J Mol Evol (1996) 43:536–540

JOURNAL OF MOLECULAR EVOLUTION

A Modular Domain of NifU, a Nitrogen Fixation Cluster Protein, Is Highly Conserved in Evolution

David M. Hwang, Adam Dempsey, Kiat-Tsong Tan, Choong-Chin Liew*

Laboratory of Molecular Cardiology, Departments of Clinical Biochemistry and Medicine, The Centre for Cardiovascular Research, The Toronto Hospital, University of Toronto, Canada

Received: 19 March 1996 / Accepted: 10 June 1996

Abstract. h*nifU*, a gene exhibiting similarity to *nifU* genes of nitrogen fixation gene clusters, was identified in the course of expressed sequence tag (EST) generation from a human fetal heart cDNA library. Northern blot of human tissues and polymerase chain reaction (PCR) using human genomic DNA verified that the h*nifU* gene represented a human gene rather than a microbial contaminant of the cDNA library. Conceptual translation of the h*nifU* cDNA yielded a protein product bearing 77% and 70% amino acid identity to NifU-like hypothetical proteins from *Haemophilus influenzae* and *Saccharomyces cerevisiae*, respectively, and 40–44% identity to the N-terminal regions of NifU proteins from several diazatrophs (i.e., nitrogen-fixing organisms). Pairwise determination of amino acid identities between the NifU-like proteins of nondiazatrophs showed that these NifU-like proteins exhibited higher sequence identity to each other (63–77%) than to the diazatrophic NifU proteins (40–48%). Further, the NifU-like proteins of non-nitrogen-fixing organisms were similar only to the N-terminal region of diazatrophic NifU proteins and therefore identified a novel modular domain in these NifU proteins. These findings support the hypothesis that NifU is indeed a modular protein. The high degree of sequence similarity between NifU-like proteins from species as divergent as humans and *H. influenzae* suggests that these proteins perform some basic cellular function and may be among the most highly conserved proteins.

Key words: Nitrogen fixation cluster — nifU gene — cDNA sequence

Introduction

The *nif* (*ni*trogen *f*ixation) cluster contains genes responsible for biological nitrogen fixation in diazatrophs (Arnold et al. 1988; Joerger and Bishop 1988). These include genes encoding enzymes responsible for reduction of diatomic nitrogen (N_2) and for other subsidiary activities such as ammonia storage and synthesis of necessary cofactors. One gene, *nifU*, encodes a protein thought to be involved in synthesizing iron-sulfur (Fe-S) metalloclusters required for nitrogen fixation (Dean et al. 1993; Zheng et al. 1993). Recently, the *nifU* gene product (NifU) was demonstrated to be a homodimer containing two identical redox Fe-S clusters, and several conserved cysteine residues were implicated in formation of these clusters (Fu et al. 1994). NifU proteins are thought to be modular proteins constructed of several domains, including a C-terminal domain common to all known nitrogen-fixing bacteria NifU proteins and an internal domain similar to a number of bacterial nitrite reductases (Masepohl et al. 1993; Ouzounis et al. 1994). While modular structure in proteins due to exon shuffling is common in eukaryotes, it is thought to be somewhat less frequent in prokaryotes, presumably due to the lack of similar shuf-

* *Present address:* Banting Institute, University of Toronto, 100 College Street, Toronto, ON M5G 1L5, Canada
Correspondence to: C.-C. Liew

```
Query= f3465
       (257 letter, both strands)

>gb|M17349|AVINIFUSV A.vinelandii nitrogen fixation genes U, S, and V complete
            cds.
            Length = 3661

  Plus Strand HSPs:

 Score = 190 (52.5 bits), Expect = 3.2e-06, P = 3.2e-06
 Identities = 70/110 (63%), Positives = 70/110 (63%), Strand = Plus

Query:   104 GAAAAGGGGAAGATTGTGGATGCTAGGTTTAAAACATTTGGCTGTGGTTCCGCAATTGCC 163
             ||||  |    |||| |||||||  | ||  | || || ||||||||||||| || ||
Sbjct:   342 GAAACCGACGTGATTCTGGATGCCGGCTTCCAGACCTTCGGCTGTGGTTCCGCCATCGCT 401

Query:   164 TCCAGCTCATTAGCCACTGAATGGGTGAAAGGAAAGACGGTGGAGGAGCC 213
             |||  |||       || ||   ||| || ||   |||  |||| ||| |
Sbjct:   402 TCCTCCTCGGCGCTGACCGAGATGGTCAAGGGCCTGACCCTGGACGAGGC 451
```

Fig. 1. Result of BLAST search for EST #F3456.

fling mechanisms in bacteria (Doolittle and Bork 1993; Ouzounis et al. 1994).

In this report, we document the identification of a human gene (h*nifU*) which exhibits striking similarity to the 5′-region of several *nifU* genes and may be among the most highly conserved genes in evolution. The hNifU protein identifies a novel domain in the N-terminal region of diazatrophic NifU proteins, further supporting the hypothesis that NifU is indeed a modular protein.

Materials and Methods

Protocols for single-pass sequencing of human fetal and adult heart cDNA libraries, as well as for analysis of EST data, have been previously described (Liew 1993; Liew et al. 1994; Hwang et al. 1994; 1995a,b). A λgt22 clone (F3465) representing the h*nifU* gene was amplified by polymerase chain reaction in the presence of vector-specific primers (Hwang et al. 1994) and Taq DNA polymerase (Pharmacia), and the resultant PCR product was subcloned into the pCR-Script vector (Stratagene). Inserts were digested with appropriate restriction endonucleases and subcloned into M13mp18 and M13mp19. Single-stranded DNA templates were isolated using *E. coli* 71–18 as a host strain and subjected to automated DNA sequencing using Pharmacia ALF and ALF-Express DNA sequencers. Sequence similarity searching was performed against the nonredundant nucleotide and protein databases at the National Center for Biotechnology Information (NCBI) using the BLAST program (Altschul et al. 1990; Gish and States 1993). The sequence was deposited in Genbank under the accession number U47101.

For Northern blot hybridizations, total RNA was extracted from human heart, liver, and placenta using the method of Chomczynski and Sacchi (1987). Northern blots were prepared using 20 μg total RNA per lane and hybridized to a radiolabeled cDNA probe derived from the purified insert of clone F3465. PCR amplification of a 234-bp fragment spanning a putative intron-exon boundary from human genomic DNA (100 ng) was performed using the following primers:

NIFU-F1: 5′-TAGTCTGTCTCCAGGATCAC-3′ (intron)
NIFU-R1: 5′-CTTGATTGCATCTTCAGCCAG-3′ (exon)

Amino acid sequence alignments were performed initially using MACDNASIS v.3.4 and alignments optimized manually. Peptide sequences of NifU and NifU-like proteins from nonhuman sources were obtained from public protein databases as follows: *Haemophilus influenzae,* gp:L45018; *Saccharomyces cerevisiae,* gp:X92441; *Anabaena sp.,* sp:P20628; *Anabaena azollae,* gp:L34879; *Azotobacter chroococcum,* sp:P23121; *Azotobacter vinelandii,* pir:S34843; *Klebsiella pneumoniae,* sp:P05343; *Nostoc commune* (fragment), gp:L23514; *Plectonema boryanum* (fragment), sp:Q00241. Sequence data from *Rattus* sp. were obtained by conceptual six-frame translation of a single EST (gb:H35834). *Candida maltosa* sequence was first identified by nucleotide similarity to the 3′-flanking region of an n-alkane-inducible cytochrome P450 gene (gb:D12718), and amino acid sequence was deduced by translation of the strand complementary to the published sequence.

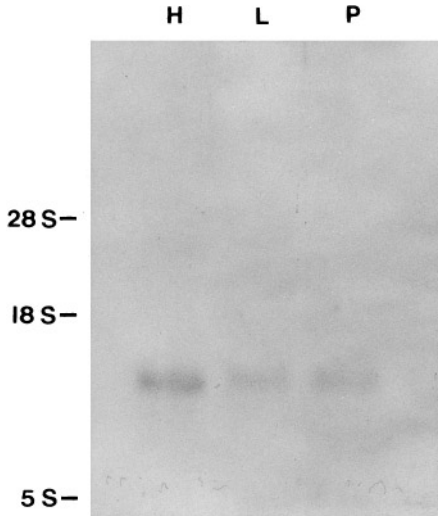

Fig. 2. Northern blot analysis of hNifU expression in various tissues. *H,* human heart; *L,* human liver; *P,* human placenta.

Results and Discussion

A clone (F3465) exhibiting a high degree of nucleotide sequence identity (63%) to the *nifU* gene of *Azotobacter vinelandii* (Genbank accession gb:M17349; Fig. 1) was first isolated in the course of sequencing a λgt22A cDNA library prepared from human fetal hearts (10–12 weeks' gestation) (Hwang et al 1994, 1995a). Northern blot detected a single species of approximately 1.1 kb in several tissues (Fig. 2), and PCR from purified human genomic DNA yielded a specific band of the expected size (234 bp; data not shown), confirming that clone F3465 did not represent a bacterial or yeast contaminant. The subsequent identification in dbEST of other matching ESTs from a variety of tissue sources (brain, liver, spleen, melanocyte, breast) further confirmed that this repre-

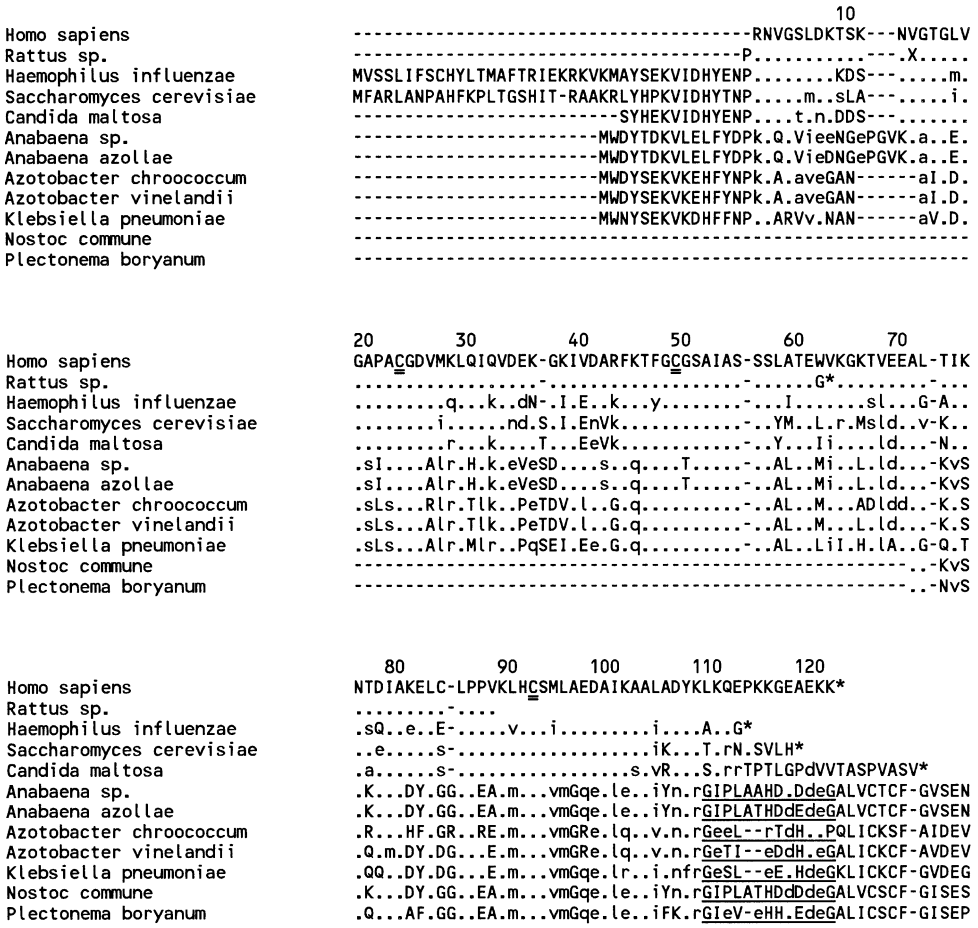

Fig. 3. Alignment of NifU and NifU-like proteins. *Dots (.)* denote residues identical to the top line; *dashes (-)* denote gaps in the sequence; *asterisks (*)* indicate stop codons. *Lower-case letters* indicate conservative substitutions. The top five sequences represent NifU-like protein sequences from non-nitrogen-fixing organisms, while the remainder represent the N-terminal regions of NifU proteins from nitrogen-fixing organisms. Conserved cysteine residues are *double-underlined,* and link regions between putative N-terminal and internal domains of the diazatrophic NifU proteins are *underlined. Numbers* correspond to the hNifU protein.

sented a human sequence that was widespread in its distribution. The identification of a putative intron and consensus splice acceptor site in several of these ESTs (gb:H18377, gb:H22470, gb:H39174, gb:H46707) suggested that this clone represented an expressed eukaryotic gene rather than a transcribed pseudogene (data not shown).

Clone F3465 was completely sequenced and its putative amino acid sequence was determined by conceptual translation. Comparison of the hNifU predicted amino acid sequence with sequences of other known NifU (diazatrophic) and NifU-like (non-diazatrophic) proteins demonstrated striking evolutionary conservation (Fig. 3). The h*nifU* gene was at least 88% identical at the nucleotide level to an EST from *Rattus* sp. (gb:H35834), and the putative protein was 97% identical over 88 amino acid residues to the predicted amino acid sequence of this EST, suggesting that gb:H35834 represents the rat homologue of the h*nifU* gene. It should be noted, however, that all three differences between the rat and human polypeptide sequences (V13X, W62G, and V63*) occurred in regions of nucleotide sequence ambiguity in the rat EST, suggesting that this may represent only a minimum degree of sequence identity. More strikingly, hNifU exhibited 77% and 70% identity (88% and 85% amino acid similarity) to hypothetical proteins from *H. influenzae* (a Gram-negative bacillus; Genpept accession gp:L45018) and *S. cerevisiae* (gp:X92441), respectively. Furthermore, hNifU showed 71% identity (84% similarity) to a partial hypothetical protein sequence of *Candida maltosa*, which was deduced by translating the antisense strand of a cytochrome P450 gene from *C. maltosa* (gb:D12718). The hNifU protein was also between 40% and 44% identical (or 67–72% similar) to known NifU proteins from nitrogen-fixing bacteria and cyanobacteria (excluding *Nostoc* and *Plectonema* fragments, the sequences of which were missing the apparently most highly conserved regions).

Pairwise determination of amino acid identities using NifU-like proteins and N-terminal domains of NifU proteins from a number of representative organisms showed the minimum identity between any two sequences to be about 40% (Table 1), indicating a high degree of evolutionary conservation. In all cases, NifU-like proteins

Table 1. Similarity matrix for selected NifU and NifU-like proteins[a]

	H. sapiens	*H. influenzae*	*S. cerevisiae*	*A. vinelandii*	*A. azollae*	*K. pneumoniae*
H. sapiens	100	77	70	44	43	41
H. influenzae		100	63	46	40	48
S. cerevisiae			100	40	40	43
A. vinelandii				100	63	71
A. azollae					100	56
K. pneumoniae						100

[a] Pairwise determination of amino acid identities were performed using MACDNASIS v.3.4 and maximized visually. Percentages were calculated for the regions of overlap shown in Fig. 3. Middle and C-terminal domains of diazatrophic NifU proteins were not included in analyses of identity and similarity

from non-nitrogen-fixing organisms exhibited much higher identity to each other (63–77%) than to the N-termini of diazatrophic NifU proteins (40–48%), which also showed higher identity to each other (56–71%) than to the NifU-like proteins (40–48%).

Analysis of global alignments (Fig. 3) identified at least 34 absolutely conserved residues (28%), suggesting these residues to be critical to NifU function. Included among these were three previously identified absolutely conserved cysteine residues (Cys23, Cys49, Cys92) (Fu et al. 1994). While these cysteines may not represent the residues previously suggested to be involved in Fe-S cluster formation in diazatrophic NifU proteins (Fu et al. 1994), their conservation, together with the conservation of flanking residues, strongly implies functional importance. Along these lines, Fu et al. (1994) postulated that NifU may function together with NifS in the sequestration of Fe and sulfide and that conserved cysteine residues other than those forming the Fe-S cluster may act to coordinate additional Fe atoms. Whether the hNifU protein is capable of binding iron merits further study and should help clarify the role of these conserved cysteines. Intriguingly, we have also recently identified a human cardiac EST (gb:N56305) exhibiting approximately 50% amino acid identity to a number of NifS proteins, suggesting that a similar system may indeed be operational in eukaryotic cells (C.C. Liew, unpublished observations).

Besides the absolutely conserved amino acids, a number of residues differed between the NifU-like and NifU proteins, but were highly conserved within each group. These included two single-residue insertions in NifU proteins relative to the hNifU sequence, the first between Lys36 and Gly37, and the second between Cys84 and Leu85. While all NifU-like proteins lacked the second insertion, interestingly, the yeast NifU-like proteins (*S. cerevisiae* and *C. maltosa*) did contain the first.

The region of similarity between the NifU-like and NifU proteins ended at Lys108 of hNifU (. . . KAALADY^{108}K). Beyond this residue, the NifU-like proteins terminated shortly and exhibited little similarity to each other. In the diazatrophic NifU proteins, however, this residue was followed by a short stretch of sequence rich in acidic residues (Fig. 3, underlined) prior to the beginning of the nitrite reductase-like internal domain (ALVCTCFGVSEN. . . in *A. azollae*) (Ouzounis et al. 1994). This acidic region may be postulated to serve as a short, mobile linker between the N-terminal and internal domains of the diazatrophic NifU proteins.

The diazatrophic NifU proteins thus appear to consist entirely of three domains: the N-terminal domain identified here, and the internal and C-terminal domains previously identified (Ouzounis et al. 1994). Interestingly, the gene encoding the NifU-like protein of *H. influenzae* is directly adjacent to a *nifS* homologue on the bacterial chromosome (gb:U32721) (Fleischmann et al. 1995). Given that the *nifU* and *nifS* genes are adjacent and co-transcribed in the *nif* clusters of *A. vinelandii* and *K. pneumoniae* (Beynon et al. 1987), this proximity in *H. influenzae* suggests that the internal and C-terminal modules may have been inserted between the ancestral *nifU* and *nifS* genes later in the evolution of the *nif* cluster. However, because *H. influenzae* contains a reduced genome (e.g., missing three citric acid cycle genes), the possibility that the NifU-like genes arose from deletion of ancestral internal and C-terminal NifU domains cannot be excluded.

It is also unclear whether the diazatrophic *nifU* and nondiazatrophic *nifU*-like genes represent paralogous (ancestral gene duplicated prior to divergence) or orthologous (no duplication) sequences (Fitch 1970). The widespread distribution of *hNifU* (Fig. 2 and EST data) and the high degree of conservation of the *NifU*-like proteins (Fig. 2) suggest that the NifU-like proteins perform an essential basic cellular function and that disruption of their function might therefore be deleterious. Duplication, though not absolutely necessary, would certainly lessen the functional constraints on modifications of such essential genes. Verification of such an event would require identification of other *nifU*-like genes in the genomes of diazatrophs.

In either case, the high degree of sequence identity between the NifU-like proteins suggests that they may be among the most highly conserved proteins known. Gupta and Golding (1993) recently suggested that the 70-kDa heat-shock protein is the most conserved protein known to date, with the minimum amino acid identity observed between any two HSP70 homologs from any of the three

domains (archaebacteria, eubacteria, eukaryotes) being approximately 42–46% (Gupta et al. 1994). Analysis of other highly conserved proteins (e.g., elongation factors 1 and 2, V/F ATPase, glutamine synthase) showed lower pairwise minimum amino acid identities (21–32%) (Gupta and Golding 1993; Gupta et al. 1994; Gogarten 1994). Strikingly, the NifU-like proteins demonstrated at least 63% amino acid identity between any pair, and a minimum of 40% identity with the N-terminal domain of any diazatrophic NifU protein, suggesting that the NifU-like proteins may be among the most highly conserved proteins known. This statement, however, should be qualified by the fact that relatively few NifU-like sequences have been characterized to date, including none from archaebacteria. Identification of other *nifU*-like genes, especially from archaebacteria, should provide further insight into the evolution of the *nifU* and *nifU*-like genes and of the *nif* cluster as a whole. Furthermore, it should also provide clues regarding the mechanisms of genetic rearrangement underlying the structure of modular proteins in prokaryotes.

Acknowledgments. This work was supported by the Canadian Genome Analysis and Technology program, the Ontario Heart and Stroke Foundation, and the Medical Research Council of Canada (MRC). D.M.H. is supported by an MRC studentship. We thank Dr. D.M. Irwin for his critical reading of the manuscript and Eva Cukerman for her excellent technical assistance.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arnold W, Rump A, Klipp W, Priefer UB, Pühler A (1988) Nucleotide sequence of a 25,206-base pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae.* J Mol Biol 203:715–738

Beynon J, Ally A, Cannon M, Cannon F, Jacobson M, Cash V, Dean D (1987) Comparative organization of nitrogen fixation-specific genes from Azotobacter vinelandii and Klebsiella pneumoniae: DNA sequence of the nifUSV genes. J Bacteriol 169:4024–4029

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Dean DR, Bolin JT, Zheng L (1993) Nitrogenase metalloclusters: structure, organization, and synthesis. J Bacteriol 175:6737–6744

Doolittle RF, Bork P (1993) Evolutionarily mobile modules in proteins. Sci Am 269:50–56

Fitch WS (1970) Distinguishing homologous from analogous proteins. Syst Zool 19:99–113

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu LI, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrman JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512

Fu W, Jack RF, Morgan V, Dean DR, Johnson MK (1994) *nifU* gene product from *Azotobacter vinelandii* is a homodimer that contains two identical [2Fe-2S] clusters. Biochemistry 33:13455–13463

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. Nat Genet 3:266–272

Gogarten JP (1994) Which is the most conserved group of proteins? Homology-orthology, paralogy, xenology, and the fusion of independent lineages. J Mol Evol 39:541–543

Gupta RS, Golding GB (1993) Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. J Mol Evol 37:573–582

Gupta RS, Golding GB, Singh B (1994) HSP70 phylogeny and the relationship between archaebacteria, eubacteria, and eukaryotes. J Mol Evol 39:537–540

Hwang DM, Hwang WS, Liew CC (1994) Single pass sequencing of a unidirectional human fetal heart cDNA library to discover genes of the cardiovascular system. J Mol Cell Cardiol 26:1329–1333

Hwang DM, Fung YW, Wang RX, Laurenssen CM, Ng SH, Lam WY, Tsui KW, Fung KP, Waye M, Lee CY, Liew CC (1995a) Analysis of expressed sequence tags from a fetal human heart cDNA library. Genomics 30:293–298

Hwang DM, Wang RX, Liew CC (1995b) Direct automated sequencing of single λ-phage plaques by exponential amplification sequencing. Anal Biochem 231:460–463

Joerger RD, Bishop PE (1988) Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii.* J Bacteriol 170:1475–1487

Liew CC (1993) A human heart cDNA library—the development of an efficient and simple method for automated DNA sequencing. J Mol Cell Cardiol 25:891–894

Liew CC, Hwang DM, Fung YW, Laurenssen C, Cukerman E, Tsui S, Lee CY (1994) Catalogue of genes in the cardiovascular system by expressed sequence tags (ESTs). Proc Natl Acad Sci USA 91: 10645–10649

Masepohl B, Angermüller S, Hennecke S, Hübner P, Moreno-Vivian C, Klipp W (1993) Nucleotide sequence and genetic analysis of the *Rhodobacter capsulatus* ORF6-*nifU*$_1$*SVW* gene region: possible role of NifW in homocitrate processing. Mol Gen Genet 238:369–382

Ouzounis C, Bork P, Sander C (1994) The modular structure of NifU proteins. Trends Biochem Sci 19:199–200

Zheng L, White R, Cash VL, Jack RF, Dean DR (1993) Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. Proc Natl Acad Sci USA 90:2754–2758