The EMBO Journal vol.4 no.11 pp.2897–2903, 1985

The structure of the human c-*fes/fps* proto-oncogene

Anton J.M.Roebroek, Jack A.Schalken[1], Joseph S.Verbeek, Ans M.W.Van den Ouweland, Carla Onnekink, Henri P.J.Bloemers and Wim J.M.Van de Ven

Molecular Oncology Section, Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, and [1]Department of Urology, St. Radboud Universital Hospital, Nijmegen, The Netherlands

Communicated by H.Bloemendal

We have determined the complete nucleotide sequence of a human DNA fragment of ~13 kbp, which was shown by Southern blot analysis to contain the entire v-*fes/fps* cellular homolog. The v-*fes/fps* homologous sequences were dispersed over 11 kbp in 18 interspersed segments which were flanked by splice junctions. Fusion of these segments created a DNA fragment in which coding regions similar to those observed in the viral oncogenes v-*fes* of the Gardner-Arnstein (GA) and Snyder-Theilen (ST) strains of feline sarcoma virus and v-*fps* found in Fujinami sarcoma virus could be identified. A potential initiation site in the first exon was found. About 200 nucleotides downstream of a translational stop codon in the v-*fes/fps* homologous region, a poly(A) addition signal was identified. The deduced amino acid sequence has a molecular weight of 93 390 dalton resembling NCP92, the recently described human c-*fes/fps* product. The topography of human c-*fes/fps* appeared to resemble that of chicken c-*fps*.

Key words: human c-*fes/fps* proto-oncogene/nucleotide sequence

Introduction

Acutely transforming retroviruses have acquired their malignant potential by capturing proto-oncogene sequences from their natural hosts (reviewed by Fishinger, 1982; Bishop and Varmus, 1982). Three independently derived feline sarcoma virus (FeSV) isolates [Gardner-Arnstein (Gardner *et al.*, 1970), Snyder-Theilen (Snyder and Theilen, 1969) and HZ1 (Hardy *et al.*, 1981)] have captured sequences from the feline c-*fes* proto-oncogene (Frankel *et al.*, 1979; Franchini *et al.*, 1981; Hardy *et al.*, 1981; Hampe *et al.*, 1982) whereas several avian sarcoma viruses [Fujinami sarcoma virus (FSV), the PRC viruses, URI virus and 16L virus (reviewed by Bishop and Varmus, 1982; Bishop, 1983)] have acquired similar sequences from the avian counterpart c-*fps* (Shibuya *et al.*, 1980; Shibuya and Hanafusa, 1982; Groffen *et al.*, 1983). The translational products of these viral transforming genes are polyproteins which possess tyrosine-specific protein kinase activity *in vitro* and are capable of autophosphorylation as well as phosphorylation of exogenous protein substrates (Ruscetti *et al.*, 1980; Van de Ven *et al.*, 1980a, 1980b; Barbacid *et al.*, 1981; Beemon, 1981; Mathey-Prevot *et al.*, 1982). Analysis of mutants has shown that the enzyme activity, which is located in the carboxy-terminal region of the polyproteins (Levinson *et al.*, 1981; Barker and Dayhoff, 1982; Weinmaster *et al.*, 1983), is essential for maintenance of the transformed state (Donner *et al.*, 1980; Pawson *et al.*, 1980; Reynolds *et al.*, 1981; Hanafusa *et al.*, 1981; Lee *et al.*, 1981).

The translational product of the c-*fes/fps* proto-oncogene has been identified in a number of species. In chicken myeloblasts, a 98 000 mol. wt. protein (NCP98) (Mathey-Prevot *et al.*, 1982) was found and in feline embryo fibroblasts and cells of epithelial or lymphoid origin (Barbacid *et al.*, 1980) a 92 000 mol. wt. protein (NCP92). NCP98 was also shown to exhibit associated protein kinase activity (Mathey-Prevot *et al.*, 1982). The murine and human c-*fes/fps* proto-oncogene products have recently been identified in myeloid cells as NCP92 and these proteins were found to be cAMP-independent protein kinases with a marked preference for Mn^{2+} over Mg^{2+} and capable of using only ATP as a donor of γ-phosphate (Feldman *et al.*, 1985). The presence of *fes/fps*-related RNA transcripts in human and chicken myeloid cells has also been described. The transcript of the human gene is 2.6 kb (Slamon *et al.*, 1984) and that of the chicken gene was reported as 3.2 kb (Huang *et al.*, 1985) and as 2.75 kb (Samarut *et al.*, 1985).

To define further the *fes/fps* proto-oncogene, we have determined the complete nucleotide sequence of a molecular clone of human c-*fes/fps* (Groffen *et al.*, 1982). The results reported in this paper provide a detailed molecular description of it. We have compared the putative coding sequences of the human c-*fes/fps* gene with those deduced from sequence data of the v-*fes* gene of GA-FeSV and ST-FeSV (Hampe *et al.*, 1982), the v-*fps* gene of FSV (Shibuya and Hanafusa, 1982) and the chicken c-*fps* proto-oncogene (Huang *et al.*, 1985). In addition, the phosphokinase domain of human c-*fes/fps* was compared with those of other members of the tyrosine kinase multigene family.

Results and discussion

Topography and nucleotide sequence of human c-fes/fps

A human DNA fragment of ~13 kbp, which was shown by Southern blot analysis to contain the entire v-*fes/fps* cellular homolog (Groffen *et al.*, 1982; Franchini *et al.*, 1982; Trus *et*

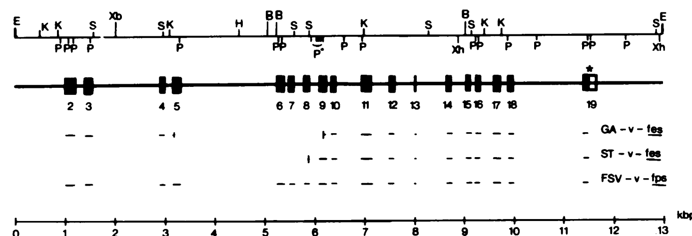

Fig. 1. Topography of the human v-*fes/fps* homologous region. A schematic restriction map of the 13 kbp *Eco*RI DNA fragment is presented. Black boxes represent the human v-*fes/fps* homologous segments. These putative exons are numbered similarly to the chicken locus (Huang *et al.*, 1985). The asterisk above exon 19 indicates a stop codon. The presence of homologous segments in GA-v-*fes*, ST-v-*fes* and FSV-v-*fps* is indicated by lines. B, *Bam*HI; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; P*, cluster of *Pst*I sites; S, *Sst*I; Xb, *Xba*I; Xh, *Xho*I.

© IRL Press Limited, Oxford, England.

Kpn1
ggtacccatggctgcgtgtgagcgaggcaggaccccacctcctccccgtctgcagtccatcctgaccctacagtccccagtctcctcgtcccatgcctccgtctccagctgctgccttgc 120
-> ex2
ctccagggatggcccctttctgtccccagAACAGCACTATGGGCTTCTCTTCCGAGCTGTGCAGCCCCCAGGGCCACGGGGTCCTGCAGCAAATGCAGGAGGCCGAGCTTCGTCTACTG 240

GAGGGCATGAGAAAGTGGATGGCCCAGCGGGTCAAGAGTGACAGGGAGTATGCAGGACTGCTTCACCACATGTCCCTGCAGGACAGTGGGGGCCAGAGCCGGGCCATCAGCCCTGACAGC 360
ex2 <-
CCCATCAGTCAGgtgggtctctatgggactctggtgggtgctggcgtatctgccttctccttcctctcctgggggccctctggggcagtggctggagatctggcaggccaatgcttggga 480
-> ex3
gccattgtgccccctccctgcctcccccatctgtgctgtatagTCCTGGGCTGAGATCACCAGCCAAACTGAGGGCCTGAGCCGCTTGCTGCGGCAGCACGCAGAGGATCTGAACTCAG 600
ex3 <-
GGCCCCTGAGCAAGCTGAGCCTGCTCATCCGGGAACGGCAGCAGCTTCGCAAGACCTACAGCGAGCAGTGGCAGCAGCTGCAGCAGGAGCTCACCAAGgtgagcgggcagcactggggct 720

tcggtcatttctgtctaaattttgagcctcgaaggggttgttttgcacaagaggccctggattcactggggaagtgtaagtccctgaccgcaggcctggcttgctctaaccttgatgtag 840
cttcctctcttccttcccctacgttgagctggcttgcagcaaggcctctctgtgcttttctgtgcctgggcaaagtgctgggagtgtaaggatgagtgaccggtcacgtgcctgggaga 960
agctcagaatcggtactcgcctccacactgtgccatctggctctgggttctgagagtcagggagaggaatgagggtcagtctgtttgccttcgacctatgcagcctcctctcagggcccg 1080
agactgggcagcagcatggccccccgaaggtcgaggactcgggccgtgaagtcagcctgcctaggttgaatcccacccagctcctccgtctagaggctgtgtgatttggaactatttatc 1200
tgggagcctagtgcccccattcagtgtgctggtcaccctccctgcaccacaccccttcctcaagtgcagagcccagccttgccatggacccacagcggcccctggtggcccaccctggcc 1320
ccattcctcgccccaaaagatcatctgattcaagggtgggcccatttttataaagttttgctggaacacagctatgccccttttgttttcatattgtctgtgactacaatgacagagttga 1440
gtaattgtgacagaggctctatggcctacaagcctaaaatatttatttactatctggccctttaagaaaaagactgatctagtcgaggaatctagctcagttacagatggggaaactgag 1560
gttgggcgcttgccccaacatatcccagcacataaacaggagaactgggacgagaacactgatctcgggctgtcatctattcctactgccaagaacataatttgcaggacccagtgcaaag 1680
tgaaattgtgggggtctttgttaaaagattgctaggaatttccaggtggcaataatggagaatgaaaccaagcacagggcccttctacatgtggagccccgtgtgactgcacaggccgtg 1800
cacacctgcaactggccctgcctgccaccaggctaccactgtcagtccaaggagggaccgttgtagcctgtagtctacctctttgcctccccaaggggtctgtctttaacaggctctctg 1920
atctttgactctcacgtcagcagccagctttcccagaagtctccaggtgctccttgcctgacgacaggacctttccagggcttcaccccaggcaagaatcttccacaactggggacctgc 2040
-> ex4
tgccccacactggcctctcctctctccctagACCCACAGCCAGGACATTGAGAAGCTGAAGAGCCAGTACCGAGCTCTGGCACGGGACAGTGCCCAAGCCAAGCGCAAGTACCAGGAGGC 2160
ex4 <-
CAGCAAAGgttcgtggcttcccttgctggcagggagggaatccgaagccagtgctgacctgtccttgggtacccagagagtgggggctgcctgggcctccatgctgtcatctatacccct 2280
-> ex5 ▽
tgcccccttctggcagACAAGGACCGTGACAAGGCCAAGGACAAGTATGTGCGCAGCCTGTGGAAGCTCTTTGCTCACCACAACCGCTATGTGCTGGGCGTGCGGGCTGCGCAGCTACA 2400
ex5 <-
CCACCAGCACCACCACCAGCTCCTGCTGCCCGGCCTGCTGCGGTCACTGCAGGACCTGCACGAGGAGATGGCTTGCATCCTgtaagcccgcagccccgtcccctggcccccacccttgag 2520

cagccctaagcccagccatcaggcccagaggcaggacccagaaaatccattgctgggaaggtgctggccatgtaaccacatgagaacgggacctgggccaaggattggaaacaggcaact 2640
tacctctgaattacactattccagggtctcattattccagggttttattacattcattgagcactgttctgggctctggattataccagagaacgatggtagacaaaaacatctgtcctc 2760
agggatctttcgtgttagtggagtgaggatgtgaggagcactaagagccatggagaaaaataaagcaagagaagtggatcgggacctgggagcacggaggcaagggaggaggtgacagtt 2880
gtccatagagtgatctgggaaagcctcttgagaggtgacattcaaagaggcccctgagaggggtacgggagtgaatcatggggctatttggagaaagaccattccagaaaggaggagcagc 3000
aattacacaggccttgaggtaggagagtaccagggactaatagccaggaaccaggtggtgcctcgagagtgagggaggggggagagtcatacacgaggctggaggaggcaggcgtcaaggg 3120
ctactgggtgatagagggtctagcagggccatggtgaggacttggctctgggtgaacaagaatggcatgatctgacctctgtttttttgtttcattttgttttaacttttttttgagtca 3240

gagtctcgctctgccgcccaggctagagtgcagtggcatgatctcggcttactgcaacctccgcctcccaggttcaagtgattcccctgcctcagcctcccgagtagctggaactacggg 3360

catgcgccaccacacccagctaatttttgtatttttagtagagatggggtttcaccatgttgcccaggctagtctctaattcctgggctcaaagcgatttgcctgcctctgcctcccaaa 3480

gtgccgggattacaggcatgagccaccatgcccagccctgacctctgttttaataaggccactctggctgctgtgctgcaaatagacttcagggagcaaggacagaagctgggaggccag 3600

agagcaggctcttgccataatccagatccaagcttttggccactaggacggggaggtagcaatggaggtgaggcgcggtcaggtcctggaaggtgaagccagtgggatttccctatggat 3720
tggaagtggggcgtgaaatagaggagtcaggggtcactctggggatttggcctggagcagctggaagatggagtggctgttaattcatgtagggaaggctgtgggaagaagaggtttagg 3840
agacaaggatagcagttcatttatttatttatttatttatttatttatttatttatttagagatgtagtctcattctttcgccaggctggagtgcagtggcgcgatcttggctcac 3960

tgcaacctccacctcccaggctcaagcgattctcttgcctcagcctcccgagtagccaagtagctgggactacaggcatgtgccaccatgcctggctaatttttgtatttgcttttcag 4080

tagagatggggtttcaccacgttagccaggctggtctcgaactgacctcaggcaatccacccgcctcgacctcccagtgttggtattataggcgtgagccactgtgcctggcccactgga 4200

tccttattacaactgccagtgtccctcttatatatatcaggaaatagaagattagggagaggttaaataatttgcctagagtggcatggctagctcgaagtgaggcaggggtcaacccca 4320
-> ex6
gccctgactccaaacccagggtcctaggcctgaactgcccagccttgcccagcctgaggctcccctgactggggatcccgtctcgggggcagGAAGGAGATCCTGCAGGAATACCTGGAG 4440
ex6 <-
ATTAGCAGCCTGGTGCAGGATGAGGTGGTGGCCATTCACCGGGAGATGGCTGCAGCTGCTGCCCGCATCCAGCCTGAGGCTGAGTACCAAGGCTTCCTGCGACAGTATGGgtaagccccg 4560
-> ex7
tccttgctcctgctgggcccagggctgctggcctgtccactgacggggcgctgtcccccacagGTCCGCACCTGACGTCCCACCCTGTGTCACGTTCGATGAGTCACTGCTTGAGGAGGG 4680
ex7 <-
TGAACCGCTGGAGCCTGGGGAGCTCCAGCTGAACGAGCTGACTGTGGAGAGCGTGCAGCACACgtgggtggtggctttgcacctgggctgcggcggggctcccagcagaccacgagtgtt 4800

tatgtaggcagggctaggtcgtggagactgtccacacagagctgtcaccaggtggccgggcttgcttggctctacagggatgcactggacctgggttgaggggggcaggagggctcggttc 4920
-> ex8
taatgctgcccttctcttgggtgcagGCTGACCTCAGTGACAGATGAGCTGGCTGTGGCCACCGAGATGGTGTTCAGGCGGCAGGAGATGGTTACGCAGCTGCAACAGGAGCTCCGGAAT 5040
▼ ex8 <-
GAAGAGGAGAACACCCACCCCCGGGAGCGgtgagtgggcccctgcctgcagcagcctcctgggcctccctccctcctacctaccctaactgctgctggctacccgccgcagaccgagccc 5160
-> ex9
ttattctcatccaccctcccacccgcccctgcctgcagGGTGCAGCTGCTGGGCAAGAGGCAAGTGCTGCAAGAAGCACTGCAGGGGCTGCAGGTAGCGCTGTGCAGCCAGGCCAAGCTG 5280
▽ ex9 <-
CAGGCCCAGCAGGAGTTGCTGCAGACCAAGCTGGAGCACCTGGGCCCCGGCGAGCCCCCGCCTGTGCTGCTCCTGCAGGATGACCGCCACTCCACGTCGTCCTCGgtgagctgccccatc 5400
-> ex10
cgcggccgctgcccgccaccggcctgcccacctgggctgcgctcctcatttcgccctcccctccctaagcctggccacccgctgacgtctgtccctggcctcagGAGCAGGAGCGAGA 5520
ex10 <-
GGGGGGAAGGACACCCACGCTGGAGATCCTTAAGAGCCACATCTCAGGAATCTTCCGCCCCAAGTTCTCGgtgagtggcgcccagctgggcccccctactgttgtgtttcgagtttaatc 5640

actgggatgtcctagagaggaggctctgcccaggctgcttgtattgggaagttcctctcttccctgggattccaggctgcagatgtccccagaccctgcccctgtgacccctccctttcc 5760
atcgccccagtgtgctaaagggaccagcaacctcgactattccatggctctccctgcttcaggagcggttgggggcctgtggcctggaggaggaggcaccagcttggtttggggtcttcc 5880
tgcctgggcttcccttcccagctctgcccagcgtgagcctgggccagtccaatgcccactccaggggcctgtggatggctctgcatgccactccatggttgtaaggctgagggcatata 6000
-> ex11
gggggagagagagacccccggctgcccccacggcctcttcaacaaggtggttaagtgactcctcctcgatcctcccttgcccagCTCCCTCCACCGCTGCAGCTCATTCCGGAGGTGCA 6120

GAAGCCCCTGCATGAGCAGCTGTGGTACCACGGGGCCATCCCGAGGGCAGAGGTGGCTGAGCTGCTGGTGCACTCTGGGGACTTCCTGGTGCGGGAGAGCCAGGGCAAGCAGGAGTACGT 6240
ex11 <-
GCTGTCGGTGCTGTGGGATGGTCTGCCCCGGCACTTCATCATCCAGTCCTTGGATgtgagtggggctgggacccgagccttccaggcctcactcttcccctcccttcccttccccaaggg 6360

aaatggcctttcagggtagggggtagctgccaggtcttggatgcctccctagcagggctggctggaagggggccacagagaccaccctgtccctgcaacaaaatagaggcttaagtgtgag 6480
tcctcccctggtggggcagcaggatgtcatgtgccatcagatggcatcttttctggaggtctctctgcccctggtcctgggcaggcccttctcccctgctgctctcctttccccctcc 6600
-> ex12
cagggctcacgcccccctcagaatggaggctgctgacccccgggtccctgccctgcagAACCTGTACCGACTGGAAGGGGAAGGCTTTCCTAGCATTCCTTTGCTCATCGACCACCTACTGA 6720
ex12 <-
GCACCCAGCAGCCCCTCACCAAGAAGAGTGGTGTTGTCCTGCACAGGGCTGTGCCCAAGgtgagcctgcacccagcctggcccatgccacctgtggcagggcttggggagtgtgggtcag 6840

Fig. 2. Nucleotide sequence of the human v-*fes/fps* homologous region. Sequence data are presented from the *Kpn*I site just upstream of the putative exon 2 to the *Pst*I site downstream of exon 19. Segments of the sequence that are homologous to v-*fes/fps* as well as the 3′ non-coding sequences in exon 19 up to a potential poly(A) addition signal are printed in capitals and indicated by arrows labeled ex2 to ex19. The AG sequence of alternative splice junctions for exon 2 are underlined; asterisk (*), termination codon of the long open reading frame of putative exons of human c-*fes/fps*; a potential poly(A) addition signal (AATAAA), is underlined; the *Alu* repeats are underlined by broken lines. Between (∇), nucleotide sequences of human c-*fes/fps* that are not represented in Ga-v-*fes*; (▼), start of nucleotide sequences homologous to ST-v-*fes*.

al., 1982) was isolated from a previously described cosmid clone (Groffen *et al.*, 1982). Nucleotide sequence analysis of this human DNA fragment and comparison with nucleotide sequences of the v-*fes* (Hampe *et al.*, 1982) and v-*fps* (Shibuya and Hanafusa, 1982) viral oncogenes and the chicken c-*fps* proto-oncogene (Huang *et al.*, 1985) revealed the distribution of the v-*fes/fps* homologous sequences over a DNA region of ~11 kbp. It should be noted that in the comparative analysis with the two v-*fes* oncogenes the complete nucleotide sequence of GA-v-*fes* and only the small unique region of ST-v-*fes* was used. Figure 1 shows a restriction map of the 13 kbp DNA fragment and, schematically, the topographical distribution of 18 v-*fes/fps* homologous genetic segments that could be identified. Numbers were assigned to the putative c-*fes/fps* exons in such a way that corresponding exons in human and chicken (Huang *et al.*, 1985) received the same number. The size and distribution of the human and chicken exons appeared highly similar from exon 3 to exon 19. However, no human DNA segment corresponding to chicken exon 1 was found and human exon 2 seemed much smaller than the chicken counterpart. In other words, 140 bp at the 5′ end of FSV-v-*fps*,

```
c-fes  acctcctccccgtctgcagtccatcctgaccctacagtccccagtctcctcgtcccatgcctccgtctccagctgctgccttgcctccagggatggccccttttctgtccccagAACAGC  156
                                                                                                                           ::::
c-fps                                                        atccatgcagGCTTCGGAGTCGCGGCACAGGCTGCCCCACTGCTCTGCAGCCCCATCCCACCAGGATCACAGT

          *
c-fes  ACTATGGGCTTCTCTTCCGAGCTGTGCAGCCCCCAGGGCCACGGGGTCCTGCAGCAAATGCAGGAGGCCGAGCTTCGTCTACTGGAGGGCATGAGAAAGTGGATGGCCCAGCGGGTCAAG  276
        : ::::::::      : ::::::::  ::::  :::: ::: : :  :::: ::   :::::::   :::::: :: :: ::::::    ::::  ::::::::: : ::::: : ::::
c-fps  GCCATGGGCTTTGGGCCGGAGCTGTGGTGCCCGAAGGGGCACAGTGAGCTGCTGCGGCTGCAGGACAGCGAGCTGCGCCTCCTGGAGCTGATGAAGAAGTGGATGTCACAGCGTGCCAAG
          *

c-fes  AGTGACAGGGAGTATGCAGGACTGCTTCACCACATGTCCCTG            CAGGACAGTGGGGGCCAGAGCCGGGCCATCAGCCCTGACAGCCCCATCAGTCAGgtgggtctctat  396
       :: ::: ::::::: :: ::  :::: :::::::::: :            :::::  :   ::: ::   ::: :::: :  ::    :::::  ::: :  ::
c-fps  AGCGACCGGGAGTACGCGGGGATGCTGCACCACATGTTCTCTCAGCTGGAGAAACAGGAGGGCCTGGGACATCTCCGTGCCACCGACCACAGCAGCCAGATCGGGGAGgtagggacact
```

Fig. 3. Comparison of the nucleotide sequence of c-*fes/fps* exon 2 and its immediate flanking sequences in man and chicken. As in Figure 2, the v-*fes/fps* homologous segment is represented by capitals. Note that the 5′ end of the human exon is not precisely defined. Possible alternatives for splice junctions are underlined. The presumed first methionine codons in human and chicken c-*fes/fps* are indicated by asterisks.

```
                                                  *
c-fes                                              NSTMGFSSELCSPQGHGVLQQMQEAELRLLEGMRKWMAQRVKSDREYAGLLHHMSLQDSGGQ    SRAISPDS   70
v-fes                                          AARADG---------------AE---------------------------------------G--RGTG       -Y-
c-fps  ASGQLHRPQPQEHTSTSAAAGTWRHTQASESRHRLPHCSAAPSHQDH-A---GP--WC-K--SE-LRL-DS------L-K---S--A--------M----FS-LEKQEGLGHL--TDHS-

c-fes  PISQSWAEITSQTEGLSRLLRQHAEDLNSGPLSKLSLLIRERQQLRKTYSEQWQQLQQELTKTHSQDIEKLKSQYRALARDSAQAKRKYQEASKDKDRDKAKDKYVRSLWKLFAHHNRYV  190
v-fes  ----------------------------------G---------------------------N-------------------R-------------
c-fps  Q-GE--WVLA----T--QT--R---E-AA---A---I---DK-----AF-------S--YAR-TQ-EM----A---S-V---T------------E-E---E---------Y-L--Q--

c-fes  LGVRAAQLHHQHHHQLLLPGLLRSLQDLHEEMACILKEILQEYLEISSLVQDEVVAIHREMAAAAARIQPEAEYQGFLRQYGSAPDVPPCVTFDESLLEEGEPLEPGELQLNELTVESVQ  310
v-fes
c-fps  -A----A---H--Y-RA--T-HE--YS-QQ--VLV-----G--CS------ED-L---Q-V-H-VEM-D-AT--SS-VQCHRYDSE---A----------T-S------------I----.

c-fes  HTLTSVTDELAVATEMVFRRQEMVTQLQQELRNEEENTHPRERVQLLGKRQVLQEALQGLQVALCSQAKLQAQQELLQTKLEHLGPGEPPPVLLLQDDRHSTSSSEQEREGGRTPTLEIL  430
v-fes                                            ----Q----A-K--------A--------------R----A---Q----------------------------------------
c-fps  -S---IEE--LASR-A-SSKEQR-WE--V---G--LALS-G---H------G----Q-Q--GLV-A-------RDM-AN--AE--SE----A-P--E--Q-VC-TD---S-V - A--TI

c-fes  KSHISGIFRPKFSLPPPLQLIPEVQKPLHEQLWYHGAIPRAEVAELLVHSGDFLVRESQGKQEYVLSVLWDGLPRHFIIQSLDNLYRLEGEGFPSIPLLIDHLLSTQQPLTKKSGVVLHR  550
v-fes  --------------------V---------------L---------T-----------------------Q--------A-----P--D--A-----V----RS---------I--N-
c-fps  -N------S-R------VP---------CQ-A--------S--Q---KC----------------------Q-------AA--------D---T---------QS---I-R---I--T-

c-fes  AVPKDKWVLNHEDLVLGEQIGRGNFGEVFSGRLRADNTLVAVKSCRETLPPDLKAKFLQEARILKQYSHPNIVRLIGVCTQKQPIYIVMELVQGGDFLTFLRTEGARLRVKTLLQMVGDA  670
v-fes  ---------------------------------------------------I--------K-----------------------------------------------M---------
c-fps  --L----------VL---R-------------------P------------E---------------N-----------------------------S---S--PH-KM-E-IK-MEN-

                                                     o
c-fes  AAGMEYLESKCCIHRDLAARNCLVTEKNVLKISDFGMSREEADGVYAASGGLRQVPVKWTAPEALNYGRYSSESDVWSFGILLWETFSLGASPYPNLSNQQTREFVEKGGRLPCPELCPD  790
v-fes  ----------------------------------------A---I--------------------------------------------------------------------------
c-fps  ----------H----------------T----------Q-E-----ST--MK-I-----------------------------A-----V--A---------AI-Q-V--EP--Q--E

c-fes  AVFRLMEQCWAYEPGQRPSFSTIYQELQSIRKRHR
v-fes  ---------------------A-------------
c-fps  D-Y---QR--E-D-RR----GAVH-D-IA------
```

Fig. 4. Similarities between the deduced amino acid sequences of the gene products of human and chicken c-*fes/fps* and feline v-*fes* (GA- and ST-v-*fes*). The amino acid sequence of human c-*fes/fps* is shown in the conventional one-letter code. Amino acid substitutions in the chicken c-*fes/fps* and feline v-*fes* products relative to the human product are indicated. Both GA- and ST-v-*fes* data were used as explained in the text. Identical amino acids are represented by hyphens (-). Positions not represented are left blank. Open circle (○), possible phosphoacceptor tyrosine. Asterisk (*), possible initiation site.

present in exon 1 and the first 57 bp of exon 2 of chicken c-*fps*, remained unaccounted for.

As indicated in Figure 1, homologous genetic sequences of some of the putative exons (exons 2 – 9) were completely present only in FSV-v-*fps* and not in the feline v-*fes* oncogenes. ST-v-*fes* homologous sequences started 14 nucleotides before the end of exon 8. GA-v-*fes* lacked sequences homologous to human c-*fes/fps* sequences between nucleotide 17 in exon 5 and nucleotide 107 in exon 9. On the other hand, 12 bp at the 5′ end of GA-v-*fes* remained unaccounted for in the human c-*fes/fps* sequences. The human c-*fes/fps* sequences that are homologous to FSV-v-*fps* and are missing in the two feline viral *fes* genes were also present in cosmid clones that contained the feline c-*fes/fps* homolog (Verbeek *et al.*, 1985) (data not shown). Apparently, these sequences that do not seem to be essential for the transforming potential of these oncogenes, have been lost during or subsequent to the generation of the feline viral oncogenes. This conclusion is in agreement with the transforming potential of v-*fps* of the virus PRCII which, as a result of a deletion, lacks sequences corresponding to the end of exon 2 to somewhere in the middle of exon 9 of c-*fes/fps* (Huang *et al.*, 1985).

In Figure 2, the nucleotide sequence of the complete human v-*fes/fps* cellular homolog is presented. As can be seen, all v-*fes/fps* homologous segments were flanked by the AG and GT splice junction sequences and, in most cases, good agreement with the complete consensus splice junction sequences (Mount, 1982) was observed. Fusion of these segments resulted in a DNA fragment with a coding region that was remarkably similar to that in FSV-v-*fps*, GA-v-*fes* and ST-v-*fes*. It also resembled quite accurately the combined putative exons of chicken c-*fps*. Although the size of most of the corresponding human and chicken exons were identical, at the 5′ end the human proto-oncogene revealed a number of differences, which are summarized in Figure 3. The first human exon that could be identified corresponded to chicken exon 2. We have tentatively placed the 5′ end of this exon at a potential splice site at a position where sequence homology between human c-*fes/fps* and chicken c-*fps* diverge. However, other splice sites could be involved, all resulting in an open reading frame in phase with the large open reading frame (mentioned below). In the interpretation presented here, exon 2 is 69 nucleotides smaller than its chicken counterpart. The first 57 nucleotides of chicken exon 2 were not represented in human exon 2. At ~50 nucleotides from the 3′ end of exon 2, 12 nucleotides found in the chicken gene are missing in the human gene. The feline c-*fes/fps* locus probably lacks 18 nucleotides at this spot, as could be deduced from v-*fes* sequence data (Hampe *et al.*, 1982) (see also Figure 4). The only other difference in size between the human and chicken c-*fes/fps* coding sequences

was observed in exon 10 where the human exon contains an additional stretch of six nucleotides. The same six nucleotides are found at the homologous site in the feline viral *fes* genes (Hampe *et al.*, 1982).

At their 3′ ends, the *fes/fps* loci of man and chicken diverge downstream of the TGA codon in exon 19. At a position of ~200 nucleotides downstream of this termination codon, a poly(A) addition signal was present. No sequence homology in the region from the termination codon to the potential poly(A) addition signal could be observed in these species. However, some sequence homology in this region was observed between man and cat, when comparison with the v-*fes* sequence data was made.

The intervening sequences were analyzed for the presence of highly repetitive sequences such as the *Alu, Eco, Hinf* and *Kpn* repeats. Only *Alu* repeats were identified. They were found clustered in the intervening sequences between exon 5 and 6, between exon 18 and 19 and downstream of exon 19. Interestingly, the human intervening sequences that contained the *Alu* repeats were much larger than the corresponding chicken introns. For instance, ~70% of the intervening sequences between exon 18 and 19 represented *Alu* repeats.

In human exon 2, a potential initiation site was found (indicated with an asterisk in Figure 3) from which an open reading frame of 2466 nucleotides extended up to a termination codon in exon 19. This open reading frame together with a non-coding region of ~200 nucleotides from the termination codon to the potential poly(A) addition signal gives a putative mRNA with a molecular size of ~2.7 kb not including a poly(A) tail and as yet unidentified 5′ sequences. This value is in the range of *fes/fps* mRNA sizes reported by others (see Introduction). Furthermore, the molecular weight of the deduced gene product is 93 390 and resembles that of the human and murine c-*fes/fps* product NCP92 recently described by Feldman *et al.* (1985). It should be noted that the assignment of the above-mentioned ATG as initiation codon would be wrong if exon 2 started at some splice site further upstream, since involvement of one of these hypothetical splice sites would lead to the presence further upstream in the exon of one or more other ATG codons in another reading frame. But interestingly, in the chicken (Huang *et al.*, 1985) and feline (Hampe *et al.*, 1982) locus a methionine codon is found in the same position (see Figure 4). For the chicken gene it was also proposed as the initiation codon (Huang *et al.*, 1985). Upstream of the putative initiation codon clear divergence between man and chicken was observed. Such divergence was not found in any of the coding segments. Nucleotide sequences homologous to the 140 bp of the 5′ end of FSV-v-*fps*, also present in chicken c-*fps* (Huang *et al.*, 1985), were not only absent in the 13 kbp *Eco*RI v-*fes/fps* homologous DNA segment but could also not be detected in hybridization experiments under conditions of reduced stringency in a human DNA region of ~9 kbp immediately upstream of v-*fes/fps* homologous segment (data not shown). Further sequence analysis of a 3 kbp segment immediately upstream of the v-*fes/fps* homologous segment did not reveal any homologous sequences either. The 12 bp at the 5′ end of GA-v-*fes* were also not found. This divergence could be explained by genetic drift upstream of the coding region of c-*fes/fps*. For these reasons, we tentatively conclude that the 140 bp are noncoding exon sequences in chicken c-*fps* because they precede a potential initiation site also found in the human c-*fes/fps* at the position where the long conserved open reading frame starts. However, sequence analysis of cDNA of human and chicken c-*fes/fps* will probably be necessary to resolve this matter.

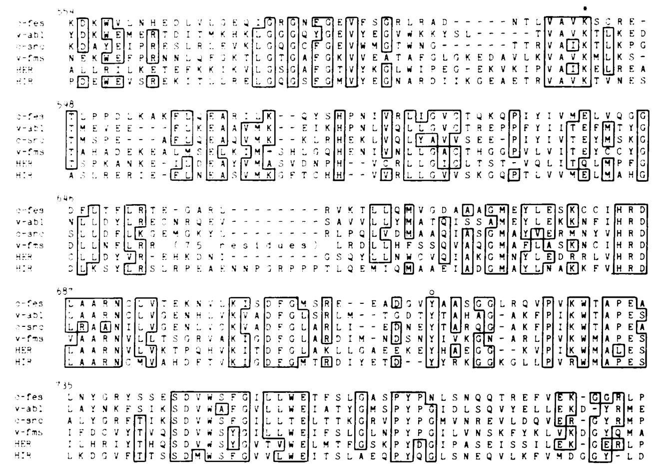

Fig. 5. Similarities between the deduced amino acid sequences of the human c-*fes/fps*-encoded tyrosine-specific protein kinase domain and other proteins. The deduced amino acid sequence of the human c-*fes/fps* product (residues 554–825) was aligned for optimal match with those deduced from v-*abl* (Reddy *et al.*, 1983), chicken c-*src* (Takeya and Hanafusa, 1983), v-*fms* (Hampe *et al.*, 1984), human epidermal growth factor receptor gene (HER) (Ullrich *et al.*, 1984) and human insulin receptor gene (HIR) (Ullrich *et al.*, 1985). Boxes, common residues among at least four of the six proteins; asterisk (*), lysine residue typifying the ATP-binding site; open circle (○), possible phosphoacceptor tyrosine.

Evolutionary conservation of the fes/fps proto-oncogene

Hybridization analysis has indicated that proto-oncogenes in general are highly conserved during evolution. The availability of nucleotide sequence data of human and chicken c-*fes/fps* enabled a more precise determination of the extent of conservation of particular segments of this proto-oncogene and its deduced gene product. We therefore compared the deduced amino acid sequences of the *fes/fps*-encoded gene products of three species, namely man, cat and chicken (Figure 4). As the feline gene product, we used GA- and ST-v-*fes* sequence data (Hampe *et al.*, 1982) since we expected these data to be highly representative of the feline c-*fes* gene. Compare for instance, the amino acid homology between chicken c-*fps* and FSV-v-*fps* which is more than 97% (Huang *et al.*, 1985). As already indicated above, the viral oncogenes of GA- and ST-FeSV captured only parts of the feline proto-oncogene and, therefore, comparison was limited. As can be seen in Figure 4, the overall homology between the feline and human coding sequences (94% at the amino acid level, 91% at the DNA level) is greater than that between chicken and human (70% at the amino acid level, 74% at the DNA level). Furthermore, it appeared that conservation was higher in the 3′ region (exon 11–exon 18). The average amino acid homology in this area between man and chicken is ~85% (80% at the DNA level). This homology is in good agreement with the results of Feldman *et al.* (1985), that showed that the human c-*fes/fps* product was detected in an immunoprecipitation analysis using a conventional antiserum as well as one prepared with a synthetic dodecapeptide corresponding to a particular amino acid sequence of (the chicken virus) FSV. In accordance with these results, our sequence data shows that the corresponding region in NCP92 shares 10 out of 12 amino acids, nine of which lie in one stretch. In this region of strong homology the protein kinase domain is

located (Barker and Dayhoff, 1982; Levinson *et al.*, 1981; Weinmaster *et al.*, 1983). These results indicate a stronger conservation of the protein kinase domain relative to other portions of the c-*fes/fps*-encoded gene product.

To investigate more specifically the shared genetic sequences of gene segments that encode tyrosine-specific protein kinases, we compared the deduced amino acid sequence of the kinase domain of the human c-*fes/fps* proto-oncogene with those encoded by v-*abl* (Reddy *et al.*, 1983), chicken c-*src* (Takeya and Hanafusa, 1983), v-*fms* (Hampe *et al.*, 1984), the human epidermal growth factor receptor gene (Ullrich *et al.*, 1984) and the insulin receptor gene (Ullrich *et al.*, 1985). As can be seen in Figure 5, there is extensive structural homology between the predicted protein portions of the different gene products. They all reveal a tyrosine phosphorylation site embedded in remarkably similar surroundings. Furthermore, they all possess in a similar position a lysine residue which is thought to be part of the ATP-binding site (Barker and Dayhoff, 1982). The presence of highly similar kinase segments in a number of different tyrosine-specific protein kinases, each widely distributed among species, indicates that an early stage of the evolution a single ancestral domain gave rise to the development of a multigene family. The members of this gene family fulfill universal, yet pluriform, tasks in cell differentiation and development. Their gene products, all being protein kinases, probably function in a mechanistically similar manner.

The precise biological role of the c-*fes/fps* gene product is not yet clear. It was recently suggested that expression of NCP92 was related to the capacity of myeloid cells to differentiate and respond to certain colony-stimulating factors (Feldman *et al.*, 1985). The functional association of tyrosine-specific protein kinases with growth factor receptors has been reported (Hunter and Cooper, 1981; Kasuga *et al.*, 1982). Whether or not the c-*fes/fps* gene product is associated with a growth factor receptor remains to be established. In this context, it should be noted that the v-*fes*-encoded tyrosine-specific protein kinase appeared to be associated with a 150 000 kd cellular protein that serves as a phosphate acceptor (Reynolds *et al.*, 1980). This apparently highly conserved cellular protein in its turn also exhibited an associated protein kinase activity, in this case with a specificity for serine and threonine. Apparently, the two proteins are links in a regulatory pathway, the elucidation of which may clarify the malignant potential of this proto-oncogene.

Materials and methods

Molecular cloning

Isolation of the human v-*fes/fps* cellular homolog from a cosmid library has been described previously (Groffen *et al.*, 1982). A 13 kbp *Eco*RI DNA fragment, which contains all v-*fes/fps* homologous genetic sequences and a 9.2 kbp *Hpa*I-*Eco*RI DNA fragment flanking the former at its 5' end, were subcloned in pSP64. The feline v-*fes/fps* cellular homolog was isolated similarly from a feline cosmid library (Verbeek *et al.*, 1985). Hybridization experiments were performed as described (Schalken *et al.*, 1985), except that hybridization under conditions of reduced stringency were performed at 42°C in buffer containing 30% formamide.

DNA sequence analysis

DNA fragments were inserted into the polylinker region of M13mp8-11 (Messing and Vieira, 1982). All of the DNA sequences were determined by the dideoxy-sequencing method as described by Sanger *et al.* (1977). All parts of the reported DNA sequence were obtained from both strands of the cloned DNA. The gel readings were recorded, edited and compared using the Staden programs (Staden, 1982).

Acknowledgements

We thank J.R.Stephenson and J.Groffen for the human v-*fes/fps* cellular homolog. We acknowledge the assistance of J.Leunissen with the computer analysis and of E.Orlemans and H.Theunissen for their help with the DNA sequencing. This work was supported by the Netherlands Cancer Foundation, Contract No. NUKC Bioch 80-1, and by the Foundation for Medical Research FUNGO which is subsidized by the Netherlands Organization for Advancement of Pure Research (Z.W.O.).

References

Barbacid,M., Beemon,K. and Devare,S.G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5158-5162.

Barbacid,M., Breitman,M.L., Lauver,A.V., Long,L.K. and Vogt,P.K. (1981) *Virology*, **110**, 411-419.

Barker,W.C. and Dayhoff,M.O. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2836-2839.

Beemon,K. (1981) *Cell*, **24**, 145-153.

Bishop,J.M. (1983) *Annu. Rev. Biochem.*, **52**, 301-354.

Bishop,J.M. and Varmus,H. (1982) in Weiss,R., Teich,N., Varmus,H.E. and Coffin,J. (eds.), *Molecular Biology of RNA Tumor Viruses: RNA Tumor Viruses*, Cold Spring Harbor Laboratory, NY, pp. 999-1108.

Donner,L., Turek,L.P., Ruscetti,S.K., Fedele,L.A. and Sherr,C.J. (1980) *J. Virol.*, **35**, 129-140.

Feldman,R.A., Gabrilove,J.L., Tam,J.P., Moore,M.A.S. and Hanafusa,H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2379-2383.

Fishinger,P.T. (1982) in Stephenson,J.R. (ed.), *Molecular Biology of RNA Tumor Viruses*, Academic Press, NY, pp. 162-250.

Franchini,G., Even,J., Sherr,C.J. and Wong-Staal,F. (1981) *Nature*, **290**, 154-157.

Franchini,G., Gelmann,E.P., Dalla Favera,R., Gallo,R.C. and Wong-Staal,F. (1982) *Mol. Cell. Biol.*, **2**, 1014-1019.

Frankel,A.E., Gilbert,J.H., Porzig,K.J., Scolnick,E.M. and Aaronson,S.A. (1979) *J. Virol.*, **30**, 821-827.

Gardner,M.B., Rongey,R.W., Arnstein,P., Estes,J.D., Sarma,P., Huebner,R.J. and Rickard,C.G. (1970) *Nature*, **226**, 807-809.

Groffen,J., Heisterkamp,N., Grosveld,F., Van de Ven,W.J.M. and Stephenson,J.R. (1982) *Science (Wash.)*, **216**, 1136-1138.

Groffen,J., Heisterkamp,N., Shibuya,M., Hanafusa,H. and Stephenson,J.R. (1983) *Virology*, **125**, 480-486.

Hampe,A., Laprevotte,I., Galibert,F., Fedele,L.A. and Sherr,C.J. (1982) *Cell*, **30**, 775-785.

Hampe,A., Gobet,M., Sherr,C.J. and Galibert,F. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 85-89.

Hanafusa,T., Mathey-Prevot,B., Feldman,R.A. and Hanafusa,H. (1981) *J. Virol.*, **38**, 347-355.

Hardy,W.D.,Jr., Zuckerman,E., Markovich,R., Besmer,P. and Snyder,H.W.,Jr. (1981) in Yohn,D. and Blakeslee,J.R. (eds.), *Advances in Comparative Leukemia Research 1981*, Elsevier, Amsterdam, pp. 205-206.

Huang,C.-C., Hammond,C. and Bishop,J.M. (1985) *J. Mol. Biol.*, **181**, 175-186.

Hunter,T. and Cooper,J.A. (1981) *Cell*, **24**, 741-752.

Kasuga,M., Karlsson,F.A and Kahn,C.R. (1982) *Science (Wash.)*, **215**, 185-187.

Lee,W.H., Bister,K., Moscovici,C. and Duesberg,P.H. (1981) *J. Virol.*, **38**, 1064-1076.

Levinson,A.D., Courtneidge,S.A. and Bishop,J.M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1624-1628.

Mathey-Prevot,B., Hanafusa,H. and Kawai,S. (1982) *Cell*, **28**, 897-906.

Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.

Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459-472.

Pawson,T., Guyden,J., Kung,T.H., Radke,K., Gilmore,T. and Martin,G.S. (1980) *Cell*, **22**, 767-775.

Reddy,E.P., Smith,M.J. and Srinivasan,A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3623-3627.

Reynolds,F.H., Van de Ven,W.J.M. and Stephenson,J.R. (1980) *J. Biol. Chem.*, **255**, 11040-11047.

Reynolds,F.H., Van de Ven,W.J.M., Blomberg,J. and Stephenson,J.R. (1981) *J. Virol.*, **37**, 643-653.

Ruscetti,S.K., Turek,L.P. and Sherr,C.J. (1980) *J. Virol.*, **35**, 259-264.

Samarut,J., Mathey-Prevot,B. and Hanafusa,H. (1985) *Mol. Cell. Biol.*, **5**, 1067-1072.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Schalken,J.A., Van den Ouweland,A.M.W., Bloemers,H.P.J. and Van de Ven,W.J.M. (1985) *Biochim. Biophys. Acta*, **824**, 104-112.

Shibuya,M. and Hanafusa,H. (1982) *Cell*, **30**, 787-795.

Shibuya,M., Hanafusa,T., Hanafusa,H. and Stephenson,J.R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6536-6540.

Slamon,D.J., deKernion,J.B., Verma,I.M. and Cline,M.J. (1984) *Science (Wash.)*, **224**, 256-262.

Snyder,S.P. and Theilen,G.H. (1969) *Nature*, **221**, 1074-1075.

Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951-2961.
Takeya,T. and Hanafusa,H. (1983) *Cell*, **32**, 881-890.
Trus,M.D., Sodroski,J.G. and Haseltine,W.A. (1982) *J. Biol. Chem.*, **257**, 2730-2733.
Ullrich,A., Coussens,L., Hayflick,J.S., Dull,T.J., Gray,A., Tam,A.W., Lee,J., Yarden,Y., Liberman,T.A., Schlessinger,J., Downward,J., Mayes,E.L.V., Whittle,N., Waterfield,M.D. and Seeburg,P.H. (1984) *Nature*, **309**, 418-425.
Ullrich,A., Bell,J.R., Chen,E.Y., Herrera,R., Petruzelli,L.M., Dull,T.J., Gray,A., Coussens,L., Liao,Y.-C., Tsubokawa,M., Mason,A., Seeburg,P.H., Grunfeld,C., Rosen,O.M. and Ramachandran,J. (1985) *Nature*, **313**, 756-761.
Van de Ven,W.J.M., Khan,A.S., Ryenolds,F.H., Mason,K.T. and Stephenson,J.R. (1980a) *J. Virol.*, **33**, 1034-1045.
Van de Ven,W.J.M., Reynolds,F.H. and Stephenson,J.R. (1980b) *Virology*, **101**, 185-197.
Verbeek,J.S., Van de Ouweland,A.M.W., Schalken,J.A., Roebroek,A.M.J., Onnekink,C., Bloemers,H.P.J. and Van de Ven,W.J.M. (1985) *Gene*, **35**, 33-43.
Weinmaster,G., Hinze,E. and Pawson,T. (1983) *J. Virol.*, **46**, 29-41.

Received on 19 June 1985; revised on 29 July 1985