1343

REPRINTED FROM:

BBA
Biochimica
et Biophysica Acta

INTERNATIONAL JOURNAL OF BIOCHEMISTRY AND BIOPHYSICS

Biochimica et Biophysica Acta, 1174 (1993) 191–194

BBAEXP 90538

Short Sequence-Paper

Cloning and sequencing of a soybean nuclear gene coding for a chloroplast translation elongation factor EF-G

Jorge Hernández Torres [a], Caroline A. Breitenberger [b],
Albert Spielmann [a] and Erhard Stutz [a]

[a] *Laboratoire de Biochimie végétale, Université de Neuchâtel, Neuchâtel (Switzerland) and*
[b] *Department of Biochemistry, Ohio State University, Columbus, OH (USA)*

(Received 20 April 1993)

ELSEVIER SCIENCE PUBLISHERS B.V.
AMSTERDAM – LONDON – NEW YORK – TOKYO

BIOCHIMICA ET BIOPHYSICA ACTA

International Journal of Biochemistry and Biophysics
Founded in 1947 by H.G.K. Westenbrink, Managing Editor 1947–1964

Authors should submit their manuscripts (four copies) for publication in BIOCHIMICA ET BIOPHYSICA ACTA by sending them to one of the Managing Editors:

G. Schatz **K. van Dam** **L.L.M. van Deenen** **P.C. van der Vliet**	**G.K. Radda**	**P.L. Dutton** **E.P. Kennedy** **D.J. Waxman** (Associate)	**Y. Kagawa**
c/o BBA Editorial Secretariat P.O. Box 1345 1000 BH Amsterdam The Netherlands	Department of Biochemistry University of Oxford South Parks Road Oxford OX1 3QU UK	c/o BBA Editorial Secretariat P.O. Box 9123 Brookline MA 02146-9123 USA	Department of Biochemistry Jichi Medical School Minamikawachi Kawachi-gun Tochigi-ken 329-04 Japan
Tel. +31 20 5803510 Fax +31 20 5803506	Fax +44 865 515059	Tel. +1 (617) 277-7919 Fax +1 (617) 730-5698	Tel. +81 285 44 2111 Ext. 3149 Fax +81 285 44 1827

Detailed information for contributors may be found in Biochim. Biophys. Acta 1075 (1991) 267–280, also published as the 1991 Edition of the booklet
INFORMATION FOR CONTRIBUTORS TO BIOCHIMICA ET BIOPHYSICA ACTA
which is available free of charge from the BBA Editorial Secretariats at P.O. Box 1345, 1000 BH Amsterdam, The Netherlands, or P.O. Box 9123, Brookline MA 02146-9123, USA.

Subscription information

1993 BBA Gene Structure and Expression, Volumes 1171–1174 (4 volumes in 12 issues), Dfl. 1364 (US$ 757.50) + Dfl. 132 (US$ 73.50) postage, packaging and handling. Total price: Dfl. 1496 (US$ 831.00).
1993 BBA full subscription, Volumes 1140–1182 (43 volumes in 121 issues), Dfl. 10664 (US$ 5924.50) + Dfl. 1419 (US$ 788.50) postage, packaging and handling. Total price: Dfl. 12083 (US$ 6713).
The Dutch guilder price is definitive. The US dollar price is subject to exchange-rate fluctuations and is given only as a guide.
Subscriptions are accepted on a prepaid basis only, unless different terms have been previously agreed upon.
Subscription orders can be entered only by calendar year (Jan.–Dec.) and should be sent to Elsevier Science Publishers B.V., Journals Department, P.O. Box 211, 1000 AE Amsterdam, The Netherlands, Tel. +31 20 5803642, Fax +31 20 5803598, or to your usual subscription agent.
Postage and handling charges include surface delivery except to the following countries where air delivery via SAL (Surface Air Lift) mail is ensured: Argentina, Australia, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, PR China, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For all other countries airmail rates are available upon request.
Claims for missing issues must be made within six months of our publication (mailing) date, otherwise such claims cannot be honoured free of charge.
In the USA and Canada: All questions arising after acceptance of a manuscript by the Editors, especially those relating to proofs, publication and reprints, should be directed to the Publisher, Elsevier Science Publishers B.V., P.O. Box 1527, 1000 BM Amsterdam, The Netherlands; Fax +31 20 5803264. For further information concerning this or any other Elsevier Science Publishers journal, contact: Elsevier Science Publishing Co., Inc., Attn: Journal Information Center, 655 Avenue of the Americas, New York, NY 10010, USA. Tel. (212) 633-3750; Fax (212) 633-3990; Telex 420-643 AEP UI.
This journal is in the *ADONIS* Service, whereby copies of individual articles can be printed out from CD-ROM on request. An explanatory leaflet can be obtained by writing to ADONIS B.V., P.O. Box 839, 1000 AV Amsterdam, The Netherlands.

BBA is cited in: Current Contents Life Sciences - Biological Abstracts - Chemical Abstracts -
Index Chemicus - Index Medicus - CSA Abstracts - Excerpta Medica (EMBASE) - Reference Update -
Current Awareness in Biological Sciences (CABS)

Printed in The Netherlands

This issue is printed on acid-free paper

Biochimica et Biophysica Acta, 1174 (1993) 191–194

BBAEXP 90538

Short Sequence-Paper

Cloning and sequencing of a soybean nuclear gene coding for a chloroplast translation elongation factor EF-G

Jorge Hernández Torres [a], Caroline A. Breitenberger [b],
Albert Spielmann [a] and Erhard Stutz [a]

[a] *Laboratoire de Biochimie végétale, Université de Neuchâtel, Neuchâtel (Switzerland) and*
[b] *Department of Biochemistry, Ohio State University, Columbus, OH (USA)*

(Received 20 April 1993)

Key words: Nuclear *fus* gene; Chloroplast; Translocation factor; EF-G; (Soybean)

A plant nuclear gene coding for a chloroplast specific translation elongation factor EF-G (cEF-G) was cloned and sequenced for the first time. We screened two partial soybean genomic libraries with a short PCR amplified pea DNA probe constructed according to the N-terminal peptide sequence of pea chloroplast EF-G. The gene is three times split, codes for a chloroplast type transit peptide and a protein very similar to bacterial translation elongation factor EF-G. The gene is expressed as evidenced by Northern hybridisations.

It was shown that chloroplast translation elongation factor EF-G activity of *Euglena gracilis* [1] and of pea [2] is light regulated. The expression of the cEF-G gene may be a key element in controlling light induced chloroplast protein synthesis and thereby chloroplast development. The corresponding gene (genes) is (are) located in the nucleus since no sequence homologous to the bacterial EF-G gene (*fus*) was yet detected in any of the completely sequenced chloroplast genomes [3–6]. However, no chloroplast specific nuclear EF-G gene was identified so far. We report here the cloning and sequencing of a soybean nuclear gene with a significant similarity to bacterial *fus* genes.

The N-terminal amino acid sequence of purified pea chloroplast EF-G was determined (Akkaya, M.S. et al., unpublished data) and accordingly short DNA probes were constructed using the PCR technique. Subsequently, we used one such DNA probe (156 bp) in Southern experiments with restricted soybean DNA to identify fragments potentially carrying *fus*-type sequences. In parallel experiments we also used DNA probes cut from the cloned *fus* gene of *Anacystis nidulans* (pAB46, kindly provided by Dr. Sugiura, Nagoya). *Eco*RI fragments in the range of 4 kb and *Bgl*II fragments in the range of 9 kb interacted with both DNA probes, however, the signal was particularly strong with the pea probe. We then constructed two separate partial genomic libraries with purified *Eco*RI fragments (about 4 kb) and *Bgl*II fragments (about 9 kb) using as vector lambda GT10 and EMBL3, respectively.

In Fig. 1 we show the sequence of a complete nuclear gene coding for a chloroplast transit peptide and a protein having a high sequence identity with EF-G of *Escherichia coli* (59%) [7] and *A. nidulans* (55%) [8], but a definitely lower sequence identity with a yeast mitochondrial EF-G (41%) [9] and a Chlorella translocation factor EF-2 (27%) [10]. The gene is split three times by introns of 330, 508 and 288 bp, respectively, which all have the canonical 5′- and 3′-ends of plant nuclear introns (5′-AG..GU-) and an AU rich domain near the 3′-end [11].

In Fig. 2 the putative soybean cEF-G protein is aligned with EF-G proteins of *E. coli* and *A. nidulans* (see also next paragraph). The soybean protein shares extensive domains with the bacterial counterparts notably three GTP-binding sequences. From these and other (not shown) alignments we conclude that the given sequence represents a chloroplast specific EF-G protein.

The N-terminal peptide (85 amino acids, Fig. 1) has typical features of a chloroplast transit peptide [12,13]: (a) zero net charge for the N-terminal 10 amino acids, (b) alanine in second position, (c) a serine-rich central domain, (d) arginine at positions −7 and −8, (e) a

---

Correspondence to: E. Stutz, Laboratoire de Biochimie végétale, Chantemerle 18, Université de Neuchâtel, CH-2000 Neuchâtel, Switzerland.

```
        10                  30                  50                  70                  90
ATCTGCTGCCTAGTCCTTACCGAAAACTAGAGGCCAGGGTCTCGGAGTGTTTGGGTTTGTTTATCTTAAACACCAAAAACACACAGAACAGAGGTGTAGA
       110                 130                 150                 170                 190
GAAGAAGCTCAAAAATATGGCGGCAGAGTCATCACTGAGGGTGGCGACTCCCACTCTTTGCAACCTCAATGGGTCTCAGAGACGACCAACCACCACCACT
                 M  A  A  E  S  S  L  R  V  A  T  P  T  L  C  N  L  N  G  S  Q  R  R  P  T  T  T  T
       210                 230                 250                 270                 290
CTCTCTCCGCTTCGCTTCATGGGTTTTCGTCCTCGACCCTCCTCTCACTCTCTCACCTCCTCTTCCCTTTCACACTTCTTTGGGAGCACAAGAATTCACT
 L  S  P  L  R  F  M  G  F  R  P  R  P  S  S  H  S  L  T  S  S  S  L  S  H  F  F  G  S  T  R  I  H S
       310                 330                 350                 370             ↓   390
CCAATTCTTCTTCTTCTTATTCTTCCATTTCACGCCAACATGCTCCGAGAAGAAACTTCTCTGTTTTCGCCATGTCTGCTGATGGTATCTATCTCTTTTT
   N  S  S  S  S  Y  S  S  I  S  R  Q  H  A  P  R  R  N  F  S  V  F  A↑ M  S  A  D  D
       410                 430                 450                 470                 490
TCTCTAAGGTTTTATTTTTATGTTTTTCTAACTACGCTAATGTTATGAAAAAAAATTATTTTTTTCTCTTTAAATTCTGTTTTTTGTTTTTTGGGTTTGG
       510                 530                 550                 570                 590
TTCATTTTATGCATCATCCTTAAAATTTGTGTGTTATGCTTGAATTTTGGCTTGAGTTATCATATTAGGGGGAAAAGGCATCATTTGTCTGTGTCATGTT
       610                 630                 650                 670                 690
GAGGTTTTAAATGAGATAAGGTGTCTTAAAAGATAATGATTTTTGTTTGGTCTTTATATTTCCTATGGTCCAATTTTATTCCATGACTATGAAAAGTAGT
       710                 730                 750    ↓            770                 790
TTGCTTTCTTTCTTTCATCCTTTTGAAAATGACCCTTTTCAATTTTGCTTCCAGACGCAAAGCGCAGTGTTCCATTGAAGGATTATCGCAACATCGGTAT
                                                         A  K  R  S  V  P  L  K  D  Y  R  N  I  G  I
       810                 830                 850                 870                 890
CATGGCTCACATAGATGCTGGAAAGACTACTACAACAGAACGAATTTTGTACTATACTGGAAGGAACTACAAAATTGGAGAGGTGCATGAGGGAACAGCA
  M  A  H  I  D  A  G  K  T  T  T  T  E  R  I  L  Y  Y  T  G  R  N  Y  K  I  G  E  V  H  E  G  T  A
       910                 930                 950                 970                 990
ACCATGGACTGGATGGAACAAGAACAAGAGAGGGGGATTACCATTACTTCTGCTGCAACCACTACATTTTGGAACAAGCACCGGATCAACATTATTGATA
 T  M  D  W  M  E  Q  E  Q  E  R  G  I  T  I  T  S  A  A  T  T  T  F  W  N  K  H  R  I  N  I  I  D T
      1010                1030                1050                1070                1090
CCCCTGGTCACGTTGACTTCACCTTAGAAGTGGAGCGTGCTCTTAGGGTGTTGGATGGAGCTATATGCTTGTTTGACAGTGTTGCTGGCGTGGAGCCGCA
   P  G  H  V  D  F  T  L  E  V  E  R  A  L  R  V  L  D  G  A  I  C  L  F  D  S  V  A  G  V  E  P  Q
      1110                1130                1150                1170                1190
ATCGGAAACTGTGTGGAGGCAGGCTGATAAATATGGAGTCCCGCGAATTTGTTTTGTCAATAAAATGGACCGTCTTGGAGCAAACTTTTATAGAACAAGA
  S  E  T  V  W  R  Q  A  D  K  Y  G  V  P  R  I  C  F  V  N  K  M  D  R  L  G  A  N  F  Y  R  T  R
      1210                1230                1250                1270                1290
GACATGATAGTAACGAATTTGGGTGCTAAACCACTTGTAATTCAGTTACCAATTGGTTCAGAAGATAACTTTAAGGGAGTTATAGACCTTGTTAGGAATA
 D  M  I  V  T  N  L  G  A  K  P  L  V  I  Q  L  P  I  G  S  E  D  N  F  K  G  V  I  D  L  V  R  N K
      1310                1330                1350                1370                1390
AGGCTATAGTTTGGTCTGGAGAAGAGCTGGGTGCCAAGTTTGATATTGTAGATATTCCAGAAGATCTTCAAGAGCAGGCTCAGGATTACCGTGCCCAGAT
   A  I  V  W  S  G  E  E  L  G  A  K  F  D  I  V  D  I  P  E  D  L  Q  E  Q  A  Q  D  Y  R  A  Q  M
      1410                1430                1450                1470                1490
GATAGAAAACATAGTTGAGTTTGATGACCAGGCTATGGAGAACTACCTGGAAGGAATTGAACCGGATGAGGAAACTATAAAAAAATTAATTAGGAAGGGA
  I  E  N  I  V  E  F  D  D  Q  A  M  E  N  Y  L  E  G  I  E  P  D  E  E  T  I  K  K  L  I  R  K  G
      1510                1530                1550                1570                1590
ACAATATCAGCCAGTTTTGTGCCAGTGATGTGTGGCTCGGCTTTCAAAAACAAGGGAGTCCAACCATTGCTCGATGCAGTAGTGGATTATCTACCATCAC
 T  I  S  A  S  F  V  P  V  M  C  G  S  A  F  K  N  K  G  V  Q  P  L  L  D  A  V  V  D  Y  L  P  S P
      1610                1630                1650                1670                1690
CACTTGACTTGCCAGCAATGAAGGGCAGTGACCCTGAAAACCCTGAAGCAACAATAGAGAGACTAGCAAGTGATGATGAACCATTTGCTGGACTAGCTTT
   L  D  L  P  A  M  K  G  S  D  P  E  N  P  E  A  T  I  E  R  L  A  S  D  D  E  P  F  A  G  L  A  F
      1710                1730                1750                1770                1790
TAAGATCATGAGTGATCCATTTGTAGGTTCCCTTACGTTTGTCAGGGTGTATGCTGGAAAGCTAGGTGCTGGGTCTTATGTACTCAATGCAAACAAAGGG
  K  I  M  S  D  P  F  V  G  S  L  T  F  V  R  V  Y  A  G  K  L  G  A  G  S  Y  V  L  N  A  N  K  G
      1810                1830                1850                1870                1890
AAAAAGGAGAGAATTGGTAGACTTTTAGAAATGCATGCAAACAGCAGAGATGATGTTAAAGTGGCTTTAGCTGGTGATATTATTGCTCTTGCAGGTTTGA
 K  K  E  R  I  G  R  L  L  E  M  H  A  N  S  R  D  D  V  K  V  A  L  A  G  D  I  I  A  L  A  G  L K
      1910                1930                1950                1970                1990
AAGATACTATTACAGGTGAAACATTGTGTGACCCTGATAATCCAATCGTGCTCGAGCGGATGGACTTCCCTGATCCTGTTATTAAGGTTGCAATTGAACC
   D  T  I  T  G  E  T  L  C  D  P  D  N  P  I  V  L  E  R  N  D  F  P  D  P  V  I  K  V  A  I  E  P
      2010                2030                2050                2070                2090
GAAAACTAAAGCTGATGTTGACAAGATGGCAACTGGTTTAATCAAACTTGCACAGGAAGACCCTTCTTTCCACTTCTCCCGTGATGAAGAGATAAACCAG
  K  T  K  A  D  V  D  K  M  A  T  G  L  I  K  L  A  Q  E  D  P  S  F  H  F  S  R  D  E  E  I  N  Q
      2110                2130                2150                2170                2190
ACAGTAATTGAAGGAATGGGAGAATTACATCTTGAAATCATTGTTGATCGGCTCAAAAGAGAATTTAAGGTAATATATCTCATTCCATTTTTAAGAATTT
 T  V  I  E  G  M  G  E  L  H  L  E  I  I  V  D  R  L  K  R  E  F  K
      2210                2230                2250                2270                2290
TCAATGATCAGTTAGCATGGTTGTGTTTTATTAGCTTTTTATGTTTTCCAAATTTTTAACTCTCCAACCTTGATGCCGTGCATGAAAATATTACAATTAG
      2310                2330                2350                2370                2390
AAATAATAATAGTTAATAAACATGGACATGATCGTATTAGATTATAATGGAATCTGTTTGCACTTTGAATTATCACTTATTACCTGGATAAAGTCTATGT
      2410                2430                2450                2470                2490
CTATCATCTGATAAAGCATTAGTATTTGTTAGGTATACTATCTTCCTTTCCTTTTTTTATATGATTATTCCCTTTGCATGACCTTTTCCTTTCTCCCCCA
      2510                2530                2550                2570                2590
TCTCCTTTTCTCTCTTTCCTTCAGTAGTTTTATGTTAGAGTGTTTTTATATATTTTAATGATTATGTAGTAATAATATAATTTTATGAAGATCACATTAT
      2610                2630                2650                2670                2690
CTATATTTGTTAATGCAGGTGGAAGCTAACGTTGGTGCCCCCCAAGTAAACTACAGGGAAAGCATTTCCAAAATCTCAGAAGTGAAGTATGTGCACAAGA
                   V  E  A  N  V  G  A  P  Q  V  N  Y  R  E  S  I  S  K  I  S  E  V  K  Y  V  H  K  K
      2710                2730                2750                2770                2790
AACAATCAGGTGGACAGGGTCAGTTTGCAGATATCACTGTCCGGTTTGAACCCATGGACCCAGGTAGTGGATATGAGTTCAAGAGTGAAATCAAAGGAGG
   Q  S  G  G  Q  G  Q  F  A  D  I  T  V  R  F  E  P  M  D  P  G  S  G  Y  E  F  K  S  E  I  K  G  G
      2810                2830                2850                2870                2890
TGCTGTACCAAGAGAATACATTCCTGGTGTGATGAAAGGATTGGAAGAGTGTATGAGCAATGGTGTGCTTGCTGGCTTTCCGGTCGTTGATGTACGAGCA
  A  V  P  R  E  Y  I  P  G  V  M  K  G  L  E  E  C  M  S  N  G  V  L  A  G  F  P  V  V  D  V  R  A
      2910                2930                2950                2970                2990
GTACTCACAGATGGTTCTTACCATGATGTAGATTCAAGCGTGTTGGCTTTCCAGTTGGCAGCAAGAGGAGCTTTTAGGGAAGGAATAAGAAAAGCTGGAC
 V  L  T  D  G  S  Y  H  D  V  D  S  S  V  L  A  F  Q  L  A  A  R  G  A  F  R  E  G  I  R  K  A  G P
      3010                3030                3050                3070                3090
CGAGGATGCTTGAGCCTATAATGAAGGTTGAAGTTGTTACTCCTGAAGAACATCTAGGTGATGTAATTGGGGATCTCAACTCAAGAAGAGGACAGATCAA
   R  M  L  E  P  I  M  K  V  E  V  V  T  P  E  E  H  L  G  D  V  I  G  D  L  N  S  R  R  G  Q  I  N
      3110                3130                3150                3170                3190
CAGTTTTGGTGACAAACCTGGTGGCCTTAAGGTATGTCAAATAACTTGGCTTCTCCTTGTCATATCATATTAGTTAGTCTATTAAATTTGAATGTATGAC
  S  F  G  D  K  P  G  G  L  K
      3210                3230                3250                3270                3290
TTGAAAGTACTTGTTCTGTATTGTTTGTCTGAAAAATGCACAAGTTTCCAAATTGTTAAGTTTTATACACAATTCTAATTTCTGGGGAGAAACACTAAAA
      3310                3330                3350                3370                3390
CATGAAAGAAAGGAAAAGTAAAATATTTTTGTTCTGTTTTTCCTTTTCTTGAATATTGATTTTATTATTTGGATTGGAGCTGGACTGATTCTTTTCTTCT
      3410                3430                3450                3470                3490
TTTCATGGCGACAAACACAGGTGGTTGATTCCCTGGTACCTCTTGCTGAGATGTTCCAGTATGTCAGCACACTCAGAGGGATGACAAAGGGCCGTGCATC
                     V  V  D  S  L  V  P  L  A  E  M  F  Q  Y  V  S  T  L  R  G  M  T  K  G  R  A  S
      3510                3530                3550                3570                3590
CTACACAATGCAATTAGCCATGTTTGATGTGGTGCCTCAGCACATTCAGAACCAACTTGCCACCAAAGAGCAAGAAGTTGCCGCTTAATGTACTTGCTTG
  Y  T  M  Q  L  A  M  F  D  V  V  P  Q  H  I  Q  N  Q  L  A  T  K  E  Q  E  V  A  A  *
      3610                3630                3650                3670                3690
GTCTCCTAACAAGGAACTTTTTACCTTTGATAGTTGTATTATGTCCATTGTCAAATAGTTGGGTTTGTTTTTGGTATTTTTTTTATTGTAATTTTCCTTG
      3710                3730                3750                3770                3790
TACTATACAATTTAACTCAACAGAAAAGTCACAGCAATGACAGCATGCAGGAAGCACCAGTAAAGAGTATAAATGGACTGCATTTGACATGCTCAAATAA
```

Fig. 1. Nucleotide sequence of a soybean nuclear gene coding for the chloroplast translation elongation factor EF-G. Sequences coding for the transit peptide and the EF-G protein are decoded. ↓ marks a presumably spliced codon (G/AC) coding for aspartic acid (D). Canonical 5′- and 3′-ends of introns are underlined. ↑ marks a possible cleavage site separating the putative transit peptide from the cEF-G protein. The proteolytic cleavage motive is in bold. .

```
       .        .         .         .         .         .         .
S   1  AKRSVPLKDYRNIGIMAHIDAGKTTTTERILYYTGRNYKIGEVHEGTATMDWMEQEQERG
E   1  MARTTPIARYRNIGISAHIDAGKTTTTERILFYTGVNHKIGEVHDGAATMDWMEQEQERG
A   1  MARSVPLEKVRNIGIAAHIDAGKTTTTERILFYSGVVHKIGEVHDGNAVTDWMEQERERG
       ..*. *.   ***** ***************.*.*   ******.* *  ******.***

       .        .         .         .         .         .         .
S  61  ITITSAATTTFWN-------KHRINIIDTPGHVDFTLEVERALRVLDGAICLFDSVAGVE
E  61  ITITSAATTAFWSGMAKQYEPHRINIIDTPGHVDFTIEVERSMRVLDGAVMVYCAVGGVQ
A  61  ITITAAAISTSW----KDY---RVNIIDTPGHVDFTIEVERSMRVLDGVVAVFCSVGGVQ
       ****.** .. *          *.************.****..***** . .. .*.**.

       .        .         .         .         .         .         .
S 114  PQSETVWRQADKYGVPRICFVNKMDRLGANFYRTRDNIVTNLGAKPLVIQLPIGSEDNFK
E 121  PQSETVWRQANKYKVPRIAFVNKMDRMGANFLKVVNQIKTRLGANPVPLQLAIGAEEHFT
A 114  PQSETVWRQADRYSVPRIVFVNKMDRTGADFFKVYGQIRDRVRANAVPIQIPIGAESDFQ
       **********..* **** ******* **.* .    *   . *... .*..**.* .*

       .        .         .         .         .         .         .
S 174  GVIDLVRNKAIVWSGEELGAKFDIVDIPEDLQEQAQDYRAQMIENIVEFDDQAMENYLEG
E 181  GVVDLVKMKAINWNDADQGVTFEYEDIPADNVELANEWHQNLIESAAEASEELMEKYLGG
A 174  GIVDLVEMKAHIYTN-DLGTDILVTDIPAELQETAAEWRSKMVEAVAETDEALLDKYFED
       *..***  **      . *  .   *** .. * * . . ...*   *  .  ...*. .

       .        .         .         .         .         .         .
S 234  IEPDEETIKKLIRKGTI---SASFVPVMCGSAFKNKGVQPLLDAVVDYLPSPLDLPAMKG
E 241  EELTEAEIKGALRQRVL---NNEIILVTCGSAFKNKGVQAMLDAVIDYLPSPVDVPAING
A 233  GDLSIEDIKAGLRKGVLIQGNDRLVPMLCGSAFKNKGVQLLLDAVVELLPSPQDIPPIQG
        .     **  .*.  .   .  .. . *********** .****.. **** *.*...*

       .        .         .         .         .         .         .
S 291  SDPENPEATIERLASDDEPFAGLAFKIMSDPFVGSLTFVRVYAGKLGAGSYVLNANKGKK
E 298  ILDDGKDTPAERHASDDEPFSALAFKIATDPFVGNLTFFRVYSGVVNSGDTVLNSVKAAR
A 293  TLPDGEVAL--RPSSDEAPFSALAFKIMADPY-GRLTFVRVYSGILQKGSYVYNATKGKK
          .   .   * .**. **..***** .**. * *** ***.* .  *  * *. *. .

       .        .         .         .         .         .         .
S 351  ERIGRLLEMHANSRDDVKVALAGDIIALAGLKDTITGETLCDPDNPIVLERMDFPDPVIK
E 358  ERFGRIVQMHANKREEIKEVRAGDIAAAIGLKDVTTGDTLCDPDAPIILERMEFPEPVIS
A 350  ERVSRLIILKADDRIEVDELRAGDLGAVLGLKDTFTGDTLCDDQNPIILESLFIPEPVIS
       ** .*.. . *. * ..    ***. *  ****  **.**** . **.** . .*.***

       .        .         .         .         .         .         .
S 411  VAIEPKTKADVDKMATGLIKLAQEDPSFHFSRDEEINQTVIEGMGELHLEIIVDRLKREF
E 418  IAVEPKTKADQEKMGLALGRLAKEDPSFRVWTDEESNQTIIAGMGELHLDIIVDRMKREF
A 410  VAVEPKTKNDMEKLSKALQALSEEDPTFRVSVDSETNQTVIAGNGELHLEILVDRMLREY
       .*.***** * .*.. .*  *. ***.*.   * * ***.* *******.*.***. **.

       .        .         .         .         .         .         .
S 471  KVEANVGAPQVNYRESIS-KISEVKYVHKKQSGGQGQFADITVRFEPMDPGS---GYEFK
E 478  NVEANVGKPQVAYRETIRQKVTDVEGKHAKQSGGRGQYGHVVIDMYPLEPGSNPKGYEFI
A 470  KVEANIGAPQVAYRETVR-KAVKAEGKFVRQSGGKGQYGHVVIELEPAEPGT---GFEFV
       .****.* *** ***..  *         .****.**.... .   * .**.   *.**

       .        .         .         .         .         .         .
S 527  SEIKGGAVPREYIPGVMKGLEECMSNGVLAGFPVVDVRAVLTDGSYHDVDSSVLAFQLAA
E 538  NDIKGGVIPGEYIPAVDKGIQEQLKAGPLAGYPVVDMGIRLHFGSYHDVDSSELAFKLAA
A 526  SKIVGGTVPKEYVGPAEQGMKETCESGVLAGYPLIDIKATLVDGSYHDVDSSEMAFKIAG
       . * ** .* **.    .*. *    * ***.*..*.   *  ********* .**..*.

       .        .         .         .         .         .         .
S 587  RGAFREGIRKAGPRMLEPIMKVEVVTPEEHLGDVIGDLNSRRGQINSFGDKPGGLKVVDS
E 598  SIAFKEGFKKAKPVLLEPIMKVEVETPEENTGDVIGDLSRRRGMLKGQESEVTGVKI-HA
A 586  SMAIKEAVRKADPVLLEPVMKVEVEVPEDFLGSVMGNLISRRGQIEGQATT-NGTATVSA
         *..*. .** * .***.*****  **.  * *.*.*  *** . .      *     .

       .        .         .         .         .         .
S 647  LVPLAEMFQYVSTLRGMTKGRASYTMQLAMFDVVPQHIQNQLATKEQEVAA
E 657  EVPLSEMFGYATQLRSLTKGRASYTMEFLKYDEAPSNVAQAVI---EARGK
A 645  KVPLAEMFGYATDLRSMTQGRGIFTMEFSQYEEVPRNVAETIIAKNKG-NA
        ***.*** * . **..*.**. .**..  ..  * .. . .
```

Fig. 2. Alignment of soybean chloroplast EF-G (S) with its *E. coli* (E) and *A. nidulans* (A) couterparts. (*) Identity in all three sequences; (·) conservative matches; overlines mark GTP binding domains. In this alignment we omited the presumed first four (five) amino acids of the soybean chloroplast EF-G, see text.

β-sheet forming sequence near the cleavage site and (f) a proteolytic cleavage site -VFA↑*M*- which closely follows the consensus motive (-V/IXA/C↑A-). We sequenced the N-terminus of the processed pea cEF-G protein as A**TEDGKR**- (Akkaya M.S. et al., unpublished data). The soybean counterpart would read SA**DD**A**KR**- (identical positions in bold) if the precursor protein is cleaved as suggested and *M* is split off in the processing step [13].

From Southern hybridisation experiments (not shown) using homologous DNA probes we gather that soybean has two different nuclear loci carrying a cEF-G gene. In Northern expriments using total RNA (not shown) we identified a transcript of about 3 kb, which may qualify as the stably processed mRNA of a *fus* gene. We do not yet know whether both genes are expressed. It is noteworthy that soybean nuclear DNA carries four genes for the chloroplast specific translation elongation factor EF-Tu [14].

We are grateful to Dr. Sugiura, Nagoya for the *A. nidulans* clone, Sophie Marc-Martin for technical help, the National Institutes of Health for support to C.A.B, the Fonds nationale suisse pour la recherche scientifique and the ROCHE Research Foundation (Basel) for support to E.S. This work is part of the PhD programm of J.H.T. which is recipient of a student fellowship of the Swiss Government.

References

1 Breitenberger, C.A., Graves, M.C. and Spremulli, L.L. (1979) Arch. Biochem. Biophys. 194, 265–270.

2 Akkaya, M.S. and Breitenberger, C.A. (1992) Plant Mol. Biol. 20, 791–800.

3 Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.

4 Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.

5 Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) Mol. Gen. Genet. 217, 185–194.

6 Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A. and Stutz, E. (1993) Nucleic Acids Res., in press.
7 Zengel, J.M., Archer, R.H. and Lindahl, L. (1984) Nucleic Acids Res. 12, 2181–2192.
8 Meng, B.Y., Shinozaki, K. and Sugiura, M. (1989) Mol. Gen. Genet. 216, 25–30.
9 Vambutas, S., Ackerman, H. and Tzagoloff, A. (1991) Eur. J. Biochem. 201, 643–652.
10 Schnelbögl, G. and Tanner, W. (1991) Plant Physiol. 97, 469–471.
11 Goodall, G.J. and Filipowicz, W. (1989) Cell 58, 473–483.
12 von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–545.
13 Gavel, Y. and Von Heijne, G. (1990) FEBS Lett. 261, 455–458.
14 Maurer, F., Spielmann, A., Bonny, C. and Stutz, E. (1993) Experientia 49, A17.

Information

BBA Information for Contributors

◆

A summary is to be found at the beginning
of each volume

◆

The full version in booklet form
is available from:

BBA Editorial Secretariat
P.O. Box 1345
1000 BH Amsterdam
The Netherlands
Fax: +31 20 5803506
Tel.: +31 20 5803510

or

BBA Editorial Secretariat
P.O. Box 9123
Brookline MA 02146-9123
USA
Fax: +1 (617) 730-5698
Tel.: +1 (617) 277-7919

For any additional advice concerning
'electronic manuscripts'
please contact a Secretariat or the Publisher

Submission

Manuscripts may be submitted to any of the
receiving centres listed on
the inside front cover of a recent issue of BBA

Short Sequence-Papers must be submitted to
the Editorial Secretariat in Amsterdam

Publication

For information concerning your accepted
manuscript, proof, etc.:

Elsevier Science Publishers B.V.
Fax: +31 20 5803264
Tel.: +31 20 5803423

Postal address:
P.O. Box 1527
1000 BM Amsterdam
The Netherlands

Courier:
Molenwerf 1
1014 AG Amsterdam
The Netherlands

Always quote "BBA" and the article number

US mailing notice – Biochimica et Biophysica Acta (*Gene Structure and Expression*) (ISSN 0167-4781) is published monthly by Elsevier Science Publishers (Molenwerf 1, Postbus 211, 1000 AE Amsterdam, The Netherlands).
Annual subscription price in the USA US$ 831.00 (subject to change), including air speed delivery. Second class postage paid at Jamaica, NY 11431.
USA POSTMASTERS: Send address changes to *Biochimica et Biophysica Acta* (*Gene Structure and Expression*), Publications Expediting, Inc., 200 Meacham Avenue, Elmont, NY 11003.
Airfreight and mailing in the USA by Publications Expediting.

Plant Molecular Biology **27**: 1221–1226, 1995.

Short communication

Isolation and characterization of *gmsti*, a stress-inducible gene from soybean (*Glycine max*) coding for a protein belonging to the TPR (*tetratrico*peptide repeats) family

Jorge Hernández Torres, Philippe Chatellard and Erhard Stutz*
Laboratoire de Biochimie végétale, Université de Neuchâtel, Rue Emile Argand 11, CH-2007 Neuchâtel, Switzerland (author for correspondence)*

Received 26 July 1994; accepted in revised form 30 January 1995

Key words: *gmsti* gene, heat stress, *tetratrico*peptide, soybean, TRP motif

Abstract

In close vicinity of two *fus* nuclear genes (chloroplast-specific translation elongation factor cEF-G) of soybean (*Glycine max*) we localized a split nuclear gene coding for a protein with *tetratrico*peptide repeats (TPR). A full-length cDNA was sequenced (1871 nucleotides). It encodes a protein (569 amino acids) with high sequence identity to the yeast STI1 stress-inducible and the human transformation-sensitive IEF SSP 3521 protein which both carry TPR elements. The soybean gene is heat-inducible. This is the first evidence for the existence of plant genes coding for proteins which belong to the TPR family. We call the gene *gmsti* and the protein GMSTI in analogy to the yeast counterpart.

While studying the structure and function of two soybean nuclear genes coding each for the chloroplast-specific translation elongation factor cEF-G (*fus*) [9, 10] we noticed in close vicinity downstream of both *fus* genes a reading frame with typical 34 amino acid repeats (*tetratrico*peptide repeat, TPR motif). The TPR motif is defined by a degenerate consensus sequence and is found as tandem repeats in diverse groups of proteins that function, for example, in mitosis, neurogenesis, transcription regulation and protein import [20].

Representatives of the TPR gene family are found in yeast and other fungi, in cyanobacteria, in *Drosophila* (review in [20]) and in man [12] but not in plants as verified by computer-assisted search of EMBL data files. We screened a soybean cDNA library using a soybean genomic DNA fragment carrying a TPR sequence. We eventually retrieved and sequenced a cDNA clone predicted to encode a protein (GMSTI) with high sequence identity to the transformation-sensitive human protein IEF SSP 3521 [12] and to the stress-inducible yeast protein STI1 [13]. Both, the human and yeast protein contain several TPR motifs but the exact function of the two proteins

---

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X79770.

is unknown. This report presents the characterization of a soybean nuclear gene sequence which encodes a protein structurally related to the human and yeast stress-inducible proteins. The steady-state concentration of the *gmsti* mRNA sharply increases in soybean plants exposed to elevated temperature. We propose to name this novel soybean gene *gmsti* (*Glycine max sti*) in analogy to the yeast counterpart.

Mapping and partial sequencing of two genome regions containing the gmsti *gene*

A partial genomic *Bgl* II library (ca. 9 kb fragments) was specifically constructed in view of retrieving *fus* genes [9, 10]. Soybean (*Glycine max* cv. Ceresia [18]) was grown in a phytotron (Sanyo) at 25 °C, with a 12 h dark/12 h light photoperiod and at 50% relative humidity. Leaves were harvested after about two weeks of growth, immediately frozen in liquid N_2 and stored at −70 °C until use. DNA and RNA were isolated from 5 to 10 g of leaves according to standard procedures and as published [4]. Total soybean DNA was digested with *Bgl* II. DNA fragments were electrophoresed in a 0.6% agarose gel and fragments of about 9 kb were isolated by electroelution. About 500 ng of 9 kb DNA fragments were ligated at 4 °C for 48 h in a ligation mix containing 1 μg of *Bam* HI λEMBL3 arms (Stratagene), 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1.5 mM ATP and 1 U of T4 DNA ligase (BRL) in a total volume of 10 μl. *In vitro* packaging reaction was done using Promega's Packagene Lambda DNA Packaging System. Library screening was carried out following standard protocols [16] using as host *Escherichia coli* strain ER 1648 [8] (gift from Biolabs). Sequencing was done as described [17] or according to the protocol for sequencing double-stranded DNA preparations using Sequenase (U.S. Biochemical) or *Taq* polymerase (*fmol* DNA sequencing kit, Promega). DNA sequence data were analysed using the Sequence Analysis Software Package from the Genetics Computer Group (GCG) [5].

Sequencing of regions downstream of *fus* revealed reading frames with TPR motifs (FASTA search, EMBL data files). Figure 1 shows a short genomic sequence containing, respectively, the 3′-terminal part of *gmsti*1 with at least one TPR element, the 3′-terminal part of *fus*1 and the non-coding part between the two genes of opposite polarity. A similar study with a second genomic clone (*fus*2) gave a similar result (not shown), however, the distance between the 3′ ends of the two genes of opposite polarity (*gmsti*2-–*fus* 2) is about 1000 nucleotides longer as determined by PCR analysis and partial sequencing. Indeed, gene 1 has a deletion in the non-coding 3′ part of *fus*1 eliminating the polyadenylation site (J. Hernández Torres and E. Stutz, to be published).

A gmsti *cDNA sequence coding for a protein with TPR motifs*

A genomic *gmsti*1 DNA probe radiolabelled by random priming [6] was used to screen a lambda Uni-ZAP XR soybean cDNA library constructed with mRNA isolated from 12-day epicotyls of soybean cv. Williams. We followed the instructions of the manufacturer (Stratagene) using the host strain *E. coli* XL1-Blue [2]. Positive pBluescript phagemids were excised *in vivo* from the lambda Uni-ZAP XR vector using the host strain SOLR (Stratagene). Several positive clones were retrieved and a full-length ORF was eventually sequenced coding for a protein structurally related to the human IEF SSP 3521 and yeast STI1 stress-inducible protein. In Fig. 2 the soybean GMSTI protein sequence derived from the cDNA sequence is aligned with the human and yeast counterparts. The soybean protein which is composed of 569 amino acids with a predicted molecular size of about 63 550 Da shares several conserved domains with the human and yeast counterparts, especially in the N- and C-terminal region. One notices, for example, a perfectly conserved β-sheet-forming [3] decamer, **NHVLYSNRSA**, which is situated at the end of the first TRP motif (positions 35 to 44). Overall sequence identities between the soybean GMSTI protein

```
   1  gatatcccttttcttttagGCTTATAGTAACAGAGCTGCATGCTATACAAAACTGGGGGCAATGCCCGAAGGCTTAAAAGATGCAGAGAAATGCATTGAG  100
                          A  Y  S  N  R  A  A  C  Y  T  K  L  G  A  M  P  E  G  L  K  D  A  E  K  C  I  E

 101  CTTGATCCAACCTTTTCAAAGGGTTATACTAGGAAAGGTGCAGTGCAGTTTTCCATGAAAGAATATGACAAAGCTTTGGAAACATACAGGGAGGGTTTGA  200
       L  D  P  T  F  S  K  G  Y  T  R  K  G  A  V  Q  F  S  M  K  E  Y  D  K  A  L  E  T  Y  R  E  G  L  K

 201  AACATGATCCTAACAATCAGGAATTGCTTGATGGCATAAGAAGgtatggtatcgtgctacattcctgctatttttgggttgaatattgccactttccata  300
         H  D  P  N  N  Q  E  L  L  D  G  I  R  R

 301  tgcaccatgacaatcaagtgcctaatcaccatagtatcatacattcagATGTGTAGAACAAATTAATAAGGCTAGCCGTGGAGATTTTACACCTGAGGAG  400
                                                        C  V  E  Q  I  N  K  A  S  R  G  D  F  T  P  E  E

 401  TTAAAGGAGAGACAGgtgtgacttgtgttctgataatctttgaactctattgctaaatcacctgaaatggcaattcatgtgtcaaatatttaattttatg  500
       L  K  E  R  Q

 501  tttttgcagGCCAAGGCAATGCAAGACCCAGAGATACAAAGCATTCTGCAAGACCCTGTTATGACACAAgtatcgtcccttgttgattgaaagttactgt  600
                A  K  A  M  Q  D  P  E  I  Q  S  I  L  Q  D  P  V  M  T  Q

 601  gtgttgtgtcaatgtaatagaggaactagctgtgctaacgagagttgggcatgtatgcagGTATTGACTGATTTCCAGGAAAATCCCAGAGCTGCAGAGG  700
                                                                     V  L  T  D  F  Q  E  N  P  R  A  A  E  E

 701  AACATGTGAAGAATCCAATGGTGATGAACAAGACTCAAAACGTGACAGTGCCGGGATGCCAGATGAGATGAATTTGATGGAAGGTGGCAATCAAGCAGAG  800
         H  V  K  N  P  M  V  M  N  K  T  Q  N  V  T  V  P  G  C  Q  M  R  *
                                                                     ↓
 801  AAACAATTGACAATTGTATTTTTCTGTGTAACTATATAATTTAAATAGATTGGATTAGAAAAAAGAGAAGGAGAATTTGCTATTTGGAGATTTTGTGTGG  900

 901  CTCTCTGCATTCGTTTGTCTTGGTGAATTCATATTATTTGAGCATGTCAAATGCAGTCCATTTATACTCTTTACTGGTGCTTCCTGCATGCTGTCATTGC  1000

1001  TGTGACTTTTCTGTTGAGTTAAATTGTATAGTACAAGGAAAATTACAATAAAAAAAATACCAAAAACAAACCCAACTATTTGACAATGGACATAATACAA  1100
                            ↑↑

1101  CTATCAAAGGTAAAAAGTTCCTTGTTAGGAGACCAAGCAAGTACATTAAGCGGCAACTTCTTGCTCTTTGGTGGCAAGTTGGTTCTGAATGTGCTGAGGC  1200
                                                   *  A  A  V  E  Q  E  K  T  A  L  Q  N  Q  I  H  Q  P
                                                            ←←←
```

Fig. 1. Genomic sequence showing the 3′-terminal region of soybean (*Glycine max* cv. Ceresia) *gmsti*1 and *fus*1, respectively. The amino acid sequence of four exons of GMSTI1 protein and of the C-terminal region of the chloroplast specific translation elongation factor cEF-G [9] are shown. A TPR motif within the topmost exon is underlined. Introns are in small letters; ↓ position of polyadenylation site in *gmsti*1; ↑↑ site of deletion in *fus*1; ←←← transcription polarity of *fus*1.

and the human and yeast counterparts are 44% and 38% respectively (GCG program GAP).

Structure and function of proteins belonging to the TRP family have been discussed in detail [1, 7, 11, 19, 20]. Sikorski *et al.* [20] compared TRP elements from different proteins and defined three classes of repeats with distinct consensus sequences. According to their data the yeast STI1 TPR elements belong together with those of the mitochondrial MAS70 protein to class III with a consensus pattern given on the bottom line of Fig. 3. In Fig. 3 soybean TPR elements are aligned with the corresponding elements of yeast STI1 and human IEF SSP 3521 protein. We find that the soybean repeats follow a consensus pattern with highly conserved positions: 1 (A), 8 (G), 20 (A), 24 (Y), 32 (P). The soybean TPR elements of block I and III show a good matching score with class III (13 out of 26 consensus positions) but the other soybean repeats in blocks II, IV and V show no preference for any of the three classes. Sikorski *et al.* [20] discuss the possibility that different classes of TPR repeats may coexist within a single protein.

Gmsti *gene is heat-inducible*

The human IEF SSP 3521 and the yeast STI1 protein accumulate under stress [12, 13]. We

```
S   1     MA-E--EAKAKGNAAFSAGDFAAAVRHLSDAIAL-SPSNHVLYSNRSAA-TLPPELRGGP
Y   1     MSLTADEYKQQGNAAFTAKDYDKAIELFTKAIEVSETPNHVLYSNRSACYTSLKKFSDAL
H   1     ME-QVNELKEKGNKALSVGNIDDALQCYSEAIKL-DPHNHVLYSNRSAAYAKKGDYQKAY
          *     * * .** *..  .   *.   . ** .    **********  .       .

S  56     SRRQKTVDLKPDWPKAYSRLGAAHLGLRRHRDASPPTKPASHSHPDNAALKSGLADAQ--
Y  61     NDANECVKINPSWSKGYNRLGAAHLGLGDLDEAESNYKKALELDASNKAAKEGLDQVHRT
H  56     EDGCKTVDLKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANNPQLKEGLQNME--
                * ..* * *.*.* .**   *    .*      .   .. *   * ** . .

S 114     AAASRPPPTSPFATAFSGPDMWARSPPTPP-HVQPPGPRVRQDHAGHPEGPQQVQPA-FE
Y 121     QQARQAQPDLGLTQLFADPNLIENLKKNPKTSEMMKDPQLVAKLIGYKQNPQAIGQDLFT
H 117     A----RLAERKFMNPFNMPNLYQKLESDPRTRTLLSDPTYRELIEQLRNKPSDLGTK-LQ
                 .   .   *  *..       *       .*          . *  .    .

S 172     YQRVMHAIGVLLNVKIQTPNHDENDH-----------DADDDVSEDEVVSQPEPEHEPEA
Y 181     DPRLMTINATLMGVDLNMDDINQSNSMPKEPETSKSTEQKKDAEPQSDSTTSKENSSKAP
H 172     DPRIMTTLSVLLGVDL-------GSM-----------DEEEEIATPPPPPPPKKETKPEP
            *.*  .. *. * .                     .   .         .  .    .

S 221     AVEVAEEEEEEEKETRDRKGQAQKEKEAGNAAYKKKDFETAIGHYSKALELDDEDISYLT
Y 241     QKEESKESEPMEVDEDDSKIEADKEKAEGNKFYKARQFDEAIEHYNKAWELHK-DITYLN
H 214     M----------EEDLPENKKQALKEKELGNDAYKKKDFDTALKHYDKAKELDPTNMTYIT
                     * .  . * .* ***  **  ** ..*. *. ** ** **.  ...*.

S 281     NRAAVYLEMGKFEDCIKDCEKAVERGKELRSDYKMIARALTRKGTALAKMAKCSKDFEPA
Y 300     NRAAAEYEKGEYETAISTLNDAVEQGREMRADYKVISKSFARIGNAYHKLG----DLKKT
H 264     NQAAVYFEKGDYNKCRELCEKAIEVGRENREDYRQIAKAYARIGNSYFKEEK----YKDA
          *.**   * * ..      . *.* *.* * **. *... .* * .  *          .

S 341     IEIFQKALTENRNPDTLKKLNEAEKAKKELEQQEYFDPKLADEAREKGNELFKQQKYPEA
Y 356     IEYYQKSLTEHRTADILTKLRNAEKELKKAEAEAYVNPEKAEEARLEGKEYFTKSDWPNA
H 320     IHFYNKSLAEHRTPDVLKKCQQAEKILKEQERLAYINPDLALEEKNKGNECFQKGDYPQA
          *. ..*.*.*.* .* * *  .***  *  *   * .*  * * .  *.* * .   *.*

S 401     TKHYTEAIKRNPKDAKAYSNRAACYTKLGAMPEGLKDAEKCIELDPTFSKGYTRKGAVQF
Y 416     VKAYTEMIKRAPEDARGYSNRAAALAKLMSFPEAIADCNKAIEKDPNFVRAYIRKATAQI
H 380     MKHYTEAIKRNPKDAKLYSNRAACYTKLLEFQLALKDCEECIQLEPTFIKGYTRKAAALE
           * *** *** * **. ******  .**     .. * .  *. .* * ..* **..

S 461     SMKEYDKALETYREG------LKHDPNNQELLDGIRRCVEQINKASRGDFTPEELKERQA
Y 476     AVKEYASALETLDAARTKDAEVNNGSSAREIDQLYYKASQQRFQPGTSNETPEETYQR--
H 440     AMKDYTKAMDVYQKA------LDLDSSCKEAADGYQRCM--MAQYNRHD-SPEDVKRR--
          ..*.*  *..    .      .  ... .*  .   .      .    . .**.   *

S 515     KAMQDPEIQSILQDPVNTQVLTDFQENPRAAEEHVKNPMVMNKTQNVTVPG-CQM-R
Y 534     -AMKDPEVAAIMQDPVMQSILQQAQQNPAALQEHMKNPEVFKKIQTLIAAGIIRTGR
H 489     -AMADPEVQQIMSDPAMRLILEQMQKDPQALSEHLKNPVIAQKIQKLMDVGLIAI-R
           ** ***.  *. ** *  .* . * .* *  **.*** . .* * .   *     *
```

Fig. 2. Comparative alignment (CLUSTAL) of aminoacid sequences from soybean cv. Williams GMSTI (S) with yeast STI1 (Y) and human IEF SSP 3521 (H) sequence. Aminoacid position numbers are on the left margin; asterisks and dots stand for perfect and conservative matches, respectively. The corresponding cDNA sequence perfectly matches the *gmsti*1 but not the *gmsti*2 gene sequence (3′-terminal region as far as sequenced).

```
       1 . . . . . . . . .10 . . . . . . . . . .20 . . . . . . . . . .30 . . . .
S   2  A E E A K A K G N A A F S A G D F A A A V R H L S D A I A L S P S N
Y   5  A E D Y K Q Q G N A A F T A K D Y D K A I E L F T K A I E V S E T P  I
H   4  V N E L K E K G N K A L S V G N I D D A L Q C Y S E A I K L D P N H

S  69  P K A Y S R L G A A H L G L R R H R D A S P P T K P A S N S N P D H
Y  74  S K G Y N R L G A A H L G L G D L D E A E S N Y K K A L E L D A S N  II
H  72  G K G Y S R K A A A L E F L N R F E E A K R T Y E E G L K H E A N N

S 242  A Q K E K E A G N A A Y K K K D F E T A I G H Y S K A L E L D D E D
Y 262  A D K E K A E G N K F Y K A R Q F D E A I E H Y N K A W E L H K D I  III
H 225  A L K E K E L G N D A Y K K K D F D T A L K H Y D K A K E L D P T N

S 381  A D E A R E K G N E L F K Q Q K Y P E A T K H Y T E A I K R N P K D
Y 396  A E E A R L E G K E Y F T K S D W P N A V K A Y T E M I K R A P E D  IV
H 360  A L E E K N K G N E C F Q K G D Y P Q A M K H Y T E A I K R N P K D

S 415  A K A Y S N R A A C Y T K C G A M P E G L K D A E K C I E L D P T F
Y 430  A R G Y S N R A A A L A K L M S F P E A I A D C N K A I E K D P N F  V
H 394  A K L Y S N R A A C Y T K L L E F Q L A L K D C E E C I Q C E P T F

       A . . . . A L G N A L F . A G D Y D E A I . A Y R K A L E L E . . N
```

Fig. 3. Comparison of conserved TPR motifs (34 amino acid repeats) from GMSTI (S) with equivalent TPR motifs (I to V) from yeast STI1 (Y) and human IEF SSP 3521 (H). Bottom line gives the consensus sequence of class III TPR motifs [20]; amino acid positions are on the left margin.

tested heat inducibility of *gmsti* transcription by monitoring the *gmsti* RNA concentration in soybean plants kept at 25 °C and after a temperature shift to 42 °C. Figure 4 shows northern blots of total leaf RNA probed with ^{32}P-labelled single-stranded *gmsti* DNA. In lane 1 (control) we see no clear signal, i.e., the *gmsti* mRNA concentration is very low at 25 °C. However, after 30 min of heat treatment (lane 2) two major size classes of mRNA strongly interact with the DNA probe. The faster moving strong band corresponds in size to the full-length *gmsti* cDNA of about 2 kb, while the upper strong signal most likely corresponds to unspliced (partially spliced) pre-mRNA. We know from genomic DNA sequencing (Fig. 1) that both *gmsti* genes are split and it was reported [22] that heat stress can inhibit the splicing reaction. After 2 h (lane 4) and 3 h (lane 5) of heat stress the *gmsti* mRNA concentration decreases. This may be due to specific changes in *gmsti* mRNA turnover upon prolonged stress or more likely to non-specific degradation of RNA in visibly deteriorating leaves. We repeated the experiment with plants undergoing two consecutive heat stress treatments (Fig. 4B). We see again that after 30 min exposure to 42 °C (lane 2) two major bands appear. Subsequently the plants were kept

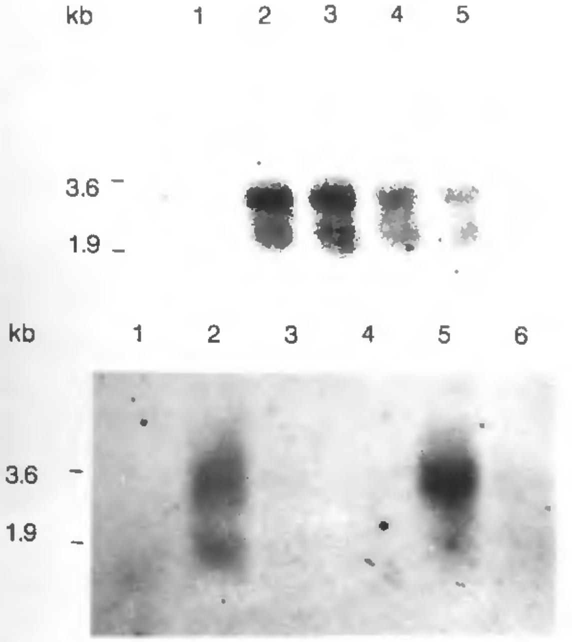

Fig. 4. Northern blot analysis of total soybean RNA. 20 μg of total RNA isolated from leaves of plants grown under different temperature conditions were loaded on each lane. Blots were probed with a 1.1 kb single-strand ^{32}P-labelled *gmsti1* cDNA probe [14]. A (top) 1, control, 25 °C; 2 to 5, 30 min, 1, 2, 3 h, respectively, at 42 °C. B. 1, control; 25 °C; 2, 30 min, 42 °C; 3, re-exposed for 30 min, 25 °C; 4, re-exposed for 1 h, 25 °C; 5, re-exposed a second time for 30 min, 42 °C; 6, re-exposed a second time for 30 min, 25 °C.

for 30 min (lane 3) and 1 h (lane 4) at 25 °C. The two major broad bands interacting with *gmsti* DNA disappear almost completely. Bringing plants back to 42 °C for 30 min induces again the synthesis of *gmsti* mRNA (lane 5) and after 30 min at 25 °C the strong signals become very faint again. Obviously the steady-state concentration of *gmsti* mRNA is very low in plants grown at 25 °C (essentially below detectability under the given experimental conditions) but its concentration sharply increases under heat stress. At least one of the two *gmsti* genes is transcribed under non-stress conditions as documented by the presence of a full-length cDNA clone.

We have not yet isolated and characterized a putative GMSTI protein. It should have a MW in the range of 63 350 and a pI of 6.07. A soybean heat shock nuclear gene of about equal size (*hsp70*) was characterized [15] but it does not have any structural relationship to the GMSTI protein. Currently we do not know the function of GMSTI protein but it may have functions similar to the yeast STI1 protein which seems to play a role in mediating the heat shock response to some HSP70 proteins. According to Nicolet and Craig [13] yeast STI1 is essential for cells growing under non-optimal conditions. Plants are exposed to several kinds of stress. The GMSTI protein may be an essential component required for activating the complex stress response as described [21].

Multiple TPR elements arranged in tandem have the capacity to form helix-turn domains and could be assembled into filamentous structures [11]. The function of TPR repeats, however, is not understood. It is tempting to speculate that the multiple repeats in GMSTI and in its yeast and human counterparts are functionally important. Otherwise it would be hard to understand why these repeats are so well conserved in yeast, plants and man.

Acknowledgements

We thank Dr A. Schori (Station fédérale de recherche agronomique, Nyon) for the soybean seeds. This work receives support from the Fonds national suisse de la recherche scientifique (E.S.). J.H.T. was recipient of a Swiss Government student fellowship. This work is part of the Ph.D. thesis submitted by J.H.T. to the *Faculté des Sciences, Université de Neuchâtel.*

References

1. Boguski MS, Silorski RS, Hieter P, Goebl M: Expanding family. Nature 346: 114–114 (1990).
2. Bullock W, Fernandez J, Short J: XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with β-galactosidase selection. Biotechniques 5: 376–378 (1987).
3. Chou PY, Fasman GD: Predication of protein conformation. Biochemistry 13: 122–145 (1974).
4. Dean C, van den Elzen P, Tamaki S, Dunsmuir P, Bedbrook J: Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multi-gene family. EMBO J 4: 3055–3061 (1985).

5. Devereux J, Haeberli P, Smithies O: Comprehensive set of sequence analysis programs for the Vax. Nucl Acids Res 12: 387–395 (1984).
6. Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137: 266–267 (1984).
7. Goebl M, Yanagida M: A novel protein repeat motif from mitosis to transcription. Trends Biochem Sci 16: 173–177 (1991).
8. Graham M, Doherty JP, Woodcock DM: Efficient construction of plant genomic libraries requires the use of mcr^- host strains and packaging mixes. Plant Mol Biol Rep 8: 18–27 (1990).
9. Hernández Torres J, Breitenberger C, Spielmann A, Stutz E: Cloning and sequencing of a soybean nuclear gene coding for a chloroplast translation elongation factor EF-G. Biochim Biophys Acta 1174: 191–194 (1993).
10. Hernández Torres J, Chatellard P, Stutz E: Analysis of nuclear *fus* genes coding for the chloroplast translation elongation factor cEF-G in soybean. Experientia 50: A24 (1994).
11. Hirano T, Kinoshita N, Morikawa K, Yanagida M: Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein *nuc* 2 +. Cell 60: 319–328 (1990).
12. Honoré B, Leffers H, Madsen P, Rasmussen HH, Vandekerckhove J, Celis JE: Molecular cloning and expression of a transformation-sensitive human protein containing the TPR motif and sharing identity to the stress-inducible yeast protein STI1. J Biol Chem 267: 8485–8491 (1992).
13. Nicolet CM, Craig E: Isolation and characterization of STI1, a stress-inducible gene from *Saccharomyces cerevisiae*. Mol Cell Biol 9: 3638–3646 (1989).
14. Orrú D, Grosso N, Angeletti B, D'Ambrossio E: An improved protocol to generate high-efficiency single-stranded DNA probes by PCR. Biotechniques 14: 905–906 (1993).
15. Roberts JK, Key JL: Isolation and characterization of a soybean *hsp70* gene. Plant Mol Biol 16: 671–683 (1991).
16. Sambrook J, Fritsch E, Maniatis T: *Molecular Cloning: A Laboratory Manual*, 2nd ed.: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
17. Sanger F, Nicklen S, Coulson A: DNA sequencing with chain-termination inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
18. Schori A, Fossati A, Soldati A, Stamp P: Cold tolerance in soybean (*Glycine max* L. Merr.) in relation to flowering habit, pod set and compensation for lost reproductive organs. Eur J Agron 2: 173–178 (1993).
19. Sikorski RS, Boguski MS, Goebl M, Hieter P: A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell 60: 307–317 (1990).
20. Sikorski RS, Michaud WA, Wootton JC, Boguski MS, Connelly C, Hieter P: TPR proteins as essential components of the yeast cell cycle. Cold Spring Harbor Symp Quant Biol 56: 663–673 (1991).
21. Sorger PK: Heat shock factor and the heat shock response. Cell 65: 363–366 (1991).
22. Yost HJ, Lindquist S: RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell: 185–193 (1986).

1343

IMPRIMATUR POUR LA THÈSE

Analysis of the nuclear genes (*fus)* coding for the
chloroplast specific translation elongation
factor cEF-G in soybean (Glycine max)

de M. Jorge Hernandez Torres

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel sur le rapport des membres du jury,

MM. E. Stutz (directeur de thèse),
P. Schürmann et A. Boschetti (Berne)

autorise l'impression de la présente thèse.

Neuchâtel, le 12 février 1997

Le doyen:

R. Dändliker

R. Dändliker