Mol Gen Genet (1989) 217:97–104

MGG

Characterization of the *str* operon genes from *Spirulina platensis* and their evolutionary relationship to those of other prokaryotes

F.R. Buttarelli [1], R.A. Calogero [2], O. Tiboni [3], C.O. Gualerzi [1,2], and C.L. Pon [2]

[1] Laboratory of Genetics, DBC, University of Camerino, I-62032 Camerino (MC), Italy
[2] Max-Planck-Institut für Molekulare Genetik, Ihnestraße 73, D-1000 Berlin 33
[3] Department of Genetics and Microbiology, University of Pavia, I-27100 Pavia (PV), Italy

Summary. A 5.3 kb DNA segment containing the *str* operon (ca. 4.5 kb) of the cyanobacterium *Spirulina platensis* has been sequenced. The *str* operon includes the structural genes *rpsL* (ribosomal protein S12), *rpsG* (ribosomal protein S7), *fus* (translation elongation factor EF-G) and *tuf* (translation elongation factor EF-Tu). From the nucleotide sequence of this operon, the primary structures of the four gene products have been derived and compared with the available corresponding structures from eubacteria, archaebacteria and chloroplasts. Extensive homologies were found in almost all cases and in the order S12 > EF-Tu > EF-G > S7; the largest homologies were generally found between the cyanobacterial proteins and the corresponding chloroplast gene products. Overall codon usage in *S. platensis* was found to be rather unbiased.

Key words: Cyanobacteria – DNA sequence – Ribosomal protein – Elongation factors – *str* operon

Introduction

Spirulina platensis is a multicellular, filamentous, helical cyanobacterium. The high content of proteins, vitamins and carotenoids and the low content of nucleic acids as well as other useful properties render the *Spirulina* biomass particularly suitable both as food and feed (Ciferri 1981; Ciferri and Tiboni 1983). A deeper knowledge of the biology and genetics of this organism is, however, a prerequisite for its better biotechnological exploitation since a severe limitation in its widespread utilization is represented by our present ignorance concerning its basic biological, genetic and physiological properties. In fact, so far, no natural or artificial genetic recombination system and no plasmids nor phages are known in this organism. Thus, the only information concerning the organization of the genetic material in *S. platensis* derives from the identification, cloning and expression of some genes involved in essential functions such as nitrogen metabolism (Riccardi et al. 1985), CO_2 fixation (Tiboni et al. 1984a) and protein biosynthesis (Tiboni et al. 1984b; Tiboni and Di Pasquale 1987).

Our long-term goal is to obtain a genetic map of *S. platensis*, to uncover, understand and make use of its natural genetic recombination mechanism(s) and to develop techniques which will allow the transfer and the manipulation of genes in this organism.

In the present paper we report the elucidation of the nucleotide sequence of the *str* operon of *S. platensis* and the homology existing between the primary structures of

Offprint requests to: C.L. Pon

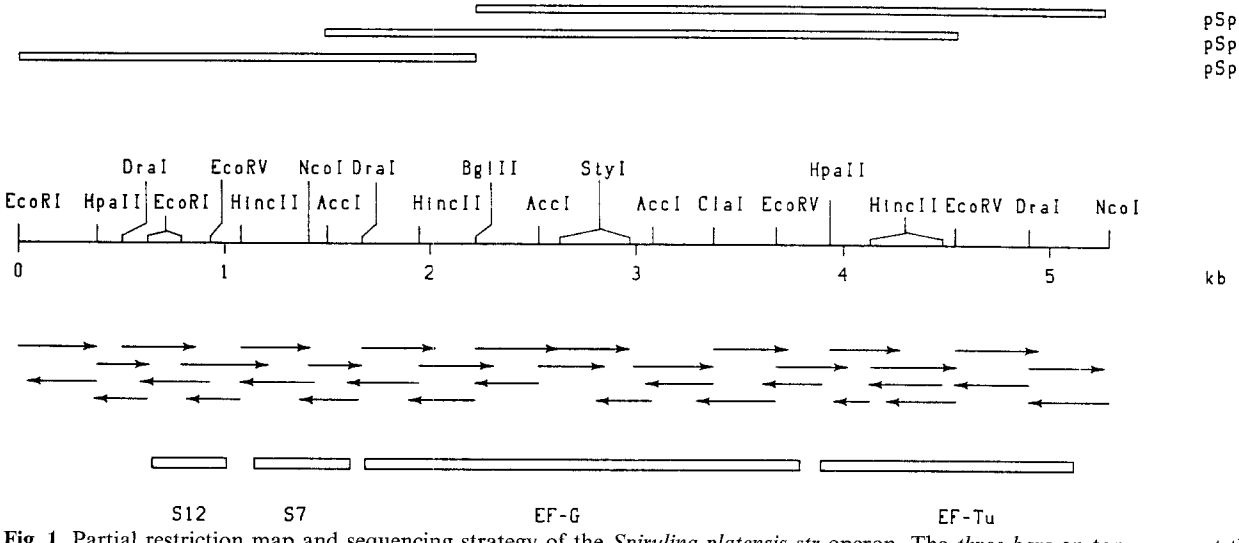

Fig. 1. Partial restriction map and sequencing strategy of the *Spirulina platensis str* operon. The *three bars* on top represent the original inserts cloned in pSp7, pSp18, pSp3 (Tiboni and Di Pasquale 1987). The *arrows* indicate the direction and the extent of the DNA regions sequenced. The restriction sites are only those used for subcloning. The *bars* at the bottom represent the length and the position of the structural genes encoding the indicated proteins

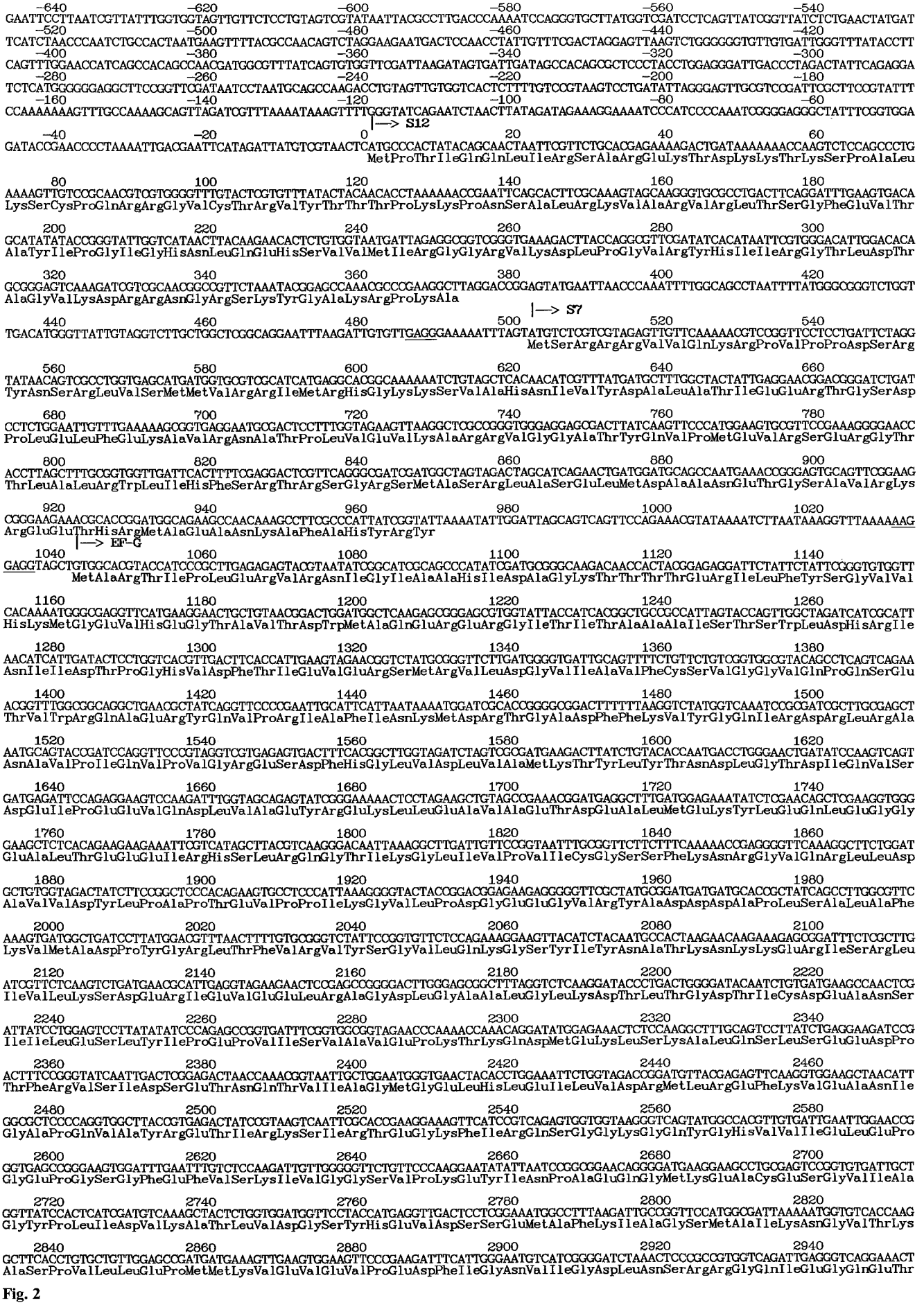

Fig. 2

Fig. 2. Nucleotide sequence of the *S. platensis str* operon and primary structure of the corresponding S12, S7, EF-G and EF-Tu proteins

the corresponding gene products (ribosomal proteins S7 and S12 and elongation factors EF-G and EF-Tu) of *S. platensis* and similar proteins from other prokaryotic sources.

Materials and methods

On the basis of the restriction map previously obtained (Tiboni and Di Pasquale 1987), the inserts carried by the three original plasmids (pSp3, pSp7 and pSp18) were subcloned; the DNA fragments ≤ 1 kb were cloned in M13mp18 and M13mp19, while the fragments >1 kb were cloned in pTZ18 and pTZ19 according to the strategy presented in Fig. 1. The recombinant vectors so obtained were used to transform *Escherichia coli* JM103 cells for the preparation of both double-stranded and single-stranded recombinant DNA. DNA was sequenced by the dideoxy chain termination method (Sanger et al. 1980) using 2′deoxy-7-deazaguanosine triphosphate in place of dGTP, in order to resolve band compression. The data obtained from the sequencing gels were analyzed and processed with the UWGCG 3.0 program on a VAX/VMS 4.5 computer.

Results and discussion

The DNA fragments used for subcloning (Fig. 1) were derived from three recombinant plasmids (pSp3, pSp7 and pSp18) previously constructed (Tiboni and Di Pasquale 1987) by cloning: (a) a 3.4 kb fragment containing the *rpsL* and *rpsG* genes and the 5′-terminal region of *fus* in the *Bgl*II site of pKC7; (b) a 5.8 kb fragment containing *fus* and *tuf* genes in the same *Bgl*II site of pKC7; and (c) a 2.7 kb fragment containing the distal part of *rpsG*, the *fus* gene and the proximal portion of *tuf* in the *Cla*I site of pBR322.

The *str* operon was sequenced by the dideoxy chain terminating method (Sanger et al. 1980) following the subcloning strategy illustrated in Fig. 1. The complete nucleotide sequence of the operon as well as the amino acid sequences derived from it are presented in Fig. 2. The structural genes within the operon are separated by intercistronic regions of 129 (*rpsL–rpsG*), 72 (*rpsG–fus*) and 114 (*fus–tuf*) nucleotides. The intercistronic regions could play important regulatory functions; it is known, for instance, that in the *E. coli str* operon, protein S7 regulates translation by binding to the spacer between the genes encoding S12 and S7 (Nomura et al. 1980). No significant homology was detected, however, between the intercistronic sequences of *S. platensis* and the corresponding ones of *E. coli* or chloroplast. Furthermore, a computer search revealed a putative transcriptional termination signal upstream from the *rpsL* gene (between approx. -80 and -60) but failed to detect any other significant sequence with potential transcriptional termination properties within the intercistronic regions.

Concerning the translational signals, the canonical initiation triplet AUG is found in S12, S7 and EF-Tu, while

```
                    20                  40                  60                  80
SPla  MPTIQQLIRS AREKTDKKTK SPALK..... .........SC PQRRGVCTRV YTTTPKKPNS ALRKVARVRL TSGFEVTAYI
ECol  .A+VN++V+K P+ARKVA+SN V+++E..... .........A+ ++K+++++++ ++++++++++ +++++C++++ +N+++++S++
BSte  .+++N++V+K G+++KVF+S+ ++++NKGYNS FKKEQTNVAS ++K+++++++ G+M+++++++ ++++Y+++++ +N+I++++++
MLut  V++++++V+K G+SPKVVN+N G+++Q..... .........GN +M++++++++ +++++T++++ +V++++++++ NG+I++++++
EuChl +++LEH+T++ P+K+IKR+++ +++++..... .........G+ ++K+AI+M++ ++++++++++ +++++T++++ S++L++++++
ToChl ++++K++++N T+QPIRNV++ ++++R..... .........G+ +++++T++++ ++I+++++++ ++++++++++ +++++I++++
LiChl +++++++++N K+QPIENR++ +++++..... .........G+ ++++++++++ ++++++++++ ++++I+++++ +++++I++++
MaChl +++++++++N K+QPIENRR+ +++++..... .........G+ ++++++++++ ++IN++++++ ++++++++++ +++++I++++

                   100                 120                 139
SPla  PGIGHNLQEH SVVMIRGGRV KDLPGVRYHI IRGTLDTAGV KDRRNGRSKY GAKRPKA..
ECol  G+E+++++++ ++IL++++++ +++++++++T V++A++CS++ +++KQA++++ +V+++++..
BSte  ++++++++++ +++L++++++ ++++++++++ +++A++++++ AN+MQ+++++ +++K+++AK
MLut  ++E+++++++ +I+LV+++++ ++++++++K+ V++A+++Q++ +N+GQA++R+ +++KE+K..
EuChl ++++++++++ +++L++++++ ++++++K++V +++C++A+S+ +N+K+A++++ +V+K++PK.
ToChl ++++++++++ +++LV+++++ ++++++++++ V+++++AV++ +++QQ+++++ +V+K++...
LiChl ++++++++++ +++LV+++++ ++++++++++ ++++++AV++ +++QQ+++++ +V+KS+...
MaChl ++++++++++ +++LV+++++ ++++++++R+ ++++++AVA+ +N+QQ+++++ +++K++K..
```

Fig. 3. Comparison of the primary structures of ribosomal protein S12 from *S. platensis* and the corresponding protein from other prokaryotic sources. Residues identical to those in *S. platensis* are indicated by +. SPla, *S. platensis*; ECol, *Escherichia coli*; Bste, *Bacillus stearothermophilus*; MLut, *Micrococcus luteus*; EuChl, *Euglena* chloroplast; ToChl, tobacco chloroplast; LiChl, liverwort chloroplast; MaChl, maize chloroplast

EF-G begins with a GUG initiation triplet. Furthermore, S12, EF-G and EF-Tu have UAG as the termination codon while S7 terminates with UAA.

The sequence of *S. platensis* 16S rRNA is not known. With the assumption, however, that its 3′ terminus is identical to that of *E. coli* and of other bacterial 16S rRNAs, we searched for regions (Shine-Dalgarno sequences) of potential complementarity between the mRNA translation initiation regions and 16S rRNA. The length of the Shine-Dalgarno (SD) region and of the spacer separating this sequence from the initiation triplet are very different for each gene. Thus, S7 (SD = GAGG, spacer = 12 nucleotides) and EF-G (SD = AAGGAGGT, spacer = 4 nucleotides) seem to have signals similar to those normally found in other bacteria. EF-Tu, on the other hand, does not have any recognizable SD sequence aside from a weak AAG complementarity 18 bases upstream from the initiation codon. This situation is reminiscent of that of chloroplast mRNAs, where SD sequences can be found up to 25 nucleotides upstream from the initiation codon (Ruf and Kössel 1988). Finally, no obvious SD sequence can be found upstream from the initiation codon of S12.

A comparison of the primary sequences of ribosomal proteins S12 and S7 (Figs. 3, 4), of elongation factors EF-G (Fig. 5) and EF-Tu (Fig. 6) from *S. platensis* with those of other prokaryotic sources known so far is presented so as to highlight the sequence homologies. As seen from the figures and from the quantitative data presented in Table 1, the homology between the *S. platensis* proteins and the equivalent ones from the other sources (with the exception of the archaebacterium) are quite large. The most conserved proteins are S12 and EF-Tu with >60% identical residues

```
                    20                  40                  60                  80
SPla  MSRRRVVQKT VPVPPDSRYN SRLVSMMVRR IMRHGKKSVA HNIVYDALAT IEERTGS.DP LELFEKAVRN ATPLVEVKAR
ECol  .P++++IGQ. RKIL++PKFG +E+LAKF+NI L+VD++++T+ ES+++S++E+ LAQ+S+K.SE ++A++V+LE+ VR+T++++S+
BSte  .P++GP+A+. RD+L++PI++ +K++TRLINK ++ID++++K+ QK+L+T+FDI +R++++K.++ MSV++Q+LK+ VM+VL++R++
MLut  +P+KGPAP+. R+LVV+PV+G +P++TQLINK VLVD++++T+ ER+++G++EG ARAKN+ARSR GHPIK++MD+ IK+AL++RS+
EuChl +++++RAK+. RIISQ+PI++ +T+A+KVINK +LLN+++TL+ QY+F+ETMKN +Q+IYKK.++ +DILR++IK+ +S+QM+TRK+
ToChl ++++GTAE+. KTAKS+PI+R N+++N+L+N+ +LK+++++L+ YQ+I+R+VKK +QQK+ET.N+ +SVLRQ+I+G V++DIT++++
LiChl +++KSIAE+. QVAK++PI+R N+++N+L+N+ +LKN++++L+ YR+L+K+MKN +KQK+KK.N+ +FVLRQ+++K V++N+T++++
MaChl ++++GTAE+. RTAKS+PIFR N+++N+V+N+ ++KD++++L+ YQ+L+R+VKK +QQK+ET.N+ +LVLRQ+I+R V++NIG++T+

                   100                 120                 140                 159
SPla  RVG.GATYQV PMEVRSERGT TLALRWLIHF SRTRSGRSMA SRLASELMDR ANETGSRVRK REETHRMAEA NKAFAHNRY
ECol  +++.+S++++ +V++.PV+RN A++M++IVEA A+K+GDK+++ L+++N++S+A +ENK+TA+K+ ++DV++++++ ++++++Y+.
BSte  +++.++N+++ +V+++PD+RV S+G++++VQY A+L+GEKT+E E+++N+I++A ++N++RT+K+ ++D++K++++ ++++++Y+W
MLut  +++.++++++ +V++KPG+S+ A++++++VG+ +KA+REKT+T E++MN+IL+A S+GL+GA+KR ++D++K++++ ++++++Y+W
EuChl +I+.+TI+++ +V++KED+++ S+++KFI+EK A+E+K++GIS TK+KN+II+A S+N++EA+K+ K++I+KT+++ ++++SNMKF
ToChl +++.+S+H++ +I+IG+TQ+K A++I+++LAA ++K+P++N++ FK+S+++V+A +KGS+DAI++ K+++++++ +R++++F+.
LiChl +ID.+S++++ +L+IK+TQ+K A++I+++LGA ++K+++QN++ FK+SY++I+A +RDN+IAI++ K++++K+++ +R++++F+.
MaChl +NKK+S+RK+ +++IG+KQ+R A++I+++LEA +QK+P++N++ FK+S+++V+A +KGS+GAI++ K+A+++++ +R+L++F+.
```

Fig. 4. Comparison of the primary structures of ribosomal protein S7 from *S. platensis* and the corresponding protein from other prokaryotic sources. Residues identical to those in *S. platensis* are indicated by +. For the abbreviations, see legend to Fig. 3

```
                    20                  40                  60                  80
                    .                   .                   .                   .
SPla  MARTIPLERV RNIGIAAHID AGKTTTTERI LFYSGVVHKM GEVHEGTAVT DWMAQERERG ITITAAAIST SWLD.....
ECol  ++++T+IA+Y +++++S++++ ++++++++++ +++T++N++I ++++D+A+TM +++E++Q+++ ++++S++TTA F+SGMAKQYE
MLut  ..VLTD+HK+ +++++M++++ +++++++++H +++T++N++L ++T+D+G+T+ +++E++K+++ ++++S++VTC F+N+.....

                   100                 120                 140                 160
                    .                   .                   .                   .
SPla  .HRINIIDTP GHVDFTIEVE RSMRVLDGVI AVFCSVGGVQ PQSETVWRQA ERYQVPRIAF INKMDRTGAD FFKVYGQIRD
ECol  P+++++++++ ++++++++++ ++++++++AV M+Y+A+++++ ++++++++++ NK+K++++++ V+++++M++N +L++VN++KT
MLut  .+Q+++++N+ ++++++V+++ ++L+++++AV +++DGKE++E ++++++++++ DK+D++++C+ V++++KL+++ +YFTVDT+VK

                   180                 200                 220                 240
                    .                   .                   .                   .
SPla  RLRANAVPIQ VPVGRESDFH GLVDLVAMKT YLYTNDLGTD IQVSDEIPEE VQDL....VA EYREKLLEAV AETDEALMEK
ECol  ++G++P++L+ LAI+A+EH+T +V++++K++A INWNDADQGV TFEYED++AD MVE+....AN +WHQN+I+SA ++AS+E++++
MLut  ++G+RPLVM+ L+I+A+N++V +V+++IS++A FVWPG+ANGI VTMGASYEI+ IRQ+QEKAEE +++NE+V+++ +++S+E++++

                   260                 280                 300                 320
                    .                   .                   .                   .
SPla  YLEQLEGGEA LTEEEIRHSL RQGTIKGLIV PVICGSSFKN RGVQRLLDAV VDYLPAPTEV PPIKGVLPDG EE..GVRYAD
ECol  ++....+++E +++A++KGA+ ++RVLNNE+I L+T+++A+++ K+++AM++++ I++++S+VD+ +A+N+I+D++ KDTPAE+H+S
MLut  +++....++E ++V+++QAGV ++L+VNAEAY ++F+++A+++ ++++PM++++ +A+++N+LDA G+V++HAVND ++VVLE+EVS

                   340                 360                 380                 400
                    .                   .                   .                   .
SPla  DDAPLSALAF KVMADPY.GR LTFVRVYSGV LQKGSYIYNA TKNKKERISR LIVLKSDERI EVEELRAGDL GAALGLKDTL
ECol  ++E+F+++++ +IAT++FV+N +++F++++++ VNS+DTVL+S V+AAR++FG+ IVQMHANK+E +IK+V++++I A++I++++VT
MLut  KE++F+++++ +IATH+FF+T +++I+++++R +ES+AQVL++ ++G+++++GK +FQMHANKEN P+D+VV++HI Y+VI+++++T

                   420                 440                 460                 480
                    .                   .                   .                   .
SPla  TGDTICDEAN SIILESLYIP EPVISVAVEP KTKQDMEKLS KALQSLSEED PTFRVSIDSE TNQTVIAGMG ELHLEILVDR
ECol  ++++L++PDA P++++RMEF+ +++++I++++ +++A+Q++MG L++GR+AK++ +S+++WT+E+ S+++I+++++ ++++D+I+++
MLut  ++++L++P++ P+++++MTF+ +++++++I++ +++G+Q++++ T+I+K+VA++ +++++NLNE+ +G++E+G+++ ++++DVF+++

                   500                 520                 540                 560
                    .                   .                   .                   .
SPla  MLREFKVEAN IGAPQVAYRE TIRKSI.RTE GKFIRQSGGK GQYGHVVIEL EPGEPGS... GFEFVSKIVG GSVPKEYINP
ECol  +K+++N++++ V+K+++++++ +++QKVTDV+ ++HAK++++R ++++++++DM Y+L++++NPK +Y++IND+K+ +VI+G+++PA
MLut  +K++++++++ V+K+++++++ ++KRKVDKVD YTHKK+T++S ++FAK+QLSF ++LD.TPRGT VY++ENA+T+ +R++R+++PS

                   580                 600                 620                 640
                    .                   .                   .                   .
SPla  AEQGMKEACE SGVIAGYPLI DVKATLVDGS YHEVDSSEMA FKIAGSMAIK NGVTKASPVL LEPMMKVEVE VPEDFIGNVI
ECol  VDK+IQ+QLK A+PL++++VV +MGIR+HF++ ++D+++++L+ ++L+A+I+F+ E+FK++K+++ +++I++++++ T++ENT+D++
MLut  VDA+IQD+MK F++L++++MV R++++SL++A ++D+++++++ +R++++Q+F+ E++R++T+II +++L+A+++R T++E+M+D++

                   660                 680                 700                 715
                    .                   .                   .                   .
SPla  GDLNSRRGQI EGQETDQSQS IAKVVAKVPL ATMFGYATDI RSKTQGRGVF SMEFSHYEEV PRSVAETIIA KSKGN
ECol  +++SR+++ML K+++SEVTG. .V+IH+E+++ SE++++++QL ++L+K++ASY T+++LK+D+A +SN++QAV+E ARGK.
MLut  ++++++++++ QI+SMEDATG VKV+N+L+++ SE++++IG+L +++++++A+Y ++T+HS+A++ +KA++DE+VQ ++Q+E
```

Fig. 5. Comparison of the primary structures of translation elongation factor EF-G from *S. platensis* and the corresponding protein from other prokaryotic sources. Residues identical to those in *S. platensis* are indicated by +. For the abbreviations, see legend to Fig. 3

suggesting that these proteins are subject to more strict structural constraints. Also, in the case of EF-G and S7, however, the percentage of identical residues is high and, in general, >50%. The homology of the *Spirulina* proteins, at least in the case of S12 and EF-Tu, is significantly greater with the chloroplast than with the bacterial counterparts. This finding is in good agreement with the conclusion that cyanobacteria and green chloroplasts form a coherent phylogenetic group and that the chloroplast lineage is contained within the cyanobacteria radiation, rather than being a sister group of these free-living organisms (Giovannoni et al. 1988).

The comparison of the primary structures of the most conserved protein (S12; Fig. 3) and of the least conserved protein (S7; Fig. 4) might be relevant for the identification of the regions of these molecules potentially important for function and/or structure. Thus, in the case of S12, there is a remarkable degree of sequence identity in the central region of the molecule where only very few amino acid changes, mainly of the conservative type, occurred. This region includes the amino acids found to influence translational fidelity and streptomycin sensitivity in *E. coli* (Funatsu et al. 1977). In contrast, many amino acid substitutions are seen in both the N-terminal and C-terminal domains

```
                    20                  40                  60                  80
SPla  MARAKFERNK PHVNIGTIGH VDHGK.TTLT AAITMTLAAS GGAKARK... .......... .YDDIDAAPE EKQRGITINT
ECol  .SKE++++T+ ++++V+++++ +++++.++++ ++++TV++KT Y+GA++A... .......... .F+Q++N+++ ++A+++++++
MeVa  ++KT.....+ +IL+VAF+++ ++A++S++VG RLLLDGG+ID PQLIV+LRKE AEEKGKAGFE FAYVM+GLK+ +RE++V++DV
MLut  ++K+++++T+ A+++++++++ +++++.++++ +++SKV+YDK YPDLNEAR.D .......... FA.T++S+++ +R++++++I
TThe  ++KGE+V+T+ ++++V+++++ +++++.++++ ++L+YVA++E NPNV.E.VKD .......... .+G+++K+++ +RA+++++++
EuChl +++Q++++T+ ++I+++++++ +++++.++++ +++++A+++T +NS++KR... .......... .+E+++S+++ ++A+++++++

                   100                 120                 140                 160
SPla  AHVEYETEQR HYAHVDCPGH ADYVKNMITG AAQMDGAILV VSAAD...GP MPQTREHILL AKQVGVPSIV VFLNKADMVD
ECol  S++++D+PT+ ++++++++++ ++++++++++ ++++++++++ +A+T+...++ ++++++++++ GR+++++Y+I +++++C++++
MeVa  ++KKFP+AKY EVTI++++++ R+FI++++++ +S+A+A+V++ +NVD+AKS+I Q++++++VF+ IRTL++RQLA +AV++M+T+N
MLut  S++++Q++K+ ++++++A+++ +++I++++++ ++++++++++ +A+T+...++ +A+++++V++ +R+++++ALL +A+++S+++E
TThe  +++++++AK+ ++S+++++++ +++I++++++ ++++++++++ +++++...++ ++++++++++ +R+++++Y++ ++M++V++++
EuChl +++++++KN+ ++++++++++ ++++++++++ ++++++++++ +++++...++ ++++K+++++ +++++++N++ +++++E+Q++

                   180                 200                 220                 240
SPla  ..DEELLELV ELEVRELLSS YDFPGDDIPI VSGSALKALD FL..TE.NPK TTRGENDWVD KIHALMDEV. DAYIPTPERD
ECol  ..++++++++ +M+++++++Q +++++++T++ +R+++++++E .......... ...+DAE+EA ++LE+.AGFL +S+++E+++A
MeVa  FSEADYN++K KMIGDQ++KM IG+NPEQ+NF +PVAS+HGDN VFKKS+R++. .......+YK G:.PTIA++I +GFQ+.++KP
MLut  ..++++++R+ +M+++Q++++ RS+DV+EA+V IRT++++++E .......... ...+DPQ++K SVED+++A+. +E+++D+V++
TThe  ..+P+++D++ +M+++D++NQ +E++++EV+V IR++++L++E QM.H..R+++ +R++++E+++ ++WE+L+AI. +E+++++V++
EuChl ..+S++++++ +++I++T++N +E+++++++V IP++++LSVE A+TK...+++ I+K+++K+++ ++LN+++Q+. +S+++++T++

                   260                 280                 300                 320
SPla  IDKGLLDGLW EDVFSITGRG TVSTAGIERG KVKVGDTVEL IGI.KDTRTT TVTGAEMFQK TLEEGMAGDN VGLLLRGIQK
ECol  +++PF+LPI. ++++++S+++ ++V+GRV+++ II+++EE++I V++.+E+QKS +C++V+++R+ L+D++R++E+ ++V+++++KR
MeVa  TNLP+RLPI. Q++YT+++V+ ++PVGRV+T+ II+P++K+VF EPAGAIGEIK ++...++HHE Q+PSAEP+++ I+FNV++VG+
MLut  K++PF+MPI. ++++T+++++ ++V+GRA+++ TL+INSE++I V++.R+VQK+ ++++I+++H+ Q+D+AW++E+ C+++V++LKR
TThe  V++PF+MPV. ++++T+++++ ++A+GR++++ ++++++E++I V+LAPE++R+ V+++V++HR+ ++Q++I++++ ++V++++VSR
EuChl TE+DF+MAI. +++L++++++ ++A+GRV+++ TI+++E++++ V+L.++++S+ +I++L+++++ S+D+AL++++ ++V+++++++

                   340                 360                 380                 400
SPla  NDVQRGMVIA .KPKSITPHT KFEAEVYILK KEEGGRHTPF FKGYRPQFYV RTTDVTGTID EFTADDGSTP EMVIPGDRIN
ECol  EEIE++Q+L+ .++GT+K+++ +++S+++++S +D++++++++ +++++++P+F ++++++++E LPEG.....V +++M+++N+K
MeVa  K+IK++D+LG HTTNPP+VA+ D+T+QIVV+Q .....HPSVL TD++T+V+HT H+AQIAC+FA +IQKKLNPAT GE+LEENPDF
MLut  D++E++Q+LV .E+G++++++ N+++N++++S +D++++++++ YSN++A+++F +++++++V+T LPEG.....T +++M+++TTE
TThe  EE+E++Q+L+ .++G++++++ ++++S++V++ ++++++++G+ +S+++++++F +++++++VVQ LPPG.....V +++M+++NVT
EuChl +++E++++L+ .++RT+N+++ ++DSQ++++T ++++++++++ +E++++++++ +++++++K+E S+RSDNDNPA Q++M+++++K

                   420                 440             453
SPla  MTVQLICPIA IEQGMRFAIR EGGRTVGAGV VAKILA..... ..........  ...
ECol  +V+T++H+++ MDD+L+++++ ++++++++++ +++V+S..... ..........  ...
MeVa  LKAGDAAIVK LIPTKPMV+E SVKEIPQL+R F+IRDMGMTV AAGMAIQVTA KNK
MLut  +S+E++Q+++ M+E+LG++++ +++++++S+R +T++TK..... ..........  ...
TThe  F++E++K+VG L+E+L+++++ ++++++++++ +T+++E..... ..........  ...
EuChl +K+E++Q+++ ++K+++++++ ++++++++++ +LS+IQ..... ..........  ...
```

Fig. 6. Comparison of the primary structures of translation elongation factor EF-Tu from *S. platensis* and the corresponding proteins from other prokaryotic sources. Residues identical to those in *S. platensis* are indicated by +. For the abbreviations, see legend to Fig. 3. Additional abbreviations: MeVa, *Methanococcus vannielii*; TThe, *Thermus thermophilus*

of S12 suggesting less stringent structural constraints on these regions of the molecule. In the case of S7, the presence of highly conserved regions within a primary structure in which many amino acid replacements have occurred in the course of evolution, possibly identifies functionally important domains. One of these is the positively charged peptide comprising approximately 20 amino acids at the C-terminus. Other highly conserved regions include the Gly_{35}-Lys-Lys-Ser peptide and the central region of the molecule which includes several basic amino acid residues and where most of the amino acid substitutions are of the conservative type. It is possible that these positively charged regions may constitute the 16S rRNA binding site of this protein. Furthermore, with the exception of S7 of *Euglena* chloroplasts, all other S7 molecules have conserved the single tryptophan found in *E. coli* B (position 101) and the methionine (position 114) found to be the point of UV-induced cross-linking between *E. coli* S7 and 16S rRNA (Möller et al. 1978). A peculiar feature of *S. platensis* S7 is the Arg/Lys ratio which is substantially higher (3.6) than that of S7 from all the other sources, which ranges between 0.5 and 1.3. Finally, comparison of the sequences of the two elongation factors (Figs. 5, 6) reveals the presence of several regions of strict conservation of the primary structure which

Table 1. Percentage of identical residues between the gene products of the *Spirulina platensis str* operon and the corresponding proteins from other prokaryotic sources

	S12	S7	EF-G	EF-Tu
Escherichia coli	73.2	52.2	57.9	69.4
Bacillus stearothermophilus	72.8	57.1	–	–
Micrococcus luteus	71.5	48.1	57.0	63.7
Thermus thermophilus	–	–	–	69.9
Methanococcus vannielii	–	–	–	25.9
Euglena chloroplast	75.8	44.9	–	77.5
Tobacco chloroplast	81.3	51.3	–	–
Liverwort chloroplast	79.7	52.3	–	–
Maize chloroplast	81.3	48.7	–	–

The sequence data of the following organisms were used: *E. coli* (Post and Nomura 1980; Reinbolt et al. 1978; Yokota et al. 1980; Zengel et al. 1984), *B. stearothermophilus* (Kimura and Kimura 1987; M. Kimura, personal communication), *M. luteus* (Ohama et al. 1987), *T. thermophilus* (Kushiro et al. 1987), *M. vannielii* (Lechner and Böck 1987), *Euglena* chloroplast (Montandon and Stutz 1983; Montandon et al. 1987), tobacco chloroplast (Fromm et al. 1986; Torazawa et al. 1986; EMBL Data Bank), liverwort chloroplast (Ohyama et al. 1986), maize chloroplast (Giese et al. 1987)

include, in particular, the peptides (Gly_{18}-His-Val-Asp-His-Gly-Lys-Thr-Thr; Val_{79}-AspCys-Pro-Gly-His; Asn_{135}-Lys-X-Asp) which have been found to constitute the GDP binding domain of *E. coli* EF-Tu (Jurnak 1985; La Cour et al. 1985).

The codon usage derived from the coding regions of the *str* operon of *S. platensis* is presented in Table 2. It has been noticed that in chloroplast genes there is a strong preference for synonymous codons with A or T in the third position (Markmann-Mulisch and Subramanian 1988). In *S. platensis*, this tendency is confirmed only for codons ending with T but not for those ending with A which are the least abundant. Overall, however, the data of Table 2 show that although a few codons are encountered rarely, codon usage in *Spirulina* is rather unbiased.

Acknowledgements. This work was supported in part by funds of the "Ente di Sviluppo della Regione Marche" and of the Italian Ministry of Public Education to C.O.G. The help of Dr. M. Brombach in the preparation of Fig. 1 as well as with the computer work is gratefully acknowledged. We are also grateful to Dr. M. Kimura for making available to us unpublished sequence data from *Bacillus stearothermophilus.*

Table 2. Codon usage within the coding region of the *str* operon of *S. platensis*

Gly	GGG	0.22	Arg	AGG	0.08	Trp	TGG	1.00	Arg	CGG	0.27
Gly	GGA	0.18	Arg	AGA	0.04	End	TGA	0.00	Arg	CGA	0.08
Gly	GGT	0.45	Ser	AGT	0.20	Cys	TGT	0.63	Arg	CGT	0.29
Gly	GGC	0.15	Ser	AGC	0.07	Cys	TGC	0.38	Arg	CGC	0.24
Glu	GAG	0.34	Lys	AAG	0.40	End	TAG	0.75	Gln	CAG	0.52
Glu	GAA	0.66	Lys	AAA	0.60	End	TAA	0.25	Gln	CAA	0.48
Asp	GAT	0.75	Asn	AAT	0.47	Tyr	TAT	0.82	His	CAT	0.33
Asp	GAC	0.25	Asn	AAC	0.53	Tyr	TAC	0.18	His	CAC	0.67
Val	GTG	0.27	Met	ATG	1.00	Leu	TTG	0.23	Leu	CTG	0.32
Val	GTA	0.25	Ile	ATA	0.03	Leu	TTA	0.10	Leu	CTA	0.14
Val	GTT	0.37	Ile	ATT	0.59	Phe	TTT	0.47	Leu	CTT	0.09
Val	GTC	0.11	Ile	ATC	0.38	Phe	TTC	0.53	Leu	CTC	0.13
Ala	GCG	0.19	Thr	ACG	0.12	Ser	TCG	0.14	Pro	CCG	0.32
Ala	GCA	0.18	Thr	ACA	0.15	Ser	TCA	0.16	Pro	CCA	0.08
Ala	GCT	0.43	Thr	ACT	0.49	Ser	TCT	0.27	Pro	CCT	0.34
Ala	GCC	0.19	Thr	ACC	0.24	Ser	TCC	0.16	Pro	CCC	0.26

References

Ciferri O (1981) *Spirulina*, the edible microorganism. Microbiol Rev 47:551–578

Ciferri O, Tiboni O (1983) The biochemistry and industrial potential of *Spirulina*. Annu Rev Microbiol 39:503–526

Fromm H, Edelman M, Koller B, Goloubinoff P, Galun E (1986) The enigma of the gene coding for ribosomal protein S12 in the chloroplasts of Nicotiana. Nucleic Acids Res 14:883–898

Funatsu G, Yaguchi M, Wittmann-Liebold B (1977) Primary structure of protein S12 from the small *Escherichia coli* ribosomal subunit. FEBS Lett 73:12–17

Giese K, Subramanian AR, Larrinua IM, Bogorad L (1987) Nucleotide sequence, promoter analysis, and linkage mapping of the unusually organized operon encoding ribosomal proteins S7 and S12 maize chloroplast. J Biol Chem 262:15251–15255

Giovannoni S, Turner S, Olsen GJ, Barns S, Lane DJ, Pace N (1988) Evolutionary relationships among cyanobacteria and green chloroplasts. J Bacteriol 170:3584–3592

Jurnak F (1985) Structure of the GDP domain of EF-Tu and location of the amino acids homologous to *ras* oncogene proteins. Science 230:32–36

Kimura M, Kimura J (1987) The complete amino acid sequence of ribosomal protein S12 from *Bacillus stearothermophilus*. FEBS Lett 210:91–96

Kushiro A, Shimizu M, Tomita K (1987) Molecular cloning and sequence determination of the *tuf* gene coding for the elongation factor Tu of *Thermus thermophilus* HB8. Eur J Biochem 170:93–98

La Cour TFM, Nyborg J, Thirup S, Clark BFC (1985) Structural details of the binding of guanosine diphosphate to elongation factor Tu from *E. coli* as studied by X-ray crystallography. EMBO J 4:2385–2388

Lechner K, Böck A (1987) Cloning and nucleotide sequence of the gene for an archaebacterial protein synthesis elongation factor Tu. Mol Gen Genet 208:523–528

Markmann-Mulisch U, Subramanian AR (1988) Nucleotide sequence of maize chloroplast *rps11* with conserved amino acid

sequence between eukaryotes, bacteria and plastids. Biochem Int 17:655–664

Möller K, Zwieb C, Brimacombe R (1978) Identification of the oligonucleotide and oligopeptide involved in an RNA-protein cross-linking induced by ultraviolet irradiation of *Escherichia coli* 30S ribosomal subunits. J Mol Biol 126:489–506

Montandon PE, Stutz E (1983) Nucleotide sequence of a *Euglena gracilis* chloroplast genome region coding for the elongation factor Tu: evidence for a spliced mRNA. Nucleic Acids Res 11:5877–5892

Montandon PE, Knuchel-Aegerter C, Stutz E (1987) *Euglena gracilis* chloroplast DNA: the untranslated leader of *tuf*A-ORF206 gene contains an intron. Nucleic Acids Res 15:7809–7822

Nomura M, Yates JL, Dean D, Post LE (1980) Feedback regulation of ribosomal protein gene expression in *Escherichia coli:* structural homology between ribosomal RNA and ribosomal protein mRNA. Proc Natl Acad Sci USA 77:7084–7088

Ohama T, Yamao F, Muto A, Osawa S (1987) Organization and codon usage of the streptomycin operon in *Micrococcus luteus*, a bacterium with a high genomic G+C content. J Bacteriol 169:4770–4777

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota SI, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Plant Mol Biol Rep 4:148–175

Post LE, Nomura M (1980) DNA sequences from the *str* operon of *Escherichia coli*. J Biol Chem 255:4660–4666

Reinbolt J, Tritsch D, Wittmann-Liebold B (1978) The primary structure of ribosomal protein S7 from *E. coli* strains K and B. FEBS Lett 91:297–301

Riccardi G, De Rossi E, Della Valle G, Ciferri O (1985) Cloning of the glutamine synthetase gene from *Spirulina platensis*. Plant Mol Biol 4:133–136

Ruf M, Kössel H (1988) Occurrence and spacing of ribosome recognition sites in mRNAs of chloroplasts from higher plants. FEBS Lett 240:41–44

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143:161–178

Tiboni O, Di Pasquale G (1987) Organization of genes for ribosomal proteins S7 and S12, elongation factors EF-Tu and EF-G in the cyanobacterium *Spirulina platensis*. Biochim Biophys Acta 908:113–122

Tiboni O, Di Pasquale G, Ciferri O (1984a) Cloning and expression of the genes for ribulose-1,5-bisphosphate carboxylase from *Spirulina platensis*. Biochim Biophys Acta 783:258–264

Tiboni O, Di Pasquale G, Ciferri O (1984b) Two *tuf* genes in the cyanobacterium *Spirulina platensis*. J Bacteriol 159:407–409

Torazawa K, Hayashida N, Obokata J, Shinozaki K, Sigiura M (1986) The 5′ part of the gene for the ribosomal protein S12 is located 30 kbp downstream for its 3′ part in tobacco chloroplast genome. Nucleic Acids Res 14:3143

Yokota T, Sugisaki H, Takanami M, Kaziro Y (1980) The nucleotide sequence of the cloned tufA gene of *Escherichia coli*. Gene 12:25–31

Zengel JM, Archer RH, Lindahl L (1984) The nucleotide sequence of the *Escherichia coli fus* gene coding for elongation factor G. Nucleic Acids Res 12:2181–2192

Communicated by K. Isono

Received October 14, 1988