Evolution of Nitrate Reductase: Molecular and Structural Variations on a Common Function

John F. Stolz[b] and Partha Basu*[a]

The biological transformation of nitrogen oxyanions is widespread in nature and gives rise to a robust biogeochemical cycle. The first step in nitrate reduction is carried out by the enzyme nitrate reductase (NR). Although NR always catalyzes the same chemical reaction (conversion of nitrate into nitrite), its location in the cell, structure, and function are organism-dependent. We use protein sequence data to determine phylogenetic relationships and to examine similarities in structure and function. Three distinct clades of NR are apparent: the eukaryotic assimilatory NR (Euk-NR) clade, the membrane-associated prokaryotic NR (Nar) clade, and a clade that includes both the periplasmic NR (Nap) and prokaryotic assimilatory NR (Nas). The high degree of sequence similarity and a phylogenetic distribution that follows taxonomic classification suggest a monophyletic origin for the Euk-NR early on in the evolution of eukaryotic cells. In contrast, sequence conservation, phylogenetic analysis, and physiology suggest that both Nar and Nap were acquired by horizontal gene transfer. Nap and Nas share a lesser degree of similarity, with Nap a subclade of Nas. Nap from strict anaerobic bacteria such as Desulfovibrio desulfuricans *is ancestral to facultative species and may provide an evolutionary link between Nap and Nas. We observed conserved binding sites for molybdenum and pterin cofactors in all four proteins. In pathways involving Euk-NR, Nas, and Nar, for which ammonia is the end product, nitrite is reduced to ammonia by a siroheme nitrite reductase. Nap, however, is coupled to a pentaheme nitrite reductase. In denitrification, whether Nar or Nap is involved, nitrite is reduced to nitric oxide by either a cytochrome* cd_1 *or a copper-containing nitrite reductase. This complexity underscores the importance of nitrate reduction as a key biological process.*

KEYWORDS:

bioinorganic chemistry · cell biology · metalloenzymes · nitrate reductase · nitrogen

Introduction

Organisms reduce nitrate for three principle reasons: to incorporate nitrogen into biomolecules, to generate energy for cellular function, and to dissipate excess energy by respiration (Figure 1).[1, 2] Although the chemical reaction is always the reduction of NO_3^- to NO_2^-, different types of nitrate reductase (NR) carry out the catalytic reaction. NRs can be distinguished by taxonomy, function, location in the cell, structure, and the end product of the pathway that they initiate. These reductases, with a few reported exceptions, are mononuclear molybdenum enzymes with a metal center at the active site coordinated by distinctive pyranopterin cofactors.[3] NR is found in plants, algae, fungi, archaea, and bacteria.[4–7] Thus, a distinction can be made at the taxonomic level, namely between eukaryotic and prokaryotic NR. This distinction is further substantiated by structural differences. Eukaryotic NR (Euk-NR) belongs to the sulfite oxidase family of mononuclear molybdenum enzymes and contains the simplified pyranopterin cofactor.[8] In contrast, all prokaryotic NRs described to date (Nas, Nap, Nar; described below) belong to the dimethylsulfoxide (DMSO) reductase family and contain a modified form of the cofactor.[8]

NR is primarily involved in nitrogen assimilation, respiration (in other words, dissimilatory nitrate reduction), or maintenance of a favorable cellular oxidation/reduction potential (redox poise).[1]

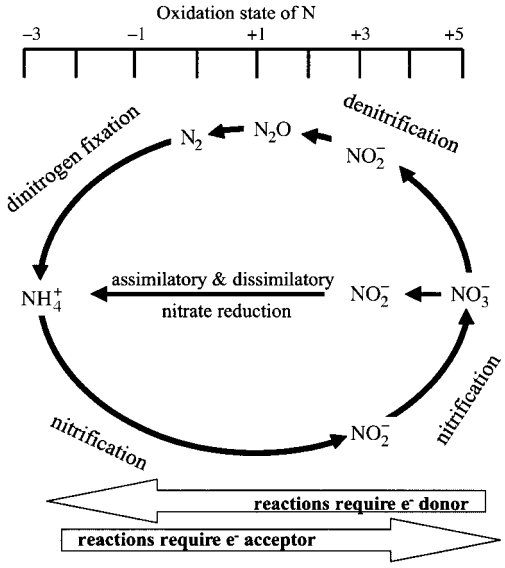


Figure 1. *The nitrogen cycle.*

[a] *Ass. Prof. P. Basu*
Department of Chemistry and Biochemistry
Duquesne University, Pittsburgh, PA 15282 (USA)
Fax: (+1) 412-396-5683
E-mail: basu@duq.edu

[b] *Prof. J. F. Stolz*
Department of Biological Sciences
Duquesne University, Pittsburgh, PA 15282 (USA)

  1439-4227/02/03/02-03 $ 17.50+.50/0 ChemBioChem **2002**, 3, 198 – 206

Assimilatory nitrate reductase is the first enzyme of an anabolic pathway for nitrogen incorporation into the biomass that maintains the bioavailability of NO_3^- to plants, algae, fungi, archaea and bacteria.[4, 9] In bacteria, the expression of Nas is induced by a lack of available NH_4^+ and the presence of NO_3^-.[9] Dissimilatory nitrate reductases catalyze the first step of a catabolic, anaerobic respiration pathway.[7, 10] Dissimilatory nitrate reduction (DNR) is a bacterial mode of energy generation under anoxic conditions and, hence, is not directly regulated by a nutritional requirement for nitrogen. Thus, DNR can result in the complete consumption of NO_3^- in the absence of other limitations such as a carbon source or an electron donor. In denitrification, nitrite is reduced to nitric oxide, which can further be reduced to nitrous oxide and dinitrogen by additional enzymes.[7, 11] In dissimilatory nitrate reduction to ammonia (DNRA), nitrite is usually reduced directly to ammonia by heme-containing nitrite reductases.[11] Although the subsequent steps involve different enzymes and intermediates, the first steps in the processes of DNRA and denitrification are identical and can be carried out by the same enzyme.

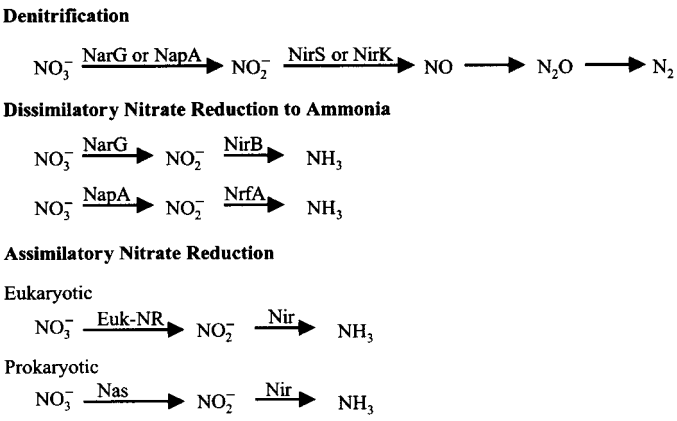
NR is localized in the cytoplasm of eukaryotes (EuK-NR) while in prokaryotes it may be cytoplasmic (prokaryotic assimilatory NR = Nas) or membrane-associated, facing either into the periplasm (prokaryotic periplasmic NR = Nap) or into the cytoplasm (membrane-associated prokaryotic NR = Nar). The location and orientation are crucial to function. Cytoplasmic NRs (both Euk-NR and Nas) are exclusively involved in nitrate assimilation and obtain their reducing equivalents from reduced nicotinamide adenine dinucleotide (NADH), reduced nicotinamide adenine dinucleotide phosphate (NADPH), ferredoxin, or flavodoxin.[2, 4, 9] Membrane-bound NRs that are oriented into the cytoplasm participate in the generation of proton motive force (PMF).[7, 11] The flow of electrons follows a conventional path in organisms like *Escherichia coli*.[11] Oxidation of the electron donor formate occurs in the periplasm. The electrons are shuttled to the intramembrane quinone pool where they are subsequently transferred to the NR active site by the *b*-type cytochrome (NarI) and iron-sulfur-containing subunit (NarH) of the NR. The nitrite produced is released into the cytoplasm. Periplasmic NRs present a dilemma in that they can be involved in redox balancing, respiration (either denitrification or DNRA), and nitrate-scavenging.[12–15] Electrons from the quinone pool are shuttled from a tetraheme cytochrome *c* (NapC) to the active site (NapA) by a biheme cytochrome *c* (NapB). Nap is believed not to contribute directly to PMF.[16] However, several nitrate-respiring bacteria possess only Nap, for example, *Pseudomonas* sp. strain G-179, and *Desulfovibrio desulfuricans*.[17, 18] Thus, the mechanism(s) of energy generation by Nap need(s) to be established.

The molecular mass, the number of subunits, and the prosthetic groups have been used to distinguish NRs structurally. Euk-NR is a multimer occurring in either a homodimeric form (for example, in plants and *Chlamydomonas*) or homotetrameric form (such as that in *Chlorella*).[4] The monomer has a mass of about 100 kDa and contains the molybdenum cofactor, a *b*-type heme, and flavin adenine dinucleotide (FAD). Electrons are believed to be transferred from NADH/NADPH to FAD and are then shuttled through cytochrome *b* to the active site. Bacterial Nas contains the molybdenum cofactor along with an iron–sulfur (Fe_4S_4) cluster. However, unlike in the Euk-NR, no heme or FAD is present in the active subunit.[9] Nas can occur in a monomeric form ranging in size from 70–95 kD, or as a heterodimer composed of the catalytic subunit (90–105 kDa) and an FAD-containing subunit called the diaphorase subunit.[16] At least three subgroups of Nas are recognized. In NAD(P)H-dependent Nas, electrons are transferred through the diaphorase subunit to the active site by iron–sulfur clusters.[1] The location and number of iron–sulfur clusters distinguish two types of NAD(P)H-dependent Nas. Two iron–sulfur clusters (Fe_4S_4, Fe_2S_2) are an integral part of the active subunit in *Klebsiella*. Only one Fe_4S_4 cluster is found in the active subunit in *Bacillus* but two Fe_2S_2 clusters are part of the FAD-containing subunit.[1] In ferredoxin-dependent Nas, which is found in *Synechococcus* sp., electrons are donated to the active site by a single iron–sulfur cluster.

Membrane-bound nitrate reductases from both nitrate-respiring bacteria such as *E. coli*, *Klebsiella pneumoniae*, and *Proteus mirabilis*, and from denitrifying bacteria such as *Paracoccus denitrificans*, *Pseudomonas stutzeri*, *Pseudomonas aeruginosa*, and *Bacillus halodenitrificans* have been characterized.[19–25] The species studied represent a wide taxonomic range, but their nitrate reductases all have similar characteristics. Several recent reviews have discussed the structural aspects of Nar in detail,[1, 10, 26] and thus only global features are highlighted here. The generalized model of the Nar consists of three subunits: a large (118–150 kDa) subunit which is the catalytically active site and contains the molybdenum cofactor, a smaller (55–64 kDa) anchor protein subunit, and an associated cytochrome *b* (19–21 kDa).[21, 22, 25, 27, 28] Periplasmic NR (Nap) has been purified from several species. The catalytic subunit ranges in size from about 70 kDa in species such as *D. desulfuricans*[29] to 90 kDa in, for example, *Rhodobacter capsulatus* and *Thiosphaera pantotropha*, now known as *Paracoccus denitrificans*.[30, 31] All the catalytic subunits contain a molybdenum cofactor as a part of the active site and an iron–sulfur cluster. In addition to this subunit, there is the 16 kDa biheme cyctochrome and an associated tetraheme cytochrome *c*.

The recent reports of non-molybdenum-containing nitrate reductases from *Pseudomonas isachenhovii*[32] and tungstate-dependent but molybdenum-inhibited nitrate reduction in the hyperthermophilic archaeon *Pyrobaculum aerophilum*[33] suggest the existence of alternative systems. *Ps. isachenhovii* contains two different molybdenum-free nitrate reductases. The periplasmic nitrate reductase has a monomer size of 55 kDa and contains vanadium. The membrane-bound nitrate reductase, in which no metal was found, is a dimer with 130 and 67 kDa subunits.[32] We have recently described a reductase complex from the dissimilatory iron-reducing bacterium *Geobacter metallireducens* that contains cytochrome *c* and can reduce both nitrate and nitrite but does not contain molybdenum.[34] However, for the purpose of this study we have focused on only molybdenum-containing nitrate reductases.

The end product of nitrate reduction may be ammonia as in assimilation and DNRA, or primarily dinitrogen as in denitrifica-



Scheme 1. *Pathways involving nitrate reduction. The nomenclature for dissimilatory nitrate and nitrite reductases is for the most part standardized: NarG = membrane-bound nitrate reductase, NapA = periplasmic nitrate reductase, NirS = cytochrome* cd_1 *nitrite reductase, NirK = copper-containing nitrite reductase, NirB = siroheme nitrite reductase, NrfA = pentaheme cytochrome* c *nitrite reductase. Assimilatory nitrate and nitrite reductases nomenclature is not standardized: Euk NR = eukaryotic nitrate reductase (Nia, NiaD, INR, Nit), Nas = cytoplasmic nitrate reductase (NasA, NasB, NasBB, NasC, NarB), Nir = siroheme nitrate reductase (Nir, NirB, NasA, NasC, NasD). Note that both NapA and NarG can be involved in nitrate reduction leading to either dinitrogen or ammonia. In contrast, Euk-NR and Nas are involved in nitrate reduction to ammonia only.*

tion (Scheme 1). The nitrite produced by NR is further reduced to either the end product ammonia or the denitrification intermediate nitric oxide. Four different types of nitrite reductases have been identified: a copper-containing type and three that contain a heme moiety.[7, 20, 35–37] Copper-containing nitrite reductase (NirK) is found exclusively in denitrifying bacteria and reduces nitrite to nitric oxide.[7] The three types of heme containing nitrite reductases are the cytochrome cd_1 type, the siroheme type, and the multiheme type that contains only heme C.[7, 11] The cytochrome cd_1 type (NirS), also found exclusively in denitrifying bacteria, is a homodimer comprised of two 62 kDa monomers that each contain one heme C and one heme D_1.[38, 39] *E. coli* has both siroheme (NirB) and multiheme (NrfA) nitrite reductase.[20, 39] NirB is coupled to Nar in DNRA, whereas the pentaheme nitrite reductase NrfA is associated with Nap (Scheme 1).[36, 37] Recent work has suggested that NrfA is a homodimer associated with an additional tetraheme cytochrome *c*,[40] although it was originally believed to be a single polypeptide of 65 kDa.

NR continues to generate a great deal of interest and several excellent reviews on the physiology, biochemistry, and molecular biology of nitrate reduction have been published in the last few years.[1, 2, 7, 9, 10, 16, 26] Previous sequence comparisons relied heavily on identification of regions of sequence identity in closely related enzymes with different functions and the inference of binding motifs to provide a broader view of mononuclear molybdenum enzymes. We used the wealth of protein sequence data that has recently become available to determine phylogenetic relationships and to identify structural/functional roles of conserved amino acids specific to NR. Our results are presented in the context of these past reviews.

Phylogenetic Analysis

Phylogenic analysis by using protein sequence provides insight into the origin and evolution of nitrate reductase. We used only complete sequences in this analysis. The increasing number of genome projects makes more sequence data available but caution must be exercised when accepting annotation without the accompanying physiological data.[41, 42] A recent review[1] of prokaryotic nitrate reductases listed 20 organisms with at least one Nar homologue, 23 organisms with a Nap homologue, and 24 organisms with a Nas homologue. Unfortunately, sequence data amenable to phylogenetic analysis was available for only about half of these species. The reason for the reduced data set was that either only a partial sequence was available (in some cases as few as 20 amino acid residues), or the identification was made from a preliminary genomic sequence without confirmation by physiological studies. It seems prudent to take a conservative approach given the great diversity of molybdenum-containing proteins and the relative paucity of sequence data for them. For our analyses, sequences were aligned by using the Clustal X program.[43] Maximum parsimony trees were generated, sequence similarities were determined with the PHYLIP program (version 3.572),[44] and neighbor-joining trees were generated by using PAUP software (version 4.0b7).[45]

The distinction between prokaryotic and eukayrotic nitrate reductases has been known for some time based on sequence data[46, 47] and the structure of the active site.[8] When all four types of NR are compared three major clades become apparent. These are the Euk-NR, Nar, and Nas/Nap clades (Figure 2). The Euk-NR clade is further divided into two subclades, a "plant" group and a "fungal" group, with further subdivisions following taxonomic classification (Figure 3). This conclusion is based on analysis of representative sequences for monocots (three species), dicots (thirteen species), algae (five species), ascomycetes (twelve species), basidiomyces (two species), yeast (two species), and an oomycete. This analysis included the two isoforms for each of *Nicotiana*, *Glycine*, and *Arabidopsis*. The isoforms had a high degree of sequence similarity (99.3, 95.5, and 94% respectively) and always branched together. In the final tree (Figure 3) only one isoform of *Arabidopsis* (Nia1) and *Nicotiana* (Nia2) were used. However, the two isoforms of *Glycine max* (INR1 and INR2) were included as an example. The "plant" group is comprised of the dicots, monocots, algae, and interestingly, the oomycete. The dicots form a tight cluster, as can be seen in Figure 3. Their sequences are highly conserved, sharing an identity of 43% and a similarity of 77%. The monocots and algae, with three species each, also appear to define their own subclades. The monocots, however, are more closely related to the dicots with 38% identity and 71% similarity. Algae have Euk-NR and not Nas (which is present in cyanobacteria) and this has implications for the symbiotic origin of chloroplasts. The chloroplast was derived from a cyanobacterial-like ancestor so it follows that Euk-NR originated from the host cell. While the Nas homologue was lost by the chloroplast, the assimilatory nitrite reductase (Nir) was retained. This process is reflected in the similarity between the Nir protein sequences and also in the location of the two processes in plants and algae: nitrate reduction occurs in the cell

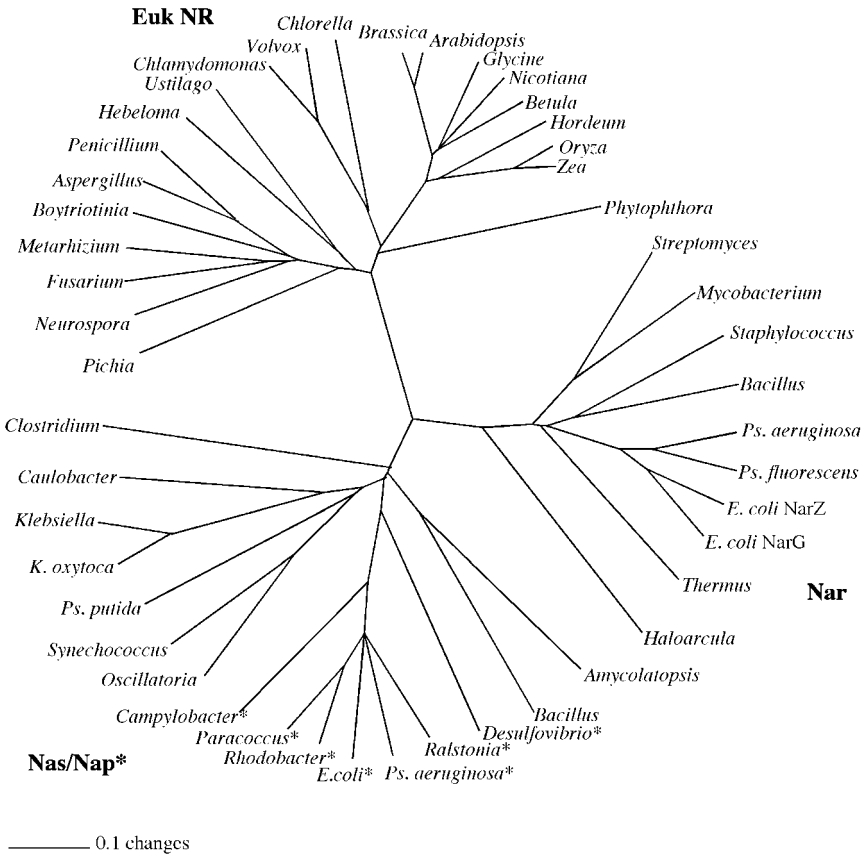


Figure 2. *Phylogeny of nitrate reductases (neighbor joining tree) showing three distinct clades: EukNR, Nar, and Nas/Nap. Nap sequences are denoted by *.*

cytoplasm whereas nitrite reduction occurs in the chloroplast.[48] Another interesting finding is the single example of an oomycete species. The fact that it forms a deep branch with the algae and plants lends support to the view that the oomycetes are more like flagellated algae (for example, chrysophytes) than fungi.[49] The "fungal" group appears to be divided into two distinct subclades with ascomycetes and yeast forming one and the basidiomycetes the other. The Euk-NR is highly conserved and follows taxonomic lines, which suggests to us that it originated from a single evolutionary event early on in the history of eukaryotes. The different number of subunits possible in the active enzyme (dimer, tetramer) and the different electron donors (ferredoxin or flavodoxin) were subsequent modifications.

Prokaryotic nitrate reductases form two major clades, the Nar clade and the Nas/Nap clade (Figure 4). *Haloarcula marismortui*, a member of the halophilic Euryarchaeota,[50] is the deepest branch of the Nar clade. Maximum parsimony trees could not distinguish a branching order for *Aeropyrum pernix*, a Crenarchaeota, and *Thermophilus thermophilum*, a deeply branching thermophilic member of the bacteria. This problem may arise because their sequences share an identity of 54% and a similarity of 81%. The high and low guanine + cytosine (G + C) Gram positive bacteria and the proteobacteria form separate subclades within the remaining species (Figure 4). The clustering together of denitrifying and DNRA bacteria is of the greatest significance. A comparison of the protein sequences of the two species of denitrifying pseudomonads (*Pseudomonas aeruginosa* and *Pseudomonas fluorescens*) with *E. coli* show a 64% identity and 86% similarity. The gene sequences share a 69% identity. Castrenana and Moreira[51] argue that this high degree of similarity indicates a common ancient lineage, an idea that is supported by Ralf and Imhoff based on their studies using NarH.[52] We believe, however, that the similarity strongly suggests horizontal gene transfer.[53] Indeed, Ramirez-Arcos et al. have demonstrated horizontal gene transfer of the *nar* cluster (the group of genes coding for Nar) in extreme thermophiles.[54] A strict aerobic strain of *Thermus thermophilus* was able to grow anaerobically on nitrate after conjugation with a nitrate-respiring strain. An origin for autonomous replication (*oriV*) that overlapped the last gene in the *nar* cluster was identified and suggested to be the mediator. A readily mobile *nar* cluster could explain the role of Nar in both denitrification and DNRA, its association with different nitrite reductases such as NirB, NirK, and NirS, the mosaic nature of denitrification gene organization,[28] and the presence of all three prokaryotic NRs in a single species.[13, 14]

Prokaryotic Nas and Nap are clearly related and together form the third clade of nitrate reductases. Nap is a distinct subclade. Each of the representatives of Nap possess the twin arginine motif, with the exception of *P. aeruginosa*. The amino acid sequence is highly conserved across phylogenetic lineages (based on 16S rRNA), with a 51% similarity. In most cases, Nap catalyzes the first step in nitrate reduction to ammonia but it can also be involved in denitrification,[17, 55] maintaining redox poise,[1] and nitrate scavenging.[15] Genetic analysis revealed that *nap* genes can be found on plasmids (for example, in *R. capsulatus, Rhodobacter sphaeroides, Ralstonia eutrophus*, and *Paracoccus denitrificans*) as well as at chromosome loci.[16] Thus Nap, like Nar, is an excellent candidate for horizontal gene transfer.[53] The divergence of Nap from the two strict anaerobes *Desulfovibrio desulfuricans* and *Campylobacteria jejuni* early on in this subclade provides an evolutionary link between Nap and Nas. As more sequences become available for strict anaerobes that can respire nitrate it will be important to see if they indeed create yet another distinct subclade. It is also tempting to speculate that if DNRA with Nap always involves NrfA, then strict anaerobes such as *Sulfurospirillum deleyianum*, *Sulfospirillum barnesii*, and *Geobacter metallireducens* should possess Nap homologues as well. Nas seems to have the least conserved sequence of all the bacterial nitrate reductases, with only 19% similarity and several deeply branching subclades (Figure 4). There also does not appear to be any specific clustering based on preferred electron

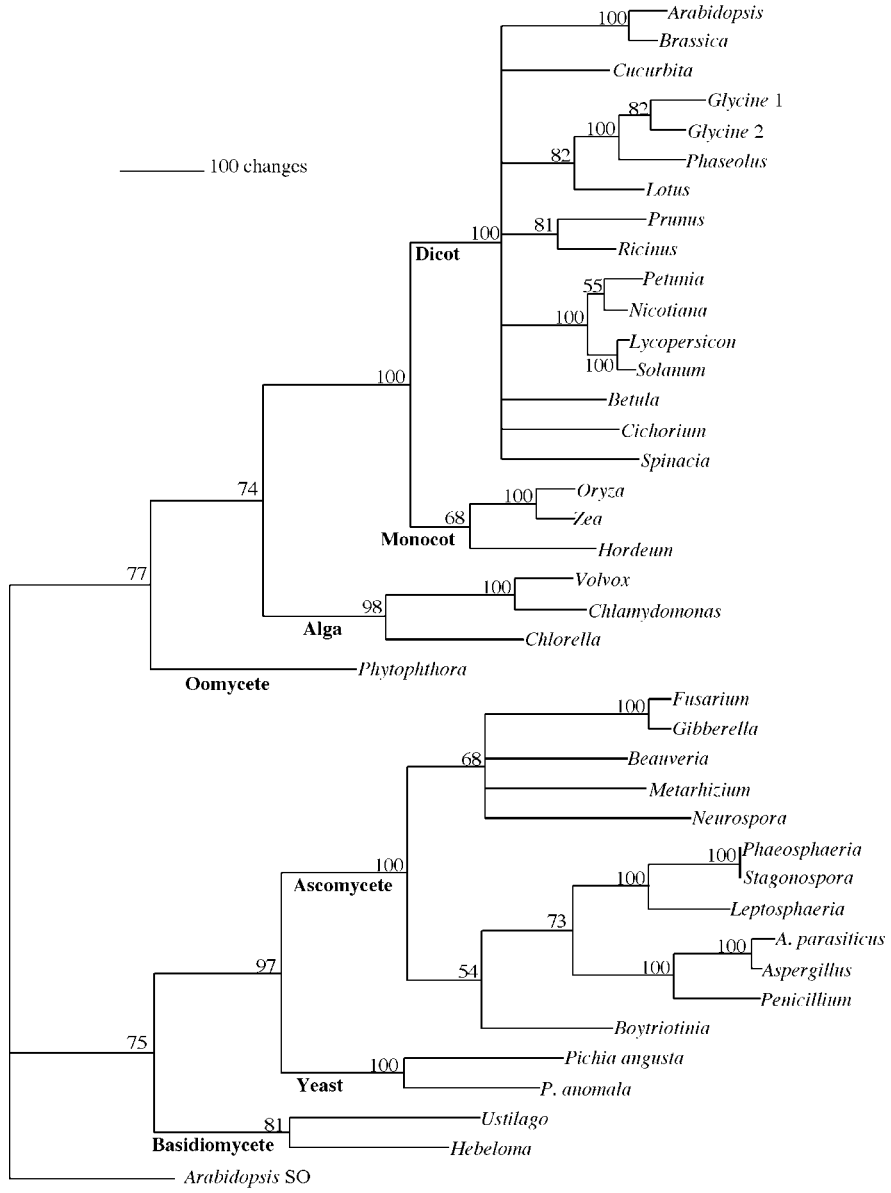


Figure 3. *Phylogeny of eukaryotic assimilatory nitrate reductase (maximum parsimony). Bootstrap values are noted at the nodes. The following sequences were used: Plants:* Arabidopsis thaliana *Nia1 Z19050,* Betula pedula *X54097,* Brassica napus *D38220,* Cichorium intybus *X84103,* Cucurbita maxima *A41667,* Glycine max *INR1 L23854,* Glycine max *INR2 U13987,* Hordeum vulgare *X60173,* Lotus japonicus *X80670,* Lycopersicon esculentum *X14060,* Nicotiana tabacum *Nia2 X14059,* Oryza sativa *P16081,* Petunia hybrida *L11563,* Phaseolus vulgaris *X53603,* Prunus persica *AB061670,* Ricinus communis *AF314093,* Solanum tuberosum *U95317,* Spinacia oleracea *D86226,* Triticum aestivum *BAB11739,* Zea mays *AF153448. Algae:* Chlamydomonas reinhardtii *AF203033,* Chlorella vulgaris *U39931,* Volvox carteri *X64136; Oomycete:* Phytophthora infestans *U14405. Fungi:* Aspergillus oryzae *D49701,* Aspergillus parasiticus *U38948,* Beauveria bassiana *X84950,* Botryotinia fuckeliana *U43783,* Fusarium oxysporum *Z22549,* Gibberella fujikuroi *X90699,* Hebeloma cylindrosporum *AJ238664,* Leptosphaeria maculans *U044405,* Metarhizium anisopliae *AJ001141,* Neurospora crassa *S16292,* Penicillium chrysogenum *U20779,* Phaeosphaeria nodorum *Y13654,* Pichia angusta *Z49110,* Pichia anomala *AF123281,* Stagonospora nodorum *AJ009827,* Ustilado maydis *X67687; Sulfite oxidase:* Arabidopsis thaliana *AF200972.*

donor (ferredoxin- or NAD(P)H-dependent), taxonomic classification, or 16S rRNA phylogeny. Cyanobacteria do cluster together, but clearly have a Nas homologue, which provides further evidence that the prokaryotic assimilatory nitrate reductase was an early victim of the chloroplast symbiosis at least for the green algae. The low degree of sequence similarity among a great diversity of bacteria suggests that Nas is a rapidly evolving protein.

Structural analysis

The majority of the nitrate reductases are mononuclear molybdenum enzymes. The molybdenum atom is coordinated by the sulfur donors of the ene-1,2-dithiolate (dithiolene) moiety from a unique pyranopterin cofactor. Crystallography indicates that the pterin cofactor adopts a tricyclic pyran ring structure rather than the originally proposed bicyclic ring structure.[56] The basic structure of the pyranopterin cofactor with a phosphate group is retained for Euk-NR and sulfite oxidase. Currently, no complete crystal structure of any Euk-NR is known, however, the crystal structure of chicken liver sulfite oxidase (CLSO)[57] has confirmed the formulation of the cofactor as proposed. In prokaryotic proteins the phosphate group originating from the 2-position is replaced by guanine diphosphate (Scheme 2). The formulation of the cofactor has been crystallographically demonstrated for Nap[58] and represents enzymes of the Nas/Nap clade. At present no crystal structure of membrane-bound nitrate reductase (Nar clade) or membrane-bound DMSO reductase is known, although these structures are expected to be similar to that of Nap. Clearly, there is a significant difference between the cofactor formulation of the proteins from eukaryotic and prokaryotic organisms. The functional ramification of this diversity has yet to be elucidated.

Although the complete crystal structure of Euk-NR has yet to be determined, the structure of the cytochrome and the FAD domains have been published.[4] Furthermore, the crystal structure of a structurally similar enzyme, sulfite oxidase, is established. We aligned the sequences of Euk-NRs with three different sulfite oxidases (chicken, rat, and *Arabidopsis*). The high degree of sequence conservation among Euk-NR (Figure 5) made comparisons useful. When only Euk-NR sequences were compared, twelve conserved residues were identified as "cofactor-binding residues" (Figure 5). These residues were identified through comparison with the 14 residues involved in cofactor binding in CLSO. The similarity is remarkable in view of the various types of organisms from which the sequences are

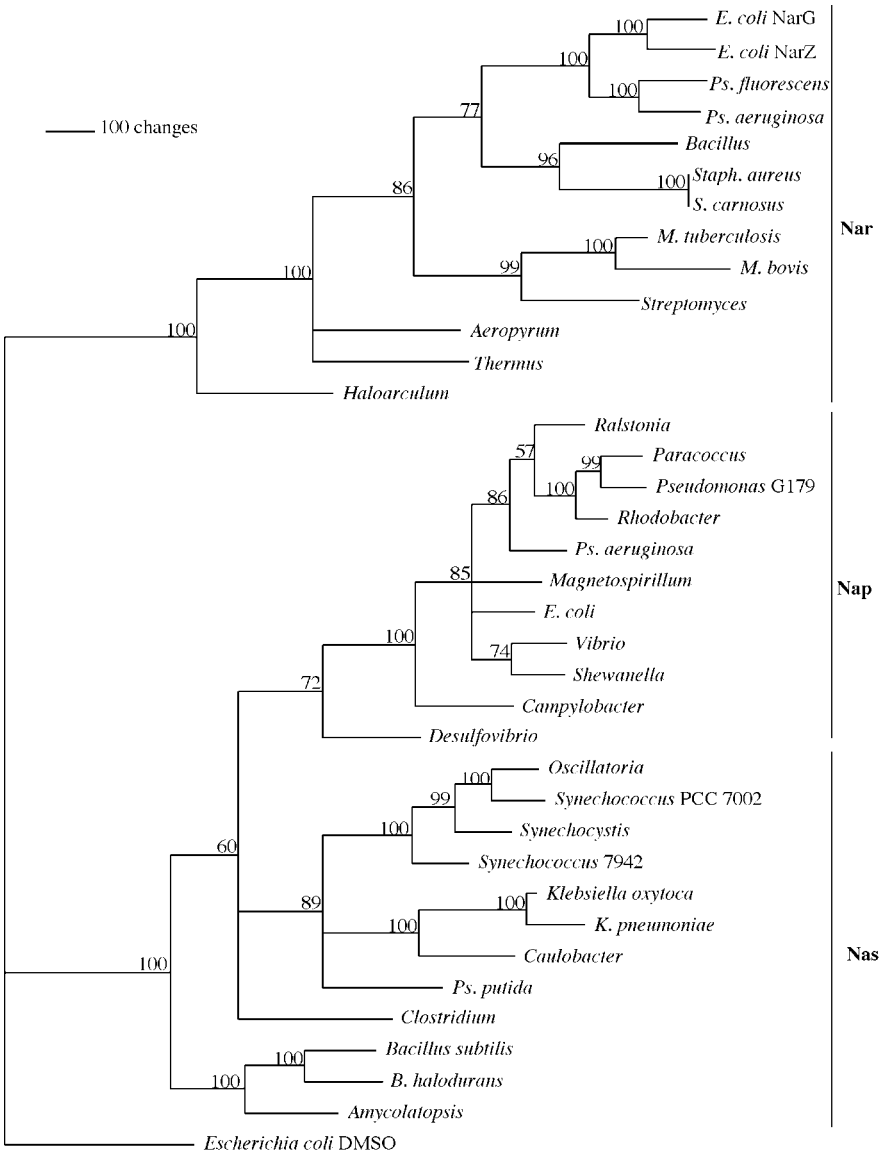
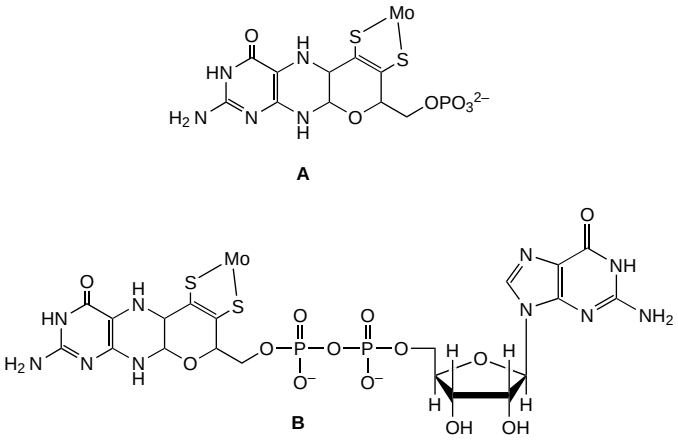


Figure 4. *Phylogeny of prokaryotic nitrate reductases (maximum parsimony). Bootstrap values are noted at the nodes. The following sequences were used: Nar:* Aeropyrum pernix *AP000061,* Bacillus subtilis *Z49884,* Escherichia coli *NarG X16181,* Escherichia coli *NarZ P19319,* Haloarcula marismortuii *AJ277440,* Mycobacterium tuberculosus *AAK45455,* Mycobacterium bovis *AF149772,* Pseudomonas aeruginosa *Y15252,* Pseudomonas fluorescens *U71398,* Staphylococcus aureus *AF029225,* Staphylococcus carnosus *AF029224,* Streptomyces coelicolor *CAC18712,* Thermus thermophilus *Y10124; Nap:* Campylobacter jejuni *Cj0780,* Desulfovibrio desulfuricans *Y18045,* Escherichia coli *P33937,* Magnetospirillum magnetotacticum *BAB59022,* Paracoccus pantatrophus *Z36773,* Pasteurella multicida *AAK03678,* Pseudomonas *G-179 AF040988,* Pseudomonas aeruginosa *F83499,* Ralstonia eutropha *X71385,* Rhodobacter sphaeroides *f. denitrificans AF06954,* Shewanella putrifaciens *TIGR,* Vibrio cholera *D82430; Nas:* Amycolatopsis mediterranei *AJ298195,* Bacillus halodurans *BAB04334,* Bacillus subtilis *D30689,* Caulobacter crescentus *AAK22602,* Clostridium perfringens *AB017192,* Klebsiella oxytoca *L08600,* Klebsiella pneumonia *Q06457,* Oscillatoria chalybea *X89445,* Pseudomonas putida *AF203789,* Synechococcus *sp. PCC7942 S36605,* Synechococcus *sp. PCC7002 AAD45942,* Synechocytis *sp. PCC6803 S77385. DMSO:* Escherichia coli *P18775.*

derived and strongly suggests that the cofactor-binding residues remain conserved during evolution. None of the EuK-NR sequences have an iron-sulfur cluster binding motif (such as CXXCXXXC; X = undefined amino acid). Physiochemical studies indicate a heme center is present as the partner prosthetic group, a feature crystallographically confirmed in CLSO. In CLSO the heme is located at the N terminus and H40 and H65 coordinate to the iron center. The sequences of eukaryotic Euk-NR do not show any conserved histidine at the N terminus that can be positively identified as a ligand to the iron. Campbell reported that the *b*-heme-binding domain is located at the C terminus and identified two histidines as the heme-binding residues.[4] Our analysis with the additional sequences concur with this prediction as these residues are conserved (Figure 5).

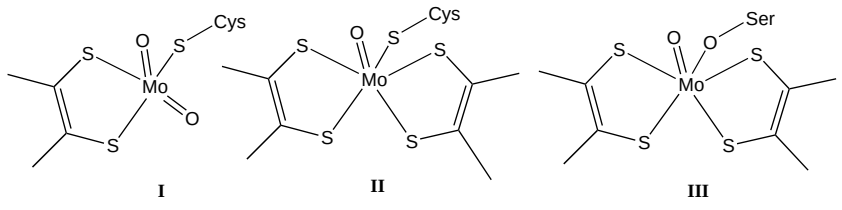
Specific structural features at the coordination sphere of molybdenum also differentiate NRs. Euk-NR is characterized by coordination of the molybdenum center by a single pyranopterin cofactor through a dithiolene linkage (Scheme 3).[59] A second defining structural feature of Euk-NR is a mono-oxo Mo center in the reduced, resting state of the enzyme. This feature was derived from the structure of CLSO and by extended X-ray absorption fine structure (EXAFS) spectroscopy at the molybdenum K-edge of Euk-NR from *Arabidopsis.*[59]

To date, the complete crystal structure of Nap has only been determined for *D. desulfuricans*. We used the *D. desulfuricans* structural data to predict the roles of the conserved residues within the Nas/Nap clade since the clade shows considerable sequence similarities. The crystal structure of Nap indicated 29 residues that are involved in binding the two cofactor molecules. Comparisons with other Nap enzymes showed that 21 of these 29 residues (72%) were conserved (Figure 6). This is significant because the overall identity between Nap sequences is only 22%. All Nap sequences have an iron–sulfur cluster with a $CX_1X_2CX_3X_4X_5C$ sequence at the N terminus,[1, 13, 26] where X_1 = R, X_2 = F or Y, X_3 = G, X_4 = T or V, and X_5 = G. The entire motif is robustly conserved. No other cofactor binding sequence could be positively identified, which suggests that in Nap enzymes there are two prosthetic groups: the pyranopterin cofactor and the iron–sulfur cluster. In addition, there is one conserved cysteine residue that coordinates the molybdenum atom. The cysteine ligation to molybdenum already predicted by other researchers[8, 13] was crystallographically confirmed in *D. desulfuricans.* All Nap proteins have a twin arginine motif that is believed to be required for placement of the protein into the periplasm,[60] with the exception of



Scheme 2. *Eukaryotic (A) and prokaryotic (B) molybdenum cofactors.*

Figure 5. *37 different Euk-NR sequences aligned with that of the* Arabidopsis *NIA1 Euk-NR. Identities are bold black; similarities (both strong and weak) are italicized bold grey; residues with structural roles are underlined; residues that are thought to be involved in cofactor binding are represented with a different (larger) font type.*



Scheme 3. *Active site structures of NRs: **I**, eukaryotic (Euk-NR); **II**, periplasmic (Nap); **III**, membrane bound (Nar).*

P. aeruginosa. The Nas proteins are the most diverse with an extremely low overall sequence conservation of less than 1% identity and 19% similarity (Figure 7). Even among the Gram negative species the sequence similarity is low (17% identity, 37% similarity). Nevertheless, a highly conserved iron–sulfur cluster binding motif, CPYCGVGCG, at the N terminus could be observed in all Nas.[9] This observation suggested it was the site of the Fe_4S_4 iron–sulfur cluster that is common to all Nas. A cysteine residue that could ligate to the molybdenum center was also found to be conserved. When Nas was compared with Nap, five residues that might be involved in cofactor binding were identified (Figure 7). Interestingly, the low number of binding residues may indicate that even though Nas clusters with Nap, there may be only one cofactor per molybdenum atom, verification of which must await structure determination.

In the absence of a crystal structure for Nar, we used the structural data of Nap to identify putative cofactor binding residues. This comparison led us to identify 16 highly conserved residues that might be involved in the cofactor binding (Figure 8). If this is true then the large number of amino acid residues that are involved in binding the cofactor supports the notion that Nar has two pyranopterin cofactors. Nar is also thought to have an iron–sulfur cluster near the N terminus. While a conventional binding motif could not be found, a highly conserved sequence HG(V)NCTGSC was detected. This sequence is not typical for an iron–sulfur cluster and may be involved in binding of the molybdenum cofactor.[61] Indeed, when the histidine was replaced with a cysteine in *E. coli*, the resulting protein failed to bind an iron–sulfur cluster.[62] Another interesting feature is that there is a conserved serine (not cysteine) that may coordinate the molybdenum. Although, X-ray absorption spectroscopy[63] has been used to investigate the active site structure, biochemical or spectroscopic confirmation of the serine coordination has yet to be found.[1] If, however, serine really does coordinate to the molybdenum atom, it represents yet another variation in the active-site structure of nitrate reductase.

The defining structural features of Nar and Nap place them in the DMSO reductase (DMSOR) family of Mo enzymes. The Mo atom is coordinated by two pyranopterin molecules through dithiolene linkages (Scheme 3) and is desoxo in the reduced, catalytically active state. The issue of whether the oxidised state is monooxo or dioxo in DMSOR has been intensely debated in recent years, in part due to different interpretations of structural studies. For example, the crystal structures of DMSOR isolated from *Rhodobacter sphaeroides* were interpreted in terms of desoxo Mo^{IV} and monooxo Mo^{VI} centers.[64] However, the crystal structures of enzymes isolated from *Rhodobacter capsulatus* were interpreted as having monooxo Mo^{IV} and dioxo Mo^{VI} centers.[65] Subsequent resonance Raman (rR) and EXAFS spectroscopy at the molybdenum K-edge on the two enzymes supports the views derived from the structural studies.[66] Recently, a very high-resolution structure indicated that plasticity of the structure is heavily dependent on experimental conditions

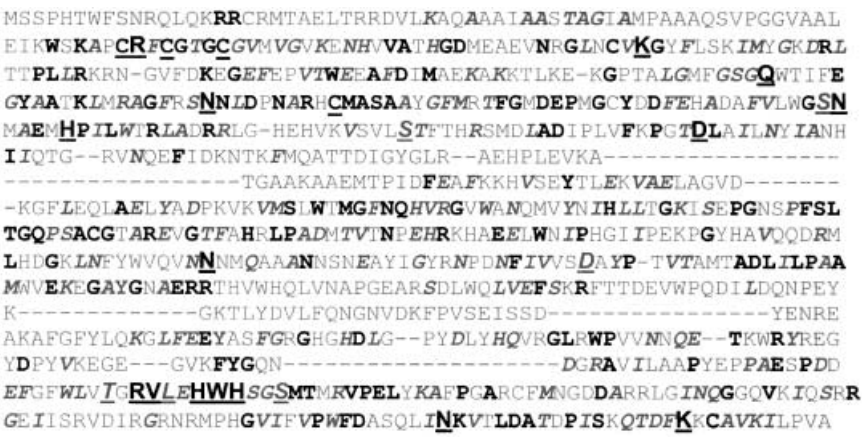


Figure 6. *Nap sequences from 10 species of bacteria aligned with* Pseudomonas *sp. G-179 Nap. See Figure 5 legend for key.*

Figure 7. *Nas sequences from 13 species of bacteria aligned with* Klebsiella oxytoca *NasA.* Klebsiella oxytoca, Klebsiella pneumonia, Caulobacter cresensus, *and* Pseudomonas aeroginosa *strain PAO1 have an extension of about 160 amino acids at the C terminus. See Figure 5 legend for key.*

Figure 8. *Nar sequences from 11 species of bacteria aligned with* Escherichia coli *NarG. See Figure 5 legend for key.*

such as the buffer.[64a] Certainly, these studies provided a basis for a catalytic transformation between a desoxo molybdenum(IV) and a monooxo molybdenum(VI) center. This important concept can also be applied to Nap.[58] Indeed, in Nap, four sulfur donors from two pyranopterin cofactors as well as the sulfur atom from the cysteine residue coordinate the molybdenum atom. Although no terminal oxo group could be located, a water molecule does coordinate the metal center.[58] No crystallographic or spectroscopic data is currently available on any Nas protein. It is tempting to speculate from the cofactor binding motif analysis that Nas has only one pyranopterin cofactor, however, investigation of this proposal must rely on experimental evidence.

Summary and Outlook

The dramatic increase in the amount of sequence data available made it possible to make a comparative study of the different nitrate reductases. Our analyses indicate three different clades of nitrate reductase: Euk-NR, Nar, and Nas/Nap. Nap from strict anaerobes provides the evolutionary link between the closely related Nas and Nap. Our results also suggest that the gene encoding Euk-NR might be useful for the study of algal, plant, and fungal evolution. Whereas Euk-NR appears to have emerged early on in the evolution of eukaryotes, Nar and Nap are prime candidates for horizontal gene transfer. Thus additional sequences of Nap and Nar especially from strict anaerobes and archaea are of great interest. Sequence alignments coupled with crystallographic information were used to predict which residues could be involved in cofactor and metal binding. Confirmation of this prediction awaits site-directed mutagenesis studies. Genome-sequencing projects offer the exciting prospect of additional sequences but need to be complemented with physiological, biochemical, structural, and spectroscopic studies. This investigation points out the serious lack of biophysical information on Nas and also of detailed kinetic studies on Nap and Nar in denitrification and DNRA. Work also needs to be done to determine the mechanism by which nitrate respiration with Nap generates PMF. Furthermore, we anxiously await the determination of the crystal structures of Euk-NR, Nar, and Nas.

Thanks are due to Dr. J. Lawrence, Dr. D. Beer Stolz, Dr. M. Melan, Dr. M. Alleman, Dr. R. A. Rothery, Dr. R. Hille, and Mr. P. R. Dowdle for stimulating discussions. We also thank the National Institutes of Health (P.B.) and the National Science Foundation (J.F.S.) for financial support to our research programs.

[1] D. J. Richardson, B. C. Berks, D. A. Russell, S. Spiro, C. J. Taylor, *Cell. Mol. Life Sci.* **2001**, *58*, 165 – 178.
[2] W. H. Campbell, *Cell. Mol. Life Sci.* **2001**, *58*, 194 – 204.
[3] Note that the pyranopterin cofactor (sometimes referred to as molybdopterin) does not contain molybdenum. The molybdenum-bound form is called molybdenum cofactor. For a detailed discussion, see: B. Fischer, J. H. Enemark, P. Basu, *J. Inorg. Biochem.* **1998**, *72*, 13 – 21.
[4] W. H. Campbell, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 277 – 303.
[5] S. Ramirez-Arcos, L. A. Fernandez-Herrero, I. Marin, J. Berenguer, *Biochim. Biophys. Acta* **1997**, *1396*, 215 – 227.
[6] P. Volkl, R. Huber, E. Drobner, R. Rachel, S. Buggraf, A. Trincone, K. O. Stetter, *Appl. Environ. Microbiol.* **1993**, *59*, 2918 – 2926.
[7] W. G. Zumft, *Microbiol. Mol. Biol. Rev.* **1997**, *61*, 533 – 616.
[8] R. Hille, *Chem. Rev.* **1996**, *96*, 2757 – 2816.
[9] J. T. Lin, V. Stewart, *Adv. Microb. Physiol.* **1998**, *39*, 1 – 30.
[10] F. Blasco, B. Guigliarelli, A. Maalon, M. Asso, G. Giordano, R. A. Rothery, *Cell. Mol. Life Sci.* **2001**, *58*, 179 – 193.
[11] B. Berks, S. J. Ferguson, J. W. B. Moir, D. J. Richardson, *Biochem. Biophys. Acta* **1995**, *1232*, 97 – 173.
[12] J. Carter, D. J. Richardson, S. Spiro, *Arch. Microbiol.* **1995**, *163*, 159 – 166.
[13] B. C. Berks, D. J. Richardson, A. Reilly, A. C. Willis, S. J. Ferguson, *Biochem. J.* **1995**, *309*, 983 – 992.
[14] D. M. M. Castillo, F. Reyes, R. Blasco, M. D. Roldan, M. Gavira, F. J. Caballero, C. Moreno-Vivian, *Curr. Microbiol.* **1996**, *33*, 341 – 346.
[15] L. C. Potter, P. Millington, L. Griffiths, G. H. Thomas, J. A. Cole, *Biochem. J.* **1999**, *344*, 77 – 84.
[16] C. Moreno-Vivian, P. Cabello, M. Martinez-Luque, R. Blasco, F. Castillo, *J. Bacteriol.* **1999**, *181*, 6573 – 6584.
[17] L. Bedzyk, T. Wang, R. W. Ye, *J. Bacteriol.* **1999**, *181*, 2802 – 2806.
[18] S. Bursakov, M-Y. Liu, W. J. Payne, J. LeGall, I. Moura, J. J. G. Moura, *Anaerobe* **1995**, *1*, 55 – 60.
[19] J. A. Cole in *The Nitrogen and Sulfur Cycles* (Eds.: J. A. Cole, S. J. Ferguson), Cambridge University Press, Cambridge, **1988**, pp. 281 – 329.
[20] J. A. Cole, *FEMS Microbiol. Lett.* **1996**, *136*, 1 – 11.
[21] V. Stewart, *Microbiol. Rev.* **1988**, *52*, 190 – 232.
[22] A. D. Moodie, W. J. Ingledew, *Adv. Micro. Phys.* **1990**, *31*, 225 – 269.
[23] H. Stouthamer in *Biology of Anaerobes* (Ed.: M. Zehnder), Blackwell Scientific, Boston, **1988**, pp. 245 – 303.
[24] J. M. Tiedje in *Biology of Anaerobes* (Ed.: M. Zehnder), Blackwell Scientific, Boston, **1988**, pp. 179 – 244.
[25] P. A. Ketchum, W. J. Payne, *Appl. Environ. Microbiol.* **1992**, *58*, 3586 – 3592.
[26] L. Philippot, O. Hojberg, *Biochim. Biophys. Acta.* **1999**, *1446*, 1 – 23.
[27] P. A. Ketchum, G. Denariaz, J. LeGall, W. J. Payne, *J. Bacteriol.* **1991**, *173*, 2498 – 2505.
[28] W. G. Zumft, H. Koerner, *Antonie van Leeuwenhoek* **1997**, *71*, 43 – 58.
[29] S. Bursakov, M.-Y. Liu, W. J. Payne, J. LeGall, I. Moura, J. J. G. Moura, *Anaerobe* **1995**, *1*, 55 – 60.
[30] D. J. Richardson, A. G. McEwan, M. D. Page, J. B. Jackson, S. J. Ferguson, *Eur. J. Biochem*. **1990**, *194*, 263 – 270.
[31] B. Berks, D. J. Richardson, C. Robinson, A. Reilly, R. T. Aplin, S. J. Ferguson, *Eur. J. Biochem*. **1994**, *220*, 117 – 124.
[32] A. N. Antipov, N. N. Lyalikova, T. V. Khijniak, N. P. L'vov, *FEBS Lett.* **1998**, *444*, 257 – 260.
[33] S. Afshar, C. Kim, H. G. Monbouquette, I. Schroeder, *Appl. Environ. Microbiol.* **1998**, *64*, 3004 – 3008.
[34] F. Martinez Murillo, T. Gugliuzza, J. Senko, P. Basu, J. F. Stolz, *Arch. Microbiol.* **1999**, *147*, 1 – 14.
[35] S. Kajie, Y. Anraku, *Eur. J. Biochem.* **1986**, *154*, 457 – 463.
[36] M.-C. Liu, H. D. Peck, Jr., *J. Biol. Chem.* **1981**, *256*, 13 159 – 13 164.
[37] W. Schumacher, U. Hole, P. M. H. Kroneck, *Biochem. Biophys. Res. Comm.* **1994**, *205*, 911 – 916.
[38] G. B. Smith, J. M. Tiedje, *Appl. Environ. Microbiol*. **1992**, *58*, 376 – 384.
[39] T. Brittain, R. Blackmore, C. Greenwood, A. J. Thomson, *Eur. J. Biochem.* **1992**, *209*, 793 – 802.
[40] O. Einsle, A. Messerschmidt, P. Stach, G. P. Bourenkov, H. D. Bartunik, R. Huber, P. M. H. Kroneck, *Nature* **1999**, *400*, 476 – 480.
[41] J. A. Gerlt, P. C. Babbitt, *Genome Biology* **2000**, *1*, 1 – 10.
[42] K. E. Nelson, I. T. Paulsen, J. F. Heidelberg, C. M. Fraser, *Nat. Biotechnol.* **2000**, *18*, 1049 – 1054.
[43] D. G. Higgins, P. M. Sharp, *Comput. Appl. Biosci.* **1989**, *5*, 151 – 153; D. G. Higgins, *Adv. Protein Chem.* **2000**, *54*, 99 – 135.
[44] J. Felsenstein, *Annu. Rev. Genet*. **1988**, *22*, 521 – 526.
[45] D. L. Swofford, PAUP* (Phylogenetic Analysis Using Parsimony; *and Other Methods), Version 4, Sinauer Associates, Sunderland, Massachusetts, **1998**.
[46] W. H. Campbell, J. H. Kinghorn, *Trends Biochem. Sci.* **1990**, *15*, 315 – 319.
[47] W. H. Campbell, *Plant Physiol*. **1996**, *111*, 355 – 361.
[48] R. Tischner, *Plant Cell Environ.* **2000**, *23*, 1005 – 1024.
[49] a) G. Van der Auwera, S. Chapelle, R. De Wachter, *FEBS Lett*. **1994**, *338*, 133 – 136; b) G. Van der Auwera, R. De Baere, Y. Van de Peer, P. De Rijk, R. De Wachter, *Mol. Biol. Evol.* **1995**, *12*, 671 – 678; c) J. H. Gunderson, H. Elwood, A. Ingold, K. Kindle, M. L. Sogin, *Proc. Natl. Acad. Sci.* **1987**, *84*, 5823 – 5827.
[50] K. Yoshimatsu, T. Sakurai, T. Fujiwara, *FEBS Lett.* **2000**, *470*, 216 – 220.
[51] J. Castresana, D. Moreira, *J. Mol. Evol.* **1999**, *49*, 453 – 460.
[52] P. Ralf, J. F. Imhoff, *Syst. Appl. Microbiol.* **2000**, *23*, 47 – 57.
[53] J. G. Lawrence, *Curr. Opin. Genet. Dev.* **1999**, *9*, 642 – 648.
[54] S. Ramirez-Arcos, L. A. Fernandez-Herrero, I. Marin, J. Berenguer, *J. Bacteriol*. **1998**, *180*, 3137 – 3143.
[55] R. A. Siddiqui, U. Warneck-Eberz, A. Hengsberger, B. Schneider, B. Friedrich, *J. Bacteriol*. **1993**, *175*, 5867 – 5876.
[56] J. L. Johnson, K. V. Rajagopalan, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6856 – 6860; J. L. Johnson, B. E. Hainline, K. V. Rajagopalan, B. H. Arison, *J. Biol. Chem.* **1984**, *259*, 5414 – 5422.
[57] C. Kisker, H. Schindelin, A. Pacheco, W. W. Wehbi, R. M. Garrett, K. V. Rajagopalan, J. H. Enemark, D. C. Rees, *Cell* **1997**, *91*, 973 – 983.
[58] J. M. Dias, M. E. Than, A. Humm, R. Huber, G. P. Bourenkov, H. D. Bartunik, S. Bursakov, J. Calvete, J. Caldeira, C. Carneiro, J. J. G. Moura, I. Moura, M. J. Romão, *Structure* **1999**, *7*, 65 – 79.
[59] G. N. George, J. A. Mertens, W. H. Campbell, *J. Am. Chem. Soc.* **1999**, *121*, 9730 – 9731.
[60] B. C. Berks, F. Sargent, T. Palmer, *Mol. Microbiol.* **2000**, *35*, 260 – 274; B. C. Berks, *Mol. Microbiol.* **1996**, *22*, 393.
[61] C. A. Trieber, R. A. Rothery, J. H. Weiner, *J. Biol. Chem.* **1996**, *271*, 4620 – 4626; J. H. Weiner, R. A. Rothery, D. Sambassivarao, C. A. Trieber, *Biochim. Biophys. Acta* **1992**, *1102*, 1 – 18.
[62] A. Magalon, M. Asso, B. Guigliarelli, R. A. Rothery, P. Bertrand, G. Giordano, F. Blasco, *Biochemistry* **1998**, *37*, 7363 – 7370.
[63] G. N. George, N. A. Turner, B. C. Bray, F. F. Morpeth, D. H. Boxer, S. P. Cramer, *Biochem. J.* **1989**, *259*, 693 – 700.
[64] a) H.-K. Li, C. Temple, K. V. Rajagopalan, H. Schindelin, *J. Am. Chem. Soc.* **2000**, *122*, 7673 – 7680; b) H. Schindelin, C. Kisker, J. Hilton, K. V. Rajagopalan, D. C. Rees, *Science* **1996**, *272*, 1615 – 1621.
[65] a) A. S. McAlpine, A. G. McEwan, S. Bailey, *J. Mol. Biol.* **1998**, *275*, 613 – 623; b) A. S. McAlpine, A. G. McEwan, A. L. Shaw, S. Bailey, *J. Biol. Inorg. Chem.* **1997**, *2*, 690 – 701; c) F. Schneider, J. Loewe, R. Huber, H. Schindelin, C. Kisker, J. Knaeblein, *J. Mol. Biol.* **1996**, *263*, 53 – 69; d) J. Knablein, H. Dobbek, S. Ehlert, F. Schneider, *J. Biol. Chem.* **1997**, *378*, 293 – 302.
[66] S. D. Garton, J. Hilton, H. Oku, B. R. Crouse, K. V. Rajagopalan, M. K. Johnson, *J. Am. Chem. Soc.* **1997**, *119*, 12 906 – 12 916; b) M. K. Johnson, S. D. Garton, H. Oku, *J. Biol. Inorg. Chem.* **1997**, 2, 797 – 803; c) A. F. Bell, X. He, J. P. Ridge, G. R. Hanson, A. G. McEwan, P. J. Tonge, *Biochemistry* **2001**, *40*, 440 – 448; d) P. E. Baugh, C. D. Garner, J. M. Charnock, D. Collison, E. S. Davies, A. S. McAlpine, S. Bailey, I. Lane, G. R. Hanson, A. G. McEwan, *J. Biol. Inorg. Chem*. **1997**, *2*, 634 – 643; e) G. N. George, J. Hilton, C. Temple, R. C. Prince, K. V. Rajagopalan, *J. Am. Chem. Soc.* **1999**, *121*, 1256 – 1266; f) G. N. George, J. Hilton, K. V. Rajagopalan, *J. Am. Chem. Soc.* **1996**, *118*, 1113 – 1117.

Received: September 13, 2001 [F 296]