# Duquesne University Duquesne Scholarship Collection

**Electronic Theses and Dissertations** 

Summer 2003

## Nitrogen metabolism in iron-respiring bacteria

Ching (Tony) Lung Hsieh

Follow this and additional works at: https://dsc.duq.edu/etd

#### **Recommended** Citation

Hsieh, C. (2003). Nitrogen metabolism in iron-respiring bacteria (Master's thesis, Duquesne University). Retrieved from https://dsc.duq.edu/etd/667

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact phillipsg@duq.edu.

## A thesis presented to

## The Department of Biological Sciences

## Bayer School of Natural and Environmental Sciences

**Duquesne University** 

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

2003

## Nitrogen metabolism in iron-respiring bacteria

by

Ching-Lung (Tony) Hsieh

Thesis Advisor: John F. Stolz, Ph.D.

Thesis Committee: Partha Basu, Ph.D. Jana Patton-Vogt, Ph.D.

© 2003 Ching-Lung Hsieh

#### ACKNOWLEDGEMENTS

I would like to thank Dr. Stolz for his guidance, patience and support and for allowing me to be part of his lab during the two years at Duquesne University. I would also like to thank Dr. Basu and Dr. Patton-Vogt for being on my thesis committee. I thank my fellow lab mates Joy, Eddie, Miru and ex-postdoc Eman for their help, especially Joy who kindly guided and helped me on many occasions. I would like to acknowledge Dr. Trun who kindly answered my hundreds of questions, Dr. Dameron from the Department of Chemistry and Biochemistry for kindly giving me tips on my research and Dr. Young who gave me a mini lecture on how to construct phylogenetic trees. Lastly, I would like to thank my family for their support and trust.

#### ABSTRACT

Nitrogen metabolism was investigated in three species of iron respiring bacteria *Geobacter metallireducens*, *Geobacter sulfurreducens*, and *Sulfurospirillum barnesii*. *G. metallireducens* and *G. sulfurreducens* were shown to grow in medium containing N<sub>2</sub>, NO<sub>3</sub><sup>-</sup>, or NH<sub>4</sub> but nitrite inhibited growth. BLAST analysis of the draft genomic sequence of *G. metallireducens* revealed only Nar, Nrf, and Nif homologues but no assimilatory genes (i.e., Nas). Two *nar* operons were identified, *narCGHJIK*<sub>1</sub>*K*<sub>2</sub>, and *narCGHI*. A putative CAP/CRP recognition binding site was found upstream and two molybdenum biosynthetic genes were found downstream of the large operon. Both operons contain a novel cytochrome *c* (NarC) that has only been described in *Thermus thermophilus* strain HB8. An attempt was made to purify the periplasmic nitrate reductase (Nap) from *S. barnesii* without success. However, a method using lysozyme to solubilize Nap and a rapid activity assay for nitrate reductase using 96 well plates were developed.

## TABLE OF CONTENTS

INTRODUCTION	1
NITROGEN CYCLE AND THE IMPORTANCE OF NITRATE AND NITRITE	1
NITRATE AND NITRITE REDUCTASE	4
MOLYBDENUM COFACTOR CONTAINING ENZYMES THAT ARE INVOLVED IN NITROGEN METABO	DLISM6
CYTOPLASMIC ASSIMILATORY NITRATE REDUCTASE	
MEMBRANE BOUND CYTOPLASMIC DISSIMILATORY NITRATE REDUCTASE	
PERIPLASMIC NITRATE REDUCTASE	15
NITRITE REDUCTASE	19
PHYLOGENETIC DIVERSITY OF IRON REDUCING BACTERIA AND NITROGEN METABOLISM	21
GEOBACTER METALLIREDUCENS AND GEOBACTER SULFURREDUCENS	
Sulfurospirillum barnesii	
OBJECTIVE	
EXPERIMENTAL APPROACH	
A. Collection of enzyme sequences and BLAST search the draft genome of $G$ .	
METALLIREDUCENS	27
B. PHYLOGENETIC ANALYSIS OF NAR GENE PRODUCTS	
C. Cell culture	
D. SOLUBILIZATION OF G. METALLIREDUCENS NARG	29
E. Lysozyme treatment	
F. SOLUBILIZATION OF S. BARNESII NAP	
G. PROTEIN CONCENTRATION DETERMINATION	
H. Electrophoresis	

1. SDS-PAGE	
2.Non denaturing (native) gel electrophoresis	31
I. ACTIVITY ASSAY FOR NITRATE AND NITRITE REDUCTASE ACTIVITY	
1. Zymogram	
2. Spectroscopic photometric assay for nitrate and nitrite reductase activity	
3. 96 wells methyl viologen assay	
J. HEME STAIN	
K. DIALYSIS	
L.COLUMN CHROMATOGRAPHY	
1. G. metallireducens	
2. S. barnesii	34
RESULTS	
Division octored cited a cutedization of a schule atody with a tody with the $\lambda$ is denoted by $C$	
PHYSIOLOGICAL CHARACTERIZATION OF ASSIMILATORY NITRATE/NITRITE REDUCTASE IN G.	25
METALLIREDUCENS AND G. SULFURREDUCENS	
MOLYBDOENZYMES IN G. METALLIREDUCENS	
ENZYMES INVOLVED IN NITROGEN METABOLISM IN G. METALLIREDUCENS	
A. Membrane bound nitrate reductase (Nar) homologues	40
B. Putative nitrite reductases	41
C. Structural genes of nitrogenase	
STRUCTURE OF NAR OPERON IN G. METALLIREDUCENS	
Comparative analysis of the two putative Nar subunits and their similarities to $T$ .	
THERMOPHILUS NAR SUBUNITS.	51
Phylogenetic trees of NarG, H, J, I, K <sub>1</sub> , and K <sub>2</sub>	
Additional analysis of <i>narG</i> gene products and putative nitrite reductases	57
Location of the putative enzymes	57
Theoretical pI, and molecular weight	57
PARTIAL PURIFICATION OF NARG FROM G. METALLIREDUCENS	
SOLUBILIZATION OF NAPA FROM S. BARNESII USING LYSOZYME TREATMENT	60

PURIFICATION OF NAPA FROM S. BARNESII	61
Development of 96 well assay	62
DISCUSSION	65
Physiological characterization of assimilatory nitrate and nitrite reduction in $G$ .	
METALLIREDUCENS AND G. SULFURREDUCENS	65
MOLYBDOENZYMES IN G. METALLIREDUCENS	65
PUTATIVE NAR OPERONS AND THEIR EVOLUTIONARY PROCESS	66
UPSTREAM AND DOWNSTREAM OF <i>NAR121</i> AND <i>NAR135</i>	68
SOLUBILIZATION OF NAPA UTILIZING LYSOZYME TREATMENT AND 96 WELLS ASSAY DEVELOPMENT	69
CONCLUSION	70
APPENDIX I	72
APPENDIX II	75
APPENDIX III	76
APPENDIX IV	77
MINERAL SALT MEDIA	77
APPENDIX V	78
MOLYBDOENZYMES BLAST RESULTS	78
NAR121, 135 OPERON BLAST RESULTS	80
APPENDIX VI	83
TRANSMEMBRANE PREDICTION FOR NARI IN <i>NAR121</i>	83
TRANSMEMBRANE PREDICTION FOR NARI IN NAR135	85
TRANSMEMBRANE PREDICTION FOR NARK <sub>1</sub> in <i>NAR135</i>	87
TRANSMEMBRANE PREDICTION FOR NARK <sub>2</sub> in <i>NAR135</i>	90

APPENDIX VII	94
PROTEIN SEQUENCE ALINGMENT OF NARC FROM G. METALLIREDUCENS AND T. THERMOPHILUS	94
PROTEIN SEQUENCE ALINGMENT OF NARG FROM G. METALLIREDUCENS AND T. THERMOPHILUS	94
PROTEIN SEQUENCE ALINGMENT OF NARH FROM G. METALLIREDUCENS AND T. THERMOPHILUS	96
PROTEIN SEQUENCE ALINGMENT OF NARI FROM G. METALLIREDUCENS AND T. THERMOPHILUS	96
REFERENCES	98

## LIST OF FIGURES

Number	Page
FIGURE 1: NITROGEN CYCLE (67)	3
FIGURE 2: MOLYBDENUM COFACTOR	5
FIGURE 3: SCHEMATIC DRAWING OF PROPOSED E. COLI MEMBRANE BOUND CYTOPLASMIC NITRATE	
REDUCTASE MODEL (52)	15
FIGURE 4: THE NAP OPERONS (51)	
FIGURE 5: PHYLOGENETIC TREE INFERRED FROM 16S RNA SEQUENCES (39)	23
FIGURE 6: PHYLOGENETIC TREE OF MOLYBDOENZYMES USING THE NEIGHBOR JOINING METHOD	39
FIGURE 7: GENE ORDER OF NAR OPERON IN G. METALLIREDUCENS	
FIGURE 8: NUCLEOTIDE SEQUENCE OF NAR OPERON IN CONTIG121	47
FIGURE 9: NUCLEOTIDE SEQUENCE OF <i>NAR</i> OPERON IN CONTIG135	51
FIGURE 10: PHYLOGENETIC TREES OF NARG, H, I, J, K <sub>1</sub> , AND K <sub>2</sub> .	
FIGURE 11: HEME STAIN AND SDS-PAGE OF G. METALLIREDUCENS ION EXCHANGE FRACTIONS	
FIGURE 12: ZYMOGRAM OF NITRITE REDUCTASE ACTIVITY	59
FIGURE 13: ZYMOGRAM OF NITRITE REDUCTASE ACTIVITY	62
FIGURE 14: SDS-PAGE OF S. BARNESII SIZE EXCLUSION COLUMN FRACTIONS	62
FIGURE 15: 96 WELL ASSAY	63

## LIST OF TABLES

Number	Page
TABLE 1: SUMMARY OF NITRATE REDUCTASE	5
TABLE 2: SUMMARY OF MOLYBDOENZYMES (31, 34)	
TABLE 3: OVERVIEW OF THE TWO COMPONENT NITRATE RESPONSE SENSOR REGULATOR SYSTEM AND	)
FUNCTIONING DETAILS OF THE RESPONSE SENSOR REGULATOR SYSTEM COMPONENTS (46)	14
TABLE 4: SUMMARY OF NITRITE REDUCTASE	
TABLE 5: LYSOZYME TREATMENT	
TABLE 6: PHYSIOLOGICAL CHARACTERIZATION OF G. METALLIREDUCENS AND G. SULFURREDUCENS	
TABLE 7: PUTATIVE MOLYBDOENZYMES IN G. METALLIREDUCENS	
TABLE 8: SUMMARY OF ALL PUTATIVE GENES INVOLVED IN NITROGEN METABOLISM IN $G$ . METALLIRE	EDUCENS
TABLE 9: SIMILARITIES BETWEEN THE PUTATIVE NARC, G, H AND I IN CONTIG121 AND 135 FROM G	S-15
AND THE CORRESPONDING NAR SUBUNITS FROM <i>T. THERMOPHILIS</i>	
TABLE 10: SUMMARY TABLE FOR LYSOZYME TREATED S. BARNESII FRACTIONS' NITRATE REDUCTASE	:
ACTIVITY	60
TABLE 11: SUMMARY OF 96 WELL ASSAY SENSITIVITY	

#### Introduction

#### Nitrogen cycle and the importance of nitrate and nitrite

Nitrogen is the most abundant gas in the atmosphere (approximately 80 %). Nitrogen can also exist in different oxidation states in the environment (Figure 1). Interconversion between these oxidation states is carried out by a variety of microorganisms and is known as the Nitrogen cycle (Figure 1). Nitrogen fixation converts gaseous dinitrogen to ammonia for assimilation. Microbial oxidation of ammonia provides energy for microbes and produces nitrite and nitrate. Denitrification (a type of dissimilatory nitrate reduction) is a bacterial respiratory process that reduces nitrate to nitrite, nitric oxide, then nitrous oxide, and finally dinitrogen gas under oxygen limiting conditions (75). Dissimilatory reduction of nitrate to nitrite then to ammonia occurs under anaerobic conditions, in which nitrate acts as the terminal electron acceptor. Assimilatory reduction of nitrate to ammonia on the other hand, can occur aerobically as well as anaerobically, and is done by plants, fungi, and algae as well as bacteria (75). Anaerobic ammonia oxidation (ANAMOX) is a relatively newly discovered process in which nitrite and ammonia combine to produce dinitrogen gas (10).

The balance of nitrogen in different oxidation states in the environment is a major concern for both humans and nature. For instance, excess agricultural nitrogen containing fertilizer may cause contamination of rivers and lakes by leaching from soil. Polluted freshwater systems and domestic wastewater are known to contain high concentrations of nitrite (80). Nitrate is a more mobile form of nitrogen compared to ammonia (or ammonium). It is because nitrate ions are not adsorbed by the negatively charged colloids

that dominate most soil (35). Thus the reduction of nitrate plays a central role in preventing leaching and polluting the groundwater (52).

Nitrate reduction also takes place in the human large intestine (49). Nitrate can have an inhibitory effect on methane production in the colon and may also have significant influence on human colonic fermentation (3). High dietary intake of nitrate may contribute to infantile methemoglobinemia and gastric/intestinal cancer through the production of carcinogenic N-nitrosoamine compounds (22). Food borne pathogens such as *Escherichia coli, Salmonella sp.* and *Proteus sp.* can utilize nitrate under anaerobic conditions. Thus even if meat is packaged anaerobically there still is a possibility that such pathogens can survive and enter the human body. Other pathogens such as *Neisseria gonococcus* (the cause of the sexualy transmitted disease gonorrhea) and *Neisseria meningitidis* (the cause of meningitis) are also able to survive anaerobic sites in human body by reducing nitrite to nitric oxide (54). Therefore studying how nitrate and nitrite are utilized by microbes is not only environmental significant but also has benefits for human health.



Figure 1: Nitrogen cycle (67)

#### Nitrate and nitrite reductase

Denitrification, dissimilatory reduction of nitrate, and assimilatory reduction of nitrate are three processes of nitrate reduction that are known to exist in organisms (Fig 1). Four different classes of nitrate reductase have been found in plants, algae, fungi, archaea and bacteria (10, 62, 76). Cytoplasmic assimilatory nitrate reductases are known to exist in eukaryotes and prokaryotes. Membrane bound and periplasmic nitrate reductases are solely found in prokaryotes (10). All four types of nitrate reductase contain a molybdopterin cofactor that has a molybdenum atom bound to it (Figure 2). In bacterial nitrate reductase, the molybdopterin is further modified by the addition of a nucleotide (guanosine) and is termed molybdopterin guanine dinucleotide (MGD) (Figure 2). A molybdenum free nitrate reductase has been isolated from *Pseudomonas isachenkovii*, a vanadate reducing bacterium (4). Molybdopterin containing nitrate reductases can be classified based on X-ray structural data, spectroscopic data, biochemical data, and sequence alignment (64). Eukaryotic nitrate reductases belong to the sulfite oxidase family whereas the other three prokaryotic dissimilatory reductases belong to Dimethyl Sulfoxide (DMSO) reductase family (64). Classification of these four nitrate reductases is summarized in Table 1.

### **Table 1: Summary of nitrate reductase**

	Eukaryotic nitrate reductase	Prokaryotic nitrate reductase		
Organisms	Plants, algae, fungi		Bacteria, Archaea	
Molybdenum	Molybdopterin	Molybdopterin guanine dinucleotide		
Molybdoenzyme family	Sulfite oxidase	Dimethyl Sulfoxide reductase		
Location of the nitrate reductase	Cytoplasmic	Cytoplasmic	Membrane bound	Periplasmic
	* *			•
Assimilatory/Dissimilatory	Assimilatory	Assimilatory	Dissimilatory	Dissimilatory
Abbreviation	INR, NiR, Nit	Nas	Nar	Nap

## Figure 2: Molybdenum cofactor

a) Molybdenum cofactor b) Molybdenum guanine dinucleotide (MGD) (67)



#### Molybdenum cofactor containing enzymes that are involved in nitrogen metabolism

Molybdenum cofactor containing enzymes are involved in nitrogen, sulfur and carbon cycles (34). These enzymes can be divided into four families, namely DMSO reductase family, Xanthine oxidase family, Sulfite oxidase family, and Aldehyde ferredoxin oxidoreductase family (Table 2) (31, 34).

Nitrate reductases and formate dehydrogenase are molybdoenzymes that are relevant to the nitrogen metabolism (xanthine oxidase is involved in uric acid production that is a part of purine and pyrimidine nucleotide degradation and will not be discussed here). Although nitrogenase is not classified as a molybdoenzyme, since it contains molybdenum and also is involved in nitrogen metabolism, it will be discussed here. Eukaryotic assimilatory nitrate reductase belongs to the sulfite oxidase family, whereas the other three prokaryotic dissmilatory nitrate reductases belong to DMSO reductase family.

Nitrogenase consists of two proteins. One is a homodimer Fe protein that contains an iron sulfur cluster (FeS) and is regarded as a transmitter of electrons from a donor protein, such as flavodoxin, to the second tetramer protein that contains molybdenum and iron (MoFe) (27). The structural genes of nitrogenase are *nifHDK*, which encode the homodimer Fe protein (*nifH*) and the tetramer MoFe (*nifDK*). Regulation of nitrogenase has been extensively studied in both phototrophic bacteria such as *Rhodobacter capsulatus* and enteric bacteria such as *Klebsiella sp.* (18). The regulatory cascade of nitrogenase is somewhat confusing due to the various nomenclatures used for the same regulatory system in different microorganisms. The regulatory cascade can be divided into three different levels, in the order of the Gln system (sensory transduction

components), Ntr (or NifR) system (general nitrogen regulation), and nifA-L (or nifAI, nifAII) system (18, 21). Nitrogenases with vanadium or iron replacing molybdenum are also known to exist (23, 27). Nitrogenase is known to be sensitive to oxygen and reduction of acetylene is a standard assay to test for nitrogenase activity. Thus, discovery of a novel nitrogenase that is oxygen tolerant and also utilizes even more reactive superoxide as its source of electron was remarkable (27, 61). Also interestingly, this nitrogenase does not reduce acetylene.

Formate dehydrogenase is known to be associated with nitrate reductase in *E. coli*. When *E. coli* is grown on media containing formate as electron donor and nitrate as electron acceptor, electrons are transported from formate to nitrate via formate dehydrogenase complex and nitrate reductase complex (33). Three different formate dehydrogenases are expressed in *E. coli*. Nitrate induced formate dehydrogenase (FdhN) is coupled to NarG that is a membrane bound nitrate reductase only expressed under anaerobic condition. FdhO is an isoenzyme of FdhN and is coupled to NarZ that is also a membrane bound nitrate reductase. Both FdhO and NarZ are expressed constitutively regardless of the presence of oxygen (1). FdhH is a membrane associated formate dehydrogenase involved in fermentation and is strongly repressed by nitrate (5, 18). Thus FdhH is not involved in nitrate reduction.

	DMSO and and any formily. Variables and the formily			Aldehyde ferredoxin
Enzyme family	DMSO reductase family	Xanthine oxidase family	Sumte oxidase family	oxidoreducase family
	DMSO reductase	Xanthine oxidase/dehydrogenase	Sulfite oxidase	Aldehyde ferredoxin
				oxidoreducase
	Trimethylamine-N-oxide	Aldehyde oxidase	Nitrate reductase ( Eukaryotic	Formaldehyde ferredoxin
	reductase		assimilatory)	oxidoreductase
	Biotin sulfoxide reducase	Aldehyde oxidoreductase		Glyceraldehyde-3-phosphate
				ferredoxin oxidoreductase
	Nitrate reductase (Assimilatory)	CO dehydrogenase		Carboxylic acid reductase
	Nas			
	Nitrate reductase	Quinoline-2-oxidoreductase		Aldehyde hydrogenase
	(dissmililatory) Nar			
	Nitrate reductase	Isoquinoline-1-oxidoreductase		Hydroxycarboxylate viologen
E	(dissmililatory) Nap			oxidoreductase
n	Formate dehydrogenase	Quinoline-4-carboxylate-2-		
z		oxidoreductase		
У	Polysulfide reductase	Quinoline-4-oxidoreductase		
m				
e	Arsenite oxidase	Quinaldic acid 4-oxidoreductase		
	Formylmethanofuran	Nicotinic acid		
	dehydrogenase	hydrogenase/dehydrogenase		
		6-hydroxynictinate hydroxylase		
		nicotine dehydrogenase		
		Picolinate hydroxylase		
		2(R)-hydroxycarboxylate		

 Table 2: Summary of molybdoenzymes (31, 34)
 Image: Comparison of the second second

#### Cytoplasmic assimilatory nitrate reductase

Eukaryotic nitrate reductase is either a dimer (plants) or tetramer (alga, *Chlorella*) (67). Each monomer contains the molybdopterin cofactor, cytochrome *b*, and FAD (67). It is thought that electrons are transferred in the order of NAD(P)H to FAD, cytochrome *b* then to molybdopterin cofactor (the active site) (67).

Prokaryotic assimilatory nitrate reductases (Nas) can be divided into two classes

(46). One is ferredoxin or flavodoxin dependent such as the one that is found in

cyanobacteria and the other is NADH dependent found in Klebsiella and Rhodobacter

(46). Another group of scientists, however, divided prokaryotic assimilatory nitrate

reductases into three classes with Klebsiella pnuemoniae (enterobacteria), Bacillus

subtilis (Gram positive bacteria) and Synechococcus (Cyanobacteria) as representative for each class (10, 62). Amino acid sequence comparison of the enzymes from these three representative bacteria indicates that the enzymes contain MGD (Figure 2) and an Nterminal iron sulfur cluster-binding site (10). Structurally, prokaryotic Nas can exist as monomeric or dimeric form (46, 67). Cyanobacteria and Azotobacter (nitrogen fixing bacteria) have a monomeric subunit whereas Klebsiella pnuemoniae and Rhodobacter capsulatus have heterodimers. The three representative Nas from K. pnuemoniae, B. subtilis, Synechococcus have been reviewed recently by Richardson et al. (62). The molybdenum binding subunit in each microorganism is termed NasA, NasC, and NarB respectively. Siroheme assimilatory nitrite reductase is also found in these organisms and the gene is located within the Nas cluster. They are termed *nasB* in *K*. *pnuemoniae*, *nasD* in B. subtilis, and nirA in Synechococcus. It is important to note that siroheme nitrite reductase can also be associated with dissimilatory nitrate reduction. In E. coli, a siroheme nitrite reductase termed NirB is associated with membrane bound dissimilatory nitrate reductase and is not involved in nitrogen assimilation (18).

Regulation of *nas* shares the same upstream regulatory systems that are observed in nitrogenase regulation. The general nitrogen regulatory Ntr system and a positive regulator (NasR) that acts by a transcription attenuation system are the two regulatory systems found in *Klebsiella* (36, 46). In *R. capsulatus, nas* is induced by nitrate and repressed at a low C/N ratio (i.e. 2-oxoglutarate/glutamine ratio), which is the same upstream regulatory system that regulates nitrogenase in this microorganism (48). Nitrogen regulation in cyanobacteria is regulated by the NtcA protein, a member of the Crp/Fnr family of transcription activators (36, 46). A second regulatory gene *ntcB* is

located immediate downstream of *ntcA* in *Synechococcus*, and its gene product is a LysR family transcription factor required to activate *nirA* operon in response to nitrite (2, 46). The Ntr system and a two component regulatory system NasS/T are responsible for Nas expression in *Azotobacter vinelandii* (46). In gram positive bacteria, *B. subtilis* has no analogue of Ntr system but rather has a positive regulator called TnrA, which is thought to bind DNA and activate transcription of nitrogen assimilation operons (36, 46). Recently, Nas has been purified from *Haloferax mediterranei*, a member of Archaea (45). This was the first Nas to be isolated from Archaea and is a heterodimer comprised of 105 and 50 KDa subunits.

As an additional note, all nitrate reductases are capable of using methyl viologen as an artificial electron donor. However, the ability to use bromophenol blue as an artificial electron donor is a characteristic of both eukaryotic and prokaryotic assimilatory enzymes (12, 30, 46).

#### Membrane bound cytoplasmic dissimilatory nitrate reductase

Membrane bound nitrate reductase (Nar) has a wide distribution and has been characterized from bacteria such as *E. coli* (52), *Bradyrhizobium japonicum* (28), and many others. Three subunits  $\alpha$  (approximately 104-150 KDa),  $\beta$  (43-63 KDa), and  $\gamma$  (19-28 KDa) constitute the enzyme (Fig 3). These three subunits are also known as NarG, NarH, and NarI respectively. The  $\alpha$  subunit contains molybdopterin and a high spin iron sulfur cluster. The  $\beta$  subunit contains four iron sulfur clusters and the  $\gamma$  subunit is a diheme cytochrome *b*. Membrane bound nitrate reductase is primarily involved in anaerobic respiration. This process is called dissimilatory reduction of nitrate. According

to the current model, the  $\gamma$  subunit accepts electrons from the quinol pool and passes them to the  $\beta$  subunit via its b hemes. Two protons are released to the periplasm, which generate transmembrane electrochemical potential. This proton gradient is utilized to generate ATP. After the electrons are passed from the  $\gamma$  subunit to the  $\beta$  subunit, they are then transferred from the  $\beta$  subunit's iron clusters to the molybdopterin in the  $\alpha$  subunit. Based on the amino acid sequence of the  $\alpha$  subunit, its N-terminal segment contains a His-X3-Cys-X3-Cys-Xn-Cys motif that bind a 4Fe-4S cluster (63). Some microorganisms have multiple membrane bound cytoplasmic nitrate reductase. Two isoenzymes are found in *E. coli*. One is only expressed under anaerobic growth conditions with nitrate (NarG) and the other is expressed constitutively (NarZ). In other words, NarZ is not under the influence of either oxygen or nitrate. Structurally, NarGHI and NarZYV are highly similar with 76 % identity for the catalytic subunit NarG and NarZ, 75 % identity for the  $\beta$  subunit (NarH and NarY), and 87 % identity for the NarI and NarY (11, 46). Genes encoding NarG comprises *narGHJI*, in which *narGHI* encode the enzyme and *narJ* may encode a chaperon that is involved in the assembly or stability of NarGH complex prior to its membrane attachment (11, 46). Upstream of the narGHJI is the *narK* gene encoding a putative nitrogen oxyanion transporter and further upstream are genes encoding the two component nitrate response sensor regulator system NarXL, in which NarX is the nitrate sensor and NarL is the DNA-binding regulator (46, 57, 62). NarZ on the other hand is encoded by the operon *narZYWV*, in which *narZYV* encode the enzyme and *narW* is the chaperon homologue to *narJ*. Upstream of *narZYWV* is *narU* gene that is the homologue to narK. No response sensory system genes are found further upstream of *narU*. However, a second set of the two component nitrate response sensor

regulator system NarQP does exist and is located at a different location on the *E. coli* chromosome (46).

Currently, the function of NarK and NarU as nitrite exporters to prevent accumulation of nitrite that is generated by nitrate reduction in the cell is widely accepted. They are also thought to transport nitrate into the cell. However, there has been no evidence that NarK and NarU are involved in nitrate transport into the cell until recently. A paper published by Clegg et al (16) showed that both NarK and NarU may be involved in nitrate uptake. Another oxyanion transporter, NirC, was also investigated in this paper. NirC is encoded by the *nir* operon (see nitrite section) and was determined to only transport nitrite.

Functions of the two component nitrate response sensor regulator system, NarXL and NarQP are summarized in table 3. It is important to note that the NarXL system is not found in all microorganisms that can reduce nitrate. For instance, denitrifiers such as *Pseudomonas aeruginosa, Pseudomonas stutzeri, Pseudomonas fluorescens*, and *Paracoccus denitrificans* seem to lack this NarXL system (51). Another well studied regulator involved in expression of nitrate metabolism genes is Fnr. It is a transcriptional regulator involved in aerobic-anaerobic regulation by controlling the expression of anaerobic metabolism genes (7). Based on sequence analysis, Fnr is very similar to the catabolite activator protein (CAP) with one notable difference. Fnr has a ferredoxin-like four cysteine cluster, Cys-X3-Cys-X2-Cys-X5-Cys, at the N-terminal (7). Activation and deactivation of Fnr is due to the presence of an iron cluster (4Fe-4S). Under anaerobic conditions, Fnr is a dimer with one 4Fe-4S in each monomer. Upon exposure to oxygen, the 4Fe-4S center disassembles to form a 2Fe-2S cluster within minutes and prolonged

exposure to oxygen causes further disassembly of the 2Fe-2S center to form apoproteins that lack iron. Fnr like factors have been identified in various bacteria (46). For instance, an Fnr-like regulator recently identified in the denitrifying bacteria *Paracoccus pantotrophus* (previously classified as *Thiospaera panotropha*) is termed NarR and is involved in the maximal expression of membrane bound nitrate reductase (83). NNR, found in *P. denitrificans*, which controls nitrite and nitric oxide reduction, is another example of an Fnr-like regulator (78). The gram positive bacterium *B.subtilis* also has a gene encoding a transcription factor of the Fnr family (63). Additionally, two membrane bound cytoplasmic nitrate reductases have also been purified from *B. japonicum*. Based on SDS-PAGE analysis, one consists of four polypeptides of molecular masses 116, 68, and 56 KDa, and the other consists of four polypeptides of molecular masses 116, 68, 59, and 56 KDa (28). There was no cytochrome *b* detected in either enzyme, and resembles *Rhodobacter spaeroides*, *P. stutzeri*, and *Haloferax denitrificans* (28). Whether these isoenzymes function similarly to *E. coli* Nar is unknown.

Table 3: Overview of the two component nitrate response sensor regulator system

and functioning details of the response sensor regulator system components (46)

	Two component nitrate response sensor regulator system			
	NarX and NarQ	NarL and NarP		
Location	NarX: Upstream of <i>narGHJI</i> NarQ: 53 min of the <i>E. coli</i> genetic map	NarL: Up stream of <i>narGHJI</i> NarP: 46min of the <i>E. coli</i> genetic map		
Function	Membrane sensor protein	Cytoplasmic response regulator		
How do they work?	NO <sub>3</sub> <sup>-</sup> or NO <sub>2</sub> <sup>-</sup> binds to a periplasmic domain, altering the confirmation of the protein to allow autophosphorylation and subsequent phosphorylation of NarL and NarP	Activated proteins bind to specific DNA target sites, the so called NarL heptamers		

	NarX NarQ		
Phosphatase activity	Dephosphorylates NarL Dephosphorylates NarP		
Kinase activity	Both NarX and NarQ can phophorylate NarL and NarP. Depends on nitrate/nitrite availability		
In the presence of nitrate	Phosphorylates both NarL and NarP, indistinguishable between NarX and NarQ		
In the presence of nitrite	Phosphorylates NarP but primarily acts as a NarL phosphataseSame as in the press nitrate (I.e. Phospho 		

	NarL	NarP
Operons that it regulates	Periplasmic nitrite reductase nrfABCDEFG	Periplasmic nitrite reductase nrfABCDEFG
	Periplasmic nitrate reductase nap	Periplasmic nitrate reductase nap
	Nitrate reductase narGHJI	
	Fumarate reductase frdABCD	
	Nitrite exporter narK	



Figure 3: Schematic drawing of proposed *E. Coli* membrane bound cytoplasmic nitrate reductase model (52)

#### Periplasmic nitrate reductase

Periplasmic nitrate reductase (Nap) is a heterodimer consisting of a larger subunit termed NapA (63-93 KDa) that contains MGD and a smaller subunit termed NapB (13-19 KDa) which is a diheme cytochrome c (52). Nap is not primarily involved in nitrate assimilation or in anaerobic denitrification. However, a genetic study of *P. pantotrophus* lacking membrane bound nitrate reductase demonstrated denitrification under anaerobic

conditions (9, 10). This indicates that periplasmic nitrate reductase might be involved in anaerobic denitrification directly. E. coli with double mutations that inhibit generation of membrane bound nitrate reductases (E. coli has two membrane bound nitrate reductase enzymes and one periplasmic nitrate reductase) showed that nitrate reduction by Nap is half the total rate of parental wild type strain (55). E. coli mutants that express only one type of nitrate reductase grown in a nitrate scarce environment showed that Nap containing E. coli had a better growth rate than membrane bound nitrate reductase containing strain (54). Thus Nap is an effective enzyme when the microbe is living under low nitrate conditions. Studies on enzyme expression in *P. pantotrophus* and *P.* denitrificans has revealed that Nap is predominantly expressed under aerobic growth condition (62). Two promoters further regulate the Nap gene and are activated under growth on different carbon sources. These studies indicate that Nap maybe involved in redox balancing. The Nap gene cluster from various bacteria shows a variety of order. Based on a recent review by Philippot (51), there are at least 10 different variations in nap gene clusters known to exist (Figure 4). Four genes, napABCD are found in all the bacteria listed and are arranged in the order of *napDABC*, though in *E. coli*, Magnetococcus MC-1, and Magnetospirillum magnetococcus MC-1, napGH are located in between of *napA* and *B* (Figure 4) (51). Genes *napAB* encode for the heterodimer periplasmic nitrate reductase as described earlier. NapC is a membrane anchored tetraheme cytochrome c that is speculated to be involved in transferring electrons from the membranous quinols to NapAB. NapD is thought to be involved in the posttranslational assembly of a functional molybdopolypeptide of NapA (51, 56). Five other

genes, napKEF and napGH, can also be found in different combinations in the nap

clusters (51, 52). The *napG* gene encodes a 20 KDa soluble protein with four putative 4Fe-4S centers, and *napH* encodes a 32 KDa membrane protein that is predicted to bind two 4Fe-4S centers. It has been proposed that a putative NapGH complex could act as a redox sensor controlling the electron flow to NapA (10, 46). A more recent study by Brondijk et al showed that NapGH are essential for electron transfer from ubiquinol to NapAB (13). This study also addressed a possible role for NapF that is a soluble protein of 16 KDa with four putative 4Fe-4S clusters. The results of the study implicates NapF in an energy conserving role coupled to the oxidation of ubiquinol (13). There are three functional quinones in E. coli, namely ubiquinone (UQ), demethylmenaquinone (DMK), and menaquinone (MK) (13). Electrons from UQ and MK, but not DMK, can be used for nitrate reduction in E. coli (13, 77, 81, 82), and NapAB is capable of accepting electrons from either UQH<sub>2</sub> and MKH<sub>2</sub>. NapK and NapE are small transmembrane proteins with unknown function (37, 51). Moreover, the Nap gene cluster is not always found in the genome. In bacteria such as Ralstonia eutropha, Ralstonia. spaeroides, and Pseudomonas sp. G-147, the nap genes are located on endogenous plasmids (51).



Figure 4: The nap operons (51)

#### Nitrite reductase

The product of nitrate reduction, nitrite is a toxic substance. In denitrifying bacteria, nitrite is rapidly converted to nitric oxide and eventually to dinitrogen gas. Whereas in assimilatory or dissimilatory nitrate reduction, nitrite is converted to ammonia.

Two nitrite reductases (NirK and NirS) are known to exist exclusively in denitrifying bacteria (67). The first one is a copper containing NirK, a trimeric enzyme with each subunit containing one type 1 Cu (blue copper) and one type 2 Cu (non-blue copper) (73). The second one (NirS) contains heme, which consists of a small cytochrome c containing subunit and a large cytochrome d1 containing subunit (29, 47, 73). Organization of *nirK* and *nirS* operons has been review recently by Philipott (51). Two other types of nitrite reductases siroheme and cytochrome c nitrite reductase reduce nitrite to ammonia. Siroheme containing nitrite reductases have been found to be involved in both assimilatory and dissimilatory nitrite reduction, whereas cytochrome c nitrite reductases has only been found to carry out dissimilatory nitrite reduction. According to Lin and Stewart (36) siroheme containing assimilatory nitrite reductase can be classified into two types; ferredoxin-dependent, found in cyanobacteria and plants; and NAD(P)H-dependent, found in other bacteria and fungi. NAD(P)H-dependent assimilatory nitrite reductase contains an iron sulfur cluster linked to siroheme and FAD. Electrons flow from NAD(P)H to FAD, the iron cluster, siroheme, and finally nitrite. On the other hand, the ferredoxin-dependent nitrite reductase contains an iron sulfur cluster linked to siroheme but lacks FAD. A siroheme containing nitrite reductase that carries out dissimilatory nitrite reduction, has been described in E. coli. It is a NADH-dependent

nitrite reductase termed NirB and is thought to be coupled to Nar. It is located in cytoplasm and reduces the nitrite generated by NarG to ammonia. Four genes *nirBDCcysG* are found in the nir operon (18). The structural genes *nirB* and *nirD* encode the subunits for the enzyme, *nirC* encodes a putative nitrite transport protein and the cysG gene product is required for the synthesis of the prosthetic group siroheme (18, 50). The most well studied cytochrome c containing nitrite reductase is found in E. coli and is formate dependent. It is located in the periplasmic space and is coupled to Nap. This cytochrome c nitrite reductase is termed NrfA, in which Nrf stands for "Nitrite reduction by formate". The X-ray crystal structure of NrfA from Sulfurospirillum deleyianum and Wolinella succinogenes revealed that the reductase is composed of two identical monomers each containing five hemes (25, 26). In these two nitrite reductases, they exhibit four conventional Cys-X1-X2-Cys-His and one unique Cys-X1-X2-Cys-Lys cytochrome c binding motifs. The X-ray crystal structure of NrfA from E. coli was made available recently and it also contains the novel Cys-X1-X2-Cys-Lys cytochrome c binding motifs (6). Seven genes, nrfABCDEF are found in the nrf operon in E. coli (18, 32). NrfB is a cytochrome c containing five heme groups, NrfC and NrfD are electron transfer proteins, and NrfEFG are involved in covalent attachment of heme c to a novel cysteine-lysine motif in NrfA (18, 24, 32). The *nrf* operon found in W. succinogenes, however, consists of *nrfHAIJ*, in which *nrfH* is a NapC type protein that transfer electrons from quinone pool to the catalytic subunit NrfA, and NrfIJ resemble proteins involved in cytochrome c biogenesis (66). According to Bamford et al, organization of nrf gene clusters can be divided into two groups among bacteria, which are E. coli like clusters

and *W. succinogenes* like clusters (6). Classification of the nitrite reductases is summarized in Table 4.

Function	Denitri	Denitrification		Assimilatory		nilatory
Organism			Bacteria/ Fungi	Cyano		
	Denitrifyir	ng Bacteria		bacteria/ Plants	Bac	teria
Cofactor	Copper	Cytochrome	Iron Sulfur	Iron Sulfur	Siroheme	Cytochrome c
		<i>cd</i> 1	Cluster	Cluster		
Classification	Periplasmic	Periplasmic	NADP-	Ferredoxin	Cytoplasmic	Periplasmic
	denitrification	denitrification	dependent	dependent		
Abbreviation	NirK	NirS	Nas	Nir	NirB	NrfA

#### Table 4: Summary of nitrite reductase

#### Phylogenetic diversity of iron reducing bacteria and nitrogen metabolism

The majority of microorganisms that have been identified as iron reducing bacteria are related to *Geobacter* and *Desulfuromonas* species (39). These microbes are strict anaerobes and based on 16S rRNA sequences they form a distinct clad (Figure 5). Other identified iron reducing bacteria such as *Shewanella putrefaciens*, *Shewanella alga*, and *Pseudomonas sp. stain Z-731* can utilize oxygen as an electron acceptor as well as iron (39). *Sulfurospirillum barnesii* (formally *Geospirillum barnesii*), which belongs to the Epsilon subgroup of the proteobacteria, is also capable of reducing iron coupled with oxidation of H<sub>2</sub> or lactate (68).

Nitrate reductases or putative nitrate reductases have been identified in some of these iron reducing bacteria. According to a recent review by Richardson et al (62), Delta

and Epsilon *Proteobacteria* only contain periplasmic nitrate reductase (Nap), whereas members of Gamma Proteobacteria can have different mixtures of assimilatory and dissimilatory nitrate reductases. This data was based on BLAST analysis of 77 complete or unfinished prokaryotic genome sequences at the NCBI Blast server at the time. However, only one bacterium each was listed as the representative for Delta and Epsilon, namely Desulfovibrio desulfuricans and Campylobacter jejuni respectively. Thus, it is possible that nitrate reductase other than Nap may be present in members of Delta and Epsilon Proteobacteria. Geobacter metallireducens strain GS-15 (referred as GS-15 beyond this point) is a member of the Delta Proteobacteria and is capable of reducing nitrate (41). Nitrate reductase activity has been found in its membrane fraction, which suggests that it has membrane bound nitrate reductase (58). Also a nitrate/nitrite complex has been characterized in GS-15 (44). This indicates that it may posses an alternative nitrate reducing pathway. Although, GS-15 is capable of reducing nitrate, a close relative of GS-15, Geobacter sulfurreducens is known not to reduce nitrate. Recently, five new bacteria were added to Geobacter genus, namely Geobacter bermensis (71), Geobacter pelophilus (71), Geobacter hydrogenophilus (17), Geobacter chapelleim (17), and Geobacter grbiciae (17). These new members of Geobacter are primarily characterized as iron reducing bacteria thus whether they reduce nitrate is yet to be studied. S. barnesii as mentioned earlier, belongs to the Epsilon *Proteobacteria*. A previous study has shown that nitrate reductase activity was found in the membrane fraction upon sonication treatment (69), however, only NapA has been found.

The presence of nitrite reductases in these iron reducing bacteria is relatively unclear compared to nitrate reductase. Cells of GS-15 harvested from media containing Fe (III) as

electron acceptor shows constitutive nitrite reductase activity but no nitrate reductase activity (58). This is also true with its close relative *G. sulfurreducens*. In the case for *S. barnesii*, a periplasmic nitrite reductase (NrfA) has been characterized from *S. deleyianum*, a close relative of *S. barnesii*. Thus it is very likely that *S. barnesii* also has NrfA (69).



Figure 5: Phylogenetic tree inferred from 16s RNA sequences (39)

The bold lines indicate iron reducing bacteria. The phylogenetic tree was inferred from 1007 positions by using the maximum likelihood method. The sequence of *Clostridium pasteurianum* was included as an out group.

#### Geobacter metallireducens and Geobacter sulfurreducens

Ferric iron (Fe (III)) reduction may cause a ground water quality problem (40). Insoluble  $Fe^{3+}$  oxide is reduced to the more soluble ferrous iron (Fe (II)), which upon reaching the surface is then oxidized back to Fe (III) oxide. Fe (III) oxide can clog wells, causing staining problems and water discoloring if it enters the plumbing system. The two most intensely studied Fe (III) reducers are Geobacter and Shewanella, which are in the Delta and Gamma subclass of the *Proteobacteria* respectively (40). GS-15 is capable of reducing nitrate to ammonia (42). A characterization study carried out by Lovley et al. (41) demonstrated GS-15 is capable of oxidizing short chain fatty acids, alcohols, and monoaromatic compounds with Fe (III) as the sole electron acceptor. The study also demonstrated reduction of manganese (Mn (IV)), uranium (U (VI)) and nitrate when acetate is utilized as electron donor. A dissimilatory nitrate reductase complex from GS-15 has been isolated and characterized previously (58). The presence of three indicative of c-type cytochrome absorption maxima at 552, 522, and 418nm were observed in the membrane fraction of nitrate grown cells. These three cytochrome c peaks were also observed in the membrane fraction of iron grown cells but at a low concentration suggesting that production of cytochrome c is enhanced when cells are grown in nitrate. A membrane bound cytochrome c containing enzyme complex exhibiting nitrate and nitrite reductase activity from GS-15 has also been characterized (44). The enzyme complex contains four different polypeptides with molecular weight of 62, 52, 36 and 16 KDa respectively. The 62 KDa polypeptide exhibited nitrite reductase activity. The known molybdoenzyme inhibitors do not inhibit nitrate reductase activity of the complex. This indicates an alternative nitrate reduction pathway may exist in GS-15. Additionally, the growth of GS-15 in the nitrate medium is significantly iron dependent (65). The nitrate/nitrite complex isolated from cells grown in iron enriched medium showed significant increase in cytochrome content (65).

Caccavo et al. (14) was the first to describe *G. sulfurreducens*. It was isolated from surface sediments of a hydrocarbon contaminated ditch in Norman, Okla. *G*.

*sulfurreducens* is capable of reducing iron, cobalt and sulfur coupled to the oxydation of acetate or hydrogen. However, *G. sulfurreducens* is unable to oxidize acetate when nitrate is used as the electron acceptor (20). A recent study demonstrated that *G. sulfurreducens* also reduces and precipitates technetium (Tc (VII)) (38). This discovery may lead to aid on controlling technetium-99, which is a fission product of uranium and is formed during nuclear reactions.

#### Sulfurospirillum barnesii

S. barnesii strain SES-3 (referred as SES-3 beyond this point) was isolated from a selenate contaminated slough in western Nevada (68). SES-3 is a motile, vibrioid shaped, gram negative bacterium, belonging the Epsilon subgroup of the Proteobacteria. SES-3 was first thought to reduce sulfate, because sulfate and selenate are closely related elements. However, the characterization of S. barnesii and other selenate reducing bacteria confirmed the reduction of selenate (Se(VI)) to selenium (Se(0)) is achieved by very different microbes using a different biochemical process (70). SES-3 can utilize selenate, arsenate, thiosulfate, elemental sulfur, trimethylamine oxide, Fe (III), nitrate, fumarate, aspartate and manganese as electron acceptor (68). Lactate, pyruvate or  $H_2$  can be used as electron donors. Acetate is required a carbon source when grown with hydrogen (68, 70). SES-3 reduces nitrate to ammonium (68) indicating that both nitrate and nitrite reductases are expressed in SES-3. A NapA homologue has been cloned and sequenced from SES-3 (Stolz and Lisak unpublished). Although it is most likely that NapA is expressed in cells of SES-3 grown in media containing nitrate as electron acceptor, the existence of Nar and Nas is a possibility. A result of western blotting using

antibodies raised against *E. coli* NapA showed a positive band at around 50 KDa with SES-3 proteins prepared from whole cell (personal communication with Dr Sudesh). This may suggest the existence of Nas, because protein sequences of Nas and Nap are similar in various bacteria (67).

#### Objective

The recent release of the GS-15 draft genome was made available to the public at DOE Joint Genome Institute website. Although, the genomic sequence was not fully annotated and not fully completed, this enabled us to investigate the presence of nitrate and nitrite reductase homologues in the organism. Three hypotheses were proposed in this thesis. Based on previous studies carried out in our lab (44, 58), the first hypothesis is then GS-15 does not have Nar. This hypothesis was based on the finding that typical nitrate reductase inhibitor did not inhibit the nitrate reductase activity and the cytochrome contents of nitrate grown cells were higher compare to iron grown cells. Also, the growth of GS-15 in the nitrate-limited medium is significantly iron dependent (65). The second hypothesis predicted that GS-15 can assimilate nitrate and thus possibly possesses an assimilatory nitrate reductase. A close relative of GS-15, G. sulfurreducens is unable to grow in media containing nitrate as electron acceptor. Whether G. sulfurreducens possesses an assimilatory nitrate is unknown and was also investigated. The third hypothesis predicted that SES-3 has a periplasmic NapA homologue. This hypothesis is based on the fact that the NapA gene has been cloned and sequenced.
#### **Experimental approach**

To investigate the first hypothesis, sequences of all types of nitrate and nitrite reductase, and also nitrogenase (i.e. Nar, Nap, Nas, NirK, NirS, NrfA, NirB, NifD) were collected from National Center for Biotechnology Information (NCBI) website. These sequences were used to BLAST search homologues in GS-15 draft genome sequence. An attempt was made to partially purify membrane bound nitrate reductase from GS-15. To investigate the second hypothesis, growth experiments were carried out in media containing various nitrogen sources (N<sub>2</sub>, NO<sub>2</sub><sup>-</sup>, NO<sub>3</sub><sup>-</sup>, and NH<sub>4</sub>) and Fe (III) as electron acceptor. To investigate the third hypothesis, an attempt was made to purify NapA from SES-3.

#### Methods

### A. Collection of enzyme sequences and BLAST search the draft genome of *G*. *metallireducens*

Sequences of enzymes involved in nitrogen metabolism (i.e. Nar, Nap, Nas, NirK, NirS, NrfA, NirB, NifD) were collected from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). The list of the enzymes and organisms are summarized in Appendix I. These sequences were used to BLAST search possible genes encoding enzymes involved in nitrogen metabolism in the GS-15 database. The BLAST search was carried out on the DOE Joint Genome Institute website (http://www.jgi.doe.gov/JGI\_microbial/html/geobacter/geobac\_mainpage.html) that contained the genome of GS-15 (Feb 28<sup>th</sup> 2003 version). The identified open reading frames of GS-15 were in turn BLAST searched at NCBI. Gene order was determined by obtaining the complete contigs.

#### B. Phylogenetic analysis of *nar* gene products

ClustalX was used for the alignment of the sequences (74). Phylogenetic trees were generated using Paup (72). Trees were generated based on neighbor joining method and maximum parsimony methods.

#### C. Cell culture

GS-15 was grown in fresh water acetate (FWA) with nitrate. The media were dispensed in 125 ml serum bottles and degassed (5 minutes medium, 2 minutes head space) with oxygen free  $CO_2$  and  $N_2$  gas (20:80 ratio). FWA-Fe (III) medium (Fe (III) 50mmol) was made separately for amending the FWA-nitrate media. After autoclaving, prior to inoculation of GS-15, FWA-Fe (III) medium was injected into FWA-Nitrate medium. The amount of FWA-Fe (III) added was calculated so that the final concentration of Fe (III) would be approximately 250 µmol in the FWA-nitrate medium.

For growth experiments, GS-15 was grown in fresh water acetate (FWA) with Fe (III) as the electron acceptors as described in Lovley and Phillips (42). FWA-Fe media were provided with ammonium salt only, ammonium salt with nitrite, nitrate only, nitrate with nitrite, or nitrite only, in different concentrations as nitrogen source (refer to Table 5 for concentration of each combination). To investigate if GS-15 contained nitrogenase, the microbe was also grown in FWA-Fe with N<sub>2</sub> nitrogen source. The media were dispensed in 125 ml serum bottles and degassed (5 minutes medium, 2 minutes head space) with oxygen free CO<sub>2</sub> and N<sub>2</sub> gas (20:80 ratio). The cells were transferred three

times subsequently to ensure the accuracy of the experiment. The inoculated medium bottles were incubated at 30 °C and color change started taking place on 4<sup>th</sup> or 5<sup>th</sup> day. If no growth was observed, the experiment was repeated to ensure the accuracy of the experiment. All experiments were done in triplicates.

Cells of SES-3 were grown in mineral salts medium (Appendix IV) containing lactate as the electron donor and nitrate as the electron acceptor. The medium was dispensed in 125 ml serum bottles and degassed (5 minutes medium, 2 minutes head space) with oxygen free CO<sub>2</sub> and N<sub>2</sub> gas (20:80 ratio). For large scale, a 14 Liter Microferm Fermenter (New Brunswich Scientific, Edison, NJ) was used. The medium was degassed longer (30 minutes medium, 15 minutes head space) and after autoclaving, prior to inoculation of cells, dithionite (10 ml of a 1 mg/ml solution) was injected to ensure the anoxic condition of the medium.

#### D. Solubilization of G. metallireducens NarG

Cells were harvested by centrifugation for 20 min at 14,000 x g and 4°C then resuspended in Tris-HCl buffer (Appendix III). The suspended cells were lysed by sonication on ice for 3 minutes and lysates were centrifuged at 14,000 x g and 4°C to remove unsonicated cells. The cell lysate was then centrifuged at 140,000 x g and 4°C for 90 minutes. The cell pellets obtained were incubated in CHAPS for 24 hr and then centrifuged at 140,000 x g at 4°C for 90 min to separate the CHAPS soluble fraction. The soluble fraction was treated with 10 % and 35 % ammonium sulfate to enrich for nitrate reductase.

#### E. Lysozyme treatment

The *S. barnesii* cell pellet was resuspended in buffer containing EDTA, 10 mM Tris PH8, protease inhibitor, 0.5 M sucrose, and treated with lysozyme (Table 5). Lysozyme treatment was done at room temperature for 1, 10, or 60 min. Spheroplasts and unbroken cells were separated from the supernatant by centrifugation at 12,500 x g at 4  $^{\circ}$ C for 15 minutes.

Lysozyme concentration	Incubation time			
(mg/ml)	(min)	EDTA (mM)	Temperature	
0.5	30	1	Ambient	
0.5	60	1	Ambient	
1	10	1	Ambient	
1	30	1	Ambient	
1	60	10	Ambient	
10	10	1	Ambient	
25	10	1	Ambient	

#### Table 5: Lysozyme treatment

#### F. Solubilization of S. barnesii Nap

Cells were harvested by centrifugation for 20 min at 14,000 x g and 4 °C then resuspended in Tris-HCl buffer (appendix III). The suspended cells were lysed by sonication on ice for 15 cycles (30 sec on, 30 sec off) and lysates were centrifuged at 14,000 x g and 4 °C to remove unsonicated cells. The supernatant was then centrifuged at

140,000 x g and 4 °C for 90 min, to separate the soluble (i.e. cytoplasmic and periplasmic fraction) from the particulate fraction (i.e. membranes, ribosomes). Activity assays for nitrate resuctase activity was done on both soluble and particulate fraction. The soluble fraction contained about 50 % of the total activity and was used for the purification steps.

#### G. Protein concentration determination

The Lowry method (43) was used to determine protein concentration. Bovine serum albumin (BSA) was used to generate a standard curve (5, 25,50,75 and 100 mg/ml).

#### **H. Electrophoresis**

#### 1. SDS-PAGE

Precast polyacrylamide Tris-HCl gels purchased from Bio-Rad (Hercules, CA) were used for SDS-gel electrophoresis. Approximately 20  $\mu$ g of protein was suspended in sample buffer for 2 min at 90 °C and loaded to each lane. The gels were run at 150V until the dye front reached the bottom of the gel. The gels were developed for the methyl viologen zymogram assay, stained for heme, or stained with coomassie blue for protein detection.

#### 2.Non denaturing (native) gel electrophoresis

Precast polyacrylamide Tris-HCl gels purchased from Bio-Rad (Hercules, CA) were used for native-gel electrophoresis. Native gel running buffer and sample buffer were used (Appendix III). 20 µg of protein was loaded to each lane. The gels were either

developed for the Methyl viologen zymogram assay, stained for heme, or stained with coomassie blue for protein detection.

#### I. Activity assay for nitrate and nitrite reductase activity

#### 1. Zymogram

A 100 ml volume of 10 mM methyl viologen dissolved in 50 mM Tris-HCl pH8 was reduced with 100 mg sodium dithionite. After electrophoresis, the polyacrylamide gel was soaked in the methyl viologen solution for 2 minutes then replaced with 20 mM nitrate or 5 mM nitrite solution to detect nitrate or nitrite reductase activity respectively.

2. Spectroscopic photometric assay for nitrate and nitrite reductase activity

The reaction mixture contained 1.3 ml 50 mM Tris-HCl pH8, 150  $\mu$ l of 50 mM electron acceptor (nitrate or nitrite), and 20  $\mu$ l of protein. The reaction mixture was degassed for 2 minutes and the head space for 1 min and the bottle was sealed. The reaction was initiated by adding 30  $\mu$ l of 50 mM methyl viologen and monitored over 10 minutes on a Perkin-Elmer (Pittsburgh, PA) Lamda2 dual beam spectrophotometer at 600 nm.

#### 3. 96 wells methyl viologen assay

The methyl viologen spectrophotometric assay was modified for the 96 well format. 150  $\mu$ l of 50 mM electron acceptor (nitrate or nitrite) and 20  $\mu$ l of protein was added to each well. The plate was transferred to the anaerobic chamber and the reaction was initiated by adding 30  $\mu$ l of 50 mM reduced methyl viologen. The color change was monitored. Both positive and negative controls were set up to ensure the accuracy of the

assay system. For the positive control, crude extract containing nitrate and nitrite reductase activity was used. The negative control contained no protein. To determine the sensitivity, different amounts of cell lysate with a known activity (in units) was used and the amount of time necessary for clearing was determined against control.

#### J. Heme stain

After proteins were run on the polyacrylamide gel, the gel was treated with 12.5 % trichloroacetic acid (TCA) for 30 min then rinsed with distilled water for 30 min to wash off excess TCA. While the gel was washing, 100 mg of DMB (o-Dianisidine) was mixed with 90 ml distilled water and stirred in the dark for 15 minutes. 10 ml of 0.5 M sodium citrate (pH 4.5) and 1 ml of hydrogen peroxide ( $H_2O_2$ ) was added to the DMB and stirred for 2 minutes. The gel was treated with the mixed DMB solution for 1 hr in the dark. After the staining, the gels were rinsed in distilled water.

#### K. Dialysis

Dialysis cassettes purchased from Pierce (Rockford, IL) were used to dialyze the protein. The 35 % ammonium sulfate fraction was dialyzed in Tris-HCl buffer (Appendix III) for 24 hours to eliminate ammonium sulfate.

#### L.Column chromatography

#### 1. G. metallireducens

The dialyzed 35 % ammonium sulfate fraction (8 mg in 2 ml) was loaded onto a DEAE-Toyopearl (Rohm and Haas, Montgomeryville, PA) ion exchange column (20 x 2 cm) with Buffer A (Appendix III) containing 1 % CHAPS (wt/vol) and eluted with a

linear NaCl gradient (0 - 0.5 M). The ion exchange column was connected to automated system Biologic LP (Bio-Rad, Hercules, CA) and 55 fractions of 1.5 ml were collected.

#### 2. S. barnesii

3 ml of the soluble fraction (5 mg/ml concentration) was loaded onto a DEAE-650 M Toyopearl (Rohm and Haas, Montgomeryville, PA) ion exchange column (20 x 2 cm) with Buffer A containing 1 % Triton X-100 (wt/vol) and eluted with a linear NaCl gradient using Buffer B that contained the same ingredients as Buffer A and 0.5 M NaCl. The ion exchange column was connected to the automated system Biologic LP (Bio-Rad, Hercules, CA) and 55 fractions of 1.5 ml were collected. This step was repeated until the soluble fraction (60 ml) was all used. The pooled fractions were concentrated by centrifugal filter device (Millipore Billerica, MA) down to approximately 50 mg in 5 ml and loaded on to Sephacryl 300 HR (Sigma, St Louis, MO) size exclusion column. The column was eluted with Buffer A containing 0.25 M NaCl at a flow rate of 1 ml/min. 3 ml fractions were eluted. The fractions that contained protein (as determined by absorbance at 280 nm) were tested for nitrate and nitrite reductase activity using the 96 well plate assay and the zymogram assay.

#### Results

### Physiological characterization of assimilatory nitrate/nitrite reductase in G. metallireducens and G. sulfurreducens

Growth of GS-15 and G. sulfurreducens was monitored by the color change of the media. Before inoculation, the color of the medium was dark brown (Fe (III)) and as the cells grew, the color of the medium turned to dark green then to clear. Growth results of both microbes in FWA-Fe medum amended with various nitrogen sources are summarized in Table 6. All media contained  $N_2$  gas as it was used for degassing. Media containing NH<sub>4</sub> was used as the control. Nitrite concentration as low as 0.5 mM inhibited the growth of GS-15 and G. sulfurreducens. Both microbes were able to grow in media containing nitrate as nitrogen source and medum in which no solid nitrogen source (i.e.  $NH_4^+$ ,  $NO_3^-$ ,  $NO_2^-$ ) was added. While these results could indicate the ability to assimilate  $NO_3^{-}$ , nitrogen fixation in both cases could not be eliminated. In fact, both microbes are capable of fixing nitrogen and nitrogenase activity and nitrogenase genes have been found in GS-15 and G. sulfurreducens respectively (8, 19). In the case for GS-15, the presence of nitrogenase was inferred by the reduction of acetylene and physiological characterization (8). The structural genes of nitrogenase were identified in G. sulfurreducens and were utilized as a reporter for constructing an expression vector (19). Growth rates for GS-15 and G. sulfurreducens have been reported (14, 42). According to the papers, the growth rate for GS-15 is  $1.5 \times 10^7$  cells per day and for G. sulfurreducens is  $1.7 \times 10^7$  cells per day at exponential growth phase.

Nitrogen source	PCA-1 Growth	GS-15 Growth
NH <sub>4</sub> (4mM)	Yes	Yes
$NH_4(4mM) + NO_2^{-}(4mM)$	N	N
$NH_4(4mM) + NO_2^{-}(1mM)$	Ν	Ν
$NH_4(4mM) + NO_2^-(0.5mM)$	Ν	Ν
$NO_3^-$ (4mM)	Yes	Yes
$NO_3^{-}(4mM) + NO_2^{-}(1mM)$	Ν	Ν
$NO_3^{-}(4mM) + NO_2^{-}(0.5mM)$	Ν	Ν
$NO_2^-$ (4mM)	Ν	Ν
NO <sub>2</sub> <sup>-</sup> (1mM)	N	N
$NO_2^-(0.5mM)$	Ν	Ν
None (N <sub>2</sub> gas)	Yes	Yes

 Table 6: Physiological characterization of G. metallireducens and G. sulfurreducens

#### Molybdoenzymes in G. metallireducens

A list of putative molybdoenzymes that were identified by using blastp from NCBI protein database is summarized in Table 7. The ten closest hits of the blastp result for each putative molybdoenzyme are listed in Appendix V. Two putative NarG homologues were identified in contig121 and 135. No other types of nitrate reductase (Nap or Nas) were identified. Four formate dehydrogenases homologues were identified. However, gene 2111 and 2112 are likely one piece rather than two because the size of gene 2111 is much too small to be a complete enzyme. The blastp results suggest gene 2111 and 2112 are fdh-N type formate dehydrogenase. Gene 498 and gene 3285 are also candidate formate dehydrogenases but which type is unclear based on the BLAST result. Gene 269 and gene 2012 are putative molybdopterin oxidoreductases.

Based on the phylogenetic tree analysis (Figure 6), gene 2111 and gene 2112 are closely related to Fdh-N and Fdh-O from *E. coli*, which is consistent with the blastp results. Gene 498 and gene 3285 are closely related to Fdh-H from *E. coli*. The two putative molybdopterin oxidoreductases, gene 269 and gene 2012 are placed closely to arsenate reductase from *Shewanella*. Gene 137 is likely a carbonmonoxide dehydrogenase, which is consistent with the BLAST result. The tree also indicates that gene 627 and gene 1636 are NarG homologues and are closely related to *Thermus thermophilus* strain HB8 (Here after strain HB8 is designated as *Thermus thermophilus* unless indicated).

Putative enzyme	Contig number	Gene number
Carbonmonoxide dehydrogenase	108	137
Formate dehydrogenase (alpha)	140	2111
Formate dehydrogenase (alpha) nitrate inducible	140	2112
Formate dehydrogenase (alpha)	119	498
Formate dehydrogenase (alpha)	148	3285
Molybdopterin oxidoreductase	113	269
Molybdopterin oxidoreductase	139	2012
NarG	121	627
NarG	135	1636

### Table 7: Putative molybdoenzymes in G. metallireducens



#### Figure 6: Phylogenetic tree of molybdoenzymes using the neighbor joining method.

Only the key abbreviations are explained. Representative from each group of the known molybdoenzymes including Nas (assimilatory), Nar (membrane bound) and Nap (periplasmic) nitrate reductases from various bacteria were collected from the NCBI database to generate the tree. Putative molybdoenzymes from GS-15 are listed as G.metgene followed by the gene number (e.g. G.metgene498). ArrA: Arsenate reductase,

COD: Carbon monoxide dehydrogenase, DMSO: DMSO reductase, Fdh: Formate dehydrogenase, XDH: Xanthine dehydrogenase,

#### Enzymes involved in nitrogen metabolism in G. metallireducens

A. Membrane bound nitrate reductase (Nar) homologues

Two putative nar operons were discovered in contig121 (nar121) and 135 (*nar135*). There are four ORFs in *nar121* and seven ORFs in *nar135* (Table 7). Both operons contain the putative membrane bound nitrate reductase structural genes narGHI. Secondary structure prediction suggests NarI in contig121 has 6 transmembrane domains whereas NarI in contig135 has 5 transmembrane domains (Appendix VI). A diheme cytochrome c was found immediate upstream of the putative narG gene. A novel cytochrome c (NarC) located upstream of narG has been described previously in T. thermophilus (84). T. thermophilus is the only other organism discovered with a cytochrome c located upstream of narG so far. A putative chaperon narJ was found only in *nar135*. Two putative *narK* homologues (*narK*<sub>1</sub> and *narK*<sub>2</sub>) were found downstream of narI in nar135. This is similar to the T. thermophilus nar operon as there are two narK located downstream of NarI (59). Secondary structure predictions suggest the two putative NarKs have 12 putative transmembrane domains (Appendix VI), which supports the putative *narK* genes products are indeed NarKs. The ten highest hits of blastp result for each Nar subunit are listed in Appendix V.

#### B. Putative nitrite reductases

Over all two putative nitrite reductases were found in the GS-15 genome available at this stage (Table 7). A putative periplasmic dissimilatory nitrite reductase (NrfA) and a putative assimilatory/dissimilatory nitrite reductase were found in contig140 and contig146 respectively. The latter was uncertain because even though it has a conserved NirB domain (dissimilatory nitrite reductase domain), the BLAST result indicated that it could be either assimilatory or dissimilatory nitrite reductase. This BLAST result was due to its NADH binding site, because both assimilatory and dissimilatory nitrite reductase (NirB) have NADH binding site. However, it is unlikely a NirB because the expected molecular weight for the ORF is only 45.5 KDa, which is much too small compared to the size of NirB that have been characterized so far (e.g. E. coli NirB is approximately 93 KDa). It is also unlikely to be an assimilatory nitrite reductase because of its size (B. subtilis NasD is approximately 88 KDa) and also the fact that GS-15 was unable to grow in the presence of nitrite as nitrogen source. The presence of NrfA is intriguing, as it is known to couple with periplasmic nitrate reductase (NapA), but no NapA homologue was found in GS-15 genome. Molecular weight prediction of the putative NrfA is approximately 54 KDa, which is a reasonable size for a NrfA homologue. Upstream and down stream of the both putative nitrite reductases have been investigated. A putative promoter site was detected at upstream of gene 2131, which indicates that gene 2130 is unlikely to be transcribed together with gene 2131. Based on BLAST results, downstream genes 2132 and 2133 have a conserved domain containing NapH (iron-sulfur protein) and PAS respectively. PAS is a DNA binding domain and has been found in NifL, NtrB and NtrC and also various histidine kinase sensor molecules in bacteria (53). Ambiguity in

genome sequence prevented further analysis of the immediate downstream region. On the other hand, upstream and downstream of the putative NADH dependent nitrite reductase are genes encoding putative glutamate synthase (i.e. gene 2980, 2982 and 2983). BLAST results also indicate that all these genes have conserved glutamate synthase domains. This may suggest that the putative NirB (gene 2981) is actually a subunit of the putative glutamate synthase encoded by the nearby genes (i.e. NADH binding domain for the glutamate synthase).

#### C. Structural genes of nitrogenase

Structural gene homologues of nitrogenase were identified in Contig145 as summarized in Table 7. This corresponds with the physiological characterization, in which GS-15 was able to grow in FWA-Fe media without nitrogen source (i.e. only  $N_2$  gas).

# Table 8: Summary of all putative genes involved in nitrogen metabolism in G. metallireducens

Putative <i>nar</i> subunits	Contig number	Gene number	Contig number	Gene number
CytC (narC)	135	1635	121	626
narG	135	1636	121	627
narH	135	1637	121	628
narJ	135	1638		
narl	135	1639	121	629
narK₁	135	1640		
narK <sub>2</sub>	135	1641		

Putative nitrite reductases	Contig number		Gene number	
nrfA		140		2131
Assimilatory or <i>nirB</i> ?		146		2981
Putative nitrogenase structural genes	Contig number		Gene number	
nifH		145		2709
nifD		145		2710
nifK		145		2711
Missing		145		2712
Hypothetical		145		2713
Missing		145		2714
nifE		145		2715

#### Structure of nar operon in G. metallireducens

A schematic drawing of each *nar* operon is shown in Figure 7. The size and the order of the each putative *nar* genes in each operon is as follow; *nar* operon in contig121 contains *narC* (886bp), *narG* (3615bp), *narH* (2288bp), and *narI* (1179bp); *nar* operon in contig135 contains *narC* (789bp), *narG* (3576bp), *narH* (1455bp), *narJ* (620bp), *narI* (689bp), *narK*<sub>1</sub> (1320bp), and *narK*<sub>2</sub> (1428bp). No putative genes encoding the NarXL system was found in either upstream or downstream of the operons. The size of *narH* in *nar121* is significantly larger than the *narH* in *nar135*. The nucleotide sequence as well as deduced amino acid sequences for *nar121* and *nar135* are shown in Figure 8 and Figure 9. Promoter sites and transcriptional regulation sites were predicted using Neural Network program and IMD program respectively (15, 79). The Neural Network program predicted one putative promoter site and four putative promoter sites in contig121 and contig135 respectively. Only the promoter that is closest to the start codon ATG is indicated in Figure 8. No putative transcriptional regulation site was found upstream of

*narC* in *nar121*, whereas a putative CAP/CRP site was found in one of the putative promoter region of *nar135*. A putative Shine-Dalgarno sequence was found upstream of every *nar* gene in both *nar* operons. The start codon of *narJ* and *narI* overlaps with *narH* and *narJ* respectively in *nar135*. Thus *narHJI* are probably translated from the same mRNA. There is a 218 base pair gap between *narI* and *narK*<sub>1</sub> but no putative promoter region was found in this gap using the promoter prediction program. This is significantly different from *T. thermophilus*, in which the ATG start codon of *narK*<sub>1</sub> overlaps the last codon of *narI* and both are likely translated together from the same mRNA.



Figure 7: Gene order of nar operon in G. metallireducens

T I P S K P T V L R M S A T 181 c + Gene626 214 atgaaaacggtgatcatcggcgggggggggggcgccagtgccgcctgc ΜΚΤΥΙΙσσνΑΑσΑΑΟ 259 gccatgacgcgaagaatgctcgttgccatggtgatcgcgctgctg ANTRRMLVAMVIALL getteegecatgeatecceaccaggggtettecaaagagagggea SAMHPHQGSSKERA 3.49 gaggggggacggaaaggeeetttteeaggegaaatgeteteeetge EGDGKALFQAKCSPC 394 cacaccataggcggtggcaggctggtgggtccggacctcaagggg ΗΤΙσσακινσροικσ 439 gtgacggcccggcgggaccggggatggctcgaacggttcatctcc TARRDRGWLERFIS gogceggacaaggtgetggeaageggggateeggttgeegeeagg A P D K V L A S G D P V A A R 529 ctgttccaggagtccaacaaggtaccgatgccgaacctggggctt LFQESNKVPMPNLGL 574 ggccaggatcaggtcaatgtcctcctcaactatctttcggccccg G Q D Q V N V L L N Y L S A P 619 gaaggaatateteeateeaaggeegaaceeetgeeteaceegtt EGISPSKAEPPASP 664 caggaggagggagatcogtogagggggggagaaattotttacoggo OEEGDPSRGEKFFTG 70.9 gegeteccettegeadaaggeggeaccecetgtategeetgteat ALPFAQGGTPCIACH 754 caggtggccggcattectcccctgggggggggggcgccctcggcccc O V A G I P P L G G G A L G P 799 gacctgaccacgatcgtcgcgacgttcggcgatacggcgctctcc LTTIVATFGDTAL totgecottgeaacgeteccettteogaceatgaaacegatette 844 SALATLPFFTMKPIF 889 ggcgadcatedeettaccegeggggggggggggggggtttgetggce DHPLTRGEQGDLL

tttttcagaaaagtggcaccgcggccgccggtgacggcaacaacg

FFREVAPRPPVTATT

ogtattgetttgtgggggggggggggggggtteetgete

IALWAGGGFLVLLL

tctacgcctgtgagaactattgcccccaccaggggagccccatgc

STPVRTIAPTRGAPC

tggcgggaatcgtcaaggacggggtcattgcctgcccgcggcacg

gctggcactacaaactggcggacgggggatgcacgggatttcoccg

M R E S S R T G S L P A R G T

AGTTNNRTGDARISP

actataccetcaaaacctaccgtgttgaggtggtcggcgacgaca

1

46

136

459

979

STHGVNCTGGCSWLV 1246 tacgtgaagggaagggategtcgcctgggagatgcaggccacggac YVKEGIVAHENQATD 1291 tatcogegtetegaagegggggetteeccectatgaacecegegge Y P R L R A G L P P Y R P R G 1336 tgecageggggggateteetttteetggtacetetacagteecete CORGISFSNYLYSP 1381 oggatcaagcatcogacgatcogggggaagcotootogacototgg RIKHPTIRGSLLDLW 1426 egegeggeaaaggaggageacagegaceegetggeegeetgggge RAAKEEHSDPLAANG 1471 tocattgtggaggacagegagaaacggegcagttaccagcaggec SIVEDSEKRRSYOOA GKGGFRRTSWDEAL 1561 gagatcacggccgcgtccacgctctatacgatcaagcggtatggc EITAASTLYTIKRYG 1606 ccggacaggatcgtggggttttcccccatcccggccatgtccatg PDRIVGFSPIPAMSH 1651 tgcagetacgccgccggaaccaggeteetttegeteetgggaggg CSYAAGTRLLSLLGG 1696 gtggtgetgagtttttacgactggtacgeogatttgeeteeegee V L S F Y D W Y A D L P P A 1741 tecceegagacetgggggggggagaagacegaegtggeggagagegee SPETHGEKTDVAESA 1786 gactggtacaacagcaagtatattgtggtcatggggacaaacete DNYNSKYIVVMGTNL 1831 tocatgaogogcaccooggaogtocatttogtogoggaggocogo SHTRTPDVHFVAEAR 1876 accaacggegecaagetggtggtgetetegeeegattteageeag TNGAKLVVLSPDFSO 1921 gtategaaatacgcegactggtggettecegteacegeeggeeac VSKYADNWLPVTAGH 1966 gacggggeettetggatggeggteaaceatgteatteteagggaa

DGAFHMAVNHVILRE

FYVERHVPYFADYLA

2011 ttttatotegagegecacgteccgtactttgecgactacetegeg

LTGALGRRRLGSVR

1111 atgcgatggatcaaggagttgacggatccgaaggcgcggcgatgg

1156 gaggaattetacegtaacegetggcagcacgacegggtegtgegg

1201 ageacceatggegtcaactgtaccgggggggtgttegtggetggte

MRWIKELTDPKARRH

EEFYRNRWQHDRVVR

TLVDEANRAGR\*

Gene627

2101 ggagaacggtatgtgccggggcggcttctgcgggcggaatcectg GERYVPGRLLRAESL 2146 ccccgttaccggagggaggaaaacggcgagtggaagttcctcgtt PRYRREENGEWKFL FDRLSGEPRMPGGSI 2236 gggttccgctggcagagggaagcggggaggtggaacctggagatg GFRNQREAGRWNLEN 2281 aaggatggeetegaoggtgeggagategageeeetteteteeett KDGLDGAEIEPLLSL LDGRDDVLPVSITDY 2371 geoggegggaccagegtegegggggggtteeggteaggtatetg AGGTSVARGVPVRYL 2416 gagacogecacogggagaacteeogtcacaacggtettegacete ETATGRTPVTTVFDL 2461 ctcatggcaaacttcggcgtcggggggggcctgtccggagactat LMANFGVGRGLSGDY 1506 coggaggattacaatgcogcogcogggaacccocgcotggcag PEDYNAAAPGTPANO 2551 gaggagtacaccgggggttggccgcgagaccctgatccgctttgcc EEYTGVGRETLIRFA 2596 cgggagtgggccggcaccgccgagagaacccggggggaaatgc REWAGTAERTRGECT 2641 atcategteggttccggcgcaaaccactggtatcacagcaacctc IIVGSGANHИУНSNL 2686 acctaccggaccggaatcaccgccctcatcctctgcggctgcgtc TYRTGITALILCGCV 2731 ggggtgaacggeggegggetcaaccactacaceggecaggagaag G V N G G G L N H Y T G Q E K 2776 gtageteeegaggegtettggteteteetegeetttgeeaaggat VAPRASNSLLAFARD 2821 tggacaaaggggggggggggggttecagaacacecegtegttecactac W Τ K G G R F Q N T P S F H Y VHSDOWRYDGGAADP 2911 cgtcccgatccgaaaggggggaccggccggccgtcaccccatggat R P D P K G G P A G R H P M D 2956 etccaggtggacgcggtccgcatgggatggetcccettttatecc LOVDAVRMGHLPFYP 3001 cagttegacegcaaccegctegaactegtgaaggaggeggaggec Q F D R N P L E L V K E A E A 

RYSDAPFLVTLDRGD

1024 ctgactggegeectggggggegeggaggetegggggggteggggggteggeeggg 2056 egetatteggaegetegttectggtgaegetegaeggggggt

AGAKGDEEIIRNVVE

46

3091 gaactggagaagaagaaacggeteogettogeegtggaagageeggat 4171 tggtategegeettoggegaacaceteecgacetteaagacaagg 5226 tatggggeteatgaecegaacettgetgeeggagegggaa

W Y R A F G E H L P T F K T R

IDTEESGHLAKSRPV

GRALMLNCLTPHGKN

нинттуурлненгт г

SRGIEPFHLNDRDAE

EIGVKDNEWVEVYND

NGVVVTRAVVSARIP

RGVGIYYHAPERTIS

V P L S P L R G N R R A G G H

GYGOLSYAFNANGPP

VLMAG

HKLEG

4351 tecoggggtategageeettetggetgaacgacegggatgeegag

4396 gagateggggtgaaggacaacgagtgggtggaagtgtacaacgac

4441 aatggggtegtggtgaccegggeegtggtaagegeeeggateeee

4486 cgaggggteggcatctactatcacgegeeggageggaccatttee

4531 gtteccetetecceceteagggggaacaggegegeggggggaac

4576 aacageetgaceegggeeggttgaageeggtgetgatggeegge

4621 ggetatggecageteteetacgeetteaatgeetggggacegeeg

4666 gggagegaceggggacacetacgtgtatgtgcacaagettgagggg

4711 gecettegetattga GGGGAGAGGGGAACGGTTATTCCCGTCGGG

4776 atgaaagtcogtgcccaggtttccatggtgttcaacctggacaag

MEVEAQVSMVFNLDK

CIGCHTCSIACKNIN

SDRRGAEYMWWNNVE

TEPGTGYPTLWEDQE

KYRGGWERTDREPRL

RLHSRGGGLAKLFFN

PRLPTIDDYYEPHSY

RYGDLLSAPAGDDOP

VARPVSLVTGDPINI

EAGPNWDDDLGGSPV

5091 cggtacggogateteeteagegeeeeggeegggggatgaceageeg

5136 gtggcacgcccggtctccctggtgaccggagaccccatcaacatc

5181 gaggeegggeeeaactgggaegaegateteggggggeteteeegte

5001 cggetecaeteceggggtggeggattggegaagetattetteaat

4866 agegacegeegggggegeegagtacatgtggtggaacaaegtggag

NELTRARLEP

GSDRDTYVYV

A L R Y \* ATCTGAGGAAGGAGAAGCGCC

Gene628

YGAHDPNLAAVSEAE

RRQLFLLERMVLFYL

PRICNHCANPGCVAA

CPSGAIYKRGEDGIV

LVSOEKCRGNRMCVS

5271 cggcgtcaactetteeteetggageggatggteetettetatett

5316 cccoggatotgcaaccactgcgcgaatccogggtgcgtggccgcc

5361 tgtccttccgggggccatctacaagcgggggggaagacggcattgtc

5406 ctggtgagtcaggagaagtgccggggctggcgcatgtgcgtttcg

5451 ggatgtccctacaagaaggtctattacaattggggtagcggtaag GCPYKKVYYNWGSGK

5496 toggaaaaatgcateetetgetateeeegeeaggagaceggeeag

SEKCILCYPROETGO 5541 gecoeggeotatttecattectgegtegggeggetecgetacetg

A P A C P H S C V G R L R Y L

G V L L V D A E R I E R S A S

RPDHDLVQAQRENIL

DPSDPEVVNKAAEEG

RFVKEWRLALPLHPE 5811 tteggeacegteeegatgetetattaegteeeteegeteetgeeg

GTVPMLYYVPPLLP

LAADRENGCYDLGGE

FFTSLEQGRLPLRYL

ASLFSAGNEEIIRAV

YRKLTAVRVHRRAAT

TGDVPAGELERFLAE

AGTTPEEAEAIYHMT

SLAGFDERHAVPPFM

REMAMEMSVDPADRK

GEAGTGFLRPPKRGN

5586 ggagtgetectetaegatgeggagegaategaaeggagtgetteg

5631 cgcoccgaccacgacctggtgcaggegcaacgggagatgattete

5676 gacccctccgatcccgaggttgtcaataaggcggcggaagagggg

5721 attcacccageggtcctggccgctgcccgtaattcgccggtctat

5766 cgattcgtcaaggagtggcggctggcactgccgcttcatcccgaa

5901 ttetteaceteetagageagggggegettgeegettegetatete

5946 geatecetetteteggeaggeaatgaagaaateateegegeggte

5991 tacoggaaactcacggeggteegtgttcacegtegageggecace

6036 acgggagatgtgccggccggcgagttggaacggttectcgcggaa

6081 gegggcaccacccoggaggaagcogaggogatctaccatatgaog

6126 teacttgeogggtttgatgageggeacgeogtteeccectteatg

6171 cgggagatggcgatggaaatgtcggtcgatccggctgaccggaaa

6261 tga TGGCCACGAT

IHPAVLAAARNSP

LAEAVDAANAILHFA 3901 cccgagaccaacggcgaggttgcctggcgcgccttccggtccctg 4956 aagtaccggggagggtggggagcgaaccggggaaccggggaccgcgcctc PETNGEVANRAFRSL 3946 gaggagaagaceggeeteaccetegeegacetegeeteegactae EKTGLŸLADLASDY RGTRYRFSDLVSQPR 4036 ogcateettaegageeeetgetggteggggggteaceaacaaegge ILTSPCWSGVTNNG 4081 coccectacocccctattocctcaacgtogagegettggtcccc RPYAPYCLNVERLVP

5046 ccccgtctcccgaccattgacgactattacgagccgtggagctac

4126 tggoggaccetgaccggccgccagcatttetatetggatcacggg

N R T L T G R Q H F Y L D H G

ESAPVREWGGKRFPS

3811 gagagtgetceggtcegegagtgggggggggggggggttteeetee 3856 ettgeggaggeggtegatgeggeaaaegegataeteeaetttgeg 4911 acgaageegggaceggetateegaeeetetgggaggaceaggag

GLRWPVADLYDELL

FVSLGEEVKREGIEE

3766 oggggaeteeggtggeeggttgeeggttgeegatetetaegaegaaetgetg 4821 tgtateggetgeeaeaegtgeageategeetgeaagaacatetgg

LAV VERDPVHLLDR 3721 tttgtctcgctgggtgaagaggtaaaacgggaggggatcgaagag

H H R G E C E A V P G T T M P

3676 cogettgecgtggttgaacgggattttgttcatetettagaccgg

LHDTPGEIAOSEVRG 

IHPPPPVRDIVAAPL 3586 ctccacgacacccccggcgagatogcccagagcgaggtgaggggg

E I F K A L A E K I S A L A R

3496 gagatattcaaggccctggcggagaagatctcggcactggcaaga 3541 atccacttecctgagccggttegggacategttgeegegecectt

LSALVPPCWESPENW

K D D L N T T D L H S F I H P 3451 ctgtcggccgccgtcccccctgctgggaatcccggagcgactgg

ΑΓΑΣΟΙΛΓΡΤΑΤΜΥΕ 3406 aaggacgatetgaataccacggatetecactogtteatacaccee

3361 gecetetactecgatategtectececacggecacetggtacgag

K L D L V V D I N F R H D T S

QVENAAWQDAAPRG

3316 aagetegatetegtegtggacateaattteeggatggacaeeteg

3271 gagcaggtggagaacgeggegtggcaggatgeegeeceeggggg

3226 tateteggeacceatgacageageategeegacgagtgegeegee YLGTHDSSIADECAA

LLSSAKGHEYFLRH

A P E N W P R L W L I H R G N 3181 gegetectetecagegecaaggggcacgaatacttectcaggcac

LEKKRLRFAVEEPD

4261 ggcaaggeteteatgeteaactgeettaceeeccaeggeaagtgg 4306 cacatgeatacgacctattacgacaateaceggatgeteaceetg

3136 geeeeggagaactggeeeeggetetggeteatetggeggggaaac 4216 ategacacagaggaategggeeacetggogaagageaggeeggte

Gene629

6277 atgeoggtegatteoggaetgteggaecatetgegeeatecteogg MPVDSGLSDHICAILR 6322 tttettgeeegteeggeeggaacageeeggaacageggtgaeeet FLARPAGTAPNSGDP 6367 atgactgactatetettttteggtgtttteeeetaegttgeeete M V D Y L F F G V F P Y V A L 6412 gttgtcgcagtggggggggggggggggtgtatcggtatgcgaaggaacgg V A V G G G V Y R Y A K E R 6457 tttteetttteaageetetetteeeagtttetggagaaeegggee FSFSSLSSQFLENRA 6502 cttttctgggggagtgttgcctggcattacggcattgtcctggtt LFWGSVAWHYGIVLV 6547 ctectegeccatetegeggeactgetecteccegeatectggggg LLAHLAALLLPABNG 6592 cgcctcctgggcgacccgctccgtctcgccgtccttgaagtgacc RLLGDPLRLAVLEVT 6637 ggeettgeeetegggetgtteaceettgeeggeategtggteett GLALGLFTLAGIVVL 6682 ctgttcaggoggatcgccagtccgaaggcgtgggcggttgtcacg LFRRIASPRAWAVVT 6727 gtcatggacttgttcctgttggccgtccttgttcttcaagtgggt V N D L F L L A V L V L Q V G 6772 accggactttacaatgetttegtgtategetggggtteeetetgg TGLYNAFVYRWGELW 6817 tatettteectggeogteecctggeteeacteectggeatgeetg YLBLAVPWLHBLACL 6862 aaccoggatatcaaatacattgttocactcoogttgcoggttaaa NPDIKYIVPLPLPVK 6907 gtgcacctgttcaacgectttgtcatctttgttetctttccggtc HLFNAFVIFVLFP 6952 accounting account of the second TRLVHIVBIPVSYLW 6997 cgcccctggcagatagctgtctggaataagcctccgcgttggatg RPWQIAVWNKPPRNM 7042 atcccaaggaagtga 7056 TPRK \* ACGGGATCGTGGTGGAGAGCGAGCAGGGGGAGGGGACGACCGTGCGGGTG CTGATTOCGOGGACOGAGCAGATOGAACGGGCACTOGCATOGCOCGCAC

CAAAGCATGGGGGGAAAACCGGAGTCGCTCCCCGGAAAAGGAGGAGAGAGTATG AAGAACTTTATTACAACCATGGCAAAGGAACTGATCGCGGCATTCGAGAG agagtavococagoogaacagtacaaaagtaccotetococaacogtea GAAGCCTTCCCGAAGGAATGACCGGCATCTGGGATGTGGATGGCAGGGAG GTCAACGTGCCGAAAGGGACCTTCATCAAGGCGGATCATGGAAAGGCCGA AGTGGGGGCTCTGTGGAGGTTGAGGGGACCCTTAGCGGCAAGACGCTGA ATGCTTACCGGATTGCCGTAACGAGGGGGGGCACGGAAGCGCAGAAAATC GCACAACGCGGGGGTGCTGGCGACGAGATGTCGCCATCAGGCAAAAG CGTGACAGGACAGAAGGTGGGGCCTGTGAGGAGGTTGAAGGCCCCAATGG GGGAGGCACGGTCGGAATCACCGGGGACAGGAGGATAACATGAAAAACTTC ATCGCCATCGTGGCACTGGGGGCTGTTCGTGGCGCCGGCGGGAGGAGTTAC ATTOGCCAGOGAGCATGGAAGGCATGAAGGAAAAGAGGGGGGGACAGTATG AAAGCACGATCTACGGAACGGTCAGAAGCCTCCCCAGCGGGATGATCGGC GTCTGGAACGTGAACGGCAGGGAGGTCAACGTAACGAAAGCCACCAGTAT CARGEARANCENTGEARANCETGEAGTGEODECCTECCTOERGTGEATG GGCACCATGCCGGCAAGGTCTTTTCCGCCGGCAAGATCGAGGTAAAAAGG GGAAAGTGATAGGCACTGACTCGGOGGCGGGGGTCTTCCCTGCCCCGGCAAC GGTTCCCTCTGGTAGAATGAAAGGAAGTCATGACCTCTGGCCTAACGGTG GCGCGAG

#### Figure 8: Nucleotide sequence of nar operon in contig121

The deduced amino acid sequences are indicated below the nucleotide sequence. The stop codon is designated by \*. The beginning of each open reading frame is designated by the gene number and an colored arrow corresponding to Figure 7. The nucleotide sequence in the box is the putative promoter region predicted by Neural Network program. Bars above the sequence indicate putative Shine-Dalgarno sequences. Red arrows indicate the inverted repeat that can form a hairpin.

1 GAAAAAGTTTGGCGCCAATTTTGATTACTTGACTAAAGTCAAGGC 1071 atggtecatgtgaaggacggcategtegggaggtecagge 2151 otgggtggogagacogggaaggtcogttcoggtttgaaggggaa LGGETGEVRFRFEAE 1116 aacgactateeccagttegacygcaccatteccaaccacgageee 2196 gyggagagegteegoggtgeeggteeggtgeeggteeggeggattgagaegaag G E S V R E V P V R R I E T K 2241 aacggccccgtgacggtggctaccgtgtacgaccttctcatggcc NGPVTVATVYDLLMA 2286 cagtteggegttteeeggggeeteggggggggggactateceaagage Q F G V S R G L G G D Y P K S 2331 tacgactcogacaagcocttcaccoccaagtggcaggagaagtac YDSDKPFTPKWQEKY 2376 accountegrogeogaatoontcatcaagatogeocgggagtgg TGIAAESLIKIAREW 2421 gccgagaacggcgagaagagcggcggccggaacatgatcatcatc AENGEKSGGRNMIII 2466 ggcgccggggtcaaccactggtaccacaacgacctcatctaccgg GAGVNHWYHNDLIYR 2511 geogecateaccgccctgatectgaccggcagcgtcggccgcaac AAITALILTGSVGRN 2556 ggcgcgggggctcgcccactacgtggggcaggagaaggtggtcccc AGLAHYVGQEEV 2601 ctggctccctggacatccatcgccatggcccaggactgggtaaag LAPWTSIAMAODWVK 2646 ccgtoccggotccagaacaccccgagcttotggtacatccactcg PSRLQNTPSFWYIHS 2691 gaccagtggegetacgaccggacattcgtcgactacctcaagecc DQWRYDRTFVDYLKP E T G E R M P L H A A D M N A 2781 aaggeggeceggeteggetggeteccettegececceactteaac KAARLGHLPFAPHFN 2826 gagageeeeetgeggetegtggaggeggeeaaggeegeegee ESPLRLVEAAKAAGA 2071 cagagegacgacgaaatcegeteetggetegtgaategeetgaag Q S D D E I R S W L V N R L K 2916 agcggagaaacccgcttcgccatcgaggacccggatggccagggg SGETRPAIEDPDGOG 2961 aactococcaaggtotggttcatctggcggggccaacgccatctcg NSPEVWFIWRANAIS 3006 tecagegeeaaggggeacgagttetteetgaageacgteategge S S A K G H E F F L K H V I G 3051 gcccccaacagcagcgtgagcgccaaggaggtggccaagggggag A P N S S V S A K E V A K G Q 3096 gtgaaggatetegtgtggcaegacaaggeeeeegaagggaagatg V K D L V W H D K A P K G K M 3141 gacetggtggtggacetcaactteeggatggacacetecaceete DLVVDLNFRMDTSTL 3186 tactoggacategtectececacggecacetggtacgagaagteg Y S D I V L P T A T W Y E K S

MVHVKDGIVGWELQA NDYPOPDGTIPNHEP 1161 cggggggggacccgggggatcagctactcctggtacctctacagc

1206 ccgctgcgggtgaagtacccctacatgcgcggcgccctcatcgat

1251 ctctggcgtgacgcccggagccgccacgaagaccccgtggaggcc

LHRDARSRHEDPVE

1296 tggaagagcatcgtggagaacccggaatcccgtcgggcctacgtg

1341 ageaageggggggaceggtggcatgegeegggeegactgggacgag

1386 gtgaccgagetgategecegectectccatetacaccgegaagaag

1431 tacggee

1791 gggcacgad

WKSIVENPESRRAYV

SKRGYGGMRRADWDE

VTELIAASSIYTAKK

YGPDRVIGPSPIPAM

SHISYAAGTRFLQLF 1521 ggeggggtegecatgagettetacgactggtactgegaceteeeg

GGVANSFYDHYCDLP

1566 coggogtogcoccaggtotggggtgagcagacogacgtgaacgag

PASPOVNGEOTDVN

1611 teogeogactggtacaacgeeteetacattgtgetetgeggetee

1656 aacgtccccatgacccggaccccggacgcccacttcatgaccgag

1746 aacategecaccaagttegeegacaactggeteeeggtggageag

NIATKPADNHLPVE

1836 aaggagtgetacgeegacegeeaggteeeettetteeaggagtat

K E C Y A D R Q V P P P Q E

1881 geostcaagtacarggacetecegtteetegtgaagetggaggag

1926 aaggacggagcgttgaagcaggggggaatteeteegggetgeegag

1971 ttogagoggagogcotccctcgagcatgcogactggaccetetgc

ALKYVDLPPLVKLEE

KDGALKQGEFLRAAE

FERSASLEHADHTLC

VADGAGDVKIPHGSL

GTRNSKQEGKWNLDH

TDLVDGTPIDCRLTL

NVPHVRTPDAHPMT

A R Y R G T K V V V M S P D Y

GHDGAFHMAVNHVIL

ggegeettetggatggeggtgaaceacgtgateete

SADNYNASYIVLCGS

1476 tecatgateagetacgeggeeggtacceggttectccagetette

oggacogggtcateggettetegecgatecoggee

atg

PLRVKYPYHRGALID

GCTRGISYSWYLYS

ACCARCETTOCARARCOCTATTITACCGGATAGACTCACTACTTREE TARTOTOCGRAPTARTAGATATTGTACACTTATCCACGTACC

Gene1635 147 atggcgagaggaattetecgagcagcagtcagcegegtatgtetg MARGILRAAVSRVCL 192 gttetggtgetgtecettgeggegeacgeatacggtgateeggge LVLSLAAHAYGDPG 237 aaggacetgttegacaaacagtgtteaggetgeeacacegtegge K D L F D K O C S G C H T V G 282 ggcggcgattcgggggggaccggacctcaagggggtgaccgacagg G D S G G P D L K G V T D R 327 aggtccgaggagtggctcgaatcgatcatcattgagccggacaag SERWLESITIEPDE Ð 372 ctggcggcgaccgatccggctcggaaggagctgatcaagaaatac LAATDPARKELIKKY 417 ggctacgagatgccaaagctcggcatcggccatgacgacgccatc GYEMPKLGIGHDDAI 462 Asgateateacetggeteagggggggggggggggggggggcceagace IITHLRGGGGG A O V 507 gggtcaccgccggcggacgaacatccggaagtcgtggtgacgccg GSPPADENPEVVVTP 552 gcgctggttgccaaggggaaagcgtttttcaccgggcaaaagccc LVAKGKAFFTGQKP LANGGAPCIACHRIS 642 acccogggcattacgggaggaaacctogcogtotocaacctgage TPGIVGGNLAVSNLS ASYQKMGEKGMRGAL 732 aaggeeetcaagtteeegaccatgaagaagatetacgaaaacege KALKFPTMKKIYENR 777 cccctdaccdatdacdaadtaacdadcctdatcdccctdtacaad

LTDDEVTSLIALYK 822 gatgeegee etcactaagagcacaccatetgcgtcettettccccDAALTKSTDSASPFD 867 geotacggegegggggaetttegtegtgtteetegteggettgaet A Y G A G T F V V F L V G L T 912 ctctacaaaaggaggaccagataa GACA LYKRRTR\*

Gene1636

936 atgacacaggaaaggeteaaggaegaeggtgeggatgtgaeege 2016 gtggeegaeggegeeggggaegtgaagateeeeeaegggageete MTQERLKDDRCGCDR 981 gugtgggaagagttetacoggaaccgetggcagtacgacgaggtg 2061 ggcacccgetggaggaaggaggggaagtggaactggaacctggact

GHEEFYRNRHQYDKV 1026 gtgcggagcacccacggggtgaactgcaccggcagctgtagctgg 2106 acggatetegtggacggtacccccatcgactgccgcctgaccctc VRSTHGVNCTGSCSH

49

DYGNLYGAKGGDDOP 4876 gtggeegaggeetttteecagateaceggggagaagatggeecat AEAFSQITGEKNAH 4921 ateaaggeeggeeceaactgggaegaegaecteteeggeteeeag IKAGPNNDDLSGSQ 4966 gtctacgcgtcggcagacgtgaacctggaggacaaaaagatc VYASADVNLEDKKII 5011 gagcagtacagcacgatgttcatgcagtacctcccccgcatctgc EQYSTMFNQYLPRIC 5056 aaccactgegegaacceggectgegtegectectgecegteecgg NHCANPACVASCPSE AIYKRGEDGVVLVDO 5146 gaggtetgeaaggggtggeggttetgeaegagegeetgeeeetae EVCKGWRFCTSACPY 5191 aagaaggtetaetteaactggaacageggcaaggeggagaagtge K K V Y F N H N S G K A E K C 5236 atettetgetteecceggacegagaceggecagtgeaacgeetge IFCFPRTETGQCNAC EIEVYNFNGVVVSRA 5281 geocacagetgegteggeogtateogetaogtggggggteeteete 4266 tgtgtctcctcccggatcccgaaaggggccgcgggtgatgtaccad AHSCVGRIRYVGVLL

ILDETEGEQGLVLNF LTPHGKWSIHSTYSD NLRMLTLGRGGPVVW LNHEDAAGAGIEDNE

CVSSRIPKGAAVMYH

3996 ttocacgaggggctccccaccttcaaggcgaagctcaaccoggcg HEGLPTPKAKLNPA 4041 attetggacgagacegaggggggaacagggggettgteetcaaette 4086 ctcaccccccccggcaagtggagcatccattccacctacagogad 4131 aacctcogaatgetcaccctgggccggggggggaccggtggtetgg 

4221 gagategaggtetacaactteaacggegtggtggteteccegggee

3951 toggggogccagtccctctacctggaccacccctactacggogag G R Q S L Y L D H P Y Y G B

T P V W S G L V N N G R P Y A 3906 contataccotgaacgtggaatacgacgtgcogtggogcaccotc PYTLNVEYDVPWETL

IKWEHVTQKPQRYCS 3861 accceggtetggageggeetegtgaacaacgggegteeetaegee

KTVADVLLSFAPESN 3726 ggggagetggeetacegggeetaceagaacetggagaagegggte G E L A Y R A Y Q N L E K R V 3771 ggcaagcocctggccgggatcgccgcggaagccgcgacctgcgg K P L A G I A A G S R D L R 

4651 acgaaacceggegteggetaccegaagaagtgggaagaccaggag TKPGVGYPKKWEDQE ggagtecaac FKGGWVADGHNLR 4741 aaggecatggggaaggggeegaegeteeteaacatgttetteeag KANGKGPTLLNNFFQ 4786 cecaacatgeogaagetggaggactactaogageeettegactte PNMPKLEDYYEPFDF

4561 tgeateggetgecaegectgeagectetectgeaagaacatetgg CIGCHACSLSCKNIW 4606 accgaccgcaagggcgccgaatacatgtggtggaacaacgtggag TDEEGAEYMWWNNVE EADDVHEQLLETMP 3636 accordacetggaacggcaagtegtteetggacategaagacgaa RTWNGKSFLDIEDE 3601 aagacogtggcogacgtgctcctctcctcgogo

4516 atggatgtcagagcacaactggtaatggtgttcaacctggacaaa 3501 gccatogtggagcgggactacgtgaacctctacaaccgctacgtg AIVERDYVNLYNRYV 3546 toggtgggggccccatatoggccacctccatgcccacggggtcagt SVGPHIGHLHAHGVS 3591 tgggaggeggaegaegteeacgageageteetggagaecatgeeg

3456 ctcggcgagtgcgaggcgatccccggcaagaccatgccgaacctg Genel637 LGECEAIPGKTNPNL

DTPGEIAQRQVKDWK

3321 ttcaccogcatogccaagaaggtgagcgagctggcggtgaagcac FTRIAKKVSELAVKH 3366 ctocccaogcoggtgaaggatatcatogcogogcogctcctccac LPTPVKDIIAAPLLH 3411 gacaccccgggcgagatcgcccagcggcaggtgaaggactggaag

3276 gccgccgtgccgccgtcctgggagtcgaaatctgactggaagatc A V P P S W E S K S D W K I 4491 gtcagaaagctgagggggtga G&CA

DLNTTDMHSPINCLD 4401

4446 aactactggggaccgaccggcgtgaaccggggacagctacgtggtc

ccgaccetgatgetoggggggetacgeecagtteagetactactte PTLMLGGYAOFSYYF

RKLRG

A P E R T I N I Q K S K K T G K R G G V H N S L S R I R L K

VRAQLVMVFNLDK

NYWGPTGVNRDSYV

5911 5956 aggggaaaacgatga 5970 RGKR Gene1638 5967 atgacageggatatetatteatecetegeeeggatgetegaetat MTADIYSSLARMLDY 6012 ccccaggagaaaaagagcetcctgtgggccatcggcaccgtccag QEKKSLLWAIGTVQ 6057 aggtatetgagegaagagggggtggagtgececetegacaeette RYLSEEGLECPLDTF 6102 ggagagttcatggccgcatcgaccctggcggagetccaggaggas GEFNAASTLAELOEE tacgtggcgacettegattteaaccoggeggtggccccetatetg YVATFDFNFAVAFYL HHLFGDNQKKGAYI 6237 attgecctgaaggggggggtatggccgccacygctacacccccctc

IALKGEYGRHGYTPL

GNELPDHLPVILWFL

6282 ggcaacgagetteecgateateteecgteatactetggtteetg

A K R L Y R L F T I G G Y N E RNIIPPQQREELDSH accegeaaggggggeeageggetteggeatteteaagaagaegggg T R K G A S G F G I L K K T G

acggacgaggegageetegggaaggeegggetegaegaggeaggg DEASLGKAGLDEAG 5821 gegaaacgtetetacogtetetteaccateggeggetacaacgag 5866 oggaacatcattoogooccagcagoggaagagetogattoocad

5731 ategeeettegeatecacatgegggeeaagaaceteaaceageee TAL RTHM RAKNINO P

GGGNRDVVAAALEKL

GGELHELVEHGTIPT 5641 ogggaggecetgegegeeceeategggtacetggegageetette EALRAPIGYLASLF 5686 ggggggggggcaaccgggacgtggtggcggccctggagaagetc

P M A Y Y I P S L S P V L 7 S

cocatggeetactacateccetecetetegeeggttetgacetee 5596 ggoggogagetecacgagetggtggageacggeacgatecegace

5506 gagttogaactogccctgccgctccaccccgagtttoggaccatg EFELALPLHPEFRTM 5551

N L E A A E K S P V Y A L I K

gaggeggtcegggaaggggggggggaggagaatggegteeeegageag A V R E G A R K N G V P E Q 5461 tggctggaggccgccgagaagtcgccggtctatgccttgatcaag

5371 cagetogtegaggeccategggggggggtgateetegateeeegggad OLVEAHRGVILDPRD

3231 gacetgaacaccaccgacatgcacteetteatcaactgeetggac 4311 gegeeggagoggaceateaacateeagaagteaaagaacgge 5326 tacgaogeeggagaggggggegeeteetgaageeggaacaac YDAEKVEQALLKPDN

YSRKNAEVPC\* CARGGARTOCCARGAAGTAAATATGGGAARCTGCGGTAGGGGCGCTGC TTGCCGCGCCCCCCCCCCTTTAGCCACCGAATCGATGGCCTAAAGAGG TGATGCC

SIAATGGPQVALY  $10010 {\tt ttcatcgccttctacgcgagetgtatcatcctcacctggtggtat}$ FIAFYASCIILTWWY 10055tattogogtaagaacgotgaagtoocotgotaa 10087 TGCCGTTTCCCAGCAGGCATCTAACCAACAAAAGGAGAACAATGTGAAAA ANACAGETATCACCATCGCCTTCGTCGCCACTTCTGCCTTTGCCGCACAT GTTTTCGCOGCCGACGTATTTGAATTCCCCCGCATCAAT9GGTAAAGTGAC APTOCCCCATAAAATGCACCAGGAGATGCTGAAGGACTGCAAGAAGTGCC ACGARRENGGACCOGGCRAGATCARGACTTCGGCRAGGACTGGGGCCCAC ANGACCTOCAAGGGGTGCCATACCGAGCTGAAGAAAGGCCCGGTCGGCTG

GGHLADELGGARVTF 9650 tggaacttcatcatcatggccggggccgtgttoggcgtgctggcg WNFIIMAGAVFGVLA 9695 tteeteeeeeaeggeggagteggeggaacttetggggatteete FLPHGGVGGNFWGFL 9740 gecatgttcatcctcctcttcgccaccaccggcatcggcaacggc AMFILLFATTGIGNG 9785 tecacettecgcatgateccggtgatettectgaccgagcggcag STFRMIPVIFLTERQ 9830 cgcgatgcggcgggaagggggaaggaggaggaggaggagcaatc RDAAGRGKEAEEQAI 9875 aaggacgecaacaaggaggeggeegteetgggetteaceteg KDANKEAAAVLGFTS 9920 gcatttgcggcctacggcgggtttttcatccccaagagettcggc A F A A Y G G F F I P E S F G 9965 acctocattgeogocacoggtggcocccaggtggcoctctacggg

9515 cogotootgaccaagtoocagttoocoggcatcaaccoggtcaag PLLTKSOFPGINPVK Y A F L G P L V G A L I R P V 

TAGAGGAGCGAATAGAGGTAGGGAACCGCCGTGTGGAGGTACCACCGGCAT TCAAGGAGAACGC Genel640 7301 atggaaaggcacgagcagaagacggcaacggcggcaacgtggggg H E R H E Q K T A T A A T W G 7346 acccaggtcaggacggcaccatggggacccacaggtacaagcag T Q V R T R T M A T H R Y K Q

G S H Y S T F P S G R E R Q R 7076 tag 7078 AGGATGGGTGGATCCTTGAGAAAATCGAGAGGCCAGTAGAGGAGGTGGACC AGCTTGGTGAACGGCAGGAGCGCCACGATCAGGAATGCCCCTGCCGCGTGC AGCTTGAGGATGAGGGGGAAGTCGGCCAGGTACTCCACCTTCGGGTTGAAG

Y S N R L T W S A Y S T O F L 6716 gagoggaagaccetetactggggggecaaccettggcactacggg RKTLYNGANPNHYG 6761 atcatecccatcatactggcccacacccttggcgttgtggccccc IIPIILAHTLGVVAI 6806 ggtgccatgaagtcgctcttgggggaacgttgaggtcctggtcttt GAMKSLLGNVEVLVF 6851 etggagagegtggggeteggeetggggetettegegetettegge LESVGLGLGLFALFG 6896 tgegtggtgeteetgetgegtegegteaacgeteecatgetgaag VVLLLRRVNAPML) 6941 cgcgtaacctatcccgccgactgggtgctcctcattctgctgacc VTYPADWVLLILLT 6986 ttccagacgctcaccggggtctacatcagcctcttcatgcgctgg FOTLTGVYISLFMRW 7031 ggeteccattactccacettcccctcgggccgggagcgccagcgg

P Å E G I G D G V L E K N R 6552 aaaaeggetgegeateeggagtaageecatgetga 6587 KTAAHPEVSPC Gene1639 6581 atgetgaacacegtgetgaacacetttatetttgtggeeetteee HLNTVLNTFIFVALP 6626 tatgtggccctggcgatgctcatcgtcgtcaccccctaccgctat YVALAMLIVVTPYRY 6671 tactogaacogtetcacetggteggeetacteeaccagtteetg

HLVR GGEDDARRSF 6372 attgaccggcacgtattgccgggggtcgagaagetcagtgacage I D R H V L P G V E K L S D S 6417 ttcgcggcatcccgccgggattccccctggaggcccgcggtcgag FAASRRDSPWRPAVE 6462 gccytccycctcatctgcggcgaggactccgcccggggagagagcac VRLICGEDSARGEH 6507 cccgcggaagggateggcgatggcgttettgaaaaaaccgcccg

7436 geggtetgggtgatgttetecatcateggcatecegatcaaggge A V W V M P S I I G I P I K G 7481 geeetaggeetgaacgagacceagtteggeateetggeegeaace ALGLNETOFGILAAT 7526 cccatteteteeggeteattgatteggetgeegetegggatgtgg PILSGSLIRLPLGMH 7571 accgacaagtacggcggacgcattgtcttcttcattctgatgctc DKYGGRIVPPILM 7616 tgcaccatcgtcccccatctacctggtgagcgccccccggcgt. C T I V P I Y L V S A A 7 A Y cacggcgtac 7661 tggcacttteteeteetggggetetteateggtettgeeggegge WHFLLLGLFIGLAGG 7706 geetteteggtggggateagetacacegeeegetggttcaceaag A F S V G I S Y T A R W F T K 7751 gecogtcagggattegecatggggatetttggcgccgggaacgcg A R Q G F A M G I F G A G N A 7796 ggggccgcggtcaccaagtteetggccccctccctggtggtggcc GAAVTKFLAPSLVVA 7841 tacggctggcagatggtcccccaaggtctatgccgtggccatgctc YGHOMVPKVYAVAML 7886 gtgacggecatectgttttgggcatteaccttcaccgagcacage VTAILPWAFTPTEHS 7931 cacaaagtgagetecaccgtgaccatcaaagaccagetggeggec K V S S T V T I K D Q L A 7976 ctcagggacccgaaggtetggegetactgecagtactacgecate L R D P K V W R Y C Q Y Y A I 8021 gtettegggggetacgtggggetetecetetggatgaccaagtac V F G G Y V G L S L W M 7 K Y tacatcaacgagtacggettegacatcoggacogcggecetggtg Y I N E Y G P D I B T A A L V 8111 geogecatettegtgetaccetcoggegtgatecgegeectggga AAIFVLPSGVIRALG 8156 gggtggetgteggacegttteggegeecacacegteacetgetgg WLSDRPGAHTVTC 8201 gtcatgtgggtetcoggegggtgcetcetgetcetetgatecce VMHVSGGCLLLSIP 8246 cacttegatgtggttgtcaccacggtgaacggtcccgtcgcgcg H F D V V V T T V N G P V A R 8291 cacgtggggctcaacgtatgggtgttcacctcccttctcttcgtt VGLNVWVFTSLLF 8336 gteggeattgegtggggatteggeaaggeateggtgtteaageae VGIANGFGKASVFKH 8381 ISDRYPHNIGVISGI

8426 gtgggggttattggeggggttggegggtttgteeteeeateatg

GLIGGLGGPVLPIM

VLTMNTFAFTVCF

ETEGKAIATRNLN SIPALFLAFAVWMVW 8840 agegtggtetgegtgaaceteeccaacateggetteaagtaegae SVVCVNLPNTGFKYD 8885 accaacaagetettetggetgacegeetteeegegettgegggg TNKLFWLTALPALAG 8930 gogacgetecgeatettetacteetteatggteeceatetteggg ATLRIFYSFMVPIFG 8975 ggacgcoggtggacogccatcagcacggcatcgctoctgattecc GRRWTAISTASLLIP 9020 gocatggggatcggctttgcggtacgggaccccgcgaccagctac ANGIGFAVEDPATSY 9065 togaccatgetgatectggcegtectetgegggtteggeggeg STMLILAVLCGFGGG 9110 aacttogeeteeageatggeeaacateagettettetateeeaag NFASSMANISFFYFX 9155 gecctgaaggggacegecetggggetgaaegeeggeetggggaae ALKGTALGLNAGLGN 9200 ctgggggtgagegecatgeagttectegtgeegetggttatetee LGVSAMQFLVPLVIS 9245 accggcatgttcgccgccctctgcggcgatccccagcagtgcgtc GMFAALCGDPQQC 9290 atcaacoggotgaccaagectatgtogetecagaacgecogette INGVTKPMNLQNAGF 9335 gtctgggtgccgttcatcgccctctccgccctggtcgcctggttc V N V P F I A L S A L V A N F 9380 ggcatgaacgacatogocagegecaaggcatectteagtgaceag GMNDIASAKASFSDO 9425 gocatcatetteeggegeaageacaactggeteatgtgeatecte AIIFRREHNWLMCIL 9470 tacctoggcaccttoggttogttcatcggctactoggcgggggctc

YLGTFGSFIGYSAGL

8660 atgccaacggacgtattgacagcaacggaaacgaaaagagtgca MPTDVLTATETKKSA 8705 teategeacgttetgacggagtggaceceggaagaecageattte SSHVLTEWTPEDQHF 8750 tgggaaaccagggggaaggccattgccaccegcaacctctggatc 8795 tegateceggeectetteetggeetttgeggtetggatggtetgg

#### CONTRACTANCO Gene1641

8561 ctggccctgaagaaagaaaacgttcatcgcaagctgaagaccgcc LALKKENVHRKLKTA 8606 gacattotogaataa CAGOCAATAACAGOCAATTOCCATGGAA DILE \*

FGALVDLTGVRSTAF 8516 atgeteetgtaeggtteeaogtgegttteeetgttetggatgtae MLLYGSTCVSLFWMY

7391 ateteggteetgaegatgaacacettegeetteactgtetgette 8471 tteggegeeetggtggaeetgaecggtgteeggagcaecgettte

#### Figure 9: Nucleotide sequence of *nar* operon in contig135

The deduced amino acid sequences are indicated below the nucleotide sequence. The stop codon is designated by \*. The beginning of each open reading frame is designated by the gene number and an colored arrow corresponding to Figure 7. The nucleotide sequence in the box is the putative promoter region predicted by Neural Network program. Red stars indicate a putative CAP/CRP recognition-binding site predicted by IMD program. Bars above the sequence indicate putative Shine-Dalgarno sequences. Red arrows indicate the inverted repeat that can form a hairpin.

### Comparative analysis of the two putative Nar subunits and their similarities to *T. thermophilus* Nar subunits.

Sequences of NarCGHI from *nar121*, *nar135* and *T.thermophilus* were aligned using ClustalX and the percentage identity and similarity were calculated (Table 9). NarGHI from contig121 has a higher similarity to *T.thermophilus* NarGHI except NarC, in which contig 135 NarC has slightly higher similarity to *T.thermophilus*. The significantly low identity of contig135 NarI to *T.thermophilus* is very intriguing. Though this low identity did not affect its position in the phylogenetic tree, in which it is still significantly similar to *T.thermophilus* (Figure 10). The protein sequence alignment of NarC, NarG, NarH, and NarI is in Appendix VII.

## Table 9: Similarities between the putative NarC, G, H and I in contig121 and 135from GS-15 and the corresponding Nar subunits from *T.thermophilis*.

Numbers are in percentages. Abbreviations indicate the following; 121: *nar121*, 135: *nar135*, Therm: *T.thermophilis* 

		121 vs 135	121 vs Therm	135 vs Therm	All 3
NarC	% identical	30.7	27.4	28.6	13.9
	% similar	60.5	56.4	61.8	37.8
NarG % iden % simi	% identical	56.2	62.7	58.3	46.8
	% similar	82.9	85.4	84.0	74.3
NarH	% identical	52.4	59.5	51.1	43.1
	% similar	79.1	83.0	77.7	69.3
NarI	% identical	15.8	40.0	12.3	5.4
	% similar	38.5	68.8	42.3	25.0

#### Phylogenetic trees of NarG, H, J, I, K<sub>1</sub>, and K<sub>2</sub>

All trees presented here were generated based on the neighbor joining method (Figure 10). The maximum parsimony method was also used on the same data set to confirm the accuracy of the neighbor joining trees (data not shown). The putative NarG, H, and I from both contigs form a distinctive clad with *T. thermophilus*. NarG, H and I from contig121 are closer related to *T. thermophilus* compared to the same subunits from contig135. NarJ from contig135 is placed in the same clad as *Haloarcula marismortui* (archaea) rather than *T. thermophilus*. The amino acid sequence comparison of NarJ between GS-15 and *T. thermophilus* is 13.9 % identical and 34 % similar. The putative NarK<sub>1</sub> and NarK<sub>2</sub> subunits were also placed closer to *Pseudomonas aeruginosa* rather

than T. thermophilus. A previous study has reported the high similarities of NarK<sub>1</sub> and NarK<sub>2</sub> between *P. aeruginosa* and *T. thermophilus*, in which NarK<sub>1</sub> protein from both organisms were 43 % identical and 54.5 % similar, and as for NarK<sub>2</sub> proteins the similarities were 41 % identical and 51.3 % similar (59). NarK<sub>1</sub> and NarK<sub>2</sub> proteins from GS-15 show much higher similarities to *P. aeruginosa* as compared to *T. thermophilus*. The comparison between GS-15 and *P. aeruginosa* NarK<sub>1</sub> proteins were 45.5 % identical and 67 % similar, and for the NarK<sub>2</sub> proteins, they were 63 % identical and 86 % similar. Comparison between GS-15 and T. thermophilus NarK<sub>1</sub> proteins were 30 % identical and 64 % similar, and for the NarK<sub>2</sub> proteins, they were 40 % identical and 72 % similar. The comparison was also made at the DNA level in which  $narK_1$  genes from GS-15 and P. aeruginosa were 52 % identical and narK<sub>2</sub> genes were 70.5 % identical, whereas GS-15 versus T. thermophilus narK<sub>1</sub> gene showed 45.5 % identity and 55.3 % identity for narK2. The phylogenetic tree using nucleotide sequences of narK from various bacteria also indicate the significant similarity between GS-15 and *P. aeruginosa*. This is an intriguing finding because in GS-15 and T. thermophilus the two in tandem  $narK_1$  and  $narK_2$  genes are located downstream of narGHJI, whereas in P. aeruginosa the  $narK_1$ and *narK*<sub>2</sub> genes are located upstream of the corresponding *narGHJI* cluster. Yet both protein and DNA sequences of GS-15 NarK<sub>1</sub> and NarK<sub>2</sub> show higher similarities to P. aeruginosa.







С



D



Е





F

G



#### Figure 10: Phylogenetic trees of NarG, H, I, J, K<sub>1</sub>, and K<sub>2</sub>.

A: NarG tree, B: narG tree, C: NarH tree, D: NarI tree, E: NarJ tree, F: NarK<sub>1</sub> and NarK<sub>2</sub>, G: narK<sub>1</sub> and

 $narK_2$  tree

All trees presented here are generated based on the neighbor joining method. Bootstrap values are

presented in circles (as percentages of 1000 replications). Organism names are abbreviated and the name of

Nar subunit is listed after the abbreviation. "Nar" indicates protein sequence and "nar" indicates DNA

sequence. The organisms' name is listed below with their abbreviation.

**Aeropyrum:** Aeropyrum pernix **B.sub**: *Bacillus subtilis* C.crescentus: Caulobacter crescentus C.perfringens: Clostridium perfringens E. coli: Escherichia coli G.met: GS-15 Halobacterium: Halobacterium H.marismortui: Haloarcula marismortui **H.pylori**: *Helicobacter pylori* M.tuber: Mycobacterium tuberculosis P.denit: Paracoccus denitrificans **P.aerug**: *Pseudomonas aeruginosa* **P.fluor**: *Pseudomonas fluorescens* **P.aerophilum**: *Pyrobaculum aerophilum* S.ruminant: Selenomonas ruminantium S.aureus: Staphylococcus aureus **S.coeli**: *Streptomyces coelicolor* S.carnosus: Staphylococcus carnosus **T.therm**: *Thermus thermophilus* 

#### Additional analysis of *narG* gene products and putative nitrite reductases

Location of the putative enzymes

The location prediction program PSORT at Expasy website (http://us.expasy.org) was used for this analysis. NarG in contig121 is predicted to reside at the cytoplasmic inner membrane (certainty 0.153) whereas NarG in contig135 is predicted to reside in the cytoplasm (certainty 0.514).

#### Theoretical pI, and molecular weight

Theoretical pI and molecular weight of the two putative *narG* and the two nitrite reductases were determined using ProtParam at Expasy website.

Putative enzyme	Isoelectric point	Molecular weight (KDa)
NarG contig121	6.2	136
NarG contig135	6.59	134
Putative NrfA contig140	8.78	54
Putative NADH dependent nitrite reductase contig146	5.49	47

#### Partial purification of NarG from G. metallireducens

After the discovery of the two putative *nar* operons (*nar121* and *nar135*), partial purification was carried out to determine which *nar* is expressed in cells of GS-15 when grown in media containing nitrate as electron acceptor. Partial purification followed by N-terminus sequencing was planned as the strategy. However, the target NarG protein (approximately 130 KDa) was not found in the fractions that contained nitrate/nitrite reductase activity (Figure 11) followed by ion exchange column. Figure 12 shows a zymogram of nitrite reductase activity. A very faint nitrate reductase activity was also detected but was unable to be captured by the camera.



**Figure 11: Heme stain and SDS-PAGE of** *G. metallireducens* **ion exchange fractions** The gel was heme stained followed by coomassie staining. Lane 1 is the heme stain standard (horse heart cytochrome *c*), lanes 2-4 are ion exchange fractions number 33-35, and lane 5 is the SDS-PAGE standards. Samples were denatured at 95 °C for 2 min. SDS-PAGE Standards: Myosin 206 KDa,  $\beta$ -galactosidase 115 KDa, Bovine serum albumin 98 KDa, Ovalbumin 54 KDa, Carbonic anhydrase 37 KDa, Soybean trypsin inhibitor 29 KDa, Lysozyme 20 KDa



#### Figure 12: Zymogram of nitrite reductase activity

The samples loaded were ion exchange fractions number 33-35 (lanes 1-3). Arrow indicates nitrite reductase activity.

#### Solubilization of NapA from S. barnesii using lysozyme treatment

Various lysozyme treatment conditions were tested as indicated in the method section. SES-3 NapA was solubilized using 10 mM EDTA and 1 mg/ml lysozyme concentration incubated at room temperature for 60 minutes. This was the most effective combination among the variety of combinations investigated. The results show that the activity of both periplasmic and spheroplast fractions were significantly greater than that determined for the whole cell sample. This indicates a permeability problem with the methyl viologen. The results also indicate that Nap complex is tightly anchored to the membrane.

# Table 10: Summary table for lysozyme treated S. barnesii fractions' nitrate reductase activity

	Volume (ml)	Protein Volume (μl)	Activity (µmol/min)	Total activity (µmol/min)	Protein (µg/ul)	Specific activity (U/mg)
Whole cell	0.09	20	0.011	0.55	6.6	0.083
Periplasm	0.81	20	0.023	1.15	1.4	0.82
Spheroplast	1.3	20	0.063	3.15	2.4	1.31

#### Purification of NapA from S. barnesii

An attempt was made to purify NapA. Sonication was utilized as an initial step to solubilize NapA instead of lysozyme treatment. This was because of the high EDTA concentration used in lysozyme treatment. The possibility of EDTA disrupting the molybdenum cofactor in NapA was the main concern. Nitrate and nitrite reductase activity was monitored by zymogram. 60 ml of solubilized fraction (10 mg/ml) was loaded to an ion exchange column followed by size exclusion column were utilized for the purification. However, only partial purification was completed. Figure 13 shows the nitrite reductase activity detected by zymogram. Nitrate reductase activity was too faint to be captured by the camera. Figure 14 shows the corresponding fraction on a SDS-PAGE.





Figure 13: Zymogram of nitrite reductase activity

Figure 14: SDS-PAGE of *S. barnesii* size exclusion column fractions

Figure 13: *S. barnesii* size exclusion fraction that contained nitrite reductase activity was detected by zymogram. Figure 14: Size exclusion column fractions that contained nitrite reductase activity were denatured by heating in a sand bath for 2 min and run on a SDS-PAGE. Lane 1 and 2 are the fractions. Lane 3 is the SDS-PAGE molecular weight standard. SDS-PAGE Standards: Myosin 206 KDa, β-galactosidase 115 KDa, Bovine serum albumin 98 KDa, Ovalbumin 54 KDa, Carbonic anhydrase 37 KDa, Soybean trypsin inhibitor 29 KDa, Lysozyme 20 KDa

#### **Development of 96 well assay**

A 96 well assay was developed by modifying the spectrophotometric methyl viologen assay protocol (Figure 15). Several false positives were detected in some fractions collected from ion exchange column. No protein bands could be detected using SDS-PAGE from these fractions. Native gel assay also indicated there was no enzyme activity from these fractions. However, the probability of getting false positive results
was relatively low (approximately 5 %). The positive control (containing crude extract) and negative control (containing no enzyme) used in the 96 well assay did not cause any problems. Depending on the condition of the anaerobic glove box (i.e. oxygen contamination), the color of methyl viologen in the negative control did not change for at least 10 - 20 min. This indicates that development of the assay system was successful.

A titration study was carried out to investigate the sensitivity of the 96 well assay. Serial dilution was carried out using the crude extract. During this experiment it was discovered that letting the plate to sit in the anaerobic glove box for several minutes before adding methyl viologen would result in a more accurate results. Color change of the media was difficult to quantify by sight, although wells containing more proteins did have a more rapid color change. The negative control for nitrate lasted for 22 min and for nitrite it lasted for 14 min before clearing. The overall sensitivity result is summarized in Table 11.



Figure 15: 96 well assay

# Table 11: Summary of 96 well assay sensitivity

Nitrate reductase Units of activity (µmol/min)	Time taken for complete methyl viologen oxidation (min)
0.0076	14
0.015	10
Nitrite reductase Units of activity	Time taken for complete methyl viologen
(µmol/min)	oxidation (min)
(µmol/min) 0.2	oxidation (min) 2
(µmol/min) 0.2 0.4	oxidation (min) 2 1
(µmol/min) 0.2 0.4 0.6	oxidation (min)           2           1           1

### Discussion

# Physiological characterization of assimilatory nitrate and nitrite reduction in G. metallireducens and G. sulfurreducens

Both GS-15 and *G. sulfurreducens* were able to grow in media containing ammonia, nitrate, and  $N_2$  (Table 6). Due to the fact that both organisms have nitrogenase genes and the activity has been confirmed previously (8, 19), it is likely that cells grown in nitrate were utilizing nitrogen gas rather than nitrate as the nitrogen source. Whether these two organisms have the ability to assimilate is still uncertain. Based on the BLAST search carried out in this study, however, no putative Nas homologue was identified. However, in order to confirm the absence of Nas in GS-15, a mutant strain lacking nitrogenase activity is required, or degassing the medium with a gas other than  $N_2$ .

### Molybdoenzymes in G. metallireducens

Nine putative molybdoenzymes were identified in the current version of GS-15 genome draft. The presence of putative nitrate induced formate dehydrogenase and fdh-H type formate dehydrogenase genes is intriguing, because GS-15 was unable to grow in media containing formate as electron donor with nitrate or Fe (III) as electron acceptor (data not shown). Three explanations can be proposed. First, the ORFs may not encode formate dehydrogenase. It is difficult to argue this unless complete protein and nucleotide analysis using prediction programs are carried out. Second, the ORFs are indeed formate dehydrogenase genes but have and cannot be transcribed or even they can be transcribed, they are not functional. Third, the growth experiments carried out did not induce a certain

factor that initiates transcription of the genes. Positions of the two putative molybdopterin oxidoreductases (gene 269 and gene 2012) in the phylogenetic tree suggest GS-15 may posses arsenate reductases (Figure 5). Currently, there is no data that suggests GS-15 can grow on arsenate thus it will be interesting to determine this.

## Putative nar operons and their evolutionary process

Discovering the existence of the two putative nar operon in GS-15 is intriguing. Multiple membrane bound nitrate reductase isoenzymes have been discovered previously in various bacteria, though these bacteria are not strictly anaerobes like GS-15. In E. coli, one of the membrane bound nitrate reductase is expressed constitutively to ensure a smooth transition from aerobic respiration to nitrate anaerobic respiration. It is possible that GS-15 also express one of the Nar isoenzyme constitutively, but the fact that no nitrate reductase activity was found present in cells grown in Fe (III) media (58) proves this to be unlikely. Putative promoter sites were found in both operons indicating that transcription can potentially be carried out on both *nar* operons. Four putative promoter sites were found upstream of *nar135* as oppose to only one putative promoter region upstream *nar121*. A putative CAP/CRP site was found and overlaps the fourth promoter that is closest to the ATG start codon of NarC in nar135 (Figure 8). Thus it is likely that *nar135* can potentially be expressed by different stimuli (such as transcription induced by heat response sigma factor) and also is regulated CAP/CRP. One of the nar operons (*nar121*) in GS-15 lacks the putative *narJ* genes suggesting that the enzyme that it encodes may not be functional. The gene product of *narJ* is a chaperone like component required for assembling functional NarGHI complex, thus lacking the gene may result in

nonfunctional enzyme. All the functional *nar* operons that have been characterized so far contain the *narJ* gene. Although a *nar* operon lacking *narJ* has been found in *Pseudomonas stutzeri ZoBell* (51), it is possible that the gene is located nearby and is yet to be discovered. Discovery of genes encoding a putative cytochrome *c* that has only been found in *T. thermophilus* is novel. Based on the BLAST search results, GS-15 and *T. thermophilus* are the only two organisms with *narC* located upstream of *narG. T. thermophilus* with mutated *narC* was unable to generate functional Nar enzyme and respire nitrate (84).

It is tempting to conclude that GS-15 may have acquired a *nar* operon from T. thermophilus. This hypothesis can also be supported by the high similarities between narGHI from the two organisms based on the results of amino acid identity/similarity comparison and the phylogenetic trees. Also, characterization of the *nar* operon from T. thermophilus has revealed that the operon is a part of a conjugative plasmid integrated into its chromosome (60). Also, growth experiment with other strains of *Thermus* indicated no anaerobic growth in nitrate (60). Moreover, the study also indicated that the nar operon could be horizontally transferred to other strains of Thermus. To add more complexity to the discussion, the amino acid sequence of *narJ* from GS-15 and T. thermophilus showed only a 13.9 % identity and 34 % similarity. Also, narK1 and narK2 genes and their translated amino acid sequence analysis from GS-15 showed that GS-15 narK1 and narK2 are closer to P. aeruginosa rather than T. thermophilus. Another difference found between *narK* genes of GS-15 and *T. thermophilus* was that the start codon ATG of  $narK_1$  in T. thermophilus overlaps with the last codon of narI gene (59), whereas a 218 base gap is located in between *narI* stop codon and *narK*<sub>1</sub> start codon in

GS-15 (Figure 8). Taking all these findings into account, plus the phylogenetic trees results, in which all of *nar121* genes are closer to *T. thermophilus*, one can argue that *nar* operon in GS-15 was originally from *T. thermophilus*. In other words, *nar121* might have been the copy that was acquired from *T. thermophilus*, and during the evolutionary process, *nar135* was duplicated from *nar121*, and both operons were eventually modified to their current state in GS-15. Finally, it would be interesting to purify NarG from GS-15 cells grown in FWA-nitrate media and sequence the N-terminus of the purified NarG. This might help determine which *nar* operon is transcribed.

### Upstream and downstream of *nar121* and *nar135*

BLAST search results of three ORFs upstream of *nar135* indicate that these three genes may encode a membrane bound three subunit enzyme (results not shown). According to the BLAST results, gene 1634 is a putative molybdopterin oxidoreductase membrane binding subunits, gene 1633 is a putative molybdopterin oxidoreductase, iron-sulfur binding subunit, and gene 1632 is a molybdopterin with very low E values. A putative Integration host factor (IHF) binding site was found at the upstream of gene 1632. Since there is a gap in between the stop codon of the three ORF and start codon of *narC*, this putative enzyme complex is not likely to be transcribed together with *nar* operon. Down stream of *nar135* on the other hand, are gene 1642 and gene 1643 encoding molybdenum cofactor biosynthesis enzyme.

Upstream of *nar121* are two hypothetical genes (gene 625 and gene 624) then a gene encoding aconitase (gene 623). Genes located down stream are three genes encoding hypothetical proteins (gene 630-632) then NtrC and NtrB (gene633-4). Overall,

no Fnr or NarXL like genes were identified upstream or downstream of *nar121* and *nar135*. Though BLAST search using amino acid sequence of NarX and NarL from *E*. *coli* did generate a dozen of positive hits. It will be interesting to find out whether any of these putative two component response sensor regulators is involved in nitrate reduction in GS-15.

### Solubilization of NapA utilizing lysozyme treatment and 96 wells assay development

Requirement of high EDTA concentration to solubilize NapA may infer an involvement of a metallic binding between NapA to the periplasmic membrane. Although lysozyme treatment seems to be successful, breakage of spheroplasts needs to be investigated. Freezing and thawing was also utilized to solubilize NapA with no success. Longer sonication time may be an alternative to lysozyme treatment. It was discovered that 15 sonication cycles (30 seconds on 30 seconds off) was an effective way to solubilize NapA. The reason is likely due to the heat that causes the dissociation of the Nap complex. Additional experiments are required to confirm the effectiveness of longer sonication time.

The success of 96 well assay development enabled the screening of many samples in a short time. However, from the fact that it sometimes gives false positive results, the assay should be used in conjunction with other assay methods.

## Conclusion

Based on the growth experiment of GS-15 and *G. Sulfurreducens* in media containing different nitrogen source, assimilation of nitrate was unable to be confirmed due to the nitrogenase activity present in both organisms.

Nine putative molybdoenzymes were identified in the GS-15 draft genome using BLAST search. The BLAST analysis also identified putative Nar, Nrf, Nir, and Nif genes. Two putative *nar* operons were identified with *nar121* consists of four genes and *nar135* consists of seven genes. A novel cytochrome *c* was identified immediately upstream of *narG* in both *nar* operons. This novel cytochrome *c* has only been identified in *T. thermophilus* previously. Protein sequence comparison between *nar* operons from GS-15 and *T. thermophilus* exhibited an interesting profile. Overall, *nar121* exhibited a higher identity and similarity to *T. thermophilus* rather than *nar135*. This was also consistent with the phylogenetic tree analysis. Protein and DNA sequence of putative *narK*<sub>1</sub> and *narK*<sub>2</sub> from *nar135* showed a higher similarity to *P. aeruginosa* rather than *T. thermophilus* based on phylogenetic tree analysis. A hypothesis that *nar* operons in GS-15 was originally from *T. thermophilus* was proposed.

Partial purification of NarG from GS-15 was attempted to investigate which *nar* operon is transcribed. However, only partial purification was completed. Thus cloning *narG* and over expressing the gene in *E. coli* may be an alternative to the direct purification of NarG from GS-15.

Purification of NapA from SES-3 was attempted but only partial purification was completed. However, cell yield was increased by successfully growing SES-3 in a fermenter. Also, it was determined that lysozyme treatment can solubilize NapA. Thus

NapA is likely a periplasmic protein as predicted. During the course of purification, a new assay system using 96 well assay plate was developed to assist a rapid screening of fractions collected from ion exchange and size exclusion column.

# Appendix I

Energy	E	Omercian
Enzyme	Enzyme subunit	Organism
Periplasmic nitrate reductase	NapA	Desulfuro Vibrio
Periplasmic nitrate reductase	NapA	Klebsiella Pneumoniae
Periplasmic nitrate reductase	NapB	E. coli
Periplasmic nitrate reductase	NapC	E. coli
Periplasmic nitrate reductase	NapF	E. coli
Membrane bound nitrate reductase	NarG	E. coli
Membrane bound nitrate reductase	NarH	E. coli
Membrane bound nitrate reductase	NarI	E. coli
Membrane bound nitrate reductase	NarJ	E. coli
Membrane bound nitrate reductase	NarK	E. coli
Assimilatory nitrate/nitrite reductase	NasA	Klebsiella oxytoca
Assimilatory nitrate/nitrite reductase	Nas	klebsiella pneumoniae
Assimilatory nitrate/nitrite reductase	NasA	Vibrio diazotrophicus
Assimilatory nitrate/nitrite reductase	NasA	Azospirillum brasilense
Assimilatory nitrate/nitrite reductase	NasA	Psudoalteromonas

Assimilatory nitrate/nitrite reductase	NasB	Klebsiella oxytoca
Assimilatory nitrate/nitrite reductase	NasC	Klebsiella oxytoca
		, , , , , , , , , , , , , , , , , , ,
Assimilatory nitrate/nitrite reductase	NasD	Klebsiella ovytoca
Assimilatory inflate/inflate reductase	Tust	Reosena oxyloca
A	NE	
Assimilatory intrate/intrite reductase	INASE	Kledstella oxytoca
Assimilatory nitrate/nitrite reductase	NasF	Klebsiella oxytoca
Assimilatory nitrate/nitrite reductase	NasA	Bacillus subtilis
Assimilatory nitrate/nitrite reductase	NasB	Bacillus subtilis
Assimilatory nitrate/nitrite reductase	NasC	Bacillus subtilis
Assimilatory nitrate/nitrite reductase	NasD	Bacillus subtilis
A ssimilatomy nitrate /nitrite reductors	NegE	Davillus subtilis
Assimilatory intrate/intrite reductase	INASE	Bactitus subitits
Assimilatory nitrate/nitrite reductase	NarB	Synechococcus sp
Assimilatory nitrate/nitrite reductase	NirA	Synechococcus sp
Periplasmic nitrite reductase	NrfA	E. coli
Periplasmic nitrite reductase	NrfB	E. coli
Periplasmic nitrite reductase	NrfC	E. coli
Perinlasmic nitrita raductasa	NrfD	F. coli
r emphasime mune reductase		L. con
Periplasmic nitrite reductase	NrfE	E. coli

Periplasmic nitrite reductase	NrfF	E. coli
Periplasmic nitrite reductase	NrfG	E. coli
Periplasmic nitrite reductase	NrfA	Sulfurospirillum deleyianum
Periplasmic nitrite reductase	NrfH	Sulfurospirillum deleyianum
Cytoplasmic dissimilatory nitrite	NirB	E. coli
Cytoplasmic dissimilatory nitrite	NirC	E. coli
Cytoplasmic dissimilatory nitrite	NirD	E. coli
Cytoplasmic dissimilatory nitrite	CysG	E. coli
Cytochrome <i>cd</i> 1 nitrite reductase	NirS	Paracoccus denitrificans
Cytochrome <i>cd</i> 1 nitrite reductase	NirI	Paracoccus denitrificans
Cytochrome <i>cd</i> 1 nitrite reductase	NirE	Paracoccus denitrificans
Cytochrome <i>cd</i> 1 nitrite reductase	NirT	Pseudomonas
Copper nitrite reductase	NirK	Bladyrhizobium japonicum
Nitrogenase	NıfA	Klebsiella Pneumoniae
NT-		
Nitrogenase	NifD	Klebsiella Pneumoniae
Nitera	NT: CTT	Kinter II a D
INITrogenase	NIIH	Kiedsiella Pheumoniae
Nitera	NT: CTZ	Kinter II a D
INitrogenase	INITK	Klebsiella Pneumoniae

# Appendix II

Distilled Water	980 ml
Ammonium Chloride	0.225 g
Sodium Phosphate (Monobasi	0.6 g
Potassium Chloride	0.1 g
Sodium Acetate	6.8 g
Sodium Bicarbonate	2.5 g
Vitamin Mix	10 ml
Trace Elements	10 ml
Electron acceptor	
Ferric Citrate	13.7 g(50mmol)
Sodium Nitrate	1.7 g (20mmol)
Vitamin Mix	
Biotin	2 mg/L
Folic Acid	2  mg/L
Pyridoxine HCl	10 mg/L
Riboflavin	5 mg/L
Thiamine	5 mg/L
Nicotinic Acid	5 mg/L
Pantothenic Acid	5 mg/L
p-Aminobenzoic Acid	5 mg/L
Thioctoc Acid	5 mg/L
B12	0.1 mg/L
Mineral Mix	
Nitrilotriacetic Acid	1.5 g/L
Magnesium Sulfate	3.0 g/L
Manganous Chloride	0.444 g/L
Sodium Chloride	1.0 g/L
Ferric Chloride	0.067 g/L
Calcium Chloride; Dihydrate	0.1 g/L
Cobalt Chloride	0.1 g/L
Zinc Sulfate 7-Hydrate	0.274 g/L
Cupric Sulfate 5-Hydrate	0.01 g/L
Aluminum Potassium Sulfate	0.01 g/L
Boric Acid; Granular	0.01 g/L
Sodium Molybdate	0.025 g/L
Nickelous Chloride 6-Hydrate	0.024 g/L
Sodium Tungstate; Hydrate	0.025 g/L

# Appendix III

Washing buffer (1 liter) Tris-HCl EDTA Protease inhibitor Distilled Water	10 mM 1 mM 1:50 volume ratio 985 ml
Native gel running buffer Glycin Tris-HCl H2O	45 g 9 g 2000 ml
Native gel sample buffer Tris-HCl PH6.8 0.5M Glycerol Bromophenol Blue	6 ml 4 ml Few crystal
Buffer A Tris-HCl EDTA PMSF Distilled Water	10 mM 1 mM 10 μM 985 ml

# Appendix IV

## Mineral salt media 980 ml Distilled water

980 ml	Distilled water
0.225 g	K <sub>2</sub> HPO <sub>4</sub>
0.225 g	$KH_2PO_4$
0.46 g	NaCl
0.225 g	$(NH_4)_2SO_4$
0.117 g	MgSO <sub>4</sub>
1 g	Yeast extract
2.24 g	Sodium lactate
4.2 g	NaHCO <sub>3</sub>
0.5 g	cysteine
10 ml	trace elements

10 ml	vitamine mix
10 1111	

# Appendix V

## Molybdoenzymes BLAST results

Contig108gene137

Sec	quences pr	odu	cing signific	ant	alignment	s:		(bits)	Value
<u>gi</u>	23055047	gb	ZP_00081176.1	l h	ypothetic	al protein	[Geobact	1123	0.0
gi gi	21226223	ref	NP_632145.1	Ca	arbon monc	xide dehyd	rogenase	349	4e-97 8e-95
gi	23114032	ref	ZP_00099357.	1	hypotheti	cal protei	n [Desulf	<u>345</u>	1e-93
gi	20092097	ref	NP_618172.1	Ca	arbon-monc	xide dehyd	rogenase,	341	2e-92
gi	23473779	ref	ZP_00129074.	1	hypotheti	cal protei	n [Desulf	325	1e-87
gi	23103291	ref	ZP_00089776.	1	hypotheti	cal protei	n [Azotob	310	4e-83
gi	23111777	ref	ZP_00097361.	1	hypotheti	cal protei	n [Desulf	304	3e-81
gi	15826756	pdb	1JJY A Chai	n A,	Crystal	Structure	Of A Ni-C	304	3e-81
gi	11095246	gb	AAG29809.1 AF	2498	<u>99_2</u> car	bon monoxi	de dehydr	299	1e-79

Score

Score

Score

Е

(bits) Value

Е

Е

Contig140gene2111

Sequences producing significant alignments:

gi	23054053	gb	ZP_00080232.1	hypothetical protein [Geobact	363	e-100
gi	16605575	emb	CAC86667.1	formate dehydrogenase alpha sub	182	2e-45
gi	28948357	pdb	1HOH A Chain	A, Tungsten Containing Formate	177	9e-44
gi	23474797	ref	ZP_00130089.1	hypothetical protein [Desulf	165	4e-40
gi	23473410	ref	ZP_00128706.1	hypothetical protein [Desulf	164	6e-40
gi	23474726	ref	ZP_00130018.1	hypothetical protein [Desulf	162	2e-39
gi	23112916	ref	ZP_00098340.1	hypothetical protein [Desulf	160	1e-38
gi	24113106	ref	NP_707616.1	formate dehydrogenase-N, nitra	133	2e-30
gi	26247761	ref	NP_753801.1	Formate dehydrogenase, nitrate	132	4e-30
gi	16129433	ref	NP_415991.1	formate dehydrogenase-N, nitra	132	4e-30
gi	15801663	ref	NP_287681.1	formate dehydrogenase-N, nitra	131	4e-30

Contig140gene2112

Sequences producing significant alignments:	Score (bits)	E Value
gi 23054054 gb ZP_00080233.1 hypothetical protein [Geobact	1640	0.0
gi 23474796 ref ZP_00130088.1 hypothetical protein [Desulf	834	0.0
gi 16605575 emb CAC86667.1 formate dehydrogenase alpha sub	810	0.0
gi 28948357 pdb 1H0H A Chain A, Tungsten Containing Formate	806	0.0
gi 23474726 ref ZP_00130018.1 hypothetical protein [Desulf	803	0.0
gi 23473409 ref ZP_00128705.1 hypothetical protein [Desulf	798	0.0
gi 16129433 ref NP_415991.1 formate dehydrogenase-N, nitra	733	0.0
gi 2506864 sp P24183 FDNG_ECOLI Formate dehydrogenase, nitr	733	0.0
gi 15801663 ref NP_287681.1 formate dehydrogenase-N, nitra	733	0.0
gi 20150976 pdb 1KQF A Chain A, Formate Dehydrogenase N Fro	732	0.0

Contig119gene498

Sequences producing significant alignments:	(bits)	Value
gi 23054608 gb ZP_00080755.1 hypothetical protein [Geobact	1546	0.0
gi 14250944 emb CAC39239.1 FdhA-II protein [Eubacterium ac	454	e-126
gi 14717798 gb AAB18330.2 formate dehydrogenase alpha subu	449	e-124
gi 16263281 ref NP_436074.1 NuoG2 NADH I CHAIN G 2 [Sinor	<u>416</u>	e-115
gi 14250938 emb CAC39234.1 FdhA-I protein [Eubacterium aci	415	e-114
gi 5650750 emb CAB51635.1 putative NADH-ubiquinone oxidore	414	e-114
gi 29375957 ref NP_815111.1 NAD-dependent formate dehydrog	<u>409</u>	e-112
gi 15679548 ref NP_276665.1 formate dehydrogenase, alpha s	401	e-110
gi 1361902 pir S57430 probable formate dehydrogenase (EC 1	386	e-105
<u>gi 26988909 ref NP_744334.1 </u> formate dehydrogenase, alpha s	<u>379</u>	e-103

#### Contig148gene3285

0000191109000200	Casho	E.
Sequences producing significant alignments:	(bits)	Value
gi 14250944 emb CAC39239.1 FdhA-II protein [Eubacterium ac	<u>556</u>	e-157
gi 14250938 emb CAC39234.1 FdhA-I protein [Eubacterium aci	527	e-148
gi 14717798 gb AAB18330.2 formate dehydrogenase alpha subu	520	e-146
gi 29375957 ref NP_815111.1 NAD-dependent formate dehydrog	502	e-141
gi 23054608 gb ZP_00080755.1 hypothetical protein [Geobact	500	e-140
gi 15679548 ref NP_276665.1 formate dehydrogenase, alpha s	478	e-133
gi 1361902 pir S57430 probable formate dehydrogenase (EC 1	451	e-125
gi 16263281 ref NP_436074.1 NuoG2 NADH I CHAIN G 2 [Sinor	442	e-122
gi 26988909 ref NP_744334.1 formate dehydrogenase, alpha s	417	e-115
gi 22987538 ref ZP_00032616.1 hypothetical protein [Burkho	409	e-112

Contig113gene269

See	quences pi	it alignments: (bits	) Value		
gi	23055700	gb	ZP_00081789.1	hypothetical protein [Geobact 1837	0.0
gi	23115051	ref	ZP_00100326.1	hypothetical protein [Desulf 479	e-134
gi	23115047	ref	ZP_00100322.1	hypothetical protein [Desulf 413	e-114
gi	11498802	ref	NP_070031.1	molybdopterin oxidoreductase, <u>339</u>	1e-91
gi	23113320	ref	ZP_00098706.1	hypothetical protein [Desulf <u>337</u>	6e-91
gi	23115035	ref	ZP_00100310.1	hypothetical protein [Desulf <u>327</u>	5e-88
gi	23114874	ref	ZP_00100154.1	hypothetical protein [Desulf 326	1e-87
gi	23113724	ref	ZP_00099076.1	hypothetical protein [Desulf 309	1e-82
gi	23115033	ref	ZP_00100308.1	hypothetical protein [Desulf 309	2e-82
gi	23113735	ref	ZP_00099086.1	hypothetical protein [Desulf 296	2e-78

Score

Score

Е

Е

Contig139gene2012

Sec	quences producing si	gnifican	t alignments:	(bits)	Value
gi	23053111 gb ZP_0007	9320.1	hypothetical protein [Geobact	<u>1419</u>	0.0
gi	23014021 gb ZP_0005	3862.1	hypothetical protein [Magneto	216	9e-55
gi	23135169 ref ZP_001	16925.1	hypothetical protein [Cytoph	202	1e-50
gi	23473865 ref ZP_001	29160.1	hypothetical protein [Desulf	192	1e-47
gi	24215966 ref NP_713	<u>447.1</u> 1	molybdopterin oxidoreductase,	182	2e-44
gi	11498802 ref NP_070	031.1 1	molybdopterin oxidoreductase,	179	1e-43
gi	23474822 ref ZP_001	30114.1	hypothetical protein [Desulf	176	1e-42
gi	22970536 gb ZP_0001	7603.1	hypothetical protein [Chlorof	166	1e-39
gi	24375549 ref NP_719	592.1	polysulfide reductase, subunit	161	4e-38
gi	23112121 ref ZP_000	97650.1	hypothetical protein [Desulf	160	9e-38

Contig121gene627

	Saoro	r
Sequences producing significant alignments:	(bits)	⊾ Value
gi[23054016 ref[ZP_00080195.1] hypothetical protein [Geobac	2337	0.0
gi 2105093 emb CAA71210.1 nitrate reductase alpha subunit	1481	0.0
gi 23056188 ref ZP_00082242.1 hypothetical protein [Geobac	1337	0.0
gi 14601314 ref NP_147849.1 nitrate reductase alpha chain	1177	0.0
gi 30020253 ref NP_831884.1 Respiratory nitrate reductase	1129	0.0
gi 21399996 ref NP_655981.1 molybdopterin, Molybdopterin o	1128	0.0
gi 1009366 emb CAA62926.1 Respiratory nitrate reductase [B	1111	0.0
gi 16080781 ref NP_391609.1 nitrate reductase (alpha subun	1108	0.0
gi 21698999 dbj BAC02732.1 nitrate reductase alpha subunit	1103	0.0
gi 18314193 ref NP_560860.1 nitrate reductase alpha subuni	1102	0.0

Contig135gene1636

	Score	E
Sequences producing significant alignments:	(bits)	Value
gi 23056188 ref ZP_00082242.1  hypothetical protein [Geobac	2476	0.0
gi 2105093 emb CAA71210.1 nitrate reductase alpha subunit	1418	0.0
gi 23054016 ref ZP_00080195.1 hypothetical protein [Geobac	1365	0.0
gi 14601314 ref NP_147849.1 nitrate reductase alpha chain	1261	0.0

gi	21399996 ref NP_655981.1	molybdopterin, Molybdopterin o	1237	0.0
gi	30020253 ref NP_831884.1	Respiratory nitrate reductase	1237	0.0
gi	1009366 emb CAA62926.1	Respiratory nitrate reductase [B	1212	0.0
gi	16080781 ref NP_391609.1	nitrate reductase (alpha subun	1210	0.0
gi	21698999 dbj BAC02732.1	nitrate reductase alpha subunit	1206	0.0
gi	18314193 ref NP_560860.1	nitrate reductase alpha subuni	1185	0.0

### nar121, 135 operon Blast results

Contig121gene626

(bits) Value Sequences producing significant alignments: gi23054015gbZP\_00080194.1(NZ\_AAAS01000003)hypothetica...gi23056187gbZP\_00082241.1(NZ\_AAAS01000025)hypothetica... <u>46</u>6 e-130 130 2e-29 gi 23058888 gb ZP\_00083905.1 (NZ\_AABA01000091) hypothetica... 94 59 57 2e-18 gi 23135168 gb ZP\_00116924.1 (NZ\_AABE01000014) hypothetica... 9e-08 gi 23136139 gb ZP\_00117866.1 (V10124) periplasmic cytochrome... gi 23136139 gb ZP\_00117866.1 (NZ\_AABE01000043) hypothetica... 2e-07 54 3e-06 <u>49</u> 44 gi 23136218 gb ZP\_00117943.1 (NZ\_AABE01000045) hypothetica... 7e-05 gi 575386 emb CAA57910.1 (X82563) c-type cytochrome [Synec... gi 17549783 ref NP\_523123.1 (NC\_003296) PROBABLE CYTOCHROM... 0.003 41 0.015 41 gi 23055766 gb ZP\_00081850.1 (NZ\_AAAS01000016) hypothetica... 0.022

Е

Score

Score

Е

Contig121gene627

Sequences producing significant alignments:	(bits)	Value
<pre>gi 23054016 ref ZP_00080195.1 hypothetical protein [Geobac gi 2105093 emb CAA71210.1 nitrate reductase alpha subunit gi 23056188 ref ZP_00082242.1 hypothetical protein [Geobac gi 14601314 ref NP_147849.1 nitrate reductase alpha chain gi 30020253 ref NP_831884.1 Respiratory nitrate reductase gi 12399996 ref NP_655981.1 molybdopterin, Molybdopterin o gi 1009366 emb CAA62926.1 Respiratory nitrate reductase [B gi 16080781 ref NP_391609.1 nitrate reductase (alpha subun gi 18314193 ref NP_560860.1 nitrate reductase alpha subuni</pre>	$\begin{array}{r} 2337\\ 1481\\ 1337\\ 1177\\ 1129\\ 1128\\ 1111\\ 1108\\ 1103\\ 1102\\ \end{array}$	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
Contig121gene628		
Sequences producing significant alignments:	Score (bite)	E
	(DICS)	varue

Contig121gene629

	Score	Е
Sequences producing significant alignments:	(bits)	Value
<u>gi 23054018 gb ZP_00080197.1</u> (NZ_AAAS01000003) hypothetica	<u>359</u>	2e-98
gi 23056191 gb ZP_00082245.1 (NZ_AAAS01000025) hypothetica	101	1e-20
gi 2105096 emb CAA71213.1 (Y10124) nitrate reductase gamma	100	1e-20
gi 21699002 dbj BAC02735.1 (AB087407) nitrate reductase ga	<u>59</u>	7e-08
<u>gi 21399999 ref NP_655984.1</u> (NC_003995) Nitrate_red_gam, N	57	2e-07
gi 3929525 gb AAC82545.1 (AF029224) NarI [Staphylococcus c	57	2e-07
gi 15925384 ref NP_372918.1 (NC_002758) nitrate reductase	56	4e-07
gi 15599067 ref NP 252561.1 (NC 002516) respiratory nitrat	55	1e-06

gi 1	6760125	ref  NP_	455742.1	(NC_003198)	respiratory nitrat	54	2e-06
gi 1	6765102	ref NP_	460717.1	(NC_003197)	nitrate reductase	54	2e-06

Contig135gene1635

Sequences producing significant a	Sco: (bits)	re E Value	
<u>gi 23056187 gb ZP_00082241.1 </u> (N	<pre>Z_AAAS01000025) hypothetica</pre> <pre>Z_AAAS01000003) hypothetica</pre>	<u>444</u>	e-124
gi 23054015 gb ZP_00080194.1  (N		113	3e-24
gi 15594131 emb CAC69865.1 (Y10	124) periplasmic cytochrome	62	8e-09
gi 23058888 gb ZP 00083905.1 (N	Z AABA01000091) hypothetica	41	0.014
gi 23136139 gb ZP_00117866.1 (N	Z_AABE01000043) hypothetica	33	2.4

Contig135gene1636

		Sco	re E
Sec	quences producing significant alignments:	(bits)	Value
gi	23056188 ref ZP_00082242.1 hypothetical protein [Geobac	2476	0.0
gi	2105093 emb CAA71210.1 nitrate reductase alpha subunit	1418	0.0
gi	23054016 ref ZP_00080195.1 hypothetical protein [Geobac	1365	0.0
gi	14601314 ref NP_147849.1 nitrate reductase alpha chain	1261	0.0
gi	21399996 ref NP_655981.1 molybdopterin, Molybdopterin o	1237	0.0
gi	30020253 ref NP_831884.1 Respiratory nitrate reductase	1237	0.0
gi	1009366 emb CAA62926.1 Respiratory nitrate reductase [B	1212	0.0
gi	16080781 ref NP_391609.1 nitrate reductase (alpha subun	1210	0.0
gi	21698999 dbj BAC02732.1 nitrate reductase alpha subunit	1206	0.0
gi	18314193 ref NP_560860.1 nitrate reductase alpha subuni	1185	0.0

Contig135gene1637

Sequences producing significant alignments:

gi	23056189 gb ZP_00082243.1 (NZ_AAAS01000025) hypothetica	1029	0.0
gi	97384 pir S21535 nitrate reductase (EC 1.7.99.4) beta c	542	e-153
gi	23054017 gb ZP_00080196.1 (NZ_AAAS01000003) hypothetica	<u>540</u>	e-152
gi	2105094 emb CAA71211.1 (Y10124) nitrate reductase beta	535	e-151
gi	17989296 ref NP_541929.1 (NC_003318) NITRATE REDUCTASE	513	e-144
gi	15599069 ref NP_252563.1 (NC_002516) respiratory nitrat	511	e-144
gi	16129188 ref NP_415743.1 (NC_000913) nitrate reductase	511	e-143
gi	15801456 ref NP_287473.1 (NC_002655) nitrate reductase	509	e-143
gi	22978247 gb ZP_00024009.1 (NZ_AAAI01000312) hypothetica	508	e-143
gi	21284047 ref NP_647135.1 (NC_003923) nitrate reductase	507	e-142

Score

(bits) Value

Score E

Е

Contig135gene1638

Sequences producing significant alignments:	(bits)	Value
gi[23056190]gb[ZP_00082244_1] (NZ_AAAS01000025) hypothetica	419	e-117
<u>gi 17549197 ref NP_522537.1</u> (NC_003296) PUTATIVE RESPIRATO	69	3e-11
gi 22978246 gb ZP_00024008.1 (NZ_AAAI01000312) hypothetica	67	1e-10
gi 21699001 dbj BAC02734.1 (AB087407) nitrate reductase pr	64	9e-10
<u>gi 16080779 ref NP_391607.1 </u> (NC_000964) nitrate reductase	<u>59</u>	4e-08
gi 1075873 pir S52917 nitrate reductase delta chain - Baci	59	5e-08
<u>gi 16765103 ref NP_460718.1</u> (NC_003197) nitrate reductase	57	1e-07
gi 2105095 emb CAA71212.1 (Y10124) chaperone protein [Ther	57	1e-07
gi 2147971 pir  S61308 nitrate reductase assembly protein n	57	2e-07
<u>gi 16760124 ref NP_455741.1 </u> (NC_003198) respiratory nitrat	_57	2e-07

Contig135gene1639

Sequences producing significant alignments:	Score (bits)	E Value
<u>gi 23056191 gb ZP_00082245.1</u> (NZ_AAAS01000025) hypothetica	353	9e-97
<u>gi 23054018 gb ZP_00080197.1 </u> (NZ_AAAS01000003) hypothetica	<u>111</u>	9e-24
gi 2105096 emb CAA71213.1 (Y10124) nitrate reductase gamma	<u>    96</u>	4e-19
<u>gi 21224834 ref NP_630613.1 </u> (NC_003888) putative nitrate r	95	7e-19
<u>gi 21399999 ref NP_655984.1</u> (NC_003995) Nitrate_red_gam, N	<u>91</u>	9e-18
<u>gi 3929525 gb AAC82545.1</u> (AF029224) NarI [Staphylococcus c	88	1e-16

gi 1955241	0 ref NP_600412.3	(NC_003450) COG2181:Nitrate re	85	7e-16
gi 1592538	4 ref NP_372918.3	(NC_002758) nitrate reductase	84	2e-15
gi 2121877	3 ref NP_624552.3	(NC_003888) putative nitrate r	78	7e-14
gi 1608077	8 ref NP_391606.	(NC_000964) nitrate reductase	77	1e-13

Contig135gene1640

See	quences producing significant alignments:	Score (bits)	E Value
gi	23056192 gb ZP_00082246.1 (NZ_AAAS01000025) hypothetica	914	0.0
gi	11344599 gb AAG34371.1 AF197465_3 (AF197465) putative ni	436	e-121
gi	22978250 gb ZP_00024012.1 (NZ_AAAI01000312) hypothetica	434	e-120
gi	17549193 ref NP_522533.1 (NC_003296) PUTATIVE NITRITE/N	412	e-114
gi	15599072 ref NP_252566.1 (NC_002516) nitrite extrusion	391	e-107
gi	17989293 ref NP_541926.1 (NC_003318) NITRITE EXTRUSION	308	2e-82
gi	18413616 dbj BAB84307.1 (AB076402) putative nitrate tra	279	7e-74
gi	14348603 gb AAK61313.1 AF295359_4 (AF295359) NarK [Parac	269	6e-71
gi	16264170 ref NP_436962.1 (NC_003078) putative nitrate t	256	8e-67
gi	6689412 emb CAB65479.1 (Y10124) NarK1 protein [Thermus	246	4e-64

Contig135gene1641

		Score	Е
Se	quences producing significant alignments:	(bits)	Value
gı	23056193 gb 2P_00082247.1 (NZ_AAAS01000025) hypothetica	804	0.0
gi	11344600 gb AAG34372.1 AF197465_4 (AF197465) putative ni	<u>527</u>	e-148
gi	22978249 gb ZP_00024011.1 (NZ_AAAI01000312) hypothetica	518	e-146
gi	17549194 ref NP_522534.1 (NC_003296) PROBABLE NITRITE/N	<u>499</u>	e-140
gi	15599071 ref NP_252565.1 (NC_002516) nitrite extrusion	<u>491</u>	e-138
gi	2765453 emb CAA75539.1 (Y15252) nitrate extrusion prote	<u>491</u>	e-138
gi	17989293 ref NP_541926.1 (NC_003318) NITRITE EXTRUSION	<u>469</u>	e-131
gi	<u> 14348603 gb AAK61313.1 AF295359_4</u> (AF295359) NarK [Parac	<u>432</u>	e-120
gi	16765106 ref NP_460721.1 (NC_003197) MFS superfamily, n	400	e-110
gi	<u>16760121 ref NP_455738.1</u> (NC_003198) nitrite extrusion	<u>399</u>	e-110

## **Appendix VI**

### Transmembrane prediction for NarI in nar121

TMpred output for NarI contig121 [ISREC-Server] Date: Sat Jul 26 0:02:37 MET 2003

Sequence: MPV...PRK, length: 260 Prediction parameters: TM-helix length between 17 and 33

\_\_\_\_\_

#### 1.) Possible transmembrane helices

The sequence positions in brackets denominate the core region. Only scores above 500 are considered significant.

Insid	le	to oi	utsid	le	helice	es :	6 found
	t	Erom			to	score	center
36	(	36)	55	(	53)	1990	45
88	(	88)	105	(	105)	2021	96
120	(	120)	138	(	138)	1987	128
146	(	150)	170	(	167)	2178	159
177	(	177)	197	(	195)	981	185
212	(	214)	232	(	230)	1615	222
Outsi	lde	e to :	insid	le	helice	es :	6 found
	t	Erom			to	score	center
36	(	36)	58	(	54)	2260	45
86	(	86)	103	(	103)	2138	95
120	(	120)	138	(	138)	2475	128
146	(	148)	173	(	165)	1886	156
172	(	175)	193	(	193)	1174	183
212	(	214)	230	(	230)	1487	222

\_\_\_\_\_

#### 2.) Table of correspondences

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.

Helices shown in brackets are considered insignificant. A "+"-symbol indicates a preference of this orientation. A "++"-symbol indicates a strong preference of this orientation.

	Ì	inside	e->out	side	outside	e->ir	nside		
36-	55	(20)	1990		36-	58	(23)	2260	++
88-	105	(18)	2021		86-	103	(18)	2138	+
120-	138	(19)	1987		120-	138	(19)	2475	++
146-	170	(25)	2178	++	146-	173	(28)	1886	
177-	197	(21)	981		172-	193	(22)	1174	+
212-	232	(21)	1615	+	212-	230	(19)	1487	

------

#### 3.) Suggested models for transmembrane topology

These suggestions are purely speculative and should be used with **extreme caution** since they are based on the assumption that all transmembrane helices have been found. In most cases, the Correspondence Table shown above or the prediction plot that is also created should be

used for the topology assignment of unknown proteins.

2 possible models considered, only significant TM-segments used

```
----> STRONGLY prefered model: N-terminus outside
6 strong transmembrane helices, total score : 11723
# from
        to length score orientation
    36
         58 (23)
                   2260 o-i
1
2
   88 105 (18)
                   2021 i-o
3 120 138 (19)
                   2475 o-i
4 146 170 (25)
                   2178 i-o
5
   172
        193 (22)
                   1174 o-i
6 212 232 (21)
                   1615 i-o
----> alternative model
6 strong transmembrane helices, total score : 10469
        to length score orientation
# from
    36
         55 (20) 1990 i-o
1
2
    86 103 (18)
                   2138 o-i
3
   120 138 (19)
                   1987 i-o
4 146 173 (28)
                  1886 o-i
5 177 197 (21)
                   981 i-o
6 212 230 (19)
                  1487 o-i
```



## Transmembrane prediction for NarI in nar135

TMpred output for NarI contig135 [ISREC-Server] Date: Sat Jul 26 0:05:27 MET 2003

Sequence: MLN...KVE, length: 231 Prediction parameters: TM-helix length between 17 and 33

#### **1.)** Possible transmembrane helices

The sequence positions in brackets denominate the core region. Only scores above 500 are considered significant.

Insid	le	to ou	utsid	le	helice	s:	5	found
	t	Erom			to	score	Ce	enter
5	(	5)	24	(	24)	2070		14
59	(	59)	78	(	76)	1158		67
93	(	93)	111	(	111)	1990		101
128	(	128)	148	(	148)	2021		138
185	(	187)	203	(	203)	1593		195
Outsi	ide	e to :	insid	le	helice	s:	6	found

-----

	t	Erom			to	score	center
9	(	13)	32	(	32)	2224	21
59	(	59)	76	(	76)	1307	68
93	(	93)	111	(	109)	2478	101
128	(	128)	146	(	146)	1803	138
152	(	152)	169	(	169)	438	160
185	(	187)	203	(	203)	1715	195

#### 2.) Table of correspondences

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.

Helices shown in brackets are considered insignificant. A "+"-symbol indicates a preference of this orientation. A "++"-symbol indicates a strong preference of this orientation.

\_\_\_\_\_

```
inside->outside | outside->inside

5- 24 (20) 2070 | 9- 32 (24) 2224 +

59- 78 (20) 1158 | 59- 76 (18) 1307 +

93- 111 (19) 1990 | 93- 111 (19) 2478 ++

128- 148 (21) 2021 ++ | 128- 146 (19) 1803

| ( 152- 169 (18) 438 ++ )

185- 203 (19) 1593 | 185- 203 (19) 1715 +
```

#### **3.)** Suggested models for transmembrane topology

These suggestions are purely speculative and should be used with **extreme caution** since they are based on the assumption that all transmembrane helices have been found.

In most cases, the Correspondence Table shown above or the prediction plot that is also created should be used for the topology assignment of unknown proteins.

```
2 possible models considered, only significant TM-segments used
----> STRONGLY prefered model: N-terminus outside
 5 strong transmembrane helices, total score : 9596
         to length score orientation
# from
                    2224 o-i
1
     9
         32 (24)
 2
    59
         78 (20)
                    1158 i-o
 3
    93 111 (19)
                    2478 o-i
 4 128
        148 (21)
                    2021 i-o
 5
   185
        203 (19)
                    1715 o-i
----> alternative model
 5 strong transmembrane helices, total score : 8763
# from to length score orientation
1
    5
         24 (20) 2070 i-o
 2
    59
         76 (18)
                    1307 o-i
```

3	93	111	(19)	1990	i-o
4	128	146	(19)	1803	o-i
5	185	203	(19)	1593	i-o



TMpred output for NarI contig135

## Transmembrane prediction for NarK<sub>1</sub> in *nar135*

TMpred output for NarK1 in contig135 [ISREC-Server] Date: Sat Jul 26 0:07:47 MET 2003

Sequence: VVP...ILE, length: 525 Prediction parameters: TM-helix length between 17 and 33

#### 1.) Possible transmembrane helices

The sequence positions in brackets denominate the core region. Only scores above 500 are considered significant.

Inside	to ou	utsić	le	helic	es :	12	found
:	from			to	score	Ce	enter
125 (	125)	144	(	142)	2333		134
154 (	158)	176	(	176)	962		166
180 (	180)	200	(	200)	2475		190

209	(	209)	227	(	227)	2395	217
241	(	243)	269	(	260)	1159	252
275	(	275)	293	(	293)	1860	284
320	(	323)	343	(	341)	1375	333
352	(	352)	370	(	368)	1400	360
382	(	382)	400	(	398)	1991	390
417	(	420)	438	(	436)	2590	428
458	(	458)	477	(	477)	2234	467
490	(	490)	507	(	507)	1714	499
Outsi	lde	e to	insid	le	helio	ces : 12	found
	1	Erom			to	score c	enter
125	(	125)	144	(	142)	2069	133
125 153	( (	125) 153)	$\begin{array}{c} 144 \\ 171 \end{array}$	( (	142) 171)	2069 944	133 163
125 153 184	( ( (	125) 153) 184)	144 171 200	( ( (	142) 171) 200)	2069 944 2290	133 163 192
125 153 184 209	( ( (	125) 153) 184) 209)	144 171 200 227	( ( (	142) 171) 200) 227)	2069 944 2290 2496	133 163 192 219
125 153 184 209 241	( ( ( (	125) 153) 184) 209) 243)	144 171 200 227 267	( ( ( (	142) 171) 200) 227) 259)	2069 944 2290 2496 1054	133 163 192 219 251
125 153 184 209 241 275	( ( ( ( (	125) 153) 184) 209) 243) 275)	144 171 200 227 267 292	( ( ( (	142) 171) 200) 227) 259) 292)	2069 944 2290 2496 1054 1945	133 163 192 219 251 283
125 153 184 209 241 275 323		125) 153) 184) 209) 243) 275) 323)	144 171 200 227 267 292 341	( ( ( ( (	142) 171) 200) 227) 259) 292) 341)	2069 944 2290 2496 1054 1945 1643	133 163 192 219 251 283 333
125 153 184 209 241 275 323 352		125) 153) 184) 209) 243) 275) 323) 352)	144 171 200 227 267 292 341 369		142) 171) 200) 227) 259) 292) 341) 369)	2069 944 2290 2496 1054 1945 1643 1710	133 163 192 219 251 283 333 360
125 153 184 209 241 275 323 352 383		125) 153) 184) 209) 243) 275) 323) 352) 383)	144 171 200 227 267 292 341 369 400		142) 171) 200) 227) 259) 292) 341) 369) 400)	2069 944 2290 2496 1054 1945 1643 1710 1884	133 163 192 219 251 283 333 360 391
125 153 184 209 241 275 323 352 383 418		125) 153) 184) 209) 243) 275) 323) 352) 383) 420)	144 171 200 227 267 292 341 369 400 438		142) 171) 200) 227) 259) 292) 341) 369) 400) 438)	2069 944 2290 2496 1054 1945 1643 1710 1884 2123	133 163 192 219 251 283 333 360 391 429
125 153 184 209 241 275 323 352 383 418 455		125) 153) 184) 209) 243) 275) 323) 352) 383) 420) 457)	144 171 200 227 267 292 341 369 400 438 473		142) 171) 200) 227) 259) 292) 341) 369) 400) 438) 473)	2069 944 2290 2496 1054 1945 1643 1710 1884 2123 2354	133 163 192 219 251 283 333 360 391 429 465

#### 2.) Table of correspondences

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.

Helices shown in brackets are considered insignificant. A "+"-symbol indicates a preference of this orientation. A "++"-symbol indicates a strong preference of this orientation.

\_\_\_\_\_

 • •	sjineer mareate	s a subing prototor	<b>e</b> manoni

	ź	inside	e->out	cside	out	side	e->ir	nside		
125-	144	(20)	2333	++	1	.25-	144	(20)	2069	
154-	176	(23)	962		1	.53-	171	(19)	944	
180-	200	(21)	2475	+	1	84-	200	(17)	2290	
209-	227	(19)	2395		2	209-	227	(19)	2496	+
241-	269	(29)	1159	+	2	241-	267	(27)	1054	
275-	293	(19)	1860		2	275-	292	(18)	1945	+
320-	343	(24)	1375		3	823-	341	(19)	1643	++
352-	370	(19)	1400		3	852-	369	(18)	1710	++
382-	400	(19)	1991	+	3	883-	400	(18)	1884	
417-	438	(22)	2590	++	4	18-	438	(21)	2123	
458-	477	(20)	2234		4	155-	473	(19)	2354	+
490-	507	(18)	1714		4	89-	509	(21)	1977	++

\_

3.) Suggested models for transmembrane topology

\_\_\_\_\_

These suggestions are purely speculative and should be used with **extreme caution** since they are based on the assumption that all transmembrane helices have been found.

In most cases, the Correspondence Table shown above or the prediction plot that is also created should be used for the topology assignment of unknown proteins.

```
2 possible models considered, only significant TM-segments used
----> STRONGLY prefered model: N-terminus inside
12 strong transmembrane helices, total score : 22762
# from
          to length score orientation
   125
         144 (20)
                      2333 i-o
1
2
   153
         171 (19)
                      944 o-i
 3
   180
                      2475 i-o
         200 (21)
 4
    209
         227 (19)
                      2496 o-i
5
    241
         269 (29)
                      1159 i-o
6
   275
         292 (18)
                     1945 o-i
7
   320
         343 (24)
                      1375 i-o
8
         369 (18)
                      1710 o-i
   352
9
   382
         400 (19)
                      1991 i-o
                      2123 o-i
10
   418
         438 (21)
11
   458
         477 (20)
                      2234 i-o
         509 (21)
12
   489
                      1977 o-i
----> alternative model
12 strong transmembrane helices, total score : 22215
# from
          to length score orientation
   125
         144 (20)
                      2069 o-i
1
 2
   154
         176 (23)
                      962 i-o
 3
   184
         200 (17)
                      2290 o-i
 4
   209
         227 (19)
                      2395 i-o
 5
    241
         267 (27)
                      1054 o-i
   275
 б
         293 (19)
                      1860 i-o
 7
   323
         341 (19)
                      1643 o-i
8
         370 (19)
   352
                      1400 i-o
9
   383
         400 (18)
                      1884 o-i
10
   417
         438 (22)
                      2590 i-o
11
   455
         473 (19)
                      2354 o-i
12 490
         507 (18)
                      1714 i-o
```



# Transmembrane prediction for NarK<sub>2</sub> in *nar135*

TMpred output for NarK2 contig135 [ISREC-Server] Date: Sat Jul 26 0:10:00 MET 2003

Sequence: MPT...VPC, length: 475 Prediction parameters: TM-helix length between 17 and 33

#### **1.)** Possible transmembrane helices

The sequence positions in brackets denominate the core region. Only scores above 500 are considered significant.

Ι	nsić	le	to ou	utsic	le	helic	es :	12	found
		1	Erom			to	scor	e ce	enter
	44	(	44)	64	(	61)	251	7	53
	78	(	81)	106	(	100)	126	7	90
	111	(	111)	128	(	128)	132	1	120
	137	(	137)	153	(	153)	188	2	145
	185	(	185)	203	(	203)	150	3	194
	223	(	225)	242	(	242)	230	9	233
	265	(	265)	283	(	283)	221	4	275

-----

301	(	301)	319	(	319)	696	311
328	(	330)	348	(	346)	2535	338
350	(	352)	372	(	369)	2340	361
414	(	414)	434	(	432)	1556	424
446	(	446)	464	(	464)	2326	456
Outsi	Lde	e to :	insid	le	heli	ces : 12	found
	f	Erom			to	score ce	enter
42	(	45)	65	(	62)	2498	54
79	(	79)	104	(	97)	1465	89
109	(	111)	127	(	127)	1534	119
138	(	142)	164	(	164)	1914	152
181	(	183)	203	(	200)	1300	192
223	(	225)	242	(	242)	2006	233
265	(	265)	288	(	283)	1701	275
301	(	301)	318	(	318)	766	309
333	(	333)	352	(	352)	2132	343
352	(	352)	372	(	372)	2013	362
415	(	415)	431	(	431)	1366	423
446	(	446)	466	(	464)	2558	456

\_\_\_\_\_

#### 2.) Table of correspondences

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.

Helices shown in brackets are considered insignificant. A "+"-symbol indicates a preference of this orientation.

A "++"-symbol indicates a strong preference of this orientation.

```
inside->outside | outside->inside
44- 64 (21) 2517
                           42- 65 (24) 2498
78- 106 (29) 1267
                           79- 104 (26) 1465
                                             +
                        109- 127 (19) 1534 ++
111- 128 (18) 1321
137-153 (17) 1882
                      138- 164 (27) 1914
                     185- 203 (19) 1503 ++
                         181- 203 (23) 1300
223- 242 (20) 2309 ++
                      223- 242 (20) 2006
265- 283 (19) 2214 ++
                      265- 288 (24) 1701
301- 319 (19) 696
                          301- 318 (18)
                                        766
328- 348 (21) 2535 ++
                          333- 352 (20) 2132
350- 372 (23) 2340 ++
                          352- 372 (21) 2013
                          415- 431 (17) 1366
414-434 (21) 1556 +
446- 464 (19) 2326
                          446- 466 (21) 2558 ++
```

\_

#### 3.) Suggested models for transmembrane topology

These suggestions are purely speculative and should be used with **extreme caution** since they are based on the assumption that all transmembrane helices have been found.

\_\_\_\_\_

In most cases, the Correspondence Table shown above or the prediction plot that is also created should be used for the topology assignment of unknown proteins.

2 possible models considered, only significant TM-segments used \*\*\* the models differ in the number of TM-helices ! \*\*\* ----> STRONGLY prefered model: N-terminus inside 12 strong transmembrane helices, total score : 22368 # from to length score orientation 1 44 64 (21) 2517 i-o 79 104 (26) 2 1465 o-i 3 111 128 (18) 1321 i-o 4 138 164 (27) 1914 o-i 5 185 203 (19) 1503 i-o б 223 242 (20) 2006 o-i 7 265 283 (19) 2214 i-o 8 301 318 (18) 766 o-i 9 328 348 (21) 2535 i-o 10 352 372 (21) 2013 o-i 11 414 434 (21) 1556 i-o 12 446 466 (21) 2558 o-i ----> alternative model 11 strong transmembrane helices, total score : 19433 # from to length score orientation 1 42 65 (24) 2498 o-i 2 78 106 (29) 1267 i-o 3 109 127 (19) 1534 o-i 4 137 153 (17) 1882 i-o 5 181 203 (23) 1300 o-i б 223 242 (20) 2309 i-o 7 265 288 (24) 1701 o-i 8 301 319 (19) 696 i-o 9 333 352 (20) 2132 o-i 10 414 434 (21) 1556 i-o 11 446 466 (21) 2558 o-i



# Appendix VII

# Protein sequence alingment of NarC from G. metallireducens and T. thermophilus

G.metacytC121 G.metacytC135 T.thermcytC	MKTVIIGGVAASAACAMTRRMLVAMVIALLASAMHPHQGSSKERAEGDGKALFQAKCSPC MARGILRAAVSRVCLVLVLSLAAHAYGDPGKDLFDKQCSGC DGKALYGTHCASC . *: : * .: ** * **********************	60 41 33
G.metacytC121 G.metacytC135 T.thermcytC	HTIGGGRLVGPDLKGVTARRDRGWLERFISAPDKVLASGDPVAARLFQESNKVPMPNL HTVGGGDSGGPDLKGVTDRRSEEWLESIIIEPDKLAAT-DPARKELIKKYG-YEMPKL HGAEGQGIPGAIPPLAGNPRVQDEAYVVRAVREGLSGPLEVGGVTYDG-VMPPLP * * * * * * * * * * * * * * * * * * *	118 97 87
G.metacytC121 G.metacytC135 T.thermcytC	GLGQDQVNVLLNYLSAPEGISPSKAEPPASPVQEEGDPSRGEKFFTGALPFAQGGTP GIGHDDAIKIITWLRGGGGGAQTGSPPADEHPEVVVTPALVAKGKAFFTGQKPLANGGAP QVSEAEARAIAQYLKGLGGAPAQAAPAPQVQGDPALGRALYLGQKALKNGGAP : : : : * . * * : *. : * .: * .: * .: : : * .: : * .: : * .: : : :	175 157 140
G.metacytC121 G.metacytC135 T.thermcytC	CIACHQVAGIPPLGGGALGPDLTTIVATFG-DTALSSALATLPFPTMKPIFGDHPLTRGE CIACHRISTPGITGGNLAVSNLSASYQKMG-EKGMRGALKALKFPTMKKIYENRPLTDDE CQACHTVAGVGFFGGGALGKDLTDAAKRLGGEAGLAALVENPAFPVMREAYKGKPLTKEE * *** :: **. :*: :*: :*: :*** *	234 216 200
G.metacytC121 G.metacytC135 T.thermcytC	QGDLLAFFRKVAPRPPVTATTRIALWAGGGFLVLLLLTGALGRRRLGSVRGTLVDEANRA VTSLIALYKDAALTKSTPSASFFPAYGAGTFVVFLVGLTLYKRRTRATALAAFLVQVSQEAPRPSSLYVGRFLVAGLVVLALLLLYQAVLWQLRPRSLAERIRAQL * *:::: . :: : :::::	294 262 260
G.metacytCl21 G.metacytCl35 T.thermcytC	GR 296  RR 262	

# Protein sequence alingment of NarG from G. metallireducens and T. thermophilus

T.thermNarG G.metacontil21 G.metacontil35	-MRDWIKEVENPAERKWEEFYRNRFQHDKRVRTTHGVNCTGSCSWEVFVKDGIVTWELQA -MR-WIKELTDPKARRWEEFYRNRWQHDRVVRSTHGVNCTGGCSWLVYVKEGIVAWEMQA MTQERLKDDRCGCDRGWEEFYRNRWQYDKVVRSTHGVNCTGSCSWMVHVKDGIVGWELQA : :*: * *******::: **:****************	59 58 60
T.thermNarG G.metacontil21 G.metacontil35	TDYPSLEAGLPPYEPRGCQRGISFSWYLYSPIRVKYPYAKGALLDLWREAKKAHGDPVAA TDYPRLEAGLPPYEPRGCQRGISFSWYLYSPLRIKHPTIRGSLLDLWRAAKEEHSDPLAA NDYPQFDGTIPNHEPRGCTRGISYSWYLYSPLRVKYPYMRGALIDLWRDARSRHEDPVEA .*** ::. :* :***** ****:***************	119 118 120
T.thermNarG G.metacontil21 G.metacontil35	WEAIQNDPHKRKRYQKARGKGGFRRASWDEVLELIAAAVVSTVKRYGPDRVIGFSPIPAM WGSIVEDSEKRRSYQQARGKGGFRRTSWDEALEITAASTLYTIKRYGPDRIVGFSPIPAM WKSIVENPESRRAYVSKRGTGGMRRADWDEVTELIAASSIYTAKKYGPDRVIGFSPIPAM * :* ::*: * . **.**:.***. *: **: : * *:********	179 178 180
T.thermNarG G.metacontil21 G.metacontil35	SQISYAAGSRFLSLLGGVPMTFYDWYCDLPNASPEIWGEQTDVHESADWYNARFIAVMGS SMCSYAAGTRLLSLLGGVVLSFYDWYADLPPASPETWGEKTDVAESADWYNSKYIVVMGT SMISYAAGTRFLQLFGGVAMSFYDWYCDLPPASPQVWGEQTDVNESADWYNASYIVLCGS * *****:*:*.*** ::**** **** ***: ***:*** ********	239 238 240
T.thermNarG G.metaconti121 G.metaconti135	NLNMTRTPDTHFIAEVRHAGAKLTVFSPDFSQVSKYADWWIPIHPGQDGAFWMAVNHVLL NLSMTRTPDVHFVAEARTNGAKLVVLSPDFSQVSKYADWWLPVTAGHDGAFWMAVNHVIL NVPMTRTPDAHFMTEARYRGTKVVVMSPDYNIATKFADNWLPVEQGHDGAFWMAVNHVIL *: ******.**:**************************	299 298 300
T.thermNarG G.metaconti121 G.metaconti135	KEFYADREVPYFQDYLKRYTDAPFLVEIKDGRPGRYLRANRLAEYAEEENGDFKL REFYVERHVPYFADYLARYSDAPFLVTLDRGDGERYVPGRLLRAESLPRYRREENGEWKF KECYADRQVPFFQEYALKYTDLPFLVKLEEKDG-ALKQGEFLRAAEFERSASLEHADWTL :* *.:*.**:* :* :*:* **** :. *. **** :. *:::::	354 358 359
T.thermNarG	${\tt LVFDEAKG-PRMPMGTLGFRWQKEKGKWNLKLEDPRTGEPLNPRLSLLGVEDEVVLVEFD}$	413

G.metacontil21 G.metacontil35	LVFDRLSGEPRMPGGSIGFRWQREAGRWNLEMKDGLDGAEIEPLLSLLDGRDDVLPVSIT CVADGAGD-VKIPHGSLGTRWSKOEGKWNLDMTDLVDGTPIDCRLTLLGGETGKVRFR	418 416
	* * . ::* *::* **.: * :: *:***. : * :	
T.thermNarG	DFAADRRLKRGVPVKYVVTKDGERVAVATVFDLLMAOFGVGRGLPGDYPASYEDDLPYTP	473
G.metacontil21	DYAGGTSVARGVPVRYLETATG-RTPVTTVFDLLMANFGVGRGLSGDYPEDYNAAAPGTP	477
G.metaconti135	-FEAEGESVREVPVRRIETKNG-PVTVATVYDLLMAQFGVSRGLGGDYPKSYDSDKPFTP	474
	:. * ***:: * **:**:****:*** *** *** .*: * **	
T.thermNarG	AWQEKWTGIHRDTLLKYARAWGENGLKTKGRNLIIIGAGINHWYHNNLMYRAGIVALMLT	533
G metacontil35		537
0.11101155	**:::*:: ::*:: ** *:: *: **:*:* ****::* **:.*	551
T.thermNarG	GSVGVNGGGLAHYVGQEKLANQASWAPIAFATDWGYPPRQQNAPSFHYVHSDQWRYERGF	593
G.metacontil21	GCVGVNGGGLNHYTGQEKVAPEASWSLLAFAKDWTKGGRFQNTPSFHYVHSDQWRYDGGA	597
G.metacontil35	GSVGRNGAGLAHYVGQEKVVPLAPWTSIAMAQDWVKPSRLQNTPSFWYIHSDQWRYDRTF *.** **.** **.***: *.*: ** * **:*** *:*****:	594
T.thermNarG	SAYDKTAKGLSDHTIDHQVRAVRKGWLPFFPQFNKSPLEVVAEAEAKGAKTEAEIVQY	651
G.metacontil21	ADPRPDPKGGPAGRHPMDLQVDAVRMGWLPFYPQFDRNPLELVKEAEAAGAKGDEEIIRW	657
G.metacontil35	VDYLKPETGERMPLHAADMNAKAARLGWLPFAPHFNESPLRLVEAAKAAGAQSDDEIRSW * * * * * * * * * * * * * * * * * * *	654
T.thermNarG	VVEALKRGDS-FAVEDPDAPENWPRVWFIWRGNAIGTSAKGHEYFLKHYLGTHTNAI-AE	709
G.metacontil21	VVEELEKKRLRFAVEEPDAPENWPRLWLIWRGNALLSSAKGHEYFLRHYLGTHDSSI-AD	716
G.metaconti135	LVNRLKSGETRFAIEDPDGQGNSPKVWFIWRANAISSSAKGHEFFLKHVIGAPNSSVSAK :*: *: **:*:**. * *::*:***.**: :*****:**: :*: .:: *.	714
T.thermNarG	EVAEGHVTEVVYRKPAPEGKLDLVVDLNFRMDTSALYSDIVLPAATWYEKDDLNTTDLHT	769
G.metacontil21	$\verb"ECAAEQVENAAWQDAAPRGKLDLVVDINFRMDTSALYSDIVLPTATWYEKDDLNTTDLHS"$	776
G.metaconti135	EVAKGQVKDLVWHDKAPEGKMDLVVDLNFRMDTSTLYSDIVLPTATWYEKSDLNTTDMHS * * :* : ::: **.**:********************	774
T.thermNarG	FINPLQAAVPPAWESKPDWEIFKAIAKKVSELARVHLPRPVKDLVMIPLQHDTPDELAQL	829
G.metacontil21		830 834
0.11101155	**: *.****:.**:.**: :*:** ** *:* **:*: ** ****:	051
T.thermNarG	EDRDWKRGEVEPIPGKTMPKFRVVERDYTKLYEKFVTLGPTVEKVGVGMHGLTIPVEDFY	889
G.metacontil21	EVRGWHRGECEAVPGTTMPRLAVVERDFVHLLDRFVSLGEEVKREGIEERGLRWPVADLY	896
G. metacont1155	QVKDWKLGECEATPGKIMPNLAIVEKDIVNLINKIVSVGPHLGHLHANGVSWEADDVH :*: ** *.:**.***.: :****:::* :::*::* : : :*: . *.:	092
T.thermNarG	KELAERQP-RLYEGEKRPSLEEARQVAEAILLLDPVSNGELAYRAFLDEEKKTGVKLTDL	948
G.metacontil21	DELLESAPVREWGGKRFPSLAEAVDAANAILHFAPETNGEVAWRAFRSLEEKTGLTLADL FOLLFTMDTPTWMCKSFLDIFDFKTVADVLLSFADFSNCFLAVRAVONLFKRVCKDLACI	956 952
0	.:* * * * : *: .: . : .: *: .:* : *:***:*: *::**	552
T.thermNarG	AEGNRHVRISFKDIVAQPRRQLTTPTWSAIINHGRAYSPYTLNVERLVPWRTLTGRQHFY	1008
G metacontil35	ASDIRGIRIRFSDLVSQPRRILISPCWSGVINNGRPIAPICLNVERLVPWRILIGRQHFI AACSPDLRIKWFHUTOKDORVOSTDUWSGLUNNGRDVADVTLNUFVDVDWRTLSGROSLV	1010
0.11121155	* . * * ::. :*:* ::* **.: *:** **** ***** ***** *****	1012
T.thermNarG	LDHPNYLAWGEHLPTYKPRPEVHMLQETEKAAKEAQGKLMNYITPHGKWSIHSTYSENHR	1068
G.metacontil21	LDHGWYKAFGEHLPTFKTRIDTEESGHLAKSRPVGKALMLNCLTPHGKWHMHTTYYDNHR	1069
G. metacontriss	*** * : * ***:*.: : . :. ::* :****** :*:**	1008
T.thermNarG	${\tt MMTLSRGGYPVWLNDKDAAELGIRDNDWVELFNDNGVFVQRAIVSARIPRGTVFVYHATE}$	1128
G.metacontil21	MLTLSRGIEPFWLNDRDAEEIGVKDNEWVEVYNDNGVVVTRAVVSARIPRGVGIYYHAPE	1136
G.metacontil35	MLTLGRGGPVVWLNHEDAAGAGIEDNEEIEVYNFNGVVVSRACVSSRIPKGAAVMYHAPE *:**.** ** ***.** *:.**: ***** ** ********	1128
T.thermNarG	RTVGIPKSPLRG-KRAGMNNSITRARLKPVLMSGGYAQFTYAFNYWGPVGVNRDTWVYVR	1187
G.metacontil25	RTINIOKSKKTG-KRGQVHNSI.SRIRIKOTIMI.GGVAOFSVVFMVWGDTGVMDDSV7777	1187
	**:.: * * :*.* :**::* ****.** ***.*::* ** *** *	1107
T.thermNarG	KLETPPEW 1195	
G.metacontil21	КЪЕ-GALKI 1204 КЪРС 1191	
0.mctac01101133	**.	

# Protein sequence alingment of NarH from G. metallireducens and T. thermophilus

T.thermNarH G.matal2lNarH G.metal35NarH	MKVRAHMSMLFHLDKCIGCHTCSIACKNLWTDRKGAEYMWWNNVETRPGAGYPTGWEDQE MKVRAQVSMVFNLDKCIGCHTCSIACKNIWSDRRGAEYMWWNNVETKPGTGYPTLWEDQE MDVRAQLVMVFNLDKCIGCHACSLSCKNIWTDRKGAEYMWWNNVETKPGVGYPKKWEDQE *.***:: *::*:**************************	60 60 60
T.thermNarH G.matal2lNarH G.metal35NarH	RFRGGWEYKDGHLDLRLHSRTQGLFRLFFNPALPSLDDYYEPFTFRYADLFTAPEGEDQP KYRGGWERTDREPRLRLHSRGGGLAKLFFNPRLPTIDDYYEPWSYRYGDLLSAPAGDDQP EFKGGWVADGHNLRLKAMGKGPTLLNMFFQPNMPKLEDYYEPFDFDYGNLYGAKGGDDQP .::*** * * * * * * * * * * * * * * * * *	120 120 120
T.thermNarH G.matal2lNarH G.metal35NarH	TAVPVSMITGEPMT-PEAGPNWDDDLSGSPLYAKNDPNLKALAPEVRAQLEEIEGVVFQY VARPVSLVTGDPIN-IEAGPNWDDDLGGSPVYGAHDPNLAAVSEAERRQLFLLERMVLFY VAEAFSQITGEKMAHIKAGPNWDDDLSGSQVYASADVNLEDKKIIEQYSTMFMQY .** :**: : :************************	179 179 175
T.thermNarH G.matal2lNarH G.metal35NarH	LPRICNHCLNPACVAACPSGAIYKRAEDGVVLVNENKCKAWRMCVAACPYKKVYYNWATG LPRICNHCANPGCVAACPSGAIYKRGEDGIVLVSQEKCRGWRMCVSGCPYKKVYYNWGSG LPRICNHCANPACVASCPSRAIYKRGEDGVVLVDQEVCKGWRFCTSACPYKKVYFNWNSG ******* **.***************************	239 239 235
T.thermNarH G.matal2lNarH G.metal35NarH	KSEKCILCFPRLETGQAPACAHSCVGRIRYMGVLLYDADRIPEAAMVPDEELVESQLSLI KSEKCILCYPRQETGQAPACFHSCVGRLRYLGVLLYDAERIERSASRPDHDLVQAQREMI KAEKCIFCFPRTETGQCNACAHSCVGRIRYVGVLLYDAEKVEQALLKPDNQLVEAHRGVI *:****:*:** ****. ** *****:*********::: .: **.:**	299 299 295
T.thermNarH G.matal2lNarH G.metal35NarH	LDPFDPEVIAAAKAEGIDEGWIQAAQKSPIYKFVKTWGLALPLHPEYRTLAMMFYVPPLS LDPSDPEVVNKAAEEGIHPAVLAAARNSPVYRFVKEWRLALPLHPEFGTVPMLYYVPPLL LDPRDEAVREGARKNGVPEQWLEAAEKSPVYALIKEFELALPLHPEFRTMPMAYYIPSLS *** * * * ::::::::::::::::::::::::::::	359 359 355
T.thermNarH G.matal2lNarH G.metal35NarH	PVVATVEKALRAGRGGEGLVRLDIPETDLDFEVYAALEKARMPVKYLANLFAAGNEALIV         PLAADRENGCYDLGGEFFTSLEQGRLPLRYLASLFSAGNEEIIR         PVLTSGGELHELVEHGTIPTREALRAPIGYLASLFGGGNRDVVA         *: : : : : : : : : : : : : : : : : : :	419 403 399
T.thermNarH G.matal21NarH G.metal35NarH	PILKKMLAVRILKRQESLEGRPTEKALKVLADAGLTLEEAEAIYRLTTLPTLEERFVLPP         AVYRKLTAVRVHRRAATTGDVPAGELERFLAEAGTTPEEAEAIYHMTSLAGFDERHAVPP         AALEKLIALRIHMRAKNLN-QPTDEASLGKAGLDEAGAKRLYRLFTIGGYNERNIIPP        *: *:*: * . *:: ***	479 463 456
T.thermNarH G.mata121NarH G.meta135NarH	YHREMAAEVWKDPLAHKGEAGLGYIQPPLRGE- 511 FMREMAMEMSVDPADRKGEAGTGFLRPPKRG 494 QQREELDSHTRKGASGFGILKKTGRGKR 484 ** *. :** :* * :: . **	

# Protein sequence alingment of NarI from G. metallireducens and T. thermophilus

T.thermNarI G.matal2lNarI G.metal35NarI	MKWNAALFGVFPYIALTLAVAVTAYRMVYR MPVDSGLSDHICAILRFLARPAGTAPNSGDPMTDYLFFGVFPYVALVVAVGGGVYRYAKE MTADIYSSLARMLDYPQEKKSLLWA .::. :* :	30 60 25
T.thermNarI	PFSVS-AQSSQLLEQKRLFFGSTAMHWGLVIVLLGHLLALLVPKGLLLWNAVPLRLYLLE	89
G.matal21NarI	RFSFS-SLSSQFLENRALFWGSVAWHYGIVLVLLAHLAALLLPASWGRLLGDPLRLAVLE	119
G.meta135NarI	IGTVQRYLSEEGLECPLDTFGEFMAASTLAELQEEYVATFDFNPAVAPYLGHHLFGDNQK	85
	: *.: ** :*. :. : :: :: * :	
T.thermNarI	ITGLGLGLWALFGTYVLLARRIAVARVRAASTPMDYLVLVVVFVSALTGVLTAVLYRYGS	149
G.matal21NarI	VTGLALGLFTLAGIVVLLFRRIASPKAWAVVTVMDLFLLAVLVLQVGTGLYNAFVYRWGS	179
G.meta135NarI	KGAYLIALKGEYGRHGYTPLGNELPDHLPVILWFLAHLVRGGEDDARRSFIDRH	139
	. :.* * : .*. : :.: *	
T.thermNarI	FWFPAVMTPYLWSVLTLNPRPELLQDLPFWTKLHVFNFWLFLALFPFSRLVHIITVPLGY	209
G.mata121NarI	LWYLSLAVPWLHSLACLNPDIKYIVPLPLPVKVHLFNAFVIFVLFPVTRLVHIVSIPVSY	239
G.metal35NarI	VLPGVEKLSDSFAASRRDSPWRPAVEAVRLICGEDSARGEHPAEGIGDGVLE	191
	. * * : : : :	

T.thermNarI	VVRPWQIVVWVRNWRGEL	227
G.matal21NarI	LWRPWQIAVWNKPPRWMIPRK	260
G.meta135NarI	KNRPKTAAHPEVSPC	206
	** .	

### References

- Abaibou, H., J. Pommier, S. Benoit, G. Giordano, and M. A. Mandrand-Berthelot. 1995.
   Expression and characterization of the *Escherichia coli* fdo locus and a possible physiological role for aerobic formate dehydrogenase. J Bacteriol 177:7141-9.
- Aichi, M., and T. Omata. 1997. Involvement of NtcB, a LysR family transcription factor, in nitrite activation of the nitrate assimilation operon in the cyanobacterium *Synechococcus sp.* strain PCC 7942. J Bacteriol 179:4671-5.
- Allison, C., and G. T. Macfarlane. 1988. Effect of nitrate on methane production and fermentation by slurries of human faecal bacteria. Journal of General Microbiology 134:1397-1405.
- Antipov, A. N., N. N. Lyalikova, T. V. Khijniak, and P. L'Vov N. 1998. Molybdenum-free nitrate reductases from vanadate-reducing bacteria. FEBS Lett 441:257-60.
- Bagramyan, K., N. Mnatsakanyan, A. Poladian, A. Vassilian, and A. Trchounian. 2002. The roles of hydrogenases 3 and 4, and the F0F1-ATPase, in H2 production by *Escherichia coli* at alkaline and acidic pH. FEBS Lett 516:172-8.
- Bamford, V. A., H. C. Angove, H. E. Seward, A. J. Thomson, J. A. Cole, J. N. Butt, A. M. Hemmings, and D. J. Richardson. 2002. Structure and spectroscopy of the periplasmic cytochrome c nitrite reductase from *Escherichia coli*. Biochemistry 41:2921-31.
- Bauer, C. E., S. Elsen, and T. H. Bird. 1999. Mechanisms for redox control of gene expression. Annu Rev Microbiol 53:495-523.
- Bazylinski DA, D. A., Schuler D, Phillips EJ, Lovley DR. 2000. N2-dependent growth and nitrogenase activity in the metal-metabolizing bacteria, *Geobacter* and *Magnetospirillum* species. Environ Microbiol 2:266-73.
- 9. Bell, L. C., M. D. Page, B. C. Berks, D. J. Richardson, and S. J. Ferguson. 1993. Insertion of transposon Tn5 into a structural gene of the membrane- bound nitrate reductase of *Thiosphaera pantotropha* results in anaerobic overexpression of periplasmic nitrate reductase activity. J Gen Microbiol 139:3205-14.
- Berks, B. C., S. J. Ferguson, J. W. Moir, and D. J. Richardson. 1995. Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. Biochim Biophys Acta 1232:97-173.
- Blasco, F., C. Iobbi, J. Ratouchniak, V. Bonnefoy, and M. Chippaux. 1990. Nitrate reductases of *Escherichia coli*: sequence of the second nitrate reductase and comparison with that encoded by the *narGHJI* operon. Mol Gen Genet 222:104-11.
- 12. **Blasco, R., F. Castillo, and M. Martinez-Luque.** 1997. The assimilatory nitrate reductase from the phototrophic bacterium, *Rhodobacter capsulatus E1F1*, is a flavoprotein. FEBS Lett **414:**45-9.
- Brondijk, T. H., D. Fiegen, D. J. Richardson, and J. A. Cole. 2002. Roles of NapF, NapG and NapH, subunits of the *Escherichia coli* periplasmic nitrate reductase, in ubiquinol oxidation. Mol Microbiol 44:245-55.
- Caccavo, F., Jr., D. J. Lonergan, D. R. Lovley, M. Davis, J. F. Stolz, and M. J. McInerney.
   1994. *Geobacter sulfurreducens sp. nov.*, a hydrogen- and acetate-oxidizing dissimilatory metalreducing microorganism. Appl Environ Microbiol 60:3752-9.
- Chen, Q. K., Hertz, J.Z. and Stormo, G.D. 1995. MATRIX SEARCH 1.0: a computer program that scans DNA sequences for transcriptional elements using a database of weight matrices. Comp. Appl. Biosciences 11:563-566.
- Clegg, S., F. Yu, L. Griffiths, and J. A. Cole. 2002. The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli K-12*: two nitrate and three nitrite transporters. Mol Microbiol 44:143-55.
- Coates JD, B. V., Achenbach LA, McInerney MJ, Lovley DR. 2001. Geobacter hydrogenophilus, Geobacter chapellei and Geobacter grbiciae, three new, strictly anaerobic, dissimilatory Fe(III)-reducers. Int J Syst Evol Microbiol. 51:581-8.
- Cole, J. 1996. Nitrate reduction to ammonia by enteric bacteria: redundancy, or a strategy for survival during oxygen starvation? FEMS Microbiol Lett 136:1-11.
- Coppi MV, L. C., Sandler SJ, Lovley DR. 2001. Development of a genetic system for Geobacter sulfurreducens. Appl Environ Microbiol 67:3180-7.

- Cord-Ruwisch, R., D. R. Lovley, and B. Schink. 1998. Growth of *Geobacter sulfurreducens* with acetate in syntrophic cooperation with hydrogen-oxidizing anaerobic partners. Appl Environ Microbiol 64:2232-6.
- 21. **Dixon, R.** 1998. The oxygen-responsive NIFL-NIFA complex: a novel two-component regulatory system controlling nitrogenase synthesis in gamma-proteobacteria. Arch Microbiol **169**:371-80.
- 22. Duncan, C., H. Li, R. Dykhuizen, R. Frazer, P. Johnston, G. MacKnight, L. Smith, K. Lamza, H. McKenzie, L. Batt, D. Kelly, M. Golden, N. Benjamin, and C. Leifert. 1997. Protection against oral and gastrointestinal diseases: importance of dietary nitrate intake, oral nitrate reduction and enterosalivary nitrate circulation. Comp Biochem Physiol A Physiol 118:939-48.
- Eady, R. R. 1996. Structureminus signFunction Relationships of Alternative Nitrogenases. Chem Rev 96:3013-3030.
- 24. Eaves, D. J., J. Grove, W. Staudenmann, P. James, R. K. Poole, S. A. White, I. Griffiths, and J. A. Cole. 1998. Involvement of products of the nrfEFG genes in the covalent attachment of haem c to a novel cysteine-lysine motif in the cytochrome c552 nitrite reductase from *Escherichia coli*. Mol Microbiol 28:205-16.
- Einsle, O., A. Messerschmidt, P. Stach, G. P. Bourenkov, H. D. Bartunik, R. Huber, and P. M. Kroneck. 1999. Structure of cytochrome c nitrite reductase. Nature 400:476-80.
- Einsle, O., P. Stach, A. Messerschmidt, J. Simon, A. Kroger, R. Huber, and P. M. Kroneck.
   2000. Cytochrome c nitrite reductase from *Wolinella succinogenes*. Structure at 1.6 A resolution, inhibitor binding, and heme-packing motifs. J Biol Chem 275:39608-16.
- 27. Ferguson, S. J. 1998. Nitrogen cycle enzymology. Curr Opin Chem Biol 2:182-93.
- Fernandez-Lopez, M., J. Olivares, and E. J. Bedmar. 1996. Purification and characterization of the membrane-bound nitrate reductase isoenzymes of *Bradyrhizobium japonicum*. FEBS Lett 392:1-5.
- Fulop, V., J. W. Moir, S. J. Ferguson, and J. Hajdu. 1995. The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome cd1. Cell 81:369-77.

- Gangeswaran, R., D. J. Lowe, and R. R. Eady. 1993. Purification and characterization of the assimilatory nitrate reductase of *Azotobacter vinelandii*. Biochem J 289 (Pt 2):335-42.
- 31. Hille, R. 1996. The Mononuclear Molybdenum Enzymes. Chem Rev 96:2757-2816.
- 32. Hussain, H., J. Grove, L. Griffiths, S. Busby, and J. Cole. 1994. A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. Mol Microbiol **12**:153-63.
- Jormakka M, T. S., Byrne B, Iwata S. 2002. Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. Science 295:1863-8.
- Kisker, C., H. Schindelin, and D. C. Rees. 1997. Molybdenum-cofactor-containing enzymes: structure and mechanism. Annu Rev Biochem 66:233-67.
- Lin BL, S. A., Shibasaki R, Suzuki M. 2001. A Modelling Approach to Global Nitrate Leaching Caused by Anthropogenic Fertilisation. Water Research 35:1961-1968.
- 36. Lin, J. T., and V. Stewart. 1998. Nitrate assimilation by bacteria. Adv Microb Physiol **39:**1-30.
- 37. Liu, H. P., S. Takio, T. Satoh, and I. Yamamoto. 1999. Involvement in denitrification of the napKEFDABC genes encoding the periplasmic nitrate reductase system in the denitrifying phototrophic bacterium *Rhodobacter sphaeroides f. sp. denitrificans*. Biosci Biotechnol Biochem 63:530-6.
- Lloyd, J. R., V. A. Sole, C. V. Van Praagh, and D. R. Lovley. 2000. Direct and Fe(II)-mediated reduction of technetium by Fe(III)-reducing bacteria. Appl Environ Microbiol 66:3743-9.
- Lonergan DJ, J. H., Coates JD, Phillips EJ, Schmidt TM, Lovley DR. 1996. Phylogenetic analysis of dissimilatory Fe(III)-reducing bacteria. J Bacteriol 178:2402-8.
- Lovley, D. R. 1997. Microbial Fe(III) reduction in subsurface environments. FEMS microbiol Reviews 20:305-313.
- Lovley, D. R., S. J. Giovannoni, D. C. White, J. E. Champine, E. J. Phillips, Y. A. Gorby, and S. Goodwin. 1993. *Geobacter metallireducens gen. nov. sp. nov.*, a microorganism capable of coupling the complete oxidation of organic compounds to the reduction of iron and other metals. Arch Microbiol 159:336-44.

- Lovley, D. R., and E. J. P. Phillips. 1988. Novel mode of microbial energy metabolism: organis carbon oxidation coupled to dissimilatory reduction of iron or manganese. Appl Environ Microbiol 54:1472-1480.
- Lowry, O. H., Rosebrough, N. J., Farr, A. L., Randall, R. J. 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem 193:265-75.
- 44. **Martinez Murillo F, G. T., Senko J, Basu P, Stolz JF.** 1999. A heme-C-containing enzyme complex that exhibits nitrate and nitrite reductase activity from the dissimilatory iron-reducing bacterium *Geobacter metallireducens*. Arch Microbiol **172:**313-20.
- 45. Martinez-Espinosa, R. M., F. C. Marhuenda-Egea, and M. J. Bonete. 2001. Assimilatory nitrate reductase from the haloarchaeon *Haloferax mediterranei*: purification and characterisation. FEMS Microbiol Lett 204:381-5.
- 46. Moreno-Vivian C, Cabello P, Martinez-Luque M, Blasco R, and C. F. 1999. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. J Bacteriol 181:6573-84.
- 47. Nurizzo, D., M. C. Silvestrini, M. Mathieu, F. Cutruzzola, D. Bourgeois, V. Fulop, J. Hajdu,
   M. Brunori, M. Tegoni, and C. Cambillau. 1997. N-terminal arm exchange is observed in the
   2.15 A crystal structure of oxidized nitrite reductase from *Pseudomonas aeruginosa*. Structure
   5:1157-71.
- Oelze, J., and G. Klein. 1996. Control of nitrogen fixation by oxygen in purple nonsulfur bacteria. Arch Microbiol 165:219-25.
- Parham, N. J., and G. R. Gibson. 2000. Microbes involved in dissimilatory nitrate reduction in the human large intestine. FEMS Microbiol Ecol 31:21-28.
- 50. Peakman, T., J. Crouzet, J. F. Mayaux, S. Busby, S. Mohan, N. Harborne, J. Wootton, R. Nicolson, and J. Cole. 1990. Nucleotide sequence, organisation and structural analysis of the products of genes in the nirB-cysG region of the *Escherichia coli K-12* chromosome. Eur J Biochem 191:315-23.
- Philippot, L. 2002. Denitrifying genes in bacterial and Archaeal genomes. Biochim Biophys Acta 1577:355-76.

- Philippot, L., and O. Hojberg. 1999. Dissimilatory nitrate reductases in bacteria. Biochim Biophys Acta 1446:1-23.
- Ponting, C. P., and L. Aravind. 1997. PAS: a multifunctional domain family comes to light. Curr Biol 7:R674-7.
- Potter, L., P. Millington, L. Griffiths, and J. Cole. 2000. Survival of bacteria during oxygen limitation. Int J Food Microbiol 55:11-8.
- 55. **Potter, L., and T. G. Millington PD, Rothery RA, Giordano G, Cole JA.** 2000. Novel growth characteristics and high rates of nitrate reduction of an *Escherichia coli strain, LCB2048*, that expresses only a periplasmic nitrate reductase. FEMS Microbiol Lett **185(1):**51-7.
- 56. Potter, L. C., and J. A. Cole. 1999. Essential roles for the products of the napABCD genes, but not napFGH, in periplasmic nitrate reduction by *Escherichia coli K-12*. Biochem J 344 Pt 1:69-76.
- 57. Rabin, R. S., and V. Stewart. 1993. Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate- and nitrite-regulated gene expression in *Escherichia coli K-12*. J Bacteriol 175:3259-68.
- 58. Naik R. R., F. M. M., Stolz, J. F. 1993. Evidence for a novel nitrate reductase in the dissimilatory iron-reducing bacterium *Geobacter metallireducens*. FEMS Microbiology Letters **106**:53-58.
- 59. Ramirez, S., Moreno, R., Zafra, O., Castan, P., Valles, C. and Berenguer, J. 2000. Two nitrate/nitrite transporters are encoded within the mobilizable plasmid for nitrate respiration of *Thermus thermophilus HB8*. J. Bacteriol 182:2179-2183.
- Ramirez-Arcos, S., L. A. Fernandez-Herrero, I. Marin, and J. Berenguer. 1998. Anaerobic growth, a property horizontally transferred by an Hfr-like mechanism among extreme thermophiles. J Bacteriol 180:3137-43.
- Ribbe, M., D. Gadkari, and O. Meyer. 1997. N2 fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-superoxide oxidoreductase that couple N2 reduction to the oxidation of superoxide produced from O2 by a molybdenum-CO dehydrogenase. J Biol Chem 272:26627-33.

- 62. **Richardson, D. J., B. C. Berks, D. A. Russell, S. Spiro, and C. J. Taylor.** 2001. Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. Cell Mol Life Sci **58**:165-78.
- Richardson, D. J., and N. J. Watmough. 1999. Inorganic nitrogen metabolism in bacteria. Curr Opin Chem Biol 3:207-19.
- Romao MJ, K. J., Huber R, Moura JJ. 1997. Structure and function of molybdopterin containing enzymes. Prog Biophys Mol Biol 68(2-3):121-144.
- Senko, J. M., and J. F. Stolz. 2001. Evidence for iron-dependent nitrate respiration in the dissimilatory iron-reducing bacterium *Geobacter metallireducens*. Appl Environ Microbiol 67:3750-2.
- 66. Simon, J., R. Gross, O. Einsle, P. M. Kroneck, A. Kroger, and O. Klimmek. 2000. A NapC/NirT-type cytochrome c (NrfH) is the mediator between the quinone pool and the cytochrome c nitrite reductase of *Wolinella succinogenes*. Mol Microbiol 35:686-96.
- 67. **Stolz, J. F., and P. Basu.** 2002. Evolution of nitrate reductase: Molecular and structural variations on a common function. Chem Biochem **3:**101-109.
- 68. Stolz, J. F., D. J. Ellis, J. S. Blum, D. Ahmann, D. R. Lovley, and R. S. Oremland. 1999. Sulfurospirillum barnesii sp. nov. and Sulfurospirillum arsenophilum sp. nov., new members of the Sulfurospirillum clade of the epsilon Proteobacteria. Int J Syst Bacteriol 49 Pt 3:1177-80.
- 69. **Stolz JF, G. T., Blum JS, Oremland R, Murillo FM.** 1997. Differential cytochrome content and reductase activity in *Geospirillum barnesii strain SeS3*. Arch Microbiol **167:**1-5.
- Stolz, J. F., and R. S. Oremland. 1999. Bacterial respiration of arsenic and selenium. FEMS Microbiol Rev 23:615-27.
- 71. **Straub KL, B.-C. B.** 2001. *Geobacter bremensis sp. nov.* and *Geobacter pelophilus sp. nov.*, two dissimilatory ferric-iron-reducing bacteria. Int J Syst Evol Microbiol **51:**1805-8.
- Swofford, D. L. 1998. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods).
   Sunderland, Massachusetts: Sinauer Associates. Version 4.
- 73. Szuki, S. K., K. Yamaguchi, K. Inoue, T. Kai, Y. 1999. Structure-function relationships of copper-containing nitrite reductases. Coordination Chemistry Reviews **190-192:**245-265.

- 74. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-82.
- Tiedje, J. M. 1988. Ecology of Denitrification and dissimilatory nitrate reduction to ammonium,
   p. 179-244. *In* Z. AJB (ed.), Biology of anaerobic Microorganisms. Wiley, New York.
- Tischner, R. 2000. Nitrate uptake and reduction in higher and lower plants. Plant, Cell and Environment 23:1005-1024.
- 77. **Tyson, K., R. Metheringham, L. Griffiths, and J. Cole.** 1997. Characterisation of *Escherichia coli K-12* mutants defective in formate-dependent nitrite reduction: essential roles for hemN and the menFDBCE operon. Arch Microbiol **168:**403-11.
- 78. Van Spanning, R. J., A. P. De Boer, W. N. Reijnders, S. Spiro, H. V. Westerhoff, A. H. Stouthamer, and J. Van der Oost. 1995. Nitrite and nitric oxide reduction in *Paracoccus denitrificans* is under the control of NNR, a regulatory protein that belongs to the FNR family of transcriptional activators. FEBS Lett 360:151-4.
- 79. Waibel, A. H., Hanazawa, T., Hinton, G. E., Shikano, K., Lang, K. J. 1989. "Phoneme Recognition Using Time-Delay Neural Networks". IEEE Transactions on Acoustic, Speech, and Signal Processing 37:328-339.
- Wetzel, M. V. d. W. a. A. 1998. Temporal and spatial dynamics of nitrite accumulation in the river Lahn. Wat. Res. 32:1653-1661.
- 81. Wissenbach, U., A. Kroger, and G. Unden. 1990. The specific functions of menaquinone and demethylmenaquinone in anaerobic respiration with fumarate, dimethylsulfoxide, trimethylamine N-oxide and nitrate by *Escherichia coli*. Arch Microbiol 154:60-6.
- 82. Wissenbach, U., D. Ternes, and G. Unden. 1992. An *Escherichia coli* mutant containing only demethylmenaquinone, but no menaquinone: effects on fumarate, dimethylsulfoxide, trimethylamine N-oxide and nitrate respiration. Arch Microbiol 158:68-73.
- 83. Wood, N. J., T. Alizadeh, S. Bennett, J. Pearce, S. J. Ferguson, D. J. Richardson, and J. W. Moir. 2001. Maximal expression of membrane-bound nitrate reductase in *Paracoccus* is induced by nitrate via a third FNR-like regulator named NarR. J Bacteriol 183:3606-13.

## 84. Zafra, O., S. Ramirez, P. Castan, R. Moreno, F. Cava, C. Valles, E. Caro, and J. Berenguer. 2002. A cytochrome c encoded by the *nar* operon is required for the synthesis of active respiratory nitrate reductase in *Thermus thermophilus*. FEBS Lett 523:99-102.