



HAL
open science

A survey of polypeptide deformylase function throughout the eubacterial lineage

Didier Mazel, Eric Coïc, Stéphane Blanchard, William Saurin, Philippe
Marlière

► **To cite this version:**

Didier Mazel, Eric Coïc, Stéphane Blanchard, William Saurin, Philippe Marlière. A survey of polypeptide deformylase function throughout the eubacterial lineage. *Journal of Molecular Biology*, 1997, 266 (5), pp.939-949. 10.1006/jmbi.1996.0835 . hal-04253842

HAL Id: hal-04253842

<https://hal.science/hal-04253842v1>

Submitted on 23 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Survey of Polypeptide Deformylase Function Throughout the Eubacterial Lineage

Didier Mazel^{1*}, Eric Coïc¹, Stéphane Blanchard², William Saurin³ and Philippe Marlière^{1*}

¹Unité de Biochimie Cellulaire (CNRS URA 1129)

²Unité de Génétique Moléculaire Humaine (CNRS URA 1968)

³Unité de Programmation Moléculaire et Toxicologie Génétique (CNRS URA 144) Institut Pasteur, 28 rue du Docteur Roux, 75724, Paris Cedex 15, France

N-terminal formylation of ribosome-synthesized polypeptides is assumed to be among the most conserved features that distinguish the eubacterial line of descent from other living phyla. In order to assess the ancientness of this trait, *def* genes encoding polypeptide deformylase were characterized from four eubacterial species, *Lactococcus lactis*, *Bacillus subtilis*, *Calothrix* PCC7601 and *Thermotoga maritima*, taking advantage of the conditional viability of the *def* mutants of *Escherichia coli*. Altogether, eight sequences of polypeptide deformylase have been obtained from all the eubacterial sources which were investigated, either through systematic genome sequence analysis or through genetic screening, yielding a highly homologous family. A gene putatively encoding Met-tRNA_i formyltransferase, *fnt*, was found downstream of the deformylase gene except in *L. lactis*, *Mycoplasma genitalium*, *Calothrix* PCC7601 and *T. maritima*. These results argue strongly for the ancestral character of N-terminal formylation in eubacteria. Most of the wide deviations of amino acid usage observed in *def*- and *fnt*-encoded proteins among species is best accounted for by the nucleotide composition of genomes. Furthermore, the species of origin of each protein appears to be more recognizable than its function, considering only its amino acid composition.

© 1997 Academic Press Limited

*Corresponding author

Keywords: *def* gene; evolution; protein synthesis; trimethoprim

Introduction

N-terminal formylation is a metabolic trait that distinguishes eubacteria from other living organisms (Kozak, 1983). Translation initiation with a formyl-methionine primer has however been directly established only in a limited number of cases (reviewed by Lengyel & Söll, 1969), in particular in the chloroplast and mitochondria of eukaryotes, corroborating their endosymbiotic origin. The generalization of a common process to the entire eubacterial line of descent was inferred from indirect evidence pertaining to features of initiator tRNAs (compiled by Sprinzl et al., 1989), primary among which the C1:A72 mismatch that is recognized by Met-tRNA_i formyltransferase (for a review, see RajBhandary, 1994). A straightforward indication of N-terminal formylation is provided by the bio-

chemical characterization of a formylated species of Met-tRNA_i (Marcker & Sanger, 1964), although this determination can be experimentally exacting. We thus opted for the genetic characterization of a related component of translational initiation with f-Met-tRNA_i, i.e. the deformylase enzyme. Activity of this enzyme was demonstrated to be vital in *Escherichia coli*, as its absence blocks further processing of N-terminal methionine (Chang et al., 1989). We reported that mutants lacking the *def* gene but expressing the *fnt* gene were unable to grow in Mueller-Hinton plus thymidine medium unless folate metabolism was blocked, e.g. with trimethoprim (Mazel et al., 1994). This conditional viability thus provided a convenient selection screen for heterologous *def* genes, with which we have now explored the eubacterial kingdom. *Thermotoga maritima*, which 16 S rRNA taxonomy analysis defines as belonging to one of the deepest branches in this line of descent (Woese, 1987), appeared as a tempting target for investigation. The cyanobacterium *Calothrix* PCC7601 was also studied because this organism is known to switch to a specialized metabolic regime for sparing methionine and other

Present address: E. Coïc, UMR 144, Institut Curie, 12 rue Lhomond, 75005 Paris, France

Abbreviations used: MH, Mueller-Hinton; LB, Luria-Bertain; ORF, open reading frame; THF, tetrahydrofolate.

sulphur metabolites (Mazel & Marlière, 1989), which may further constrain translation initiation. The firmicute *Bacillus subtilis* was also selected for investigation, because deformylase activity is bound to ribosomes in its cells unlike *E. coli* (Takeda & Webster, 1968), a feature that might require additional determinants in the structure of the enzyme. All attempted clonings resulted in the characterization of a *def* gene whose products were similar to *E. coli* deformylase. The whole set of deformylase sequences, which include, in addition to the sequences elucidated by complete genome analysis of *Haemophilus influenzae* (Fleischmann *et al.*, 1995) and *Mycoplasma genitalium* (Fraser *et al.*, 1995), those of *Lactococcus lactis* (Cancilla *et al.*, 1995), *Thermus thermophilus* (Meinzel & Blanquet, 1994), and those elucidated in this work, form a homologous family. This strongly argues for the ancestral state of N-terminal formylation in eubacteria. In addition, analysis of deformylase sequences reveals an amino acid usage strongly biased by constraints of DNA metabolism.

Results

Selection screen for deformylase genes and identification of ORF211 from *Lactococcus lactis* as a *def* gene

An *E. coli* strain encoding an active Met-tRNAⁱ transformylase grows only if deformylase is also active or if the endogenous supply of its 10-formyl-THF is blocked (Mazel *et al.*, 1994); otherwise, growth is arrested by accumulation of N-formylated proteins. The strain β 2174, bearing a deletion of the *def* and *fnt* genes in its chromosome and carrying the multicopy plasmid pFMT expressing *fnt* and *tet* can be propagated in MH medium plus thymidine, trimethoprim and tetracycline. This medium includes nutrients which can be converted into the five metabolites derived from the tetrahydrofolate coenzyme (see Figure 1). Trimethoprim inhibition of dihydrofolate reductase ensures that formyl-THF is not synthesized, thereby preventing N-terminal formylation, and tetracycline ensures that the *fnt* bearing plasmid is not lost over successive generations. The multiplicity of the plasmid expressing *fnt* in strain β 2174 prevented loss of formyltransferase by mutation. Transforming the β 2174 Def⁻ strain with a compatible plasmid carrying foreign DNA inserts will not result in viable clones in medium containing tetracycline but lacking trimethoprim unless accumulation of N-formylated proteins is prevented, e.g. by expression of a deformylase gene (Figure 1).

This selection screen was applied to the ORF211 from *L. lactis*, neighbouring the *gap* gene encoding the glyceraldehyde-3-phosphate dehydrogenase (Cancilla *et al.*, 1995), which displays a clear resemblance to the *E. coli def* gene. The host plasmid of ORF211, pJDC9, was reconstructed by deleting the *EcoRI* insert from the ORF211 bearing plasmid pMU5521 (Cancilla *et al.*, 1995). Transformation

of the β 2174 Def⁻ strain with pMU5521 resulted in clones able to grow without trimethoprim in MH medium plus tetracycline, unlike pJDC9 transformants of β 2174 which remained trimethoprim dependent. The identity of a gene encoding a deformylase gene in the β 2174 Def⁻ strain can be confirmed by taking advantage of the methionine auxotrophy of this strain. Indeed, deformylase activity is required for retrieving methionine from certain formylated peptides (Mazel *et al.*, 1994). It was found accordingly that pMU5521, unlike pJDC9, conferred to β 2174 the ability to grow on minimal medium with f-Met-Ala as a methionine source. These results strongly suggest that ORF211 from the genome of *L. lactis* is a deformylase gene.

Bacillus subtilis

A genomic library was constructed from *B. subtilis* strain 168 into plasmid pTZ18R and propagated in the *E. coli* strain GT869, from which it was retrieved for subsequent selection screening after transforming the β 2174 Def⁻ strain. Transformants were plated on MH medium containing tetracycline, ampicillin, thymidine and trimethoprim to estimate the total population size of transformants, which amounted to about 15,000 (see Table 1). Plating on the same medium but lacking thymidine and trimethoprim afforded two clones that could prevent or withstand accumulation of N-formylated proteins. Both clones were found to grow on minimal medium with f-Met-Ala as a source of methionine, indicating that they harboured a deformylase gene. The inserts from the two plasmids were found to be identical as established by restriction analysis and spanned 3.9 kbp (Figure 2). Further analysis was conducted using one of these two plasmids, pBS1. Sequencing of the insert revealed the presence of four open reading frames (ORFs) in the same orientation (Figure 2), 577, 160, 317 and 239 codons long, respectively. The first one displays 37% identity with the *priA* gene from *E. coli*, encoding the factor Y of the primosome (Nurse *et al.*, 1990). The second ORF started with a TTG codon, a situation often encountered in *B. subtilis* (Green & Vold, 1993). The 160 amino acid long protein sequence, deduced from it, showed 44% identity with the *def* gene product of *E. coli* (Figure 3). The protein sequence of the third ORF shared 44% identity with Met-tRNAⁱ transformylase of *E. coli* (Figure 4). The insert ended with a fourth 3'-truncated ORF of 239 codons showing 28% identity with the protein sequence deduced from ORF HI0624 of *H. influenzae*, 451 codons long. Incidentally, sequencing of the *fnt* downstream region from *E. coli* revealed a unique gene of 429 codons, which we designate as *sun*, showing extensive similarity to ORFs equivalently located in *H. influenzae* and *B. subtilis*, and corresponding to a fusion of the *fmu* and *fmv* ORFs, recognized previously (Meinzel & Blanquet, 1994). Moreover, the protein sequences deduced from these genes are

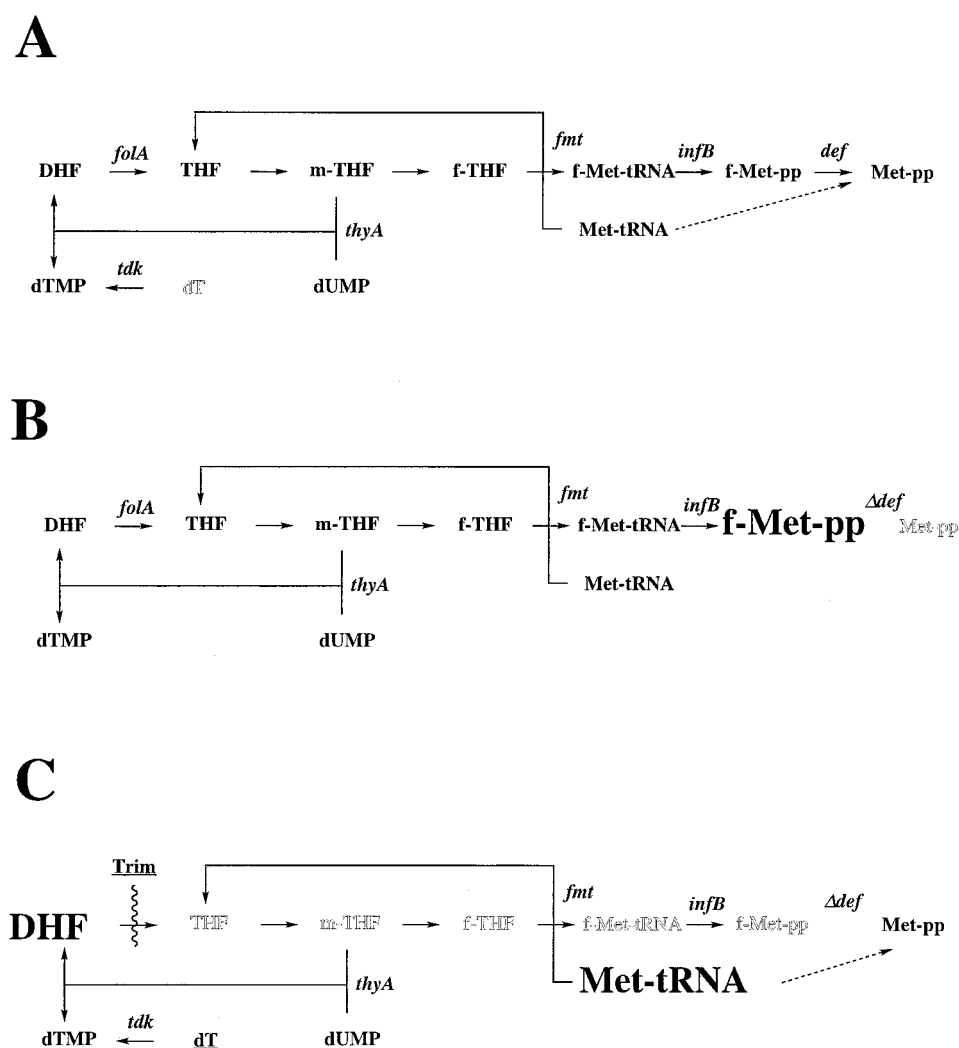


Figure 1. Conditional viability of deformylase-deficient strains of *Escherichia coli*. A, The metabolic pathway leading from dihydrofolate (DHF) to amino free polypeptides (pp) condensed on ribosomes *via* tetrahydrofolate (THF), 5,10-methylene-tetrahydrofolate (m-THF), 10-formyl-tetrahydrofolate (f-THF), initiator formyl-methionyl-tRNA (f-Met-tRNA) and N-formyl-polypeptides (f-Met-pp) is shown, together with afferent reactions producing thymidylate (dTMP) from deoxyuridylate (dUMP) and from thymidine (dT, written in ghost letters to indicate its trace quantities in metabolism). The broken arrow represents the direct condensation of unformylated initiator Met-tRNA_i into polypeptides, presumed to occur marginally. *E. coli* genes commanding the catalysis of individual reactions of the pathway are shown in italics. B, Deletion of the *def* gene results in cell death by accumulation of formylated polypeptides (written in large letters). C, The growth of a *def* mutant is however restored if the production of tetrahydrofolate is blocked by inhibition of *folA*-encoded dihydrofolate reductase by trimethoprim (Trim, written as underlined to indicate its exogenous supply). Growth under such conditions of tetrahydrofolate shortage requires an exogenous supply of thymidine as well as other factors (glycine, pantoate, inosylate, methionine) which can be synthesized from the nutrients of Mueller-Hinton medium, and are omitted here for the sake of clarity. Translation is primed by unformylated Met-tRNA_i. Active *def* genes from heterologous origins can thus be cloned in a *def* mutant of *E. coli*, propagated in Mueller-Hinton medium plus trimethoprim and thymidine as in C, by selection for growth of the transformants in the absence of the two latter compounds as in A.

similar to the *E. coli nusB* gene product, which is known to interact with the ribosomal protein S10 in the transcription antitermination complex (Nodwell & Greenblatt, 1993). Shortening the 3.9 kbp insert by subcloning with *EcoRV* and *PstI*, which bracket the *def*-homologue, resulted in a 1035 bp insert still conferring the Def⁺ phenotype (pBS10 in Figure 2). On the other hand, truncating the *def*-homologue by subcloning the *DraI*-*SacI* fragment in pTZ18 (pBS11 in Figure 2) abolished

this phenotype. The same construct expresses a full size *fmt*-homologue, which, just as *E. coli fmt*, cannot be propagated in the $\Delta(def-fmt)$ strain β 2098 unless trimethoprim and thymidine are added to the growth medium (Figure 1). Genomic DNA from *B. subtilis* strain 168 cut by *SacI* was probed with the 1035 bp *EcoRV*-*PstI* fragment from the pBS10 insert, resulting in only one detectable band of about 3.5 kbp by Southern blot hybridization (data not shown).

Table 1. Deformylase gene cloning

	Species			
	<i>Bacillus subtilis</i> strain 168	<i>Calothrix</i> PCC7601	<i>Thermotoga maritima</i>	<i>Lactococcus lactis</i>
Genetic source	Genomic library	Genomic library	Genomic library	Recombinant plasmid
Restriction enzyme	<i>Sac</i> I	<i>Sau</i> 3AI partial	<i>Sau</i> 3AI partial	<i>Eco</i> RI
Estimated number of different transformants	1.5 × 10 ⁴	2 × 10 ⁵	2 × 10 ⁵	1
Number of clones complementing Def ⁻ Fmt ⁺ strains	2	22	19	1
Number of clones able to grow with fMet-Ala as source of methionine	2	4	7	1
Representative plasmid	pBS1	pCAL164	pTM9	pMU5521
Length of insert (kb)	3.9	4.2	0.7	1.9

Calothrix PCC7601

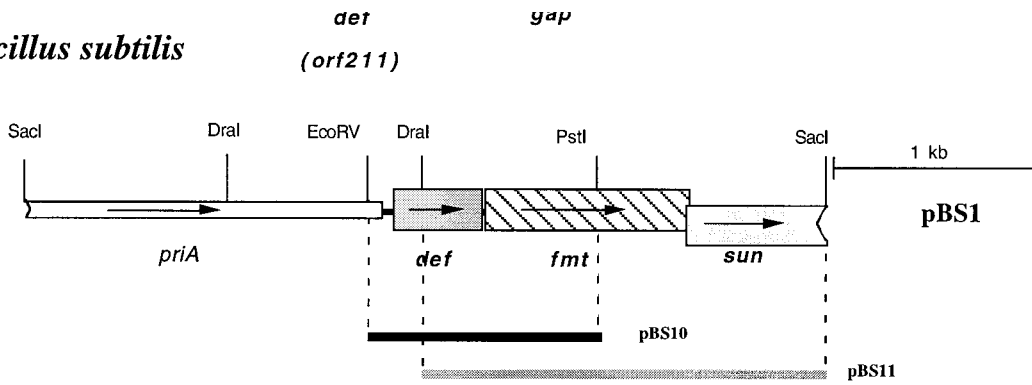
A genomic library was constructed from *Calothrix* strain PCC7601 into the plasmid pUC13 and propagated in the *E. coli* strain GT869, from which it was retrieved for subsequent selection screening after transforming the β 2174 Def⁻ strain. Transformants were plated on MH medium containing tetracycline, ampicillin, thymidine and trimethoprim to estimate the total population size of transformants, which amounted to about 200,000 (Table 1). Plating on the same medium, but lacking thymidine and trimethoprim, afforded 22 clones that could prevent or withstand accumulation of N-formylated proteins. Among those, four clones could also grow on minimal medium with f-Met-Ala as source of methionine, indicating that they harboured deformylase genes. Those clones unable to grow under these conditions were found to contain no *fmt* bearing plasmid pFMT even though they were resistant to tetracycline. An illegitimate recombination event leading to integration of the pFMT plasmid and inactivating *fmt*, but maintaining *tet*, is likely at the origin of these clones, but was not further investigated. It was established, by restriction analysis, that the inserts from the four Def⁺ plasmids overlapped, sharing a 1.3 kbp *Eco*RV-*Kpn*I fragment. A 1656 bp segment encompassing this overlap was sequenced from the plasmid pCAL164, as shown in Figure 2. A large ORF, corresponding to a putative gene of 187 codons starting with ATG, was recognized as specifying a protein sequence with 42% identity the *def* gene product of *E. coli* (Figure 3). This ORF was bracketed by two truncated ORFs in opposite orientation. One of these matched an ORF pre-

viously described in *E. coli* but whose function is unknown (accession number P32701). Shortening the 4.2 kbp fragment from pCAL164 by subcloning with *Xba*I resulted in a 1561 bp insert still conferring the Def⁺ phenotype. The *Eco*RV-*Kpn*I fragment from pCAL164, encompassing the *def* gene, was used to probe genomic DNA from *Calothrix* 7601 cut by *Xba*I, resulting in only one detectable band of about 1.6 kbp, by Southern blot hybridization, even at low stringency (data not shown).

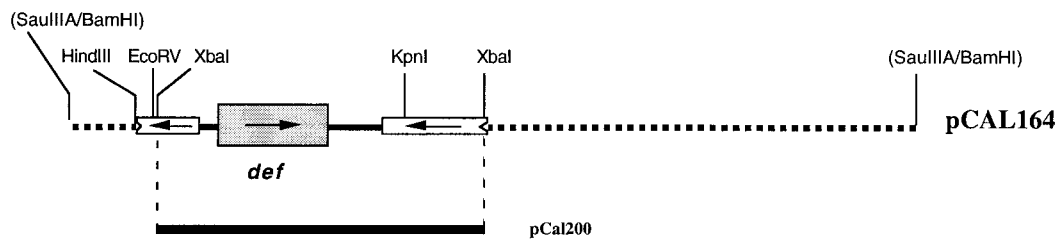
Thermotoga maritima

A genomic library was constructed from *T. maritima* genomic DNA into the plasmid pUC13 and propagated in the *E. coli* strain GT869, from which it was retrieved for subsequent selection screening after transforming β 2174. Transformants were plated on MH medium containing tetracycline, ampicillin, thymidine and trimethoprim to estimate the total population size of transformants, which amounted to about 200,000 (Table 1). Plating on the same medium but lacking thymidine and trimethoprim afforded 19 clones that could prevent or withstand accumulation of N-formylated proteins. Among those, seven clones could also grow on minimal medium with f-Met-Ala as source of methionine, indicating that they harboured deformylase genes. The clones unable to grow under these conditions were found to contain no pFMT plasmid even though they were resistant to tetracycline. Presumably, an illegitimate recombination event, leading to genomic integration of the pFMT plasmid and inactivation of *fmt*, but maintaining *tet*, is likely at the origin of these clones, but was not further considered. Of the seven plasmids iso-

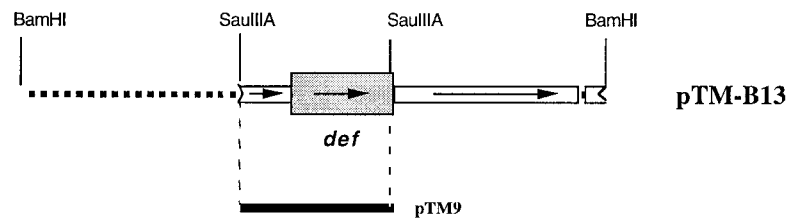
Bacillus subtilis



Calothrix PCC 7601



Thermotoga maritima



Lactococcus lactis

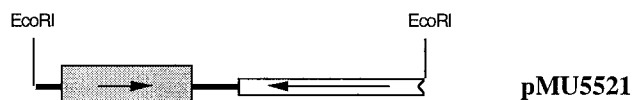


Figure 2. Physical maps of homologous *def* loci. Unsequenced parts of the different inserts are shown by dotted lines. Genes and ORFs are indicated by boxes, and identified by their names when known. Each bold line indicates the smallest insert generated by recloning and conferring the Def⁺ phenotype. The grey line indicates the *B. subtilis* subclone with Def⁻ Fmt⁺ phenotype, as discussed in the text.

	1				50
Tth			MV	YPIRLYGDV	LRRKAREV
Tma			M	YRIRVFGDPV	LRRKRAKFTV
Eco			MSV	LQVLHLPDER	LRRKARFV
Hin			MTA	LNVLIIYFDDH	LKVVCPEFVT
Bsu			MAV	KKVVTTHPAEV	LETFAETVVT
Cal		MPSEIA	VEKSKLKNPP	LQLHLYLGDV	LRRQPAKRIA
Mge	MTKILPWLFT	SIVRIILITLL	FLSMTFQPTK	TWLVFDDNAL	LNKPEAVNPF
Lla		MNRDEIQAN	LIKASHMISM	DDIIRREGYPT	LREAVANDVTL
	51				100
Tth	DFSG	IKRLA	EDMLETMFEAKGVGLAAP
Tma	KFDENLKKTI	ERMIEITMYHYDVGGLAAP	QVGLSQRFF
Eco	EVNAEIQRIV	DDMFETMYAEEGIGLAAT	QVDIHQRII
Hin	KVNDAIRKIV	DDMFDTMYQEKGIIGLAAP	QVDILQRII
Bsu	VFDKLLKLL	DDMYDTMLMDVGGLAAP	QIGILKRAA
Cal	KVDDDELRLI	RDMLOQMYSKDGIIGLAAP	QVGIHKQLI
Mge	FIDEQIETCI	KKMIAVDAS	YDG	KAQEYD	IIPGIGIAAN
Lla	PLSDEDIILG	EKMLQFLHNS	QDPVMAEKMG	LRGGVGLAAN	QLGLLKKVIA
	101				150
Tth	..VAVEYADE	PEGEEERPLR	ELVRRRVVVA	NPVITYREGL	VEGT..EGCL
Tma	..V.MDVNG	P.....	..VAVIN	PELLEIDPET	EVAE..EGCL
Eco	..V.IDVSEN	RDE.....	..RLVLIN	PELLEKSGET	GI..E..EGCL
Hin	..T.IDVGD	KQN.....	..QFVLIN	PELLESEGET	GI..E..EGCL
Bsu	..V.VEIGDD	R.G.....	..RIDLVN	PELLEKSGET	TG..I..EGCL
Cal	..V.IDLEPD	NPAN.....	..PPLVLIN	PTIKQVSKET	CVAQ..EGCL
Mge	IHL.....	..NDLMEK	KC.....	LLIN	PKIIDQSENK
Lla	VLIPNEPEVD	EENETIPPKE	AYKMRBIMYN	AKVVSHSVQD	AAVEGEGGCL
	151				200
Tth	SLPGLYSEEV	PRAERIRVEY	QDEEG..RGR	VLELEGYMAR	VFQHEIDHLD
Tma	SPPEIFV	EI	ERSKRIKVKY	QNTFRG..EYV	EELLEGYAAR
Eco	SIPQRA	LV	PRAEKVKIRA	LDRDG..KPF	ELEADGLLAI
Hin	SIPGFRA	LV	PRKEKVTVRA	LDRDG..KEF	TLDADGLLAI
Bsu	SFPNVYG	DV	TRADYKVVRA	FNRQG..KPF	ILEARGFLAR
Cal	SIPNVYM	DV	KRPEVVELAY	KDENG..RPK	TLKATDLLAR
Mge	SVKQKHQYV	IRSEWITIKG	YDW	FEKKEI	TIKATGLFGM
Lla	SVDREVPYV	VRHARVTVEY	YNKEGKCKKI	RLK..DFPAI	CVQHEIDHTN
	201				241
Tth	GILFFERLPK	PKREAFLEAN	RAELVRFQKE	ARALLKELSQ	G
Tma	GVLIIDRISF	AKR	LLLRKK	LMDIARTVKR	
Eco	GKLFMDYLSF	LKQQRIRQKV	EKLDRLKARA		
Hin	GILFVDYLSF	LKRQRILKEL	TKYKQKAKS		
Bsu	GVLPSTKISK	YYTEDELADM	EG		
Cal	GVVFDVRVDN	SLTLAQELSK	NGFSYQAVK	VA	
Mge	GRFFYQRINP	LNPFWKKPEW	KVINPTLKT	NG	
Lla	GVMFYDHINM	NDPWEIKDGM	IIVK		

Figure 3. Comparison of polypeptide deformylase sequences. Tth, *Thermus thermophilus*; Tma, *Thermotoga maritima*; Eco, *Escherichia coli*; Hin, *Haemophilus influenzae*; Bsu, *Bacillus subtilis*; Cal, *Calothrix PCC7601*; Mge, *Mycoplasma genitalium*; Lla, *Lactococcus lactis*. Identical residues are shown by *, and conservative replacement by -.

lated as possibly carrying a deformylase gene, pTM9 contained a 0.7 kbp insert that was spanned by the six other plasmids, which contained larger inserts. This insert was sequenced, yielding a truncated putative gene starting with an ATG preceded by a segment complementary to the *T. maritima* 16 S rRNA 3' terminus (accession number M21774) and interrupted after 139 codons at a *Sau3AI* site fused with the *Bam*HI of the cloning vector. The deduced protein sequence matched the N terminus of the *E. coli def* gene product. None of the longer inserts in the remaining six plasmids extended farther than 0.6 kbp after that *Sau3AI* site. In order to complete the sequence of the *def* gene and of its 3' surroundings, a *Bam*HI library was prepared from *T. maritima* DNA and cloned in pUC13. The resultant transformants in the Def⁻Met⁻ strain β 2174 were scored as above for inserts conferring growth on MH medium lacking thymidine and trimethoprim, as well as on minimal medium supplemented with f-Met-Ala as a methionine source. Eleven clones out of about 50,000 transformants

	1				50
Eco	MSESLRIIFA	GTPDFAAARHL	DALL	SSGHN	VVGVFQTPDR
Hin	MKSLNIIIFA	GTPDFAAQHL	QAIL	NSQHN	VIAVVTQPKD
Bsu	MTRIVFM	GTPDFSVFVL	RTLIL	EDGYE	VVGVTQDR
Tth	MRVAFF	GTPFLNAFVL	DAL	RKRHQ	VVLVVSQPKD
Mge	MFKIVFP	GTSTLSKKCL	EQLFYDNDPE	ICAVVTQPKD	INHNRNKIVP
	51				100
Eco	SPVKVLAEEK	GLPVFQPVSL	RPOENQQLVA	ELQADVMMVV	AYGLILPKAV
Hin	SPVKQLAEQN	NIPVYQPKSL	RKEEAQSELK	ALNADVMVV	AYGLILPKAV
Bsu	PPVKEEALRH	GIPVLQPEKV	RLTEBIEKVL	ALKPDLIVTA	AFGQILPKEL
Tth	SPVARYAEAE	GLPPLLRPARL	REEAFLEALR	QAAPAVAVVA	AYGKILPKEA
Mge	SDVKSFCLEK	NITFFQPK..	QSSISAKALE	KLKADIGICV	SFGQYLHQDI
	101				150
Eco	LEMPLRGLIN	VHGSLLPRWR	GAAPIQRSWL	AGDAETGVTI	MQMDVGLDTG
Hin	LDAPRLGCLN	VHGSILPRWR	GAAPIQRSWL	AGDVQTVTI	MQMDVGLDTG
Bsu	LDSPKYGCIN	VHASLLPELR	GGAPITHYSL	QGGKKTGITI	MYMVEKLDG
Tth	LIDIPPHGLN	LHPSLLPKYR	GAAPVQRALL	AGERETGVSI	MRLDEGLDTG
Mge	IDLFPNKVIN	LHPSKLLPLR	GAAPLHWITI	NGFKKSALVS	IQLVKKMDAG
	151				200
Eco	DMLYKLSCEPI	TAEDTSGTLY	DKLAEELGPOG	LITTLKQLAD	GTAKPEVQDE
Hin	DMLHKVYCDI	LPTETSTSLY	NKLAEALPSA	LIDVLDNLEN	GKFAIEKQDG
Bsu	DMISKVEVDI	EETDNVGLH	DKLSVAGAKL	LSETVFNVA	GSISPEKQDG
Tth	PLYAVWRTPI	LPDEDAVALG	NRLRDKGVEL	LLEVLERLPE	LTFRP...QE
Mge	PIWKQDFLV	NNDWNTGDS	IYVEHSPSF	LIECTKELIN	KKGKWFQI
	201				250
Eco	TLVITYAEKLS	KEEARIDNSL	SAAQLERCIR	AFNPFWMSWL	EIE...GQ
Hin	SQSNTYAEKLS	KEEAQLNWSL	SAMQLERNIR	AFNPFPIAYF	STEDKGNH
Bsu	EKATYAPNIK	REQELEDWSR	TGEELYNQIR	GLNPFWVAYT	TL...NGQ
Tth	GEVSYAPPLS	KEEGRLDVPE	SABALYRRHR	AVQFWPGSYF	FHRGRRVKAL
Mge	GEPTFGLNIR	KEQEHLDLNQ	IYKSLNWKV	GLAPKPGGWL	SFE...GK
	251				300
Eco	PVKVWASVI	DTAT.NAAPG	TILEANKQGI	QVATG.DGIL	NLLSLQPAK
Hin	TLKVYQAKVL	PHQD.KPA.G	TILSADKNGI	QIATV.DGVL	NLLQLQSAGK
Bsu	KLENMGVEKN	SGTN.NSEPG	TVVAVEKEGI	IVATGNETAL	LLTELQPGK
Tth	RLR.....	PEFG.EGEPG	VVARVGPVGV	AVGTA.SGLL	LLLEVQPEGR
Mge	NIKIFKAKYV	SKSNYKHQLG	EIVNISRKGI	NIALKSNEII	SIEKIQIPGK
	301				329
Eco	KAMSAQDLLN	SRREWFVPGN	RLV		
Hin	KPMSAQDLLN	GRAEWFITGK	VLA		
Bsu	KRMKGEDFV	..RGARHVEAGD	VLGVNNEKN		
Tth	RAMPAADW	..ARGYVAPCT	RLGQV		
Mge	RVMEVSEIIN	GKHPPVV.V	GK	CFK	

Figure 4. Comparison of Met-tRNAi transformylase sequences. See abbreviations and symbols in Figure 3.

passed this double screen. They were found to contain an identical insert of 2.9 kbp. The sequence of the putative *T. maritima def* homologue could therefore be completed. This gene encodes a 164 amino acid protein showing 40% similarity to *E. coli def* deformylase (Figure 3). That the *T. maritima def* homologue could be shortened while remaining functional is not without precedent, as we previously showed that the last six codons of the *E. coli def* gene carried on a multicopy plasmid can be removed without apparent change of phenotype (Mazel *et al.*, 1994). A downstream putative gene in the same orientation starting with GTG had the capacity to encode a 311 amino acid protein resembling the protease lysostaphins from *Staphylococcus* (Recsei *et al.*, 1987) and homologous proteins (Karow & Georgopoulos, 1992; Ichikawa *et al.*, 1994; Kovach *et al.*, 1994; Lange & Hengge-Aronis, 1994). There was also a gene upstream of the *def* homologue in the same orientation which could be detected from its 97 3'-terminal codons as highly similar (36% identity) to *surE* from *E. coli*, a gene whose mutants were shown to be impaired in stationary phase survival (Li *et al.*, 1994). The 0.7 kbp insert from the pTM9 insert was labelled to probe genomic DNA from *T. maritima* cut by

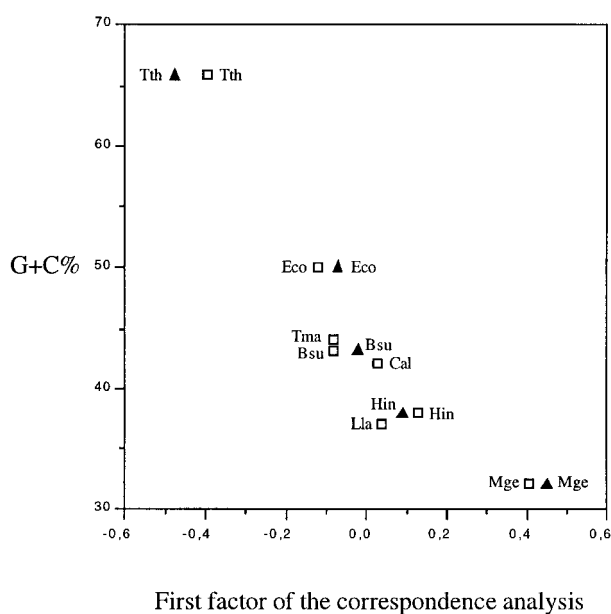


Figure 5. Correlation between the amino acid composition of *def* and *fnt* gene products and the nucleotide composition of genomes. The genome G + C% of each organism has been plotted against the first factor of the correspondence analysis of the amino acid composition of each protein. Open squares, *def* products; filled triangles, *fnt* products.

*Bam*HI, resulting in only one detectable band of about 2.9 kbp by Southern blot hybridization (data not shown).

Sequence analysis

Deformylase sequences from the eight eubacterial species so far elucidated exhibit three blocks of elevated resemblance. The N and C termini appear variable in length and sequence. Two of the three conserved blocks within the sequences were previously identified as corresponding to a zinc coordination motif, whose critical residues, Cys149, His194 and His198 in Figure 3, were confirmed by systematic mutagenesis (Meinzel *et al.*, 1995). The third conserved block encompasses a motif with an aliphatic sequence following a glycine doublet GI/VGLAAxQI/V/L, reminiscent of an external turn (see Figure 3). Beyond their sequence resemblance, the eight deformylase chains display a wide variation of amino acid usage (Table 2). The amino acid compositions of *def* and *fnt* encoded proteins have been subjected to correspondence analysis. The first two factors of the correspondence analysis accounted for 65.7% of the inertia of the amino acid compositions, 47% and 18.7% respectively. It is shown in Figure 5 that most of the variability in amino acid composition of *def* and *fnt* encoded proteins is correlated to the nucleotide composition

Table 2. Chemical composition of deformylase proteins

	Species							
	Mge	Lla	Hin	Cal	Bsu	Tma	Eco	Tth
Genome G + C (%)	32	37	38	42	43	44	50	66
Amino acids								
Gly	12	15	10	8	13	10	8	15
Ala	9	13	10	13	13	9	12	15
Ser	8	7	4	7	4	4	5	5
Pro	10	10	7	14	6	8	7	11
Val	7	20	13	18	16	16	14	17
Thr	13	4	7	5	9	6	3	3
Cys	4	2	3	3	1	1	2	1
Ile	25	18	19	13	8	14	15	7
Leu	20	15	18	21	15	12	20	23
Asn	12	12	5	9	3	5	3	2
Asp	10	15	15	13	14	8	12	7
Gln	10	6	8	11	4	4	9	6
Lys	23	15	15	16	12	12	11	7
Glu	13	21	12	11	14	18	20	29
Met	4	11	3	5	6	5	5	4
His	4	7	3	4	3	3	4	2
Phe	13	3	5	2	6	8	3	8
Arg	4	9	8	9	8	15	14	22
Tyr	8	7	4	5	5	6	2	8
Trp	7	1	0	0	0	0	0	0
Total	216	211	169	187	160	164	169	192
Elements								
H	1795	1675	1397	1530	1265	1374	1410	1573
C	1169	1045	853	929	789	855	849	986
N	288	286	227	258	209	236	242	277
O	313	319	251	272	241	241	255	288
S	8	13	6	8	7	6	7	5

Items were counted regardless of post-translational modifications.

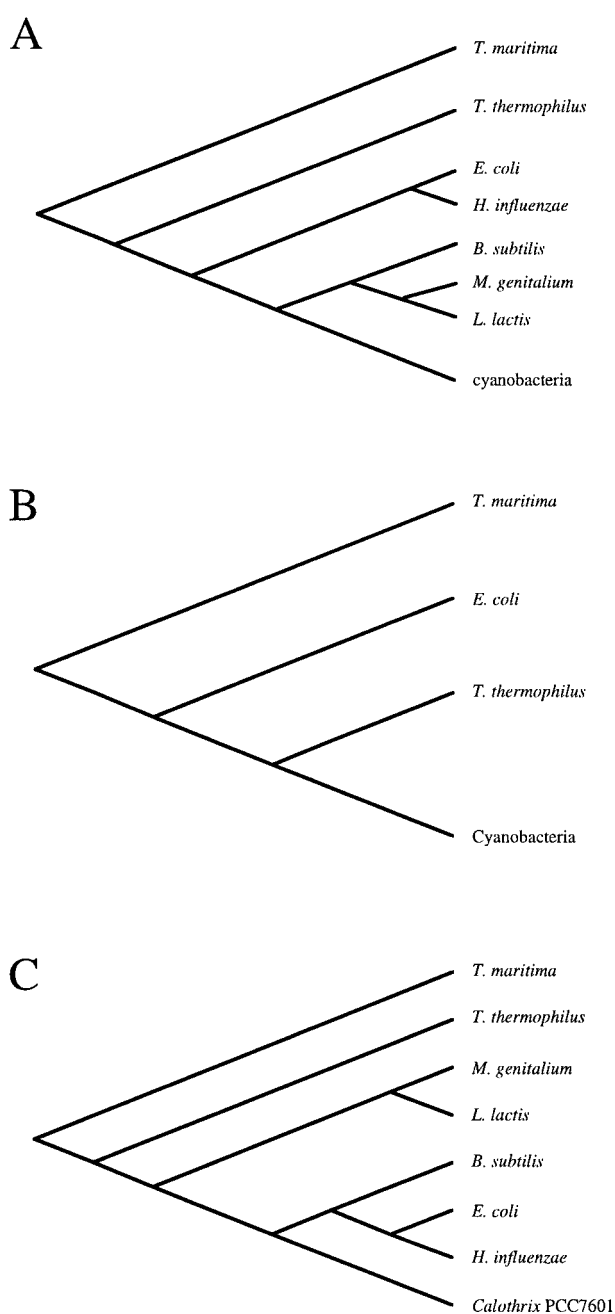


Figure 6. Comparative phylogenies reconstructed from small: A, subunit rRNA (Woese, 1987); B, elongation factors (Baldauf *et al.*, 1996); C, protein sequences encoded by the *def* gene. Branch lengths are not proportional to the computed evolutionary distances.

of genomes. Indeed, the value of the correlation coefficient between the G + C content and the coordinates of the first factor of the correspondence analysis is 0.93. Significantly the second factor of the correspondence analysis discriminates between *def* and *fnt* encoded proteins (data not shown). Variability associated with the genome chemical composition is thus twice as large as that associated with function.

Phylogenetic analysis

The currently available set of eight deformylase sequences was aligned using the program PILEUP and the least variant stack of 140 conserved residues subjected to classification using the neighbour joining method (Saitou & Nei, 1987). Three groups are well defined, according to the bootstraps scores comprising, respectively, *T. thermophilus* and *T. maritima* (bootstrap score 954), *M. genitalium* and *L. lactis* (bootstrap score 968), and *E. coli* and *H. influenzae* (bootstrap score 1000). Ambiguities in the placement of firmicutes and cyanobacteria are recurrently encountered in separate studies. As apparent in Figure 6, phylogenetic trees of eubacterial species derived from sequences of small subunit rRNAs (Woese, 1987), translation elongation factors (Baldauf *et al.*, 1996) and deformylases are not consistent, even though these molecules act in the same biological function. The only firm conclusion that can be drawn from these trees is the early branching of the *Thermotoga* line of descent.

Discussion

A genetic segment capable of replacing the *def* gene in *E. coli* could be cloned from the genomes of the three eubacterial species that we investigated, *B. subtilis*, *Calothrix PCC7601* and *T. maritima*. These genetic segments also allowed a *def metA* double mutant to grow with f-Met-Ala as the methionine source just as the *E. coli def* gene itself. ORF of *L. lactis* also satisfied this criterion. The genes thus elucidated from *B. subtilis*, *Calothrix*, *T. maritima* and *L. lactis* were found to be similar in size and sequence to those previously characterized in *E. coli* and *T. thermophilus* as well as to the HI0622 gene from *H. influenzae* and the MG106 gene from *M. genitalium* discovered through systematic genome sequencing (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995). *def* homologues in the two latter species were found to be unique within their entire genomes. No additional homologues in the *B. subtilis*, *Calothrix* and *T. maritima* genomes could be detected by hybridization experiments, just as for *E. coli* (Mazel *et al.*, 1994). It can be inferred from the present survey that a deformylase gene has been present in the genomes of eubacteria since the divergence of the *Thermotoga* branch (Woese, 1987). Whether the last common ancestor of all extant organisms used a formyl group remains an open question. In this context, we note that no gene homologous to *def* or to *fnt* could be found in the entire genome of the archaeobacterium *Methanococcus jannaschii* (Bult *et al.*, 1996). Thorough computer searches for homologous proteins failed to match proteins beyond the zinc peptidase family, as previously mentioned by Meinel & Blanquet (1995) (data not shown). The genetic origin of deformylase thus remains in limbo.

The neighbourhood of the *def* genes of the eight eubacterial species appears to vary, *def* being concatenated in a 5'-*def-fnt*-3' structure with the Met-

tRNAi formyltransferase gene in *T. thermophilus*, *B. subtilis*, *E. coli* and *H. influenzae*. That this organisation constitutes an operon in those organisms, as demonstrated in *E. coli* (Meinzel *et al.*, 1993), stems from the proximity or overlap of genes in the same orientation (Figure 2). One may reasonably suppose that the *def-fmt* operon was already present in the last common ancestor of *Thermus*, proteobacteria and firmicutes (Figure 6). According to this hypothesis, the operon structure would have disintegrated during evolution of the lactic bacteria-*Mycoplasma* cluster in the firmicute phylum as well as in the cyanobacterial phylum. A third gene, *sun*, similar to *nusB* from *E. coli* is found to closely follow the *def-fmt* operon in *E. coli*, *H. influenzae* and *B. subtilis*, but not in *T. thermophilus*. Again, there is no evidence to decide whether the 5'-*def-fmt-sun*-3' operon structure was already present in the last common ancestor of *T. thermophilus*, proteobacteria and firmicutes, but it was probably in the last common ancestor of proteobacteria and firmicutes.

Polypeptide deformylase is expected to be an abundant protein of eubacteria considering that the bulk of the proteins condensed on ribosomes is processed by this enzyme. As such, it provides a probe for studying metabolic constraints in the natural design of proteins (Mazel & Marlière, 1989; Gandbhir *et al.*, 1995). One expects that statistics of elemental composition will reveal such metabolic constraints. *B. subtilis* deformylase stands as the most frugal version, requiring the lowest amounts of N, C and S atoms (Table 2). Its small size does not entirely account for this economy, because the average numbers of N, C and S atoms per amino acid also appear lower for *B. subtilis* deformylase. Considering that organisms of the *Bacillus* lineage have been evolving as soil inhabitants for geological times, metabolic optimisation seems a plausible explanation for the trimming of their deformylase. By contrast, the deviation of the nucleotide composition of genes can be invoked for explaining the richness of other species in certain amino acids, as discovered by Sueoka (1961). The correlation found between the amino acid composition of *fmt* and *def* products and the G + C genome content (Figure 5) leaves little doubt that DNA metabolism is a prominent force shaping protein structure. It can apply to the extent of accumulating as much as an overall 37.5% of Asn, Lys, Ile, Tyr and Phe codons (all encodable with only A and T) in *M. genitalium* deformylase to be compared to 16.6% for the enzyme of the G + C-biased *T. thermophilus*. The high methionine and cysteine content of *Calothrix* deformylase may appear surprising if one considers that it must be expressed under conditions of sulphur scarcity (Mazel & Marlière, 1989). However, it cannot be excluded that N-terminal formylation is suspended under such conditions, with concomitant switching off of deformylase production. Other metabolic constraints on amino acid composition may deserve investigation. For instance, the fact that *M. genitalium* and *L. lactis* obtain their amino

acids from their eukaryotic hosts rather than by synthesizing them may reverberate in the organization of their genomes, as suggested by the excess of carbon content in their deformylases (Table 2). It will be interesting to see whether imprints of amino acid usage can be tracked along food chains of ecosystems by systematic sequence analysis of abundant proteins. Nevertheless, the present analysis suggests that protein composition is primarily determined by genome nucleotide composition and secondarily by protein function.

Experimental Procedures

Culture conditions

E. coli strains were routinely propagated in Luria-Bertani (LB) or Mueller-Hinton (MH) rich media (Miller, 1972). Growth tests were performed in MS minimal medium (Richaud *et al.*, 1993), to which were added 2 g/l of glucose and, when necessary, amino acids (L-methionine and L-threonine) or peptides at 1 mM concentration, thiamine at 3×10^{-6} M. When required, antibiotics were used at the following concentrations: ampicillin, 100 mg/l; kanamycin 25 mg/l; tetracycline 15 mg/l; chloramphenicol, 25 mg/l. Trimethoprim was used at 100 mg/l in LB medium supplemented with 0.3 mM thymidine. Agar (Difco) was added at the concentration of 15 g/l for preparing plates. *T. maritima* was cultivated as recommended by the institution from which it was obtained (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-3300 Braunschweig, Germany). Culture conditions of *Calothrix* PCC 7601 were as described previously (Mazel *et al.* 1986)

Escherichia coli strains and plasmids

The *E. coli* K-12 strain GT869 (Parsot, 1986) was used for routine propagation of plasmids. β 2174 was constructed by transformation of β 2098 (relevant genotype: *metA::kan thrB thi* Δ (*def-fmt*)::*cat*; Mazel *et al.*, 1994) with plasmid pFMT, a pRK415 derivative (*tet*) bearing *fmt* (Mazel *et al.*, 1994). pUC18 and pTZ18R were purchased from Pharmacia. pMU5521 (Cancilla *et al.*, 1995) was kindly provided by Dr B. Davidson (University of Melbourne, Australia).

Construction of genomic libraries

DNA from the *B. subtilis* strain 168 was a gift from Dr T. Msadek (Unité de Biochimie Microbienne, Institut Pasteur, France). This DNA was digested by *SacI* and mixed with an approximately equimolar amount of pTZ18R digested with *SacI* and dephosphorylated. The DNA of *T. maritima* was isolated as described for *M. sociabilis* (Lauerer *et al.*, 1986) except that French press disruption of cells was omitted. The DNA of *Calothrix* PCC7601 was isolated as described previously (Mazel *et al.*, 1986). The DNA from *T. maritima* and *Calothrix* were partially digested with restriction endonuclease *Sau3AI*, and independently mixed with an approximately equimolar amount of pUC18 digested with *Bam*HI and dephosphorylated (Pharmacia). An amount (1 μ g of each of the three DNA mixtures was incubated with phage T4 DNA ligase for 18 hours at 8°C, and then 1/10 of the mix was used to transform the *E. coli* strain GT869 by electropora-

tion following a standard procedure (Dower *et al.*, 1988). A total of about 5×10^5 ampicillin-resistant transformants were thus obtained and pooled together for each library. Total plasmid DNA was prepared on QIAGEN-tip 500 according to the manufacturer's instructions (QIAGEN GmbH). Samples of each library were then used to transform β 2174 by electroporation (*T. maritima* and *Calothrix* libraries) or by the CaCl_2 method (*B. subtilis* library), and subjected to selection as described in Results. Restriction enzymes and T4 DNA ligase were purchased from either Boehringer or Biolabs. Mini-preparation of plasmid DNA, transformation, and restriction enzyme analyses were carried out according to the methods described by Sambrook *et al.* (1989).

DNA sequence determination and computer analysis

DNA fragments were excised from plasmids at convenient restriction sites and ligated in pTZ18R in both orientations. Unidirectional deletions were performed with the Rapid Deletion System (IBI) adapted to pTZ18R as previously described (Mazel *et al.*, 1988). The nucleotide sequences of deleted clones in both orientations were obtained by the dideoxy chain termination method using *Taq* DNA polymerase (Perkin-Elmer) fluorescent dideoxynucleotides and selected oligonucleotide primers. The sequences were determined on an Applied Biosystems 373 ABI DNA sequencer. Sequence data were stored and combined using the software of the Institut Pasteur computer facilities. Similarities of nucleic acid sequences and their putative translation products were searched in the EMBL data library, GenBank and Swissprot using the FASTA (Lipman & Pearson, 1985) or BLAST (Altschul *et al.*, 1990) programs. Amino acid protein sequence alignments have been produced using the PILEUP program (Genetics Computer Group, 1994). Phylogenetic analysis were performed using the NJ method (Saitou & Nei, 1987), and bootstrap scores were computed from 1000 independent trials.

The sequences have been deposited to the EMBL database under the following accession numbers: Y10304 (*B. subtilis*); Y10305 (*Calothrix* PCC7601); Y10306 (*T. maritima*); Y10307 (*E. coli* *sun* gene).

Statistical methods

The amino acid compositions of Def and Fmt were subjected to correspondence analysis as implemented in the BMDP Statistical Software (Dixon *et al.*, 1990).

Acknowledgements

We are grateful to Drs M. Schwartz, M. Goldberg and C. Petit for their constant interest, and to Drs W. Beatty, E. Jonhson, T. Rose and R. Mutzel for helpful comments. We thank Dr T. Msadek for the gift of *B. subtilis* DNA; Dr B. Davidson for the gift of pMU5521; R. de Lecubari and A. Meier for technical assistance. This work was supported in part by research funds from the University of Paris 7 and by a grant from the Ministère de l'Enseignement Supérieur et de la Recherche (Action "Essor des Biotechnologies").

References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. (1996). The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl. Acad. Sci. USA*, **93**, 7749–7754.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G. & *et al.*, (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058–1073.
- Cancilla, M. R., Hillier, A. J. & Davidson, B. E. (1995). *Lactococcus lactis* glyceraldehyde-3-phosphate dehydrogenase gene, *gap*: further evidence for strongly biased codon usage in glycolytic pathway genes. *Microbiology*, **141**, 1027–1036.
- Chang, S.-Y. P., McGary, E. & Chang, S. (1989). Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. *J. Bacteriol.* **171**, 4071–4072.
- Dixon, W. J., Brown, M. B., Engelman, L. & Jennrich, R. I. (1990). In *BMDP Statistical Software Manual* ((to accompany the 1990 software release)). University of California Press Berkeley, Los Angeles, CA.
- Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucl. Acids Res.* **16**, 6127–6145.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R. & *et al.*, (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.
- Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D. & *et al.*, (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403.
- Gandbhir, M., Rascherd, I., Marlière, P. & Mutzel, R. (1995). Convergent evolution of amino acid usage in archaeobacterial and eubacterial lineages adapted to high salt. *Res. Microbiol.* **146**, 113–120.
- Genetics Computer Group, (1994). Program Manual for the Wisconsin Package. In *Version 8*, September 1994, 575 Science Drive, Madison, WI 53711, USA.
- Green, C. J. & Vold, B. S. (1993). tRNA, tRNA processing, and aminoacyl-tRNA synthetases. In *Bacillus subtilis and Other Gram-positive Bacteria* (Sonenshein, L. L., Hoch, J. A. & Losick, R., eds), chapt. 47, pp. 683–698. American Society for Microbiology, Washington D.C.
- Ichikawa, J. K., Li, C., Fu, J. & Clarke, S. (1994). A gene at 59 minutes on the *Escherichia coli* chromosome encodes a lipoprotein with unusual amino acid repeat sequences. *J. Bacteriol.* **176**, 1630–1638.
- Karow, M. & Georgopoulos, C. (1992). Isolation and characterization of the *Escherichia coli* *msbB* gene, a multicopy suppressor of null mutations in the high-temperature requirement gene *htrB*. *J. Bacteriol.* **174**, 702–710.
- Kovach, M. E., Hughes, K. J., Everiss, K. D. & Peterson, K. M. (1994). Identification of a ToxR-activated gene, *tagE*, that lies within the accessory colonization factor gene cluster of *Vibrio cholerae* O395. *Gene*, **148**, 91–95.
- Kozak, M. (1983). Comparison of initiation of protein synthesis in prokaryotes, eucaryotes and organelles. *Microbiol. Rev.* **47**, 1–45.

- Lange, R. & Hengge-Aronis, R. (1994). The *nlpD* gene is located in an operon with *rpoS* on the *Escherichia coli* chromosome and encodes a novel lipoprotein with a potential function in cell wall formation. *Mol. Microbiol.* **13**, 733–743.
- Lauerer, G., Kristjansson, J. K., Langworthy, T. A., König, H. & Stetter, K. O. (1986). *Methanothermus sociabilis* sp. nov., a second species within the *Methanothermaceae* growing at 97°C. *System. Appl. Microbiol.* **8**, 100–105.
- Lengyel, P. & Söll, D. (1969). Mechanism of protein biosynthesis. *Bacteriol. Rev.* **33**, 264–301.
- Li, C., Ichikawa, J. K., Ravetto, J. J., Kuo, H.-C., Fu, J. C. & Clarke, S. (1994). A new gene involved in stationary-phase survival located at 59 minutes on the *Escherichia coli* chromosome. *J. Bacteriol.* **176**, 6015–6022.
- Lipman, D. J. & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, **227**, 1435–1441.
- Marcker, K. & Sanger, F. (1964). N-Formyl-methionyl-S-RNA. *J. Mol. Biol.* **8**, 835–840.
- Mazel, D. & Marlière, P. (1989). Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature*, **341**, 245–248.
- Mazel, D., Guglielmi, G., Houmard, J., Sidler, W., Bryant, D. A. & Tandeau de Marsac, N. (1986). Green light induces transcription of the phycoerythrin operon in the cyanobacterium *Calothrix* 7601. *Nucl. Acids Res.* **14**, 8279–8290.
- Mazel, D., Houmard, J. & Tandeau de Marsac, N. (1988). A multigene family in *Calothrix* sp. PCC 7601 encodes phycocyanin, the major component of the cyanobacterial light harvesting antenna. *Mol. Gen. Genet.* **211**, 296–304.
- Mazel, D., Pochet, S. & Marlière, P. (1994). Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. *EMBO J.* **13**, 914–923.
- Meinzel, T. & Blanquet, S. (1994). Characterization of the *Thermus thermophilus* locus encoding peptide deformylase and methionyl-tRNA^{Met} formyltransferase. *J. Bacteriol.* **176**, 7387–7390.
- Meinzel, T. & Blanquet, S. (1995). Enzymatic properties of *Escherichia coli* peptide deformylase. *J. Bacteriol.* **177**, 1883–1887.
- Meinzel, T., Guillon, J.-M., Mechulam, Y. & Blanquet, S. (1993). The *Escherichia coli* *fnt* gene, encoding methionyl-tRNA^{Met} formyltransferase, escapes metabolic control. *J. Bacteriol.* **175**, 993–1000.
- Meinzel, T., Lazennec, C. & Blanquet, S. (1995). Mapping of the active site zinc ligands of peptide deformylase. *J. Mol. Biol.* **254**, 175–183.
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Nodwell, J. R. & Greenblatt, J. (1993). Recognition of boxA antiterminator RNA by *E. coli* antitermination factors NusB and ribosomal protein S10. *Cell*, **72**, 261–268.
- Nurse, P., DiGate, R. J., Zavitz, K. H. & Mariani, K. J. (1990). Molecular cloning and DNA sequence analysis of *Escherichia coli* *priA*, the gene encoding the primosomal protein replication factor Y. *Proc. Natl Acad. Sci. USA*, **87**, 4615–4619.
- Parsot, C. (1986). Evolution of biosynthetic pathways: a common ancestor for threonine synthase, threonine dehydratase and D-serine dehydratase. *EMBO J.* **5**, 3013–3019.
- RajBhandary, U. L. (1994). Initiator transfer RNAs. *J. Bacteriol.* **176**, 547–552.
- Recsei, P. A., Gruss, A. D. & Novick, R. P. (1987). Cloning, sequence, and expression of the lysostaphin gene from *Staphylococcus simulans*. *Proc. Natl Acad. Sci. USA*, **84**, 1127–1131.
- Richaud, C., Mengin-Lecreux, D., Pochet, S., Johnson, E. J., Cohen, G. N. & Marlière, P. (1993). Directed evolution of biosynthetic pathways. Recruitment of cysteine thioethers for constructing the cell wall of *Escherichia coli*. *J. Biol. Chem.* **268**, 26827–26835.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). In *Molecular Cloning. A Laboratory Manual* (2nd edit.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sprinzl, M., Hartmann, T., Weber, J., Blank, J. & Zeidler, R. (1989). Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.* **17** (Sequences Suppl.), 1–172.
- Sueoka, N. (1961). Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. *Proc. Natl Acad. Sci. USA*, **47**, 1141–1149.
- Takeda, M. & Webster, R. E. (1968). Protein chain initiation and deformylation in *B. subtilis* homogenates. *Proc. Natl Acad. Sci. USA*, **60**, 1487–1494.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.

Edited by J. Karn

(Received 8 July 1996; received in revised form 25 November 1996; accepted 4 December 1996)