Volume 12 Number 6 1984 Nucleic Acids Research

The genes for the ribosomal proteins S12 and S7 are clustered with the gene for the EF-Tu protein on the chloroplast genome of *Euglena gracilis*

Paul-Etienne Montandon and Erhard Stutz

Laboratoire de Biochimie, Université de Neuchâtel, Ch. de Chantemerle 18, CH-2000 Neuchâtel, Switzerland

Received 30 January 1984; Accepted 27 February 1984

ABSTRACT

We characterize a DNA segment of the Euglena gracilis chloroplast DNA fragment Eco·N by nucleotide sequencing and S1 nuclease analysis. We show that this region, which is upstream of the previously sequenced tuf A gene, contains the genes for the ribosomal proteins S12 and S7. The gene arrangment is 5'-rps 12-80 bp spacer-rps 7-174 bp spacer-tuf A, somewhat similar to the str operon of E. coli. The chloroplast S12 and S7 proteins contain 124 and 155 aminoacids, respectively, and are to 68% and 38% homologous with the corresponding E. coli proteins. The region is transcribed into a distronic mRNA of about 1.1 to 1.2 kb. The rps 12 and rps 7 genes, contrary to the tuf A gene, are not split.

INTRODUCTION

We have recently sequenced the Euglena gracilis chloroplast gene for the elongation factor EF-Tu (tuf A) (1) which was previously mapped on the DNA fragment Eco·N (2). The chloroplast tuf A gene is expressed as a 1.95 kb mRNA which is the product of three splicing events (1). The coding part of this chloroplast gene is to about 55% homologous with the E. coli tuf A gene which is known to be part of a large operon containing in the 5' to 3' direction the genes for the small subunit ribosomal proteins S12 and S7 (rps 12, rps 7), the elongation factor EF-G (fus) and the elongation factor EF-Tu (tuf A) (3). The bacterial tuf A (4), rps 12 and parts of rps 7 have been sequenced (5) as well as the S12 (6) and S7 proteins (7). The S12 protein is of particular interest since it is involved in the translation initiation step and controls the fidelity of the reading process (8). Streptomycin interacts with S12 abolishing normal protein synthesis (9 and references there). Streptomycin also interacts with chloroplast ribosomes, preventing normal chloroplast growth what can lead to irreversibly bleached cells. Streptomycin resistant Euglena strains have been described (10,11), i.e.,

© IRL Press Limited, Oxford, England.

streptomycin resistance is most likely a convenient chloroplast bound genetic marker.

Considering these facts it became of interest to sequence the 5' flanking region of the tuf A gene and to size its stable transcript. We show in this report that adjacent and upstream of the tuf A gene there are in the 5' to 3' direction the rps 12 and rps 7 genes. The two genes are transcribed into a dicistronic mRNA. A preliminary account of this work was given at the eleventh A. Katzir-Katchalsky conference, Jerusalem, 1984.

MATERIALS

Enzymes were purchased from Boehringer-Mannheim and used following instructions of the supplier. [α-^{32}P]-ATP (400 Ci/mmole) was from Radiochemical Center, Amersham.

METHODS

Cell culture and preparation of chloroplast RNA

E. gracilis (Z-strain) was grown as published (1). Chloroplast RNA was extracted (12) from purified chloroplasts (13).

DNA cloning

The chloroplast DNA fragment Eco·N was previously cloned into pBR322 (14). Selected subfragments of Eco·N were isolated according to published procedures (15).

S1 nuclease mapping and DNA sequencing

S1 nuclease protection analysis was according to (16,17). Protected DNA fragments were detected as described (18,1). For DNA sequencing subfragments of Eco·N were blunt ended with DNA polymerase (Klenow), cloned into the HincII site of M13 mp 9 (19,20) and sequenced according to (21). DNA fragments with HhaI ends were blunt ended with S1 nuclease (22).

RESULTS AND DISCUSSION

1. Mapping and sequencing of the rps 12 and rps 7 genes

We have totally sequenced the chloroplast DNA fragment Eco·N (3067 positions) and we found upstream and in close vicinity to the previously analyzed tuf A gene (1) the chloroplast rps 12 and rps 7 genes. As shown in Fig. 1 the overall gene arrangment is 5'-rps 12-rps 7-tuf A what is very similar to the situation found in the str operon of E. coli (5), the crucial difference being that the chloroplast genome lacks the fus gene. The gene for

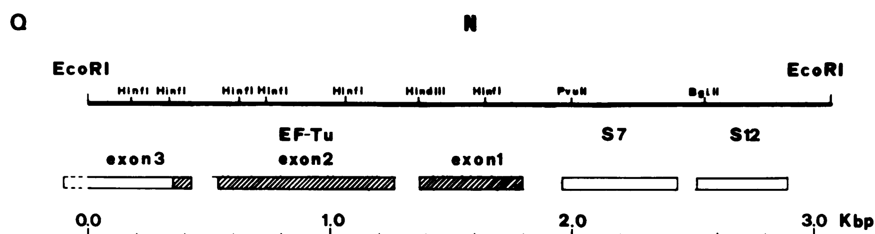

Fig. 1. Arrangment of the genes for the chloroplast ribosomal proteins S12 (rps 12), S7 (rps 7) and the EF-Tu protein (tuf A) on the DNA fragment Eco·N. The three genes have the same polarity starting 5' with the gene for S12. The coding regions are boxed. In order to correlate the sequence data given in Fig. 2 with the results previously reported (1) we show also the split tuf A gene (exon 1, exon 2, part of exon 3). [-- indicates continuation of an ORF on Eco·Q (1). Some restriction sites are given as position markers.

the Euglena chloroplast EF-G protein seems to be located in the nuclear genome (23); our data support this assumption indirectly. It is noteworthy that the spinach chloroplast genome seems to contain the genetic information for the EF-G protein (24).

In Fig. 2 we give the nucleotide sequence of the relevant DNA region. The sequence starts with the 5' end of the EcoRI site (N/J') and continues towards the tuf A gene, including 28 codons of the first exon of tuf A. The aminoacid sequence is given for the coding parts and aligned with it is the aminoacid sequence of the corresponding E. coli proteins.

The chloroplast rps 12 gene has an open reading frame for 124 aminoacids (not counting the initiation codon). The deduced chloroplast protein is one position longer than the bacterial S12 protein. The chloroplast S12 protein has a calculated Mr of 13'828 and is to 68% homologous with the bacterial protein. We find several important conserved regions, especially the domains around position 42 (Lys) and 87 (Lys) which are involved in streptomycin binding (9). Streptomycin resistant strains of E. gracilis were analyzed and compared with the wild type strain on a ribosomal protein level (11). The strain Sm^r_2 BNgL seems to contain a modified acidic protein in the small ribosomal subunit (two-dimensional gel-electrophoresis). Since the Euglena chloroplast S12 protein (Fig. 2) like the E. coli S12 protein belongs to the basic ribosomal proteins the observed resistance of strain Sm^r_2 BNgL seems not to be due to a mutated S12 protein. The analytical data need, however, to be verified. Future tests will show whether the same type of mutations as

```
↓
GAATTCTTTTAGAATATAAATATAAATATAAAAAAGTAAATTTATAAATTGTTATTTCGAAAGAAAACCTACGCACATA
        10        20        30        40        50        60        70        80

ACAAAAAGATGATTAAAAAAAAAAAACGCAAGAAAAAAATAGATAAAAATATTTTAAAGCAAAATAAATATTACCCTAAAAT
        90       100       110       120       130       140       150       160

                     *  A  *  V  N  Q  *  V  *  K  *  *  A  R  K  V  A  *  S  N  V
                     M  P  T  L  E  H  L  T  R  S  P  R  K  K  I  K  R  K  T  K  S
AGAATAAAAAAACTAAAATGCCTACATTAGAACATTTAACACGATCACCGAGAAAAAAAATAAAACGAAAAACTAAATCA
       170       180       190       200       210       220       230       240
                                                                  42
 *  *  *  E  A  *  *  *  *  *  G  V  *  T  *  *  *  *  *  *  *  *  *  *  *  *  *
 P  A  L  K  G  C  P  Q  K  R  A  I  C  M  R  V  Y  T  T  T  P  K  K  P  N  S  A
CCAGCATTAAAAGGATGCCCGCAAAAACGCGCAATATGCATGCGGGTTTACACAACAACACCAAAAAAACCAAATTCCGC
       250       260       270       280       290       300       310       320

  *  *  *  *  C  *  *  *  *  T  N  *  F  *  *  *  S  *  *  G  *  E  *  *  *  *
  L  R  K  V  T  R  V  R  L  S  S  G  L  E  V  T  A  Y  I  P  G  I  G  H  N  L
CTTGCGCAAAGTAACAAGGGTAAGACTTTCTTCAGGGTTAGAAGTTACAGCTTATATACCAGGAATAGGGCATAATTTAC
       330       340       350       360       370       380       390       400
                                          87
*  *  *  *  *  I  *  *  *  *  *  *  *  *  *  *  *  *  *  R  *  *  T  V  *  *  A
Q  E  H  S  V  V  L  I  R  G  G  R  V  K  D  L  P  G  V  K  Y  H  V  I  R  G  C
AAGAACATTCAGTAGTCCTTATCAGAGGTGGAAGAGTTAAAGATTTACCAGGAGTAAAGTACCACGTAATACGTGGATGT
       410       420       430       440       450       460       470       480

 *  *  C  S  G  *  *  D  *  *  Q  *  *  *  *  *  *  *  *  R  *  *  A  TER
 L  D  A  A  S  V  K  N  R  K  N  A  R  S  K  Y  G  V  K  K  P  K  P  K TER
TTAGACGCAGCAAGTGTAAAAAATCGCAAAAATGCAAGATCTAAATACGGTGTAAAAAAACCAAAACCCAAATAAAACCA
       490       500       510       520       530       540       550       560

                                                                           *  P  *
                                                                           M  S  R
ACCCAAATTAACTAAAATAAAACACTAAAGTTGAAAAAGAAATTTAGAAAAATAGAAATTTAACTTATATATATGTCTCG
       570       580       590       600       610       620       630       640

  *  *  V  I  G  Q  *  K  *  L  P  *  *  K  F  G  *  E  *  L  A  *  F  V  *  I
  R  R  R  A  K  K  R  I  I  S  Q  D  P  I  Y  N  S  T  L  A  S  K  V  I  N  K
AAGAAGAAGAGCAAAAAAAAAGAATAATATCACAAGATCCTATCTATAACAGTACTTTAGCAAGTAAAGTAATTAATAAAA
       650       660       670       680       690       700       710       720

L  M  V  D  *  *  *  S  T  *  E  S  *  V  *  S  A  L  E  T  L  A  Q  R  S  G  *
I  L  L  N  G  K  K  T  L  A  Q  Y  I  F  Y  E  T  M  K  N  I  Q  E  I  Y  K  K
TATTATTGAATGGAAAAAAAACTCTAGCTCAGTATATTTTTTATGAAACAATGAAAAATATACAGGAAATTTATAAAAAA
       730       740       750       760       770       780       790       800

 S  E  *  E  A  F  E  V  *  L  E  *  V  R  *  T  V  *  V  K  S  *  *  V  *  *  S
 D  P  L  D  I  L  R  K  A  I  K  N  A  S  P  Q  M  E  T  R  K  R  R  I  G  G  T
GACCCCTTAGACATTCTAAGAAAAGCGATAAAAAACGCATCTCCACAAATGGAAACAAGAAAGCGTCGTATCGGAGGAAC
       810       820       830       840       850       860       870       880

 T  *  *  *  *  *  *  *  P  V  R  *  D  A  L  A  M  R  W  I  V-Δ-*  A  *  *
 I  Y  Q  V  P  V  E  V  K  E  D  R  G  T  S  L  A  L  K  F  I  I  E  K  A  R
AATATATCAAGTTCCTGTAGAAGTAAAAGAAGATCGTGGAACTAGCTTAGCATTAAAATTCATAATAGAAAAAGCTAGAG
       890       900       910       920       930       940       950       960

K  *  G  D  K  S  M  A  L  R  *  A  D  *  L  S  *  *  A  E  *  K  *  T  *  *  *
E  R  K  G  R  G  I  S  T  K  L  K  N  E  I  I  D  A  S  N  N  T  G  E  A  V  K
AAAGAAAAGGAAGAGGAATATCCACAAAACTAAAAAATGAAATAATTGATGCATCGAATAACACAGGTGAAGCTGTAAAA
       970       980       990      1000      1010      1020      1030      1040
```

```
 *  R  *  D  V  A  R  M  *  *  *  *  *  *  *  A  H  Y  R  TER
 K  K  E  E  I  H  K  T  A  E  A  N  K  A  F  S  N  M  K  F TER
AAAAAAGAGGAGATACATAAAACAGCTGAAGCAAATAAAGCATTCTCAAATATGAAATTTTAATATATATAAATTTTTTA
        1050      1060      1070      1080      1090      1100      1110      1120

AGCGTTAAAAAAAATAAAAAAATAAAATTTGACCAAAATCTTTCTAATAAAATGAGTTTAATTAAATATTTTAAATAAAAT
        1130      1140      1150      1160      1170      1180      1190      1200

                                                                              *  S
                                                                              M  A
AAAAATTTTATTAATTTTAAGATAATAAAATTTTTTCAAAAAAAAAGTTTCATAAAAGAAAACAAAAAATAAAAATGGCT
        1210      1220      1230      1240      1250      1260      1270      1280

 K  E  *  *  *  *  *  *  *  *  V  *  V  *  *  *  *  *  *  *  *  *  *  *  *  *
 R  Q  K  F  E  R  T  K  P  H  I  N  I  G  T  I  G  H  V  D  H  G  K  T  T  L
CGTCAGAAATTTGAAAGAACTAAGCCACATATTAACATAGGCACAATTGGACATGTTGACCATGGAAAAACTACTTTA
        1290      1300      1310      1320      1330      1340      1350      1360
```

Fig. 2. Nucleotide sequence of the region containing the rps 12, rps 7 and parts of the tuf A gene.
A DNA stretch of 1358 positions, starting with the right hand EcoRI restriction site is given (arrow). We only show the RNA like strand. Within the coding parts the deduced aminoacid sequences are given for the S12 and S7 proteins and aligned with them are the aminoacid sequences of the corresponding E. coli proteins (6,7). For convenience we add 28 codons of the first exon of the tuf A gene. Asterisks indicate aminoacid identity. -△- (base position 945-947) indicates a deletion. The deletion may be placed alternatively at nucleotide positions 912-914 increasing somewhat the degree of aminoacid sequence homology. Aminoacids positions 42 and 87 of the E. coli S12 protein which are involved in streptomycin binding are marked.

observed in bacteria are also found in certain streptomycin resistant Euglena strains.

The rps 7 gene codes for a protein of 155 aminoacids. This gives a calculated molecular weight of 17'831. The S7 protein occurs in E. coli in two sizes depending on the strain (7). The Euglena chloroplast S7 protein is two positions longer than the E. coli B S7 (note the insertion at nucleotide position 945), and 22 aminoacids shorter than the E. coli K S7 protein. The aminoacid sequence homology between the chloroplast and bacterial S7 protein is very low, only 38%. As a consequence there exist only a few short conserved regions, e.g., aminoacid positions 85 to 91 and 145 to 151. We may add that the same domains are also conserved in the S7 protein of soybean chloroplasts (J.M. von Allmen, personal communication) what strongly suggests that these regions are functionally important.

Table I shows codon usages for translating the chloroplast and bacterial S12 proteins. Euglena chloroplasts prefer codons rich in A and/or U while

TABLE 1. A comparison of the codon usage in the *Euglena gracilis* chloroplast and *E. coli** *rps* 12 gene

Phe	UUU	0	(0)	Ser	UCU	2	(0)	Tyr	UAU	1	(2)	Cys	UGU	1	(1)
	UUC	0	(1)		UCC	1	(5)		UAC	3	(2)		UGC	2	(3)
Leu	UUA	7	(0)		UCA	4	(0)	Ter	UAA	1	(1)	Ter	UGA	0	(0)
	UUG	1	(0)		UCG	0	(0)		UAG	0	(0)	Trp	UGG	0	(0)
Leu	CUU	2	(1)	Pro	CCU	1	(3)	His	CAU	3	(0)	Arg	CGU	1	(13)
	CUC	0	(1)		CCC	1	(0)		CAC	1	(3)		CGC	3	(2)
	CUA	0	(0)		CCA	6	(1)	Gln	CAA	2	(1)		CGA	2	(0)
	CUG	0	(6)		CCG	2	(3)		CAC	0	(3)		CGG	1	(0)
Ile	AUU	0	(0)	Thr	ACU	1	(5)	Asn	AAU	4	(0)	Ser	AGU	1	(0)
	AUC	1	(3)		ACC	0	(2)		AAC	0	(5)		AGC	0	(1)
	AUA	5	(0)		ACA	7	(1)	Lys	AAA	18	(9)	Arg	AGA	5	(0)
Met	AUG	1	(0)		ACG	0	(0)		AAG	1	(4)		AGG	1	(0)
Val	GUU	3	(6)	Ala	GCU	1	(3)	Asp	GAU	1	(0)	Gly	GGU	2	(7)
	GUC	1	(0)		GCC	1	(0)		GAC	1	(3)		GGC	0	(4)
	GUA	7	(5)		GCA	5	(2)	Glu	GAA	3	(3)		GGA	5	(0)
	GUG	0	(4)		GCG	0	(4)		GAG	0	(1)		GGG	2	(0)

*Initiation codon not included; reference, see text.

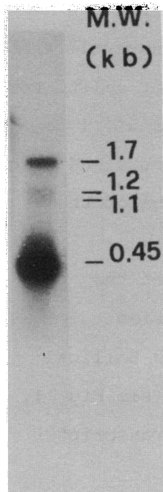

Fig. 3. S1 nuclease protection analysis of the DNA region containing the rps 12, rps 7 genes and parts of the tuf A gene. Chloroplast RNA was hybridized to the 1.7 kb DNA fragment HindIII-EcoRI (right side of Eco·N) and the protected DNA fragments electrophoresed under alkaline conditions in a 2% agarose gel. The blotted DNA fragments (nitrocellulose paper) were detected by hybridization with ^{32}P-labelled Eco·N DNA.

E. coli prefers codons rich in G and/or C. As a result the sequence divergence already seen on the aminoacid level is accentuated on the nucleotide level, being only 52% for the S12 genes. For the chloroplast and bacterial rps 7 genes sequence homologies on the nucleotide and aminoacid level are about equal (38%) at least for the first 81 codons, as far as data are available (5). Also in this case a difference in codon usage is observed (data not shown) similar to that for the rps 12 genes. But for statistical reasons the difference in codon usage is not apparent.

2. Transcription of the chloroplast rps 12 and rps 7 genes

We have previously shown that the chloroplast tuf A gene is split (1). It was important to test whether the adjacent rps 12 and rps 7 genes are also split or whether the stable transcripts are colinear with the genes like in E. coli. We used as hybridization probe the 1.7 kb HindIII-EcoRI fragment in S1 nuclease protection analysis (see Fig. 1). The experiments were performed under neutral and alkaline conditions. In Fig. 3 we only show the result obtained under alkaline conditions since the result obtained under neutral conditions was identical. We see three bands on the radiograph, the 0.45 kb fragment corresponds to a DNA segment containing the first exon of the tuf A gene (consult Fig. 1 and ref. 1). The weak double band of 1.1 to 1.2 kb can only stem from a contiguous DNA fragment protected by a mRNA coding for the S12 + 7 protein including the intergenic spacer (375 + 80 + 468 bases). The 1.7 kb band represents some renatured DNA probe. No smaller protected DNA

fragments can be detected, what strongly suggests that the two genes are co-transcribed into a dicistronic mRNA. Furthermore, since the "neutral" and "alkaline" data coincided, the two ribosomal protein genes are not split. For the very moment we cannot unequivocally explain the appearance of a 1.1 to 1.2 kb double band in Fig. 3. Some preliminary data (not shown) obtained with other DNA probes suggest that the size difference is at the 5' end of the transcripts.

Koller and Delius (25) showed by EM analysis very recently that the entire Eco·N is transcribed in one direction only, but transcription switches to the other strand on either side of Eco·N. Karabin and Hallick (26) showed that indeed on Eco·Q and in close vicinity to Eco·N (see Fig. 1, ref. 1) two tRNA gene clusters of opposite polarity exist. The transcripts on the other side of Eco·N in Eco·J' are not yet characterized.

In E. coli the rps 12 and rps 7 genes are part of the str operon and co-transcribed with the genes for EF-G and EF-Tu proteins (3). In E. gracilis chloroplasts this gene cluster is transcribed into a stable dicistronic 1.1-1.2 kb mRNA (rps 12 + 7) and a 1.95 kb spliced mRNA (tuf A plus ORF for unknown proteins). Under the prevailing growth and analytical conditions, the 1.95 kb transcript is much more abundant than the 1.1 - 1.2 kb mRNAs as estimated from the relative intensities of the signals in the radiograph (Fig. 3). This suggest that transcription of the two regions is under separate control (two promoters), but it is also possible that the entire region is co-transcribed (one promoter) and the steady state concentration of stable transcripts depends primarily on subsequent transcript processing. Nothing specific is yet known about promoter sequences in E. gracilis chloroplast genomes. Therefore, we do not mark in Fig. 2 any possible promoter like "boxes" upstream of the coding parts of the genes.

ACKNOWLEDGMENTS

We are grateful to Drs A.R. Subramanian for helpful advices and B. Koller for providing results prior to publishing. We thank C. Aegerter for technical and Ch. Bachmann for secretarial help. This research is supported by Fonds national suisse de la Recherche scientifique (3.183.82 E.S.).

REFERENCES

1. Montandon,P.E. and Stutz,E. (1983) Nucleic Acids Res. 11, 5877-5892.
2. Passavant,C.W., Stiegler,G.L. and Hallick,R.B. (1983) J. Biol. Chem.

258, 693-695.
3. Jaskunas,S.R., Lindahl,L., Nomura,M. and Burgess,R.R. (1975) Nature 257, 458-462.
4. Yokota,T., Sugisaki,H., Takanami,M. and Kaziro,Y. (1980) Gene 12, 25-31.
5. Post,L.E. and Nomura,M. (1980) J. Biol. Chem. 255, 4660-4666.
6. Funatsu,G., Yaguchi,M. and Wittmann-Liebold,B. (1977) FEBS Lett. 73, 12-16.
7. Reinbolt,J., Tritsch,D. and Wittmann-Liebold,B. (1978) FEBS Lett. 91, 297-300.
8. Nomura,M., Mizushima,S., Ozaki,M., Traub,P. and Lowry,C.V. (1969) Cold Spring Harbor Symp. Quant. Biol. 34, 49-61.
9. Funatsu,G. and Wittmann,H.G. (1972) J. Mol. Biol. 68, 547-550.
10. Schwartzbach,S.D. and Schiff,J.A. (1974) J. Bacteriol. 120, 334-341.
11. Morlé,F., Freyssinet,G. and Nigon,V. (1979) Plant Sci. Lett. 16, 41-49.
12. Sagher,D., Grosfeld,H. and Edelman,M. (1976) Proc. Natl. Acad. Sci. USA 73, 722-726.
13. Ortiz,W., Reardon,E.M. and Price,C.A. (1980) Plant Physiol. 66, 291-294.
14. Rutti,B., Keller,M., Ortiz,W. and Stutz,E. (1981) FEBS Lett. 134, 15-19.
15. Graf,L., Roux,E., Stutz,E. and Kössel,H. (1982) Nucleic Acids Res. 10, 6369-6381.
16. Berk,A.J., Sharp,P.A. (1977) Cell 12, 721-732.
17. Favaloro,J., Treisman,R. and Kaamen,R. (1980) Methods in Enzymology 65 (Grossman,L., Moldave,K. eds), Academic Press, New York, pp. 718-749.
18. Rigby,P.W.J., Dieckman,H., Rhodes,C. and Berg,P. (1977) J. Mol. Biol. 113, 237-251.
19. Messing,J., Crea,R. and Seeberg,P.H. (1980) Nucleic Acids Res. 9, 309-321.
20. Messing,J. and Vieira,J. (1982) Gene 19, 269-276.
21. Sanger,F., Coulson,A.R., Barrel,B.G., Smith,A.J.H. and Roe,B.A. (1980) J. Mol. Biol. 143, 161-178.
22. Wiegand,R.C., Godson,G.N. and Radding,C.M. (1975) J. Biol. Chem. 250, 8848-8855.
23. Breitenberger,C.A., Graves,M. and Spremulli,M. (1979) Arch. Biochem. Biophys. 194, 265-270.
24. Tiboni,O., Di Pasquale,G. and Ciferri,O. (1978) Eur. J. Biochem. 92, 471-477.
25. Koller,B. and Delius,H. (1984) Cell in press.
26. Karabin,G.D. and Hallick,R.B. (1983) J. Biol. Chem. 258, 5512-5518.