JOURNAL OF VIROLOGY, Mar. 1985, p. 879–884
0022-538X/85/030879-06$02.00/0
Copyright © 1985, American Society for Microbiology
Vol. 53, No. 3

Nucleotide Sequence of Avian Sarcoma Virus UR2 and Comparison of Its Transforming Gene with Other Members of the Tyrosine Protein Kinase Oncogene Family

WENDI S. NECKAMEYER AND LU-HAI WANG*
The Rockefeller University, New York, New York 10021

Received 25 October 1984/Accepted 10 December 1984

The genome of avian sarcoma virus UR2 was completely sequenced and found to have a size of 3,165 nucleotides. The UR2-specific transforming sequence, *ros*, with a length of 1,273 nucleotides, is inserted between the truncated *gag* gene coding for p19 and the *env* gene coding for gp37 of the UR2AV helper virus. The deduced amino acid sequence for the UR2 transforming protein P68 gives a molecular weight of 61,113 and shows that it is closely related to the oncogene family coding for tyrosine protein kinases. P68 contains two distinctive hydrophobic regions that are absent in other tyrosine kinases, and it has unique amino acid changes and insertions within the conserved domain of the kinases. These characteristics may modulate the activity and target specificity of P68.

Avian sarcoma virus (ASV) UR2 is a replication-defective virus which was isolated together with its associated helper virus, UR2AV, from a spontaneous chicken tumor (2). The genome of UR2 contains a stretch of 1.2 kilobases of transformation-specific sequence, called *ros* (31), which has a homologous counterpart, c-*ros*, in normal chicken cellular DNA (25). UR2 was presumably generated by recombination between UR2AV and c-*ros* at the expense of certain replicative sequences in UR2AV. As a result, *ros* was fused to the 5′ region of the UR2AV sequence which coded for part of the viral structural protein p19 (31). The fused p19 and *ros* sequences in UR2 code for a 68,000-molecular-weight polyprotein called P68, which was found to be associated with a tyrosine-specific protein kinase activity (8).

UR2 transforms chicken embryo fibroblasts (CEF) and efficiently induces sarcomas in chickens (2). The UR2 transforming protein P68 has biochemical properties similar to those of the protein kinases encoded by several other oncogenic viruses, and this functional conservation implies that the *ros* gene must be closely related to the tyrosine protein kinase-encoding oncogene family. Yet at the level of nucleic acid hybridization, no significant homology was found among *ros* and oncogenes present in other ASVs with the exception of the 3′ region of the transforming sequence of Fujinami sarcoma virus (FSV) (18).

Cells transformed by other ASVs or by Abelson murine leukemia virus contain elevated levels of phosphotyrosine compared with uninfected cells. However, UR2-transformed CEF showed no significant increase in phosphotyrosine relative to control CEF (5). Additionally, it has been shown that P68 differs from the protein kinases encoded by FSV, ASV Y73, and Rous sarcoma virus (RSV) in many enzymatic properties, including pH optimum, cation preference, and specificity of phosphate donors (8).

Perhaps the most striking difference between UR2 and other ASVs is the distinctive, extremely elongated morphology of UR2-transformed CEF (2). Recently, two studies showed that UR2-transformed CEF still maintained a significantly higher level of organized cytoskeleton than Schmidt-Ruppin (SR) RSV-transformed cells (1, 19) and displayed increased surface fibronectin in comparison with normal cells (19). These phenomena indicate that the *ros* gene product P68 interacts in a distinct manner with its cellular targets despite the similarity of its biochemical properties to those of kinases encoded by other oncogenes.

To elucidate the basis for functional conservation of *ros* as well as the differences between *ros* and other oncogenes, we have sequenced the entire genome of UR2 and compared the predicted amino acid sequence of P68 with other members of the tyrosine protein kinase family.

MATERIALS AND METHODS

Molecular cloning. pUR2, a biologically active UR2 DNA pBR322 plasmid clone described recently (18), was used for the nucleotide sequencing. The following DNA fragments from pUR2 were subcloned into the *Eco*RI site of M13mp8 (16) with *Eco*RI linkers (Fig. 1): *Eco*RI-*Pvu*II, 840 base pairs (bp); *Eco*RI-*Hin*dIII (3′ site), 1,515 bp; *Hin*dIII (3′ site)-*Sst*I, 1,450 bp; *Sst*I-*Eco*RI, 1,000 bp; *Eco*RI-*Nci*I, 510 bp; *Nci*I-*Pvu*II, 330 bp; *Pvu*II-*Hin*dIII (5′ site), 605 bp; and *Hae*II-*Eco*RI, 470 bp.

DNA sequencing. The DNA sequence was determined by the methods of Maxam and Gilbert (15) and Sanger (22). M13mp8 recombinant clones of both polarities were sequenced at least twice by the dideoxy method.

Hydrophilicity analysis. The hydrophilicity profile of the 400-amino acid *ros* region of $P68^{gag\text{-}ros}$ was determined by the computer program of Hopp and Woods (11) with the hydrophilicity values for each amino acid determined by Levitt (13).

RESULTS AND DISCUSSION

Nucleotide sequence of the UR2 genome. The sequencing strategy of pUR2 is shown in Fig. 1. Since pUR2 was derived from a UR2 circular DNA molecule cut with *Eco*RI, the UR2 genome in this clone is permuted with respect to the *Eco*RI site in *ros* (Fig. 1). However, we present the complete UR2 genomic sequence (Fig. 2, lower-case letters) as colinear to the viral RNA genome which begins at the predicted cap site (nucleotide 1) and ends at the polyadenylation site (nucleotide 3,165). The *gag-ros* fusion protein P68 begins

* Corresponding author.

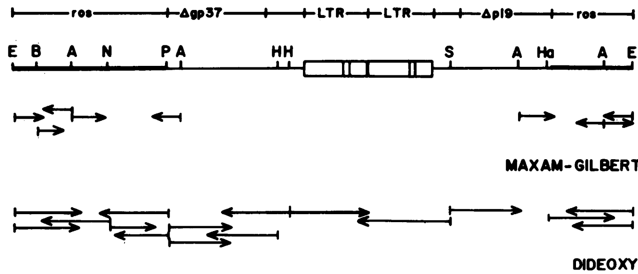

FIG. 1. Restriction enzyme cleavage map and nucleotide sequencing strategy of the UR2 genome. The transforming sequence *ros* of the permuted UR2 molecular clone pUR2 (18) is defined by the heavy line. The arrows indicate the direction and approximate extent of sequence determined by either the Maxam-Gilbert (15) or Sanger dideoxy (22) method. Abbreviations for restriction enzymes: A, *Ava*I; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; Ha, *Hae*II; N, *Nci*I; P, *Pvu*II; S, *Sst*I.

with the initiator amino acid methionine of p19 at nucleotides 380 through 382. The *ros*-specific sequence is fused to *gag* at nucleotide position 831. The coding sequence of the transforming protein P68 ends with an ochre stop codon at nucleotides 2036 through 2038. The v-*ros* noncoding region continues down to nucleotide 2103, and the Δgp37 envelope sequence resumes afterwards. The UR2 long terminal repeat is 326 bp long.

Previous restriction enzyme analysis of UR2 DNA showed that UR2 shares 0.8 kilobases of 5′ and 1.4 kilobases of 3′ sequences with its helper virus UR2AV (18). Since we have not sequenced UR2AV, we compared the UR2 sequence with those of Prague C RSV (23) in the regions of *gag* and *env* to determine the sites of recombination between UR2AV and cellular *ros* in the generation of UR2. The UR2 sequence diverges from that of PR-C RSV downstream from nucleotide 830 in the 5′ region of *ros* and converges after nucleotide 2103 in the 3′ *ros* region (Fig. 3). We conclude that the 5′ and 3′ recombination sites between UR2AV and c-*ros* must be at or very close to nucleotides 831 and 2103 in the p19 and gp37 regions, respectively. The total length of *ros* is 1,273 nucleotides.

Extensive homology was found between the UR2 sequences outside *ros* and the corresponding sequences of other avian viruses. Within the 5′ 380-bp noncoding sequence of UR2, there are 22, 30, and 19 scattered single-base changes relative to RAV-2 (4), Y73 (12), and PR-C (23), respectively. These changes include one deletion at nucleotide 371. This change (A in UR2 replaces TG in PR-C) occurs immediately upstream from the initiation site for translation. It has been suggested (4) that this upstream region is important for efficient mRNA translation; we do not know whether this change in any way affects the translation of UR2 genomic RNA. Sequences within the binding site for the tRNA primer are completely identical to other avian viruses.

The p19 amino acid sequence of UR2 differs from those of PR-C and Y73 by three and two amino acids, respectively. It is not clear whether these amino acid changes affect any of the p19 antigenic determinants; however, P68 could still be efficiently precipitated by the anti-*gag* serum apparently by the remaining p19 peptide. The gp37 sequence of UR2 is well conserved relative to PR-C, Y73, and RAV-2 except for the carboxy-terminal region. The termination codons of gp37 in UR2 (at nucleotides 2627 through 2629), Y73, and RAV-2 occur 10 bp further downstream than that in PR-C, and there is considerable divergence in this region, especially in the stretch spanning the 13 carboxy-terminal amino acids of gp37 in UR2.

In the U3 region of UR2, 18 to 31 base changes were found when compared with those of other avian viruses including RAV-2, Y73, PR-C RSV, and SR-A RSV (28). In addition, there are stretches of deletions and insertions. In particular, a stretch of 17 bp in the middle of U3 (nucleotides 3023 to 3040 in UR2) is marked by variable deletions in Y73, PR-RSV, and SR-RSV but not in RAV-2 and UR2. These changes in UR2 apparently have no pronounced effect on the efficiency of transcription, since UR2 RNA is produced in comparable amounts in UR2-infected cells relative to the amounts of viral RNAs produced in other ASV-transformed cells (18, 31; unpublished data). The universal signal sequences for transcription and processing of mRNAs are conserved in all viruses. Sequences in the inverted repeat regions found at both ends of the long terminal repeat and in the polypurine tract are also conserved in UR2. In addition, a region corresponding to the sequence suggested to be involved in RNA packaging is also present in UR2 (27).

An additional *Hin*dIII site 67 bp upstream from the *Hin*dIII site 3′ to *ros* detected previously by restriction enzyme mapping was discovered after the entire 3′ region of the UR2 genome had been sequenced (Fig. 1).

Structural domains and reading frames in the UR2 genome. The location of open reading frames in the region spanning nucleotides 380 through 3000 of UR2 is shown in Fig. 4. A long open reading frame in frame 2, beginning with an AUG codon at nucleotide 380 and ending in an ochre stop codon at nucleotide 2036, encodes the UR2 transforming protein P68. The other open frame in frame 2 represents the Δgp37 sequence. The next largest sequence with an open reading frame is the 111-amino acid stretch in frame 1; all other open frames are less than 40 amino acids in length. All the deduced amino acid sequences of these reading frames except that for gp37 begin with an initiator methionine. It is unlikely that those open reading frames other than the one coding for P68 are translated because subgenomic mRNAs have not been detected in UR2-transformed cells (31).

The size of the *ros* coding region is 1,205 bases. A noncoding sequence of 65 bp follows the ochre stop codon of *ros* and contains four termination codons, three of which are immediately adjacent to each other. The termination codon and the 3′ noncoding *ros* sequence are most likely derived from the chicken c-*ros* sequence, since uninterrupted homology between UR2 *ros* and chicken c-*ros* DNA in this region was found by heteroduplex mapping (W. S. Neckameyer, M. Shibuya, M.-T. Hsu, and L.-H. Wang, unpublished data).

P68 contains 552 amino acids, of which 150 are encoded by *gag* and 402 are encoded by *ros*. The transforming protein has an apparent molecular weight of 68,000 as estimated from sodium dodecyl sulfate-polyacrylamide gel electrophoresis (8), but the predicted molecular weight of the deduced amino acid sequence is 61,113. It is unlikely that this discrepancy is due to glycosylation of the protein, because in vitro translation of the genomic 24S UR2 RNA yielded a protein product of 68,000 (8). Unusual amino acid composition such as a high percentage of cysteine or proline could cause extensive folding and change the relative mobility of the protein in sodium dodecyl sulfate-polyacrylamide gels (9, 30). The amino acid composition of $P68^{gag\text{-}ros}$ shows that proline and cysteine comprise only 5.2 and 2.35% of the protein, respectively, so this probably does not account for the 6,800-higher apparent molecular weight.

5'
gccatttaccactcaccacattggtgtgcacctgggttgatggccggaccgtcgattccctgacgactacgaacaccta 80
R → U5

aatgaagcggaaggcttcatttggtgaccccgacgtgatagttagggaatagtggtcggccacagacggcgtggcgatcc 160
U5 ← → PBS ←

tgtcctcatccgtctcgcttgttcggggagcgggcgatgaccctagtagaggggcttcggcttaggagggcgaaaagct 240

gagtggcgtcggagggagctctactgcagggagccaacataccctaccgagcactcagagagtcgttggaagacgggaag 320

gaagcccgacgactgagcggtccaccccagacgtgattctggtcgcccggagatcaagcatggaagccgtcataaaggtg 400
M E A V I K V
→Δp19

atttcgtccgcgtgtaaaacctattgcggaaaaacctctccttctaagaaggaaatcggggccatgttgtccctcttaca 480
I S S A C K T Y C G K T S P S K K E I G A M L S L L Q

aaaagaagggttgcttatgtctccctcagacttatattccccggggtcctgggatcccatcactgcggcgctctcccagc 560
K E G L L M S P S D L Y S P G S W D P I T A A L S Q

ggctaatggtacttgggaaatcgggagagttaaaaacctggggattggtttgggggcattgaaagcggcccgagaggaa 640
R L M V L G K S G E L K T W G L V L G A L K A A R E E

caggttacatctgagcaagcaaagttttggttgggattaggggagggagggtctctcccccaggtccggaatgcatcga 720
Q V T S E Q A K F W L G L G G G R V S P P G P E C I E

gaaaccagcaacggagcggcgaatcgacaaaggggaggaagtgggagaaacaactgtgcagcgagatgcgaagatggcgc 800
K P A T E R R I D K G E E V G E T T V Q R D A K M A

cggaggaaacggccacacctaaaaccgttgatactgtgacctctccagatatcactgctattgttgctgtgattggagca 880
P E E T A T P K T V D T V T S P D I T A I V A V I G A
Δp19 ← → ros

gttgtactgggcttgactagcttgactataatcatactgtttggttttgtatggcaccaaagatggaaatccagaaaacc 960
V V L G L T S L T I I I L F G F V W H Q R W K S R K P

agcctcaactgggcagattgtgcttgtcaaggaagataaagaattagctcaacttaggggaatggctgagacagtgggat 1040
A S T G Q I V L V K E D K E L A Q L R G M A E T V G

tagccaatgcttgttatgctgtcagcactcttccttctcaagcagagattgagtcattgccagcttttcctcgggacaaa 1120
L A N A C Y A V S T L P S Q A E I E S L P A F P R D K

ctgaacttacacaagttgttaggaagtggagcatttggagaggtgtatgaagggactgcattagatatcctggcagatgg 1200
L N L H K L L G S G A F G E V Y E G T A L D I L A D G

aagtggagaatccagagtagcagtcaagactttgaagagaggtgcaacagaccaagagaagagtgaattcttgaaggagg 1280
S G E S R V A V K T L K R G A T D Q E K S E F L K E

cacacttaatgagtaaatttgatcatccccacattctgaagctacttggagtgtgtctgttaaatgaacctcagtacctt 1360
A H L M S K F D H P H I L K L L G V C L L N E P Q Y L

atactggagctgatggaaggaggagatctgcttagctatttacgaggagccagaaagcaaaagttccagagtcccttact 1440
I L E tL M E G G D L L S Y L R G A R K Q K F Q S P L L

gacattgactgatctcttggatatatgcttggatatttgcaaaggttgtgtctatttagagaaaatgcgtttcatacaca 1520
T L T D L L D I C L D I C K G C V Y L E K M R F I H

gggacctggctgctcgcaactgccttgtgtctgagaagcaatatgggagctgctcccgagtggtaaagattggtgatttt 1600
R D L A A R N C L V S E K Q Y G S C S R V V K I G D F

ggacttgccagagatatctataaaaatgattactacaggaaaagaggagaaggcctactccctgtcagatggatggctcc 1680
G L A R D I Y K N D Y Y R K R G E G L L P V R W M A P

tgaaagcctcattgatggcgtctttacaaatcactctgatgtttgggcttttggagtcttagtgtgggaaacattaactt 1760
E S L I D G V F T N H S D V W A F G V L V W E T L T

tgggtcaacagccatatccgggtctctccaatatagaagttttacaccatgtacgatcaggaggaaggctggaatctccg 1840
L G Q Q P Y P G L S N I E V L H H V R S G G R L E S P

aataactgtcctgatgacatacgtgatttaatgacaagatgctgggcccaagatcctcacaacagacctactttcttta 1920
N N C P D D I R D L M T R C W A Q D P H N R P T F F Y

tattcagcacaaactgcaagagataaggcactctccactgtgcttcagctacttccttggagacaaagagtcagtggctc 2000
I Q H K L Q E I R H S P L C F S Y F L G D K E S V A

cactgcggattcagacagcattctttcaaccactctaatggaagcaagggatcaagaaggcctgaattatttagtagtag 2080
P L R I Q T A F F Q P L .

tcaaagaaagcaaccaggatcaaggtagcagctgcgcaagccttaagagaaatcgagagactagcctgctggtccgttaa 2160
V A A A Q A L R E I E R L A C W S V K
ros ← → Δgp37

acaggctaacttgacaacatcactcctcggggacttattggatgatgtcacgagtattcgacacgcggtcctacaaaacc 2240
Q A N L T T S L L G D L L D D V T S I R H A V L Q N

gagcggctattgacttcttgcttctagctcacggtcatggctgtgaggacattgccggaatgtgttgtttcaatctgagt 2320
R A A I D F L L L A H G H G C E D I A G M C C F N L S

gatcacagtgagtctatacagaaaaagttccaactaatgaaggaacatgtcaataagatcggcgtggacagcgacccaat 2400
D H S E S I Q K K F Q L M K E H V N K I G V D S D P I

cggaagttggctgcgaggactattcggaggaataggagaatggccgttcatttactgaaaggactgcttttggggcttg 2480
G S W L R G L F G G I G E W A V H L L K G L L L G L

tagttattttgctgctagtagtgtgcttgccttgcctttgcaatttgtatccagtagcatccgaaagatgattgataat 2560
V V I L L L V V C L P C L L Q F V S S S I R K M I D N

tcactcggctatcgcgaggagcgtaagaaatttcaggaggcttataagcaacccgaaagagtagtttgagcgggttcttg 2640
S L G Y R E E R K K F Q E A Y K Q P E R V V .

tattccgtgtgataactggttggattgataatcgatcggctggcatacggagtataggaggtcgctgagtggtaagcttg 2720

cagacttggctgaagcatagagtatctcctatagcttcgataactgctagagaataagctaagcttgcgaatcgggctgt 2800

aacggggaaggcttgactaggggaccatggtatgtataggcggaaggcggggcttcggttgtacgcggttaggagtcccc 2880

tcaggatatagtagttgcgcttctgcatagggagggggaaatgtagtcttatgcaatactcttatgtaacgatgaaacag 2960
pp tract → U3

caaaatgccttataaggggaaaagaggcatgtacatgttgattggtggaagtaggtggtatgatcgtggtacgatcgtgc 3040

cttattaggaaggcaacagacgggtctaacagggattggataaaccccttagttccgcattgcaagagataatatattta 3120
TATA

agtacctggcttgatgtaataaacgccatttaccactcaccaca 3165 3'
PA R

FIG. 2. Complete nucleotide sequence of the UR2 genome and deduced amino acid sequence of P68 and Δgp37. The nucleotide sequence from the predicted cap site (nucleotide 1) to the polyadenylation site (nucleotide 3165) in UR2 DNA is shown by the lower-case letters. Each amino acid is represented by a capital letter shown under the second nucleotide of the triplet codon. The R, U5, and U3 regions of the long terminal repeat, the primer binding site (PBS), the polypurine tract (pp tract), the polyadenylation signal (PA), and the TATA box are indicated by heavy lines underneath these sequences. Stop codons are boxed and indicated by a period at the end of the amino acid sequences. The dashed line under amino acids 8 through 36 of the *ros* region indicates the hydrophobic region of the *ros* polypeptide presumed to be important in membrane association. Nucleotide numbers are given at the end of each row of sequence, and every 10 nucleotides are marked by a period over the 10th nucleotide. The one-letter symbols for the amino acids are used: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

The hydrophilicity profile of $P68^{gag\text{-}ros}$ (Fig. 5) revealed one 5′ and one 3′ hydrophobic region (shaded area). The 5′ hydrophobic region is 29 amino acids long (underlined in Fig. 2), is characterized by a stretch of 13 consecutive hydrophobic amino acids, and is flanked by an acidic residue (aspartic acid) immediately upstream and a basic residue (arginine) two amino acids downsteam of this sequence. The 3′ hydrophobic region (Fig. 6, amino acids 448 through 456) contains nine hydrophobic amino acids flanked by an aspartic acid and a glutamic acid at N and C termini, respectively. With the exceptions of the *ros* and *abl* (21) proteins, this hydrophobic region is interrupted by serine in all other tyrosine kinase oncogene products (Fig. 6). P68 in UR2-transformed CEF remained associated with the plasma membrane during cellular fractionation even under conditions of high salt (300 mM) (E. Garber et al., personal communica-

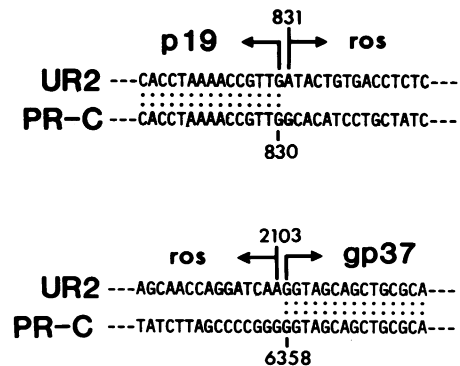

FIG. 3. Junctions of *ros*- and UR2AV-derived sequences. The numbers refer to the nucleotide positions of UR2 sequence shown in Fig. 2 and to the published PR-C RSV sequence (23).

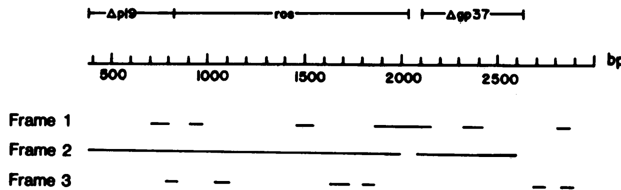

FIG. 4. Location of sequences with open reading frames within the UR2 genome. Considering the predicted cap site as nucleotide 1 of the sequence presented in Fig. 2, the genome was translated in all three frames. Only the region spanning nucleotides 380 through 3000 is shown here.

tion). The 5′ hydrophobic region is long enough to span the membrane, and we speculate that this region may be responsible for the association of P68 with the plasma membrane.

Comparison of P68 *gag-ros* with other protein kinases. The conserved region of the deduced amino acid sequence of P68 was compared with the analogous domains of the proteins encoded by the viral oncogenes *src* (6, 7, 29), *yes* (12), *fps* (24), *fms* (10), *erb* B (32), *fgr* (17), and *abl* (21) as well as the catalytic subunit of the cyclic AMP-dependent bovine protein kinase (3, 26) (Fig. 6). Certain highly conserved regions are immediately apparent. The lysine residue which is the proposed ATP-binding site of BPK is conserved in *ros* (amino acid 283), and so is the characteristic sequence Leu-Gly-X-Gly-X-Phe-Gly-X-Val 15 to 20 residues upstream from the binding site (3). This region is highly conserved in all the kinases under comparison and is likely to be an important structural domain. Two other highly conserved regions include a 26-residue stretch 28 amino acids upstream from the putative phosphotyrosine acceptor site (Fig. 6, amino acid 419, marked with an asterisk), and the 28-amino acid stretch 6 amino acids downstream from this site (Fig. 6, rows 5 through 7). Among the kinase-related oncogenes, *erbB*, *fgr*, and *raf* (not listed in Fig. 6) lack demonstrable kinase activity (14, 17, 32) although they share regions of conserved amino acid sequences. It is interesting to note that the circled amino acids of *ros*, which denote amino acids different from the rest of the tyrosine protein kinases, occur

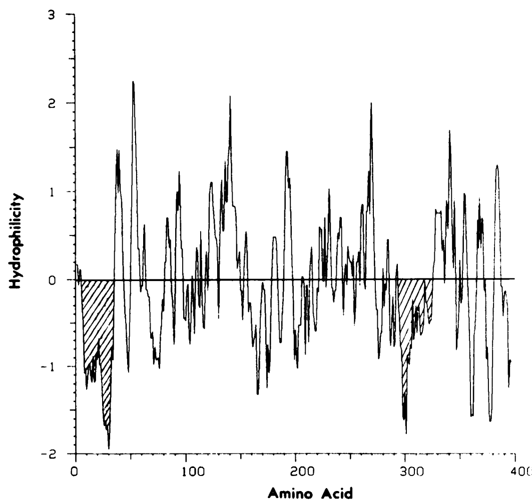

FIG. 5. Hydrophilicity analysis of the deduced amino acid sequence of *ros*. Positive values on the *y* axis denote hydrophilic stretches of amino acids, and negative values denote hydrophobicity. The two cross-hatched areas indicate the extensively hydrophobic regions of *ros*.

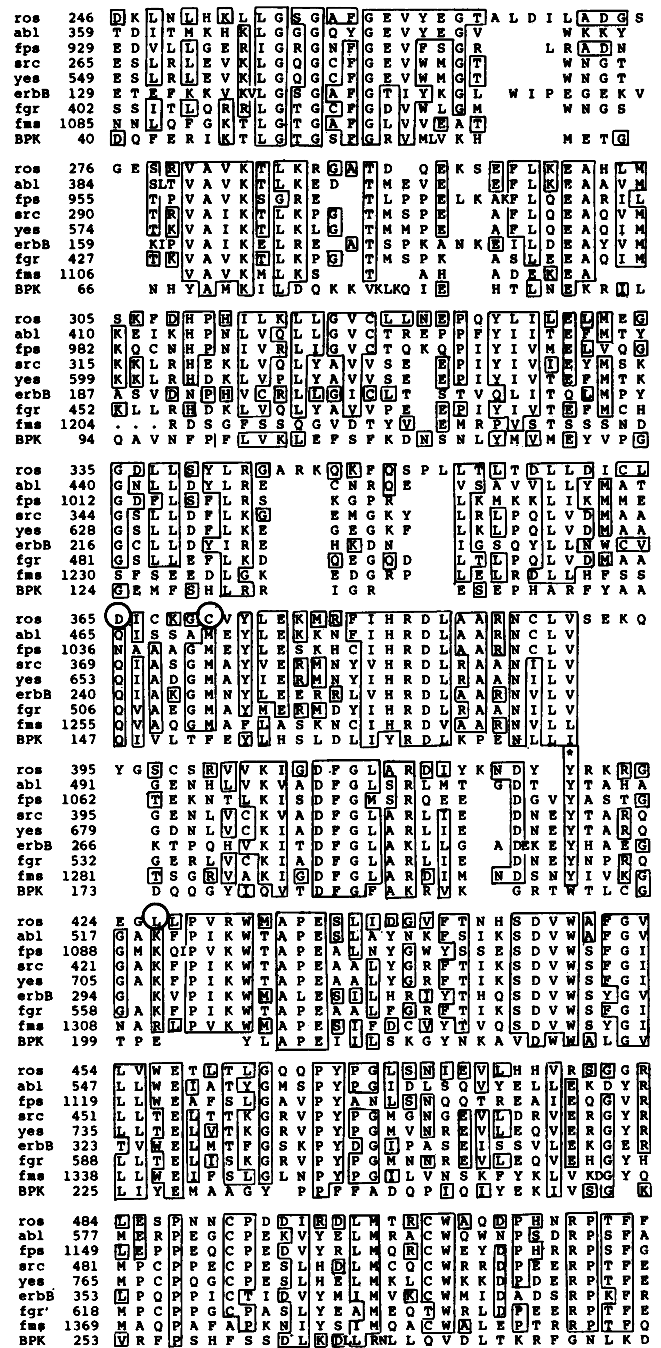

FIG. 6. Comparison of the amino acid sequences within the conserved region of protein kinases encoded by *ros* and other oncogenes and the catalytic subunit of the bovine protein kinase gene. The 3′ amino acids 246 through 513 of P68*gag-ros* have been aligned to show homology with other protein kinases from the following sources: *abl*, Abelson murine leukemia virus (21); *fps*, Fujinami sarcoma virus (24); *src* SR-A Rous sarcoma virus (6, 7, 29); *yes*, Y73 sarcoma virus (12); *erbB*, avian erythroblastosis virus AEV-H (32); *fgr*, Gardner-Rasheed feline sarcoma virus (17); *fms*, McDonough strain of feline sarcoma virus (10); and BPK, catalytic subunit of the cyclic AMP-dependent protein kinase from bovine cardiac muscle (26). Amino acid numbers are given preceding each row of sequence. The putative phosphotyrosine acceptor site of *ros* (amino acid 419) is marked with an asterisk. Regions of conserved amino acids among the proteins are boxed. Amino acids of *ros* at positions 365, 370, and 426 are circled to show the unique differences from the rest of the tyrosine protein kinases.

at the N-terminal regions of both of the two highly conserved sequences. Furthermore, *ros* contains a unique six-amino acid insertion (amino acids 391 through 396) within the highly conserved region, in addition to two other insertions at positions 344 through 346 and 352 through 353.

We have aligned the tyrosine residue at amino acid 419 of *ros* with the known tyrosine phosphoacceptors of the other protein kinases. The region surrounding the tyrosine phosphoacceptor site appears to be somewhat conserved, although its essentiality is not clear, in several tyrosine protein kinases (Y73 P90, PR-CII P105, $p60^{src}$, and ST-FeSV P85) in that the phosphotyrosine is located seven residues to the carboxy-terminal side of a basic amino acid (arginine or lysine) and either four residues to the carboxy-terminal side of or adjacent to a glutamic acid residue (20). However, the likely phosphoacceptor tyrosine residue at position 419 of *ros* is four and two residues to the carboxy-terminal side of a lysine and an aspartic acid, respectively. No tyrosine-containing peptides in the deduced amino acid sequence of P68 concur with the above proposed consensus sequence. This implies that there is some degree of flexibility with respect to primary amino acid sequence at the phosphotyrosine acceptor site. Although $P68^{gag\text{-}ros}$ is phosphorylated on serine residues in vivo (8), we cannot speculate which residues in particular are possible phosphorylation sites.

The N-terminal 95 amino acids of *ros* upstream from the region compared in Fig. 6 vary greatly from the corresponding domains of other protein kinases and are characterized by a long stretch of hydrophobic sequence mentioned above.

Previous hybridization of *ros* DNA to *fps*, *src*, *yes*, and *abl* DNAs detected significant homology only to *fps* (18). However, there is greater amino acid homology among *ros*, *abl*, and *src* than between *ros* and *fps*. This is due to degeneracy of the triplet code. The longest stretch of nucleotide homology between *ros* and *fps* has been localized to the region upstream of the tyrosine phosphoacceptor site, where the sequences are identical for 40 nucleotides, with the exception of a single base change at the 18th position within this sequence.

It is evident from the deduced amino acid sequence that P68 is a member of the tyrosine protein kinase family. Aside from the unique amino acid changes and insertions within the conserved domain of the protein kinases compared here, P68 contains two distinctive highly hydrophobic regions, particularly in the 5′ region of *ros*. Any combination of these changes could be responsible for modulating the P68 activity and for the specific interactions between P68 and its cellular targets that lead to the unique elongated transformed CEF morphology.

ACKNOWLEDGMENTS

We thank D. Foster, E. Garber, S.-M. Jong, and R. Jove for reading the paper and, particularly, H. Hanafusa for help in comparing amino acid sequences of protein kinases and for comments on the paper. We also thank P. Model for help with the hydrophilicity analysis.

This work was supported by Public Health Service grant CA29339 to L.-H.W. from the National Cancer Institute. W.S.N. was supported by Public Health Service training grant T32A107233. L. H. W. is a recipient of Public Health Service research career development award CA00574 from the National Cancer Institute.

LITERATURE CITED

1. **Antler, A., M. E. Greenberg, G. M. Edelman, and H. Hanafusa.** 1985. Increased phosphorylation of tyrosine in vinculin does not occur upon transformation by some avian sarcoma viruses. Mol. Cell. Biol. **5**:263–267.
2. **Balduzzi, P. C., M. F. D. Notter, H. R. Morgan, and M. Shibuya.** 1981. Some biological properties of two new avian sarcoma viruses. J. Virol. **40**:268–275.
3. **Barker, W. C., and M. O. Dayhoff.** 1982. Viral *src* gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. **79**:2836–2839.
4. **Bizub, D., R. A. Katz, and A. M. Skalka.** 1984. Nucleotide sequence of noncoding regions in Rous-associated virus-2: comparisons delineate conserved regions important in replication and oncogenesis. J. Virol. **49**:557–565.
5. **Cooper, J. A., and T. Hunter.** 1983. Regulation of cell growth and transformation by tyrosine-specific protein kinases: the search for important cellular proteins. Curr. Top. Microbiol. Immunol. **107**:125–161.
6. **Czernilofsky, A. P., A. Levinson, H. Varmus, J. Bishop, E. Tischer, and H. Goodman.** 1980. Nucleotide sequence of an avian sarcoma virus oncogene (*src*) and proposed amino acid sequence for gene product. Nature (London) **287**:198–203.
7. **Czernilofsky, A. P., A. Levinson, H. Varmus, J. Bishop, E. Tischer, and H. Goodman.** 1983. Corrections to the nucleotide sequence of the *src* gene of Rous sarcoma virus. Nature (London) **301**:736.
8. **Feldman, R. A., L.-H. Wang, H. Hanafusa, and P. C. Balduzzi.** 1982. Avian sarcoma virus UR2 encodes a transforming protein which is associated with a unique protein kinase activity. J. Virol. **42**:228–236.
9. **Gingeras, T., D. Sciaky, R. Gelinas, J. Bing-Dong, C. Yen, M. Kelly, P. Bullock, B. Parsons, K. O'Neill, and R. Roberts.** 1982. Nucleotides sequences from the adenovirus-2 genome. J. Biol. Chem. **257**:13475–13491.
10. **Hampe, A., M. Gobet, C. J. Sherr, and F. Galibert.** 1984. Nucleotide sequence of the feline retroviral oncogene v-*fms* shows unexpected homology with oncogenes encoding tyrosine-specific protein kinases. Proc. Natl. Acad. Sci. U.S.A. **81**:85–89.
11. **Hopp, T. P., and K. R. Woods.** 1982. Prediction of antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
12. **Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and M. Yoshida.** 1982. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. Nature (London) **297**:205–208.
13. **Levitt, M.** 1976. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. **104**:59–107.
14. **Mark, G. E., and U. R. Rapp.** 1984. Primary structure of v-*raf*: relatedness to the *src* family of oncogenes. Science **224**:224–226.
15. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
16. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digested restriction fragments. Gene **19**:269–276.
17. **Naharro, G., K. C. Robbins, and E. P. Reddy.** 1984. Gene product of v-*fgr onc*: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase. Science **223**:63–66.
18. **Neckameyer, W. S., and L.-H. Wang.** 1984. Molecular cloning and characterization of avian sarcoma virus UR2 and comparison of its transforming sequence with those of other avian sarcoma viruses. J. Virol. **50**:914–921.
19. **Notter, M. F. D., and P. C. Balduzzi.** 1984. Cytoskeletal changes induced by two avian sarcoma viruses: UR2 and Rous sarcoma virus. Virology **136**:56–68.
20. **Patschinsky, T., T. Hunter, F. Esch, J. Cooper, and B. Sefton.** 1982. Analysis of the sequence of amino acids surrounding sites of tyrosine phosphorylation. Proc. Natl. Acad. Sci. U.S.A. **79**:973–977.
21. **Reddy, E. P., M. J. Smith, and A. Srinivasan.** 1983. Nucleotide sequence of Abelson murine leukemia virus genome: structural similarity of its transforming gene to other *onc* gene products with tyrosine-specific kinase activity. Proc. Natl. Acad. Sci. U.S.A. **80**:3623–3627.
22. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
23. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide

sequence of Rous sarcoma virus. Cell **32**:853–869.

24. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other avian sarcoma viruses. Cell **30**:787–795.
25. **Shibuya, M., H. Hanafusa, and P. C. Balduzzi.** 1982. Cellular sequences related to three new *onc* genes of avian sarcoma virus (*fps*, *yes*, and *ros*) and their expression in normal and transformed cells. J. Virol. **42**:143–152.
26. **Shoji, S., D. Parmelee, R. Wade, S. Kumar, L. Ericsson, K. Walsh, H. Neurath, G. Long, J. Demaille, E. Fischer, and K. Titani.** 1981. Complete amino acid sequence of the catalytic subunit of bovine cardiac muscle cyclic AMP-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. **78**:848–851.
27. **Sorge, J., W. Ricci, and S. H. Hughes.** 1983. *cis*-Acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. J. Virol. **48**:667–675.
28. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U.S.A. **78**:124–128.
29. **Takeya, T., R. A. Feldman, and H. Hanafusa.** 1982. DNA sequence of the viral and cellular *src* gene of chickens. I. Complete nucleotide sequence of an *Eco*RI fragment of recovered avian sarcoma virus which codes for gp37 and pp60^{src}. J. Virol. **44**:1–11.
30. **Verma, I.** 1984. From c-*fos* to v-*fos*. Nature (London) **303**:317.
31. **Wang, L.-H., H. Hanafusa, M. F. D. Notter, and P. C. Balduzzi.** 1982. Genetic structure and transforming sequence of avian sarcoma virus UR2. J. Virol. **41**:833–841.
32. **Yamamoto, T., T. Nishida, N. Miyajima, S. Kawai, T. Ooi, and K. Toyoshima.** 1983. The *erb*B gene of avian erythroblastosis virus is a member of the *src* gene family. Cell **35**:71–78.