Origin and phylogenetic relationships of [4Fe-4S]-containing O$_2$-sensors of bacteria

C. Barth$^1$, M. C. Weiss$^2$, M. Roettger$^2$, W. F. Martin$^2$, G. Unden$^1$

$^1$Microbiology and Wine Research, Institute for Molecular Physiology, Johannes Gutenberg University Mainz, Mainz, Germany

$^2$Institute for Molecular Evolution, Heinrich Heine University of Düsseldorf, Düsseldorf, Germany

* Corresponding author
Gottfried Unden,
Microbiology and Wine Research, Institute for Molecular Physiology, Johannes Gutenberg University, Becherweg 15, 55099 Mainz, Germany
unden@uni-mainz.de, Phone +49-6131-3923550
Fax +49-6131-3922695

Running title:
Origin of iron-sulfur containing O$_2$-sensors

Keywords:
O$_2$-sensor, iron-sulfur, FNR, NreB, Escherichia coli, Bacillus subtilis

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.14411
Originality-Significance statement

Bacterial sensors for O$_2$ developed early in evolution, and a significant number of sensors (FNR$_{Ec}$, FNR$_{Bs}$, NreB, WhiB3) use [4Fe4S] to [2Fe2S] cluster conversion for sensing. The manuscript shows that the sensors have no common origin and developed independently. Information about the chemistry of the iron-sulfur clusters suggests that their functional sensory mechanism (but not the signal output mechanism) is independent of the protein type.

Summary

The advent of environmental O$_2$ about 2.5 billion years ago forced microbes to metabolically adapt and to develop mechanisms for O$_2$-sensing. Sensing of O$_2$ by [4Fe-4S]$^{2+}$ to [2Fe-2S]$^{2+}$ cluster conversion represents an ancient mechanism that is used by FNR$_{Ec}$ (Escherichia coli), FNR$_{Bs}$ (Bacillus subtilis), NreB$_{Sa}$ (Staphylococcus aureus), and WhiB3$_{Mt}$ (Mycobacterium tuberculosis). The phylogenetic relationship of the sensors was investigated. FNR$_{Ec}$ homologs are restricted to the proteobacteria and a few representatives from other phyla. Homologs of FNR$_{Bs}$ and NreB$_{Sa}$ are located within the Bacilli, of WhiB3 within the Actinobacteria. Archaea contain no homologs. The data reveal no similarity between the FNR$_{Ec}$, FNR$_{Bs}$, NreB$_{Sa}$ and WhiB3 sensor families on the sequence and structural levels. These O$_2$ sensor families arose independently in phyla that were already present at the time O$_2$ appeared, their members were subsequently distributed by lateral gene transfer. The chemistry of [4Fe-4S] and [2Fe-2S] cluster formation and interconversion appears to be shared by the sensor protein families. The type of signal output is, however, family-specific. The homologs of FNR$_{Ec}$ and NreB$_{Sa}$ vary with regard to the number of Cys residues that coordinate the cluster. It is suggested that the variants derive from lateral gene transfer and gained other functions.

Introduction

In modern microbes, molecular oxygen is important for many aspects of bacterial physiology, mainly catabolism (oxidation, respiration) and stress response. Cyanobacteria started producing O$_2$ about 2.5 billion years ago (Fischer et al., 2016). Prior to that time
there were only anaerobes on Earth. The appearance of an O$_2$-enriched atmosphere confronted microbes with the presence of a strong oxidant that could readily inactivate or otherwise poison the active site and reactive groups of ancient enzymes, mostly metalloenzymes, that we today view as oxygen sensitive (Martin & Sousa, 2015; Raymond & Segre, 2006). This forced microbes to find mechanisms of detecting and dealing with O$_2$, mechanisms that have persisted to the present. Access to aerobic or anaerobic niches requires extensive remodeling of energy conserving metabolic and anabolic pathways (see review by Unden and Bongaerts (1997). Additionally, during aerobic growth reactive oxygen species such as (hydrogen)peroxide, superoxide and hydroxyl radicals are formed which require a protective response by the bacteria (for reviews see Imlay (2002) and (2006)).

Because of its ecological and physiological importance, most bacteria contain sensors for detecting O$_2$ and reactive oxygen species. Sensors for O$_2$ are essentially restricted to facultative anaerobic and (micro)aerobic bacteria. About 2.5 x 10$^9$ years before present, atmospheric O$_2$ rose to estimated levels of about 0.02 to 0.04 atm (Holland, 2006), close to the Pasteur point (Engelhardt, 1974) for the onset of aerobic metabolism. The Pasteur point coincides with the switch point (approx. 0.5 to 2% of air saturation) of O$_2$-regulated genes in facultative metabolism of E. coli that respond to FNR$_{Ec}$ and other O$_2$-sensors (Becker et al., 1996; Becker et al., 1997; Tseng et al., 1996). O$_2$ levels of 0.02 to 0.04 atm resulted in mild oxygenation of ocean surface waters, a situation that persisted for almost 2 billion years (Lenton et al., 2016) because deep ocean oxygenation was not completed until roughly 580-430 million years ago (Stolper & Keller, 2018). In response to environmental oxygenation, bacteria evolved a variety O$_2$ sensing systems, classified as direct and indirect O$_2$ sensors, that recruited different biochemical mechanisms for monitoring O$_2$ levels (Green & Paget, 2004; Unden et al., 2010).

Direct O$_2$ sensors use either O$_2$-sensitive [4Fe-4S] clusters (such as the FNR proteins of E. coli and Bacillus subtilis, or NreB of Staphylococcus carnosus), heme B (such as FixL of Sinorhizobium meliloti, or Dos of E. coli), or FAD (such as NifL of Azotobacter vinelandii) as O$_2$-reactive prosthetic groups (for review see Green and Paget (2004) and Unden et al. (2010)). In Pseudomonas species, a system related to the hypoxia-inducible transcription factor (HIF) of animals has been described that employs O$_2$-dependent hydroxylation reactions to sense...
decreased O₂ availability (Schmidt et al., 2016; Scotti et al., 2014).

Indirect O₂-sensors respond to metabolites and pathway intermediates that change their cellular concentration or redox state in response to O₂ availability. The ArcAB two-component system of E. coli responds to the redox state of the respiratory quinones or changed quinone/quinol ratios (Malpica et al., 2004). The Rex transcriptional regulator of Gram-positive bacteria on the other hand measures the cellular NADH/NAD ratio, which changes as a function of O₂-reducing respiratory activity (Brekasis & Paget, 2003; Sickmier et al., 2005; Wang et al., 2008).

**O₂-labile [4Fe-4S] clusters as universal cofactors for O₂-sensing**

Iron-sulfur clusters are widespread and ancient metal cofactors of proteins that are composed of iron ions and sulfide. The clusters are coordinated via weak covalent bonds with cysteine thiol sidechains of the protein. Two-iron-sulfur [2Fe-2S], three-iron-sulfur [3Fe-4S] and four-iron-sulfur [4Fe-4S] clusters are common in proteins (Beinert et al., 1997). Proteins with [4Fe-4S] clusters are most versatile and have roles in protein-bound electron transfer such as in ferredoxin, fumarate reductase and many other redox or respiratory enzymes (Beinert, 1976; Lancaster et al., 1999; Malkin & Rabinowitz, 1967), as catalytically active sites in hydratases such as aconitase (Beinert et al., 1996), and in iron and redox responsive regulatory proteins (Beinert et al., 1996; Mettert & Kiley, 2015). All direct O₂-sensors with FeS clusters use [4Fe-4S] clusters. For this reason only [4Fe-4S] cluster binding sensors will be considered here.

In aconitase and related enzymes the [4Fe-4S] cluster is required for binding and activation of the substrate (Beinert et al., 1996). In addition to its catalytic role, the cluster has a structural role in aconitase (Beinert et al., 1996). Thus apo-aconitase is catalytically inactive but has a regulatory function (Beinert et al., 1996; Rouault et al., 1992). Loss of the iron-sulfur cluster causes structural rearrangements that allow binding of the apo-enzyme (iron regulatory protein, or IRP) to iron responsive elements (IREs), which are located in the 5’ region of the mRNA of iron homeostasis genes (Commichau & Stulke, 2008; Kaptain et al., 1991; Tang & Guest, 1999; Tang et al., 2005). IRE binding controls gene expression through mRNA stability.
By contrast, the transcriptional bacterial regulators FNR\textsubscript{Ec}, FNR\textsubscript{Bs}, NreB, WhiB3, and NsrR (Volbeda \textit{et al}., 2017) contain [4Fe-4S] clusters that react chemically with molecular O\textsubscript{2} or NO, which controls function of the sensors by cluster conversion and modification. The physiological role of FNR\textsubscript{Ec}, FNR\textsubscript{Bs}, NreB is essentially that of O\textsubscript{2}-sensors whereas that of WhiB3 can be either that of an O\textsubscript{2} or NO sensor depending on physiological conditions (Singh \textit{et al}., 2007). For each of the sensors, reaction with O\textsubscript{2} has been characterized in detail (see below). NsrR, on the other hand, is essentially an NO sensor and the [4Fe-4S] cluster is used for NO response (Volbeda \textit{et al}., 2017). For this reason in the present work only FNR\textsubscript{Ec}, FNR\textsubscript{Bs}, NreB, and WhiB3 will be discussed. The iron-sulfur clusters are surface exposed and are able to react with O\textsubscript{2}. FNR\textsubscript{Ec}, FNR\textsubscript{Bs}, and WhiB3 represent transcriptional regulators (Fig. 1). DNA binding and their function as transcriptional regulators is controlled by the [4Fe-4S] clusters which react directly with O\textsubscript{2}. In the FNR proteins, the reaction of the cluster with O\textsubscript{2} causes cluster degradation. As a consequence FNR\textsubscript{Ec} monomerizes and loses the ability for DNA binding and transcriptional activation (Lazazzera \textit{et al}., 1996) whereas FNR\textsubscript{Bs} is a permanent dimer but loses DNA binding due to conformational changes after degradation of the FeS cluster (Reents \textit{et al}., 2006b).

In WhiB3, the [4Fe-4S] cluster is required for complex formation with the sigma factor \sigma\textsubscript{A}. Degradation of the FeS cluster disassembles the complex and transcriptional activity (Kudhair \textit{et al}., 2017). In the sensor kinase NreB, the sensory PAS domain controls the activity of the kinase domain in response to O\textsubscript{2} via the iron-sulfur cluster. Auto-phosphorylation of NreB leads to the phosphorylation of the response regulator NreC that activates in the phosphorylate state (NreC-P) the expression of target genes.

The Cys residues that ligate the iron-sulfur clusters in FNR\textsubscript{Ec}, FNR\textsubscript{Bs}, WhiB3, and NreB are located in clusters that differ in sequence, spacing and location within the sensors (Fig. 1B). In FNR\textsubscript{Bs} one of the ligands of the iron-sulfur cluster is replaced by an Asp residue (Gruner \textit{et al}., 2011). FNR\textsubscript{Ec} was the first O\textsubscript{2}-sensor of this type and represents the prototype of this type of O\textsubscript{2}-sensors (Green & Paget, 2004; Shaw & Guest, 1982; Shaw \textit{et al}., 1983). The crystal structure of the FNR\textsubscript{Ec} type FNR\textsubscript{Af} from \textit{Aliivibrio fischeri} was solved (Volbeda \textit{et al}., 2015). The protein consists of two domains that provide the sensory and the DNA-binding function.
The N-terminal sensory domain ligates under anoxic conditions a $[4\text{Fe}-4\text{S}]^{2+}$ cluster and under oxic conditions a $[2\text{Fe}-2\text{S}]^{2+}$ cluster.

The $[4\text{Fe}-4\text{S}]^{2+}$ cluster binding sensors FNR$_{Ec}$ and FNR$_{Af}$ are dimers with an α-helical dimer interface, and the dimeric state is required for site specific DNA-binding (Khoroshilova et al., 1995; Kiley & Beinert, 1998; Lazazzera et al., 1993). Conversion to the $[2\text{Fe}-2\text{S}]^{2+}$ form results in a rearrangement of the dimer interface causing monomerization and loss of specific DNA-binding (Khoroshilova et al., 1995; Volbeda et al., 2015). The $[4\text{Fe}-4\text{S}]^{2+}/[2\text{Fe}-2\text{S}]^{2+}$ cluster conversion was identified by Mössbauer and EPR spectroscopy (Khoroshilova et al., 1995; Khoroshilova et al., 1997; Kiley & Beinert, 1998; Lazazzera et al., 1993; Lazazzera et al., 1996). Combination of visible absorbance and EPR spectroscopy and time resolved electrospray ionization mass spectrometry allowed a very detailed analysis of the reactions. The $\text{O}_2$ triggers a reaction in two steps (Crack et al., 2008; Crack et al., 2007; Crack et al., 2006):

\[
\begin{align*}
\text{Step 1:} & \quad [4\text{Fe-4S}]^{2+} + \text{O}_2 \rightarrow [3\text{Fe-4S}]^{1+} + \text{Fe}^{2+} + \text{O}_2^- \\
\text{Step 2:} & \quad [3\text{Fe-4S}]^{1+} \rightarrow [2\text{Fe-2S}]^{2+} + \text{Fe}^{3+} + 2\text{S}^2- 
\end{align*}
\]

Step 2 apparently involves a second oxidation by $\text{O}_2$ (Crack et al., 2017; Zhang et al., 2012).

The two sulfide ions in step 2 are not released into aqueous solution but oxidized to sulfane ($\text{S}^0$) and form a persulfide with two of the Cys ligands (RS) of the cluster. Up to two of the Cys ligands exist then in the persulfide (RSS') state. The reaction of step 2 has to be reformulated accordingly (Step2*):

\[
\text{Step 2*:} \quad [3\text{Fe-4S}](\text{RS})_3 + \text{RS}^- + \text{O}_2 + 4\text{H}^+ \rightarrow [2\text{Fe-2S}](\text{RS})_2(\text{RSS})_2 + \text{Fe}^{3+} + 2\text{H}_2\text{O}
\]

Formation of the Cys-persulfide provides a mechanism for storing the sulfur released from the iron-sulfur cluster during the $[4\text{Fe-4S}]/[2\text{Fe-2S}]$ cluster conversion rather than releasing it to the water space. This mechanism allows reversion of the $\text{O}_2$-inactivated FNR and to the anaerobic $[4\text{Fe-4S}]^{2+}$ form by reduction and repair without involvement of the iron-sulfur biosynthesis machinery (Crack et al., 2017; Zhang et al., 2012).
In NreB, a \([4\text{Fe-4S}]^{2+}/[2\text{Fe-2S}]^{2+}\) cluster conversion is the basis for \(O_2\)-sensing (Mullner \textit{et al.}, 2008) which suggests a reaction sequence similar to \(\text{FNR}_{\text{EC}}\). In the anoxic form, \(\text{FNR}_{\text{BS}}\) and WhiB3 contain a \([4\text{Fe-4S}]^{2+}\) cluster as well (Crack \textit{et al.}, 2009; Jakimowicz \textit{et al.}, 2005; Reents \textit{et al.}, 2006b). It appears therefore that \(\text{FNR}_{\text{EC}}, \text{FNR}_{\text{BS}}, \text{NreB}\) and WhiB3 use the same cofactor for \(O_2\)-sensing, and that the reactions occurring at the \([4\text{Fe-4S}]^{2+}\) during response to \(O_2\) are similar.

Here we investigate the phylogenetic relationships of bacterial \(O_2\) sensors using \([4\text{Fe-4S}]^{2+}\) clusters for sensing. The protein sequences of \(\text{FNR}_{\text{EC}}, \text{FNR}_{\text{BS}}, \text{S. aureus NreB}\) and \(\text{M. tuberculosis WhiB3}\) were used to identify homologues among gene families (clusters) generated from 1,981 sequenced prokaryotic genomes and screened for retention of the consensus Cys clusters for binding the iron-sulfur clusters. The latter represent the most characteristic feature of the proteins in order to characterize their potential for \(O_2\)-sensing and their distribution and variation among prokaryotes.

RESULTS

\textit{Clustering and phylogenetic distribution of FNR}_{\text{EC}}, FNR_{\text{BS}}, \text{NreB}_{\text{Sa}} \text{and WhiB3 homologs}

The presence absence matrix in Fig. 2 summarizes the occurrence of \(\text{FNR}_{\text{EC}}, \text{FNR}_{\text{BS}}, \text{NreB}_{\text{Sa}}, \text{WhiB3}_{\text{Mt}}\) and their homologs in the bacterial taxa indicated. No homologues were detected in archaea. The occurrence of CRP, a global regulator and homologue of \(\text{FNR}_{\text{EC}}\) that lacks \(O_2\) sensing clusters, is also indicated. Each column represents a cluster, ticks summarizing the results of BLAST searches and clustering to detect homologues (black ticks indicate presence; white ticks indicate absence) from prokaryote genomes in the RefSeq 2012 database. The clusters were generated using the standard Markov cluster algorithm (Enright, A. J. \textit{et al.} 2002) (MCL) at a 25\% global identity threshold as previously described (Nelson-Sathi \textit{et al.}, 2015; Weiss \textit{et al.}, 2016). Each black tick indicates the presence of one (or more) homologous protein in the corresponding species (rows) with a BLAST hit with \(\geq 25\%\) local identity and a pairwise alignment with \(\geq 25\%\) global identity. Protein sequence similarity was scored without weighting of specific motifs. To distinguish between homologous proteins that could function as oxygen sensor and false positive sensors, the presence of the cluster-ligating Cys motif was scored. FNR homologs that lack N- or C-terminal Cys motifs
were scored as CRP-type regulators due to the absence of the oxygen-responsive FeS cluster required for sensor function. The NreB-type regulators are identified by similarity of PAS and kinase domain.

FNR of the *E. coli* type (FNR<sub>Ec</sub>) predominantly occurs among the α-, β- and γ-Proteobacteria but some homologs are found in the Clostridia, Spirochaetes, Bacteriodetes and other phyla. By contrast, the *B. subtilis*-type FNR (FNR<sub>Bs</sub>) occurs mainly within the Bacilli and has no homologs in the Proteobacteria. Few additional BLAST hits are found within the Clostridia, Negativicutes, Bacteriodetes and Actinobacteria. The sensor kinase NreB from *S. aureus* and its homologs are mostly restricted to the Bacilli but there are few similar proteins in other Firmicutes, and in the Actinobacteria, Proteobacteria, Spirochaetes, Planctomycetes as well as in the Deinococcus-Thermus groups. The *M. tuberculosis* WhiB3 sensor is limited to the Actinobacteria. The global regulator CRP from *E. coli* that is related to FNR<sub>Ec</sub> and FNR<sub> Bs</sub> (Korner et al., 2003) has homologs in almost all bacterial phyla. Most of the CRP homologs are found within the Actinobacteria and the Proteobacteria whereas the Bacilli that harbor most of the FNR<sub> Bs</sub> proteins are mostly devoid of CRP. In general, the oxygen sensors FNR<sub>Ec</sub>, FNR<sub> Bs</sub>, NreB and WhiB3 are restricted to specific bacterial phyla whereas CRP occurs in most bacterial phyla. Several phyla lack specific O<sub>2</sub>-sensors of this type (Aquifex, Thermotoga, Fusobacteria, ε-Proteobacteria, Acidobacteria, Chalmydiae, Chlorobi, Cyanobacteria, and the Green Filamentous bacteria) or only show sporadic occurrence (Bacteroides, Planctomyces, Tenericutes, δ-Proteobacteria, Negativicutes, and Deinococcus-Thermus). Notably, the archaea, including the aerobic Haloarchaea (Euryarchaeota) and the (facultatively) aerobic hyperthermophilic Crenarchaeota (*Sulfolobus, Acidianus, Pyrobaculum, Metallosphaera*) are devoid of FNR<sub>Ec</sub>, FNR<sub> Bs</sub>, NreB or WhiB3 type sensors.

The iron-sulfur clusters of FNR<sub>Ec</sub>, FNR<sub> Bs</sub>, NreB<sub> Sc</sub> and WhiB3 are co-ordinated by Cys and occasionally Asp residues. FNR<sub>Ec</sub> and FNR<sub> Bs</sub> are distant homologs of the CRP protein (Korner et al., 2003; Shaw et al., 1983). FNR<sub>Ec</sub> contains a short N-terminal extension to CRP of about 29 AA (Fig. 1A) that are specific for FNR<sub>Ec</sub> (Shaw et al., 1983). Three of the four Cys residues (C1 to C3) of FNR<sub>Ec</sub> for ligating the iron-sulfur cluster are placed in the FNR-typic extension (Fig. 1B), only the fourth residue (C122) is located in the CRP homologous region. FNR<sub> Bs</sub> that is also
distantly homologous to CRP, carries the Cys-cluster in a C-terminal extension downstream the CRP homologous region. Only the first ligand (D141) is located in the CRP homologous region. Replacement of single residues of the binding clusters by other residues generally inactivates their capacity for O$_2$-sensing (Green et al., 1993; Gruner et al., 2011; Kamps et al., 2004; Melville & Gunsalus, 1990), demonstrating the significance of the residues for the basic function of the proteins. The proteins of the searches (Fig. 2) were therefore screened for the presence of the respective Cys/Asp clusters and the presence of sensory domain. Refined overviews of FNR$_{Ec}$, FNR$_{Bs}$, NreB$_{Sc}$ and WhiB3 homologs are presented in the following sections including an evaluation of their presumptive properties in O$_2$-sensing.

**FNR-Ec-type sensors in the prokaryotic kingdom**

FNR$_{Ec}$ is a major regulator of the aerobic/anaerobic switch in *E. coli* and responsible for the induction of genes of anaerobic and microaerobic respiration, fermentation and anaerobiosis related genes. Transcriptional activation by FNR$_{Ec}$ depends on its dimeric state which is controlled by the [4Fe-4S]/[2Fe-2S] cluster conversion in response to O$_2$ presence. The matrix of Fig. 2 contains altogether 414 FNR$_{Ec}$ homologs after deletion of redundant or closely related hits from related species. From the remaining, 95% of the homologs were located within the α-, β- and γ-Proteobacteria (Fig. 3), with the γ-proteobacteria enclosing the largest number. The residual hits (5 % of 414) were in diverse phyla outside the proteobacteria, that is in the Clostridia (6 strains), Bacteriodetes (6 strains), the Spirochaetes (3 strains) and others (Fig. 4). The homologs with the ‘diverse’ origin showed a closed clustering separate from the proteobacteria. For all homologs within the α-, β- and γ-proteobacteria the phylogenetic tree of the homologs agrees with that of their hosts (Fig. S1), suggesting that the proteins co-evolved with their phyla.

The 414 FNR-Ec type homologs were investigated for the conservation of the Cys clusters by recording the number and spacing of Cys residues. Spacing variants delineate subfamilies of FNR proteins, whereas conservation or loss of Cys residues indicates conservation of O$_2$-sensing. The majority of FNR$_{Ec}$ homologs (79%) retained a 4Cys cluster (Fig. 3), the remaining 21% were lacking one or more of the Cys residues (Fig. 3). The variants containing less than 4 Cys residues of the cluster are mostly scattered in the phyla with some specific preferences.
The 4Cys variants represent the majority of the homologs in the α-, β- and γ-proteobacteria, whereas in the other bacteria the variants with ≤3 Cys residues predominate (12 from 20 homologs). The function of these ≤3 Cys variants has not been studied, but their prevalence suggests a modified function.

The 3Cys variants are rare and are found only in the γ-proteobacteria and the "diverse lineages" group. The 2Cys and 1Cys variants are found in all protobacterial phyla at a frequency of ≤10% of phylum members each. The 0Cys variants are characteristic for the α-proteobacteria and represent 35% of the FNR Ec homologs (Fig. 3). In the following, the FNR Ec Cys variants will be discussed with respect to their possible function, but we note that only 4Cys and 0Cys variants of FNR Ec have been characterized by experiments.

4Cys variants of FNR Ec

Within the 4Cys FNR Ec proteins, three types of spacing are observed. The C1-X2-C2 spacing is conserved in all types, but for the C2/C3 and C3/C4 pairs some variation can be seen (Table 1, and Fig. S2 and S3). The prototypic spacing C1-X2-C2-X5-C3-X92-C4 (Table 1; Fig. S2) from FNR Ec is present in the FNR proteins of the γ- and β-proteobacteria and represent the largest group of 4Cys FNR Ec proteins. FNR Ec of E. coli represents the prototype of 4Cys FNR where all details in O2-sensing and cluster biochemistry have been studied (Crack et al., 2017; Green et al., 1996; Khoroshilova et al., 1995; Khoroshilova et al., 1997; Shaw & Guest, 1982). In addition to FNR Ec, the proteins of other γ-proteobacteria have been verified as FNR Ec-type O2-sensors, including Fnr or Anr from Pseudomonas spp (Galimand et al., 1991; Ibrahim et al., 2015; Sawers, 1991), EtrA from Shewanella spp (Cruz-Garcia et al., 2011; Maier & Myers, 2001), Fnr from Vibrio strains (Kado et al., 2017; Septer et al., 2010), FnrP of Pasteurella (Uhlich et al., 1999), HlyX of Actinobacillus pleuropneumoniae (MacInnes et al., 1990) and Fnr of Klebsiella pneumoniae (Grabbe et al., 2001) (Table 1). FNR proteins of the β-proteobacteria contain the same Cys spacing as the γ-proteobacteria, but the biochemistry of iron-sulfur cluster has not been studied in vitro. The FNR proteins of the β-proteobacteria control expression of catabolic processes like anaerobic respiration and fermentation (Table 1) similar to FNR from the γ-proteobacteria and are significant for the virulence of some pathogenic strains like Bordetella, Neisseria and Burkholderia (Bannan et al., 1993; Bartolini et al., 2006; Edwards et al., 2010; Sass et al., 2013; Wood et al., 1998).

This article is protected by copyright. All rights reserved.
The α-proteobacteria contain 4Cys variants with a slightly different spacing of the C2/C3 and the C3/C4 residues (C1-X2-C2-X7-C3-X97-C4) (Table 1, and Fig. S2). For some 4Cys FNR_{Ec} variants of the α-proteobacteria O_2-sensitivity and their capability for complementing FNR_{Ec} function in vivo has been verified but detailed biochemical studies on the properties of the modified Cys clusters are missing. The 4Cys FNR_{Ec} homologs of the α-proteobacteria often function as transcriptional activator for genes of anaerobic metabolism many of which are related to bacteria/plant association like nitrate and microaerobic respiration, N_2-fixation, virulence and functionally related genes (Table 1). Representatives of this class have been characterized mainly in vivo, such as Fnrl of *Rhodobacter capsulatus* (Zeilstra-Ryalls et al., 1997), FnRN of *Rhizobium leguminosarum* (Schluter et al., 1992), SinR of *Agrobacterium tumefaciens* (Ramey et al., 2004), FixK_1 of *Bradyrhizobium japonicum* (Anthamatten et al., 1992) and AadR of *Rhodopseudomonas palustris* (Dispensa et al., 1992). Remarkably, the 4Cys FNR of the plant pathogenic α-proteobacteria *Xanthomonas* (and *Pseudoxanthomonas*) contain the same Cys cluster indicating a relation of this type of FNR to a plant associated biotope.

The 4Cys FNR_{Ec} proteins of the ‘diverse’ bacteria group outside the proteobacteria (Fig. 4) differ in the 4Cys cluster slightly from that of the γ/β- and the α-proteobacteria (Table 1, Fig. S3). The difference in Cys-spacing and their separate phylogenetic clustering stresses their diversification from the proteobacterial FNR_{Ec}-proteins. None of the FNR_{Ec} type proteins of this group has been tested for iron-sulfur cluster properties or function, but the presence of a modified 4Cys cluster suggests that the proteins may be O_2-sensitive transcriptional regulators similar to FNR_{Ec}. Cluster type and the protein sequence give no indication on the phylogenetic origin of this group of FNR_{Ec} proteins.

### 3Cys variants of FNR_{Ec}

3Cys variants of FNR_{Ec} are scarce (Fig. 3). Most variants (6) are placed within the ‘diverse’ group non-proteobacteria, and only one further representative is found within the (γ-) proteobacteria (Fig. 3; Fig. S1). None of the variants has been characterized genetically or biochemically. In the 3Cys variants generally C3 is missing (C1-X2-C2-X97.98-C4; Table 1, and Fig. S4) whereas C1, C2 and C4 are conserved. In some variants C3 is replaced by a Ser, Asn
or Pro residue, or a Cys residue in a modified position. These residues could serve as the fourth ligand for the iron-sulfur cluster, with or without conservation of the function (Muraki et al., 2010). The 3Cys variants therefore might be functionally similar to the 4Cys FNR Ec proteins, or represent precursors of other FNR variants.

2Cys variants of FNR Ec

2Cys variants of FNR Ec are present in small numbers (18 representatives in total) in the proteobacterial and the diverse groups (Fig. 3). The 2Cys variants uniformly lack C1. Most of the homologs contain conserved C3 and C4, but other combinations of conserved Cys residues are present as well (Table 1). Thus in Burkholderia (β-proteobacteria) and Acholeplasma (Tenericutes) the N-terminal sequence with C1 is deleted whereas in the variants of Chromholoaobacter (γ-proteobacteria), Streptococcus (Bacilli), Clostridium cellulolyticum and others the Cys residues are not conserved. None of the 2Cys variants has been functionally or biochemically characterized. The Flp proteins from lactic acid bacteria (Gostick et al., 1998; Scott et al., 2000) serve as models for the function of FNR-like proteins with two Cys residues (Table 1). FlpA from Lactococcus lactis is a member of the FNR/CRP family of transcriptional regulators (Korner et al., 2003) with low similarity (22 % identity with FNR Ec) and is not part of the FNR Ec like protein cluster of Fig. S1. The Cys pair of FlpA (residues C15 and C112) assembles in the FlpA dimer an O2-labile [4Fe-4S] cluster that abolished DNA binding (Scott et al., 2000). Flp from Lactobacillus casei on the other hand uses the Cys pair for an intra-molecular disulphide-dithiol redox switch (Gostick et al., 1998). The reactions of the Flp proteins in vivo are not clear, however. The Flp proteins respond to oxidative stress, and control redox stress reactions, zinc uptake and the arginine deiminase pathway. It is feasible that the 2Cys FNR Ec proteins employ similar reactions or functions.

1Cys variant of FNR-Ec

The 1Cys variants of FNR Ec have high global sequence similarity to FNR Ec and are located within the proteobacteria and Opitutus terrae from the diverse group (Fig. 3, Fig. S1). O. terrae and some proteobacteria including Paracoccus denitrificans contain only the 1Cys variant of FNR Ec, whereas other strains carry additionally 4Cys FNR Ec. The conserved Cys residue is mostly C4, whereas in some representatives C1 (Methyllobacterium) or C3 (O. Accepted Article Accepted Article

This article is protected by copyright. All rights reserved.
*terrae* and *V. paradoxus*) is conserved. The bacteria with conserved C4 generally lack the N-terminal part of FNR<sub>Ec</sub> with aa 1 to 25 and C1 to C3 of FNR<sub>Ec</sub>. Presence of a single redox sensitive Cys residue is reminiscent of the *B. subtilis* OhrR protein. The thiolate of Cys15 from OhrR is oxidized (reversibly) to Cys15-SOH (sulphenic acid) by treatment with hydroperoxide. The oxidation inhibits DNA-binding of OhrR and induces expression of the organic hydroperoxidase gene (Fuangthong & Helmann, 2002). Generally, conservation of single Cys residues suggests redox regulation by reversible protein S-thiolation or thiol-based redox switches (Hillion & Antelmann, 2015; Loi *et al.*, 2015).

**0Cys variants of FNR-Ec**

Most of the 0Cys variants of FNR<sub>Ec</sub> are located in the α-proteobacteria (Fig. 3) with few examples within the δ- (*Desulfomicrobium*) and β-proteobacteria (*Burkholderia*). Some of the 0Cys variants show deletions in the N-terminal C1 to C3 region of FNR<sub>Ec</sub> whereas other cover the region without conservation of the Cys residues. The residual part of the protein is conserved which defines the proteins as FNR<sub>Ec</sub> homologs.

Prototypes of the 0Cys variants are represented by FixK<sub>2</sub> of *Rhodopseudomonas (Rp.)* *japonicum* and FixK of *Rhizobium meliloti* (Batut *et al.*, 1989; Fischer, 1994). FixK<sub>2</sub> and FixK are part of an O<sub>2</sub>-regulatory cascade which induces nitrogen fixation and nitrate respiratory genes under anoxic conditions. FixK<sub>2</sub> and FixK are no O<sub>2</sub>-sensors on their own. Expression of the *fixK<sub>2</sub>* and *fix* genes is, however, under the transcriptional regulation of the FixL-FixJ O<sub>2</sub>-sensing system that stimulates expression (and function) of FixK<sub>2</sub> and FixK under anoxic conditions (for overview see Fischer (1994)). In summary, the 0Cys proteins are members of the FNR<sub>Ec</sub> family, but represent in the form of FixK and FixK<sub>2</sub> indirect O<sub>2</sub>-sensors without iron-sulfur cluster.

**FNR *B. subtilis* (FNR<sub>Bs</sub>)**

The search for homologs of FNR<sub>Bs</sub> yielded 96 hits which were mostly (82 %) in the Bacilli phylum (Fig. 1) and among those most (63 %) in the genus *Bacillus. B. subtilis* FNR<sub>Bs</sub>.
coordinates the [4Fe-4S]-cluster by Asp D141 and three Cys residues. The Cys residues are located in a C-terminal domain (Gruner et al., 2011; Reents et al., 2006b) that represents an extension to the CRP homologous region (see Fig. 1). The basic function of this type of FNR in O$_2$-sensing has been studied for FNR$_{Bs}$ (Gruner et al., 2011; Reents et al., 2006b; Reents et al., 2006a). FNR$_{Bs}$ and FNR$_{Bl}$ of B. licheniformis (Klinger et al., 1998; Rey et al., 2004) are required for anaerobic induction of the nar genes coding for the nitrate respiratory system narGHJI and narK as well as for arfM, a fermentation regulator.

18% of the FNR$_{Bs}$-type proteins lack the C-terminal domain and the Cys residues for cluster binding, other variants of the Cys cluster were not identified. Most of the 0Cys variants lack also the conserved Asp residue. The function of the 0-Cys/Asp variants of FNR$_{Bs}$ has not been analyzed, but most of the strains are facultatively anaerobic and capable of fermentation or nitrate respiration (Table S1). The 0Cys variants of FNR$_{Bs}$ could represent indirect, iron-sulfur deficient O$_2$-sensors, such as the 0Cys variants of FNR$_{Ec}$ (compare Table 1), or other regulators of the FNR-CRP family (Korner et al., 2003). The FNR$_{Bs}$ variant from aerobic N. koreensis has the conserved Asp residue, but lacks the C-terminal Cys cluster of FNR$_{Bs}$. Instead, it contains, an N-terminal Cys-cluster similar to that of FNR$_{Ec}$. The protein appears to be an FNR$_{Bs}$ homolog with an FNR$_{Bs}$/FNR$_{Ec}$ hybrid iron-sulfur binding 3Cys/Asp site.

Fig. 5B shows a matrix for the FNR$_{Bs}$ proteins (80 variants) excluding 0Cys-FNR$_{Bs}$ variants. The occurrence of the FNR$_{Bs}$ homologs is presented in a simplified tree where strains are combined on the genus level (Fig. 5A). FNR$_{Bs}$ is confined to the Bacilli phylum, with few exceptions of homologs in Clostridium botulinum (Clostridia) and Selenomonas (Sm) ruminantium (Negativicutes). FNR$_{Bs}$ is therefore restricted to the Firmicutes with strong predominance in the genus Bacillus. Homologs are present in most of the Bacilli genera, including Paenibacillus, Geobacillus, Exiguobacterium, Brevibacillus, Lysinibacillus, Macrooccus, Staphylococcus and Anoxybacillus.

The bacteria of the matrix of Fig. 5, and the corresponding FNR$_{Bs}$ proteins can be functionally sub-grouped by the controlled genes. (i) The typical form represented by FNR$_{Bs}$ of B. subtilis and B. licheniformis controls anaerobic induction of the nitrate respiratory system (Cruz

This article is protected by copyright. All rights reserved.
The branch comprising 17 non-pathogenic strains with B. subtilis, B. licheniformis, B. artrophaeus and the plant growth promoting rhizobacterium B. amyloliquefaciens (Chen et al., 2007; He et al., 2012) and B. sp (Song et al., 2012) appears to be part of this class. FNR from the Geobacilli and B. megaterium could be part of the same cluster due to the presence of nitrate respiration and colocalization with nar genes (Brumm et al., 2015; Feng et al., 2007; Muhd Sakaff et al., 2012). Interestingly Paenibacillus terrae, polymyxa and mucilaginosus contain two FNR-like proteins each. (ii) FNR$_{Bc}$ of the pathogenic B. cereus modulates under anaerobic conditions glucose fermentation and other catabolic genes in a carbohydrate-dependent manner, but is dispensable for nitrate respiration. FNR$_{Bc}$ activates also the expression of enterotoxins (Esbelin et al., 2012; Messaoudi et al., 2010; Zigha et al., 2007). (iii) The FNR homologs of B. selenitireducens are more distantly related and not part of the FNR$_{Bs}$ and FNR$_{Bc}$ branches. The bacteria are non-pathogenic and not nitrate respiring (Eppinger et al., 2011; Switzer Blum et al., 1998), suggesting that their FNR proteins fulfill a different function in regulation. (iv) The role of FNR$_{Bs}$-type proteins in bacteria outside the genus Bacillus is mostly unknown, e.g. in Exiguobacterium sibiricum and Anoxybacillus flavithermus (Rodrigues et al., 2008; Saw et al., 2008). FNR of A. flavithermus FNR might control arginine metabolism, that of Macrococcus caseolyticus and Staphylococcus pseudintermedius nitrate respiration (Baba et al., 2009). The latter strains contain also a homolog of the NreABC system which regulates in Staphylococcus nitrate respiration in response to O$_2$ availability (Fedtke et al., 2002; Schlag et al., 2008). The role of FNR$_{Bs}$ in the bacteria is therefore not clear. The role of FNR$_{Bs}$ in the pathogenic C. botulinum and the non-pathogenic Sm. ruminantium is also not known. Both bacteria grow anaerobically, and both proteins are most similar to the FNR$_{Bs}$ like proteins of S. pseudintermedius and Exiguobacterium with unknown function.

**NreB**

For S. aureus NreB, 73 homologous proteins were identified in the BLAST search. Most (84%) of the homologs occur among the Bacilli, and of those 54% among the Staphylococci, documenting predominance of NreB in Bacilli and Staphylococci. The sensor kinase NreB coordinates the sensory iron-sulfur cluster by four conserved Cys residues of the N-terminal PAS domain (Fig. 1, and (Mullner et al., 2008)). Proteins of the matrix (12% of the homologous)
that showed similarity only in the kinase domain of HisKA_3 sensor kinases (Huynh et al., 2010) but not in the sensory PAS domain, were deleted in the revised matrix (Fig. 6B). The revised matrix includes 65 NreB-like proteins with similarity including the PAS domain. NreB is part of the NreABC two component system (nreA nreB nreC gene cluster) (Schlag et al., 2008). The nreB genes of the revised matrix were accompanied mostly by nreA and nreC genes. Among the 65 strains with NreB-like proteins, 47 encoded the complete Cys cluster for binding the iron-sulfur cluster, 12 with only one of the Cys residues, and few others with two (4) or three (2) conserved Cys residues. Most of the Cys variants are found outside the Staphylococci or Bacilli genera and lack also the nreA and nreC genes. Fig. 6 shows the schematic tree of NreB proteins after combining branches on genus level. Branches were only combined for proteins with the same amount of conserved Cys residues. The Cys residues are essential for the binding of the [4Fe-4S] cluster and function of NreB as an O$_2$-sensor (Kamps et al., 2004; Mullner et al., 2008). Therefore conservation of the Cys cluster was inspected as described above for FNREc. Notably, all homologs having the complete Cys motif are grouped on one main branch of the tree whereas the variants (highlighted in green, blue and red) are dispersed, with Brevibacillus brevis as the only exception.

4Cys variants of NreB

Most homologs of NreB (40 from 47 proteins with 4 Cys) are found in the genus Staphylococcus. Studies on the regulatory properties and on the biochemistry of the NreABC system were performed in the non-pathogenic S. carnosus and in S. aureus (Fedtke et al., 2002; Mullner et al., 2008; Schlag et al., 2008; Unden et al., 2013).

Some bacteria encode NreB and FNR of the Bs type in parallel which are responsible for anaerobic regulation of nitrate respiration in Staphylococcus and Bacillus, respectively (Cruz Ramos et al., 1995; Fedtke et al., 2002; LaCelle et al., 1996; Schlag et al., 2008). The differential roles for both sensors in bacteria like Macrococcus caseolyticus is not known (Baba et al., 2009). Bacillus clausii encodes NreABC but lacks FNR$_{Bs}$. It was suggested that NreABC regulates expression of nitrate respiration in this Bacillus strain (Mullner et al., 2008). NreB homologs are found also in some Paenibacilli including Pb. sp, Pb. terrae and Pb. polymyxa. Their nreB gene is located near the nar genes. Paenibacillus sp that encodes FNR of the FNR-Bs type and NreABC is capable of nitrate respiration (Mead et al., 2012). Gene clustering
suggests that NreABC rather than FNR_Bs regulates expression of nitrate respiration. *Brevibacillus brevis* representing a member of the Paenibacillaceae (Chen *et al.*, 2012) that carries also FNR_Bs and NreB, on the other hand lacks genes for the (nitrate sensory component) NreA and for dissimilatory nitrate reduction, suggesting a different role of NreB in the bacteria.

3Cys and 2Cys variants of NreB

Of the 65 NreB-like proteins six showed variation of the Cys cluster with 2 or 3 conserved residues (Fig. 6, and Fig. S6). None of the Cys variants has been studied genetically or biochemically. The 3Cys homologs of NreB contain an alternative Cys residue close to the N-terminus and may represent 4Cys variants (Fig. S6). The role of the 3Cys NreB of the aerobic *Thermus* species lacking anaerobic nitrate respiration (Henne *et al.*, 2004) is not known.

The 2Cys variants are represented by NreB homologs in *Lactobacillus reuteri*, *L. plantarum* and *L. fermentum* (Fig. 6, Fig. S6). In the 2Cys variants only the C-terminal Cys pair is conserved, but the proteins share high sequence similarity with NreB PAS and kinase domains. The Lactobacilli possess genes for dissimilatory nitrate reduction (Kleerebezem *et al.*, 2003; Morita *et al.*, 2008). *L. fermentum* and *L. plantarum* encode also the nitrate binding NreA-like protein (Mullner *et al.*; Unden *et al.*, 2013). It can be speculated that a dimer of the 2Cys variant of NreB is required for coordinating one iron-sulfur cluster as hypothesized for nitrogenase NfID subunit in *Methanocaldococcus jannaschii* (Staples *et al.*, 2007) or FlpA (Scott *et al.*, 2000). Alternatively, the 2Cys variants function by using a regulatory inter- or intra-molecular thiol/disulfide switch (Gostick *et al.*, 1998; Scott *et al.*, 2000).

1Cys variants of NreB

Twelve NreB variants with only one conserved cysteine residue are present in Bacilli (*Paenibacillus* and *Halobacillus*) and *Clostridium sp, Planctomyces limnophilus* and *Opitutus terrae*. Within the Paenibacilli strains with 4Cys or 1Cys NreB or FNR_Bs-like proteins are found, indicating different functions for the variants. The 1Cys-homologs are supposed to
fulfill a role in redox-sensing like the 1Cys FNR<sub>Ec</sub> proteins. Either C1, or C2 or C4 of the Cys cluster can be conserved. In most systems NreB is not associated with NreA and NreC, which argues against involvement in transcriptional and anaerobic nitrate regulation.

**WhiB3 M. tuberculosis**

BLAST search identified 124 proteins that are homologs of WhiB3 of *M. tuberculosis*. The WhiB3 homologs were exclusively present in the Actinobacteria (Fig. 1). Approximately one third of the homologs were detected in the Mycobacteria and about ten percent belong each to Streptomycyes and Corynebacteria. The remaining homologs are widespread in the Actinobacteria including *Rhodococcus, Frankia, Nocardia* as well as *Amycolatopsis* with broad branching of the phylogenetic tree (not shown).

In general, Actinobacteria contain multiple WhiB-like proteins which share similar structures including the Cys and the DNA-binding motif but have only moderate sequence similarity. Most Mycobacteria possess seven WhiB-like proteins (WhiB1 to B7) which control expression of different gene clusters (Geiman *et al.*, 2006; Saini *et al.*, 2012; Soliveri *et al.*, 2000). WhiB3 to WhiB7 respond to redox stress (O<sub>2</sub> and NO), WhiB1 only to NO, and the function of WhiB2 is redox-independent. WhiB5 is lacking in non-pathogenic Mycobacteria (Saini *et al.*, 2012). The response of WhiB3 to O<sub>2</sub> is very similar to that of FNR<sub>Ec</sub> converting the [4Fe-4S]<sup>2+</sup> cluster first to [3Fe-4S]<sup>1+</sup> with a concomitant release of Fe<sup>2+</sup> and one electron which are then used for the two-electron reduction of O<sub>2</sub> to H<sub>2</sub>O<sub>2</sub> (Singh *et al.*, 2007). In a following step the [3Fe-4S]<sup>1+</sup> is converted in a nonredox reaction to [2Fe-2S]<sup>2+</sup>. Overall, the reactions are very similar to those of Step 1 and 2 for FNR<sub>Ec</sub> (see above) with an overall reaction [4Fe-4S]<sup>2+</sup> + O<sub>2</sub> + 2H<sup>+</sup> → [2Fe-2S]<sup>2+</sup> + 2 Fe<sup>3+</sup> + 2 S<sup>2-</sup> + H<sub>2</sub>O<sub>2</sub>. According to Singh *et al.* (2007), the H<sub>2</sub>O<sub>2</sub> may destroy the [2Fe-2S] cluster.

The structure of WhiB1, which is NO-responsive, shows that WhiB1 is a four-helix bundle (Kudhair *et al.*, 2017). The core of the protein is formed by three α-helices that are held together by the [4Fe-4S] cluster, which is required for formation of a complex with the major sigma factor σ<sup>A</sup>. Reaction of the cluster with NO disassembles the complex and DNA-binding...
as well as gene expression at target genes.

The redox sensor WhiB3 of *M. tuberculosis* controls expression of genes that are important for maintenance of intracellular redox homeostasis as well as virulence, pathogenesis and persistence (Singh *et al.*, 2009; Singh *et al.*, 2007). 5% of the WhiB3 homologs lack the first Cys residue of the cluster-binding motif due to a N-terminal deletion of 20 to 30 amino acids whereas the residual part of the protein is homologous to WhiB3<sub>Mt</sub> e.g. in the proteins of *M. intracellulare*, *M. sp*, *Streptomyces hygroscopicus*, *Amycolicoccus subflavus* and *Thermonospora curvata*. WhiB4 of *M. intracellulare* having only three cysteine residues was proposed to take over a role in the adaptation to peroxide stress comparable to that of OxyR (Lewis & Falkinham, 2015). WhiB3 homologs with complete cluster-binding motif are found in pathogenic (*M. bovis* and *M. leprae*) and non-pathogenic (*M. smegmatis*) mycobacteria. Function and regulation of WhiB3 appears to be similar in all *Mycobacterium* species.


Overall, WhiB-like proteins are restricted to Actinobacteria which are all aerobic, indicating the evolution of WhiB proteins with the [4Fe-4S] cluster after accumulation of oxygen on earth.

**DISCUSSION**

*Origin and relation of the sensor proteins FNR<sub>Ec</sub>, FNR<sub>Bs</sub>, NreB and WhiB3*
Sequence comparison and distribution indicate ancient origin and independent evolution of FNR\textsubscript{Ec}, FNR\textsubscript{Bs}, NreB and WhiB3/D. FNR\textsubscript{Ec} and FNR\textsubscript{Bs} proteins are derived from the ancient carbon-regulator CRP (Saier \textit{et al.}, 1996). The global identity between FNR and CRP is low (17.9\% between FNR\textsubscript{Ec} and CRP\textsubscript{Ec}), and the proteins fall for this reason into separate clusters and alignments. The role of CRP as the precursor and the FNR proteins as the derivatives is suggested by the broad distribution of CRP in many phyla, and its full-length conservation in FNR\textsubscript{Ec} and FNR\textsubscript{Bs}. Korner \textit{et al.} (2003) and Green \textit{et al.} (2001) describe the CRP-FNR family of transcriptional regulators that includes in addition to CRP and FNR also other redox regulators like Yeil and Flp. The latter have low global identity with FNR and CRP and are not considered in the present study.

FNR\textsubscript{Ec} and FNR\textsubscript{Bs} are distinguished from CRP by the presence of short N- and C-terminal sequences, respectively, containing each three Cys ligands for ligation of the iron-sulfur clusters. Together with an internal fourth Cys (C4) or Asp residue, the Cys residues of the terminal extensions ligate the [4Fe-4S] cluster in FNR\textsubscript{Ec} or FNR\textsubscript{Bs}. In addition, FNR\textsubscript{Ec} and FNR\textsubscript{Bs} differ by the mode of signal output. In FNR\textsubscript{Ec} the signal output is produced by monomerization of the protein after reaction with O\textsubscript{2}, the oligomerization state then affects DNA binding (Lazazzera \textit{et al.}, 1993). Within the CRP/FNR family of transcriptional regulators inactivation by monomerization is unique for FNR\textsubscript{Ec} whereas other members like CRP (Anderson \textit{et al.}, 1971; Takahashi \textit{et al.}, 1980) and FNR\textsubscript{Bs} retain their dimeric state in the presence of their effector or in the active or inactive state, respectively. Thus FNR\textsubscript{Bs} is a permanent dimer, and DNA binding is regulated by conformational changes within FNR\textsubscript{Bs} (Reents \textit{et al.}, 2006b). The broad overlap of co-occurrence of CRP with FNR\textsubscript{Ec} in the \(\gamma\)-proteobacteria suggests the origin for FNR\textsubscript{Ec} within this group.

NreB belongs to the large family of histidine sensor kinases that show broad variation in their domain composition. In NreB and related His kinases sensing occurs by an N-terminal PAS domain. The PAS domain of the closely related FixL sensor kinase binds hemeB at a position homologous to the [4Fe-4S] binding site of NreB (Mullner \textit{et al.}, 2008; Unden \textit{et al.}, 2013). Remarkably, both NreB and FixL represent direct O\textsubscript{2} sensors. NreB and FixL are present, however, essentially within the Bacilli and the proteobacteria, respectively.
Therefore the hemeB/[4Fe-4S] binding sites appear to have a common origin, and an occasional but rare lateral transfer to other bacterial phyla could be detected.

The WhiB3/D proteins are restricted to Actinobacteria. The Actinobacteria contain no FNR_{Ec}, FNR_{Bs} or NreB-type O_2-regulators, but a significant variation in the WhiB3/D regulators which take over multiple functions. Their unique presence in aerobic Actinobacteria suggests development of the WhiB3/D proteins after separation or development of the group and after establishment of oxic conditions.

Recent reports for the structures of FNR_{Ec}- and Wbl-type proteins show binding and interaction of the [4Fe-4S] clusters with the respective proteins (Kudhair et al., 2017; Volbeda et al., 2015). In FNR_{Ec}, binding of the [4Fe-4S] cluster is achieved by a pocket formed from two α-helices, one β-sheet and the N-terminal loop of FNR whereas in WhiB1 proteins the [4Fe-4S] cluster is coordinated by three α-helices that are part of a four-helix bundle. WhiB1 is basically of the NO sensing type, but the overall structure is conserved in the Wbl/Whi family. NreB accommodates the cluster in a PAS domain that binds the [4Fe-4S] cluster by the α-helical PAS core and the α-helical connector which links this region to the β-scaffold and the kinase domain (Miyatake et al., 2000; Mullner et al., 2008). Therefore the structural arrangements around the iron-sulfur clusters and their binding share no obvious similarity apart from the Cys ligands which is again strong indication for their independent origin and evolution.

**Functional diversity and diversification**

Despite general restriction of FNR_{Ec}, FNR_{Bs}, NreB and WhiB3/D to diverse bacterial phyla, some bacteria contain multiple sensor proteins of these types in various combinations (Table 2). (i) Some bacteria contain multiple sensors of the same type. Thus *Pseudomonas putida* comprises three FNR-Ec type proteins that differ in their O_2-sensitivity (Ibrahim et al., 2015). Actinobacteria contain up to seven WhiB3/D proteins that differ in their response to O_2 but also in the target promoters for the individual WhiB3/D proteins (Choi et al., 2009; Crack et al., 2009; Jakimowicz et al., 2005; Kim et al., 2005; Lee et al., 2017; Lee et al., 2012; Molle et al., 2000; Steyn et al., 2002). There are also strains such as *Pb. mucilaginosus* with more than one 4Cys-FNR_{Bs} variant. (ii) Other bacteria like *Bradyrhizobium japonicum* and *Rhizobium*...
*meliloti* contain multiple variants of the same type of regulator, such as AadR (4Cys FNR$_{Ec}$, a direct O$_2$-sensor) and FixK$_2$ or FixK (0Cys FNR$_{Ec}$, an indirect O$_2$-regulator (Fischer, 1994) (Table 1, and S4). (iii) Other strains contain different types of predicted O$_2$-sensors like 4Cys-FNR$_{Bs}$ and 4Cys-NreB that are predicted to form functional O$_2$-sensors. The function and dual roles in O$_2$-sensing have not been verified, however. This list is extended by bacteria with 4Cys-FNR$_{Bs}$ (or 4Cys-FNR$_{Ec}$) proteins that are combined with 1Cys- or 0Cys-variants of an alternative sensor (FNR$_{Bs}$, FNR$_{Ec}$, NreB). (iv) Other strains contain FNR hybrid sensors with FNR$_{Ec}$/FNR$_{Bs}$ hybrid Cys clusters. Thus the FNR$_{Bs}$ homologous protein of *C. perfringens* lacks the D/Cys cluster of FNR$_{Bs}$ type sensors but contains an N-terminal Cys-cluster of the FNR$_{Ec}$ type including the central C4 residue. The spacing of the 4 Cys-residues is slightly different from that of FNR$_{Ec}$.

**Origin of the [4Fe-4S] clusters and [4Fe-4S]/[2Fe-2S] cluster conversion in O$_2$-sensing: The role of the cluster biochemistry**

The FNR$_{Ec}$, FNR$_{Bs}$, WhiB3/D and NreB proteins use [4Fe-4S]$^{2+}$ clusters for O$_2$-sensing (Crack et al., 2009; Crack et al., 2017; Jakimowicz et al., 2005; Khoroshilova et al., 1997; Mullner et al., 2008; Reents et al., 2006b; Singh et al., 2007; Zhang et al., 2012). For FNR$_{Ec}$, NreB and to some extent also for WhiB3/D, the response to O$_2$ appears to be similar by the conversion to a [2Fe-2S]$^{2+}$ cluster, despite the use of unrelated protein types and Cys clusters for ligation that differ in spacing of the Cys residues and sequence (Fig. 1). It is not known whether the similarity extends to details such as the formation of an [4Fe-3S] intermediate and of the Cys persulfides as described for FNR$_{Ec}$ (Crack et al., 2017).

Cluster conversion is the basis for signal transmission to the output domains. The response triggers conversion of dimeric to monomeric FNR$_{Ec}$ and loss of specific DNA binding (Lazazzera et al., 1996; Volbeda et al., 2015), activation of the kinase domain in NreB (Mullner et al., 2008; Nilsens et al., 2014), and loss of DNA binding by the permanent dimer FNR$_{Bs}$ (Reents et al., 2006b). It appears therefore that the response of the iron-sulfur cluster depends to a large extent on the chemistry and properties of the cluster rather than on the protein, whereas the signal output is governed by the surrounding protein.
Iron-sulfur clusters can be produced in protic and apoproteinogenic systems, and $[4\text{Fe}-4\text{S}]^{2+}$ and $[2\text{Fe}-2\text{S}]^{2+}$ clusters have been produced under anoxic conditions in the thiol-ligated form with various thiol ligands (for review see (Venkateswara Rao & Holm, 2004)). The cubane-type $[4\text{Fe}-4\text{S}]^{2+}$-thiolate clusters are very common. Formation of the $[4\text{Fe}-4\text{S}]^{2+}$ cluster represents a thermodynamic sink and the $[4\text{Fe}-4\text{S}]^{2+}$ cluster is among the most stable under anoxic conditions (Ogino et al., 1998; Venkateswara Rao & Holm, 2004). The clusters can be incorporated in cysteinyl peptides (Ohno et al., 1991; Ueyama et al., 1985). The $[2\text{Fe}-2\text{S}]$ cluster, in particular the more stable $[2\text{Fe}-2\text{S}]^{2+}$ form, can also be generated in many thiolates and cysteinyl peptides (references in (Hagen et al., 1983; Venkateswara Rao & Holm, 2004). The $[2\text{Fe}-2\text{S}]^{1+}$ cluster is less stable, but can be produced by reduction of the $[2\text{Fe}-2\text{S}]^{2+}$ cluster. Biotin and lipoate synthases, anaerobic ribonucleotide reductase and the pyruvate formate-lyase activating enzyme contain $[4\text{Fe}-4\text{S}]$ clusters that serve as electron donors in the radical catalysis of the enzymes (Buis & Broderick, 2005; Duin et al., 1997; Lotierzo et al., 2005; Ollagnier-De Choudens et al., 2000; Ugulava et al., 2000). The enzymes are sensitive to inactivation by $O_2$ that causes a side reaction with $[4\text{Fe}-4\text{S}]^{2+}$ to $[2\text{Fe}-2\text{S}]^{2+}$ conversion (compare (Venkateswara Rao & Holm, 2004)), resembling the reaction at the iron-sulfur cluster of FNR$_{Ec}$. Thus, the information from cluster (bio)chemistry is compatible with the observation that $[4\text{Fe}-4\text{S}]^{2+}$ and $[2\text{Fe}-2\text{S}]^{2+}$ clusters are present and interconverted in different protein environments. The conformational changes of cluster conversion can then be used for largely different responses, depending on the surrounding protein. The $[4\text{Fe}-4\text{S}]^{2+}$ clusters therefore represent a unique molecular device that can be modulated in evolution to bring forth $O_2$-sensors in different protein background and different output reactions. Remarkably, this mode of $O_2$ sensing that has been used for developing independent lines of $O_2$-sensors in bacteria, is apparently missing in archaea and eukaryotes.

**Experimental procedures**

**Sequences**

Protein sequences involved in oxygen sensing and containing a $[4\text{Fe}-4\text{S}]$ cluster were downloaded in July 2016 from UniProt (Consortium, 2017) and GenBank (Benson et al., 2005): Fumarate and nitrate reduction regulatory protein (FNR) from *Escherichia coli* (UniProt: P0A9E5) and *Bacillus subtilis* (GenBank: KIX83509), oxygen sensor histidine kinase
NreB from *S. aureus* (GenBank: EFW34334), redox-responsive transcriptional regulator WhiB3 from *Mycobacterium tuberculosis* (GenBank: KPU49338), and cAMP-activated global transcriptional regulator (CRP) from *E. coli* (UniProt: P0ACJ8). Furthermore, all protein sequences of 1,981 complete prokaryotic genomes were downloaded from the NCBI RefSeq database (version June 2012) and clustered into protein families as previously described (Nelson-Sathi *et al.*, 2015; Weiss *et al.*, 2016).

**Identification of homologous protein families**

Homologous sequences were identified by sequence comparisons of the five reference sequences involved in oxygen sensing, without any weighting of specific residues or motifs, with all 6.1 million proteins of the 1,981 complete prokaryotic genomes using BLASTp (Altschul *et al.*, 1997) with an E-value threshold ≤ $10^{-10}$ and local identity cutoff ≥ 25%.

Remaining BLAST hits with global amino acid identity not smaller than 25%, calculated with needle from EMBOSS 6.6.0 (Rice *et al.*, 2000), were subsequently compared to the predefined protein families. The protein family showing the highest number of BLAST hits for the respective reference sequence was identified as the homologous protein family.

**Identification of the number of cysteine residues**

Conserved cysteine residues linked to the [4Fe-4S] cluster were counted manually in sequence comparisons performed with Clustal Omega v1.2.4 (Sievers *et al.*, 2011) for each protein. Because of their low sequence conservation, *B. subtilis* FNR- and NreB-like proteins without conserved cysteine residues were excluded alignments used to generate trees.

**Multiple sequence alignment and phylogenetic tree reconstruction**

In the identified homologous protein family for FNR (*Escherichia coli*), for all sequences showing global identity ≥ 80% on the species level, only the longest sequence was retained in the protein family. Sequences in each identified homologous protein family were aligned together with the respective reference sequence using MAFFT v7.299b (Katoh & Standley, 2013) with the options --localpair, --maxiterate=1000 and --anysymbol. Maximum likelihood trees were reconstructed using RAxML v8.2.8 (Stamatakis, 2014) using the PROTCATWAG model. The clades in the trees for NreB and FNR (*B. subtilis*) were collapsed on the genus...
level using Figtree v1.3.1, when the corresponding [4Fe-4S] binding motif contained the same number of cysteine residues. No schematic tree was generated for WhiB3, where all sequences of the protein family belong to the taxonomic group Actinobacteria.

**Acknowledgements.** Financial support by Deutsche Forschungsgemeinschaft to GU (grant UN 49/18-1) and by the European Research Council (grant 666053 to WM) as well as the Volkswagen Foundation (Grant 93046 to WM) is gratefully acknowledged.

**Reference**


This article is protected by copyright. All rights reserved.


Hagen, K.S., Watson, A.D., and Holm, R.H. (1983) Synthetic routes to iron sulfide (Fe2S2, Fe3S4, Fe4S4, and Fe6S9), clusters from the common precursor tetrakis(ethanethiolate)ferrate(2-) ion ([Fe(SC2H5)4]2-): structures and properties of [Fe3S4(SR)4]3- and bis(ethanethiolate)nonathioxohexaflate(4-) ion ([Fe6S9(SC2H5)2]4-), examples of the newest types of Fe-S-SR clusters. Journal of the American Chemical Society 105: 3905-3913.


Figure legends

Fig. 1: Domain structure (A) and conserved cluster-binding motif (B) of the \([4\text{Fe}-4\text{S}]^{2+}\)-containing sensors FNR, NreB and WhiB3. A) The cysteine-carrying sensory domains (grey), the output domains as well as the kind of response (DNA binding or autophosphorylation) are depicted. Approximate positions of cluster coordinating conserved cysteine and aspartate residues as well as phosphorylation residue His159 of sensor kinase NreB of S. carnosus are marked. B) The \([\text{Fe-S}]\)-cluster binding sequence motifs consisting of conserved cysteine (C) and aspartate (D) residues are shown as well as their position in the protein. Variable amino acids residues between cluster ligands are indicated with x and the corresponding number. Modified after Unden et al. (2013).

Fig. 2: Matrix of \([\text{Fe-S}]\)-containing sensors FNR, NreB, WhiB3 and global regulator CRP. The matrix shows the distribution of the identified homologous proteins for each reference protein during BLAST search. One black line corresponds to one homologous protein in the respective bacteria. On the vertical axis the predefined bacterial phyla (sometimes classes) are displayed.

Fig. 3: Distribution of FNR_{Ec} homologs and Cys variants. The homologs (664 proteins) were identified by the BLAST search followed by removal of close homology (> 80% global identity) in order to delete hits from closely related strains. The resulting homologs (414 proteins) represent the basis for the data in A and B. A) shows the distribution of the FNR_{Ec} homologs in the bacterial phyla. B) presents the number of Cys variants within the bacterial phyla from the total of 414 FNR_{Ec} homologs. The α-, β- and γ-proteobacteria of A) show the same coloring as in the phylogenetic tree of Fig. S1. The Cys variants in B) are labeled by color code 0Cys (grey), 1Cys (red), 2Cys (blue), 3Cys (green) and 4Cys (white).

Fig. 4: FNR_{Ec} variants outside the proteobacteria phylogenetic group. Labelled in black, green, blue, red and grey are variants with four, three, two, one or zero conserved Cys residues, respectively. The digits in brackets correspond to the amount of FNR homologs in the respective genus or phylum.

Fig. 5: Schematic tree of B. subtilis FNR_{Bs} (A) and matrix for strains carrying FNR_{Bs}-like proteins (B). In (B) strains are listed after removing the 0Cys/D variants (see main text), other details as for Fig. 2. In (A) the clades were collapsed on genus level. The digits in brackets correspond to the amount of FNR homologs in the respective genus.

Fig. 6: Schematic tree of S. aureus NreB (A) and matrix of strains carrying NreB_{Sa} like proteins (B). In the matrix the 0Cys variants were deleted (compare main text). Clades were collapsed on genus level. The digits in brackets correspond to the amount of FNR homologs in the respective genus.
Homologs with three conserved residues of the [Fe-S]-binding motif are highlighted in green, whereas two cysteine residues are marked in blue and only one cysteine residue is marked in red.
Table 1: Occurrence and properties of FNR<sub>Ec</sub> variants in relation to their Cys clusters.

<table>
<thead>
<tr>
<th>FNR&lt;sub&gt;Ec&lt;/sub&gt; type</th>
<th>Cys clusters</th>
<th>Bacteria (examples)</th>
<th>Function or regulated genes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>FNR (4Cys)</td>
<td>C1 - X₂ - C2 - X₅ - C3 - X₉₂ - C₄ (Alignment, Fig. S2)</td>
<td>y-proteobacteria</td>
<td>Transcriptional regulation in response to O₂ (anaerobic and microaerobic respiration; fermentation; related genes)</td>
<td>(Galimand et al., 1991; Salmon et al., 2003; Sawers, 1991; Shaw &amp; Guest, 1982; Spiro &amp; Guest, 1988; Unden &amp; Bongaerts, 1997)</td>
</tr>
<tr>
<td>FNR (4Cys)</td>
<td>C1 - X₂ - C2 - X₅ - C3 - X₉₂ - C₄ (Alignment, Fig. S2)</td>
<td>β-proteobacteria (Bordetella pertussis; Neisseria meningitides)</td>
<td>Anaerobic growth; virulence; anaerobic regulation fumarate and nitrate respiration; sugar fermentation</td>
<td>(Bartolini et al., 2006; Edwards et al., 2010; Wood et al., 1998)</td>
</tr>
<tr>
<td>FNR (4Cys)</td>
<td>C1 - X₂ - C2 - X₇ - C3 - X₈₇ - C₄ (Alignment, Fig. S2)</td>
<td>α-proteobacteria (Rhizobium spp, Rhodobacter spp, Bradyrhizobium spp and others) y-proteobacteria (Pseudo-) Xanthomonas</td>
<td>N₂-Fixation, anaerobic respiration</td>
<td>(Anthamatten et al., 1992; Batut &amp; Boistard, 1994; Schluter et al., 1992; Zeilstra-Ryalls et al., 1997)</td>
</tr>
<tr>
<td>FNR (4Cys)</td>
<td>C1 - X₂ - C2 - X₃ - C₃ - X₈₈-₉₈ - C₄ (Alignment, Fig. S3)</td>
<td>Diverse bacteria (Clostridia, Spirochaeta, Leptospira)</td>
<td>Not characterized</td>
<td>Not characterized</td>
</tr>
<tr>
<td>FNR (3Cys)</td>
<td>C1 - X₂ - C2 - X₉₇ - C₄ (Alignment, Fig. S4)</td>
<td>y-proteobacteria (Acidithiobacillus caldus) Diverse bacteria (Bacteriodetes, Clostridium spp)</td>
<td>Not characterized</td>
<td>Not characterized</td>
</tr>
<tr>
<td>FNR (2Cys)</td>
<td>C₃ - X₉₂ - C₄ (most common) C2 - X₅ - C₃ C2 - X₉₇ - C₄</td>
<td>proteobacteria α-proteobacteria (Methylobacterium spp) Diverse bacteria (Bacteriodetes, Clostridia, Bacilli, Tenericutes)</td>
<td>Not characterized</td>
<td>Not characterized</td>
</tr>
<tr>
<td>FNR (1Cys)</td>
<td>C4 (or C1)</td>
<td>α-, β-, γ-proteobacteria Diverse bacteria (<em>Opitutus</em>)</td>
<td>Not characterized</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------------------------------------------------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>FNR (0Cys) (FixK&lt;sub&gt;2&lt;/sub&gt;, FixK)</td>
<td>No C (Alignment in Fig. S5)</td>
<td>α-proteobacteria (often in addition to 4Cys-FNR) <em>Bradyrhizobium japonicum</em> (FixK&lt;sub&gt;2&lt;/sub&gt;), <em>Rhizobium meliloti</em> (FixK)</td>
<td>N&lt;sub&gt;2&lt;/sub&gt;-fixation (<em>fix</em> and <em>nif</em> genes), nitrate respiration (Batut <em>et al.</em>, 1989; Fischer, 1994)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Species containing multiple forms of FNR<sub>Ec</sub>, FNR<sub>Bs</sub>, and NreB. The type of FNR or NreB (4Cys, 1Cys or 0Cys variant) is indicated. Abbreviations for the bacterial phyla in the species column: Bacilli B, Negativicutes, Neg; Clostridia, C; α-, β- or γ-Proteobacteria, α, β, or γ. *, homologs of FNR or NreB that are supposed to regulate anaerobic (nitrate) respiration.

<table>
<thead>
<tr>
<th>Species</th>
<th>FNR&lt;sub&gt;Ec&lt;/sub&gt;</th>
<th>FNR&lt;sub&gt;Bs&lt;/sub&gt;</th>
<th>NreB</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Multiple sensors within one species</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudomonas (γ)</td>
<td>4Cys* (3x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paenibacillus spp (B)</td>
<td>4Cys (2x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) Multiple variants of one sensor type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bradyrhizobium japonicum (α)</td>
<td>4Cys*, 0Cys*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhizobium meliloti</td>
<td>4Cys*, 0Cys*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhodoverax ferrireducens (β)</td>
<td>4Cys*, 1Cys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paenibacillus spp (B)</td>
<td></td>
<td>4Cys*, 1Cys</td>
<td></td>
</tr>
<tr>
<td>(iii) Different sensor classes in one species</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. caseolyticus (B)</td>
<td>-</td>
<td>4Cys</td>
<td>4Cys*</td>
</tr>
<tr>
<td>Paenibacillus sp JDR2</td>
<td>-</td>
<td>4Cys</td>
<td>1Cys</td>
</tr>
<tr>
<td>Sm. ruminantium (Neg)</td>
<td>-</td>
<td>4Cys*</td>
<td>0Cys</td>
</tr>
<tr>
<td>D. hafniense DCB2</td>
<td>4Cys</td>
<td>0Cys</td>
<td>-</td>
</tr>
<tr>
<td>Shewanella piezotolerans (γ)</td>
<td>4Cys*</td>
<td>-</td>
<td>0Cys</td>
</tr>
<tr>
<td>(iv) Hybrid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. perfringens str 13</td>
<td>-</td>
<td>No D/Cys, but FNR-Ec type</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-terminal Cys cluster with variant spacing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5–x&lt;sub&gt;2&lt;/sub&gt;–x&lt;sub&gt;3&lt;/sub&gt;–C–x&lt;sub&gt;9&lt;/sub&gt;–C113</td>
<td></td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved.
A) FNR(EC)  C1(20) -X1 -C2 -X3 -C3 -X4 -C4
FNR(Bs)  D(141) -X5 -C1 -X2 -C2 -X3
NreBSc  C1(58) -X4 -C2 -X1 -C3 -X3 -C4
WhiB3Mt  C1(23) -X5 -C2 -X1 -C3 -X4 -C4

Cys (●), Asp (▲), His (■)