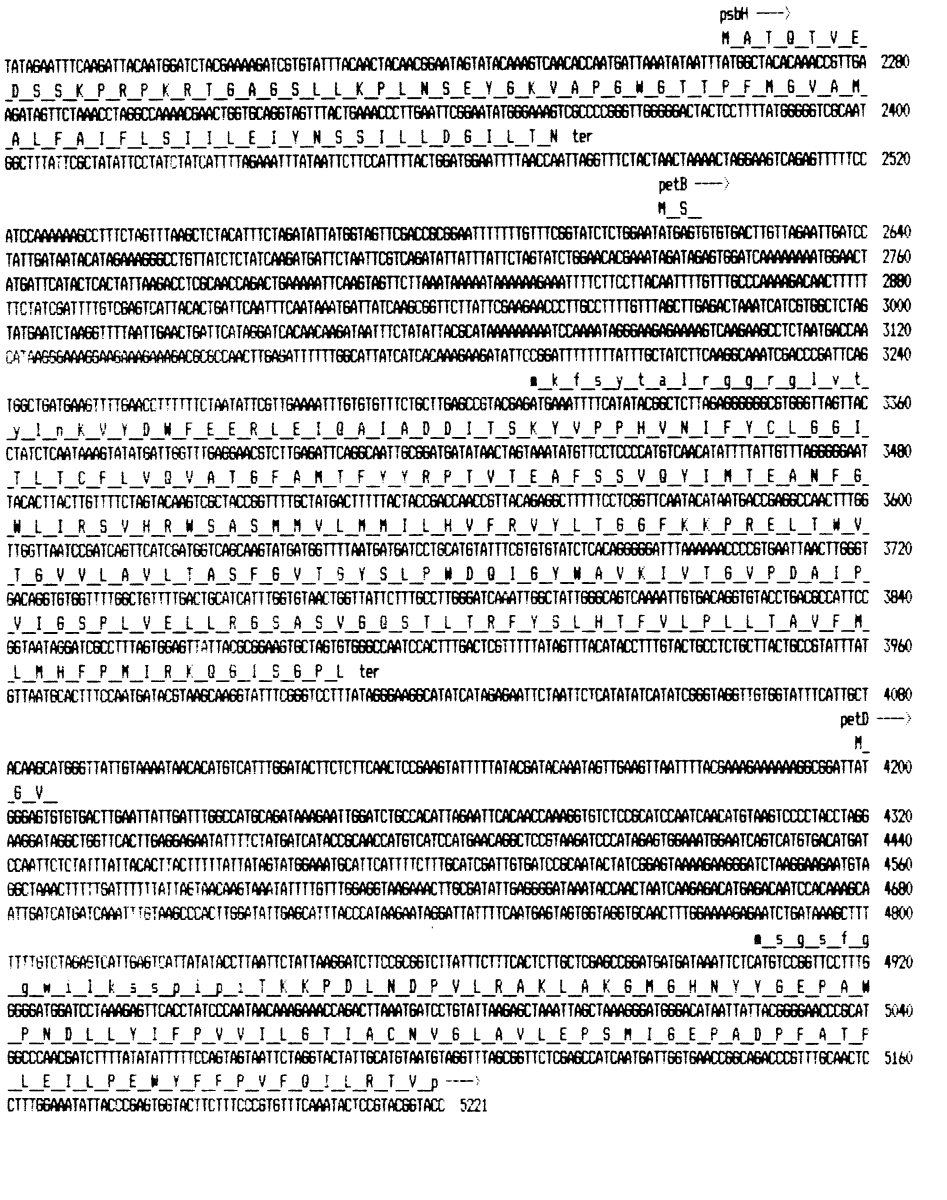
# Nucleotide sequence of the 5.2 kbp barley chloroplast DNA fragment, containing *psbB-psbH-petB-petD* gene cluster

S.V.Reverdatto, A.V.Andreeva, A.A.Buryakova, O.G.Chakhmakhcheva and V.A.Efimov

Shemyakin Institute of Bioorganic Chemistry, ul. Miklukho-Maklaya 16/10, Moscow 117871, USSR
Submitted February 23, 1989 EMBL accession no. X14107

This sequence is a 5.2 kbp Pst I - Kpn I fragment of barley chloroplast DNA, containing *psbB*, *psbH*, *petB* and *petD* genes. This genes code for 47 kDa protein and 10 kDa phosphoprotein of photosystem II, cytochrome b6 and subunit 4 of b6/f cytochrome complex, respectively. Sequence showed close similarity to the corresponding maize chloroplast DNA fragment ( see Curr.Genet., 1987, 12:69 ) not only in the coding, but also in the intergenic regions. Deduced amino acid sequences are shown. Supposed N-terminal extensions (alternative splicing) of *petB* and *petD* are shown in small letters.

```
CTGCAGTTGCCATGTTTTACGGCAGTGAGAGCAGAGATAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACAGA 120
                                                    psbB ---->
                                                    M  G  L  P  W  Y  R  V  H  T  V  V  L  N  D  P  G  R  L  L  S
AGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAGCATCAATGGGTTTGCCTTGGTATCGTGTTCATACTGTCGTATTGAATGATCCGGGTCGATTGCTTT 240
  V  H  I  M  H  T  A  L  V  S  G  W  A  G  S  M  A  L  Y  E  L  A  V  F  D  P  S  D  P  V  L  D  P  M  W  R  Q  G  M  F
CGGTGCATATAATGCACACAGCTCTAGTTTCTGGTTGGGCTGGCTCAATGGCTTTATACGAATTAGCAGTTTTTGATCCCTCTGATCCTGTTCTGGATCCAATGTGGAGACAAGGTATGT 360
  V  I  P  F  M  T  R  L  G  I  T  D  S  W  G  G  W  S  I  S  G  G  T  V  T  N  P  G  I  W  S  Y  E  G  V  A  A  T  H  I
TCGTAATTCCCTTCATGACTCGTTTAGGAATAACGGATTCGTGGGGTGGTTGGAGTATTTCAGGAGGAACTGTAACAAATCCGGGTATTTGGAGTTATGAAGGTGTGGCAGCTACGCATA 480
  V  F  S  G  L  C  F  L  A  A  I  W  H  W  V  Y  W  D  L  E  I  F  S  D  E  R  T  G  K  P  S  L  D  L  P  K  I  F  G  I
TTGTGTTTTCTGGCTTGTGTTTCTTGGCAGCGATCTGGCATTGGGTATATTGGGACCTAGAAATATTCTCTGATGAGCGGACGGGAAAACCCTCTTTGGATTTGCCCAAGATCTTTGGAA 600
  H  L  F  L  A  G  V  A  C  F  G  F  G  A  F  H  V  T  G  L  Y  G  P  G  I  W  V  S  D  P  Y  G  L  T  G  K  V  Q  A  V
TTCATTTATTTCTTGCAGGGGTGGCTTGCTTTGGCTTTGGGGCATTTCATGTAACGGGTTTGTATGGTCCTGGGATATGGGTATCCGATCCTTATGGACTAACTGGAAAAGTACAAGCTG 720
  N  P  A  W  G  A  E  G  F  D  P  F  V  P  G  G  I  A  S  H  H  I  A  A  G  T  L  G  I  L  A  G  L  F  H  L  S  V  R  P
TAAATCCAGCGTGGGGTGCAGAAGGTTTGATCCTTTTGTTCCGGGGGGAATAGCTTCTCATCATATTGCTGCGGGTACATTGGGTATATTAGCGGGCTTATTCCATCTTAGTGTCCGTC 840
  P  Q  R  L  Y  K  G  L  R  M  G  N  I  E  T  V  L  S  S  S  I  A  A  V  F  F  A  A  F  V  V  A  G  T  M  W  Y  G  S  A
CGCCTCAACGTCTATATAAAGGATTACGTATGGGCAATATTGAAACTGTACTTTCCAGTAGTATCGCTGCTGTTTTTTTTGCAGCTTTCGTAGTTGCTGGAACTATGTGGTATGGGTCAG 960
  T  T  P  I  E  L  F  G  P  T  R  Y  Q  W  D  Q  G  Y  F  Q  Q  E  I  Y  R  R  V  S  N  G  L  A  E  N  L  S  L  S  E  A
CAACGACCCCAATCGAATTATTTGGGCCTACTCGTTATCAGTGGGATCAGGGATACTTTCAGCAAGAAATATATCGAAGAGTTAGCAATGGTTTAGCCGAAAATCTTAGTTTATCAGAAG 1080
  W  S  K  I  P  E  K  L  A  F  Y  D  Y  I  G  N  W  P  A  K  G  G  L  F  R  A  G  S  M  D  N  G  D  G  I  A  V  G  W  L
CTTGGTCTAAAATTCCCGAAAAATTAGCCTTTTATGATTATATTGGTAATAATCCGGCAAAAGGGGGATTATTCAGAGCGGGCTCAATGGACAATGGGGATGGAATAGCTGTTGGATGGT 1200
  G  H  P  V  F  R  D  K  E  G  R  E  L  F  V  R  R  M  P  T  F  F  E  T  F  P  V  V  L  V  D  E  E  G  I  V  R  A  D  V
TAGGACATCCCGTCTTAGAGATAAAGAAGGACGTGAGCTTTTTGTACGCCGTATGCCTACTTTTTTTGAAACATTTCCGGTTGTTTTGGTAGATGAGGAGGGAATTGTTAGAGCGGACG 1320
  P  F  R  R  A  E  S  K  Y  S  V  E  Q  V  G  V  T  V  E  F  Y  G  G  E  L  N  G  V  N  Y  S  D  P  A  T  V  K  K  Y  A
TTCCTTTTAGAAGAGCAGAATCCAAATATAGTGTTGAACAAGTAGGCGTAACGGTGGAGTTCTATGGTGGCGAACTTAATGGAGTAAATTATTCTGATCCTGCTACTGTAAAAAAATATG 1440
  R  R  S  Q  L  G  E  I  F  E  L  D  R  A  T  L  K  S  D  G  V  F  R  S  S  P  R  G  W  F  T  F  G  H  A  T  F  A  L  L
CGAGGCGTTCTCAATTAGGGGAAATTTTTGAATTAGACGGGCTACTTTGAAATCCGATGGTGTTTTTCGCAGCAGTCCAAGGGGTGGTTCACTTTTGGTCATGCTACCTTTGCTTTGC 1560
  F  F  F  G  H  I  W  H  G  A  R  T  L  F  R  D  V  F  A  G  I  D  P  D  L  D  A  Q  V  E  F  G  T  F  Q  K  V  G  D  P
TCTTCTTTTTCGGACACATTTGGCATGGCGCTAGAACATTGTTCCGAGATGTTTTTGCTGGTATTGATCCAGATTTGGATGCTCAAGTGGAATTTGGAACATTCCAAAAAGTGGGAGATC 1680
  T  T  K  K  Q  A  V  ter
CAACTACAAAGAAACAGGCAGTCTGATACACATTGTTATGGTATCTTTCACCTCTCTTTTTTGATTTGACATCTCCCATCCTTTCTTTGACTCTTTTTCTTTCTTTATATGGGAAATTCT 1800
CCCAAATGACAAATGAATAGGTGTGGAAGTTATAATTGTAAATAAACCACGATCGAATCTATGGAAGCATTGGTTTATACGTTCCTTTTAGTTTCGACTTTAGGGATAATTTTTTTTCGCT 1920
ATCTTCTTCCGAGAACCACCTAAGGTTCCACCAACTCCAACTAAAAGAATAAAATAATTTCATTTAAGTAAGAAGTCTCCCAGATAGGGGGACTTCTTACTTAAATTAGTCTCCGTGTTC 2040
TTCGAATGGATCTCTTAATTGTTGAGAGGGTTGCCCAAACGCGGTATATAAGGCATACCCAGTAAAGCTTACAGTAAACCAGATATGGAGATGGCGACTAAAGTTGCTGTTTCCATTTT 2160
```


psbH ---->
M A T Q T V E
TATAGAATTTCAAGATTACAATGGATCTACGAAAAGATCGTGTATTTACAACTACAACGGAATAGTATACAAAGTCAACACCAATGATTAAATATAATTTATGGCTACACAAACCGTTGA 2280
D S S K P R P K R T G A G S L L K P L N S E Y G K V A P G W G T T P F M G V A M
AGATAGTTCTAAACCTAGGCCAAAACGAACTGGTGCAGGTAGTTTACTGAAACCCTTGAATTCGGAATATGGGAAAGTCGCCCCGGGTTGGGGGACTACTCCTTTTATGGGGGTCGCAAT 2400
A L F A I F L S I I L E I Y N S S I L L D G I L T N ter
GGCTTTATTCGCTATATTCCTATCTATCATTTTAGAAATTTATAATTCTTCCATTTTACTGGATGGAATTTTAACCAATTAGGTTTCTACTAACTAAAACTAGGAAGTCAGAGTTTTTCC 2520
petB ---->
M S
ATCCAAAAAAGCCTTTCTAGTTTAAGCTCTACATTTCTAGATATTATGGTAGTTCGACCGCGGAATTTTTTTGTTTCGGTATCTCTGGAATATGAGTGTGTGACTTGTTAGAATTGATCC 2640
TATTGATAATACATAGAAAGGGCCTGTTATCTCTATCAAGATGATTCTAATTCGTCAGATATTATTTATTCTAGTATCTGGAACACGAAATAGATAGAGTGGATCAAAAAAATGGAACT 2760
ATGATTCATACTCACTATTAAGACCTCGCAACCAGACTGAAAAATTCAAGTAGTTCTTAAATAAAAATAAAAAAGAAATTTTCTTCCTTACAATTTTGTTTGCCCAAAAGACAACTTTTT 2880
TTCTATCGATTTTGTCGAGTCATTACACTGATTCAATTTCAATAAATGATTATCAAGCGGTTCTTATTCGAAGAACCCTTGCCTTTTGTTTAGCTTGAGACTAAATCATCGTGGCTCTAG 3000
TATGAATCTAAGGTTTTAATTGAACTGATTCATAGGATCACAACAAGATAATTTCTATATTACGCATAAAAAAAATCCAAAATAGGGAAGAGAAAAGTCAAGAAGCCTCTAATGACCAA 3120
CATAAGGGAAAGGAAGAAGAAAGAACGCGCCAACTTGAGATTTTTTGGCATTATCATCACAAAGAAGATATTCCGGATTTTTTTATTTGCTATCTTCAAGGCAAATCGACCCGATTCAG 3240
a k f s y t a l r g g r g l v t
TGGCTGATGAAGTTTTGAACCTTTTTTCTAATATTCGTTGAAAATTTGTGTGTTCTGCTTGAGCCGTACGAGATGAAATTTTCATATACGGCTCTTAGAGGGGGGCGTGGGTTAGTTAC 3360
y l n K V Y D W F E E R L E I Q A I A D D I T S K Y V P P H V N I F Y C L G G I
CTATCTCAATAAAGTATATGATTGGTTGAGGAACGTCTTGAGATTCAGGCAATTGCGGATGATATAACTAGTAAATATGTTCCTCCCCATGTCAACATATTTTATTGTTTAGGGGGAAT 3480
T L T C F L V Q V A T G F A M T F Y Y R P T V T E A F S S V Q Y I M T E A N F G
TACACTTACTTGTTTTCTAGTACAAGTCGCTACCGGTTTTGCTATGACTTTTTACTACCGACCAACCGTTACAGAGGCTTTTTCCTCGGTTCAATACATAATGACCGAGGCCAACTTTGG 3600
W L I R S V H R W S A S M M V L M M I L H V F R V Y L T G G F K K P R E L T W V
TTGGTTAATCCGATCAGTTCATCGATGGTCAGCAAGTATGATGGTTTTAATGATGATCCTGCATGTATTTCGTGTGTATCTCACAGGGGGATTTAAAAAACCCCGTGAATTAACTTGGGT 3720
T G V V L A V L T A S F G V T G Y S L P W D Q I G Y W A V K I V T G V P D A I P
GACAGGTGTGGTTTTGGCTGTTTGACTGCATCATTTGGTGTAACTGGTTATTCTTTGCCTTGGGATCAAATTGGCTATTGGGCAGTCAAAATTGTGACAGGTGTACCTGACGCCATTCC 3840
V I G S P L V E L L R G S A S V G Q S T L T R F Y S L H T F V L P L L T A V F M
GGTAATAGGATCGCCTTTAGTGGAGTTATTACGCGGAAGTGCTAGTGTGGGCCAATCCACTTTGACTCGTTTTATAGTTTACATACCTTTGTACTGCCTCTGCTTACTGCCGTATTTAT 3960
L M H F P M I R K Q G I S G P L ter
GTTAATGCACTTTCCAATGATACGTAAGCAAGGTATTTCGGGTCCTTTATAGGGAAGGCATATCATAGAGAATTCTAATTCTCATATATCATATCGGGTAGGTTGTGGTATTTCATTGCT 4080
petD ---->
M
ACAAGCATGGGTTATTGTAAAATAACACATGTCATTTGGATACTTCTCTTCAACTCCGAAGTATTTTTATACGATACAAATAGTTGAAGTTAATTTTACGAAAGAAAAAAGGCGGATTAT 4200
G V
GGGAGTGTGTGACTTGAATTATTGATTTGGCCATGCAGATAAAGAATTGGATCTGCCACATTAGAATTCACAACCAAAGGTGTCTCCGCATCCAATCAACATGTAAGTCCCCTACCTAGG 4320
AAGGATAGGCTGGTTCACTTGAGGAGAATATTTTCTATGATCATACCGCAACCATGTCATCCATGAACAGGCTCCGTAAGATCCCATAGAGTGGAAATGGAATCAGTCATGTGACATGAT 4440
CCAATTCTCTATTATTACACTTACTTTTTATTATAGTATGGAAATGCATTCATTTTCTTTGCATCGATTGTGATCCGCAATACTATCGGAGTAAAGAAGGGATCTAAGGAAGAATGTA 4560
GGCTAAACTTTTTGATTTTTATTAETAACAAGTAAATATTTTGTTTGGAGGTAAGAAACTTGCGATATTGAGGGGATAAATACCAACTAATCAAGAGACATGAGACAATCCACAAAGCA 4680
ATTGATCATGATCAAATTTGTAAGCCCACTTGGATATTGAGCATTTACCCATAAGAATAGGATTATTTTCAATGAGTAGTGGTAGGTGCAACTTTGGAAAAGAGAAATCTGATAAAGCTTT 4800
s g s f g
TTTTGTCTAGAGTCATTGAGTCATTATATACCTTAATTCTATTAAGGATCTTCCGCGGTCTTATTTCTTTCACTCTTGCTCGAGCCGGATGATGATAAATTCTCATGTCCGGTTCCTTTG 4920
g w i l k s s p i p T K K P D L N D P V L R A K L A K G M G H N Y Y G E P A W
GGGGATGGATCCTAAAGAGTTCACCTATCCCAATAACAAAGAAACCAGACTTAAATGATCCTGTATTAAGAGCTAAATTAGCTAAAGGGATGGGACATAATTATTACGGGGAACCCGCAT 5040
P N D L L Y I F P V V I L G T I A C N V G L A V L E P S M I G E P A D P F A T P
GGCCCAACGATCTTTTATATATTTTTCCAGTAGTAATTCTAGGTACTATTGCATGTAATGTAGGTTTAGCGGTTCTCGAGCCATCAATGATTGGTGAACCGGCAGACCCGTTTGCAACTC 5160
L E I L P E W Y F F P V F Q I L R T V p ---->
CTTTGGAAATATTACCCGAGTGGTACTTCTTTCCCGTGTTTCAAATACTCCGTACGGTACC 5221