The nucleotide sequence of potato virus X RNA

K.G.Skryabin, A.S.Kraev, S.Yu.Morozov[1], M.N.Rozanov, B.K.Chernov[1], L.I.Lukasheva[1] and J.G.Atabekov[1]*

The complete nucleotide sequence of potato virus X (PVX, a potexvirus) genomic RNA was determined using dideoxy method on M13-cloned cDNA templates. The RNA is 6435 nucleotides long (excluding poly(A)). Computer analysis of the (+)RNA primary structure revealed five open reading frames coding for the proteins with Mr 165,405 (165K), Mr 24,563 (24.5K), Mr 12,201 (12K), Mr 7,595 (8K) and Mr 25,055 (25K). The corresponding gene positions are: 85-4452 (165K); 4486-5163 (24.5K); 5147-5490 (12K); 5427-5636 (8K); 5650-6360 (25K). Amino acid sequence analysis predicts following functions of these polypeptides: 25K - coat protein (1); 12K and 8K polypeptides are the potential membrane-bound proteins (2); 165K - RNA-replicating protein (3); the central domain of the 165K and 24.5K proteins - putative NTP-binding helicase(s) (4,5). The 165K protein is translated directly from the PVX genomic RNA (6). Expression of the coat protein gene requires the specific subgenomic RNA which starts at position 5645, as it was shown by primer extension analysis; start sites for the two minor subgenomic mRNAs were predicted at the positions 4479 and 5089 (S.Yu.Morozov, unpublished).

```
   1 GAAAACTAAACCATACACCAACAACACAACCAAACCCACCACGCCCAATTGTTACACACCCGCTTGGAAAAGTAAGTCTAACAAATGGCCAAAGTGCGCG
 101 AGGTTTACCAATCCTTTACAGACTCCACCACAAAAACTCTCATCCAAGATGAGGCTTATAGAAATATTCGTCCCATCATGGAAAAACATAAACTAGCTAA
 201 CCCGTACGCTCAAACGGTTGAAGCGGCTAATGATCTAGAGGGGTTCGGCATAGCCACCAATCCCTATAGCATTGAATTGCATACACATGCAGCCGCTAAG
 301 ACCATAGAGAACAAACTTCTAGAGGTGCTTGGTTCCATCCTACCACAAGAACCTGTTACATTTATGTTCCTTAAACCCAGGAAGCTAAACTACATGAGAA
 401 GAAACCCGCGGATCAAGGACATTTTCCACAATGTTGCCATTGAACCGAGAGATGTAGCAAGGTACCCCAAGGAAACAATAATTGACAAACTCACAGAGAT
 501 CACAACGGACACAGCATACATCGGTGACACTCTGCACTTCTTGGATCCGAGCTACATAGTGGAGACATTCCAAAACTGCCCAAAATTGCAAACATTGTAT
 601 GCAACCTTAGTTCTCCCCGTTGAGGCCGCCTTCAAAATGGAAAGCACTCACCCGAATATATACAGCCTTAAATACTTCGGAGATGGTTTCCAGTATATAC
 701 CAGGCAACCATGGTGGTGGAGCATACCATCATGAATTTTCTCATTTACAATGGCTCAAAGTGGGAAAGATCAAGTGGAGGGACCCCAAGGATAGCTTTCT
 801 CGGGCATCTCAATTACACGACTGAGCAGGTGAGATGCACACAGTGACAGTACAGTTGCAGGAATCGTTCGCGGCAAACCACTTGTACTGCATCAGGAGA
 901 GGAGACTTACTCACACCGGAGGTACGCACTTTCGGCCAACCTGACAGGTATGTGATTCCACCACAGATCTTTCTTCCGAAAGTTCACAACTGCAAGAAGC
1001 CGATTCTTAAGAAAACTATGATGCAGCTCTTCTTGTATGTTAGGACAGTCAAGGTCGCAAAAAATTGTGACATTTTGCCAAAGTCAGACAATTAATTAA
1101 ATCATCTGACTTGGACAAATATTCTGCTGTGGAACTGGTTTACTTAGTAAGCTATATGGAGTTCCTTGCCGATTTACAAGCCACCACCTGCTTCTCAGAC
1201 ACACTTTCTGGTGGCTTGCTAACAAAGACCCTTGCACCGGTGAGGGCTTGGATACAAGAGAAAAAGATGCAACTATTTGGTCTTGAGGACTACGCGAAGI
1301 TAGTCAAAGCAGTTGATTTCCACCCAGTGGATTTTTCTTTCAAGGTGGAAACTTGGGACTTCAGATTCCACCCATTGCAAGCGTGGAAAGCCTTCCGACC
1401 AAGGGAAGTGTCGGACGTAGAGGAAATGGAAAATTTGTTCTCTGATGGGGACCTGCTCGATTGCTTCACGAGAATGCCAGCTTATGCAGTAAACGCAGAG
1501 GAAGATTTGGCTGCAATCAGGAAAACGCCGGAGATGGATGCCGGTCAAGAAGTTAAGGAACCTGCAGGAGACAGAAATCAATACTCAAACCCTGCAGAAA
1601 CTTTCCTCAGCAAGCTCCACAGGAAACACAGTAGGGGAGGTGAAACATCAGGCCGCAAAGAAAGCTAAACGCCTAGCCGAAATCCAGGAGTCCATGAGAGC
1701 TGAAGGTGAGGCCGAATCAAATGAGATGAGCGGGGGCATGGGGGCAATACCCAGTAACGCCGAACTTCCCAGCACAAGTGGTGCCAGACAAGAACTCACA
1801 CTCCCAACCACTAAGCCTGTTCCTGCAAGGTGGGAAGATGCCTCATTCACAGACTCTAGTGTGGAAGAGGAGCAGGTGAAACTCCCTGGAAAAGAAGCCG
1901 TTGAGACAGCAACGCAACAAGTCATCGAAGGACTCCCTTGGACCACTGGATTCCTCAACTAAATGCTGTTGGATTCAAGGCGCTGGAGATTCAGAGGGGA
2001 TCGAAGTGGAACAATGATCATGCCCATCACAGAAATGGTCTCCGGGTTGGAAAAAGAGGATTTCCCGGAAGGAACTCCAAAAGAGTTGGCACGAGAATTG
2101 CTCGCTATGAACAGAAGCCCTGCCACCATCCCTTTGGACCTGCTTAGAGCCAGAGACTACGGCAGTGATGTGAAAAACAAGAGAATTGGTGCCATCACAA
2201 AGACACAAGCAGCAAGTTGGGGCGAGTACCTAACAGGAAAGATAGAAAGCCTGACTGAGAGGAAAGTTGCGGCCTGCGTCATTCATGGAGCTGGAGGCTC
2301 AGGGAAAAGTCATGCCATCCAGAAGGCATTGAGAGAAATTGGCAAGGGCTCAGACATCACTGTCGTCCTGCCGACCAATGAACTGCGACTAGATTGGAGT
2401 AAGAAAGTGCCTAACACTGAACCATATATGTTCAAAACCTATGAAAAGGCGTTAATTGGGGGAACTGGCAGTATAGTCATCTTTGACGATTACTCAAAAC
2501 TTCCTCCCGGTTACATCGAAGCCTTAATCTGTTTCTACTCTAAAATCAAGCTAGTCATTCTAACAGGAGATAGCAGACAGAGCGTCTATCACGAAACTGC
2601 TGAAGACGCCTCCATCAGGCATTTGGGACCGGCGACAGAGTACTTCTCAAAATACTGCAGATACTATCTCAATGCTACACACCGCAATAAGAAGGACCTT
2701 GCAAACATGCTTGGTGTCTACAGTGAGAGAACGGGAGTCACCGAAATCAGCATGAGCGCCGAGTTCTTAGAAGGAATCCCAACTTTAGTACCCTCGGATG
```

```
2801  AGAAGAGAAGGCTGTACATGGGCACTGGGAGGAATGACACATTCACATACGCTGGATGTCAGGGGCTAACTAAGCCGAAAGTACAAATAGTGTTGGACCA
2901  CAACACCCAAGTGTGTAGCGCGAATGTGATGTACACGGCACTTTCTAGAGCCACCGACAGGATTCACTTCGTGAACACAAGTGCAAACTCTTCGGCCTTC
3001  TGGGAAAAGTTAGACAGCACCCCTTATCTCAAGACTTTCCTATCAGTGGTGAGAGAACAAGCGCTCAAGGAGTACGAGCCGGCAGAGGCAGAGCCAATTC
3101  GAGAGCCTGAGCCCCAGACACACATGTGTGTCGAGAATGAGGAGTCCGTGCTAGAGGAGTACAAAGAGGAACTCTTGGAAAAGTTCGACAGAGAGATCCA
3201  CTCTGAATCCCATGGTCATTCGAACTGTGTCCAAACTGAAGACACAACCATTCAGTTGTTTTCGCATCAACAAGCAAAAGATGAGACCCTCCTCTGGGCA
3301  ACTATAGATGCGCGGCTCAAGACCAGCAATCAAGAAACAAACTTCCGAGAATTCTTGAGCAAGAAGGACATTGGGGACGTTCTGTTTTTGAACTACCAAA
3401  AAGCTATGGGTTTGCCCAAAGAGCGTATTCCCTTCTCACAAGAGGTCTGGGAAGCTTGTGCCCACGAAGTACAGAGCAAGTACCTCAGTAAGTCAAAGTG
3501  CAACTTAATCAATGGGACTGTGAGACAGAGTCCAGACTTTGACGAAAACAAAATTATGGTATTCCTCAAGTCGCAGTGGGTCACAAAGGTGGAAAAGCTA
3601  GGTCTACCCAAAATTAAGCCAGGTCAAACCATAGCAGCTTTTTATCAGCAGACTGTGATGCTTTTTGGAACTATGGCCAGGTACATGCGATGGTTCAGAC
3701  AGGCTTTCCAGCCAAAAGAAGTCTTCATAAACTGTGAGACTACGCCAGAGGACATGTCTGCATGGGCCTTGAACAACTGGAACTTCAGCAGACCTAGCTT
3801  GGCTAATGACTATACAGCTTTCGACCAGTCTCAGGATGGAGCTATGCTGCAATTTGAGGTGCTCAAGGCCAAGCACCACTGCATACCAGAGGAAATCATC
3901  CAAGCATACATAGACATTAAGACCAATGCACACATTTTCCTAGGCACATTGTCAATTATGCGCCTGACTGGTGAAGGTCCCACTTTTGATGCAAACACTG
4001  AGTGTAACATAGCGTACACCCATACAAAGTTCGACATCCCAGCCGGAACTGCCCAAGTGTATGCAGGAGATGACTCAGCACTGGATTGCGTTCCAGAGGT
4101  GAAGCATAGTTTCCACAGGCTTGAAGACAAATTACTCCTCAAGTCAAAGCCCGTAATCACACAACAAAAGAAAGGCAGTTGGCCTGAATTTTGTGGTTGG
4201  CTGATTACACCAAAAGGGGTAATGAAAGACCCAATTAAGCTCCATGTTAGCTTAAAATTGGCCGAAGCTAAGGGTGAACTCAAGAAATGTCAAGATTCCT
4301  ATGAAATTGATCTGAGTTATGCTTATGACCACAAGGACTCTCTGCATGACTTGTTCGATGAGAAACAGTGTCAGGCACATACACTCACTTGCAGGACACT
4401  AATCAAGTCAGGGAGAGGCACTGTCTCACTTCCCCGCCTCAGGAACTTTCTTTAACCGTTAAGTTACCTTAGAGATTTGAATAAGATGGATATTCTCATC
4501  ATTAGTTTGAAAAGTTTAGGTTATTCTAGGACTCATAAATCTTTAGATTCAGGACCTTTGGTAGTACATGCAGTAGCCGGAGCCGGTAAGTCCACAGCCC
4601  TAAGGAAGTTGATCCTCAGACACCCAACATTCACTGTGCATACACTCGGTGTCCCTGACAAGGTGAGCATCAGAACTAGAGGCATACAGAAGCCAGGACC
4701  TATTCCTGAGGGCAATTTCGCAATCCTCGATGAGTATACTTTGGACAACACCACAAGGAACTCATACCAGGCACTTTTTGCTGACCCTTATCAGGCACCT
4801  GAGTTTAGCCTAGAGCCCCACTTCTACTTGGAAACCTCATTTCGAGTTCCGGAGAAAGTGGCAGATTTGATAGCAGGCTGTGGCCTTCGATTTCGAGACCA
4901  ACTCACAGGAAGAAGGGCACTTAGAGATCACTGGCATATTTAAAGGGCCCCTACTTGGGAAGGTGATAGCCATTGACGAGGAGTCCGAGACAACACTGTC
5001  CAAACATGGTGTTGAGTTTGTTAAGCCCTGCCAAGTGACTGGACTTGAGTTCAAAGTAGTCACTGTTGTGTCTGCCGCACCAATAGAGGAAATTGGCCAG
5101  TCCACAGCTTTCTACAACGCCATCACCAGGTCAAAGGGACTGACATATGTCCGCGCAGGGACATAGACTCACCGCTCCGGTCAATTCTGAAAAAGTGTAC
5201  ATAGTATTAGGTCTATCATTTGCTTTAGTTTCAATTACCTTTCTGCTTTCTAGAAATAGTTTACCCCACGTAGGTGATAACATTCACAGTTTACCACACG
5301  GAGGAGCTTACAGAGACGGCACCAAAGCAATCTTGTACAACTCACCAAATCTAGGGTCACGAGTGAGTCTACACAACGGAAAGAACGCAGCATTTGCTGC
5401  CGTTTTACTACTGACTTTTACTGATCTATGGAAGCAAAATACATATCTCAACGCAATCATACTTGTGCTTGTGGTAACAATCATAGCAGTCATTAGTACTTC
5501  CTTAGTGAGGACTGAACCTTGTGTCATCAAGATTACTGGAGAATCAATCACAGTGTTGGCTTGCAAATTAGATGCAGAAACCATCAGAGCCATTGCCGAT
5601  CTCAAGCCACTCTCCGTTGAACGGTTAAGTTTCCATTGATACTCGAAAGATGTCAGCACCAGCTAGCACAACACAGGCCACAGGGTCAACTACCTCAACT
5701  ACCACAAAAACTGCAGGCGCAACTCCTGCCACAGCTTCAGGACTGTTCACCATCCCGGATGGGGATTTCTTTAGTACAGCCCGTGCTGTAGTAGCCAGCG
5801  ATGCCGTTGCGACGAATGAGGACCTCAGCGAGATTGAGGCTGTCTGGAAGGACATGAAGGTGCCCACAGACACTATGGCACAGGCTGCTTGGGACTTAGT
5901  CAGACACTGTGCTGATGTGGGCTCATCTGCTCAAACAGAAATGATAGATACGGGTCCCTACTCCAACGGCATCAGCAGAGCCAGACTGGCAGCAGCAATC
6001  AAAGAGGTGTGCACACTTAGGCAATTTTGCATGAAGTATGCCCCAGTGGTATGGAACTGGATGCTGACTAACAACAGTCCACCTGCTAACTGGCAAGCGC
6101  AAGGTTTCAAGCCTGAGCACAAATTCGCTGCATTCGACTTCTTCAATGGAGTCACCAACCCAGCTGCCATCATGCCCAAAGAGGGGCTCATTCGGCCACC
6201  GTCTGAAGCTGAAATGAAATGCTGCCCAAACTGCTGCCTTTGTGAAGATTACAAAGGCCAGGGCACAATCCAACGACTTTGCCAGCCTAGATGCAGCTGTC
6301  ACTCGAGGTCGTATCACTGGAACAACAACCGCTGAGGCTGTTGTCACTCTACCACCACCATAACTACGTCTACATAACCGACGCCTACCCCAGTTTCATA
6401  GTATTTTCTGGTTTGATTGTATGAATAATATAAAT
```

*To whom correspondence should be addressed

References: 1. Morozov S.Yu. et al. (1983) Dokl.Akad.Nauk SSSR 271, 211-215. 2. Morozov S.Yu. et al. (1987) FEBS Letters 213, 438-442. 3. Quadt R. et al. (1988) Virology 165, 256-261. 4. Gorbalenya A.E. et al. (1988) Nature 333, 22. 5. Hodgman T.C. (1988) Nature 333, 22-23. 6. Miroshnichenko N.A. et al. (1988) FEBS Letters 234, 65-68.