Brazilian Journal of Medical and Biological Research (2001) 34: 1379-1395
ISSN 0100-879X

Sequencing and promoter analysis of the *nifENXorf3orf5fdxAnifQ* operon from *Azospirillum brasilense* Sp7

D.P. Potrich[1], T.A. Bressel[2], I.S. Schrank[2] and L.M.P. Passaglia[1]

Departamentos de [1]Genética and [2]Biologia Molecular e Biotecnologia, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

Abstract

A 40-kb DNA region containing the major cluster of *nif* genes has been isolated from the *Azospirillum brasilense* Sp7 genome. In this region three *nif* operons have been identified: *nifHDKorf1Y*, *nifENXorf3orf5fdxAnifQ* and *orf2nifUSVorf4*. The operons containing *nifENX* and *nifUSV* genes are separated from the structural *nifHDKorf1Y* operon by about 5 kb and 10 kb, respectively. The present study shows the sequence analysis of the 6045-bp DNA region containing the *nifENX* genes. The deduced amino acid sequences from the open reading frames were compared to the *nif* gene products of other diazotrophic bacteria and indicate the presence of seven ORFs, all reading in the same direction as that of the *nifHDKorf1Y* operon. Consensus σ^{54} and NifA-binding sites are present only in the promoter region upstream of the *nifE* gene. This promoter is activated by NifA protein and is approximately two-times less active than the *nifH* promoter, as indicated by the ß-galactosidase assays. This result suggests the differential expression of the *nif* genes and their respective products in *Azospirillum*.

Correspondence
L.M.P. Passaglia
Centro de Biotecnologia
Departamento de Genética, UFRGS
Av. Bento Gonçalves, 9500
Prédio 43421
Caixa Postal 15005
91501-970 Porto Alegre, RS
Brasil
E-mail: luciane@dna.cbiot.ufrgs.br

Research supported by CNPq and FAPERGS.

Received April 12, 2001
Accepted July 25, 2001

Key words
- *Azospirillum brasilense*
- *nif genes*
- *nifENX* operon
- Promoter activity

Introduction

The biological process of nitrogen fixation is catalyzed by the nitrogenase enzyme, an enzyme complex containing the nitrogenase reductase (Fe-protein) and dinitrogenase (MoFe-protein). The Fe-protein (NifH) is a dimer of identical subunits, which contains a single 4Fe-4S cluster. The MoFe-protein is an $\alpha_2ß_2$ tetramer (NifD and NifK) including two iron-molybdenum cofactors (FeMoco) and about 16 additional iron and acid-labile sulfur irons (1). Twenty *nif* genes have been described in the extensively studied nitrogen-fixing organism *Klebsiella pneumoniae*, but they are not all essential for N_2-fixing activity (2). These genes are located on a contiguous 24-kb chromosomal DNA fragment and organized into several operons (2). The observation that the nucleotide sequence and protein structure of dinitrogenase and dinitrogenase reductase are conserved among all nitrogen-fixing organisms allowed the identification and cloning of homologous sequences from several other diazotroph organisms (2-6). However, the organization of the *nif* genes has been observed to diverge among the different diazotrophic bacteria. In some cases, the *nifHDK* genes are transcribed as a single unit, as

observed in several fast-growing rhizobia including *Sinorhizobium meliloti* (7) and *Rhizobium leguminosarum* (8) and in the newly described *nif* cluster from *Gluconacetobacter diazotrophicus* (9). In other cases, the *nifHDK* operon contains other *nif* genes, as reported for *Azotobacter vinelandii*, where they constitute an operon together with *nifTY*, *orf1* and *orf2* (3), *Herbaspirillum seropedicae*, where they form a large operon composed by *nifHDKENXorf1orf2* (10), and *Azospirillum brasilense*, where they are clustered with *orf1* and *nifY* (4). In other cases, they are separated into two different regions (*nifH* and *nifDK*), as shown in the slow-growing species *Bradyrhizobium japonicum* (11) and *Bradyrhizobium* sp (Vigna) (cowpea *Bradyrhizobia*) (12,13).

In all diazotrophics several other genes have been reported to be essential for producing an active nitrogenase. The gene products of *nifE* and *nifN* are involved in synthesis of the FeMoco of nitrogenase and are structurally related to the products of the nitrogenase genes *nifDK* (14). The *nifEN* genes from *K. pneumoniae* are clustered into a single operon, together with *nifX*, downstream from the *nifHDKTY* operon (2), and are expressed from the promoter located upstream of the *nifE* gene (14,15). In *Rhodobacter capsulatus*, the *nifENXorf4orf5nifQ* operon contains the *nifX* and *nifQ* genes (16). Transposon insertions into the *nifE* and *nifN* genes yielded the Nif$^-$ phenotype in *K. pneumoniae* (17) and *A. brasilense* (18,19). However, transposon insertion in *R. capsulatus nifX* showed that its product is not essential for nitrogen fixation (20). Recent results indicate that the *Azotobacter vinelandii* NifX protein participates in FeMoco synthesis *in vitro* (21). The *orf4* of *R. capsulatus* encodes a ferredoxin-like protein and is homologous to *orf3* found in *A. vinelandii* as part of the operon *nifENXorf3orf4* (3).

Bacteria of the genus *Azospirillum* are diazotrophs capable of fixing nitrogen in free-living state or associated with roots of economically important grasses (22). In *A. brasilense* three *nif* operons, *nifHDKorf1Y*, *nifENXorf3orf5fdxAnifQ* and *orf2nifUSV orf4*, have been identified and their sequences have been determined (4,23, and the present study). The operons encompassing the *nifENX* and *nifUSV* genes are separated from the structural *nifHDKorf1Y* operon by about 5 and 10 kb, respectively (23). A putative -24/-12 promoter element has been found in the promoter region of the *nifH*, *nifE* and *nifU* genes. The promoter site of *nifH* was studied in more detail and showed two overlapping NifA-binding sites, where the examination of activation of the mutant *nifH* promoter by NifA revealed that the integrity of the NifA-binding site closer to *nifH* is required for the most efficient activation (24).

In the present study, we have determined the nucleotide sequence of the *A. brasilense* Sp7 *nifENX* genes, three open reading frames (*orf3*, *orf5* and *fdxA*) and the *nifQ* gene, which constitute an operon transcribed from a single promoter upstream of the *nifE* gene. We have also shown that this promoter is activated by NifA protein and is less active than the *nifH* promoter, as measured by ß-galactosidase promoter fusion.

Material and Methods

Bacterial strains, plasmids and growth conditions

The bacterial strains and plasmids used in the present study are listed in Table 1. LB medium (25) was used for growing *E. coli* strains at 37ºC. *A. brasilense* Sp7 strain was grown at 30ºC in either LB medium or MMAb minimal medium (26) using 0.5% malate as carbon source. For the ß-galactosidase assays, A medium was used for growing *E. coli* MC1061 strain at 28ºC, as indicated by Sambrook et al. (25). The medium was supplemented with the antibiotics tetracycline and/or ampicillin when necessary, at

concentrations of 10 and 100 g/ml, respectively.

DNA manipulations and sequencing

Plasmid DNA preparation, restriction enzyme analysis, transformation and electrophoresis on agarose or polyacrylamide gel were performed as described by Sambrook et al. (25). Restriction endonucleases and other enzymes were purchased from Pharmacia (Uppsala, Sweden) or Gibco/BRL (Gaithersburg, MD, USA) and used according to the manufacturer's instructions.

The nucleotide sequence was determined by the chain-termination method of Sanger et al. (as described in Ref. 27) using ^{33}P-ddNTPs and the ThermoSequenase radiolabeled terminator cycle sequencing kit (Amersham Pharmacia Biotech, Uppsala, Sweden). Defined restriction fragments from the *Eco*RI/*Pst*I *A. brasilense* DNA region shown in Figure 1 were subcloned to generate several sequencing plasmids (Table 1). The junctions of all subclones were checked by sequencing directly from the larger parental plasmid, pWY1 (this laboratory), using oligonucleotides generated from the sequences already obtained. All manual sequencing data were confirmed by using the Molecular Genetics Instrumentation Facility, University of Georgia, Atlanta, GA, USA. Analysis of DNA sequences and comparison with nucleotide and deduced protein sequences from

Table 1. Bacterial strains and plasmids used in the present study.

Bacteria	Strain	Relevant characteristics	Reference
Azospirillum brasilense	Sp7	ATCC 29145, Amp^R	22
Escherichia coli	XL1-Blue	*supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac-F'[proAB$^+$ lacI^q lacZΔ15 Tn10(tet^r)]*	25
	MC1061	*hsdR mcrB araD 139Δ (araABC-leu) 7679 ΔlacX74 galU galK rpsL thi*	28

Plasmid	Relevant characteristics	Reference
pBluescript	Amp^R	Stratagene
pUC18	Amp^R	25
pCK3	pRK290 derivative containing the *Klebsiella pneumoniae nifA* gene	29
pMC1403	translational fusion *lacZ* plasmid	28
pMCH	pMC1403 derivative containing the *A. brasilense nifH* promoter	24
pMCE	pMC1403 derivative containing the *A. brasilense nifE* promoter	Present paper
pWY1	pUC18 derivative containing a 5.8-kb *Eco*RI/*Pst*I *A. brasilense* DNA fragment	This laboratory
pKS2.2	pBSKS+ derivative containing a 2.2-kb *Eco*RI/*Hin*dIII DNA fragment	Present paper
pKS0.8	pBSKS+ derivative containing a 0.8-kb *Hin*dIII/*Sal*I DNA fragment	Present paper
pKS1.5	pBSKS+ derivative containing a 1.5-kb *Sal*I DNA fragment	Present paper
pKS1.3	pBSKS+ derivative containing a 1.3-kb *Sal*I/*Pst*I DNA fragment	Present paper

other organisms were performed using the GCG (Wisconsin Package Version 9.0, Genetics Computer Group, Madison, WI, USA) computer programs (licensed to CENARGEN-EMBRAPA, Brasília, DF, Brazil).

PCR

The 590-bp DNA fragment containing the promoter region of the *nifE* gene was obtained by PCR with *A. brasilense* Sp7 total DNA as a template. The oligonucleotides used for PCR were 5' CGCCGCCAAC GACGAGGTCAAGAA 3', which corresponds to the C-terminal region of the *mcpAb* gene (Schneider C, unpublished results), and 5' GGTGGAGCAACCCGGCTCGTT 3', which corresponds to the beginning of the *nifE*-coding region. The amplified fragment was cloned into the *Hin*cII site of the pBluescript KS+ vector and completely sequenced to check the absence of mutations.

Plasmid construction and ß-galactosidase assays

The strategy used to obtain the pMCE plasmid is shown in Figure 1. DNA from the pBluescript KS+ plasmid carrying the 590-

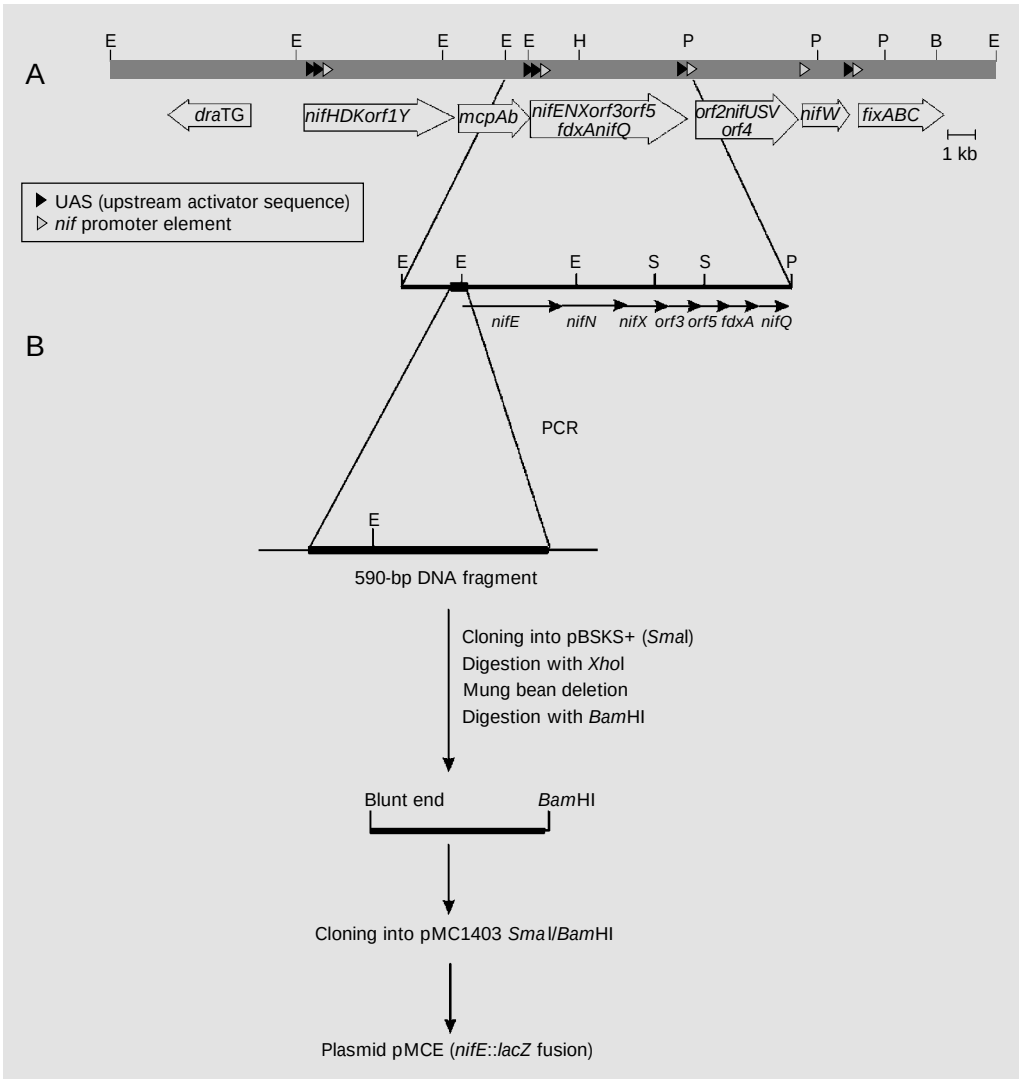


Figure 1. Physical map of the *nif* gene cluster in the genome of *Azospirillum brasilense* (A) and cloning strategy of the *nifE* promoter region (B). *A*, Represents the organization of *nif* and other genes in the chromosome of *A. brasilense*. Arrows represent the positions and direction of transcription of the operons. The *nifENXorf3orf5fdxAnifQ* operon is shown in detail in the lower part of the figure. Each restriction fragment was cloned into the pUC18 plasmid, as described in Table 1. *B*, Cloning strategy of the *nifE* promoter region into the pMC1403 plasmid. E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; B, *Bam*HI; S, *Sal*I.

bp PCR amplified DNA fragment (described above) was digested with *Xho*I (vector site) followed by mung bean deletion to generate a shortened blunt-end fragment. This linear pBluescript recombinant plasmid was further digested with *Bam*HI and the blunt-end/*Bam*HI fragment was cloned into the pMC1403 *Sma*I/*Bam*HI-digested vector containing the *lacZ* translational fusion. The recombinant pMCE was used to monitor the activity of the *nifE* promoter by measuring ß-galactosidase activity in *E. coli* MC1061 (28). The *K. pneumoniae* NifA protein synthesized constitutively was provided by the pCK3 plasmid (29). Plasmid pMCH, containing a *nifH::lacZ* fusion (24), was used as a positive control.

The ß-galactosidase activities of strains carrying the *nif-lacZ* promoter fusions were determined according to the procedure of Miller et al. (as described in 25) at 28ºC.

Nucleotide sequence accession number

The nucleotide sequence of the *nifENX orf3orf5fdxAnifQ* operon of *A. brasilense* Sp7 has been deposited in Genbank under accession No. AF361867 along with the predicted amino acid sequence.

Results

The *nifENXorf3orf5fdxAnifQ* genes of *A. brasilense* Sp7 are in the same transcriptional unit and are expressed from a σ^{54} promoter upstream of *nifE*.

Tn5 mutagenesis of the *A. brasilense* DNA regions downstream from the *nifHDK* operon revealed a region of approximately 6000 bp containing the *nifENX* genes (19). In this region seven complete open reading frames (ORFs) were identified, all reading in the same direction as that of the *nifHDK* operon, as indicated in Figure 1. The assignment of genes was based on deduced amino acid sequence identities to the *nif* gene products of *K. pneumoniae* and several other diazotrophic bacteria. The likely initiation codon for all seven genes or ORFs is an ATG preceded in each case by a characteristic AG-rich ribosome-binding site.

Only one region within the 6045-bp sequence displays close similarity to the σ^{54} recognition consensus sequence (15) and occurs at 45 bp upstream from the translational initiation codon of *nifE*. We identified two *nif*-specific upstream activator sequences, TGT-N_{10}-ACA, characteristic of NifA-dependent promoters (30), located 72 and 45 bp upstream of the putative σ^{54}-dependent promoter, respectively.

The *A. brasilense nifE*-, *nifN*- and *nifX*-coding regions are 1413, 1398 and 471 nucleotides long, respectively, and predict polypeptides of 417 residues corresponding to NifE, 466 residues corresponding to NifN, and 157 residues corresponding to NifX proteins. An overlapping coding region was observed between *nifN* and *nifX*. The predicted amino acid sequences from *A. brasilense nifE*, *nifN* and *nifX* genes were compared to their counterpart sequences from other diazotrophic organisms, as shown in Appendix 1 (I, II and III) (see pages 1388-1393). The identity of the deduced *A. brasilense* NifE, NifN and NifX amino acid sequences is distributed over the entire length of corresponding proteins. Conserved cysteine residues (marked by black dots in Appendix 1) are present in all NifE, NifN and NifX proteins, except for some cysteine residues that were not found in the *H. seropedicae* NifE protein (11; Appendix 1(I)). The highest similarity level was found between *A. brasilense* and *G. diazotrophicus* NifE (9) proteins (60.8%), *A. brasilense* and *B. japonicum* NifN (31) proteins (50.3%), and *A. brasilense* and *H. seropedicae* NifX (10) proteins (55%).

In addition to the *nifENX* genes, four other ORFs were identified (Figure 1). The first one, *orf3*, revealed high similarity with *G. diazotrophicus orf1* (59.6%) (9). The contiguous ORF, *orf5*, showed 61.8% similarity

to *Azorhizobium caulinodans orf1* (32) and 61.2% similarity to *Acetobacter diazotrophicus orf2* (9). The third ORF was homologous to a ferredoxin-like protein from *R. capsulatus* (54.9% similarity) (16) and a ferredoxin III protein from the cyanobacterium *Plectonema boryanum* (57.3% similarity) (33) and was assigned to *fdxA.* The comparison between these ORFs and their homologous counterparts is shown in Appendix 2 (see pages 1394 and 1395).

The last ORF identified within the sequenced region of *A. brasilense* encoded a protein of 196 amino acids and the alignment of the deduced amino acid sequence of this *A. brasilense* ORF with that of NifQ proteins from *K. pneumoniae* (34), *R. capsulatus* (16), *Acetobacter diazotrophicus* (9), *Enterobacter agglomerans* (35), *Azotobacter vinelandii* (4) and *Rhizobium* sp (36) revealed an overall ranging of homology from 30.5 to 37.5% of similarity (Appendix 1(IV)). The identity between the NifQ proteins was mainly restricted to the C-terminal part, including a typical cysteine motif (marked in Appendix 1) found in all NifQ proteins identified to date.

Inverted repeat structures were detected only in two regions at 110 and 173 nucleotides from the *nifQ* stop codon. Messenger RNA transcribed from these regions could potentially form a characteristic stem-and-loop secondary structure and may be involved in the termination of transcription. No other ORF was found in the region between the end of the *nifQ* gene and the beginning of the *orf2nifUSVorf4* operon.

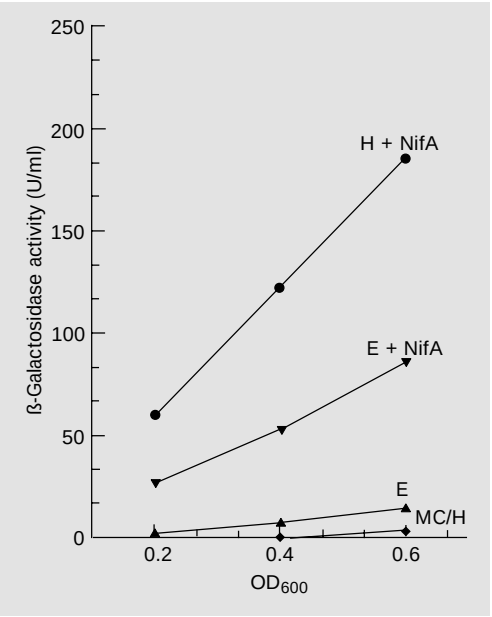


Figure 2. *In vivo* expression of *nif'-'lacZ* from the *nifE* (E) and *nifH* (H) promoters in the *E. coli* MC1061 strain in the presence (+ NifA) or absence of the NifA activator protein provided by pCK3 plasmid. Expression of *nif'-'lacZ* fusion was measured as a function of the OD_{600} for cultures grown in A medium (see Material and Methods). In each case, three data points (between OD_{600} of 0.2 and 0.6) were used to determine the slope of the line, which reflects the differential rate of *nif'-'lacZ* expression. Curve labeled MC (which overlaps the curve from the *nifH* promoter in the absence of NifA) gives data for the *E. coli* MC1061 host strain (negative control).

The *nifE* promoter activity is dependent on NifA protein

To verify the functionality of the putative *nif* promoter identified in the region upstream of the *nifE* initiation codon we amplified by PCR a 590-bp DNA fragment containing the regulatory region of the *nifENX orf3orf5fdxAnifQ* operon. This fragment was cloned into the pMC1403 translational fusion plasmid and the recombinant pMCE was used to monitor the activity of the *nifE* promoter by measuring ß-galactosidase activity in *E. coli* MC1061 (as described in Material and Methods). As shown in Figure 2, the ß-galactosidase activity driven by the *A. brasilense nifE* promoter was only detected if the NifA protein was provided in trans via the pCK3 plasmid that promotes constitutive expression of the *K. pneumoniae* NifA. In fact, *E. coli* MC1061 harboring pMCE, but not pCK3, showed only background levels of ß-galactosidase activity (Figure 2). These results demonstrate that the activity of the *nifE* promoter is dependent on NifA.

We also used the pMCH plasmid (24) as a positive control and we were able to compare the activities of both *A. brasilense nif* promoters. The *nifE* promoter is approximately two times less active than the *nifH* promoter, as indicated by the ß-galactosidase assays (Figure 2, compare E + NifA with H + NifA).

Discussion

Nitrogen fixation genes, *nif* genes, are

frequently clustered into different transcriptional units. The overall organization of *nif* operons also shows some conservation among the genomes of diazotrophic bacteria. In *A. brasilense* the distribution of the *nifENX* genes within the operon resembles the organization described for other diazotrophic organisms. NifE and NifN proteins were found to be similar to those present in other α-group Proteobacteria. NifX was more similar to the gene product of *H. seropedicae*, a member of the ß-group of Proteobacteria, than the product of the γ-group members. However, NifQ was similar to the NifQ from *Azotobacter vinelandii* which belongs to the γ-group of Proteobacteria.

In some members of the Proteobacteria the *nif* genes, initially described in *K. pneumoniae*, are organized in operons together with different ORFs. In *A. brasilense* the organization of *nifENXorf3orf5fdxAnifQ* is similar to that found in *G. diazotrophicus*. In fact, the organization of the *nif* gene in *G. diazotrophicus* and *A. brasilense* seems to be highly similar: the *nifENXorf3orf5fdx AnifQ* operon containing orthologous proteins similarly organized, the presence of the *mcpA* gene in the surroundings of the *nif* cluster and the organization of the *orf2nifUSV* operon may indicate that both microorganisms share common characteristics, especially concerning their ability to enhance plant growth through the transfer of bacterially fixed nitrogen and the production of plant growth-stimulating factor(s) (9,22). In addition, *A. brasilense* ORF3 was found to be similar to ORF1 from *H. seropedicae* present within the related *nifENXorf1orf2* gene cluster (10) and to ORF3 from *A. vinelandii* also located downstream to the *nifX* gene (3).

Several studies have demonstrated that *nifEN* are essential for the nitrogen fixation process. However, determination of the role of NifX during nitrogen fixation shows some differences concerning the diazotrophic bacteria. Araújo et al. (19), using insertional mutagenesis, obtained five Tn5 insertions in the region adjacent to the *nifHDK* genes of the *A. brasilense* genome. Four of them were located in the region corresponding to the *nifEN* genes and gave a Nif$^-$ phenotype. The fifth insertion which gave a Nif$^+$ phenotype was within the corresponding region of the *A. brasilense nifX* gene, as further confirmed by sequencing analysis (present paper), and like the *nifX* gene from the ß-group member *H. seropedicae* it proved not to be essential for nitrogen fixation (10). In contrast, NifX was shown to be essential for nitrogenase activity in an *in vitro* system in *A. vinelandii* (21), although under laboratory conditions, *A. vinelandii nifX* showed a Nif$^+$ phenotype (3). To date, the *nifX* mutant in all diazotrophs described has wild-type nitrogenase activity (3,10,20).

Transcriptional and translational organization of the *nif* gene cluster of *A. brasilense* revealed conserved features. As found for the other *A. brasilense nif/fix* operons, sequence analysis of the *nifENXorf3orf5fdx AnifQ* revealed σ^{54} and NifA-binding sites upstream of *nifE*, which are required, respectively, for *nif* promoter recognition and for *nif* gene transcriptional activation. Similarly to other *nif* genes, the *A. brasilense nifE* promoter is positively controlled by the activator NifA protein.

The nucleotide sequence and promoter analysis of the *nifENXorf3orf5fdxAnifQ* operon in *A. brasilense* revealed the presence of typical features in the deduced protein common among the related proteins in other organisms, as well as sequences upstream of *nifE* indicating a NifA-dependent transcriptional activation. Moreover, ORFs similar to *orf3*, *orf5*, and the putative *fdxA* were also described in other groups of the Proteobacteria.

References

1. Orme-Johnson WH (1985). Molecular basis of nitrogen fixation. *Annual Review of Biophysics and Biophysical Chemistry*, 14: 419-459.
2. Arnold W, Rump A, Klipp W, Priefer O & Pühler A (1988). Nucleotide sequence of a 24,206 base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *Journal of Molecular Biology*, 203: 715-738.
3. Jacobson MR, Brigle KE, Bennett LT, Setterquist RA, Wilson MS, Cash VL, Beynon JL, Newton WE & Dean DR (1989). Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *Journal of Bacteriology*, 171: 1017-1027.
4. Passaglia LMP, Nunes CP, Zaha A & Schrank IS (1991). The *nifHDK* operon in the free-living nitrogen-fixing bacteria *Azospirillum brasilense* sequentially comprises genes H,D,K, an 353 bp orf and gene Y. *Brazilian Journal of Medical and Biological Research*, 24: 649-675.
5. Machado IMP, Yates MG, Machado HB, Souza EM & Pedrosa FO (1996). Cloning and sequencing of the nitrogenase structural genes *nifHDK* of *Herbaspirillum seropedicae*. *Brazilian Journal of Medical and Biological Research*, 29: 1599-1602.
6. Kessler PS & Blank CLJA (1998). The *nif* gene operon of the methalogenic archaeon *Methanococcus maripaludis*. *Journal of Bacteriology*, 180: 1504-1511.
7. Corbin D, Barran L & Ditta GS (1983). Organization and expression of *Rhizobium meliloti* nitrogen fixation genes. *Proceedings of the National Academy of Sciences, USA*, 80: 3005-3009.
8. Krol AJM, Hontelez JGJ, Roozendaal B & Kammen AV (1982). On the operon structure of the nitrogenase genes of *Rhizobium leguminosarum* and *Azotobacter vinelandii*. *Nucleic Acids Research*, 10: 4147-4157.
9. Lee S, Reth A, Meletzus D, Sevilla M & Kennedy C (2000). Characterization of a major cluster of *nif*, *fix*, and associated genes in a sugarcane endophyte, *Acetobacter diazotrophicus*. *Journal of Bacteriology*, 182: 7088-7091.
10. Klassen G, Pedrosa FO, Souza EM, Yates MG & Rigo LU (1999). Sequencing and functional analysis of the *nifENXorf1orf2* gene cluster of *Herbaspirillum seropedicae*. *FEMS Microbiology Letters*, 181: 165-170.
11. Fischer HM & Hennecke H (1984). Linkage map of the *Rhizobium japonicum nifH* and *nifDK* operons encoding the polypeptides of the nitrogenase enzyme complex. *Molecular and General Genetics*, 196: 537-540.
12. Jagadish M, Yun AC, Noti JD, Folkerts O & Szalay AA (1985). Structural and functional organization of the *nif* region in the slow-growing, broad host-range cowpea *Rhizobium* strain IRc78. In: Szalay A & Legocki RP (Editors), *Advances in Molecular Genetics of Bacterial-Plant Interactions*. Cornell University Publishers, Ithaca, NY, 27-31.
13. Yun AC & Szalay AA (1984). Structural genes of dinitrogenase and dinitrogenase reductase are transcribed from two separate promoters in the broad host range cowpea *Rhizobium* strain IRc78. *Proceedings of the National Academy of Sciences, USA*, 81: 7358-7362.
14. Brigle KE, Weiss MC, Newton WE & Dean DR (1987). Products of the iron-molybdenum cofactor-specific biosynthetic genes, *nifE* and *nifN*, are structurally homologous to the products of the nitrogenase molybdenum-iron protein genes, *nifD* and *nifK*. *Journal of Bacteriology*, 169: 1547-1553.
15. Beynon JL, Cannon MC, Buchanan-Wollaston AV & Cannon FC (1983). The *nif* promoters of *Klebsiella pneumoniae* have a characteristic primary structure. *Cell*, 34: 665-671.
16. Moreno-Vivian C, Hennecke S, Pühler A & Klipp W (1989). Open reading frame 5 (ORF5), encoding a ferredoxin-like protein, and *nifQ* are cotranscribed with *nifE*, *nifN*, *nifX*, and ORF4 in *Rhodobacter capsulatus*. *Journal of Bacteriology*, 171: 2591-2598.
17. Roberts GP, MacNeil T, MacNeil D & Brill WJ (1978). Regulation and characterization of protein products coded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*. *Journal of Bacteriology*, 136: 267-279.
18. Galimand M, Perroud B, Delorme F, Paquelin A, Vieille C, Bozouklian H & Elmerich C (1989). Identification of DNA regions homologous to nitrogen fixation genes *nifE*, *nifUS* and *fixABC* in *Azospirillum brasilense* Sp7. *Journal of General Microbiology*, 135: 1-13.
19. Araújo EF, Zaha A, Schrank IS & Santos DS (1988). Characterization of DNA segments adjacent to the *nif*HDK genes of *Azospirillum brasilense* Sp7 by Tn*5* site-directed mutagenesis. In: Klingmüller W (Editor), *Azospirillum IV: Genetics, Physiology and Ecology*. Proceedings of the Fourth Bayreuth *Azospirillum* Workshop, Bayreuth, Germany, 16-25.
20. Moreno-Vivian C, Schmehl M, Masepohl B, Arnold W & Klipp W (1989). DNA sequence and genetic analysis of the *Rhodobacter capsulatus nifENX* gene region: homology between NifX and NifB suggests involvement of NifX in processing of the iron-molybdenum cofactor. *Molecular and General Genetics*, 216: 353-363.
21. Shah VK, Rangaraj P, Ranjini C & Allen RM (1999). Requirement of NifX and other *nif* proteins for *in vitro* biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Journal of Bacteriology*, 181: 2797-2801.
22. Tarrand JJ, Krieg NR & Döbereiner J (1978). A taxonomic study of the *Spirillum lipoferum* group, with descriptions of a new genus, *Azospirillum* gen. nov. and two species, *Azospirillum lipoferum* (Beijerinck) comb. nov. and *Azospirillum brasilense* sp nov. *Canadian Journal of Microbiology*, 24: 967-980.
23. Frazzon J & Schrank IS (1998). Sequencing and complementation analysis of the *nifUSV* genes from *Azospirillum brasilense*. *FEMS Microbiology Letters*, 159: 151-158.
24. Passaglia LMP, Schrank A & Schrank IS (1995). The two overlapping *Azospirillum brasilense* upstream activator sequences have differential effects on *nifH* promoter activity. *Canadian Journal of Microbiology*, 41: 849-854.
25. Sambrook J, Fritsch EF & Maniatis T (1989). *Molecular Cloning: A Laboratory Manual, 2*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
26. Vanstockem M, Michiels KW, Vanderleyden J & Gool APV (1987). Transposon mutagenesis of *Azospirillum brasilense* and *Azospirillum lipoferum*: physical analysis of Tn*5* and Tn*5*-Mob insertion mutants. *Applied and Environmental Microbiology*, 53: 410-415.
27. Sanger F, Nicklen S & Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA*, 74: 5463-5467.
28. Casadaban MJ, Martinez-Arias A, Shapira S & Chou JH (1983). ß-galactosidase gene fusions for analysing gene expression in *Escherichia coli* and yeast. *Methods in Enzymology*, 100: 293-308.
29. Kennedy C & Drummond MH (1985). Use

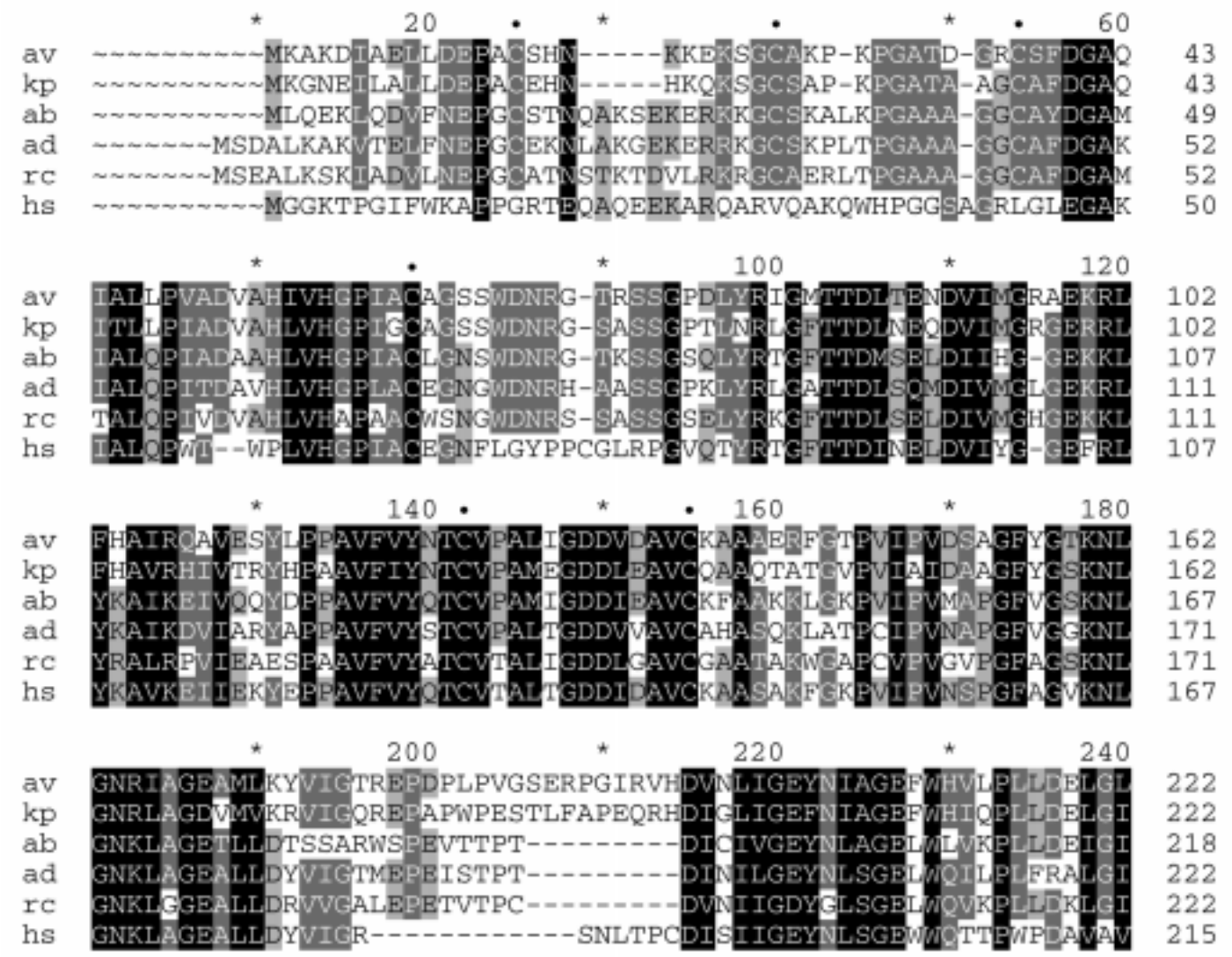
of cloned *nif* regulatory elements from *Klebsiella pneumoniae* to examine *nif* regulation in *Azotobacter vinelandii*. *Journal of General Microbiology*, 131: 1787-1795.

30. Buck M, Miller S, Drummond MH & Dixon RA (1986). Upstream activator sequences are present in the promoters of nitrogen fixation genes. *Nature*, 320: 374-378.
31. Aguilar OM, Taormino JP, Thöny B, Ramseier T, Hennecke H & Szalay AA (1990). The *nifEN* genes participating in FeMo cofactor biosynthesis and genes encoding dinitrogenase are part of the same operon in *Bradyrhizobium* species. *Molecular and General Genetics*, 224: 413-420.
32. Arigoni F, Kaminski PA, Hennecke H & Elmerich C (1991). Nucleotide sequence of the *fixABC* region of *Azorhizobium caulinodans* ORS571: similarity of the *fixB* product with eucaryotic flavoproteins, characterization of *fixX*, and identification of *nifW*. *Molecular and General Genetics*, 225: 514-520.
33. Schrautemeier B, Neveling U & Schmitz S (1994). Characterization of the genome region encoding an *fdxH*-type ferredoxin and a new 2[4Fe-4S] ferredoxin from the nonheterocystous, nitrogen-fixing cyanobacterium *Plectonema boryanum* PCC 73110. *Journal of Bacteriology*, 176: 1037-1046.
34. Buikema WJ, Klingensmith JA, Gibbons SL & Ausubel FM (1987). Conservation of structure and location of *Rhizobium meliloti* and *Klebsiella pneumoniae nifB* genes. *Journal of Bacteriology*, 169: 1120-1126.
35. Siddavattan D, Singh M & Klingmüller W (1993). Structure of the *nifQ* gene from *Enterobacter agglomerans* 333 and its overexpression in *Escherichia coli*. *Molecular and General Genetics*, 239: 435-440.
36. Freiberg C, Fellay R, Bairochi A, Broughton WJ, Rosenthal A & Perret R (1997). Molecular basis or symbiosis between *Rhizobium* and legumes. *Nature*, 387: 394-401.
37. Setterquist RA, Brigle KE, Beynon JL, Cannon MC, Ally A, Cannon FC & Dean DR (1988). Nucleotide sequence of the *nifE* and *nifN* genes from *Klebsiella pneumoniae*. *Nucleic Acids Research*, 16: 5215.
38. Beynon JL, Cannon MC, Buchanan-Wollaston AV, Ally A, Setterquist RA, Dean DR & Cannon FC (1988). The nucleotide sequence of the *nif*T, *nif*Y, *nif*X and *nif*W genes of *K. pneumoniae*. *Nucleic Acids Research*, 16: 9860.
39. Nicholas KB & Nicholas Jr HB (1997). GeneDoc: Analysis and visualization of genetic variation. Available at http://www.cris.com/~Ketchup/genedoc.shtml. Accessed March 2001.
40. Schrautemeier B, Neveling U & Schmitz S (1995). Distinct and differently regulated Mo-dependent nitrogen-fixing systems evolved for heterocysts and vegetative cells of *Anabaena variabilis* ATCC 29413: characterization of the *fdx*H1/2 gene regions as part of the *nif*1/2 gene clusters. *Molecular Microbiology*, 18: 357-369.

Appendix 1

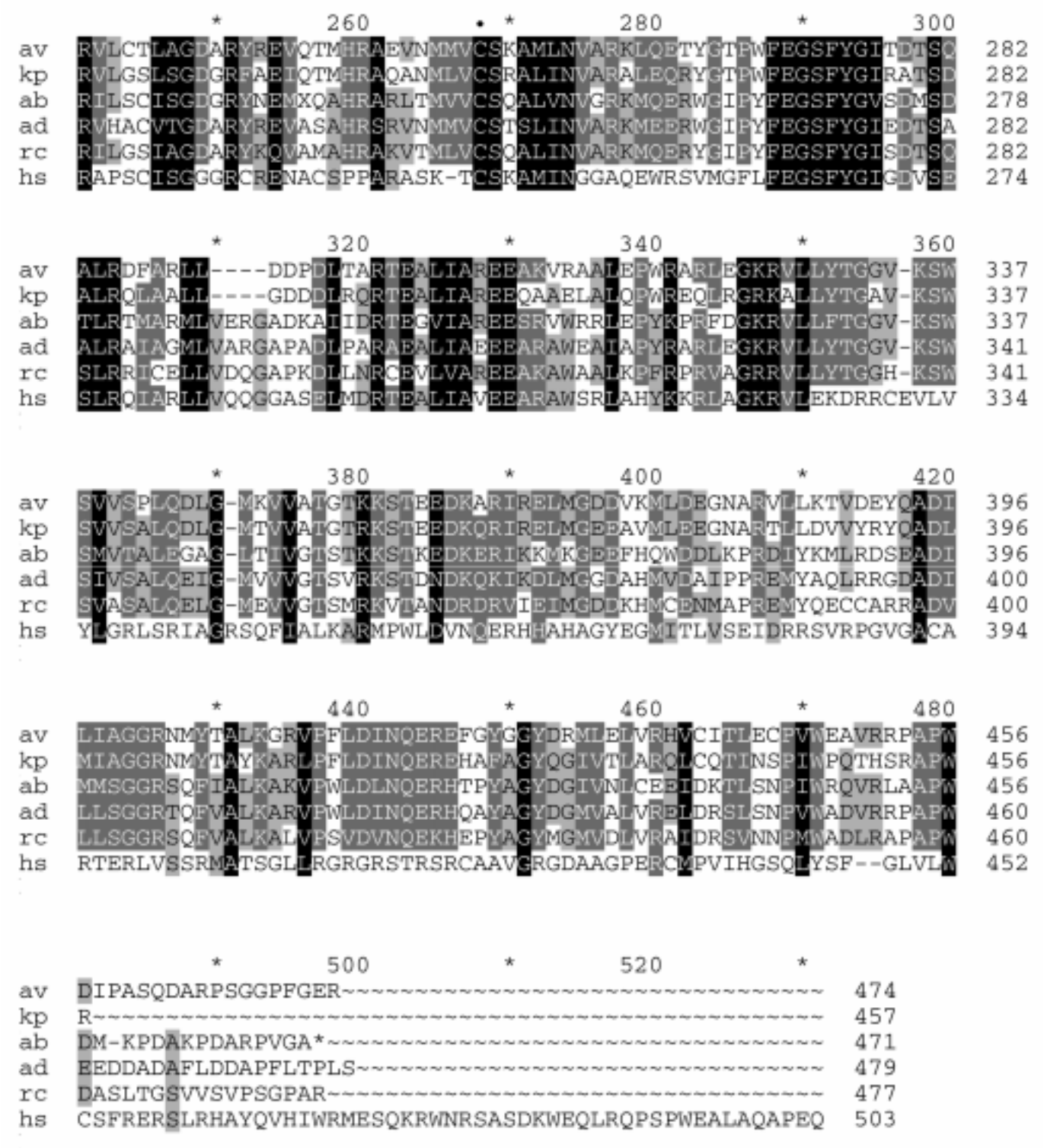
Comparison of the predicted amino acid sequences of the *Azospirillum brasilense* (ab) NifE (I), NifN (II), NifX (III) and NifQ (IV) proteins with analogous gene products from *Azotobacter vinelandii* (av) (4), *Klebsiella pneumoniae* (kp) (37,38), *Gluconacetobacter diazotrophicus* (ad) (9), *Rhodobacter capsulatus* (rc) (16), *Bradyrhizobium japonicum* (bj) (31), *Herbaspirillum seropedicae* (hs) (10), *Rhizobium* sp (rsp) (36), and *Enterobacter agglomerans* (ea) (35). A black background indicates conserved residues in all aligned sequences, dark grey indicates conserved residues in at least 80% of the aligned sequences, and light grey indicates conserved residues in at least 60% of the aligned sequences. Conserved cysteine residues are indicated by black dots. Multiple alignment was done using the PILEUP program, University of Wisconsin Genetics Computer Group. The alignment editing was done using the GeneDoc program and the Dayhoff PAM 250 score table (39).

APPENDIX 1(I)



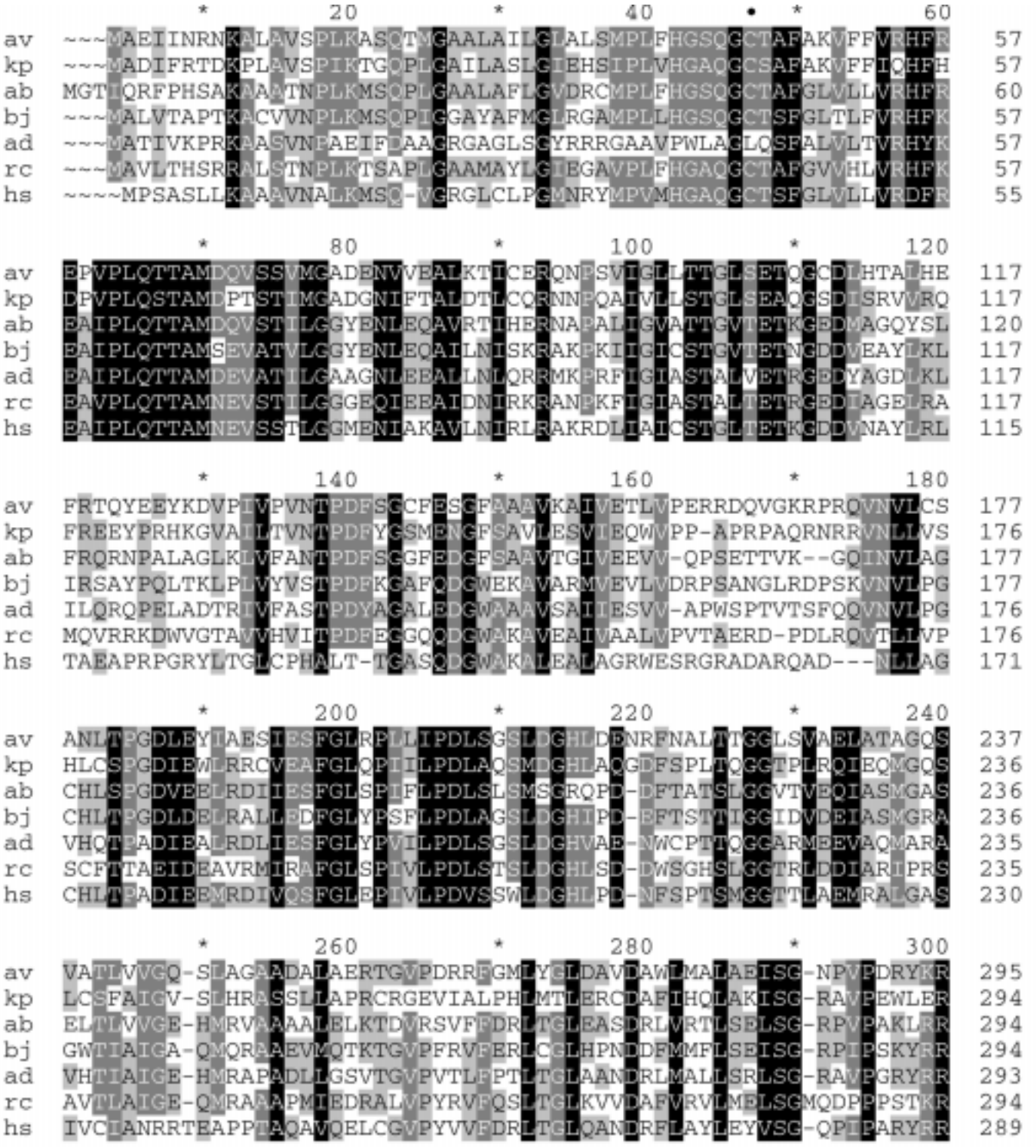
Appendix 1(I) continued on next page

Appendix 1(I) continued



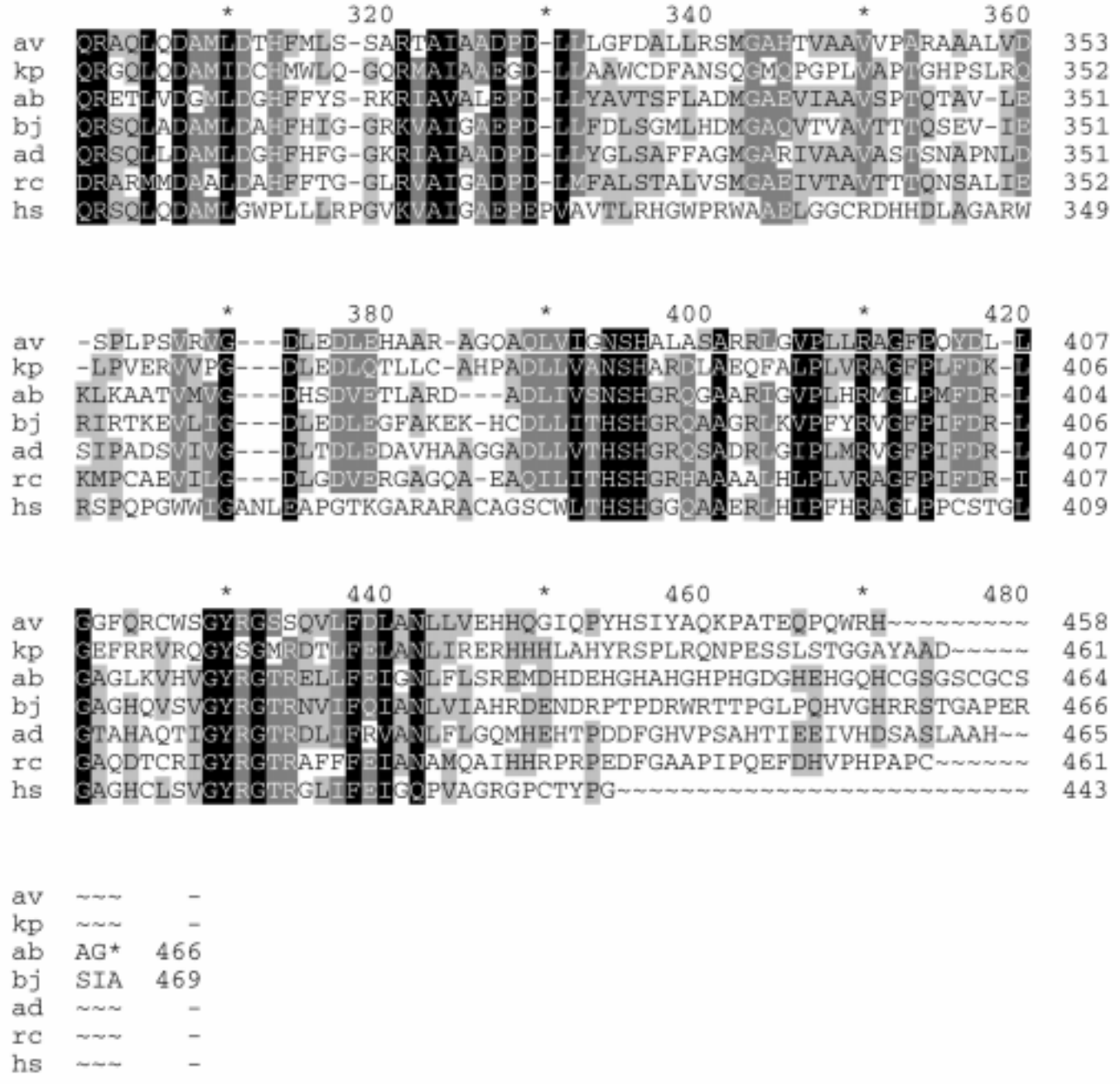
Appendix 1 continued on next page

APPENDIX 1(II)



Appendix 1(II) continued on next page

Appendix 1(II) continued



Appendix 1 continued on next page

APPENDIX 1(III)

```
                 *        20         *        40         *        60
ab  ~~MRMQRRLSVVGQAEEGRPRKGGSMKVAFCTQDMQQHVDAHFGWAKNIVVYEVDKAGYT   58
hs  ~~~~~~~~~~~~MPWRWPRPFKEKAMKVAFATQELQR-VEAHFGWAKNLAVFELWPNGYS   47
ad  MTARRLQLIEPEAGGDGAAAGVVPRLRIAIATQDMK-ALNAHFGSARRFAVWDVTPDDAH   59
rc  ~MSRTLRLVEPA----GPAPGEKP-LRVAIASNDLE-NLDAHFGSARQIAVYEVWKTGAR   53
av  ~~~~~~MSSPTRQLQVLDSEDDGTLLKVAFASSD-RELVDQHFGSSRSFAIYGVNPERSQ   53
kp  ~~~~~~~~MPPINRQFDMVHSDEWSMKVAFASSD-YRHVDQHFGATPRLVVYGVKADRVT   51

                 *        80         *    •  100         *       120
ab  MVETCQFGGSMFEDGN-----EDKLIPKLDALADCAIVYLSAIGASAAARVVAKKIHPVK  113
hs  FVQTHSFDGDLKEDGD-----EDKLAPKIEAIKECAILYVAAIGGSGAARVVANRIHPVK  102
ad  FVEAVGFDDVSDESGAHKVDVDDRIGPKVDALAGCNLLFVLAIGGPAAAKVVGAHIHPVK  119
rc  FVEVHQFSSATDQKGRHD-ELEDRIGPKLEALSGCTLVFALAVGGPSAARMVRAGMHPIK  112
av  LLSVVEFGEL------EQDGNEDKLARKIDLLDGCVAVYCCACGASAVRQLMAIGVQPIK  107
kp  LIRVVDF-SV------ENGHQTEKIARRIHALEDCVTLFCVAIGDAVFRQLLQVGVRAER  104

                 *       140         *       160         *
ab  VEATETITALLDRLVETINGNPPPWLRKALNAGQPQELAFDEED*~~~~~~~~~~~~~  157
hs  VAQAEPILDILDKLQEVLKGTPAPWLRKAMQKGQERVINFEEEV~~~~~~~~~~~~~~  146
ad  LPAPQSIASVIERVQTMMKGNPPPWLRKVMGAAVPRSMDFLDEED~~~~~~~~~~~~~  164
rc  RKEPEPISAVIEQVQVMLNGTPPPFLRKVLGTWEKPDFTADFEEEEV~~~~~~~~~~~  159
av  VSEGARIAELIEALQVELREGPSAWLAKAIQRTRGPDMRRFDAMAAEGWDE~~~~~~~  158
kp  VPADTTIVGLLQEIQLYWYDKGQR------KNTRQRDPERFTRLLQEQEWHGDPDPRR  156
```

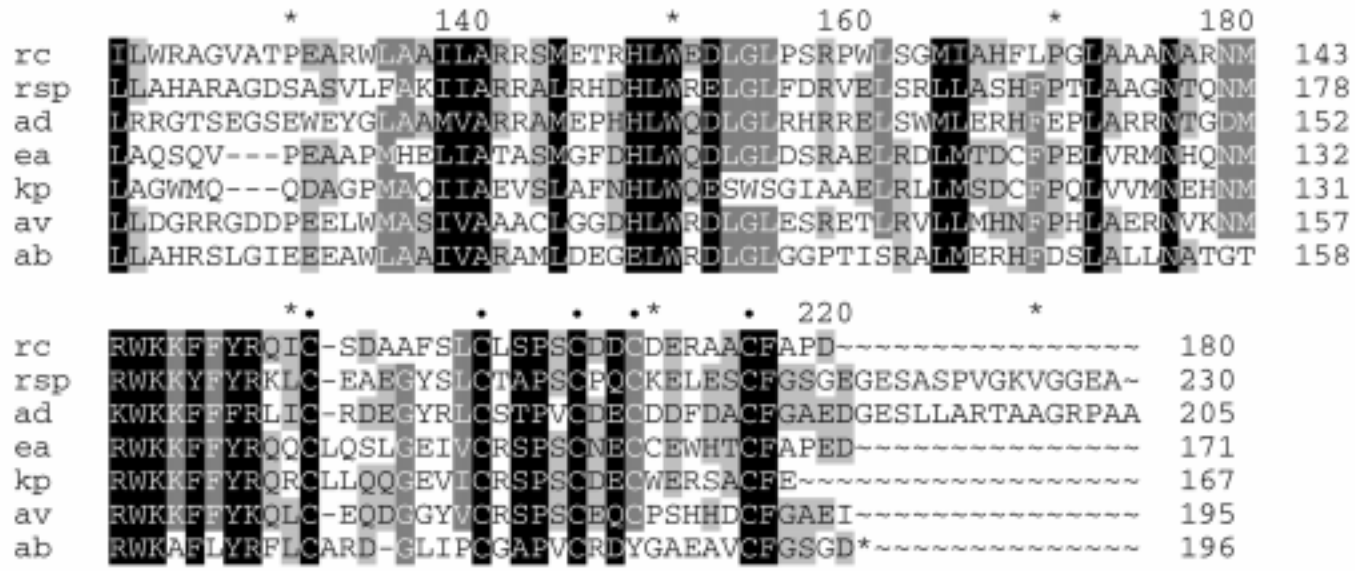
APPENDIX 1(IV)

```
                 *        20         *        40         *        60
rc  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MLHFPDPFATGPGPV-IPAEALGFILAQ   27
rsp MSNLAQVRALSKGRVTLGIRMTDRPGWRQLSELLDLGPWPPTDLEMD-FDQYVFACVLSR   59
ad  ~~~~~~~~~~~~~~~~~~~~~~~~~~~MRAEDLHAWLMAQGTGTECDRFDVHVLASILAI   33
ea  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGAQGWLSRLLSLHLTGRSR   21
kp  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MPPLDWLRRLWLLYHAGKGS   20
av  ~~~~~~~~~~~~~~~~~~~~~~MGSAAAHRGDTTQAVRHDRANHLWLERIVRSQRDGLSC   38
ab  ~~~~~~~~~~~~~~~~~~~~~~MGILHAAPPGAGDTTRLYRWLTDRQGRSNVFDAHLFAC   38

                 *        80         *       100         *       120
rc  GLRECAAGLGPLTARLGLSGADLAALRDRF--APGLELPDLD-LPR-PEAGPDQQAIETL   83
rsp ALEEIDAGEATATEATGLSQVELRDILNRSFPAPTIHVFRLE-EVRDSEPGPEEALLRGL  118
ad  ALIQSRERGLPLPGLVGLGGTDFVALVGAMLPGALSRFQTMA-DLPAPVPDENESILRDL   92
ea  FPPQMGLGDVAWQALLQH-----TGRAAPVLSTLQFEQQKLL-GLLQQARTCEREQLAQW   75
kp  FPLRMGLSPRDWQALRRR-----LGEVETPLDGETLTRRRLM-AELNATREEERQQLGAW   74
av  LPFHLGLDERSYAELIRTHFPELAGQTSASLGSLAHECSELR-EDLLEMRRDEWEELRVL   97
ab  ILSRRWSAGPGALGLDDRALGQLLDRYFPGAFAAGLPVPDSSPTPLPPLLRSEADDIATL   98
```

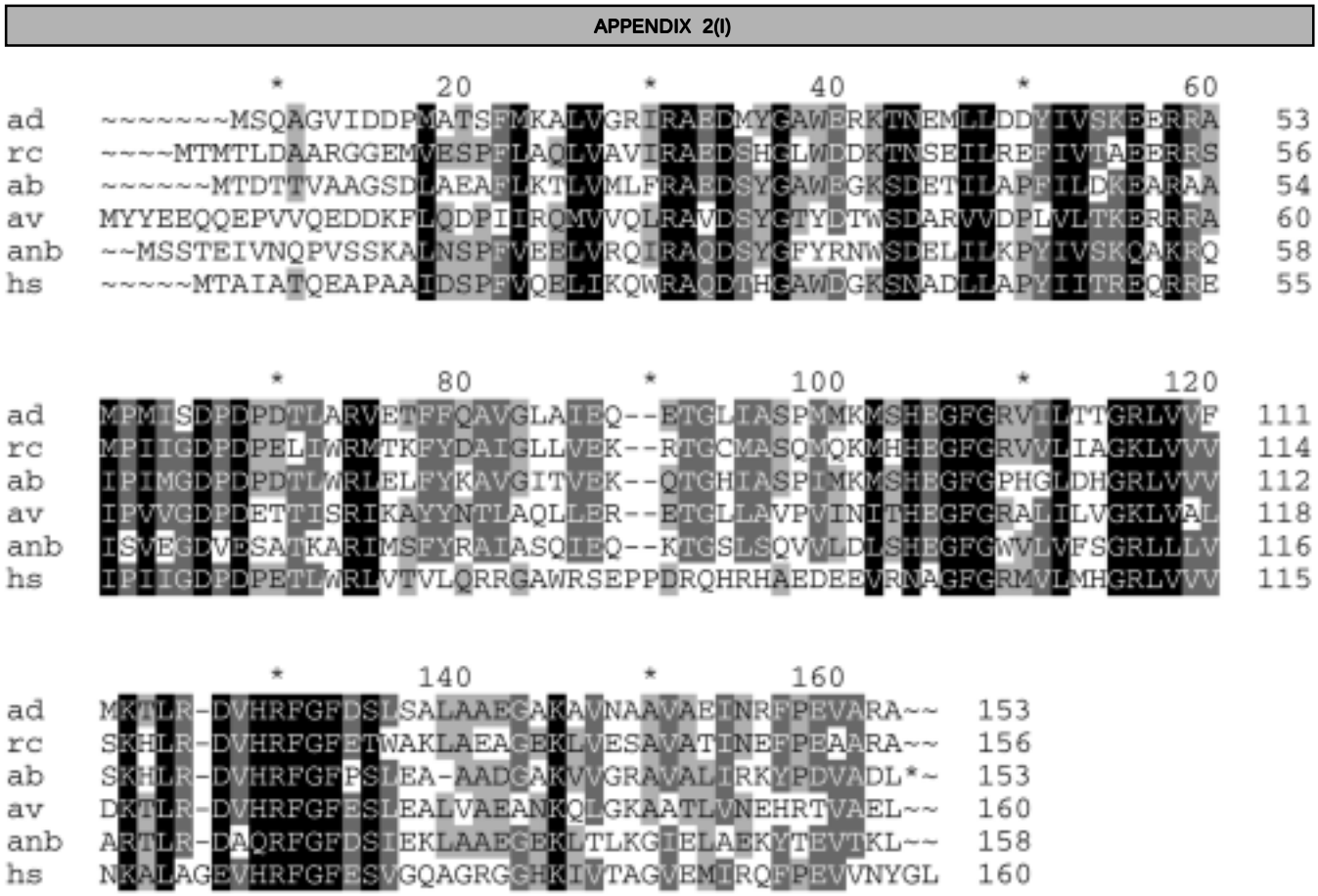

Appendix 1(IV) continued on next page

Appendix 1(IV) continued



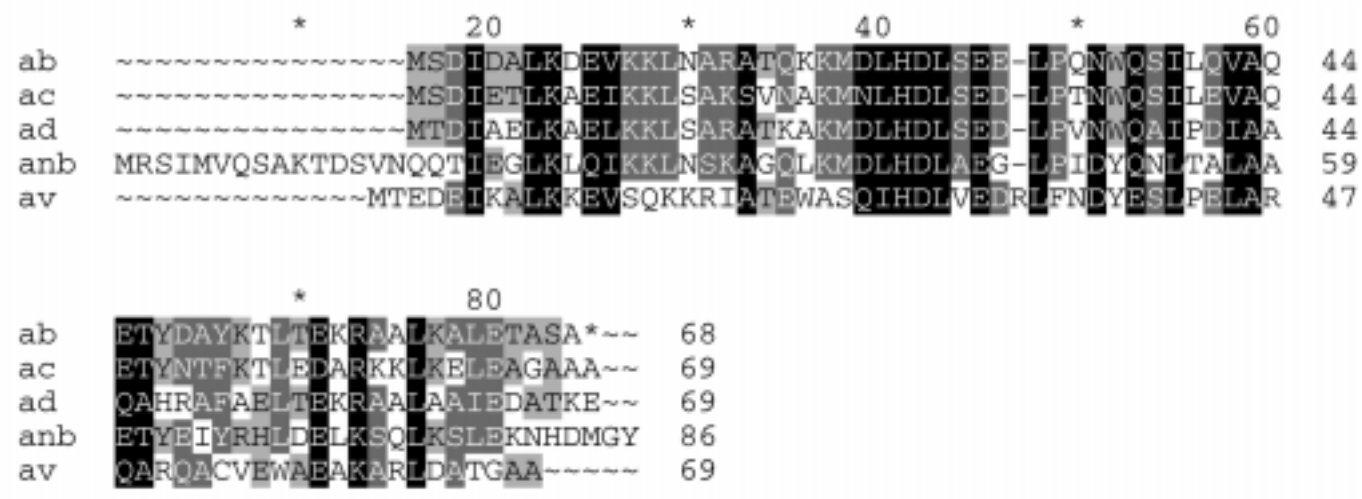
Appendix 2

Comparison of the predicted amino acid sequences of the *Azospirillum brasilense* (ab) Orf3 (I), Orf5 (II) and FdxA (III) proteins with analogous gene products from *Azotobacter vinelandii* (av) (3), *Anabaena variabilis* (anb) (40), *Plectonema boryanum* (pb) (33), *Gluconacetobacter diazotrophicus* (ad) (9), *Rhodobacter capsulatus* (rc) (16), *Azorhizobium caulinodans* (ac) (32) *Rhizobium* sp NGR234 (rsp) (36), and *Herbaspirillum seropedicae* (hs) (10). A black background indicates conserved residues in all aligned sequences, dark grey indicates conserved residues in at least 80% of the aligned sequences, and light grey indicates conserved residues in at least 60% of the aligned sequences. Multiple alignment was done using the PILEUP program, University of Wisconsin Genetics Computer Group. The alignment editing was done using the GeneDoc program and the Dayhoff PAM 250 score table (39).



Appendix 2 continued on next page

APPENDIX 2(II)



APPENDIX 2(III)