, and our BLASTp analyses revealed the possible ORFs encoding for enzymes (marked red) of pathways I, IV and V in the genome sequence of *R. erythropolis*. As seen in Fig. 5.1, the pathways differ in the energy and cofactor requirements. We used our genome scale *in silico* model to investigate the preferred path for conversion of sulfite to sulfate. With all five paths active, our *in silico* cell seems to prefer paths based on ATP need. With

zero ATP needed, pathways V (EC 1.8.4.9) is the most preferred, while pathway III (EC 1.8.99.2) and pathway IV (EC 1.8.4.10) are preferred over pathway II (EC 1.8.4.8). That is, III is taken, only when V is blocked. If pathways V and III are blocked, cells utilize sulfate via IV, and II is taken only when pathways I, III, IV, and V are all blocked. ATP requirement seems to be the key, as the ATP needed is the highest for pathway II and zero for pathways V.

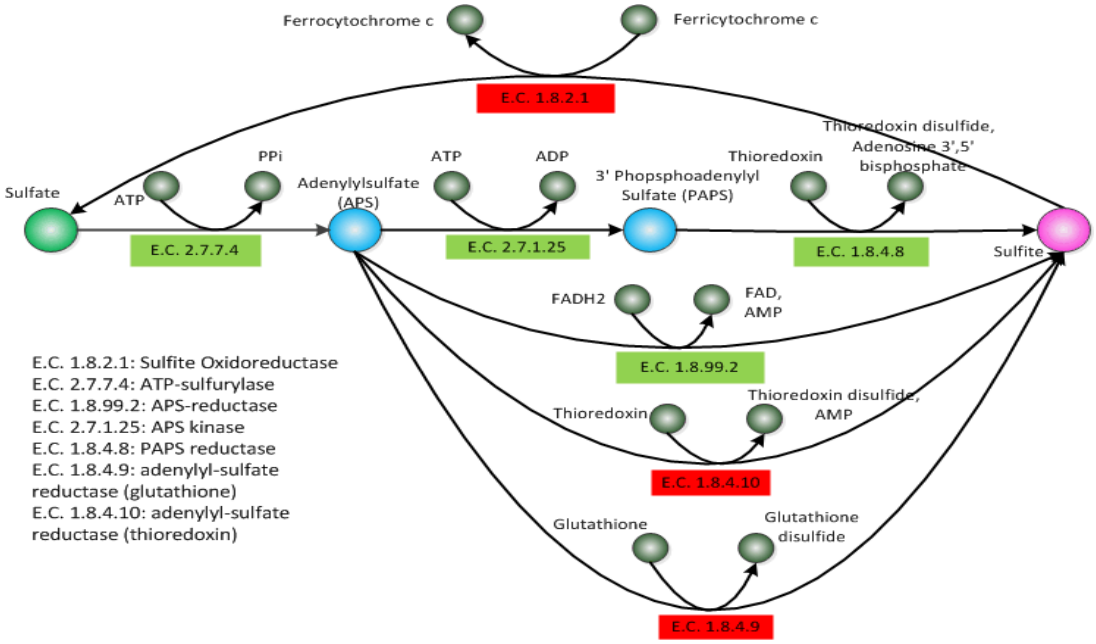


Fig. 5.1 Schematic showing pathways for oxidation of sulfite to sulfate with associated enzymes

The five pathways shown in Fig. 5.1 have different directionalities. It may be that one of the pathways is used for the conversion of intracellular sulfite into extracellular sulfate, while another pathway is used for the conversion of extracellular sulfate into intracellular sulfite that leads to incorporation into biomass. Research is needed to explore these possibilities.

5.3.3 **Role of SOR**

To obtain microbial cultures with higher desulfurization activity, it appears essential that *R. erythropolis* desulfurize DBT in excess of its nutritional requirement for sulfur,

and excrete extra sulfite as sulfate. Our *in silico* model presented in chapter 4 without SOR confirms this easily. Without an active SOR, it is not possible to force extra flux through the 4S pathway for a fixed carbon source uptake in their steady state model, as sulfite accumulation is not possible due to toxicity. Therefore, to study the role of SOR, we updated and revalidated their genome-scale model by adding the following reaction catalyzed by SOR and simulating the experiments of Omori et al. [203] and Folsom et al. [103].

Sulfite + H2O + 2 Ferricytochrome c <-> Sulfate + 2 Ferrocytochrome c + 2 H+

Omori et al. [203] observed sulfate excretion along with DBT uptake in an ethanol-based media. To simulate these observations, we allowed unlimited ethanol uptake, but fixed the DBT uptake and sulfate production rates at their experimental values to compute maximum cell growth rates. An active SOR enables sulfate formation and thus subsequent excretion. Fig. 5.2 shows that our cell growth predictions are in reasonable agreement with experimental measurements of Omori et al. [23]. Folsom et al. [103] studied DBT desulfurization using *R. erythropolis* I-19, a recombinant strain of *R. erythropolis* IGTS8 containing multiple copies of *dsz* genes. They observed that I-19 exhibited desulfurization rates (5.0 μmol/gdcw-h) that were nearly 25 times those exhibited by IGTS8 (0.1-0.2 μmol/gdcw-h). In the absence of uptake rates and other data in Folsom [15], we proceed to estimate the maximum possible desulfurization activities using our model.

We first study the impact of SOR on desulfurization activity (DA) using our *in silico* model [204]. We fix an arbitrary ethanol uptake rate (say 1 mmol/gdcw-h). Then, we set SOR activity (sulfate formation flux) to zero and compute the maximum cell growth rate as 0.048 h^{-1}. At this growth, DA is $4.5*10^{-3}$ mmol HBP/gdcw-h, which is the maximum DA for the wild type strain with zero SOR activity. It represents

growth related sulfur requirements for *R eryhthropolis*. Since our main goal is to remove sulfur from DBT beyond these growth needs, we now allow unlimited *dsz* gene activity, and maximize DA at various SOR activity levels for the ethanol uptake of 1 mmol/gdcw-h. As seen in Fig. 5.3, DA increases linearly with SOR activity, as the active SOR provides a mechanism for removing the surplus sulfite. Fig. 5.3 clearly suggests that SOR controls DA in the absence of other factors such as *dsz* genes.

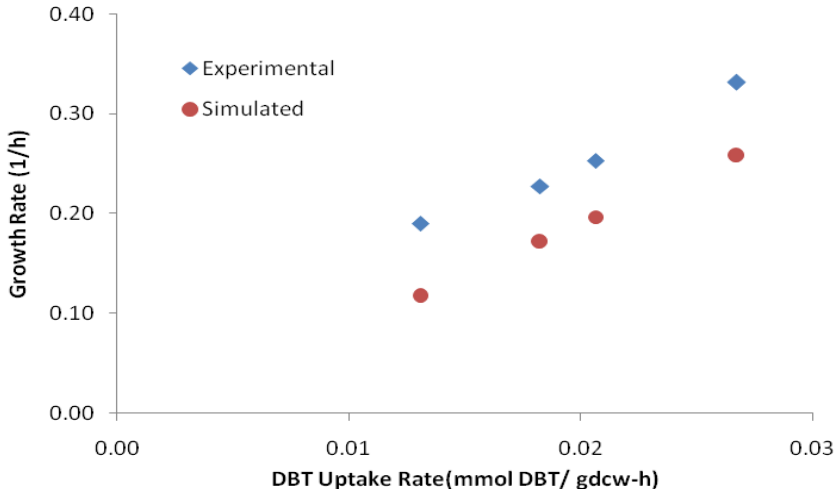
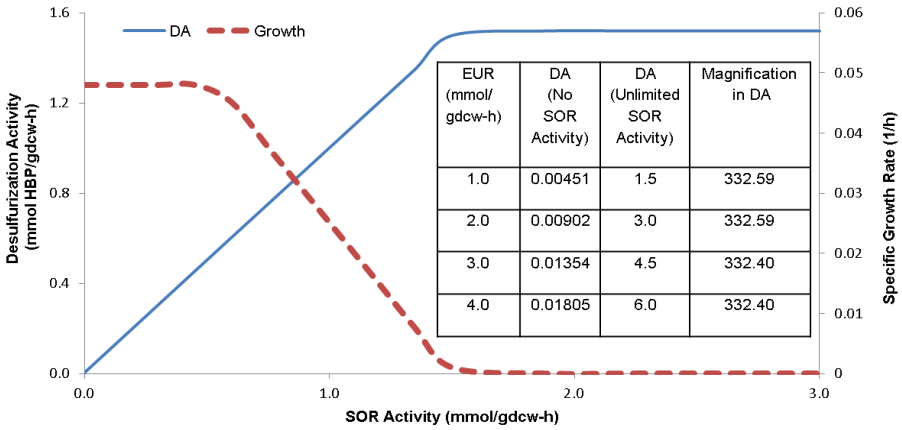


Fig. 5.2 Cell growth vs. DBT uptake for the data from Omori et al. [23]. The simulations involve fixing DBT uptake rates and sulfate production rates at their experimental values and then maximizing cell growth.



EUR (mmol/ gdcw-h)	DA (No SOR Activity)	DA (Unlimited SOR Activity)	Magnification in DA
1.0	0.00451	1.5	332.59
2.0	0.00902	3.0	332.59
3.0	0.01354	4.5	332.40
4.0	0.01805	6.0	332.40

Fig. 5.3 Plot showing effects of SOR activity on desulfurizing activity (DA) and specific growth rate for ethanol uptake rate of 1 mmol/gdcw-h. The table shows maximum DA achievable for different uptake rates (1-4 mmol/gdcw-h) with zero or unlimited SOR activity

Fig. 5.3 also shows that DA levels off at about 332 times its base value at the zero SOR activity, once SOR activity crosses a threshold. Thus, the maximum possible magnification in DA attainable by increasing the copy numbers of the *dsz* and SOR genes seems to be 332 at an ethanol uptake of 1 mmol/gdcw-h. While the maximum DA increases with ethanol uptake (1-4 mmol/gdcw-h), its magnification from the base level remains nearly unchanged. Since Omori et al. [23]'s experiments point to a non-zero activity of SOR in the wild type *R. erythropolis*, it seems that the highest possible magnification in DA in the wild type *R. erythropolis* must be lower than the 332 with unlimited copy numbers of *dsz* and SOR genes. In I-19 *R. erythropolis* with an elevated *dsz* copy number, Folsom et al. [15] observed only a 25-fold increase in DA. This is much lower than the 332 predicted by our model, and suggests that the SOR activity in Folsom's experiments was very likely limited.

Fig. 5.3 also shows the impact of SOR activity on cell growth. For DA $\leq$ 0.46 mmol HBP/gdcw-h (magnification of about 120 from the base DA), cell growth remains unaffected at its maximum of 0.048 h^{-1}. However, as DA increases beyond this threshold, cell growth decreases linearly with SOR activity, until it becomes zero at the maximum possible DA. This interaction between cell growth and DA can be explained as follows. Initially, ethanol supply is sufficient for the cofactor needs of both cell growth and DA. When DA increases to 120 times its base level, ethanol supply becomes limiting. Cell growth must now compete with DA for its cofactor needs. Thus, it continues to reduce with subsequent increase in DA, until it becomes zero at the maximum DA level. This observation is interesting for obtaining a strain that has lower growth rate, but higher DA. We will discuss a strategy to design such a strain later in this article.

Next we used our *in silico* model to simulate Folsom et al. [15] experiments with glucose as a carbon source. We observed that the magnification in DA due to increased SOR activity was up to a factor of 221 with glucose (1-4 mmol/gdcw-h) (data not shown) which is much lower than that observed with ethanol (332). This also points to the critical role that the carbon source and its metabolism may play in limiting DA by way of NADH regeneration [204].

5.3.4 **Role of SR**

As our main aim is to remove sulfur from DBT by the action of *R. erythropolis* and not cell growth per se, it is desirable to have recombinant strains that can show higher levels of DA at reduced growth rates. As discussed earlier, while SOR converts sulfite into sulfate, SR assimilates it into biomass. As such, the flux through SR directly affects the growth (Fig. 5.4) by limiting the supply of sulfur for biomass synthesis. Our earlier simulations showed that the improvement in DA beyond 120 fold is accompanied by a reduction in growth for a fixed ethanol uptake. Therefore, limiting the SR activity could divert sulfite from biomass to sulfate. To examine the effect of lowering SR activity on DA with unrestricted SOR activity, we fixed ethanol uptake at 1 mmol/gdcw-h as earlier, and maximized growth. At the maximum cell growth, SR activity is the highest at $4.51*10^{-3}$ mmol/gdcw-h. Then, keeping ethanol uptake at 1 mmol/gdcw-h, we decreased the flux through SR ($0\text{-}4.51*10^{-3}$ mmol/gdcw-h) and maximized DA. Fig. 5.4 shows that DA increased, as flux through SR decreased. This suggests that an efficient desulfurizing strain of *R. erythropolis* can be obtained by using higher copy numbers of *dsz* and SOR genes with a reduced activity of SR.

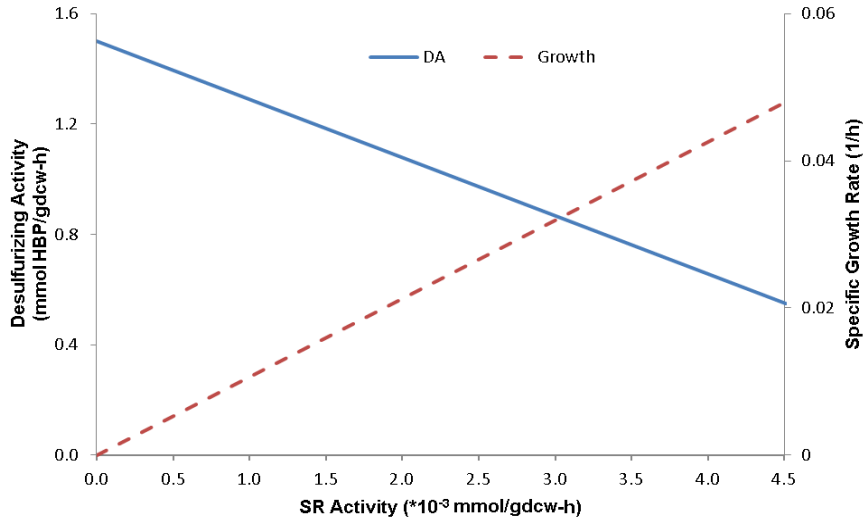


Fig. 5.4 Effect of SR activity on desulfurizing activity and specific growth rate for an ethanol uptake rate of 1 mmol/gdcw-h with unlimited SOR activity.

5.3.5 **SOR activity**

Our hypothesized central role of SOR in limiting desulfurization led us to explore the factors that may limit or impact its activity. Kappler et al. [206] reported that sulfate non-competitively inhibits SOR. Omori et al. [16] observed enhanced desulfurization and improved growth by removing sulfate from the medium. It is also reported [206, 207] that pH has a strong effect on SOR activity. The activity is the highest at pH = 7.5, zero at pH = 3, and decreases as pH decreases [206]. Thus, we decided to focus on the two most critical factors (sulfate and pH) affecting SOR activity.

5.3.5.1 **Effect of sulfate**

Some studies [2] have reported sulfate to retard DA. Li et al. [50] observed that DA decreased with increasing sulfate presence. They proposed the binding of some repressor protein to the *Dsz* promoter (which lies upstream of *dszA*) as a possible explanation. Later, Franchi et al. [26] reported that a recombinant strain with a

combination of strong promoters that are not regulated by sulfate showed improved DA. However, when they cultured these strains in the presence of sulfate, DA was still 50% lower than that in the absence of sulfate. They conjectured that some additional sulfate-regulated factors may be controlling desulfurization. In addition, they observed that NADH in the cell extracts of their desulfurizing strains depleted more rapidly in the presence of DBT than sulfate. They suggested that sulfate was lowering the activity of DszD (oxidoreductase) that provides NADH for the 4S pathway, as Hirasawa et al. [27] had already shown that sulfate has no effect on the transcription from the *dszD* promoter.

In chapter 3 we proposed an alternate/additional explanation for the effect of sulfate on desulfurization. We argued that any repression effects by sulfate notwithstanding, the cell naturally prefers sulfate over DBT. By using sulfate instead of DBT, the cell saves NADH for other growth-related metabolic activities. Indeed, the observations of Franchi et al. [26] strongly support this explanation. We also used the same argument in chapter 4 to explain the higher cell growth with sulfate than DBT [24].

In addition to our previous arguments, we now propose the inhibitory effect of sulfate on SOR as an additional contributing factor for lower cell growth and reduced DA. The removal of sulfate from the medium by Omori et al. [16] prevented the inhibition of SOR activity by sulfate, which resulted in enhanced desulfurization and improved growth. Clearly, the removal of non-toxic sulfate cannot directly increase cell growth, but as discussed earlier, sulfate can non-competitively inhibit SOR activity [206], which may lead to the accumulation of toxic sulfite and hence lower growth. The cell may also naturally limit DBT uptake to its bare minimum to slow the accumulation of toxic sulfite. That means DA decreases.

5.3.5.2 **Effect of pH**

Several studies [11, 20, 98, 203] have reported that decreased medium pH decreases DBT desulfurization by *R. erythropolis*. Furthermore, Wang et al. [98] and Olmo et al. [23] observed that medium pH strongly affected desulfurization, as well as cell growth. A medium pH of 6.5 seemed the best for growth, while 6.5-6.0 was optimum for desulfurization. When pH was not controlled at its initial value of 6.5, it decreased gradually with substantial reductions in both growth and desulfurization. We can explain the above using the activity of SOR. Since SOR activity decreases with decreasing pH, the conversion of sulfite to sulfate slows and sulfite begins to accumulate in the cell. The toxicity of accumulating sulfite reduces cell growth, which in turn reduces its need for sulfur. Furthermore, the cell's reduced ability to remove sulfite forces it to import only the sulfur needed for growth. All these lead to reduced DA. Clearly there may be other effects of pH on cell metabolism besides the activity of SOR, but this is an example of hypotheses for improving DA that can be generated by metabolic flux analysis, and subsequently tested by experimental investigation.

5.4 **Summary**

This chapter highlights the central and critical roles of sulfate, pH, and SOR in desulfurization. We hypothesize that apart from the copy number of *dsz* genes and supply of NADH, the inactivity of SOR limits the extent of DBT desulfurization by *R. erythropolis*. Thus, efforts at increasing the desulfurization rate in biocatalysts may benefit by focusing on SOR. While increasing the expression and/or copy number of genes encoding SOR is obviously essential, several other factors must be kept in mind. Sulfate must be continuously removed from the medium, medium pH must be

maintained at 6.5 and adding molybdenum salts may help, as SOR utilizes molybdenum as a cofactor.

Another possibility for improved desulfurization activity suggested by our *in silico* analysis is the simultaneous decrease in SR activity and increase in SOR activity. SR is essential for the utilization of sulfite or sulfate to support cell growth. A lower SR activity would effectively starve cells for sulfur and result in an increased rate of utilization of DBT provided that elevated levels of SOR are maintained to avoid the toxic accumulation of sulfite. This approach should result in desulfurization biocatalysts that have a slower growth rate, but higher desulfurization activity.

6 *In silico* evaluation of *Gordonia alkanivorans* as an attractive microbe for biodesulfurization

6.1 Introduction

Biodesulfurization is considered an attractive alternate method for desulfurization of fossil fuels. Most desulfurization studies in the literature have used DBT as the model compound. While several rhodococci strains exhibit non-destructive desulfurization of DBT, *R. erythropolis* IGTS8 was the first to be identified [10] and has received the most attention. However, most rhodococci are unable to show high activity for the alkyl derivatives of DBT and show no activity for BT and other thiophenic compounds. Since fossil fuels do contain these compounds in significant amounts, it is critical to study microbes that possess activity for compounds other than DBT. Furthermore, because desulfurization of BT and DBT requires distinct pathways, bacterial strains that possess the associated genes for all these pathways are clearly desirable. *Gordonia* is an attractive genus in this regard, because its members exhibit much metabolic versatility [9].

Numerous gordoniae strains exhibit higher desulfurization activities [9, 35] than the rhodococci and from a broader range of PASHs [31-35, 125]. Of them, *G. alkanivorans* [31, 35, 36, 138] is widely studied for its ability to desulfurize various thiophenic compounds, BT, DBT, and their derivatives. It desulfurizes DBT via the well-known 4S pathway similarly to *R. erythropolis* [7]. However, unlike *R. erythropolis*, it can also specifically cleave the C-S bond in BT and other thiophenes besides DBT. Because of its ability to desulfurize a wider range of PASHs, *G.*

alkanivorans appears to offer some advantage over *R. erythropolis* for biodesulfurization. Moreover, *G. alkanivorans* strains are reported [124] to show nearly 2-10 times higher desulfurization activities than the desulfurizing *R. erythropolis* strains. In other words, it has the greater ability to reduce the overall sulfur content of the fossil fuels.

In spite of its promise, desulfurization studies with *G. alkanivorans* are far more limited than those with *R. erythropolis.* Although it does offer higher desulfurizing activity than *R. erythropolis*, the activity levels are still not acceptable for commercial application. Thus, there is a need to identify and study the factors and host functions that may play key roles in controlling the extent of desulfurization by *G. alkanivorans*. However, the complexity of metabolic networks makes it difficult to predict or identify such host functions intuitively or using a trial-and-error experimental approach. Since cellular activities are invariably and intricately coupled, a holistic study of the various metabolic functions occurring within *G. alkanivorans* besides the desulfurization of PASHs is essential to understand the interactions between the various components of its metabolic network. Such a study would also allow one to compare gordoniae with rhodococci in a theoretical and comprehensive manner. However, no such holistic study for gordoniae exists in the literature.

This chapter discusses the first *in silico* genome scale metabolic model for *G. alkanivorans*. It covers the key metabolic pathways such as central metabolism, amino acids biosyntheses, nucleotide metabolism, and sulfur metabolism that describes the assimilation of sulfur into biomass. It can help in understanding the metabolic architecture of *G. alkanivorans*, and its host functions related to desulfurization. We validate the model using the available desulfurization and growth data in the literature [125], and use it to study the effects of various medium components such as carbon

sources, amino acids, and vitamins on the desulfurization activity of *G. alkanivorans*. We assess the properties of its metabolic network such as flexibility and robustness using flux variability [210] and gene essentiality analyses. Finally, we use flux sum analyses [211] to study qualitatively and quantitatively the effect of intracellular metabolites on growth and desulfurization activity and propose several experimentally testable conditions and modifications that may help enhance the desulfurizing activity of *G. alkanivorans.*

6.2 Materials and methods

6.2.1 Model reconstruction

The reconstruction of an *in silico* genome scale metabolic model for an organism requires the identification and classification of its metabolite reactions and the establishment of their appropriate gene-protein-reaction (GPR) associations. It is an iterative process involving the collection and processing of diverse information about cellular metabolism, biochemistry, and various strain-specific parameters of the organism [43]. We reconstructed the genome scale metabolic model of *G. alkanivorans* in three steps, i) constructing an initial draft model based on genome annotations, ii) model improvement to enable cell growth, and iii) model analysis for identifying and filling network gaps based on biochemical information.

For reconstructing an initial draft model of *G. alkanivorans*, we annotated the genome sequence of *G. alkanivorans* using the tools available on the online annotation server RAST [212]. We manually processed this information to establish the GPR associations and assign appropriate gene(s) to the various enzymes and their corresponding reactions in the metabolic network. We also checked all reactions for

elemental balancing. Then, we cross checked the GPR associations and the reaction directionality with the information available for *G. alkanivorans* in KEGG [170] and MetaCyc [172]. We incorporated any additional reactions or pathways that were available in MetaCyc and KEGG. We removed all the reactions that accounted for the polymerization of monomers and conversion of general class compounds such as ROH, RCOOH, etc.

After this, we identified several broken pathways, dead end metabolites (DEMs), and missing reactions in the model, which arise mainly due to the lack of metabolite connectivity and presence of gaps in the network [49]. To complete and enhance our model, we employed several means. First, we looked for additional reactions based on the literature evidence and other biochemical information. Second, we used optimization-based automated procedures of GapFill and GapFind, proposed by Kumar et al. [49], to identify and restore the connectivity of the DEMs and to identify and fill the remaining network gaps. We used GAMS/CPLEX 10.0 [194] to execute these procedures and systematically determine and eliminate these network gaps by restoring the connectivity within the metabolic network. All these required adding new reactions into the model, for which no genetic evidence is currently available. Therefore, we tried to identify and assign possible ORFs that may potentially encode for these missing functions. For this, we performed BLASTp searches between the translated set of genes associated with these additional reactions in various databases and the genome of *G. alkanivorans*. While we used a high e cut-off of 10^{-30} for most network improvement reactions, we used a low cut-off of 10^{-5} for some reactions to enable the essential activity of biomass generation.

6.2.2 **Experimental studies**

We used the experimental data of Rhee et al. [125] for validating our *in silico* model. Rhee et al. [125] isolated and studied DBT desulfurization characteristics of *Gordonia* sp. CYSK1. They used a minimal salt medium (MSM) consisting of 5.0 g glucose, 5.0 g K_2HPO_4, 1.0 g NaH_2PO_4, 1.0 g NH_4Cl, 0.2 g $MgCl_2$, 0.01 g $CaCl_2.2H_2O$, 1 ml of sulfur-free trace element solution dissolved in EDTA, and 1 ml of vitamin solution. The MSM was supplemented with DBT dissolved in ethanol at a concentration of 100 mM. They cultured CYSK1 in 250 ml Erlenmeyer flasks containing 50 ml of MSM at 30 ºC on a gyratory shaker. During the culture, they measured concentration profiles of glucose via reverse-phase HPLC, DBT and HBP via GC, and cell growth via absorbance at 600 nm. They observed that the cells gave HBP as the final desulfurized product and did not re-assimilate it as a carbon source.

We used the experimental data of Iida et al. [213] to study the utilization of various carbon sources. Iida et al. [24] studied substrate utilization patterns of several *Gordonia* strains.

6.2.3 **Model analysis**

For analyzing and predicting phenotypes, the constraint-based genome scale metabolic models typically solve the following linear optimization problem (Flux Balance Analysis or FBA), which is based on the assumption that intracellular metabolites as discussed earlier. To solve the FBA model, we need a cellular objective (Z). Several cellular objectives such as maximum cell growth, minimum substrate utilization, minimum maintenance energy, etc. [47] have been used in the past literature. Cell growth is one of the most commonly used objectives, as microbial cells have evolved to maximize growths. It can be expressed as a synthetic reaction consuming multiple

biomass precursor metabolites in some ratios, which can be determined from cell composition. However, no information is available in the literature on the cellular composition of *G. alkanivorans*. Therefore, we adapted information from the metabolic models of a related organism, *Corynebacterium glutamicum* [214, 215]. Such adaptation from related organisms is an established practice in the reconstruction of metabolic models [159].

We used MetaFluxNet [174] and GAMS/CPLEX 10.0 [194] to solve and analyze our FBA model.

6.2.4 **Flux variability analysis**

The cellular metabolism is quite robust due to the redundancy of the pathways. This is reflected in the underdetermined nature of genome scale models which leads to the multiple solutions. Under a given set of conditions and constraints the flux distributions may not be unique. As such there may be alternate optimal solutions for flux distributions that lead to the same value of objective function. The flux variability analysis [210] is a method used to obtain the ranges over which the flux values can vary under a given condition for the same value of the objective function. The analysis of the variability in the flux values in a metabolic model helps in understanding the flexibility exhibited by the metabolic network. It can guide in deciding which fluxes can be manipulated without having an undesirable effect on the desired objective.

6.2.5 **Flux sum analysis**

The metabolites are key players in the metabolism. They determine the connectivity and the flow of nutrients in the metabolic networks. However, Flux Balance Analysis (FBA) [142], the commonly used method for metabolic model analysis is based on the

reaction centric approach. It mainly emphasizes on the role of reaction fluxes and characterizes the metabolic system based on the constraints imposed on the reaction fluxes. The role of metabolites can be elucidated using the flux sum analysis method [211] based on mixed integer programming. Like FBA, it is based on the assumption of steady state for intracellular metabolites. Nevertheless, the "flux sum" i.e., the net turnover rate of a metabolite has to be non-zero for it to participate in the metabolism. The quantitative analysis of the flux sum of a metabolite can help in understanding and elucidating its role in the cellular metabolism and determining the effects of perturbation in its intracellular generation.

6.3 Results and discussions

6.3.1 Reconstructed *in silico* genome scale metabolic model of *G. alkanivorans*

Our initial draft model comprised 739 reactions. However, it failed to show any cell growth on substrates known to be utilized by *G. alkanivorans*. We found that nearly 60% of all precursor metabolites could not be synthesized by the draft model. For instance, the draft model was unable to produce amino acids such as methionine and histidine, although *G. alkanivornas* is known [35] to synthesize them and survive without any external supply. Our draft model also showed zero growth with DBT and BT as sole sulfur sources, although *G. alkanivorans* is known [35] to metabolize them. Therefore, we added relevant pathways and reactions based on the information available in the literature and databases [7, 35, 123, 216].

GapFind revealed 279 dead-end metabolites (DEMs) that could not be produced in the initial draft of our model, as they were disconnected from the rest of

the network either upstream or downstream. GapFill could identify the possible candidate reactions to restore the connectivity for only 140 of the 279 DEMs. We performed BLASTp analyses for assigning putative ORFs to the enzymes associated with these reactions. To ensure that we do not add reactions indiscriminately to our model, we used a high e cut-off of 10^{-30} to include only the reactions with a strong evidence of ORFs. We could locate ORFs for only 62 (~22%) DEMs, thus we did not include other reactions.

Our final curated model consists of 881 unique metabolites and 922 reactions associated with 568 ORFs/genes and 544 unique enzymes. Of these 922 reactions, 67 account for the transport of various metabolites across the membrane, while the rest (855) account for intracellular metabolic activities. Table 6.1 lists the features of our GSM model. The BLASTp analyses identified possible annotations for 55 ORFs in *G. alkanovirans*, which are given in Table 6.2. However, the model still has 217 DEMs, which warrants further biochemistry studies. The reactions and metabolites details for the model are given in Appendix C

Table 6.1 Features of the reconstructed genome scale model of *G. alkanivorans*

Features	Properties
Reactions in genome scale mode	923
No of ORFs included	568
No of enzymes included	544
Intracellular reactions	856
Transport reactions	67
Metabolites in genome scale model	881
Internal metabolites	814

6.3.2 **Model validation**

We mimicked the experiments of Rhee et al. [125] and compared the growth rates predicted by our model with their experimental values. From their experiments, we inferred glucose to be limiting, as its complete depletion from the medium triggered the stationary phase. So we computed and used specific uptake rates for glucose at several time points to constrain our model. We assumed unlimited supply for other medium components, as they were in excess. We solved our model to maximize biomass growth at each time point. As seen in Fig. 6.1, our predicted biomass growth rates are in close agreement with the experimental data of Rhee et al. [10].

6.3.3 **Gene essentiality analysis**

Most cells can withstand disturbances at the genetic as well as metabolic levels by utilizing alternate genes, enzymes, and pathways depending on the conditions. However, the non-functionality of certain reactions and genes may be lethal for a cell. Therefore, it is desirable to determine which genes and reactions are essential for a cell's survival.

We studied the robustness of *G. alkanivorans* metabolism by assessing its ability to exhibit *in silico* growth in case of gene knockouts or mutations. The utilization of a pathway, and thus the essentiality of its reactions and genes, will in general depend on medium components. We used five media with different carbon sources (ethanol, fumarate, oxoglutarate, pyruvate, and glutamate) and evaluated the essentialities of genes and reactions for each medium. For reactions, we removed one reaction at a time by setting its flux as zero and maximized cell growth. If the model could not produce cell mass, then we classified the reaction as essential else non-essential. Similarly, for genes, if the removal of a gene prevented cell growth, then we

classified that gene as essential otherwise non-essential. To remove a gene, we set the fluxes of all its associated reactions to zero in our *in silico* model. However, if a reaction was controlled by two or more isozymes, then the reaction was kept active in the absence of any one of the associated genes.

Table 6.2 New Possible Genome Annotations for *G. alkanivorans*

E.C. No.	Enzyme Name	Current Annotation	NCBI Accession N	E value
EC 1.1.1.103	L-threonine 3-dehydrogenase	alcohol dehydrogenase	ZP_08767112.1	3.00E-29
EC 1.1.1.29	glycerate dehydrogenase	D-3-phosphoglycerate dehydrogenase	ZP_08766341.1	1.00E-20
EC 1.1.1.81	hydroxypyruvate reductase	putative oxidoreductase	ZP_08767993.1	3.00E-22
EC 1.2.7.6	glyceraldehyde-3-phosphate dehydrogenase (ferredoxin)	putative dehydrogenase	ZP_08767188.1	3.00E-04
EC 3.1.4.17	3',5'-cyclic-nucleotide phosphodiesterase	putative LuxR family transcriptional regulator	ZP_08766296.1	3.00E-12
EC 3.2.1.122	maltose-6'-phosphate glucosidase	molybdenum cofactor biosynthesis protein A	ZP_08767656.1	8.00E-05
EC 3.5.4.21	creatinine deaminase	putative hydrolase	ZP_08765243.1	7.00E-04
EC 1.1.1.17	mannitol-1-phosphate 5-dehydrogenase	putative phosphoribosylglycinamide formyltransferase 2	ZP_08768014.1	0.0001
EC 1.1.1.26	glyoxylate reductase	D-3-phosphoglycerate dehydrogenase	ZP_08766341.1	1.00E-42
EC 1.1.1.36	acetoacetyl-CoA reductase	3-oxoacyl-[acyl-carrier-protein] reductase	ZP_08768232.1	6.00E-44
EC 1.1.1.60	2-hydroxy-3-oxopropionate reductase	3-hydroxyisobutyrate dehydrogenase	ZP_08765584.1	3E-44
EC 1.1.1.65	pyridoxine 4-dehydrogenase	putative aldo/keto reductase	ZP_08764692.1	7E-11
EC 1.1.1.79	glyoxylate reductase (NADP+)	D-3-phosphoglycerate dehydrogenase	ZP_08766341.1	1.00E-42
EC 1.1.1.83	D-malate dehydrogenase (decarboxylating)	3-isopropylmalate dehydrogenase	ZP_08766340.1	8.00E-81
EC 1.1.5.8	quinate dehydrogenase (quinone)	putative non-ribosomal peptide synthetase	ZP_08767228.1	8.00E-04
EC 1.17.3.2	xanthine oxidase	putative xanthine dehydrogenase	ZP_08766184.1	2.00E-52
EC 1.18.6.1	nitrogenase	chromosome partitioning protein ParA	ZP_08768173.1	7.00E-11
EC 1.2.1.22	lactaldehyde dehydrogenase	succinate-semialdehyde dehydrogenase	ZP_08765499.1	4.00E-91
EC 1.2.7.5	aldehyde ferredoxin oxidoreductase	hypothetical protein GOALK_120_00670	ZP_08768084.1	3.00E-06
EC 1.2.7.7	3-methyl-2-oxobutanoate dehydrogenase (ferredoxin)	putative oxidoreductase	ZP_08766668.1	3.00E-05
EC 1.21.4.3	sarcosine reductase	putative acetyl-CoA acetyltransferase	ZP_08764801.1	7.00E-05
EC 1.3.1.78	arogenate dehydrogenase (NADP+)	prephenate dehydrogenase &	ZP_08765663.1	2.00E-26
EC 1.3.99.10	isovaleryl-CoA dehydrogenase	acyl-CoA dehydrogenase	ZP_08765489.1	4.00E-79
EC 1.4.3.21	primary-amine oxidase	adenosylcobinamide kinase	ZP_08763719.1	0.0007
EC 1.4.99.5	glycine dehydrogenase (cyanide-forming);	putative ferredoxin reductase	ZP_08764700.1	2.00E-09
EC 1.7.2.2	nitrite reductase (cytochrome; ammonia-forming)	ethanolamine ammonia-lyase large subunit	ZP_08767599.1	0.0002
EC 1.7.7.2	ferredoxin---nitrate reductase	putative nitrate/sulfite reductase	ZP_08764724.1	1.00E-175
EC 2.1.1.2	guanidinoacetate N-methyltransferase	hypothetical protein GOALK_050_00300	ZP_08765250.1	0.016
EC 2.1.3.1	methylmalonyl-CoA carboxytransferase	pyruvate carboxylase	ZP_08766189.1	5.00E-13
EC 2.1.4.1	glycine amidinotransferase	isocitrate lyase	ZP_08765259.1	9.00E-04
EC 2.3.1.182	(R)-citramalate synthase	putative 4-hydroxy-2-oxovalerate aldolase	ZP_08765376.1	1.00E-13
EC 2.3.3.10	hydroxymethylglutaryl-CoA synthase	3-oxoacyl-[acyl-carrier-protein] synthase III	ZP_08764811.1	6.00E-07
EC 2.4.2.28	S-methyl-5'-thioadenosine phosphorylase	purine nucleoside phosphorylase	ZP_08766163.1	6.00E-17
EC 2.5.1.82	hexaprenyl diphosphate synthase [geranylgeranyl-diphosphate specific]	putative polyprenyl diphosphate synthase	ZP_08765134.1	2E-33
EC 2.5.1.83	hexaprenyl-diphosphate synthase [(2E,6E)-farnesyl-diphosphate specific]	putative polyprenyl diphosphate synthase	ZP_08765134.1	2.00E-33
EC 2.5.1.84	all-trans-nonaprenyl-diphosphate synthase [geranyl-diphosphate specific]	putative polyprenyl diphosphate synthase	ZP_08765134.1	3.00E-42
EC 2.7.1.100	S-methyl-5-thioribose kinase	hypothetical protein	ZP_08767309.1	3.00E-06
EC 2.7.1.48	uridine kinase	uracil phosphoribosyltransferase	ZP_08766161.1	2.00E-18
EC 3.1.1.17	gluconolactonase	hypothetical protein	ZP_08765725.1	6.00E-12
EC 3.1.2.4	3-hydroxyisobutyryl-CoA hydrolase	hypothetical protein	ZP_08764807.1	3.00E-66
EC 3.2.1.93	alpha,alpha-phosphotrehalase	alpha-glucosidase	ZP_08767019.1	3.00E-82
EC 3.2.2.1	purine nucleosidase	putative ribonucleoside hydrolase	ZP_08767439.1	4.00E-25
EC 3.5.1.59	N-carbamoylsarcosine amidase	putative hydrolase	ZP_08765823.1	2.00E-40
EC 3.5.2.10	creatininase	putative creatininase family protein	ZP_08767265.1	7E-18
EC 3.5.2.15	cyanuric acid amidohydrolase	hypothetical protein	ZP_08768158.1	0.006
EC 3.5.3.9	allantoate deiminase	putative M20D family peptidase	ZP_08766098.1	1.00E-08
EC 3.5.4.1	cytosine deaminase	putative cytosine deaminase	ZP_08764308.1	2.00E-67
EC 3.5.4.12	dCMP deaminase	tRNA-specific adenosine deaminase	ZP_08765661.1	2.00E-19
EC 3.5.5.1	nitrilase	putative carbon-nitrogen hydrolase	ZP_08767356.1	1.00E-11
EC 4.1.2.20	2-dehydro-3-deoxyglucarate aldolase	putative citrate lyase beta subunit	ZP_08765089.1	1.00E-05
EC 4.2.1.66	cyanide hydratase	putative amidohydrolase	ZP_08767351.1	4.00E-10
EC 4.2.1.84	nitrile hydratase	thiocyanate hydrolase gamma subunit	ZP_08768164.1	2E-48
EC 5.1.3.6	UDP-glucuronate 4-epimerase	UDP-glucose 4-epimerase	ZP_08763819.1	2E-17
EC 6.2.1.25	benzoate---CoA ligase	putative fatty-acid--CoA ligase	ZP_08763669.1	1.00E-60
EC 6.2.1.4	succinate---CoA ligase (GDP-forming)	succinyl-CoA synthetase beta subunit	ZP_08766733.1	2.00E-70

We identified 116 reactions and 75 genes to be essential irrespective of the medium. As seen in Fig. 6.2, most essential reactions belong to the amino acids metabolism followed by nucleotides metabolism, central metabolism, and cell wall metabolism. Therefore, these parts of the metabolic network are most susceptible to perturbations. Any environmental factor that alters any of these parts in *G.*

alkanivorans may prove lethal. The difference in the numbers of essential reactions and essential genes is due to isozymes, as several reactions are catalyzed by enzymes that multiple genes encode. These reactions are essential at the metabolic level but not the genetic level.

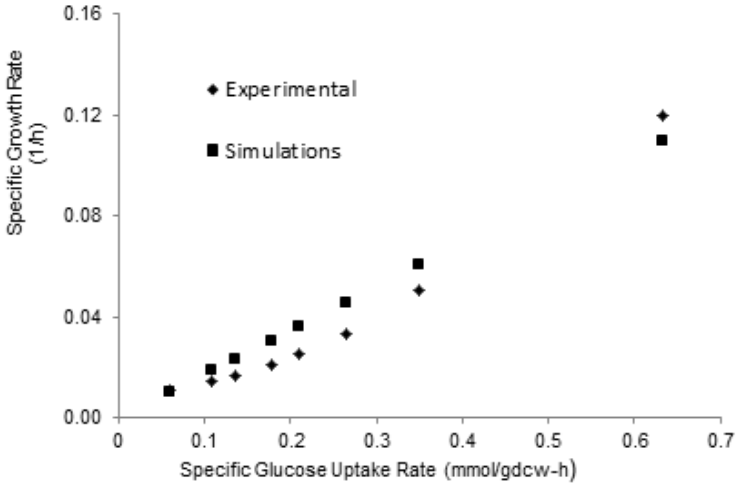


Fig. 6.1 Experimental and simulated growth rates at various glucose uptake rates from Rhee et al. [8]

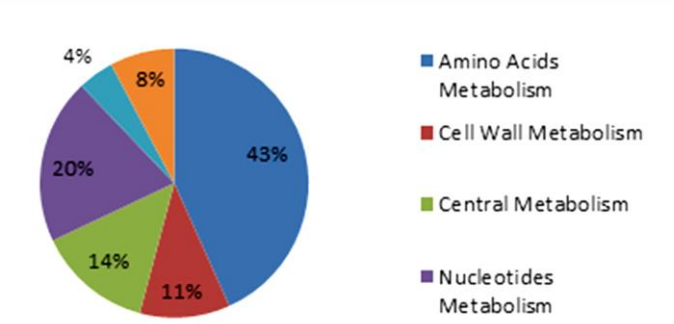


Fig. 6.2 Distribution of essential reactions over various cellular subsystems in *G. alkanivorans*

As discussed earlier, *G. alkanivorans* also has the *dszABC* genes for the 4S pathway, which are essential for survival on DBT.

6.3.4 **Utilization of alternate carbon sources**

Gordonia strains are known for their metabolic versatility. They are able to metabolize and utilize several compounds to satisfy their nutritional requirements. Iida et al. [213] experimentally examined the ability of various species of *Gordonia* to utilize 31 different carbon sources. We simulated their experimental conditions to examine the ability of our *in silico* organism to exhibit growth on these 31 compounds when provided as sole carbon sources. For this, we performed 31 different simulations by supplying fixed amount (1 mmol/gdcw-h) of each of the carbon source individually and maximizing growth. Table 6.1 shows the comparison of the experimental and *in silico* results for the growth of *G. alkanivorans* on these carbon sources. The *in silico* growth was correctly observed for 16 carbon sources while there were disagreements for the rest of the 15 carbon sources.

The disagreements involve both the false positive and the false negative results. A false positive result is one in which the *in silico* growth is observed on a carbon source while the experiments show the inability of the organism to grow on it. Such results may arise due to the errors in reconstruction or due to the lack of regulatory mechanisms in the model which may otherwise be functional in the organism and therefore, do not allow the organism to grow on the respective compound in the wet lab. However, the model predictions in this case may be improved by incorporating regulatory information. A false negative result is one in which the model is unable to predict any *in silico* growth on a carbon source while experimentally *G. alkanivornas* shows growth on it. These errors arise due to the lack of complete biochemical information of *G. alkanivorans*. Such limitations of the model can be eliminated by acquiring further information about *G. alkanivorans* using various biochemical characterization techniques.

6.3.5 **Effects of medium composition on desulfurizing activity**

The studies have shown that the medium composition has a prominent effect on the desulfurizing activity exhibited by *R. erythropolis* [20]. Several reports are available in literature which have studied the effects of various medium components on the desulfurizing activity of *R. erythropolis* [18-20, 24]. However, such studies have not been largely performed for *G. alkanivorans*. Therefore, we performed *in silico* analyses to study the effects of the medium components such as alternate carbon sources and addition of various vitamins and amino acids on the desulfurizing activities of *G. alkanivorans*.

6.3.5.1 **Effect of carbon sources**

Biodesulfurization is considered as a potential alternative as it is expected to significantly reduce the fuel processing costs incurred during the hydrodesulfurization process [2]. Franchi et al. [26] reported that the type and cost of the carbon source used will have a significant effect on the overall cost of biodesulfurization. Therefore, it is critical to study the effects of various possible carbon sources on the desulfurizing activity of a strain and identify the most suitable carbon source. We examined 16 different carbon sources (acetate, citrate, ethanol, formate, fructose, fumarate, gluconate, glucose, glutamate, glycerol, lactate, malate, oxaloacetate, oxoglutarate, pyruvate, and succinate) for their effect on the desulfurizing activity and growth of *G. alkanivorans.* For this we performed 16 different simulations in each of which we sequentially provided equal amounts (20 mg/gdcw-h) of these carbon sources one at a time and calculated maximum growth rate and corresponding desulfurization activity. We found that ethanol is the best of all the carbon sources considered for achieving higher growth and desulfurizing activity. As seen in Fig. 6.3, the suitability of the

various carbon sources for higher growth and desulfurizing activity is in the order ethanol (0.027 h^{-1} as 100% and 3.3 µmol HBP/gdcw-h as 100%, respectively) > fumarate (80%) > oxoglutarate (78.79%) > pyruvate (78.43%) > glutamate (78.24%) > succinate (78%) > acetate ≈ fructose ≈ glucose ≈ lactate (76.86%) > glycerol (75%) > citrate (71.88%) > oxaloacetate (69.70%) > malate (69.11%) > formate (50%). The suitability of the carbon sources can be explained on the basis of NADH production as discussed by in our previous chapters 3 and 4. The DBT utilization as sulfur source via 4S pathway is energy intensive as metabolism of 1 mol of DBT requires 4 moles of NADH. However, NADH is essential for growth and is also additionally required to metabolize DBT as a sulfur source. A carbon nutrient is the main source of energy and also affects the cofactor regeneration in cellular metabolism of an organism. Therefore, any carbon source that provides more NADH during its metabolism is likely to support higher desulfurization and growth rates of *G. alkanivorans*. Metabolism of each mole of ethanol is accompanied by the generation of 2 additional moles of NADH and thus, it supports higher growth and desulfurizing activity.

6.3.5.2 **Effects of various vitamins and amino acids**

There are several studies available which have made attempts to improve desulfurizing activity of *R. erythropolis* by adding different components such as vitamins to the medium. Yan et al. [19] found that addition of nicotinamide and riboflavin improved the desulfurization activity of *R. erythropolis* significantly. Since there are no such studies available for *G. alkanivorans*, we tried to explore the effects of the addition of vitamins (nicotinamide and riboflavin) and the twenty amino acids on the desulfurizing activity of *G. alkanivorans*.

Table 6.3 Utilization of various carbon sources examined by Iida et al. [29] and as predicted by model. A '+' means the compound can be utilized as a sole carbon source while a '-' means that it cannot be

Carbon Source	Experimental Utilization	*In Silico* Utilization
D-galactose	+	-
L-Rhamnose	-	-
D-Ribose	+	+
Sucrose	+	+
Turanose	+	-
Arabitol	+	-
Inositol	+	+
Glucarate	+	-
Gluconate	+	+
D-Glucosaminic acid	+	+
Caprate	+	-
Citrate	+	+
4-Aminobutyrate	-	+
2-Hydroxyvalerate	+	-
2-Oxoglutarate	+	+
Pimelate	+	+
Succinate	+	+
Benzoate	+	-
3-Hydroxybenzoate	+	-
4-Hydroxybenzoate	+	-
Phenylacetate	+	-
Quinate	+	-
L-Alanine	+	+
L-Aspartate	+	+
L-Leucine	+	-
L-Proline	+	+
L-Serine	-	+
L-Valine	-	-
Putrescine	+	+
Tyramine	+	-
Acetamide	-	-

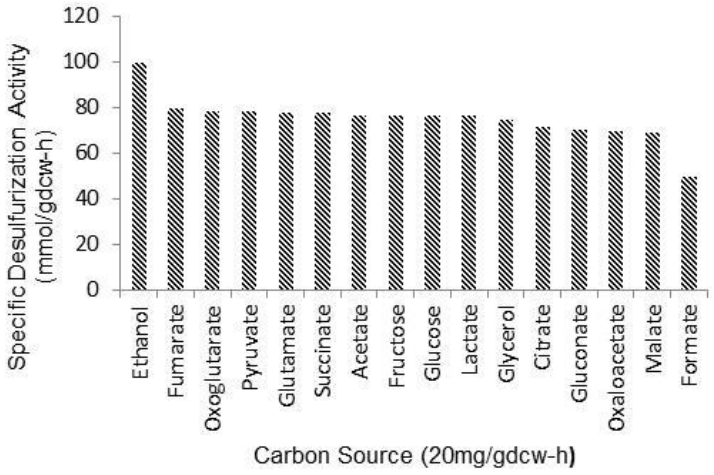


Fig. 6.3 Specific desulfurizing activities for an uptake rate of 20mg/gdcw-h of various carbon sources

For this we supplied a fixed amount of glucose (1 mmol/gdcw-h) as the carbon source with excess of other components of the minimal medium and obtained the maximum growth (0.17 h^{-1}) and corresponding desulfurization rate (0.021 mmol/gdcw-h). We next performed a number of simulations to evaluate the growth rates and desulfurizing activities of *G. alkanivorans* when the medium was supplemented with same amounts (1 mmol/gdcw-h) of various vitamins and amino acids individually. When the *in silico* medium was supplemented with vitamins, there was no effect on the growth rate and desulfurizing activity. The inability of vitamins to enhance desulfurization *in silico* may arise due to the possibility that they act at the regulatory or transcriptional level as hypothesized by Yan et al. [19] but not at the metabolic level. Since our model does not include the regulatory and transcriptional effects, it is unable to capture any such effect of the two vitamins on the desulfurizing activity.

The presence of different amino acids showed varying levels of effects on the growth rate and desulfurizing activity of *G. alkanivorans*. The presence of nine of the amino acids (arginine, histidine, isoleucine, leucine, lysine, phenylalanine, tryptophan,

tyrosine, and valine) did not significantly affect the growth and desulfurization rates. However, in the presence of cysteine the desulfurization activity was not observed. To find an explanation for this we examined our model and found that cysteine can be utilized as a sole sulfur source by *G. alkanivorans.* The utilization of cysteine as a sulfur source is less energy expensive than DBT as the metabolism of 1 mole of DBT requires additional 4 moles of NADH, as explained earlier in previous works [204]. Therefore, when cysteine, a readily usable sulfur source is available the cell consumes it in preference to the DBT and therefore, no DBT desulfurization is observed. It was also seen that in the presence of methionine the desulfurization activity is reduced by 63%. This arises due to the fact that the *in silico G. alkanivorans* is unable to utilize methionine solely as the sulfur source. Therefore, it has to utilize an additional sulfur source (here DBT) when methionine is present. However, since metabolism of methionine is energetically less demanding cell prefers to utilize it to synthesize most of the sulfur containing metabolic precursors that it can. Thereby, the consumption of DBT and hence, desulfurization activity is lowered in the presence of methionine.

The growth rates and desulfurizing activities were greatly improved in the presence of the remaining nine amino acids (alanine, asparagine, aspartate, glutamine, glutamate, glycine, proline, serine, and threonine). Further investigations showed that these amino acids can support *in silico* growth as sole carbon sources while those from the former sets cannot. Therefore, when the amino acids from the latter set are supplied to the medium with limited amounts of glucose, they provide additional carbon flux. As such the additional supply of any utilizable carbon source(s) with the availability of other medium components allows higher growth rates and higher cofactor regeneration. Since sulfur is essential for growth, higher growth rates are accompanied

by higher sulfur requirements and thus, higher desulfurization rates. Fig. 6.4 shows relative effects of the addition of various amino acids on desulfurizing activity.

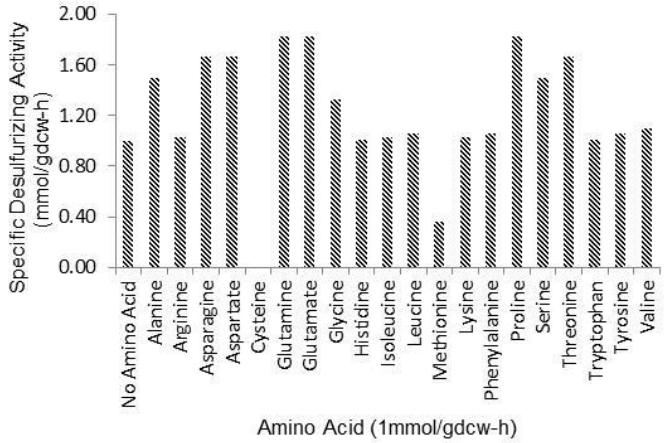


Fig. 6.4 Specific desulfurizing activities for an uptake rate of 1mmol/gdcw-h of various amino acids

6.3.6 **Utilization of PASHs**

The fossil fuels contain several PASHs that are usually recalcitrant to the hydrodesulfurization treatment. These are mostly the thiophenic compounds such as BT, DBT and their derivatives. In most of the studies DBT has been used as the model sulfur compound as it is considered to be a representative of most of the PASHs present in the fossil fuels. However, since BT is not utilized by all the DBT desulfurizing strains [119, 123, 125] and is desulfurized via a different pathway than the 4S pathway, the utilization of one does not ensure utilization of the other. Therefore, it is critical to study desulfurization of both BT and DBT by the various desulfurizing strains. *G. alkanivorans* is known to utilize both BT and DBT as sole sulfur sources [35]. We studied the *in silico* metabolism of both BT and DBT by *G. alkanivorans* using our GSM model. Alves et al. [35] studied the desulfurization of BT and DBT when supplied simultaneously in a glucose based medium to *G. alkanivorans*. They found that the two sources were not utilized simultaneously but sequentially with BT as the preferred sulfur source. BT was completely consumed first

and DBT was consumed only when BT was exhausted from the medium. They proposed that this occurred probably due to the presence of a non-specific uptake system for thiophenic compounds which has a higher affinity for BT than DBT.

We used our GSM model to study the *in silico* utilization of BT and DBT when supplied simultaneously to *G. alkanivorans*. For this we allowed a fixed amount of glucose (20 mg/dcw-h) with unlimited supply of BT and DBT simultaneously and maximized biomass growth. It was observed that only BT was consumed and DBT was not taken up. The corresponding growth rate and BT desulfurizing activity were 0.021 h^{-1} and 2.55 µmol/gdcw-h, respectively. This amount of BT corresponds to the minimum *in silico* sulfur requirements of the cell for maximizing growth when supplied with limited glucose (20 mg/gdcw-h). In the next set of simulations we kept the same fixed supply of glucose with unlimited DBT but reduced the supply of BT in steps from 2.55µmol/gdcw-h to 0µmol/gdcw-h and maximized biomass. We observed that as the supply of BT was lowered from the minimum requirement for the fixed glucose supply, there was an increase in DBT uptake (Fig. 6.5).

This observation shows that the *in silico* cell prefers BT over DBT as a sulfur source and is in accordance with the experimental results of Alves et al. [35]. Since our model does not include any regulatory mechanism and also the *in silico* uptake systems for DBT and BT have been modeled to exhibit equal preferences, its ability to predict preferential usage of BT shows that the regulation/uptake system may not be the only determining factors. The preference of BT over DBT as a sulfur source can be explained on the basis of the energy requirements for their metabolism. The complete consumption of 1 mole of BT as a sulfur source requires 2 moles of NADH while the complete consumption of 1 mole of DBT requires 4 moles of NADH. Since NADH is a high energy molecule required for growth as well as for other essential metabolic

activities, the cell prefers to utilize a sulfur source which consumes lesser NADH. As such the cell prefers BT over DBT for its growth when supplied with both of them simultaneously.

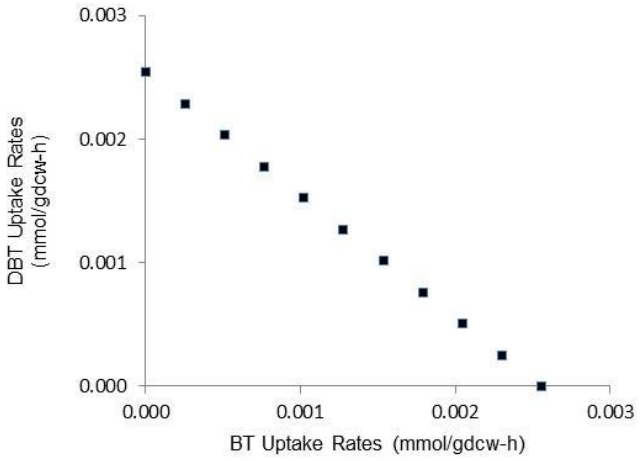


Fig. 6.5 Effect of increasing the BT uptake rate on specific DBT desulfurization

6.3.7 **Flux variability analysis (FVA)**

The metabolic network models are underdetermined in nature. As a result, under any given condition the model can have a number of alternate optimal solutions for flux distributions that give the same value for the desired objective. A particular optimal flux distribution would be a point in an otherwise large allowable flux space. Therefore, the entire allowable flux space under a particular set of constraints can be explored by estimating the possible values that each flux can take under defined constraints. FVA [210] is a technique that allows the computation of the range over which a particular flux can vary for the same objective value under the defined conditions. We used FVA to determine the allowable flexibility in reaction flux values under the given conditions for the same value of the biomass growth rate. For this, we fixed the supply of ethanol at 10mmol/gdcw-h with unrestricted amounts of other components of the minimal media and calculated the maximum growth achievable.

Next, for the same conditions of the medium we sequentially calculated the maximum and minimum values that each reaction flux can take under the additional constraint that the growth rate should be equal to that calculated in previous simulation. This allowed us to study the robustness of *G. alkanivorans*'s metabolism by examining variations in reaction fluxes that can be allowed liberally without affecting the growth deleteriously. As expected, there were several reactions for which the flux values varied over a wide range. However, there are certain reaction fluxes (~15% of the total) which did not show considerable variations. They mainly belong to the amino acids and cell wall metabolic pathways. These reactions represent the inflexible part of the metabolism when a desired amount of biomass production is required. As such any uncontrolled perturbations in these reactions and pathways may lead to undesirable effects on the growth of *G. alkanivorans*.

6.3.8 **Flux sum analysis (FSA)**

The FBA method elucidates the significance of various reactions in representing the particular phenotype of an organism. Nevertheless, the metabolites are the key players in determining the metabolic state of a cell. As such it is crucial to the study the roles of various metabolites in an organism's metabolism. FBA is based on the steady state assumption under which there is no net accumulation or depletion of any intracellular metabolite. However, the net turn over number (i.e., net generation or net utilization) of any metabolite can be non-zero [211]. The net generation of a metabolite in turn determines its availability in the network and thus, the extent to which the associated activities can be exhibited by a cell under the given condition. FSA [211] is an optimization based algorithm that determines the net generation of a metabolite under given conditions and constraints. It also helps in determining the effect of

attenuating/amplifying the net generation of an internal metabolite on the desired metabolic activity/ phenotype of a cell. We analysed our model to study the importance of various metabolites in the metabolism of *G. alkanivorans*.

For performing FSA, we first assumed the maximization of cellular growth as the objective function for a fixed ethanol supply (10 mmol/gdcw-h) and solved the model. Based on the resulting flux distribution values we calculated the flux sum (i.e., net generation) of each metabolite and used it as the basal flux sum value for the metabolites. Next we maximized the flux sum of each of the metabolites under the same conditions. Of the 814 internal metabolites only 34% had a non-zero basal flux sum value while the remaining 60% had a zero basal flux sum. The metabolites with high basal flux sum values were mainly the essential cofactors that had shown participation in relatively higher number of reactions than others. During the maximization of flux sum for various metabolites we found that nearly 26% of them had a zero value for the maximum flux sum achievable. The ones that had zero values for both the basal and maximum flux sums were the dead-end metabolites that could not be eliminated from the model using the available procedures. Such metabolites point to the need for a more thorough analysis of the biochemistry and metabolism of *G. alkanivorans* to account for the missing functionalities that may otherwise be associated with these dead-end metabolites. There were certain metabolites that had an infinite amount of generation irrespective of the type of the medium components and the constraints used to define the environmental conditions. These cyclic compounds arose due to the generation and consumption in equal amounts within the cycles formed by certain reactions [211]. However, the other metabolites had non-zero, measurable and small to very high values for their flux sum which were mainly the cofactors. During the flux sum attenuation, we found that only 24% of the metabolites

were essential for growth. Most of these essential metabolites were found associated with the essential reactions identified in the previous studies. We also found that when DBT was supplied as the sole sulfur source the biomass generation was linearly dependent on the flux sum values of the intermediate metabolites from the 4S pathway. As such any limitation on the amount of generation of any of the 4S pathway intermediates will lead of a reduced biomass growth of *G. alkanivorans* when DBT is supplied as sole sulfur source. Further, we repeated the flux sum analysis to study the effects of the various metabolites on the desulfurization activity. We found that only 1.5% of the metabolites were essential for DBT desulfurization. The key metabolites that exerted prominent effect on DBT desulfurization activity during the flux sum attenuation were NADH, oxygen and ferricytochrome c besides the 4S pathway intermediates. Hence, it will be essential to maintain higher levels for the regeneration of NADH and ferricytochrome c and oxygen supply for obtaining higher desulfurization activity with *G. alkanivorans*.

6.4 **Summary**

In this chapter, we presented the first genome scale metabolic model of *G. alkanivorans.* We have identified appropriate genome annotations to account for the functionalities missing from the currently available annotations for *G. alkanivorans*. The model has been successfully validated using the experimental growth data from the literature. The *in silico* analyses of the effects of various medium components have given interesting testable scenarios for improving desulfurizing activity of *G. alkanivorans*. The studies with the carbon sources and the desulfurization of BT and DBT in the mixture point to the importance of the NADH regeneration in desulfurizing activity. The modifications in the metabolic architecture of *G. alkanivorans* for

improved supply/regeneration of NADH are likely to give enhanced desulfurization rates. The flux variability and gene essentiality analyses show the essentiality of the cell wall for the survival of *G. alkanivorans*. Any external factor that may lead to the undesirable alterations in the cell wall may be lethal for *G. alkanivorans*. The utility of the model is further shown by the identification of the roles of the various metabolites. The metabolites such as NADH, ferricytochrome, etc. that exhibit a positive effect on the growth and desulfurizing activity may be supplemented in the medium for obtaining improved desulfurization *G. alkanivorans*. Our model appropriately captures the inter-relationships between the various metabolic activities occurring within *G. alknaivorans*. It can further be used to study other properties of its metabolic network and devise metabolic engineering strategies for obtaining improved strains.

7 An improved genome-scale metabolic model of *Rhodococcus erythropolis* to understand its metabolism for application as a biocatalyst for fossil fuels refining

7.1 Introduction

Among the well-known industrially important strains, rhodococci are of special interest owing to their metabolic versatility. The ability of rhodococci to metabolize/utilize a wide variety of compounds arises from the presence of large linear plasmids and multiple homologues of enzymes [175]. This enables them to be used for a large number of applications under wide range of conditions. Moreover, the faster growth, simpler developmental cycle [7], environmental persistence, and remarkable ability of the rhodococci strains to degrade diverse organic compounds and produce surfactants make them highly attractive for biotransformation and bioremediation purposes. Furthermore, they exhibit considerable tolerance to different solvents and are able to adapt to solvent toxicity by altering their membrane fluidity [217]. These characteristics make them preferred organisms for transforming or metabolizing compounds that need to be dissolved in organic solvents. In addition, the presence of mycolic acids in their cell walls facilitates the uptake of hydrophobic compounds, making them the microbial species of interest for bioremediation that involve organic compounds [7]. They have been used for various industrial level applications such as production of acrylamide, acrylic acid, bioactive steroid compounds, etc. [175].

Besides they have been well studied and characterized for their ability to degrade xenobiotics such as poly chlorinated biphenyls.

An important application of rhodococci which has been widely studied over the past two decades is the desulfurization of fossil fuels. There have been piecewise experimental studies [7] as well as holistic systems level studies [189, 204] of its metabolism to understand the properties of its metabolic network. In the chapters 3 and 4 we have presented the metabolic models of *R. erythropolis* for holistic study of its sulfur metabolism and DBT desulfurization. These holistic studies have complemented the experimental studies on desulfurization by providing plausible explanations for the various phenotypes and proposing various testable hypotheses. However, these model do not include thermodynamic constraints on the reactions. Here we present a more comprehensive model of *R. erythropolis* updated with latest genome annotations which includes more functionality and thermodynamic constraints.

In this chapter, we present an expanded model of *R. erythropolis*. It accounts for a higher number of ORFs (i.e., 1080) associated with 1285 unique reactions compared to 573 ORFs and 821 reactions included in the previous model. The predictability of the model is further improved by inclusion of thermodynamic constraints. In addition, we have identified and establish connections for the dead end metabolites (DEMs) which have enabled us to incorporate additional functionalities in the metabolic network. During the model reconstruction process we have also identified a few possible genome annotations that account for missing functionalities and thus aid in better characterization of metabolism in *R. erythropolis*. Further, we have analyzed the model to study the metabolism of several additional sulfur sources known to be metabolized by *R. erythropolis* to give a more comprehensive coverage of its sulfur metabolism. We have then analyzed our extended model using the

established methods such as flux variability analysis [210] and flux sum analysis [211] to canvas the properties of the metabolic network and identify the key metabolic components that determine the extent of the desired phenotypes. We then designed and studied a modified *in silico* strain of *R. erythropolis* for biorefining (i.e., simultaneous denitrification and desulfurization) of fossil fuels. Finally, we did a comparative analysis of *in silico* desulfurizing strains of *Pseudomonas*, *Corynebacterium*, *Mycobacterium*, *Gordonia*, and *Rhodococcus* for their activities.

7.2 Materials and methods

7.2.1 Model reconstruction

For the reconstruction of the expanded model of *R. erythropolis* we first obtained annotations of its latest genome sequence using the automated tools available on the online annotation server RAST [212]. Next, using our existing genome scale model [218] presented in chapter 4, as the initial framework we compared the metabolic activities given by the new annotations and incorporated the additional functionalities that were absent in our previous models. Since the first model of *R. erythropolis* is based on the genome annotations as available in KEGG [170] and MetaCyc [172] at the time of reconstruction, we now included the updated information from these two databases in the model expansion process. We established the appropriate gene, protein and reaction (GPR) associations and cross checked them in the existing model and the new annotations. In case of any discrepancy, we used the new and updated annotations to assign functionalities in the model with the associated genes. The reactions were elementally and charged balanced by considering the formula and charge state of the metabolites as available in SEED [219]. We assigned the directionality of the reactions

based on the information available in literature or using the thermodynamic information/analysis [220]. Since there have been no further advances in the characterization of biomass components of *R. erythropolis*, we did not make any change in the biomass equation. The generation of a computations ready model required the identification of missing reactions to allow the production of certain biomass precursors. The resulting model consisted of several DEMs that arose due to the lack of connectivity with the rest of the metabolic networks. These DEMs were identified and fixed using the established GapFind and GapFill methods [49]. The reactions thus identified to fix the DEMs were included only when an appropriate ORF in the genome of *R. erythropolis* could be identified/assigned for the related function using the BLASTp analysis.

7.2.2 **Experimental data**

An essential feature of any model is its ability to successfully predict the experimental observations both qualitatively and quantitatively. We validated our model using the data from Izumi et al. [11] and Davoodi-Dehagahani et al. [99]. The details of the process and data generation can be found in chapter 4. We use their specific DBT (HBP) uptake (output) rates and specific growth rates for the validation of our model.

7.3 **Results and discussions**

7.3.1 **Reconstructed model**

We reconstructed an expanded and improved model of *R. erythropolis*. It accounts for the various metabolic activities and the industrially important biochemical transformations occurring within *R. erythropolis*. We built the expanded model using

our genome scale model given in chapter 4, annotations of its latest genome sequence and the literature based biochemical information. The expanded model has improvements over the existing model in the following aspects: (i) alignment with the latest genome annotation, (ii) incorporation of thermodynamic constraints, (iii) analysis for dead end metabolites, and (iv) increased scope.

The entire genome sequence for the chromosomal DNA (RefSeq: NC_012490.1) and the three plasmids (RefSeq: NC_007486.1, NC_007487.1, and NC_007491.1) of *R. erythropolis* were taken from NCBI's GenBank Database [221]. We annotated these sequences using the automated tools available on the online annotation server RAST [212]. The new annotations accounted for 1214 reactions associated with 1291 ORFS. During the comparison of the existing model and updated annotations, we mapped the compounds with different identifiers to ensure proper assignment of reactions and omit any redundancies. We compared the reactions in the existing model and the new annotations using both the manual and automated methods. The reactions present in the new annotations but not in the existing model were included for the model extension. However, there were several reactions present in the new annotations that accounted for the inter-conversions of the general compound classes. These reactions represent the catalysis by enzymes with broad substrate specificity such that they are capable of transforming various compounds of a particular class into compounds of other class. In these reactions the reactants and substrates are represented by general functional groups or names instead of specific compound names. We eliminated such reactions as they cannot be included in the model for computation purposes. There were certain reactions in the existing model and new annotations that brought about the conversion of same compound but involved different cofactors. In such cases we consulted the literature for deciding the

appropriate usage of cofactors. In case of discrepancy, separate reactions with different cofactors were included in the model.

The genome scale metabolic models are solved using optimization methods by applying suitable constraints and objective function(s). These constraints represent the physicochemical, topological, and environmental limitations of the cellular networks. The genome scale models include most of these constraints but the thermodynamic constraints are usually not evaluated. The thermodynamic constraints are a part of physicochemical constraints that can be evaluated using established methods [220] and be imposed in terms of the directionalities of the reactions. In this model we applied the thermodynamic constraints on reactions by using/evaluating the associated standard Gibbs energy of reactions, $\Delta_r G'^{\circ}$. We used the values of the standard Gibbs energy of formation ($\Delta_f G'^{\circ}$) for most of the metabolites and $\Delta_r G'^{\circ}$ for reactions, as provided in SEED database [219]. The Appendix D provides the values of $\Delta_f G'^{\circ}$ and $\Delta_r G'^{\circ}$ used for the various metabolites and reactions, respectively. We have incorporated the thermodynamic constraint for 1158 (~90%) reactions.

A computations ready model should be able to show *in silico* growth when simulated with conditions representing the experimentally determined minimal media. The *in silico* growth is included in the model in terms of a biomass equation that is a hypothetical equation consuming all biomass precursors in proportions as determined from the cell's composition. However due to presence of gaps in the draft model the biomass equation could not carry any flux. Our model analysis showed that 30 (~50%) out of 66 biomass precursor could not be produced by the draft model. We used GapFill method [49] to identify the possible reactions that could be included to fill these gaps in the network to enable the production of biomass. The addition of these putative reactions to the model was justified by performing BLASTp analysis to

identify the possible ORFs in *R. erythropolis* that can encode for these missing functionalities. Since the restoration of biomass formation is critical for the reconstruction the expectation cut off value was kept low i.e., at 10^{-5}. The resulting working model still consisted of DEMs that arose due to the lack of connectivity with the rest of the metabolic network. As such, the reactions associated with these DEMs can never carry any flux through them under any condition. We applied the GapFind procedure [49] to identify all the DEMs (245 in number) present in the model. Then we used the GapFill procedure [49] to find the potential reactions that when added to the model can restore the connectivity of the DEMs. The procedure suggested possible reaction sets to restore connectivity for only 102 (~42 %) of these DEMs. Next, we performed BLASTp analyses to assign possible ORFs for these additional reactions using a more stringent expectation cut of value of 10^{-25}. Since BLASTp searches allowed us to assign ORFs only for 40 % of the identified reactions, we could restore connectivity for only 78 (32 %) of the DEMs. The presence of the remaining DEMs in the model shows a clear need for more detailed studies and characterization of the biochemistry and metabolism of *R. erythropolis*. The new genome annotations identified using the BLASTp analyses are shown in Table 7.1.

The expanded model is improved in scope as it accounts for the 464 additional reactions and 334 additional metabolites as compared to our previous model genome scale model [218] reported in chapter 4. Furthermore, it is an improvement over the previous one as the generic reactions present in the previous version have been eliminated. The model has been scrutinized to establish an improved connectivity by identifying and restoring connections for the DEMs. It accounts for additional ORFs and genes in *R. erythropolis*. The final expanded version accounts for 1080 ORFs i.e,

nearly 17% of the known 6437 ORFs corresponding to the 1287 reactions. The Table 7.2 shows the overall characteristics of the expanded model.

Table 7.1 New possible genome annotations for additional functionalities in *R. erythropolis*

Enzyme	Enzyme Name	ORF	Current Annotation	NCBI accession no.	E value
EC 1.1.1.312	2-hydroxy-4-carboxymuconate semialdehyde hemiacetal dehydrogenase	RER_09450	oxidoreductase	YP_002764392	1*10^-20
EC 1.1.2.4	lactic acid dehydrogenase	RER_27890	FAD-linked oxidase	YP_002766236	1*10^-126
EC 1.14.13.63	3-hydroxyphenylacetate 6-hydroxylase	pREC1_0055	cytochrome P450	YP_345116	2*10^-15
EC 1.14.16.4	tryptophan 5-monooxygenase	RER_01870	prephenate dehydratase	YP_002763634	3*10^-7
EC 1.2.1.22	lactaldehyde dehydrogenase	RER_05180	aldehyde dehydrogenase	YP_002763965	2*10^-97
EC 1.2.1.31	L-aminoadipate-semialdehyde dehydrogenase	RER_18370	Phaeosphaeria nodorum	YP_002765284.1	2*10^-57
EC 1.2.1.5	aldehyde dehydrogenase	RER_07680	aldehyde dehydrogenase	YP_002764215	2*10^-175
EC 1.2.3.1	aldehyde oxidase	RER_33910	xanthine dehydrogenase	YP_002766838	1*10^-42
EC 1.5.1.34	6,7-dihydropteridine reductase	RER_42720	nitroreductase	YP_002767719	5*10^-6
EC 1.5.3.1	sarcosine oxidase	RER_14560	hypothetical protein	YP_002764311	3*10^-9
EC 1.5.3.17	non-specific polyamine oxidase	RER_44090	putrescine oxidase	YP_002767856	9*10^-14
EC 2.1.1.17	phosphatidylethanolamine N-methyltransferase	RER_14200	hypothetical protein	YP_002764867	2*10^-15
EC 2.3.1.4	phosphoglucosamine transacetylase	RER_23190	hypothetical protein	YP_002765766.1	6*10^-6
EC 2.3.1.42	glycerone-phosphate O-acyltransferase	RER_16370	hypothetical protein	YP_002765084	6*10^-9
EC 2.3.1.57	diamine N-acetyltransferase	RER_07140	acetyltransferase	YP_002765828	1*10^-10
EC 2.3.3.10	hydroxymethylglutaryl-CoA synthase	RER_11360	3-oxoacyl-[acyl-carrier-protein] synthase III	YP_002764583	3*10^-8
EC 2.5.1.90	octaprenyl-diphosphate synthase	RER_17090	polyprenyl diphosphate synthase	YP_002765156	1*10^-47
EC 2.6.1.21	D-amino-acid transaminase	RER_36130	branched-chain amino acid aminotransferase	YP_002767060	1*10^-8
EC 2.6.1.39	2-aminoadipate transaminase	RER_20820	aminotransferase	YP_002765529.1	6*10^-51
EC 2.6.1.57	aromatic-amino-acid transaminase	RER_03070	aspartate aminotransferase	YP_002763754	7*10^-56
EC 2.7.1.3	ketohexokinase	RER_37820	ribokinase	YP_002767229	1*10^-6
EC 2.7.1.74	deoxycytidine kinase	RER_22740	hypothetical protein	YP_002765721	8*10^-5
EC 2.7.1.94	acylglycerol kinase	RER_50140	hypothetical protein/ diacylglycerol kinase	YP_002768461	1*10^-7
EC 2.8.2.11	galactosylceramide sulfotransferase	RER_17880	ABC transporter ATP-binding protein	YP_002765235	5*10^-5
EC 3.1.2.4	3-hydroxyisobutyryl-CoA hydrolase	RER_39530	enoyl-CoA hydratase	YP_002767400	5*10^-63
EC 3.5.4.12	dCMP deaminase	RER_14660	deaminase	YP_002764913	2*10^-7
EC 3.5.5.1	acetonitrilase	RER_44150	hydrolase	YP_002767862	3*10^-16
EC 4.1.1.22	histidine decarboxylase	RER_35380	glutamate decarboxylase	YP_002766985	6*10^-7
EC 4.1.1.28	tryptophan decarboxylase	RER_58030	lyase	YP_002769250	5*10^-8
EC 4.1.1.41	methylmalonyl-CoA decarboxylase	echA	enoyl-CoA hydratase	YP_345131	9*10^-28
EC 4.1.1.68	5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase	RER_24020	fumarylacetoacetate hydrolase family protein	YP_002765849	1*10^-27
EC 4.1.2.20	2-dehydro-3-deoxyglucarate aldolase	RER_05190	aldolase	YP_002763966	1*10^-34
EC 4.2.1.-	UDP-N-acetylglucosamine 4,6-dehydratase	RER_02500	dTDP-glucose 4,6-dehydratase	YP_002763697	3*10^-9
EC 4.2.1.40	glucarate dehydratase	RER_54220	muconate cycloisomerase	YP_002768869	2*10^-23
EC 5.4.2.3	phosphoacetylglucosamine mutase	RER_19060	phosphoglucosamine mutase	YP_002765353	3*10^-9
EC 2.3.1.182	(R)-citramalate synthase	RER_04110	2-isopropylmalate synthase	YP_002763858	6*10^-14

Table 7.2 Features and statistics of the extended and existing genome scale models of *R. erythropolis*

Feature	Extended Model	Previous Model
Reactions	1287	823
No of ORFs	1080	618
Intracellular reactions	1142	792
Transport reactions	145	31
Metabolites	1120	786

7.3.2 **Model validation**

We performed the *in silico* experiments with our expanded model to mimic conditions defined by Izumi et al. [11] and Davoodi-Dehaghani et al. [99]. We allowed

unrestricted fluxes through the other nutrients of the medium except the limiting sulfur source and calculated the maximium specific cell growth rates. In our simulations we took the growth associated maintenance (GAM) energy of 40 mmol ATP/ gdcw-h and non-growth associated maintenance energy of 8.4 mmol ATP/ gdcw-h based on those used for other closely related organisms [161, 178]. As seen in Fig. 7.1, the *in silico* growth rates predicted by our model are in close agreement with the experimental growth rates.

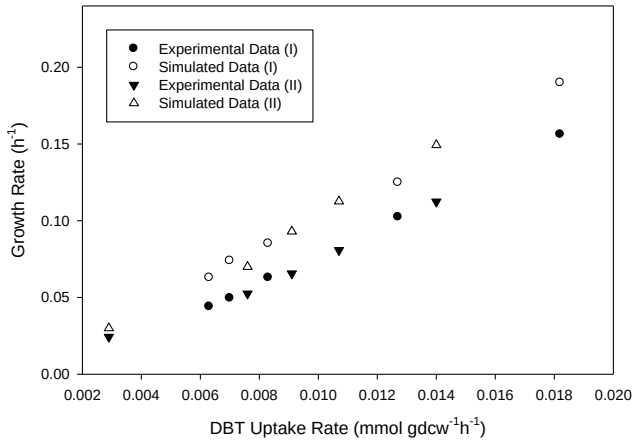


Fig. 7.1Experimental and predicted growth rates of *R. erythropolis* based on data from Izumi et al. [10] (I) and Davoodi-Dehaghani et al. [99] (II)

7.3.3 **Alternate sulfur sources**

A few experimental studies have examined the ability of *R. erythropolis* to utilize a number of sulfur compounds [11, 16]. We evaluated our model's ability to show *in silico* growth on these compounds when provided as sole sulfur sources. For this we performed 38 different simulations by supplying 38 different sulfur compounds as sole sources. Of the 38 sulfur sources examined, the *in silico* growth was observed only on 23 sources (~61%) as shown in Table 7.3. The remaining 15 sulfur sources could not be utilized as the associated metabolic pathways for their utilization are not been reported/available in literature for inclusion in the model. However, the experimental

evidences for their utilization point to the presence of appropriate pathways and metabolic capabilities in *R. erythropolis*. The lack of knowledge of their metabolic pathways points to the need for elaborate studies on the sulfur metabolism in *R. erythropolis*. Such studies will help in elucidation and better understanding of the metabolism of various polyaromatic sulfur heterocycles present in fossil fuels and thus, will help in improving biodesulfurization.

7.3.4 **Flux variability analysis (FVA)**

The genome scale metabolic models are underdetermined in nature as such there exist alternate optimal solutions that give the same value of the objective function. These arise due to the redundancy of pathways in the metabolic network. The analysis of the metabolic network to determine the ability of the organisms to use alternate pathways/reactions for the same objective function value is a critical step. It can aid in redesigning of its metabolic architecture without affecting the desired phenotypes. We used FVA to determine the ranges over which the various reaction flux values can vary for a given environmental condition. For this we fixed the supply of ethanol at 10mmol/gdcw-h and maximized biomass growth as the objective function. Next, we fixed the biomass at the calculated optimal value and sequentially tried to obtain the maximum and minimum values that each flux can take under the defined constraints. Most of the reaction fluxes could vary over a wide range of values. This showed high flexibility exhibited by the metabolism in *R. erythropolis*. Significantly high variations were seen in the reaction flux values from the central metabolism. It shows that although central metabolism plays an important role in the generation of energy and important precursors, it is flexible. However, 18% (232) of the total reactions did not allow variations in their flux values for attaining desirable level of biomass. They

represent the more stringent parts of the metabolism in *R. erythropolis*. These include mainly the reactions from cell wall metabolism and amino acids. This shows that the amino acid and cell wall metabolism may not be altered without affecting the growth of the biocatalyst for desulfurization. The exposure of the *R. erythropolis* cells to any conditions that lead to variations or mutations in these pathways would lead to reduced growth rate. This is an important result as we intend to use the cells for the desulfurization of petroleum that is a complex mixture of hydrocarbons. Hydrocarbons are known to exert toxicity to bacterial cells by altering the cell wall fluidity [222]. However, *R. erythropolis* offers the advantage as a biocatalyst for desulfurization as it shows the capacity to reduce toxic effects of hydrocarbons by altering its cell wall fluidity to minimize the harmful effect [223].

Next, we performed FVA under the same defined conditions but used the specific desulfurization activity as the objective function. As expected, we found that a different part of metabolism showed stringency in flux values than the one when biomass growth was used as the objective. Only 1.8% of the total reactions did not exhibit any considerable variation in their flux values. The reactions that mainly showed more or less fixed flux values belong to sulfur metabolism and central metabolism. The stringency in the flux through the central metabolism is seen associated with reactions that lead to the regeneration of NADH. This observation shows the important role of central metabolism in energy generation. The alterations/improvements in these parts of the metabolism may hold the key for obtaining the strains with high desulfurizing activity. However, it was seen that the biomass growth was reduced to zero when the desulfurizing activity was maximized. This shows that the biomass and desulfurizing activity are decoupled. This can be explained on the basis of the fact that both the growth and desulfurizing activities

require NADH. As such the competitive requirements for NADH lead to reduced growth when a high flux is forced through desulfurization pathway.

Table 7.3 Utilization of various sulfur sources by R. erythropolis as examined by Izumi et al. [10] and Omori et al. [15]and as predicted by the model. A 'y' means the compound can be utilized as a sole sulfur source while a 'n' means that it cannot be.

Compound	*In silico* Utilization	Experimental Utilization
1-napthalenesulfonate	n	y
2-ethylthiophene	n	y
2-mercaptoethanol	n	y
2-methylthiophene	n	y
2-napthalenesulfonate	n	y
3-methylthiophene	n	y
ammonium sulfate	y	y
benzenesulfinic acid	n	y
benzenesulfonate	n	y
benzenesulfonic acid	n	y
benzothiophene	y	y
cysteine	y	y
DBT	y	y
DBT sulfone	y	y
dibenzyl sulfoxide	n	y
dimethyl sulfate	y	y
dimethyl sulfide	y	y
dimethyl sulfone	y	y
dimethyl sulfoxide	y	y
hexanesulfonate	y	y
isethionate	y	y
methane sulfonic acid	y	y
methionine	y	y
picrylsulfonate	n	y
p-toluenesulfonic acid	y	y
sulfinoalanine	y	y
sulfipyruvate	y	y
sulfite	y	y
sulfoacetate	y	y
sulfopyruvate	y	y
sulfur	y	y
taurine	y	y
thianthrene	n	y
Thiophene	n	y
thiophene-2-acetic acid	n	y
thiophene-2-carboxylic acid	y	y
thiosulfate	y	y
thioxanthen-9-one	n	y

7.3.5 **Flux sum analysis**

We performed Flux Sum Analysis [211] using the expanded model of *R. erythropolis* to study the role of various metabolites. For this we first fixed the ethanol supply at 10mmol/gdcw-h while keeping all other nutrients at relatively higher values. Then we

maximized the biomass growth to obtain the corresponding flux distribution. These reaction flux values were then used to calculate the 'flux sum' i.e., the net generation/consumption of each metabolite in network. This represented the basal flux sum values for each of the metabolites. It was observed that only 36% of the metabolites had a non-zero basal flux sum value while the remaining metabolites had a zero basal flux sum value for biomass maximization. Next, under the same defined conditions we calculated the maximum flux sum achievable for each of the metabolites. The metabolites showed varying values for their maximum flux sum. Some of the metabolites (~21%) had a zero value for the maximum flux sum. These are the unresolved DEMs that cannot be generated in the model due to the lack of the connectivity with the rest of the metabolic network. There are some metabolites that have an infinite value for their maximum flux sum. They are the metabolites that are produced from the cyclic loops present in the model. Remaining metabolites showed finite flux sum values ranging from very low to very high. The cofactors mainly showed a high value for their flux sum as they are involved in a large number of reactions. Using the calculated flux sum values, we studied the effects of perturbations in metabolites generation levels on the biomass growth. For this, we performed repeated simulations by constraining the flux sum values of metabolites between zero and maximum values in steps and maximized biomass. There were certain metabolites that were essential for biomass generation as constraining their flux sum at zero inhibited the biomass growth. These metabolites were mainly associated with the reactions that were essential for the survival of *R. erythropolis*.

7.3.6 **Denitrogenation of fossil fuels**

The fossil fuels contain several compounds of nitrogen besides those of sulfur. The combustion of nitrogen containing organic compounds can lead to the release of oxides of nitrogen but they have not been subjected to environmental regulation. However, such compounds act as potential poisons to the catalysts used in refineries and therefore, need to be treated before the processing of fuels [5]. Thus, denitrogenation of the fuels can be an important step in the pre-processing of the fossil fuels. Carbazole is used as the model nitrogen compound to study the denitrogenation of petroleum. It is a good representative of poly aromatic nitrogen compounds commonly present in petroleum as it is the most abundant nitrogen containing organic compound found in several fractions of petroleum. Several bacterial strains [224-226] have been found to metabolize carbazole via different pathways. However, so far none of the studies have reported the denitrogenation of carbazole by the wild type strain of *R. erythropolis*. Nevertheless, it would be interesting to design and study a biocatalyst that can upgrade fossil fuels by simultaneously lowering their sulfur and nitrogen content. Therefore, we designed an *in silico* modified strain of *R. erythropolis* capable of simultaneous denitrogenation and desulfurization of fossil fuels.

We introduced the carbazole metabolism (Fig. 7.2) pathway consisting of 10 distinct steps based on the details available in different sources [172, 216]. It enables the *in silico* organism to utilize carbazole as a sole nitrogen source. The complete pathway leads to the metabolism of the carbazole to catechol and ammonium. The ammonium is utilized by the organism for its growth via the nitrogen metabolism pathway while catechol is further broken down to acetaldehyde by a number of reactions and is finally consumed by the TCA cycle. The presence of carbazole as a

nitrogen source did not inhibit or reduce the DBT desulfurizing activity of *R. erythropolis*. It was completely metabolized and utilized as a sole carbon and nitrogen source. For the specific uptake rate of 10mmol/gdcw-h of carbazole, a specific desulfurization activity of 0.06 mmol/gdcw-h and a specific growth rate of 0.70 h^{-1} were observed. However, this complete pathway led to the undesirable destruction of the carbazole skeleton. The denitrogenation of carbazole via this pathway is accompanied by the rupture of the carbon skeleton into anthralite and 2-hydroxypenta-2,4-dianoate. However, the elimination of the reactions in the pathway after the breakdown of anthralite to catechol and ammonia reduced the loss of calorific value. Furthermore, it was observed that the utilization of carbazole lead to a tremendous increment in the oxygen uptake. Since in our simulations we did not put stringent limits on the supply of oxygen, there was no competition for oxygen and hence, no effect on biodesulfurization. However, in practical purposes the high oxygen requirements in presence of carbazole metabolism may pose a problem for achieving higher desulfurization as oxygen supply is known to limit desulfurization [227]. Clearly, the carbazole denitrogenation via the aforementioned pathway could not be achieved without a considerable loss of calorific value of fuels. Therefore, it is desirable to identify genes that encode for the non-destructive pathway for denitrogenation of carbazole. It will greatly help in designing efficient and non-destructive strains of *R. erythropolis* for improved biorefining of petroleum.

7.3.7 **Comparative studies with other desulfurizing strains**

We reconstructed and analyzed the *in silico* models of five desulfurizing strains belonging to different genus, *Rhodococcus, Gordonia, Mycobacterium*, *Pseudomonas*, and *Corynebacterium*. The *in silico* desulfurizing strains of *Mycobacterium*,

Pseudomonas and *Corynebacterium* were reconstructed based on the available genome scale models of *M. tuberculosis* [161, 162], *P. putida* [178, 179], and *C. glutamicum* [214, 215], respectively. For the *in silico* desulfurizing strain of *Mycobacterium*, we considered the nutritional requirements and biomass composition given for its *in vitro* growth [161, 162] as it does not include the components required for pathogenesis by *M. tuberculosis*. Next, we compared the reaction sets in each of the three *in silico* models with the annotations provided in KEGG [170]. Then we added the 4S pathway [7] for DBT desulfurization to *in silico* desulfurizing strains of *Mycobacterium* and *Corynebacterium* based on literature evidence [228-231]. In addition we added the benzothiophene desulfurization pathway [123] to the *Mycobacterium* desulfurizing strain as it is known to desulfurize BT. It is well known from literature that *Pseudomonas* is capable of metabolizing DBT via Kodama pathway [2] and our *in silico* model confirmed this. However, Kodama pathway is destructive in nature and leads to the loss of calorific value of fuels when used for desulfurization. Therefore, we knocked out the Kodama pathway genes and reconstructed an *in silico* recombinant desulfurizing strain of *Pseudomonas* by inserting the 4S pathway genes. The recombinant *Peudomonas* strain was able to successfully desulfurize DBT non-destructively via the 4S pathway. We next compared the five *in silico* desulfurizing strains. The desulfurizing strains of *Rhodococcus, Gordonia*, *Mycobacterium*, and *Corynebacterium* showed several similarities in their cell wall metabolism and amino acids metabolism. However, they were considerably different than *Pseudomonas*.

Next, we performed simulations to compare the *in silico* desulfurizing activities of the five strains. For this, we provided 1 mmol/gdcw-h of glucose as the sole carbon source with unlimited amounts of other nutrients. In addition we allowed unlimited supply of DBT as the sole sulfur source. Next we maximized biomass and calculated

the corresponding desulfurization activity. We observed that the growth rates of the strains with the corresponding desulfurizing activity are in the order *Pseudomonas* (0.17 h^{-1}, 40.39 µmol HBP/gdcw-h) > *Gordonia* (0.15 h^{-1}, 18.13 µmol HBP/gdcw-h) > *Rhodococcus* (0.14 h^{-1}, 13.54 µmol HBP/gdcw-h) > *Corynebacterium* (0.09 h^{-1}, 21 µmol HBP/gdcw-h) > *Mycobacterium* (0.05 h^{-1}, 4 µmol HBP/gdcw-h). We then calculated the minimum sulfur (i.e., DBT) requirements for a unit growth rate for each of the *in silico* strains. For this, we supplied all the nutrients in excess, constrained the biomass growth rate to be 1 h^{-1} and minimized the DBT uptake. We found the sulfur requirements of sulfur for the strains were in the order *Pseudomonas* (233.06 mmol/gdcw-h) = *Corynebacterium* (233 µmol/gdcw-h) > *Gordonia* (120 µmol/gdcw-h) > *Rhodococcus* (93.90 µmol /gdcw-h) > *Mycobacterium* (81 µmol/gdcw-h). It is evident from these results that the sulfur requirements for growth determine the desulfurization activity in the strains. The *Pseudomonas* shows highest growth rate and thus, the highest desulfurizing activity owing to its high sulfur requirements. Although the minimum sulfur requirements for unit growth rate of *Corynebacterium* are almost same as *Pseudomonas*, the growth rate and corresponding desulfurization for a unit uptake rate of glucose is lower for *Corynebacterium*. The lower growth rate of *Corynebacterium* can be explained on the basis of the observation that the *in silico* growth was accompanied by excretion of acetate. This leads to diversion of a part of the carbon flux away from growth. From these observations *Pseudomonas* appears to be most interesting candidate for desulfurization however, it needs to be genetically engineered for desulfurization of DBT in a non-destructive manner. Nonetheless, the high sulfur requirements of *Corynebacterium* make it an equally interesting desulfurizing biocatalyst. Any evidence of acetate formation by *Corynebacterium* under DBT desulfurization condition and consequent reduced growth is not reported in

literature. Therefore, it will be thought-provoking to study conditions that support higher growth rates of *Corynebacterium* without the loss of carbon flux as any undesirable side product such as acetate.

As seen in the previous studies the DBT desulfurization activity by any bacterial strain is determined by its sulfur requirements for growth. However, the sulfur content of bacterial cells is usually low. Therefore, in order to obtain higher desulfurizing rates it is desirable to have strains that can metabolize DBT beyond their growth requirements. We designed recombinant *in silico* desulfurizing strains for the above five candidates with high copy numbers of *dsz* genes by allowing unrestricted flux through the '4S' pathway. Next, we allowed fixed glucose uptake (1 mmol/gdcw-h) with unlimited amounts of other nutrients (DBT, oxygen, etc.) and calculated the maximum desulfurization activity possible for each of the recombinant strain. The order of desulfurization activity obtained is in the order *Mycobacterium* > *Pseudomonas* > *Gordonia* > *Corynebacterium* >*Rhodococcus*. An important observation is that for maximum desulfurization activity for each of the strains, the corresponding biomass growth was zero. This decoupling of the fluxes through growth and '4S' pathway arises due to the competitive demands for NADH as explained earlier in chapters 3, 4, and 5. The FSA with the five desulfurizing strains showed that NADH regeneration and oxygen supply are essential for the desulfurization activity.

The fossil fuels contain various PASHs of which DBT and BT are the major representatives. For an efficient desulfurizing process it is desirable to have strains that exhibit high desulfurizing activities for a wide range of compounds. However, of the above five desulfurizing strains only *Gordonia* and *Mycobacterium* showed desulfurization of both DBT and BT naturally. Therefore, it will be interesting to carry

out more studies with them for obtaining higher desulfurization for a wider range of compounds.

7.4 **Summary**

We built an improved model of *R. erythropolis* by updating our previous model to include additional functionalities based on new genome annotations. The model predictability is further improved by including the thermodynamic constraints in terms of the reaction directionality. The model has been successfully validated and analyzed using the experimental studies reported in literature. We next designed and analysed *in silico* modified strain of *R. erythropolis* for bio-refining (i.e., simultaneous desulfurization and denitrogenation) of fossil fuels. It is found that oxygen supply may act as limiting factor for the bio-refining due to high oxygen requirements by both the desulfurization and denitrogenation pathways. Further, the reconstruction and comparative analysis of the various *in silico* wild type desulfurizing strains of *Corynebacterium, Mycobacterium, Gordonia* and *Rhodococcus* show that *Gordonia* exhibits higher desulfurizing activity than the other.

8 *In silico* metabolic modeling of co-culture of *Rhodococcus erythropolis* and *Gordonia alkanivorans* for efficient biodesulfurization of benzothiophene and dibenzothiophene mixtures in fossil fuels

8.1 Introduction

The various thiophenic compounds, BT, DBT and their derivatives show recalcitrance to the hydrodesulfurization process. However, in order to obtain fossil fuels with lower sulfur content it is essential to desulfurize all of these compounds. Various desulfurization biocatalysts show activities for some or all of these compounds [2]. Most biodesulfurization studies have considered DBT as the model compound. In this regard, *R. erythropolis* strains are most popularly studied [7]. However, most of them do not exhibit activity for BT and its derivatives. Since BT and its derivatives contribute to a considerable portion of the total sulfur content in fossil fuels, it is critical to study their desulfurization. Moreover, as DBT and BT are desulfurized via distinct pathways [7, 123] it is substantive to study microbes that desulfurize BT. In this regards, several gordoniae strains have been studied for their ability to desulfurize BT and its derivatives [120, 123, 133] but most of them are unable to utilize DBT. Nevertheless, for an efficient desulfurization process, it is desirable to use a

desulfurizing biocatalyst that exhibits activity for a wider range of PASHs. In this regard, *G. alkanivorans* 1B is of interest as it exhibits desulfurization of thiophene, DBT, BT and their derivatives [35]. However, when supplied with a mixture of BT and DBT, it does not utilize them simultaneously. It preferentially uses BT and takes up DBT only after the BT has been exhausted from the medium. This results in longer time required to utilize both the sulfur substrates i.e, the complete desulfurization of the mixture. Similar observations have been made with other strains that exhibit desulfurization of multiple PASHs [102, 232]. However, a competent desulfurizing strain should be able to metabolize various PASHs simultaneously. In spite of dedicated years of research no such efficient strain has been obtained. Moreover, the level of desulfurizing activity exhibited by the known desulfurizing strains is too low for commercial applications [218]. Even the genetically engineered strains have been unable to give desired levels of activity. Although different metabolic engineering strategies can be used for introducing non-native pathways for achieving higher activity levels for diverse sulfur compounds, they exert detrimental metabolic burden on the strains [233]. An alternative to overcome this problem is the use of mixed microbial cultures of strains that exhibit desulfurization activity for different PASHs.

The mixed cultures can act more efficiently as each participating strain can utilize its preferred source thus, reducing the metabolic burden on others. Furthermore, in mixed cultures each of the strain can be engineered separately to improve its natural abilities than the introduction of non-native activities [234, 235]. Thus, it will be easier to optimize the flux through the native pathways of each strain than forcing non-native activities using metabolic engineering studies. However, it is usually challenging to establish co-cultures with desired levels of activities. Moreover, there is only a limited literature available on the co-culture based desulfurization studies [33]. *In silico*

metabolic modeling strategies can be used to study the combinations of mixed cultures and evaluate their performance for obtaining efficient desulfurizing systems. Such *in silico* approaches have been successfully used to study the microbial consortia for the utilization of mixtures of pentose and hexose sugars for ethanol production [234, 235]. In case of mixed substrates, the organisms exhibit sequential uptakes of substrates but simple FBA is unable to capture this phenotype. Mahadevan et al. [236] proposed an extension of FBA, Dynamic Flux Balance Analysis (DFBA) that can incorporate the rates of changes of fluxes as constraints explicitly in the formulation. DFBA can be used to predict the time profiles of the substrate, products, and the biomass and gives more complete information on the metabolism. It is also capable of predicting the sequences in which the substrates from a mixture are consumed.

In this chapter we have performed DFBA to study the desulfurization of DBT and BT provided in a mixture to *R. erythropolis* and *G. alkanivorans*. For this, we included the Michaelis-Menten kinetics for the uptake of DBT by *R. erythropolis* based on the literature [237]. We assumed the similar kinetics for the consumption of both BT and DBT by *G. alkanivorans* and estimated the relevant parameters using the available experimental data [35]. First, we performed DFBA to study the consumptions of the substrates by the pure cultures of *R. erythropolis* and *G. alkanivorans*. We further extended our approach to model a co-culture of *R. erythropolis* and *G. alkanivorans* for improved biodesulfurization. We modeled the system assuming that there is no interaction between the two strains. We used our co-culture model to study the desulfurization of BT and DBT when provided in a mixture. We used our models to compare the desulfurization ability of the mixed culture with those of the pure cultures. Further, we used our co-culture model to determine the concentrations of the two strains in the initial inoculum for efficient desulfurization.

8.2 **Materials and methods**

8.2.1 **Experimental data**

We used the experimental data from Izumi et al. [11] and Alves et al. [35] to validate the dynamic models for the pure cultures of *R. erythropolis* and *G. alkanivorans*, respectively. The details of the experimental setup of Izumi *et al.* [11] is discussed in chapter 4.

Alves et al. [35] isolated *G. alkanivorans* 1B strain for its ability to utilize BT, DBT, and other thiophenic compounds as sole sulfur sources. In their study, they used a glucose based sulfur free medium consisting of 1.22 g of NH_4Cl, 2.5 g KH_2PO_4, 2.5 g $Na_2HPO_4 \cdot 2H_2O$, 0.17 g $MgCl_2 \cdot 6H_2O$, and 1 L milli-Q water. To this they added 0.5 mL of a trace elements solution without sulfur containing 25 g/L EDTA, 2.14 g/L of $ZnCl_2$, 2.5 g/L $MnCl_2 \cdot 4H_2O$, 0.3 g/L $CoCl_2 \cdot 6H_2O$, 0.2 g/L $CuCl_2 \cdot 2H_2O$, 0.4 g/L $NaMoO_4 \cdot 2H_2O$, 4.5 g/L $CaCl_2 \cdot 2H_2O$, 2.9 g/L $FeCl_3 \cdot 6H_2O$, 1.0 g/L H_3BO_3, and 0.1 g/L KI. They supplemented their glucose based medium with 1mM of either BT or DBT as the sole sulfur source. They grew the culture at 30°C, pH 6.9–7.0, with 150-rpm shaking. During cultivation, they measured cell growth (via turbidometry at 660 nm) and BT/DBT concentrations (via gas chromatography) at various times.

8.2.2 **Stoichiometric models**

For this work we used the improved extended model of *R. erythropolis* presented in chapter 7 of this thesis and the genome scale model of *G. alkanivorans* reported in chapter 6. The *R. erythropolis* model consists of 1287 reactions associated with 1080 ORFs. It consists of 1120 metabolites that are compartmentalized either as intracellular or extracellular. The *G. alkanivorans* model consists of 922 reactions associated with

568 ORFs. In includes 881 metabolites of which 814 are compartmentalized as intracellular while the remaining 67 metabolites are available as extracellular metabolites. *R. erythropolis* is unable to utilize BT as the sole sulfur source while *G. alkanivorans* is modeled to utilize both BT and DBT as has been reported in the literature.

We assumed the two species to be non-interacting. Therefore, the flux balance model for the co-culture was formulated as

$$\text{Maximize} \qquad Z_R + Z_G \tag{8.1}$$

$$\text{Subject to} \qquad \begin{bmatrix} S_{ij} & 0 \\ 0 & S_{i\prime j\prime} \end{bmatrix} \begin{bmatrix} v_j \\ v_{j\prime} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8.2}$$

$$\begin{bmatrix} v_{j,\min} \\ v_{j',\min} \end{bmatrix} \leq \begin{bmatrix} v_j \\ v_{j\prime} \end{bmatrix} \leq \begin{bmatrix} v_{j,\max} \\ v_{j',\max} \end{bmatrix} \tag{8.3}$$

where, Z_R and Z_G are the growth rates of *R. erythropolis* and *G. alkanivorans*, respectively. $S_{i,j}$ and $S_{i',j'}$ are the associated stoichiometric matrices while v_j and $v_{j'}$ are the reaction flux vectors for the two species. The objective function has been taken to be a combination of their growth rates as microorganisms are known to be evolved for maximizing their growth. The resulting objective function represents the attempt of both the species to maximize their growth using the available resources. The models were solved using GAMS/CPLEX [194].

8.2.3 **Dynamic flux balance analysis for the co-culture**

The flux balance analysis formulation discussed in previous section gives only the instantaneous changes in fluxes [155, 166, 236] and does not give any information on the metabolites concentration or dynamic behavior of the metabolic fluxes. Therefore, we extended our models to perform dynamic flux balance analysis. For this, we modified our models by incorporating the substrate uptakes kinetics. DBT uptake is

known to follow the Michaelis-Menten kinetics [237] and we assumed the same for BT. We included the following equations for the representing the uptake of BT and DBT,

$$v_{D,R} = v_{D,max,R} \frac{D}{K_{D,R}+D} \quad (8.4)$$

$$v_{D,G} = v_{D,max,G} \frac{D}{K_{D,G}+D} \quad (8.5)$$

$$v_{B,G} = v_{B,max,G} \frac{B}{K_{B,G}+B} \quad (8.6)$$

where, D and B are the extracellular concentrations of DBT and BT, respectively. $v_{D,max,R}$ and $v_{D,max,G}$ are the maximum DBT uptake rates for *R. erythropolis* and *G. alkanivorans*, respectively and $K_{D,R}$ and $K_{D,G}$ are the corresponding saturation constants. $v_{B,max,G}$ is the maximum BT uptake rate for *G. alkanivorans* and $K_{B,G}$ is the corresponding saturation constant.

In addition, we described the rates of changes of the fluxes for extracellular biomass, DBT, BT, 2-hydroxybiphenyl (HBP) and o-hydroxystyrene (HSty) as follows,

$$\frac{dX_R}{dt} = Z_R X_R \quad (8.7)$$

$$\frac{dX_G}{dt} = Z_G X_G \quad (8.8)$$

$$\frac{dD}{dt} = -v_{D,R} X_R - v_{D,G} X_G \quad (8.9)$$

$$\frac{dB}{dt} = -v_{B,G} X_G \quad (8.10)$$

$$\frac{dHBP}{dt} = v_{HBP,R} X_R + v_{HBP,G} X_G \quad (8.11)$$

$$\frac{dHSty}{dt} = v_{HSty,G} X_G \quad (8.12)$$

where, X_R and X_G are the biomass concentrations of *R. erythropolis* and *G. alkanivorans* respectively. $v_{HBP,R}$ and $v_{HBP,G}$ are the HBP production rates for *R. erythropolis* and *G. alkanivorans*, respectively. $v_{HSty,G}$ is the HSty production rate for

G. alkanivorans. The values of various parameters used for the simulations are given in Table 8.1.

Table 8.1 Kinetic parameter values used for DFBA

Parameter	Value
$v_{D,max,R}$	0.0542 mmol/gdcw·
$v_{D,max,G}$	0.209 mmol/gdcw-
$v_{B,max,G}$	0.029 mmol/gdcw-
$K_{D,R}$	0.365 mM
$K_{D,G}$	6.61 mM
$K_{B,G}$	0.0439mM

8.3 Results and discussions

8.3.1 Utilization of DBT by *R. erythropolis*

Our model analysis showed that the wild type *in silico* strain of *R. erythropolis* is capable of utilizing DBT as a sole sulfur source but not BT. Therefore, we performed DFBA to simulate the experiments of Izumi et al. [11] to study the utilization of DBT by a pure culture of *R. erythropolis* in a batch culture. For the simulations we took the initial concentrations of biomass, DBT, and HBP from their experimental data. Since sulfur was the limiting substrate in their experiments we did not restrict the supply of other nutrients in the *in silico* medium. In accordance with their medium conditions, we supplied glucose as the sole carbon source. The values for maximum uptake rate ($v_{D,max,R}$) and substrate constant ($K_{D,R}$) were taken from literature [237] and are given in Table 8.1. During the simulations we broke the total time period into small time intervals and sequentially calculated the total concentrations of the various substrates and products up to the end of each time interval. We assumed maximization of the

biomass growth rate of *R. erythropolis* to be the objective function. Fig. 8.1 shows the predicted time profiles of DBT, HBP, and biomass. As evident in figure, the model successfully shows similar trends for the utilization of DBT and the production of biomass and HBP. The batch time i.e., the time taken for the complete consumption of the sulfur substrate(s) present in the medium, predicted by the model is close to the experimental time taken for complete exhaustion of DBT. However, there is slight error for the consumption of the DBT. These deviations observed arise due to the various assumptions made during the reconstruction of the model, estimation of kinetic parameters, and presence of unknown regulatory mechanism. During the model reconstruction, due to the unavailability of the strain-specific parameters such as biomass composition and maintenance energy values for *R. erythropolis* we adapted them from a related organism, *M. tuberculosis*. Therefore, it will be interesting to determine these values for *R. erythropolis*. It is likely that the use of actual measurements of these parameters for *R. erythropolis* will help in improving the model predictions.

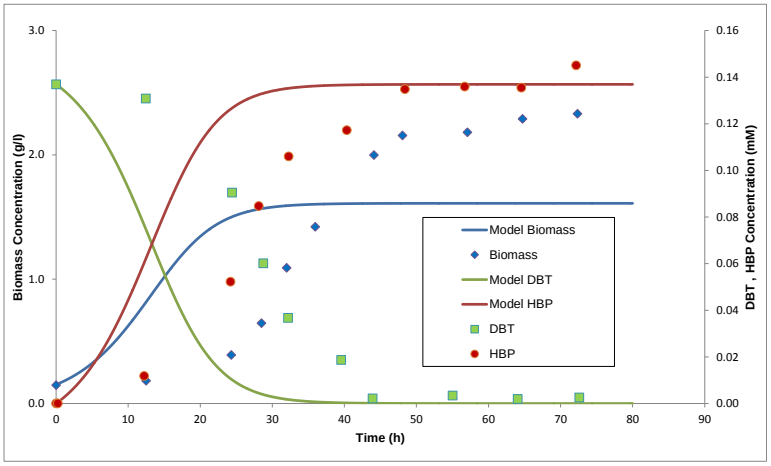


Fig. 8.1 Comparison of model predictions and experimental data on consumption of DBT by in a batch culture by *R. erythropolis*

8.3.2 **Utilization of DBT and BT in a mixture by *G. alkanivorans***

G. alkanivorans is known to utilize both BT and DBT as sole sources of sulfur [35] and this is successfully reflected by our *in silico* genome scale model. We next analyzed our model to study the utilization of BT and DBT when supplied together in the medium to *G. alkanivorans*. We incorporated the Michaelis Menten kinetics for the uptake of these substrates into our genome scale model. Since the kinetic parameter values for BT and DBT uptake by *G. alkanivorans* are not available in the literature, we approximated them using the experimental data of Alves et al. [35]. We used the linear regression method for the parameter fitting. We performed DFBA to study utilization of BT and DBT by simulating the experiments of Alves et al. [35]. For our *in silico* experiments, we used the values of initial concentrations for BT, DBT, and biomass to predict their time concentration profiles. We assumed maximization of the biomass growth rate of *G. alkanivorans* to be the objective function. Fig. 8.2 shows the comparison of the model predictions and experimental data for the two substrates and biomass. As seen in the figure, the trends for the *in silico* consumption of BT and DBT are similar to the experimental trend of their consumption by *G. alkanivorans*. It is clearly seen that there is a limited consumption of DBT prior to the complete exhaustion of BT. The *in silico* model uses BT in preference to DBT as observed experimentally. However, there is a significant difference in the batch time, i.e., the total time taken for the complete exhaustion of both the sulfur substrates, BT and DBT. This can be attributed to the errors associated in the calculation of the Michaelis Menten kinetics parameters for the consumption of BT and DBT by *G. alkanivorans*. Since the experiments were out of the scope of our work, we used the experimental data from the literature. However, the data set was too small for the appropriate

calculation of the desired parameters but for this preliminary work we approximated the values. It is therefore desirable to perform experiments to determine the kinetics and the associated parameters for the utilization of BT and DBT by *G. alkanivorans*.

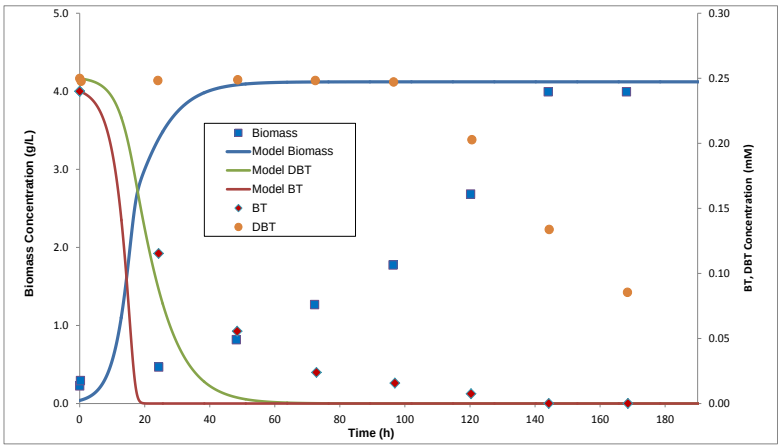


Fig. 8.2 Comparison of model predictions and experimental data on consumption of BT and DBT in a mixture by *G. alkanivorans*

8.3.3 *In silico* co-culture of *G. alkanivorans* and *R. erythropolis* for the utilization of a mixture of BT and DBT

We next used an *in silico* co-culture of *G. alkanivorans* and *R. erythropolis* for the utilization of BT and DBT in a mixture. For modeling of the co-culture, each of the species was modeled as a distinct compartment, an approach similar to the modeling of different organelles in eukaryotes. Such approaches have been used to model microbial consortia [238]. We implemented it by incorporating the reactions and metabolites for each of the species as parts of separate participating networks. The common intracellular metabolites for each network are distinguished by using different abbreviations or identifiers so that the intracellular metabolites are not shared by the two species. The extracellular compounds supplied in the *in silico* medium were made available to both the species through separate transport reactions defined for each of the species. We did not allow any interactions between *G. alkanivorans* and *R.*

erythropolis by direct exchange of their intracellular metabolites but they could interact by exchanging the components through the extracellular space. We then incorporated the kinetics for the uptake of BT and DBT for both the species using appropriate parameters.

Then we performed DFBA to study the desulfurization of BT and DBT when supplied in a mixture to the co-culture. We do not have any experimental data for this kind of study. For our analysis we assumed the mixture of BT and DBT with the same initial concentrations as those used in previous studies with the pure culture *G. alkanivorans*. The initial inoculum was assumed to be consisting of equal amounts (0.05 g/l) of both the species in the co-culture. We solved the model to obtain the time concentration profiles of BT, DBT, and biomass concentrations of both the species. We maximized the function representing the sum of the individual biomass growth rates of *R. erythropolis* and *G. alkanivorans*. This is equivalent to modeling that both the species are trying to maximize their growth using the available substrates in the mixture. Fig. 8.4 shows the time profiles for the *in silico* consumption of the DBT and BT by the co-culture. It is evident from the figure that there is a simultaneous consumption of BT and DBT by the co-culture as opposed to the sequential consumption observed with the pure culture of *G. alkanivorans*. In the co-culture, *G. alkanivorans* utilizes BT while *R. erythropolis* utilizes DBT for the growth. Nearly both the substrates are consumed at the same time. As such the batch time for the complete *in silico* consumption of both BT and DBT from the medium is lower for the co-culture (34 h) as compared to the *in silico* batch time observed with the pure culture, *G. alkanivorans* (100 h). This study shows that a co-culture will act as more efficient biocatalyst for the reduction in overall sulfur content of fossil fuels in a relatively shorter period of time.

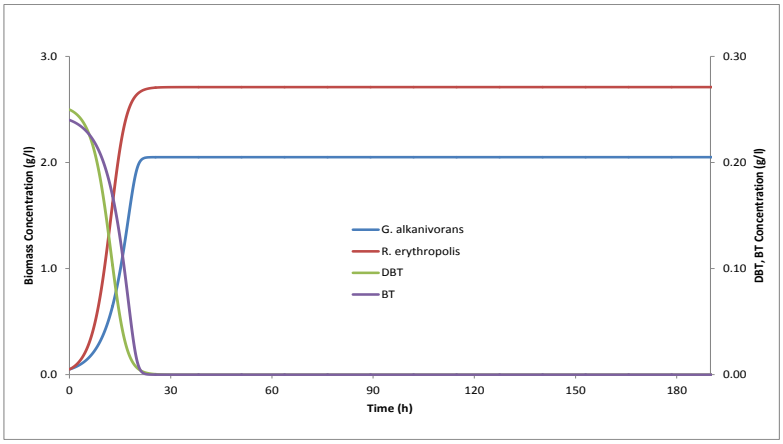


Fig. 8.3 Time profiles for consumption of BT and DBT in a mixture by a co-culture of *R. erythropolis* and *G. alkanivorans*

Next we studied the effect of varying the amounts of the two species in the initial inoculum of the co-culture on batch time for complete consumption of BT and DBT supplied together in a mixture. For this, we assumed the total co-culture concentration to be 0.1 g/l while we varied the concentration of each strain to vary between 0.01 to 0.09 g/l respectively. We then performed DFBA to obtain the concentration profiles for BT, DBT, and biomass of the two species. We used the same initial concentration of the BT and DBT in the mixture as used previously. It was seen that the batch time changed with the change in relative amounts of the two species in the initial inoculum. Of the eight different combinations assumed we found that the shortest batch time was observed when the concentrations of the strains used were 0.02 g/l for *G. alkanivorans* and correspondingly 0.08 g/l for *R. erythropolis* as seen in Fig. 8.4. Moreover, it was seen that as the amount of *G. alkanivorans* increased in the inoculum with relative decrease in the amount of *R. erythropolis*, the batch time increased. This can be explained on the basis of the fact that *G. alkanivorans* prefers BT over DBT as the sulfur source while *R. erythropolis* uses only DBT. As such, when there is more *G. alkanivorans* in the initial inoculum, the BT is consumed rapidly

while the consumption of DBT is lowered due to lower requirements posed by *R. erythropolis*. This leads to an increased batch time over which both the sulfur substrates are completely consumed and removed from the medium.

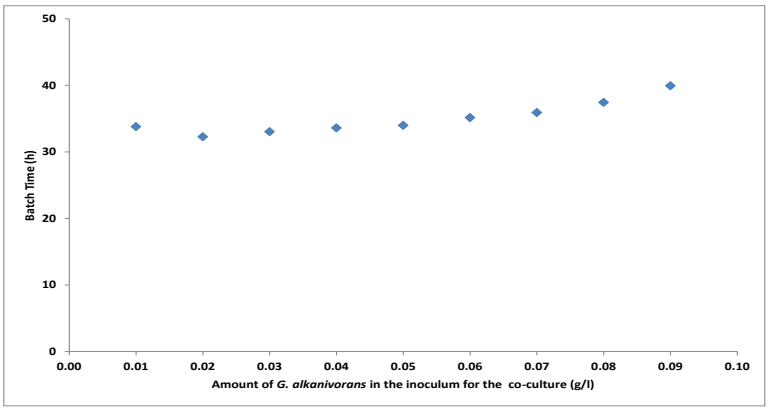


Fig. 8.4 Effect of changes in amounts of *G. alkanivorans* in the inoculum on batch time. The total inoculum concentration is assumed to be 0.1 g/l

8.4 **Summary**

We performed DFBA with the pure culture of *R. erythropolis* to study the time profiles for the consumption of DBT and production of HBP. The trends followed by the predicted concentration profiles were in close agreement with the experimental data. Further, the DFBA with *G. alkanivorans* to study utilization of BT and DBT in a mixture successfully showed that BT is preferentially used up first as the sulfur source followed by DBT. As expected it is also seen that co-culture helps in the complete desulfurization of a mixture of BT and DBT in a shorter period of time than the pure culture. The relative amounts of two species in a co-culture have a prominent effect on the total time required for complete desulfurization of a given mixture of BT and DBT.

9 Conclusions and Future Recommendations

9.1 Conclusions

Biodesulfurization is widely studied as an alternate economical desulfurization method. Nevertheless, the lack of efficient desulfurizing strains has limited the development of biodesulfurization as a commercial technology. A major limitation in the currently available studies is the lack of a holistic study of the cellular metabolism of desulfurizing strains. As a result, it is unclear how the sulfur from DBT gets incorporated into biomass. In addition, host functions that are likely to control the extent of desulfurization activity are unknown. These missing links in the literature have motivated us to develop *in silico* metabolic models of the two commonly studied desulfurizing strains, *R. erythropolis* and *G. alkanivorans* to study their various metabolic activities as an integrated network. The following section summarizes the key conclusions of our work.

***In silico* model of sulfur metabolism in *R. erythropolis*:** We reconstructed an *in silico* stoichiometric model for sulfur metabolism in *R. erythropolis*. It gives a quantitative elucidation of the process by which the sulfur from DBT gets incorporated into biomass. The model successfully predicts cell growth and several known/unknown phenotypes. Our analysis shows that NADH plays a critical role in desulfurization activity. Any changes in medium design or genetic manipulations, etc., which increase NADH regeneration and supply within the cellular metabolism, are likely to enhance desulfurization activity.

Reconstruction of an *in silico* genome scale model of *R. erythropolis*: We reconstructed the first genome-scale model for the metabolic network of *R. erythropolis*. It successfully predicts and explains several experimental phenotypes reported in the literature on *R. erythropolis*. The design of appropriate *in silico* mutants of *R. erythropolis* indicates that cysteine, like methionine, may not repress *dsz* genes. The gene essentiality analyses highlight that the maintenance of cell membrane's integrity is crucial for the growth and thus, functioning of *R. erythropolis*. A comparative study of *in silico* desulfurizing strains of *E. coli* and *R. erythropolis* shows that sulfur requirements for growth play an important role in determining the extent of desulfurization.

Role of SOR and SR for Enhanced Desulfurization by *R. erythropolis*: We have investigated the possible routes taken by sulfur from DBT to convert into biomass or other metabolites. We propose two alternate hypotheses. In the first, we hypothesize that the cell can convert via sulfite reductase (SR) the sulfite from the metabolism of DBT into sulfide that can be assimilated into biomass. However, in the process, it may convert any excess sulfite into extracellular sulfate via sulfite oxidoreductase (SOR) to avoid the toxic effects of sulfite. In the second, we speculate that the cell cannot assimilate the sulfite directly into biomass via SR. It must first use SOR to produce extracellular sulfate, and then recapture that sulfate into biomass via SR. Thus, either way, we propose that SOR and SR activities, in addition to dsz genes and cofactors, may be critical in increasing desulfurization levels significantly. In particular, we suggest that the simultaneous increase in SOR activity and decreased SR activity can give enhanced desulfurization activity.

Reconstruction of an *in silico* genome scale model of *G. alkanivorans*: We have presented the first genome scale metabolic model of *G. alkanivorans*. The model has

been successfully validated using the experimental growth data from the literature. The studies with the carbon sources and the desulfurization of BT and DBT in a mixture point to the importance of the NADH regeneration in determining the extent of desulfurizing activity achievable. The modifications in the metabolic architecture of *G. alkanivorans* for improved supply/regeneration of NADH are likely to give enhanced desulfurization rates. The model analysis shows that the metabolites such as NADH, ferricytochrome, etc. exhibit a positive effect on the growth and desulfurizing activity. As such a supplement of these metabolites in the medium may give enhanced desulfurization with *G. alkanivorans*.

An improved *in silico* model of metabolic network of *R. erythropolis* for bio-refining of fossil fuels: We developed an updated genome scale metabolic model of *R. erythropolis* by including additional functionalities based on the new genome annotations. It also incorporates thermodynamic constraints that improve its predictability. The design and analysis of an *in silico* modified strain of *R. erythropolis* for bio-refining (i.e., simultaneous desulfurization and denitrogenation) of fossil fuels shows that oxygen supply may act as limiting factor due to high oxygen requirements by both the desulfurization and denitrogenation pathways. Further, the reconstruction and comparative analysis of the various *in silico* wild type desulfurizing strains of *Corynebacterium, Mycobacterium, Gordonia* and *Rhodococcus* show that *Gordonia* exhibits higher desulfurizing activity than the other.

***In silico* metabolic model of a co-culture of *G. alkanivorans* and *R. erythropolis*:** We presented an *in silico* co-culture of *G. alkanivorans* and *R. erythropolis* for desulfurization of a mixture of BT and DBT. It is seen that the co-culture helps in simultaneous consumption of the two sulfur substrates as opposed to the sequential consumption when supplied to the pure culture of *G. alkanivorans*. Further, it is also

seen that initial amount of each species used for the inoculum of the co-culture also has a prominent effect on the desulfurization efficiency.

Overall, this work reports the *in silico* metabolic models of *R. erythropolis* and *G. alkanivorans* that quantitatively elucidate the interactions between the various metabolic activities and the desulfurization pathways. The analyses of the models provide insights into their cellular metabolic networks and help in providing possible explanations for the inexplicable experimental observations reported in literature. In addition, the various *in silico* experiments with the models give new testable hypotheses that can aid in the design of the improved strains.

9.2 Future recommendations

9.2.1 Expanding the horizon of metabolism in *R. erythropolis*

R. erythropolis strains are well known for their metabolic versatility. They are capable of carrying out a number of industrially important chemical transformations. However, the reported models have been mainly considered for the application of *R. erythropolis* as a biocatalyst for fuel refining. Therefore, the utility of the model can be increased by adding the metabolic activities associated with other commercially interesting biotransformation(s). This will enable the systems level study of *R. erythropolis* as a useful biocatalyst for other applications. In addition, the predictability of the model can be improved by adding the regulatory information.

Inclusion of regulatory information: The bacterial cells are usually exposed to constantly changing environments. They react to such fluctuations by changing the types and levels of genes and their associated proteins and enzymes in order to adapt to the new environment. As such, the bacterial strains have evolved diverse sophisticated

mechanisms to regulate their metabolism that ensure their survival under the dire or unusual conditions. These mechanisms usually involve series of complex interactions between the various components of the metabolic network that result in the control of gene expressions. Hence, any approach that incorporates regulatory information with the cellular metabolism will significantly improve our understanding of the metabolism in *R. erythropolis*

Exploring *R. erythropolis* for the production of value added products: Over the years, there has been deliberation on utilizing the complex hydrocarbons present in petroleum for synthesis of useful 'petrochemicals' than just burning them for fuels [6]. The alkylated HPBSs are examples of such value added petrochemicals. They are valuable as they exhibit hydrotrope properties and can be used as effective detergents. They are the penultimate products obtained from the desulfurization of DBT and its derivatives via '4S' pathway. Since DBT and its derivatives are usually abundant in low value high sulfur petroleum streams, the production of the HPBS and its derivatives using economical biological processes under ambient conditions appears to be an interesting application. Therefore, the desulfurizing *R. erythropolis* strains can be suitably modified and used for the commercial level production of these high value products.

Engineering of *R. erythropolis* for enhanced desulfurizing activities: A major utility of an *in silico* metabolic model of an organism is the characterization and elucidation of the complex interactions between the various metabolic components. Such insights can guide the steps to engineer the metabolic network for enhancing the fluxes through the desired pathways. The *in silico* analysis of the genome scale model of *R. erythropolis* in chapter 5 hypothesizes the role of SOR and SR in determining the extent of desulfurization activity. Experiments can be carried out to engineer *R.*

erythropolis strains to modulate the amounts and activity levels of SOR and SR for obtaining improved strains. Further, the expanded model can be analyzed using the established *in silico* methods for designing more metabolic engineering strategies for improved desulfurizing strains. The results of these *in silico* analyzes can then be used to engineer the *R. erythropolis* strains for enhanced desulfurization.

Experimental studies to determine strain specific parameters of *R. erythropolis*: The reconstruction of a genome scale model requires certain strain specific parameters such as biomass composition, maintenance energy, etc. However, these values are currently not available for *R. erythropolis*. The current models have adapted this information from related organisms. Measurements of these parameters and values will help in improving the predictability of the genome scale models.

9.2.2 Improved *in silico* models and strains of *G. alkanivorans*

The members of the genus *Gordonia* have attracted much interest of researchers over the past one and a half decades. They exhibit the ability to metabolize and transform a broad range of chemicals [9]. As such the gordoniae can be used as biocatalysts for the industrial and environmental biotechnology. As stated earlier, any cellular phenotype results from complex interactions of the various metabolic activities. The *in silico* model of *G. alkanivorans* reported in this thesis can be expanded and analyzed to study other applications. The model can be further improved by including additional constraints that can reflect the properties of the metabolic network in *G. alkanivorans*, more appropriately.

Integration of regulatory and thermodynamics constraints: It has been shown that every organism responds to any environmental or genetic perturbation by substantial variations in its gene expression patterns. The development of high throughput

technologies has enabled the study of gene expression profiles of a cell under different conditions. It helps in elucidating the regulation mechanism that the cell uses to overcome the stress caused by any unwanted perturbation. The integration of these regulatory conditions such as variable gene expressions, protein-protein interactions, etc. into the *in silico* model will help in determining which ORFs can be transcribed under the given conditions. This will help in restricting the solution space by making a limited number of pathways available under the given conditions and will represent the flux distributions more appropriately. The thermodynamic constraints represent the physicochemical constraints that help in assessing the reversibility of the reactions. Thus, the inclusion of regulatory and thermodynamics constraints can help in providing better insight into the cellular metabolism of *G. alkanivorans.*

Engineering of *G. alkanivorans* for enhanced desulfurization: The desulfurizing strains of *G. alkanivorans* offer the advantage of exhibiting activity for a broader range of refractory sulfur compounds present in fossil fuels. However, their application for commercial level biodesulfurization is limited by the low activities that the wild type gordoniae exhibit. The *in silico* genome scale model can be analyzed using different optimization procedures to determine the genetic manipulations that can help in enhancing the flux through the desulfurization pathways. In addition, it has been observed that the substituted sulfur heterocycles are desulfurized at lower rates with lower preference than the non-substituted ones due to the steric hinderances. The protein engineering strategies can be used to modify the active sites of the desulfurizing enzymes in *G. alkanivorans* so as to improve their activity for bulkier sulfur heterocycles.

Experiments to study strain specific parameters of *G. alkanivorans*: There are only limited experimental studies available for desulfurization by gordoniae strains. The

measurements on the biomass composition, maintenance energy, kinetic parameters for consumption of various substrates, etc. for *G. alkanivorans* will help in improving the features and predictability of the existing models and development of improved models.

APPENDICES

Appendices are available in the attached CD-ROM. The list includes the following.

Appendix A: *In silico* model for the sulfur metabolism in *R. erythropolis*

Appendix B: Genome Scale Model for *R. erythropolis*

Appendix C: Genome Scale Model for *G. alkanivorans*

Appendix D: Expanded and Improved Genome Scale Model for *R. erythropolis*

LIST OF PUBLICATIONS

- Aggarwal, S., I. A. Karimi, Lee, D-Y (2011). "Flux-based analysis of sulfur metabolism in desulfurizing strains of *Rhodococcus erythropolis*." *FEMS Microbiology Letters* 315(2): 115-121
- Aggarwal, S., I. A. Karimi, Lee, D-Y (2011). "Reconstruction of a Genome-Scale Metabolic Network of *Rhodococcus erythropolis* for desulfurization studies." *Molecular Biosystems* 7(11):3122-31
- Aggarwal S, I A Karimi, J J Kilbane, Lee, D-Y. "Roles of Sulfite Oxidoreductase and Sulfite Reductase in Improving Desulfurization by *Rhodococcus erythropolis.*" *Molecular Biosystems* 8: 2724-2732.
- Aggarwal S, I A Karimi, Ivan G. "*In Silico* Modeling and Evaluation of *Gordonia alkanivorans* for Biodesulfurization." Manuscript submitted to *Molecular Biosystems*
- Aggarwal S, I A Karimi. "Expanded Genome Scale Model of *R. erythropolis* for Application as a Biocatalyst for Refining of Fossil Fuels." Manuscript in preparation

REFERENCES

1. Song, C., *An overview of new approaches to deep desulfurization for ultra-clean gasoline, diesel fuel and jet fuel.* Catalysis Today, 2003. **86**(1-4): p. 211-263.
2. Soleimani, M., A. Bassi, and A. Margaritis, *Biodesulfurization of refractory organic sulfur compounds in fossil fuels.* Biotechnology Advances, 2007. **25**(6): p. 570-596.
3. Kilbane II, J.J. and S. Le Borgne, *Petroleum biorefining: The selective removal of sulfur, nitrogen, and metals.* Studies in Surface Science and Catalysis, 2004. **151**: p. 29-65.
4. *BP Statistical Review of World Energy*

June 2012.

5. Kilbane II, J.J., *Microbial biocatalyst developments to upgrade fossil fuels.* Current Opinion in Biotechnology, 2006. **17**(3): p. 305-314.
6. Monticello, D.J., *Biodesulfurization and the upgrading of petroleum distillates.* Current Opinion in Biotechnology, 2000. **11**(6): p. 540-546.
7. Gray, K.A., et al., *Molecular mechanisms of biocatalytic desulfurization of fossil fuels.* Nature Biotechnology, 1996. **14**(13): p. 1705-1708.
8. Larkin, M.J., et al., *Applied aspects of Rhodococcus genetics.* Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 1998. **74**(1-3): p. 133-153.
9. Arenskötter, M., D. Bröker, and A. Steinbüchel, *Biology of the metabolically diverse genus Gordonia.* Applied and Environmental Microbiology, 2004. **70**(6): p. 3195-3204.
10. Kilbane II, J.J. and K. Jackowski, *Biodesulfurization of water-soluble coal-derived material by Rhodococcus Rhodochrous IGTS8.* Biotechnology and Bioengineering, 1992. **40**(9): p. 1107-1114.
11. Izumi, Y., et al., *Selective desulfurization of dibenzothiophene by Rhodococcus erythropolis D-1.* Applied and Environmental Microbiology, 1994. **60**(1): p. 223-226.
12. Kayser, K.J., et al., *Utilization of organosulphur compounds by axenic and mixed cultures of Rhodococcus rhodochrous IGTS8.* Journal of General Microbiology, 1993. **139**(12): p. 3123-3129.
13. Kirkwood, K.M., et al., *Sulfur from benzothiophene and alkylbenzothiophenes supports growth of Rhodococcus sp. strain JVH1.* Biodegradation, 2007. **18**(5): p. 541-549.
14. Ohshiro, T., Y. Hine, and Y. Izumi, *Enzymatic desulfurization of dibenzothiophene by a cell-free system of Rhodococcus erythropolis D-1.* FEMS Microbiology Letters, 1994. **118**(3): p. 341-344.
15. Ohshiro, T., T. Hirata, and Y. Izumi, *Desulfurization of dibenzothiophene derivatives by whole cells of Rhodococcus erythropolis H-2.* FEMS Microbiology Letters, 1996. **142**(1): p. 65-70.
16. Omori, T., et al., *Desulfurization of alkyl and aromatic sulfides and sulfonates by dibenzothiophene-desulfurizing Rhodococcus sp. strain SY1.* Bioscience, Biotechnology and Biochemistry, 1995. **59**(7): p. 1195-1198.

17. Onaka, T., et al., *Selective cleavage of the two C-S bonds in asymmetrically alkylated dibenzothiophenes by Rhodococcus erythropolis KA2-5-1.* Journal of Bioscience and Bioengineering, 2001. **92**(1): p. 80-82.
18. Yan, H., et al., *Increase in desulfurization activity of Rhodococcus erythropolis KA2-5-1 using ethanol feeding.* Journal of Bioscience and Bioengineering, 2000. **89**(4): p. 361-366.
19. Yan, H., et al., *Effects of nicotinamide and riboflavin on the biodesulfurization activity of dibenzothiophene by Rhodococcus erythropolis USTB-03.* Journal of Environmental Sciences, 2008. **20**(5): p. 613-618.
20. Del Olmo, C.H., et al., *Modeling the production of a Rhodococcus erythropolis IGTS8 biocatalyst for DBT biodesulfurization: Influence of media composition.* Enzyme and Microbial Technology, 2005. **37**(2): p. 157-166.
21. Kirkwood, K.M., J.M. Foght, and M.R. Gray, *Selectivity among organic sulfur compounds in one- and two-liquid-phase cultures of Rhodococcus sp. strain JVH1*. Biodegradation, 2007. **18**(4): p. 473-480.
22. Patel, S.B., J.J. Kilbane II, and D.A. Webster, *Biodesulphurisation of dibenzothiophene in hydrophobic media by Rhodococcus sp. strain IGTS8.* Journal of Chemical Technology and Biotechnology, 1997. **69**(1): p. 100-106.
23. Olmo, C.H.D., et al., *Production of a Rhodococcus erythropolis IGTS8 biocatalyst for DBT biodesulfurization: Influence of operational conditions.* Biochemical Engineering Journal, 2005. **22**(3): p. 229-237.
24. Honda, H., et al., *High cell density culture of rhodococcus rhodochrous by pH-stat feeding and dibenzothiophene degradation.* Journal of Fermentation and Bioengineering, 1998. **85**(3): p. 334-338.
25. Hirasawa, K., et al., *Improvement of Desulfurization Activity in Rhodococcus erythropolis KA2-5-1 by Genetic Engineering* Biosci. Biotechnol. Biochem., 2001. **65**(2): p. 8.
26. Franchi, E., et al., *Vector development, isolation of new promoters and enhancement of the catalytic activity of the Dsz enzyme complex in Rhodococcus sp. strains.* Oil Gas Sci. Technol., 2003. **58**(4): p. 515-520.
27. Hirasawa, K., et al., *Improvement of desulfurization activity in Rhodococcus erythropolis KA2-5-1 by genetic engineering.* Bioscience, Biotechnology and Biochemistry, 2001. **65**(2): p. 239-246.
28. Kaufman, E.N., et al., *Sulfur specificity in the bench-scale biological desulfurization of crude oil by Rhodococcus IGTS8.* Journal of Chemical Technology and Biotechnology, 1999. **74**(10): p. 1000-1004.
29. Yoshikawa, O., et al., *Enhancement and stabilization of desulfurization activity of Rhodococcus erythropolis KA2-5-1 by feeding ethanol and sulfur components.* Journal of Bioscience and Bioengineering, 2002. **94**(5): p. 447-452.
30. Yu, B., et al., *Microbial desulfurization of gasoline by free whole-cells of Rhodococcus erythropolis XP*. FEMS Microbiology Letters, 2006. **258**(2): p. 284-289.
31. Alves, L., et al., *Sequencing, cloning and expression of the dsz genes required for dibenzothiophene sulfone desulfurization from Gordonia alkanivorans strain 1B.* Enzyme and Microbial Technology, 2007. **40**(6): p. 1598-1603.
32. Kilbane II, J.J. and J. Robbins, *Characterization of the dszABC genes of Gordonia amicalis F.5.25.8 and identification of conserved protein and DNA sequences.* Applied Microbiology and Biotechnology, 2007. **75**(4): p. 843-851.

33. Li, G.Q., et al., *Improved biodesulfurization of hydrodesulfurized diesel oil using Rhodococcus erythropolis and Gordonia sp.* Biotechnology Letters, 2008. **30**(10): p. 1759-1764.
34. Santos, S.C.C., et al., *Characterization of Gordonia sp. strain F.5.25.8 capable of dibenzothiophene desulfurization and carbazole utilization.* Applied Microbiology and Biotechnology, 2006. **71**(3): p. 355-362.
35. Alves, L., et al., *Desulfurization of dibenzothiophene, benzothiophene, and other thiophene analogs by a newly isolated bacterium, Gordonia alkanivorans strain 1B.* Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology, 2005. **120**(3): p. 199-208.
36. Shavandi, M., et al., *Biodesulfurization of dibenzothiophene by recombinant Gordonia alkanivorans RIPI90A.* Bioresource Technology, 2008. **100**(1): p. 475-479.
37. Price, N.D., et al., *Genome-scale microbial in silico models: The constraints-based approach.* Trends in Biotechnology, 2003. **21**(4): p. 162-169.
38. Sang, Y.L., D.Y. Lee, and Y.K. Tae, *Systems biotechnology for strain improvement.* Trends in Biotechnology, 2005. **23**(7): p. 349-358.
39. Tang, W.L. and H. Zhao, *Industrial biotechnology: Tools and applications.* Biotechnology Journal, 2009. **4**(12): p. 1725-1739.
40. Covert, M.W., et al., *Metabolic modeling of microbial strains in silico.* Trends in Biochemical Sciences, 2001. **26**(3): p. 179-186.
41. Fong, S.S. and B.Ø. Palsson, *Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes.* Nature Genetics, 2004. **36**(10): p. 1056-1058.
42. Chin, J.W., et al., *Analysis of NADPH supply during xylitol production by engineered escherichia coli.* Biotechnology and Bioengineering, 2009. **102**(1): p. 209-220.
43. Saha, R., P.F. Suthers, and C.D. Maranas, *Zea mays irs1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism.* PLoS ONE, 2011. **6**(7).
44. Oberhardt, M.A., B.Ø. Palsson, and J.A. Papin, *Applications of genome-scale metabolic reconstructions.* Molecular Systems Biology, 2009. **5**.
45. Suthers, P.F., et al., *Genome-scale metabolic reconstruction Of mycoplasma genitalium, iPS189.* PLoS Computational Biology, 2009. **5**(2).
46. Zomorrodi, A.R. and C.D. Maranas, *Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data.* BMC Systems Biology, 2010. **4**.
47. Burgard, A.P. and C.D. Maranas, *Optimization-based framework for inferring and testing hypothesized metabolic objective functions.* Biotechnology and Bioengineering, 2003. **82**(6): p. 670-677.
48. Pharkya, P., A.P. Burgard, and C.D. Maranas, *OptStrain: A computational framework for redesign of microbial production systems.* Genome Research, 2004. **14**(11): p. 2367-2376.
49. Satish Kumar, V., M.S. Dasika, and C.D. Maranas, *Optimization based automated curation of metabolic reconstructions.* BMC Bioinformatics, 2007. **8**.
50. Li, M.Z., et al., *Genetic analysis of the dsz promoter and associated regulatory regions of Rhodococcus erythropolis IGTS8.* Journal of Bacteriology, 1996. **178**(22): p. 6409-6418.

51. Tanaka, Y., et al., *The cbs mutant strain of Rhodococcus erythropolis KA2-5-1 expresses high levels of Dsz enzymes in the presence of sulfate.* Archives of Microbiology, 2002. **178**(5): p. 351-357.
52. http://www.epa.gov/region7/air/quality/health.htm. Available from: http://www.epa.gov/region7/air/quality/health.htm.
53. http://www.epa.gov/air/caa/peg/acidrain.html. Available from: http://www.epa.gov/air/caa/peg/acidrain.html.
54. König, A., et al., *Current tasks and challenges for exhaust aftertreatment research. A viewpoint from the automotive industry.* Topics in Catalysis, 2001. **16-17**(1-4): p. 23-31.
55. EPA, *Regulatory Announcement - Clean Air Nonroad Diesel Rule.* May 2004. **EPA420-F-04-032**: p. www.epa.gov/nonroad-diesel.
56. EPA, *Rules and Regulations - Control of Emissions of Air Pollution From Nonroad Diesel Engines and Fuel.* Federal Register, June 2004. **69**(124).
57. Song, C., *Fuel processing for low-temperature and high-temperature fuel cells: Challenges, and opportunities for sustainable development in the 21st century.* Catalysis Today, 2002. **77**(1-2): p. 17-49.
58. Shekhawat, D., et al., *Fuel constituent effects on fuel reforming properties for fuel cell applications.* Fuel, 2009. **88**(5): p. 817-825.
59. Latz, J., et al., *Hydrodesulfurization of jet fuel by pre-saturated one-liquid-flow technology for mobile fuel cell applications.* Chemical Engineering Science, 2009. **64**(2): p. 288-293.
60. Labana, S., G. Pandey, and R.K. Jain, *Desulphurization of dibenzothiophene and diesel oils by bacteria.* Letters in Applied Microbiology, 2005. **40**(3): p. 159-163.
61. Song, C. and X. Ma, *New design approaches to ultra-clean diesel fuels by deep desulfurization and deep dearomatization.* Applied Catalysis B: Environmental, 2003. **41**(1-2): p. 207-238.
62. Lewis, R.E., *Safety and Integrity of Marine Fuel Pumps Operating on 0.1% Sulfur Marine Gas Oil.* http://www.infomarine.gr/attachments/047_Safety%20and%20Integrity%20of%20Marine%20Fuel%20Pumps.pdf, 2009.
63. Alcalde, M., et al., *Environmental biocatalysis: from remediation with enzymes to novel green processes.* Trends in Biotechnology, 2006. **24**(6): p. 281-287.
64. Bull, A.T., A.W. Bunch, and G.K. Robinson, *Biocatalysts for clean industrial products and processes.* Current Opinion in Microbiology, 1999. **2**(3): p. 246-251.
65. Schmid, A., et al., *Industrial biocatalysis today and tomorrow.* Nature, 2001. **409**(6817): p. 258-268.
66. Sutherland, T.D., et al., *Enzymatic bioremediation: From enzyme discovery to applications.* Clinical and Experimental Pharmacology and Physiology, 2004. **31**(11): p. 817-821.
67. Ahuja, S.K., G.M. Ferreira, and A.R. Moreira, *Utilization of enzymes for environmental applications.* Critical Reviews in Biotechnology, 2004. **24**(2-3): p. 125-154.
68. Linguist, L. and M. Pacheco, *Enzyme-based diesel desulfurization process offers energy, CO2 advantages.* Oil and Gas Journal, 1999. **97**(8).
69. Payzant, J.D., D.S. Montgomery, and O.P. Strausz, *Sulfides in petroleum.* Organic Geochemistry, 1986. **9**(6): p. 357-369.

70. Bhadra, A., J.M. Scharer, and M. Moo-Young, *Microbial desulphurization of heavy oils and bitumen.* Biotechnology Advances, 1987. **5**(1): p. 1-27.
71. Choudhary, T.V., et al., *Towards clean fuels: Molecular-level sulfur reactivity in heavy oils.* Angewandte Chemie - International Edition, 2006. **45**(20): p. 3299-3303.
72. Kodama, K., K. Umehara, and K. Shimizu, *Identification of microbial products from dibenzothiophene and its proposed oxidation pathway.* Agricultural and Biological Chemistry, 1973. **37**(1): p. 45-50.
73. Yamada, K., et al., *Microbial Conversion of Petro-sulfur Compounds Part I. Isolation and Identification of Dibenzothiophene-Utilizing Bacteria** Agricultural and Biological Chemistry, 1968. **32**(7): p. 6.
74. Monticello, D.J., D. Bakker, and W.R. Finnerty, *Plasmid-mediated degradation of dibenzothiophene by Pseudomonas species*, in *Applied and Environmental Microbiology*1985. p. 756-760.
75. Mormile, M.R. and R.M. Atlas, *Mineralization of the dibenzothiophene biodegradation products 3-hydroxy-2-formyl benzothiophene and dibenzothiophene sulfone.* Applied and Environmental Microbiology, 1988. **54**(12): p. 3183-3184.
76. Kilbane, J.J., *Desulfurization of coal: the microbial solution.* Trends in Biotechnology, 1989. **7**(4): p. 97-101.
77. Dahlberg, M.D., et al., *Biodesulfurization of dibenzothiophene sulfone by Arthrobacter sp. and studies with oxidized Illinois No. 6 coal.* Fuel, 1993. **72**(12): p. 1645-1649.
78. Van Afferden, M., et al., *Degradation of dibenzothiophene by Brevibacterium sp.DO.* Archives of Microbiology, 1990. **153**(4): p. 324-328.
79. Armstrong, S.M., B.M. Sankey, and G. Voordouw, *Evaluation of sulfate-reducing bacteria for desulfurizing bitumen or its fractions.* Fuel, 1997. **76**(3): p. 223-227.
80. Crawford, D.L. and R.K. Gupta, *Oxidation of dibenzothiophene by Cunninghamella elegans.* Current Microbiology, 1990. **21**(4): p. 229-231.
81. Bezalel, L., et al., *Initial oxidation products in the metabolism of pyrene, anthracene, fluorene, and dibenzothiophene by the white rot fungus Pleurotus ostreatus.* Applied and Environmental Microbiology, 1996. **62**(7): p. 2554-2559.
82. Gürtler, V., B.C. Mayall, and R. Seviour, *Can whole genome analysis refine the taxonomy of the genus Rhodococcus?* FEMS Microbiology Reviews, 2004. **28**(3): p. 377-403.
83. Gallagher, J.R., E.S. Olson, and D.C. Stanley, *Microbial desulfurization of dibenzothiophene: A sulfur-specific pathway.* FEMS Microbiology Letters, 1993. **107**(1): p. 31-36.
84. Denome, S.A., E.S. Olson, and K.D. Young, *Identification and cloning of genes involved in specific desulfurization of dibenzothiophene by Rhodococcus sp. strain IGTS8.* Applied and Environmental Microbiology, 1993. **59**(9): p. 2837-2843.
85. Denome, S.A., et al., *Characterization of the desulfurization genes from Rhodococcus sp. strain IGTS8.* Journal of Bacteriology, 1994. **176**(21): p. 6707-6716.
86. Oldfield, C., et al., *Elucidation of the metabolic pathway for dibenzothiophene desulphurization by Rhodococcus sp. strain IGTS8 (ATCC 53968).* Microbiology, 1997. **143**(9): p. 2961-2973.

87. Denis-Larose, C., et al., *Conservation of plasmid-encoded dibenzothiophene desulfurization genes in several rhodococci.* Applied and Environmental Microbiology, 1997. **63**(7): p. 2915-2919.
88. Wolf, B.P., et al., *Characterization of proteins utilized in the desulfurization of petroleum products by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.* Analytical Biochemistry, 1998. **260**(2): p. 117-127.
89. Ohshiro, T., et al., *Purification and characterization of dibenzothiophene (DBT) sulfone monooxygenase, an enzyme involved in DBT desulfurization, from Rhodococcus erythropolis D-1.* Journal of Bioscience and Bioengineering, 1999. **88**(6): p. 610-616.
90. Piddington, C.S., B.R. Kovacevich, and J. Rambosek, *Sequence and molecular characterization of a DNA region encoding the dibenzothiophene desulfurization operon of Rhodococcus sp. strain IGTS8.* Applied and Environmental Microbiology, 1995. **61**(2): p. 468-475.
91. Ohshiro, T. and Y. Izumi, *Purification, characterization and crystallization of enzymes for dibenzothiophene desulfurization.* Bioseparation, 2000. **9**(3): p. 185-188.
92. Gray, K.A., G.T. Mrachkott, and C.H. Squires, *Biodesulfurization of fossil fuels.* Current Opinion in Microbiology, 2003. **6**(3): p. 229-235.
93. Kamali, N., et al., *Site-directed mutagenesis enhances the activity of NADH-FMN oxidoreductase (DszD) activity of Rhodococcus erythropolis.* Biotechnology Letters, 2010. **32**(7): p. 921-927.
94. Ohshiro, T., et al., *Flavin reductase coupling with two monooxygenases involved in dibenzothiophene desulfurization: Purification and characterization from a non-desulfurizing bacterium, Paenibacillus polymyxa A-1.* Applied Microbiology and Biotechnology, 2002. **59**(6): p. 649-657.
95. Wang, P. and S. Krawiec, *Kinetic analyses of desulfurization of dibenzothiophene by Rhodococcus erythropolis in batch and fed-batch cultures.* Applied and Environmental Microbiology, 1996. **62**(5): p. 1670-1675.
96. Kertesz, M.A., *Riding the sulfur cycle - Metabolism of sulfonates and sulfate esters in Gram-negative bacteria.* FEMS Microbiology Reviews, 2000. **24**(2): p. 135-175.
97. Noda, K.I., K. Watanabe, and K. Maruhashi, *Cloning of a rhodococcal promoter using a transposon for dibenzothiophene biodesulfurization.* Biotechnology Letters, 2002. **24**(22): p. 1875-1882.
98. Wang, P., A.E. Humphrey, and S. Krawiec, *Kinetic analyses of desulfurization of dibenzothiophene by Rhodococcus erythropolis in continuous cultures.* Applied and Environmental Microbiology, 1996. **62**(8): p. 3066-3068.
99. Davoodi-Dehaghani, F., M. Vosoughi, and A.A. Ziaee, *Biodesulfurization of dibenzothiophene by a newly isolated Rhodococcus erythropolis strain.* Bioresource Technology, 2010. **101**(3): p. 1102-1105.
100. Ohshiro, T., et al., *Characterization of dibenzothiophene desulfurization reaction by whole cells of Rhodococcus erythropolis H-2 in the presence of hydrocarbon.* Journal of Fermentation and Bioengineering, 1996. **82**(6): p. 610-612.
101. Ohshiro, T., T. Hirata, and Y. Izumi, *Microbial desulfurization of dibenzothiophene in the presence of hydrocarbon.* Applied Microbiology and Biotechnology, 1995. **44**(1-2): p. 249-252.

102. Matsui, T., et al., *Alkylated Benzothiophene Desulfurization by Rhodococcus sp. Strain T09.* Bioscience, Biotechnology and Biochemistry, 2000. **64**(3): p. 596-599.
103. Folsom, B.R., et al., *Microbial desulfurization of alkylated dibenzothiophenes from a hydrodesulfurized middle distillate by Rhodococcus erythropolis I-19.* Applied and Environmental Microbiology, 1999. **65**(11): p. 4967-4972.
104. McFarland, B.L., *Biodesulfurization.* Current Opinion in Microbiology, 1999. **2**(3): p. 257-264.
105. Feng, J., et al., *The surfactant tween 80 enhances biodesulfurization.* Applied and Environmental Microbiology, 2006. **72**(11): p. 7390-7393.
106. Borole, A.P., et al., *Comparison of the emulsion characteristics of Rhodococcus erythropolis and Ecsherichia coli SOXC-5 cells expressing Biodesulfurization genes.* Biotechnology Progress, 2002. **18**(1): p. 88-93.
107. Setti, L., G. Lanzarini, and P.G. Pifferi, *Whole cell biocatalysis for an oil desulfurization process.* Fuel Processing Technology, 1997. **52**(1-3): p. 145-153.
108. Kawaguchi, H., H. Kobayashi, and K. Sato, *Metabolic engineering of hydrophobic Rhodococcus opacus for biodesulfurization in oil-water biphasic reaction mixtures.* Journal of Bioscience and Bioengineering, 2012. **113**(3): p. 360-366.
109. Naito, M., et al., *Long-term repeated biodesulfurization by immobilized Rhodococcus erythropolis KA2-5-1 cells.* Applied Microbiology and Biotechnology, 2001. **55**(3): p. 374-378.
110. Ansari, F., et al., *DBT Degradation Enhancement by Decorating Rhodococcus erythropolis IGST8 With Magnetic Fe3O4 Nanoparticles.* Biotechnology and Bioengineering, 2009. **102**(5): p. 1505-1512.
111. Kirimura, K., et al., *Identification and functional analysis of the genes encoding dibenzothiophene-desulfurizing enzymes from thermophilic bacteria.* Applied Microbiology and Biotechnology, 2004. **65**(6): p. 703-713.
112. Wang, Z., et al., *Enhanced biodesulfurization by expression of dibenzothiophene uptake genes in Rhodococcus erythropolis.* World Journal of Microbiology and Biotechnology, 2011. **27**(9): p. 1965-1970.
113. Watanabe, K., K.I. Noda, and K. Maruhashi, *Enhanced desulfurization in a transposon-mutant strain of Rhodococcus erythropolis.* Biotechnology Letters, 2003. **25**(16): p. 1299-1304.
114. Matsui, T., et al., *Microbial desulfurization of alkylated dibenzothiophene and alkylated benzothiophene by recombinant Rhodococcus sp. strain T09.* Applied Microbiology and Biotechnology, 2001. **56**(1-2): p. 196-200.
115. Matsui, T., et al., *Optimization of the copy number of dibenzothiophene desulfurizing genes to increase the desulfurization activity of recombinant Rhodococcus sp.* Biotechnology Letters, 2001. **23**(20): p. 1715-1718.
116. Matsui, T., et al., *Recombinant Rhodococcus sp. strain T09 can desulfurize DBT in the presence of inorganic sulfate.* Current Microbiology, 2002. **45**(4): p. 240-244.
117. Li, G.Q., et al., *Improvement of dibenzothiophene desulfurization activity by removing the gene overlap in the dsz operon.* Bioscience, Biotechnology and Biochemistry, 2007. **71**(4): p. 849-854.
118. Li, G.Q., et al., *Genetic rearrangement strategy for optimizing the dibenzothiophene biodesulfurization pathway in Rhodococcus erythropolis.* Applied and Environmental Microbiology, 2008. **74**(4): p. 971-976.

119. Gilbert, S.C., et al., *Isolation of a unique benzothiophene-desulphurizing bacterium, Gordona sp. strain 213E (NCIMB 40816), and characterization of the desulphurization pathway.* Microbiology, 1998. **144**(9): p. 2545-2553.
120. Kim, S.B., et al., *Gordonia desulfuricans sp. nov., a benzothiophene-desulphurizing actinomycete.* International Journal of Systematic Bacteriology, 1999. **49**(4): p. 1845-1851.
121. Kim, S.B., et al., *Gordonia amicalis sp. nov., a novel dibenzothiophene-desulphurizing actinomycete.* International Journal of Systematic and Evolutionary Microbiology, 2000. **50**(6): p. 2031-2036.
122. Li, G.Q., et al., *Isolation and characterization of a benzothiophene-desulfurizing bacterium.* Huanjing Kexue/Environmental Science, 2006. **27**(6): p. 1181-1185.
123. Matsui, T., et al., *Benzo[b]thiophene desulfurization by Gordonia rubropertinctus strain T08.* Applied Microbiology and Biotechnology, 2001. **57**(1-2): p. 212-215.
124. Mohebali, G., et al., *Biodesulfurization potential of a newly isolated bacterium, Gordonia alkanivorans RIPI90A.* Enzyme and Microbial Technology, 2007. **40**(4): p. 578-584.
125. Rhee, S.K., et al., *Desulfurization of dibenzothiophene and diesel oils by a newly isolated Gordona strain, CYKS1.* Applied and Environmental Microbiology, 1998. **64**(6): p. 2327-2331.
126. Li, W., et al., *Biodesulfurization of dibenzothiophene by growing cells of Gordonia sp. in batch cultures.* Biotechnology Letters, 2006. **28**(15): p. 1175-1179.
127. Santos, S.C.C., et al., *Characterization of Gordonia sp. strain F.5.25.8 capable of dibenzothiophene desulfurization and carbazole utilization.* Applied Microbiology and Biotechnology, 2006. **71**(3): p. 355-362.
128. Aminsefat, A., B. Rasekh, and M.R. Ardakani, *Biodesulfurization of dibenzothiophene by Gordonia sp. AHV-01 and optimization by using of response surface design procedure.* Microbiology, 2012. **81**(2): p. 154-159.
129. Alves, L., et al., *Sequencing, cloning and expression of the dsz genes required for dibenzothiophene sulfone desulfurization from Gordonia alkanivorans strain 1B.* Enzyme and Microbial Technology, 2007. **40**(6): p. 1598-1603.
130. Raheb, J., et al., *A novel indigenous gordonia RIPI species identified with a reduction in energy consuming in the desulfurization process.* Energy Sources, Part A: Recovery, Utilization and Environmental Effects, 2011. **33**(23): p. 2125-2131.
131. Shavandi, M., et al., *Genomic structure and promoter analysis of the dsz operon for dibenzothiophene biodesulfurization from Gordonia alkanivorans RIPI90A.* Applied Microbiology and Biotechnology, 2010. **87**(4): p. 1455-1461.
132. Kim, Y.J., et al., *A physiological study on growth and dibenzothiophene (DBT) desulfurization characteristics of Gordonia sp. CYKS1.* Korean Journal of Chemical Engineering, 2004. **21**(2): p. 436-441.
133. Alves, L., et al., *Evidence for the role of zinc on the performance of dibenzothiophene desulfurization by Gordonia alkanivorans strain 1B.* Journal of Industrial Microbiology and Biotechnology, 2008. **35**(1): p. 69-73.
134. Alves, L., et al., *Dibenzothiophene desulfurization by Gordonia alkanivorans strain 1B using recycled paper sludge hydrolyzate.* Chemosphere, 2008. **70**(6): p. 967-973.

135. Mohebali, G., et al., *Dimethyl sulfoxide (DMSO) as the sulfur source for the production of desulfurizing resting cells of Gordonia alkanivorans RIPI90A.* Microbiology, 2008. **154**(3): p. 878-885.
136. Je Hwan, C., et al., *Production of a desulfurization biocatalyst by two-stage fermentation and its application for the treatment of model and diesel oils.* Biotechnology Progress, 2001. **17**(5): p. 876-880.
137. Chang, J.H., et al., *Desulfurization of light gas oil in immobilized-cell systems of Gordona sp. CYKS1 and Nocardia sp. CYKS2.* FEMS Microbiology Letters, 2000. **182**(2): p. 309-312.
138. Alves, L. and S.M. Paixão, *Toxicity evaluation of 2-hydroxybiphenyl and other compounds involved in studies of fossil fuels biodesulphurisation.* Bioresource Technology, 2011. **102**(19): p. 9162-9166.
139. Irani, Z.A., et al., *Analysis of petroleum biodesulfurization in an airlift bioreactor using response surface methodology.* Bioresource Technology, 2011. **102**(22): p. 10585-10591.
140. Mohebali, G., et al., *Stabilization of water/gas oil emulsions by desulfurizing cells of Gordonia alkanivorans RIPI90A.* Microbiology, 2007. **153**(5): p. 1573-1581.
141. Ok, K.C., et al., *Enhancement of phase separation by the addition of de-emulsifiers to three-phase (diesel oil/biocatalyst/aqueous phase) emulsion in diesel biodesulfurization.* Biotechnology Letters, 2003. **25**(1): p. 73-77.
142. Raman, K. and N. Chandra, *Flux balance analysis of biological systems: Applications and challenges.* Briefings in Bioinformatics, 2009. **10**(4): p. 435-449.
143. Mazzocchi, F., *Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory.* EMBO Reports, 2008. **9**(1): p. 10-14.
144. Wolkenhauer, O., H. Kitano, and K.H. Cho, *Systems biology.* IEEE Control Systems Magazine, 2003. **23**(4): p. 38-48.
145. Schmid, A. and L.M. Blank, *Systems biology: Hypothesis-driven omics integration.* Nature Chemical Biology, 2010. **6**(7): p. 485-487.
146. Dhar, P.K. and R. Weiss, *Enabling the new biology of the 21st century.* Systems and Synthetic Biology, 2007. **1**(1): p. 1-2.
147. Klipp, E. and W. Liebermeister, *Mathematical modeling of intracellular signaling pathways.* BMC Neuroscience, 2006. **7**(SUPPL. 1).
148. Stelling, J., *Mathematical models in microbial systems biology.* Current Opinion in Microbiology, 2004. **7**(5): p. 513-518.
149. Varner, J. and D. Ramkrishna, *Mathematical models of metabolic pathways.* Current Opinion in Biotechnology, 1999. **10**(2): p. 146-150.
150. Heijnen, J.J., *Approximative kinetic formats used in metabolic network modeling.* Biotechnology and Bioengineering, 2005. **91**(5): p. 534-545.
151. Teusink, B., et al., *Can yeast glycolysis be understood terms of vitro kinetics of the constituent enzymes? Testing biochemistry.* European Journal of Biochemistry, 2000. **267**(17): p. 5313-5329.
152. Wright, B.E. and P.J. Kelly, *Kinetic models of metabolism in intact cells, tissues, and organisms.* Current topics in cellular regulation, 1981. **19**: p. 103-158.
153. Theobald, U., et al., *In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: I. Experimental observations.* Biotechnology and Bioengineering, 1997. **55**(2): p. 305-316.

154. Schilling, C.H., J.S. Edwards, and B.O. Palsson, *Toward metabolic phenomics: Analysis of genomic data using flux balances.* Biotechnology Progress, 1999. **15**(3): p. 288-295.
155. Lee, J.M., E.P. Gianchandani, and J.A. Papin, *Flux balance analysis in the era of metabolomics.* Briefings in Bioinformatics, 2006. **7**(2): p. 140-150.
156. Varma, A. and B.O. Palsson, *Metabolic flux balancing: Basic concepts, scientific and practical use.* Bio/Technology, 1994. **12**(10): p. 994-998.
157. Feist, A.M. and B.O. Palsson, *The biomass objective function.* Current Opinion in Microbiology, 2010. **13**(3): p. 344-349.
158. Schuetz, R., L. Kuepfer, and U. Sauer, *Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli.* Molecular Systems Biology, 2007. **3**.
159. Durot, M., P.Y. Bourguignon, and V. Schachter, *Genome-scale models of bacterial metabolism: Reconstruction and applications.* FEMS Microbiology Reviews, 2009. **33**(1): p. 164-190.
160. Reed, J.L. and B.Ã. Palsson, *Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: Assessment of correlated reaction subsets that comprise network states.* Genome Research, 2004. **14**(9): p. 1797-1805.
161. Beste, D.J.V., et al., *GSMN-TB: A web-based genome-scale network model of Mycobacterium tuberculosis metabolism.* Genome Biology, 2007. **8**(5).
162. Jamshidi, N. and B.Ø. Palsson, *Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets.* BMC Systems Biology, 2007. **1**.
163. Davoodi-Dehaghani, F., M. Vosoughi, and A.A. Ziaee, *Biodesulfurization of dibenzothiophene by a newly isolated Rhodococcus erythropolis strain.* Bioresource Technology.
164. Johnson, S., et al. *Biodesulfurization of petroleum by Rhodococcus erythropolis*. 1994.
165. Emanuele, J.J. and K.A. Gray, *Physical and kinetic properties of an nadh:fmn oxidoreductase isolated from rhodococcus erythropolis strain igts8.* FASEB Journal, 1997. **11**(9).
166. Orth, J.D., I. Thiele, and B.Ø. Palsson, *What is flux balance analysis?* Nature Biotechnology, 2010. **28**(3): p. 245-248.
167. Thiele, I. and B.Ø. Palsson, *A protocol for generating a high-quality genome-scale metabolic reconstruction.* Nature Protocols, 2010. **5**(1): p. 93-121.
168. Gonzalez, O., et al., *Reconstruction, modeling & analysis of Halobacterium salinarum R-1 metabolism.* Molecular BioSystems, 2008. **4**(2): p. 148-159.
169. Park, J.M., T.Y. Kim, and S.Y. Lee, *Constraints-based genome-scale metabolic simulation for systems metabolic engineering.* Biotechnology Advances, 2009. **27**(6): p. 979-988.
170. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
171. Oldfield, C., et al., *Desulphurisation of benzothiophene and dibenzothiophene by actinomycete organisms belonging to the genus Rhodococcus, and related taxa.* Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 1998. **74**(1-3): p. 119-132.
172. Caspi, R., et al., *The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases.* Nucleic Acids Research, 2008. **36**(SUPPL. 1): p. D623-D631.

173. Rawat, M. and Y. Av-Gay, *Mycothiol-dependent proteins in actinomycetes.* FEMS Microbiology Reviews, 2007. **31**(3): p. 278-292.
174. Lee, D.Y., et al., *MetaFluxNet: The management of metabolic reaction information and quantitative metabolic flux analysis.* Bioinformatics, 2003. **19**(16): p. 2144-2146.
175. Van Der Geize, R. and L. Dijkhuizen, *Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications.* Current Opinion in Microbiology, 2004. **7**(3): p. 255-261.
176. Feist, A.M., et al., *A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.* Molecular Systems Biology, 2007. **3**.
177. Reed, J.L., et al., *An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).* Genome Biology, 2003. **4**(9).
178. Nogales, J., B.Ø. Palsson, and I. Thiele, *A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory.* BMC Systems Biology, 2008. **2**.
179. Puchałka, J., et al., *Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology.* PLoS Computational Biology, 2008. **4**(10).
180. Sohn, S.B., et al., *In silico genome-scale metabolic analysis of Pseudomonas putida KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival.* Biotechnology Journal, 2010. **5**(7): p. 739-750.
181. Oberhardt, M.A., et al., *Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1.* Journal of Bacteriology, 2008. **190**(8): p. 2790-2803.
182. Duarte, N.C., M.J. Herrgård, and B.Ø. Palsson, *Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model.* Genome Research, 2004. **14**(7): p. 1298-1309.
183. Famili, I., et al., *Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(23): p. 13134-13139.
184. Förster, J., et al., *Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network.* Genome Research, 2003. **13**(2): p. 244-253.
185. Chung, B.K.S., et al., *Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement.* Microbial Cell Factories, 2010. **9**.
186. Sohn, S.B., et al., *Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production.* Biotechnology Journal, 2010. **5**(7): p. 705-715.
187. Lee, K.Y., et al., *The genome-scale metabolic network analysis of Zymomonas mobilis ZM4 explains physiological features and suggests ethanol and succinic acid production strategies.* Microbial Cell Factories, 2010. **9**.
188. Widiastuti, H., et al., *Genome-scale modeling and in silico analysis of ethanologenic bacteria Zymomonas mobilis.* Biotechnology and Bioengineering, 2011. **108**(3): p. 655-665.
189. Aggarwal, S., I.A. Karimi, and D.Y. Lee, *Flux-based analysis of sulfur metabolism in desulfurizing strains of Rhodococcus erythropolis.* FEMS Microbiology Letters, 2011. **315**(2): p. 115-121.

190. Sekine, M., et al., *Sequence analysis of three plasmids harboured in Rhodococcus erythropolis strain PR4.* Environmental Microbiology, 2006. **8**(2): p. 334-346.
191. Pramanik, J., et al., *Development and validation of a flux-based stoichiometric model for enhanced biological phosphorus removal metabolism.* Water Research, 1999. **33**(2): p. 462-476.
192. Schilling, C.H., et al., *Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems.* Biotechnology and Bioengineering, 2000. **71**(4): p. 286-306.
193. Bonarius, H.P.J., G. Schmid, and J. Tramper, *Flux analysis of underdetermined metabolic networks: The quest for the missing constraints.* Trends in Biotechnology, 1997. **15**(8): p. 308-314.
194. Brooke, A., et al., GAMS Development Corporation, 1998.
195. Thiele, I., et al., *Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): An in silico genome-scale characterization of single- and double-deletion mutants.* Journal of Bacteriology, 2005. **187**(16): p. 5818-5830.
196. Bornemann, C., et al., *Biosynthesis of mycothiol: Elucidation of the sequence of steps in Mycobacterium smegmatis.* Biochemical Journal, 1997. **325**(3): p. 623-629.
197. Finnerty, W.R., *The biology and genetics of the genus Rhodococcus.* Annual Review of Microbiology, 1992. **46**: p. 26.
198. Daffe, M., M. McNeil, and P.J. Brennan, *Major structural features of the cell wall arabinogalactans of Mycobacterium, Rhodococcus, and Nocardia spp.* Carbohydrate Research, 1993. **249**(2): p. 383-398.
199. Fischer, E. and U. Sauer, *Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism.* Nature Genetics, 2005. **37**(6): p. 636-640.
200. Carnicer, M., et al., *Macromolecular and elemental composition analysis and extracellular metabolite balances of Pichia pastoris growing at different oxygen levels.* Microbial Cell Factories, 2009. **8**.
201. Konishi, M., et al., *Effect of sulfur sources on specific desulfurization activity of Rhodococcus erythropolis KA2-5-1 in exponential fed-batch culture.* Journal of Bioscience and Bioengineering, 2005. **99**(3): p. 259-263.
202. Yu, B., et al., *Selective biodegradation of S and N heterocycles by a recombinant Rhodococcus erythropolis strain containing carbazole dioxygenase.* Applied and Environmental Microbiology, 2006. **72**(3): p. 2235-2238.
203. Omori, T., et al., *Desulfurization of dibenzothiophene by Corynebacterium sp. strain SY1.* Applied and Environmental Microbiology, 1992. **58**(3): p. 911-915.
204. Aggarwal, S., I.A. Karimi, and D.Y. Lee, *Reconstruction of a genome-scale metabolic network of Rhodococcus erythropolis for desulfurization studies.* Molecular BioSystems, 2011. **7**(11): p. 3122-3131.
205. Chang, I.S., B.H. Kim, and P.K. Shin, *Use of sulfite and hydrogen peroxide to control bacterial contamination in ethanol fermentation.* Applied and Environmental Microbiology, 1997. **63**(1): p. 1-6.
206. Kappler, U. and C. Dahl, *Enzymology and molecular biology of prokaryotic sulfite oxidation.* FEMS Microbiology Letters, 2001. **203**(1): p. 1-9.
207. Suzuki, I., *Sulfite: Cytochrome c oxidoreductase of thiobacilli.* Methods in Enzymology, 1994. **243**: p. 447-454.

208. Frederiksen, T.M. and K. Finster, *Sulfite-oxido-reductase is involved in the oxidation of sulfite in Desulfocapsa sulfoexigens during disproportionation of thiosulfate and elemental sulfur.* Biodegradation, 2003. **14**(3): p. 189-198.
209. Van Der Ploeg, J.R., M. Barone, and T. Leisinger, *Functional analysis of the Bacillus subtilis cysK and cysJI genes.* FEMS Microbiology Letters, 2001. **201**(1): p. 29-35.
210. Mahadevan, R. and C.H. Schilling, *The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.* Metabolic Engineering, 2003. **5**(4): p. 264-276.
211. Chung, B.K.S. and D.Y. Lee, *Flux-sum analysis: A metabolite-centric approach for understanding the metabolic network.* BMC Systems Biology, 2009. **3**.
212. Aziz, R.K., et al., *The RAST Server: Rapid annotations using subsystems technology.* BMC Genomics, 2008. **9**.
213. Iida, S., et al., *Gordonia otitidis sp. nov., isolated from a patient with external otitis.* International Journal of Systematic and Evolutionary Microbiology, 2005. **55**(5): p. 1871-1876.
214. Kjeldsen, K.R. and J. Nielsen, *In silico genome-scale reconstruction and validation of the corynebacterium glutamicum metabolic network.* Biotechnology and Bioengineering, 2009. **102**(2): p. 583-597.
215. Shinfuku, Y., et al., *Development and experimental verification of a genome-scale metabolic model for Corynebacterium glutamicum.* Microbial Cell Factories, 2009. **8**.
216. Gao, J., L.B.M. Ellis, and L.P. Wackett, *The University of Minnesota Biocatalysis/Biodegradation Database: Improving public access.* Nucleic Acids Research, 2010. **38**(SUPPL.1): p. D488-D491.
217. De Carvalho, C.C.C.R. and M.M.R. Da Fonseca, *The remarkable Rhodococcus erythropolis.* Applied Microbiology and Biotechnology, 2005. **67**(6): p. 715-726.
218. Aggarwal, S., I.A. Karimi, and D.Y. Lee, *Reconstruction of a Genome-Scale Metabolic Network of Rhodococcus erythropolis for desulfurization studies.* Mol. BioSyst., 2011: p. 10.1039/C1MB05201B.
219. Overbeek, R., et al., *The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.* Nucleic Acids Research, 2005. **33**(17): p. 5691-5702.
220. Mavrovouniotis, M.L., *Estimation of standard Gibbs energy changes of biotransformations.* Journal of Biological Chemistry, 1991. **266**(22): p. 14440-14445.
221. (NCBI), N.C.f.B.I., http://www.ncbi.nlm.nih.gov/sites/genome.
222. Sardessai, Y. and S. Bhosle, *Tolerance of bacteria to organic solvents.* Research in Microbiology, 2002. **153**(5): p. 263-268.
223. De Carvalho, C.C.C.R., L.Y. Wick, and H.J. Heipieper, *Cell wall adaptations of planktonic and biofilm Rhodococcus erythropolis cells to growth on C5 to C16 n-alkane hydrocarbons.* Applied Microbiology and Biotechnology, 2009. **82**(2): p. 311-320.
224. Li, Y.G., et al., *Biodegradation of carbazole in oil/water biphasic system by a newly isolated bacterium Klebsiella sp. LSSE-H2.* Biochemical Engineering Journal, 2008. **41**(2): p. 166-170.
225. Li, S.S., et al., *Microbial denitrogenation of fuel oil.* Wei sheng wu xue bao = Acta microbiologica Sinica, 2006. **46**(6): p. 1023-1027.

226. Benedik, M.J., et al., *Microbial denitrogenation of fossil fuels.* Trends in Biotechnology, 1998. **16**(9): p. 390-395.
227. Santos, V.E., et al., *Oxygen uptake rate measurements both by the dynamic method and during the process growth of Rhodococcus erythropolis IGTS8: Modelling and difference in results.* Biochemical Engineering Journal, 2006. **32**(3): p. 198-204.
228. Kayser, K.J., et al., *Isolation and characterization of a moderate thermophile, Mycobacterium phlei GTIS10, capable of dibenzothiophene desulfurization.* Applied Microbiology and Biotechnology, 2002. **59**(6): p. 737-745.
229. Li, F., et al., *Biodesulfurization of dibenzothiophene by a newly isolated bacterium mycobacterium sp. X7B.* Journal of Chemical Engineering of Japan, 2003. **36**(10): p. 1174-1177.
230. Okada, H., et al., *Cultivation of a desulfurizing bacterium, Mycobacterium strain G3.* Biotechnology Letters, 2001. **23**(24): p. 2047-2050.
231. Maghsoudi, S., et al., *Selective desulfurization of dibenzothiophene by newly isolated Corynebacterium sp. strain P32C1.* Biochemical Engineering Journal, 2000. **5**(1): p. 11-16.
232. Okada, H., et al., *Analyses of microbial desulfurization reaction of alkylated dibenzothiophenes dissolved in oil phase.* Biotechnology and Bioengineering, 2003. **83**(4): p. 489-497.
233. Lee, S.K., et al., *Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels.* Current Opinion in Biotechnology, 2008. **19**(6): p. 556-563.
234. Hanly, T.J. and M.A. Henson, *Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures.* Biotechnology and Bioengineering, 2011. **108**(2): p. 376-385.
235. Hanly, T.J., M. Urello, and M.A. Henson, *Dynamic flux balance modeling of S. cerevisiae and E. coli co-cultures for efficient consumption of glucose/xylose mixtures.* Applied microbiology and biotechnology, 2012. **93**(6): p. 2529-2541.
236. Mahadevan, R., J.S. Edwards, and F.J. Doyle Iii, *Dynamic Flux Balance Analysis of diauxic growth in Escherichia coli.* Biophysical Journal, 2002. **83**(3): p. 1331-1340.
237. Kobayashi, M., et al., *Kinetic analysis of microbial desulfurization of model and light gas oils containing multiple alkyl dibenzothiophenes.* Bioscience, Biotechnology and Biochemistry, 2001. **65**(2): p. 298-304.
238. Taffs, R., et al., *In Silico approaches to study mass and energy flows in microbial consortia: A syntrophic case study.* BMC Systems Biology, 2009. **3**.