The nucleotide sequence of the herpes simplex virus type 1 late gene ICP18.5 of strain Angelotti

C.W.Knopf

Insititute of Virus Research, German Cancer Research Center, Im Neuenheimer Feld 280, 6900 Heidelberg, FRG

The nucleotide sequence of the gene region encoding the herpes simplex virus type 1 (HSV-1) late infected cell polypeptide 18.5 (ICP18.5) of strain Angelotti which plays an important role in the translocation of HSV glycoproteins to membranes, reveals striking differences to that obtained from strain F (1,2). The presented sequence spans from nucleotide pos. 596 of the glycoprotein B gene (3) to pos. 751 of the major DNA binding protein gene (4). The sequence of strain F shows 22 nucleotide deletions (⊔) (18 within the ICP18.5 coding sequences), 16 substitutions (—) and 3 insertions (⊤). This leads to an 11.6% difference in the predicted amino acid composition. The 2355 bp long open reading frame of the ANG ICP18.5 codes for a predicted M_r 85575 protein. The F protein is shorter by 4 amino acids (deleted residues: 167, 338, 434, 435) and displays reading frame differences at residue pos. 43 to 68, 215 to 239, 256 to 279, 472 to 480 and 555 to 557 as well as substitutions at residues 109, 437, 438 and 458. In contrast to the previous analysis (1) and as shown below, the hydropathic profiles (determined as described (5)) of both, the ANG and Epstein-Barr virus BALF3 protein (6), are quite similar and reveal a major hydrophilic domain in the centre of the molecule. Of interest is a predicted α-helical domain comprising 52 residues between pos. 328 to 379 in the ANG protein which is less prevalent in the BALF3 protein.
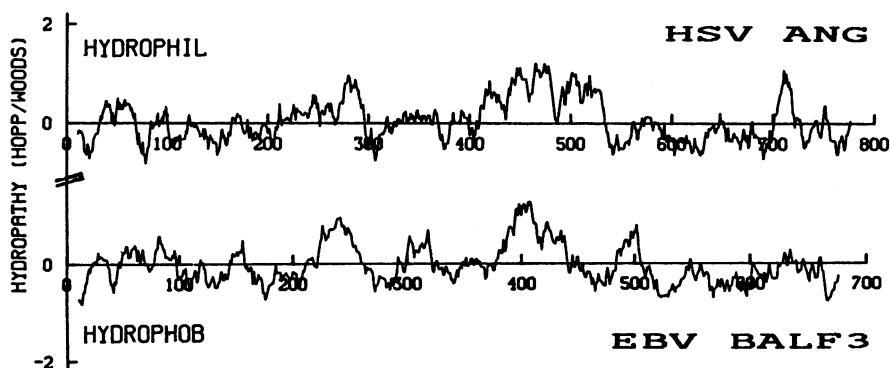
```
   1 CCCATACGACGAACCACCGGCGCCCCCACGAGGGGGCGCCCTGGTGCATGGCGGGACTACGGGGGCCCGTCGTGCCCCCCGTCAGGTAGCCTGGGGGCG
 785                                          *  P  P  G  D  H  G  G  D  P  L  R  P  A

 100 AGGTGCTGGAGGACCGAGTAGAGGATCGAGAAAACGTCTCGGTCGTAGACCACGACCGACCGGGGGCCGATACAGCCGTCGGGGGCGCTCTCGACGATG
 772 L  H  Q  L  V  S  Y  L  I  S  F  V  D  R  D  Y  V  V  V  S  R  P  G  I  C  G  D  P  A  S  E  V  I

 199 GCCACCAGCGGACAGTCGGAGTCGTACGTGAGATATACGCCGGGCGGGTAACGGTAACGACCTTCGGAGGTCGGGCGGCTGCAGTCCGGGCGGCGCAAC
 739 A  V  L  P  C  D  S  D  Y  T  L  Y  V  G  P  P  Y  R  Y  R  G  E  S  T  P  R  S  C  D  P  R  R  L

 298 TCGAGCTCCCCGCACCGGTAGACCGAGGCAAAGAGTGTGGTGGCGATAATCAGCTCGCGAATATATCGCCAGGCGGCGCGCTGAGTGGGCGTTATTCCG
 706 E  L  E  G  C  R  Y  S  A  F  L  T  T  A  I  I  L  E  R  I  Y  R  W  A  A  R  Q  T  P  T  I  G

 397 GAAATGCCGTCAAAACAGTAAAACCTCTGAAATTCGCTGACGGCCCAATCAGCACCCGAGCCCCCGCCCCCATGATGAACCGGGCGAGCTCCTCCTTC
 673 S  I  G  D  F  C  Y  F  R  Q  F  E  S  V  A  W  D  A  G  S  G  G  A  G  M  I  F  R  A  L  E  E  K

 498 AGGTGCGGCAGGAGCCCCACGTTCTCGACGCTGTAATACAGCGCGGTGTTGGGGGGCTGGGCGAAGCTGTGGGTGGAGTGATCAAAGAGGGGCCCGTTG
 840 L  H  P  L  L  G  V  N  E  V  S  Y  Y  L  A  T  N  P  P  Q  A  F  S  H  T  S  H  D  F  L  P  G  N

 595 ACGAGCTCGAAGAAGCGATGGGTGATGCTGGGGAGCAGGGCCGGGTCCACCTGGTGTCGCAGGAGAGACGCTCGCATGAACCGGTGCGCGTCGAACACG
 607 V  L  E  F  F  R  H  T  I  S  P  L  L  A  P  D  V  Q  H  R  L  L  S  A  R  M  F  R  H  A  D  F  V

 694 CCCGGCGCCGAGCGGTTGTCGATGACCGTGCCCGCGCCCGCCGTCAGGGCGCAGAAGCGCGCGCGGCGCGCCGCAAAGCCGTTGGCGACCGCGGCGAACGTC
 574 G  P  A  S  R  N  D  I  V  T  G  A  G  A  T  L  A  C  F  R  A  R  A  A  F  G  N  A  V  A  A  F  T

 793 GCGGGCAGCACCTCGCCGTGGACAGCTGACCCGCAGCATCTTCTCGAGCTCCCCGCGCTGCTCGCGGACGCAGCGCCCCAGGCTGGCCAACGACCGCTTC
 541 A  P  L  V  E  G  H  V  S  V  R  L  M  K  E  L  E  G  R  Q  E  R  V  C  R  G  L  S  A  L  S  R  K

 892 GTCAGGCGGTCCGCGTACAGCCGCCGTCGCTCCCGCACGTCCGCGGCCGCTTGCGTGGCGATGTCCCCCCACGTCTCGGGCCCCTGCCCCCCGGGCCCG
 508 T  L  R  D  A  Y  L  R  R  R  E  R  V  D  A  A  A  Q  T  A  I  D  G  W  T  E  P  G  Q  Q  G  P  G

 991 CGGCGACGGTCTTCGTCCTCGCCCCCGCCCCCGGGAGCTCCCAACCCCCGTGTCCCTTCCTCTACGGCGACACGGTCCCCGTCGTCGTCGGGGCCCGCG
 475 R  R  R  D  E  D  E  G  G  G  G  P  A  G  L  G  R  T  G  E  E  V  A  V  R  D  G  D  D  D  P  G  A

1090 CCGCCCTTGGGCGCGTCCGCCGCGCCCCCCGCCCCCATGCGCGCCAGCACGCGACGCAGCGCCTCCTCGTCGCACTGTTCGGGGCTGACGAGGCGCCGC
 442 G  G  K  P  A  D  A  A  G  G  A  G  M  R  A  L  V  R  R  L  A  E  E  D  C  Q  E  P  S  V  L  R  R

1189 AAGAGCGGCGTCGTCAGGTGGTGGTCGTAGCACGCGCGGATGAGCGCCTCGATCTGATCGTCGGGTGACGTGGCCTGACCGCCGATTATTAGGGCGTCC
 409 L  L  P  T  T  L  H  H  D  Y  C  A  R  I  L  A  E  I  Q  D  D  P  S  T  A  Q  G  G  I  I  L  A  D

1288 ACCATATCCAGCGCCGCCAGGTGGCTCCCGAACGCGCGATCGAAATGCTCCGCCCGCCGCGCCGCGCCCGAACAGCGCCAGTTCCACGGCCACCGCGGTCTCC
 378 V  M  D  L  A  A  L  H  S  G  F  A  R  D  F  H  E  A  R  R  G  F  L  A  L  E  V  A  V  A  A  T  E

1387 TGCTGCAACTCGCGCCCGCGCCCAGCGCGGTCAGGTTGCTGGCAAACGCGTCCATGGTGGTCTGGCCGGCGCGGTCGCCGGACGCGAGCCAGAATCGCAAT
 343 Q  Q  L  E  R  R  A  L  A  T  L  N  S  A  F  A  D  M  T  T  Q  G  A  R  D  G  S  A  L  W  F  R  L

1486 TCGCTGATGGCGTACAGGCCGGGCGTGGTGGCCTGAAACACGTCGTGCGCCTCCAGCAGGGCGTCGGCCTCCTTGCGGACCGAGTCGTTCTCGGGCGAC
 310 E  S  I  A  Y  L  G  P  T  T  A  Q  F  V  D  H  A  E  L  L  A  D  A  E  K  R  V  S  D  N  E  P  S

1585 GGGTGGGGCTGCCCGTCGCCCCCCGCGGTCCGGGCCAGCGCATGGTCCAACACGGAGAGCGCCCGCGCGGGTCGGCGTCCGACAGCCCGGCGGCATGG
 277 P  H  P  Q  G  D  G  A  T  R  A  L  A  H  D  L  V  S  L  A  R  A  R  D  A  D  S  L  G  A  A  H

1684 GGCAGGTACCGCCGCAGCTCGTTGGCGTCCAGCCGCACCTGCGCCTGCTGGGTGACGTGGTTACAGATACGGTCCGCCAGGCGGCGGGCGATCGTCGCC
 244 P  L  Y  R  R  L  E  N  A  D  L  R  V  Q  A  Q  Q  T  V  H  N  C  I  R  D  A  L  R  R  A  I  T  A

1783 CCCTGGTTCGCCGTCACACACAGTTCCTCGAAACAGACCGCGCAGGGGTGGGACGGGTCGCTAAGCTCCGGGGGGACGATAAGGCCCGACCCCACCGCC
 211 G  Q  N  A  T  V  C  L  E  E  F  C  V  A  C  P  H  S  P  D  S  L  E  P  P  V  I  L  G  S  G  V  A

1882 CCCACCATAAACTCCCGAACGCGCTCCAGCGCGGCGGTGGCGCCGCGCGAGGGGGTGATGAGGTGGCAGTAGTTTAGCTGCTTTAGAAAGTTCTCCACG
 178 G  V  M  F  E  R  V  R  E  L  A  A  T  A  G  R  S  P  T  I  L  H  C  Y  N  L  Q  K  L  F  N  E  V

1981 TCGTGCAGGAAACACAGCTCCATATGGACGGTCCCGCCATACGTATCCAGCCTGACCCGTTGGTGATACGGACAGGGTCGGGCCCAGGCCCATGGTCTCG
 145 D  H  L  F  C  L  E  M  H  V  T  G  G  Y  T  D  L  R  V  R  Q  H  Y  P  C  P  R  A  L  G  M  T  E

2080 GTGAAAAACGCCGCGACGTCTCCCGCGGTCGCGAACGTCTCCAGGCTGCCCAGGAGCCGCTCGCCCTCGCGCCACGCGTACTCTAGCAGCAACTCCAGG
 112 T  F  F  A  A  V  D  G  A  T  A  F  T  E  L  S  G  L  L  R  E  G  E  R  W  A  Y  E  L  L  L  E  L

2179 GTGACCGACAGCGGGGTGAGAAAGGCCCCGGCCTGGGCCTCCAGGCCCGGCCTCAGACGACGCCGCAGCGCCCGCACCTGAAGCGCGTTCAGCTTCAGT
  79 T  V  S  L  P  T  L  F  A  G  A  Q  A  E  L  G  P  R  L  R  R  R  L  A  R  V  Q  L  A  N  L  K  L

2278 TGGGGGAGCTTCCCCCGTCCGATGTGGGGGTCGCACCGCCGGAGCAGCTCTATCTGAAACACATAGGTCTGCACCTGCCCGAGCAGGGCTAACACTTT
  46 Q  P  L  K  G  R  G  I  H  P  D  C  R  R  L  L  E  I  Q  F  V  Y  T  Q  V  Q  G  L  L  A  L  L  K

2377 TGACGGGCCACGGTGGGCTCGGACACCGGGGCGGCCATCTCGCGGCGCCGGAGTATGCGGTGGACCGAGGCGGTCCGTACGCT
  13 Q  R  A  V  T  P  E  S  V  P  A  A  M

2476 ACCCGGTGTCTGGCTGAGCCCCGGGGTCCCCCTCTTCGGGGCGGCCTCCCGCGGGCCCGCCGACCGGCAAGCCGGGAGTCGGCGGCGCGTGCGTTCTG

2575 CTCTATTCCCAGACACCGCGGAGAGGAATCACGGCCCGCCCAGAGATATAGACACGGAACACAAACAAGCACGGATGTCGTAGCAATAATTTATTTTAC

2674 ACACATTCCCCGCCCCGCCCTAGGTTCCCCCACCCCCCCAACCCCTCACAGCATATCCAACGTCAGGTCTCCCTTTTTGTCGGGGGGCCCCTCCCCAAAC

2773 GGGTCATCCCCGTGGAACGCCCGTTTGCGGCCGGCAAATGCCGGTCCCGGGGCCCCGGGCCGCCGAACGGCGTCGCGTTGTCGTCCTCGCAGCCAAAA

2872 TCCCCAAAGTTAAACACCTCCCCGGCGTTGCCGAGTTGGCTGACTAGGGCCTCGGCCTCGTGCGCCACCTCCAGGGCCGCGTCCGTCGACCACTCGCCG

2971 TTGCCGCGCTCCAGGGCACGCGCGGTCAGCTCCATCATCTCCTCGCTTAGGTACTCGTCCTCCAGGAGCGCCAGCCAGTCCTCGATCTGCAGCTGCTGG

3070 GTGCGGGGCCCCAGGCTTTTCACGGTCGCCACGAACACGCTACTGGCGACGGCCGCCCCGCCCTCGGAGATAATGCCCCGGAGCTGCTCGCACAGCGAG

3169 CTTTCGTGCGCTCCGCCGCCGAGGCTCGAGGCCGCGCACACAAACCCGGCCCGGGGACAGGCCAGGACGAACTTGCGGGTGCGGTCAAAAATAAGGAGC
```

REFERENCES: (1) Pellett,P.E. et al.(1986) J.Virol.60,1134. (2) Pellett, P.E. et al.(1985) J.Virol. 56,807. (3) Bzik,D.J. et al. (1984) Virol.133,301. (4) Quinn,J.P.& McGeoch,D.J.(1985) NAR 13, 8143. (5) Knopf,C.W.(1987) J.gen.Virol.68,1429. (6) Baer,R.et al.(1984) Nature 310,207.