Short sequence-paper

Sequence analysis of an internal 9.72-kb segment from the 30-kb denitrification gene cluster of *Pseudomonas stutzeri* [1]

Andrea B. Glockner, Walter G. Zumft *

Lehrstuhl für Mikrobiologie, Universität Fridericiana, Kaiserstrasse 12, D-76128 Karlsruhe, Germany

Abstract

The DNA segment was sequenced that links the *nir-nor* and *nos* gene clusters for denitrification of *Pseudomonas stutzeri* ATCC 14405. Of 10 predicted gene products, four are putative membrane proteins. Sequence similarity was detected with the subunit III of cytochrome-*c* oxidase (ORF175), PQQ3 of the biosynthetic pathway for pyrrolo-quinoline quinone (ORF393), *S*-adenosylmethionine-dependent uroporphyrinogen-III C-methyltransferase (ORF278), the cytochrome cd_1 nitrite reductase and the NirF protein involved in the biosynthesis of heme d_1 (ORF507), LysR type transcriptional regulators (ORF286), short-chain alcohol dehydrogenases (ORF247), and a hypothetical protein, YBEC, of *Escherichia coli* (ORF57). The current data together with previous work establish a contiguous DNA sequence of 29.2 kb comprising the supercluster of *nos-nir-nor* genes for denitrification in this bacterium.

A denitrification gene cluster extends in *P. stutzeri* over ≈ 30 kb and encodes multiple functions for the respiration of nitrite, nitrogen monoxide (NO), and nitrous oxide (N_2O). The structural genes for cytochrome cd_1 nitrite reductase, *nirS*, and NO reductase, *norCB*, together with functions for electron donation, regulatory components and enzyme processing have been found to be closely linked on the chromosome [1,2]. The *nos* region, with the structural gene for N_2O reductase, *nosZ*, and genes encoding copper-processing proteins, *nosDFY*, is ≈ 10 kb

distant from the cluster of *nir* and *nor* genes. The intergenic region, anticipated to harbor further genes for denitrification, was sequenced and analyzed. Preliminary accounts of this work have appeared [3,4].

The 10062 bp *Hind*III fragment, carrying at its extremes the 3' ends of the genes *nirQ* [5] and *nosL* [6], was isolated from cosmid cDEN1 [2]. *nirQ* affects both nitrite reductase and NO reductase, and *nosL* encodes a putative protein disulfide isomerase acting on NosZ in the periplasm. The *Hind*III fragment was cloned into pBR322 to give plasmid pBRH97. Sequencing of subfragments of pBRH97 cloned into pUC18 was done by the dideoxy chain-termination method using the femtomol sequencing kit (Promega) and [^{33}P]dATP. The nucleotide sequence was determined with sequence-specific primers on both strands. The sequence was analyzed

---

* Corresponding author. Fax: +49 721 694172/6084290; e-mail: zumft@mbkSUN1.bio-geo.uni-karlsruhe.de.

PII S0005-2728(96)00108-9

Table 1
Coding regions and properties of the derived gene products

Gene designation [a]	Coding region [b]	Position of RBS [c]	M_r of gene product	Specific signature of derived protein and position in nucleotide sequence	Homologous protein(s) by BLAST search [44]	Reference(s)
orf175	8-535	1118-1121 [d]	19485	5 predicted transmembrane segments	COXIII, ORF2_PSEAE, NORE_PARDE	[8-10]
orf82	546-794	533-541	8523	3 predicted transmembrane segments	ORF3_PSEAE	[8]
orf393	964-2145	953-956	44411		PQQ3_ACICA, PQQ_ERWHE, PQQE_KLEPN	[18-20]
orf278	2216-3052	2203-2209	29577	SUMT [e] motifs I: 2282-2326, II: 2507-2608	CYSG_ECOLI, SUMT_BACME and PSEDE, NIRE_PARDE	[22,24,25,28]
orf507	3043-4566	3034-3038	55482	Heme c-binding site, 3175-3189	NIRF_PSEST, NIRS_PSEST, NIRC_PSEST	[29,45]
orf286	4669-5529	4658-4665	32700	DNA-binding site, HTH [e], 4720-4797	Regulators of the LysR family	[34]
orf378 [f]	5539-6675	6680-6686	41794	5 predicted transmembrane segments		
orf247 [f]	7055-7798	7803-7808	25840	SCAD [e] consensus motif, 7325-7357	NODG_RHIME, ENTA_ECOLI, FABG_HAEIN, BNZE_PSEPU	[46-49]
orf396 [f]	7923-9113	9121-9128	43051	12 predicted transmembrane segments	YTFF_ECOLI	[50]
orf57 [f]	9433-9606	9612-9617	6082		YBEC_ECOLI, YIGT_HAEIN	[37,48]

[a] Designation according to number of amino acids of the derived gene product.
[b] Includes stop codon.
[c] Ribosome-binding site.
[d] Located at the indicated positions within the previously published nirQ sequence [5].
[e] For explanation see text.
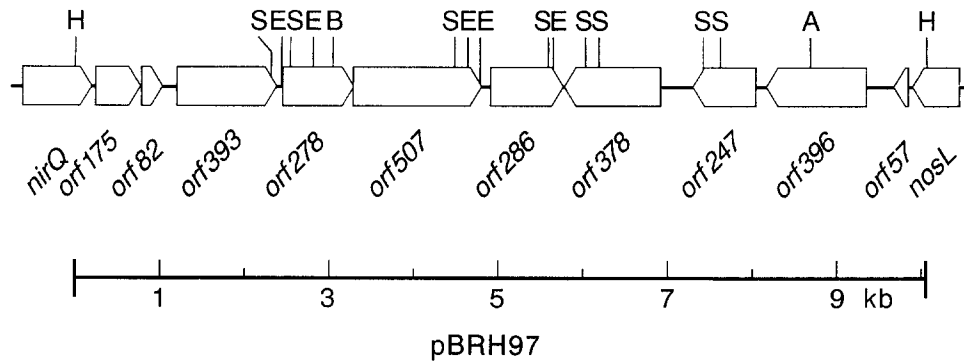[f] Located on the complementary strand.

Fig. 1. Physical map and open reading frames of the intergenic DNA region between the *nir* and *nos* clusters of *P. stutzeri*. The transcriptional direction is indicated by open arrows. The pBR322 derivative, pBRH97, was used for subcloning and sequencing. Restriction sites: A, *Acc*65I; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; S, *Sma*I.

with the software packages HUSAR, Deutsches Krebsforschungszentrum, Heidelberg, and PC/GENE, IntelliGenetics, Mountain View, CA. Signatures of distinct protein families were scanned using the PROSITE dictionary [7].

Fig. 1 shows the restriction map of the sequenced region with the location and transcriptional direction of the identified open reading frames (ORF). The previously reported sequences for *nirQ* (accession no. Z17423) and *nosL* (accession no. Z69589) overlap with the first 31 and last 106 nucleotides of the current sequence, respectively. The coding strands were identified from the codon preference of *P. stutzeri* genes known to encode components for denitrification (Fig. 2). Overlap of consecutive ORFs was not considered when no acceptable ribosome-binding site was found for the start codon further upstream, and the codon preference was atypical for a pseudomonadal gene. ORFs thus obtained and features of the derived gene products are listed in Table 1.

The derived protein of *orf175* is hydrophobic with five predicted membrane-spanning domains (Fig. 3) and is homologous with the ORF2 protein from *Pseudomonas aeruginosa* (66.9% identity) [8] and NorE from *Paracoccus denitrificans* (44.6% identity) [9]. These proteins have sequence similarity with the C-terminal part covering helices 3 to 7 of subunit III (COXIII) of *P. denitrificans* cytochrome-*c* oxidase

[10]. The similarity is strongest for helices 6 and 7. From the crystal structure of COXIII from *Paracoccus* it was established that the seven membrane-spanning helices of this protein are divided into two bundles, one formed by the first two helices, and the other by helices 3 to 7 [11]. The domain structure of the ORF175 protein and its homologues has been proposed on the basis of the COXIII structure to consist in a 5-helix bundle [10]. Since it was noted earlier that NO reductase may be an ancient member of the family of heme-copper oxidases [12,13], it is tempting to assume that ORF175 and its homologues ORF2 and NorE have a role analogous to COXIII. COXIII is thought to be required for the assembly and/or stability of the oxidase [14,15], but a role in energy conservation has recently been reconsidered [16].

orf82 has the coding potential for a hydrophobic protein with three predicted membrane-spanning domains (Fig. 3). It is 52.4% identical with a hypothetical protein, ORF3, from *P. aeruginosa* [8] but has no further similarity to current entries in the SWISS-PROT data bank.

The derived protein of *orf393* is hydrophilic without any specific signature. Sequence alignment by BESTFIT [17] showed highest similarity (50.3%) with the PQQ3 protein of *Acinetobacter calcoaceticus* [18], and 46.9 and 45.4% similarity with the homolo-

Fig. 2. Codon preference plot and frequency of rare codons for the 9.72-kb DNA segment. The analysis is shown for the three possible phases in both directions (A and B) with ATG and GTG as the start codons; window size 50 [43]. Numbers identify the ORFs described in the text and in Table 1.
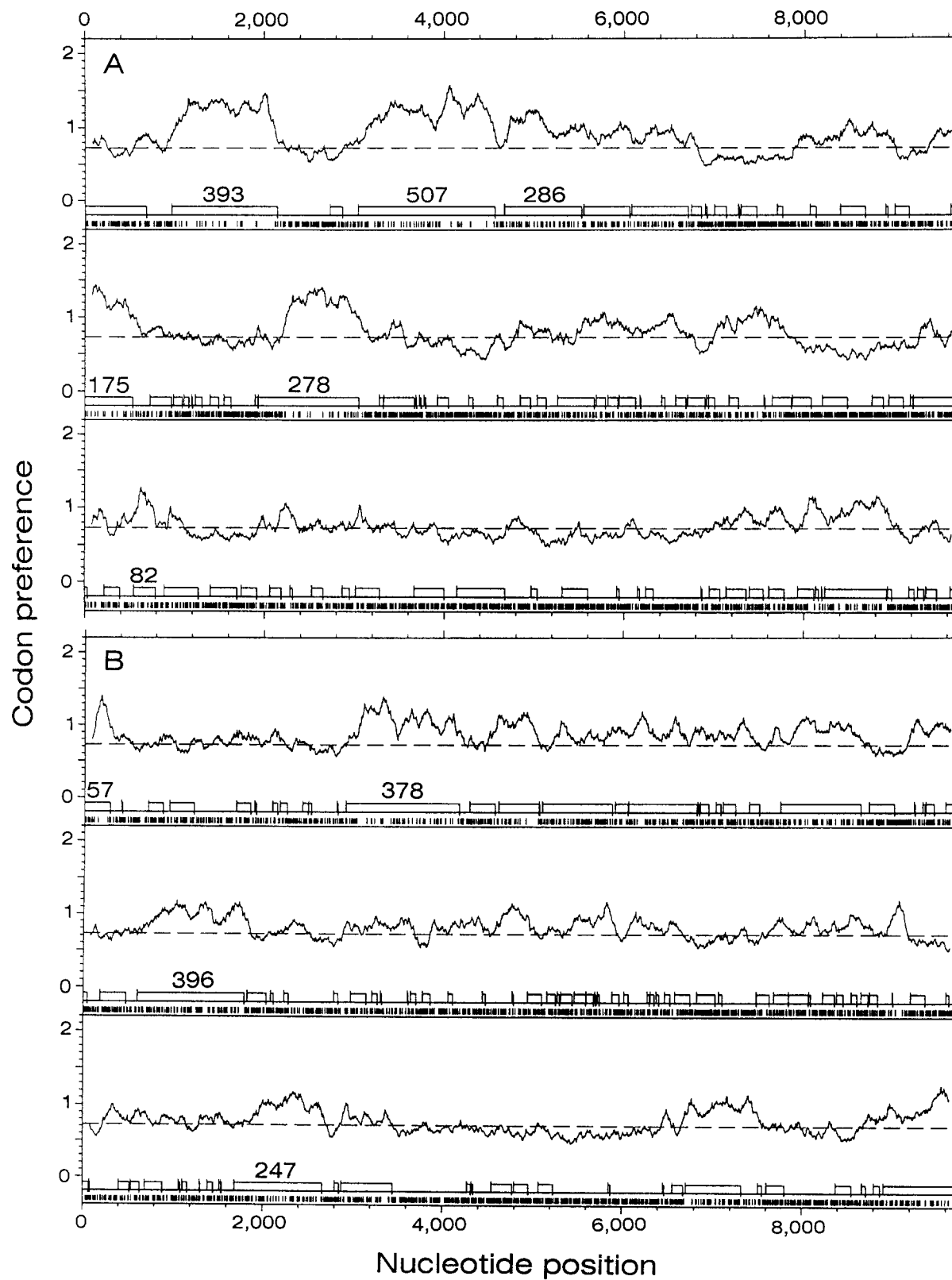
gous PQQ and PQQE proteins from *Erwinia her-bicola* [19] and *Klebsiella pneumoniae* [20], respectively (identity range 20.7–23.8%). The highest score was observed with the sequence 83-SGGEPLLRAD which exists in slightly modified forms in all four proteins. PQQ3 is involved in the biosynthesis of the pyrrolo-quinoline quinone coenzyme, but its precise function is still unknown (see Refs. in Ref. [21]).

orf278 encodes a hydrophilic protein with homology to the siroheme synthase of *E. coli* [22] and *Salmonella typhimurium* [23] and to the S-adenosyl-L-methionine-dependent uroporphyrinogen-III C-methyltransferase (SUMT, EC 2.1.1.107)) of *Bacillus megaterium* [24], *Pseudomonas denitrificans* [25], *Methanobacterium ivanovii* [26], and other bacteria. These proteins belong to a family that groups SUMT and CysG with methyltransferases of the cobalamine biosynthetic pathway and share sequence motifs important for substrate-binding [27]. The NirE protein of *P. denitrificans*, which was proposed to be another member of the SUMT protein family [28], is 43.3% identical with the *orf278* product. Mutagenesis of *nirE* leads to a heme d_1-free nitrite reductase. The ORF278/NirE proteins are likely to transform uroporphyrinogen-III to precorrin-II in the biosynthetic pathway of heme d_1 [28,29]. Canonical recognition half-sites for the transcription factor FNR begin 74 bp (TTGAT) and 55 bp (ATCAA) upstream of the ATG start codon. They are spaced versus the usual 4 nucleotides by exactly one additional turn in the DNA helix and require investigation whether they are involved in anaerobic regulation of this gene by binding a trancription factor.

orf507 overlaps *orf278* by 10 bp. The deduced ORF507 protein is hydrophilic except for a hydrophobic segment from positions 3-19. The protein may have an export signal sequence of 30 amino acids with a cleavage site in accordance with the − 3, − 1 rule [30]. ORF507 shows 30.5% positionally identical residues with NirF from *P. stutzeri* (overall similarity 53.4%). NirF, together with several other proteins, is necessary for the biosynthesis of heme d_1 [28,29,31]. ORF507 has also 26.0% identical residues (52.0% similarity) with NirS, the tetraheme nitrite reductase from *P. stutzeri*, and has a binding site for heme c, 45-CQSCH. The heme-binding domain of ORF507 is in turn related to NirC (31.8% identity, 50.4% similarity), a c-type cytochrome with a puta-

tive role in heme d_1 synthesis or the maturation of cytochrome cd_1 [28,32]. The similarity of ORF507 with NirF and NirS may have its origin in a gene duplication and is indicative of a Nir-related function for ORF507. We have previously pointed out several gene duplication events in the *nirD* region of *P. stutzeri*, which encodes proteins required for heme d_1 biosynthesis [29].

The deduced protein of *orf286* carries the consensus motif of a transcriptional regulator belonging to the LysR family [33]. These regulators are hydrophilic proteins of about 300 amino acids with a helix-turn-helix (HTH) motif in the N-terminal domain. They do not share notable structure outside of this domain [34]. LysR type regulators often affect genes transcribed in the opposite direction whose promoters overlap that of the regulator gene, but other members also regulate the transcription of groups of genes at a distant location [35]. The potential regulatory function of ORF286 and identification of its target gene(s) is currently under investigation.

The subsequent four ORFs are oriented in the opposite transcriptional direction. *orf378* encodes a protein composed of a hydrophobic domain with five membrane-spanning segments and a hydrophilic domain of about the same size (Fig. 3). The deduced protein has no notable similarity with known bacterial proteins. Codon usage and presence of a ribosome-binding site, 5′-AAGcAGG-3′, suggests translation of this ORF.

orf247 encodes a hydrophilic protein. We chose the third of three contiguous ATG triplets as start codon since it is preceded by a ribosome-binding site, 5′-GGAGaT-3′, at the near optimal distance of four nucleotides. ORF247 shows similarity to numerous members of the short-chain alcohol dehydrogenase (SCAD) family [36]. Proteins of this family have limited homology among each other but share structural features around two conserved sites. The consensus motif of ORF247 for binding of the NAD(P) coenzyme, 14-GASSGIG, resides within an N-terminal hydrophobic stretch of amino acids. The active site motif of ORF247, 148-Y*I*ASKAALAA*Q* (where Y and K are essential), deviates in the two underlined positions from the consensus [36]. Examples of related proteins are shown in Table 1.

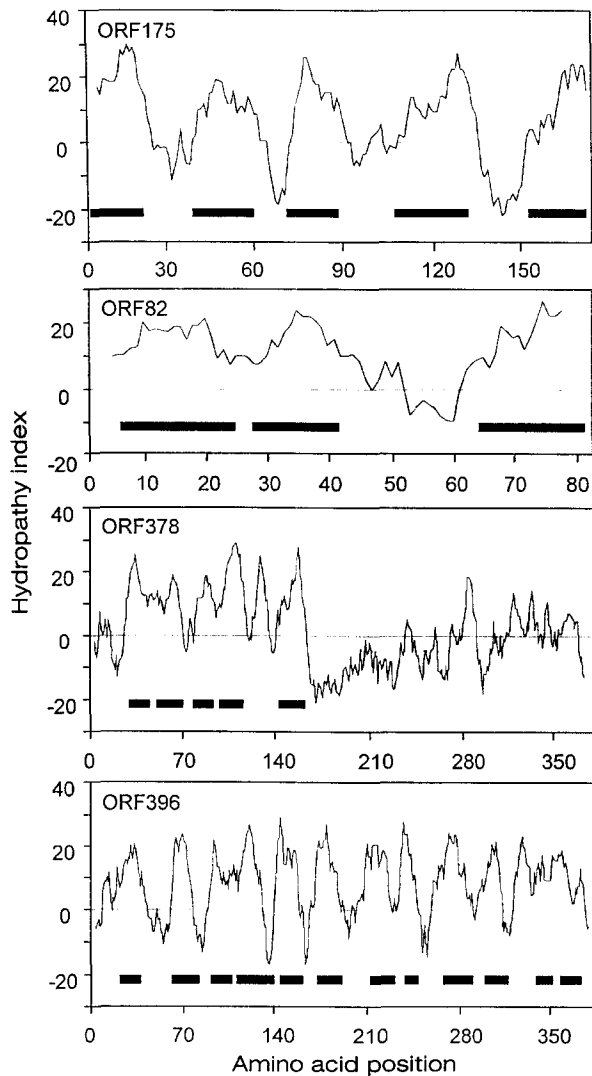The deduced protein of *orf396* is highly hydrophobic with 12 membrane-spanning segments (Fig.

Fig. 3. Hydropathy (Kyte-Doolittle) plots of putative membrane proteins. The location of transmembrane segments is indicated by black bars. Window size 9 residues.

role of YBEC is unknown; the gene is located downstream of *lipA* which is involved in the biosynthesis of lipoic acid [37].

The data described here and our previous work result in a contiguous DNA sequence of 29.2 kb that encodes essential components for three steps of the denitrification process of *P. stutzeri*. A total of 30 genes have been identified within this segment. Most of them are transcribed in the same direction, with only 7 contiguous genes, *nirQ* through *orf286*, organized between *nirS* and *orf378* in the opposite direction. Further genes for denitrification are located downstream of *norCB*, among them *fnrD*, encoding the previously postulated FNR-like transcriptional regulator [38]. The genes necessary for nitrate respiration (*nar*) are located elsewhere on the chromosome (unpublished data). Gene clusters comprising functions both for respiratory nitrite reduction (*nir*) and NO reduction (*nor*) exist in *P. denitrificans* [9,28] and in *P. aeruginosa* [39]. *P. stutzeri* appears to be distinct having the *nos* genes as part of a *nos-nir-nor* supercluster, whereas they are separate in *P. aeruginosa* [40] and located on a plasmid in *Alcaligenes eutrophus* [41] and *Rhizobium meliloti* [42]. The novel genes identified here are likely to encode ancillary functions required for nitrite and NO respiration as deduced from similarity with known proteins. These data now provide a basis to study the gene products and regulatory circuits for their coordinate function.

3). The likely start codon of this ORF is GTG which is preceded by a ribosome-binding site, 5'-TttG-GAGG-3'. Similarity was found with the hypothetical YTFF protein from *E. coli* of unknown function.

ORF57 is homologous to the small hypothetical *E. coli* protein, YBEC. The two proteins share 47.2% positional identity. *orf57* has a lower GC content (60.8%) than other genes of *P. stutzeri* (average 64.4%); however, the similarity of its product with YBEC and a ribosome-binding site, 5'-TAAGcA-3', are indicative for its coding function. The N-terminal half of both predicted proteins is hydrophobic. The

References

[1] Jüngst, A., Braun, C. and Zumft, W.G. (1991) Mol. Gen. Genet. 225, 241–248.
[2] Braun, C. and Zumft, W.G. (1992) J. Bacteriol. 174, 2394–2397.
[3] Zumft, W.G. and Körner, H. (1995) in Beijerinck Centennial. Microbial Physiology and Gene Regulation: Emerging Principles and Applications (Scheffers, W.A. and Van Dijken, J.P., eds.), pp. 122–123, Delft University Press, Delft.
[4] Glockner, A.B. and Zumft, W.G. (1996) Biospectrum 1996, 70.

[5] Jüngst, A. and Zumft, W.G. (1992) FEBS Lett. 314, 308–314.

[6] Dreusch, A., Riester, J., Kroneck, P.M.H. and Zumft, W.G. (1996) Eur. J. Biochem. 237, 447–453.

[7] Bairoch, A. (1993) Nucleic Acids Res. 21, 3097–3103.

[8] Arai, H., Igarashi, Y. and Kodama, T. (1994) Biosci. Biotechnol. Biochem. 58, 1286–1291.

[9] De Boer, A.P.N., Reijnders, W.N.M. and Stouthamer, A.H. (1995) EMBL database, accession no. U28078.

[10] Zumft, W.G. and Körner, H. (1996) Antonie Leeuwenhoek, in press.

[11] Iwata, S., Ostermeier, C., Ludwig, B. and Michel, H. (1995) Nature 376, 660–669.

[12] Van der Oost, J., De Boer, A.P.N., De Gier, J.-W.L., Zumft, W.G., Stouthamer, A.H. and Van Spanning, R.J.M. (1994) FEMS Microbiol. Lett. 121, 1–10.

[13] Saraste, M. and Castresana, J. (1994) FEBS Lett. 341, 1–4.

[14] Haltia, T., Saraste, M. and Wikström, M. (1991) EMBO J. 10, 2015–2021.

[15] Echabe, I., Haltia, T., Freire, E., Goñi, F.M. and Arrondo, J.L.R. (1995) Biochemistry 34, 13565–13569.

[16] Wu, S., Sanchez-Moreno, R. and Rottenberg, H. (1995) Biochemistry 34, 16298–16305.

[17] Smith, T.F. and Waterman, M.S. (1981) J. Mol. Biol. 147, 195–197.

[18] Goosen, N., Horsman, H.P.A., Huinen, R.G.M. and Van de Putte, P. (1989) J. Bacteriol. 171, 447–455.

[19] Liu, S.-T., Lee, L.-Y., Tai, C.-Y., Hung, C.-H., Chang, Y.-S., Wolfram, J.H., Rogers, R. and Goldstein, A.H. (1992) J. Bacteriol. 174, 5814–5819.

[20] Meulenberg, J.J.M., Sellink, E., Riegman, N.H. and Postma, P.W. (1992) Mol. Gen. Genet. 232, 284–294.

[21] Springer, A.L., Ramamoorthi, R. and Lidstrom, M.E. (1996) J. Bacteriol. 178, 2154–2157.

[22] Peakman, T., Crouzet, J., Mayaux, J.F., Busby, S., Mohan, S., Harborne, N., Wooton, J., Nicolson, R. and Cole, J. (1990) Eur. J. Biochem. 191, 315–323.

[23] Wu, J.-Y., Siegel, L.M. and Kredich, N.M. (1991) J. Bacteriol. 173, 325–333.

[24] Robin, C., Blanche, F., Cauchois, L., Cameron, B., Couder, M. and Crouzet, J. (1991) J. Bacteriol. 173, 4893–4896.

[25] Crouzet, J., Cauchois, L., Blanche, F., Debussche, L., Thibaut, D., Rouyez, M.-C., Rigault, S., Mayaux, J.-F. and Cameron, B. (1990) J. Bacteriol. 172, 5968–5979.

[26] Blanche, F., Robin, C., Couder, M., Faucher, D., Cauchois, L., Cameron, B. and Crouzet, J. (1991) J. Bacteriol. 173, 4637–4645.

[27] Kagan, R.M. and Clarke, S. (1994) Arch. Biochem. Biophys. 310, 417–427.

[28] De Boer, A.P.N., Reijnders, W.N.M., Kuenen, J.G., Stouthamer, A.H. and Van Spanning, R.J.M. (1994) Antonie Leeuwenhoek 66, 111–127.

[29] Palmedo, G., Seither, P., Körner, H., Matthews, J.C., Burkhalter, R.S., Timkovich, R. and Zumft, W.G. (1995) Eur. J. Biochem. 232, 737–746.

[30] von Heijne, G. (1986) Nucleic Acids Res. 14, 4638–4690.

[31] Kawasaki, S., Arai, H., Igarashi, Y. and Kodama, T. (1995) Gene 167, 87–91.

[32] Ye, R.W., Arunakumari, A., Averill, B.A. and Tiedje, J.M. (1992) J. Bacteriol. 174, 2560–2564.

[33] Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988) Proc. Natl. Acad. Sci. USA 85, 6602–6606.

[34] Schell, M.A. (1993) Annu. Rev. Microbiol. 47, 597–626.

[35] Schell, M.A. and Sukordhaman, M. (1989) J. Bacteriol. 171, 1952–1959.

[36] Persson, B., Krook, M. and Jörnvall, H. (1991) Eur. J. Biochem. 200, 537–543.

[37] Hayden, M.A., Huang, I., Bussiere, D.E. and Ashley, G.W. (1992) J. Biol. Chem. 267, 9512–9515.

[38] Cuypers, H. and Zumft, W.G. (1993) J. Bacteriol. 175, 7236–7246.

[39] Arai, H., Igarashi, Y. and Kodama, T. (1995) FEBS Lett. 371, 73–76.

[40] Vollack, K.-U., Römling, U. and Zumft, W.G. (1996) Biospectrum 1996, 70.

[41] Zumft, W.G., Dreusch, A., Löchelt, S., Cuypers, H., Friedrich, B. and Schneider, B. (1992) Eur. J. Biochem. 208, 31–40.

[42] Holloway, P., McCormick, W., Watson, R.J. and Chan, Y.-K. (1996) J. Bacteriol. 178, 1505–1514.

[43] Gribskov, M., Devereux, J. and Burgess, R.R. (1984) Nucleic Acids Res. 12, 539–549.

[44] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.

[45] Jüngst, A., Wakabayashi, S., Matsubara, H. and Zumft, W.G. (1991) FEBS Lett. 279, 205–209.

[46] Debellé, F. and Sharma, S.B. (1986) Nucleic Acids Res. 14, 7453–7472.

[47] Liu, J., Duncan, K. and Walsh, C.T. (1989) J. Bacteriol. 171, 791–798.

[48] Fleischmann, R.D. and Venter, J.C. (1995) Science 269, 496–512.

[49] Irie, S., Doi, S., Yorifuji, T., Takagi, M. and Yano, K. (1987) J. Bacteriol. 169, 5174–5179.

[50] Burland, V., Plunkett, G., Sofia, H.J., Daniels, D.L. and Blattner, F.R. (1995) Nucleic Acids Res. 23, 2105–2119.