GENETIC ORGANIZATION
OF THE HUMAN AND
FELINE c-fes/fps
PROTO-ONCOGENES

Anton Roebroek
GENETIC ORGANIZATION OF THE HUMAN AND
FELINE c-fes/fps PROTO-ONCOGENES
The studies presented in this thesis were supported by the Netherlands Cancer Foundation, Contract No. NUKC Biochem. 85-9 and by the Foundation for Medical Research FUNGO which is subsidized by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

This thesis was sponsored by Holland Biotechnology bv.
GENETIC ORGANIZATION OF THE HUMAN AND FELINE c-fes/fps PROTO-ONCOGENES

Proefschrift

ter verkrijging van de graad van doctor in de wiskunde en natuurwetenschappen aan de Katholieke Universiteit te Nijmegen op gezag van de Rector Magnificus Prof. Dr. B.M.F. van Iersel volgens besluit van het College van Decanen in het openbaar te verdedigen op donderdag 19 november 1987 des namiddags te 3.30 uur

door

Antonius Johannes Maria Roebroek
Geboren te Vleuten-De Meern

Nijmegen, 1987
Promotor: Prof. Dr. H.P.J. Bloemers
Co-referent: Dr. W.J.M. Van de Ven
Aan mijn vader, mijn moeder
en Marjan
Een van de eerste pagina's van mijn proefschrift heb ik gereserveerd voor een woord van dank aan alle mensen die op enigerlei wijze hebben bijgedragen aan de totstandkoming van dit proefschrift.

Speciaal wil ik Carla Onnekink bedanken voor de ondersteuning bij het experimentele werk. Daarnaast wil ik Jack Schalken bedanken voor de vruchtbare samenwerking. Jack Leunissen wil ik roemen voor zijn geduld en begeleiding bij mijn intrede in het informatica tijdperk.

Verder wil ik de andere leden van de werkgroep Moleculaire Oncologie niet onvernoemd laten: Adrie van Bokhoven, Marion Bussemakers, Jan van Groningen, Maddy de Jong, Cor Lamers, Han van Neck, Ans van den Ouweland en Sjef Verbeek. Ook wil ik Pieter Magusin, Lex Menco, Eric Orlemans en Henri Theunissen bedanken voor hun bijdrage aan mijn onderzoek in het kader van hun doctoraalstudie.

Tenslotte een woord van dank aan alle overige medewerkers van de afdeling Biochemie voor de uitstekende werksfeer.
CONTENTS

CHAPTER 1 Introduction
  1. Proto-oncogenes
  2. The c-fes/fps proto-oncogene
  3. Protein kinases
    3.1. Perspectives
    3.2. Serine/threonine-specific kinases
    3.3. Tyrosine-specific kinases
    3.4. Network of signal pathways
  4. Outline of the investigations described in this thesis

CHAPTER 2 The structure of the human c-fes/fps proto-oncogene

CHAPTER 3 Characterization of human c-fes/fps reveals a new transcription unit (fur) in the immediately upstream region of the proto-oncogene

CHAPTER 4 Evolutionary conserved close linkage of the c-fes/fps proto-oncogene and genetic sequences encoding a receptor-like protein

CHAPTER 5 Structure of the feline c-fes/fps proto-oncogene: genesis of a retroviral oncogene

SUMMARY

SAMENVATTING

LIST OF PUBLICATIONS

CURRICULUM VITAE
CHAPTER 1

Introduction
INTRODUCTION

1. Proto-oncogenes

The discovery that the viral oncogene v-src, the gene responsible for the malignant properties of Rous sarcoma virus, had entered the viral genome through a genetic recombination with a so-called cellular proto-oncogene (c-onc) had a profound impact on tumor biology. More than 20 other proto-oncogenes were subsequently detected (Bishop and Varmus, 1982; Bishop, 1983). It became apparent that proto-oncogenes are highly conserved genes which under normal physiological conditions fulfill essential functions (Gonda et al., 1982; Müller et al., 1982, 1983) but, under certain aberrant conditions, may exhibit activities contributing to the development of malignancies such as sarcomas, carcinomas, leukemias and lymphomas. Transduction of proto-oncogene sequences to a retrovirus resulting in a recombinant virus with a viral oncogene (v-onc) as integral part of its genome as mentioned above for Rous sarcoma virus is not the only manner of tumorigenesis mediated by proto-oncogenes. Another mode of activation is the proviral integration by a retrovirus in the host's cellular genome nearby or in a proto-oncogene. Integration of viral genomes, through the action of viral regulatory sequences such as promoters or enhancers, may influence the expression of genes in the neighborhood of the integration site and, in the case that a proto-oncogene happens to be triggered in this way, may cause malignant activation. Retroviruses which induce malignancies by this so-called promoter or enhancer insertion model of activation are often referred to as nonacutely transforming retroviruses, because of the long latency period between infection and the appearance of the tumor. Retroviruses with a transduced viral oncogene as integral part of their genome are referred to as acutely transforming retroviruses, because of the relative short period of latency. This difference in latency period is explained by the fact that in the former case the actual activation of a proto-oncogene, a rare event, generally requires many cell infections, whereas in the latter case the activated oncogene forms part of the retrovirus and will be expressed almost immediately in each infected cell. Malignant activation without interference of (retro)viruses can be caused by somatic changes within proto-oncogenes such as point mutations, chromosome translocations, gene amplifications or insertions. In general, malig-
nant activation of proto-oncogenes involves alteration in regulation of their expression, alteration of their primary gene product, or both (Bishop, 1985).

As above mentioned, proto-oncogenes encode proteins with properties essential for normal cellular functions. Based upon (partly) identified or deduced functions of the proto-oncogenes they can be divided in several classes:

- the c-ras genes, encoding GTP-binding proteins belonging to the adenylate cyclase (AC) system or phosphatidylinositol 4,5-biphosphate (PIP<sub>2</sub>) phosphodiesterase (PDE) system within the plasma membrane (Fleischman et al., 1986);
- genes encoding DNA-binding proteins regulating gene expression (e.g. c-myc, c-fos) (Cochran et al., 1984; Kruijer et al., 1984; Müller et al., 1984);
- a gene encoding a growth factor, namely platelet-derived growth factor chain 2, PDGF-2 (c-sis) (Deuel et al., 1983; Doolittle et al., 1983; Waterfield et al., 1983);
- genes encoding serine- and/or threonine-specific protein kinases (e.g. c-mos, c-raf/mil) (Moelling et al., 1984; Maxwell and Arlinghaus, 1985);
- genes encoding tyrosine-specific protein kinases not being a growth factor receptor (e.g. c-fes/fps, c-abl) (Van de Ven et al., 1980a; 1980b; Mathey-Prevott et al., 1982; Feldman et al., 1985; Roebroek et al., this thesis, chapter 2; Konopka et al., 1985);
- genes encoding receptors with tyrosine-specific protein kinase activity, namely the colony stimulating factor-1 (CSF-1) receptor (c-fms) (Sherr et al., 1985) and the epidermal growth factor (EGF) receptor (c-erbB) (Downward et al., 1984), or a receptor without kinase activity, namely the thyroid hormone receptor (c-erbA) (Sap et al., 1986; Weinberger et al., 1986).

In general, proto-oncogene products seem to belong to signal transducing systems between cells or within cells. Therefore it is plausible, that aberrant expression of proto-oncogenes could lead to aberrant cell proliferation and differentiation and eventually to tumorigenesis.
The c-fes/fps proto-oncogene was first detected using the viral oncogene of feline sarcoma virus (FeSV) isolates (v-fes) and of several avian sarcoma viruses (v-fps) as probes. Independently, three FeSV isolates [Gardner-Arnstein (Gardner et al., 1970), Snyder-Theilen (Snyder and Theilen, 1969) and HZ1 (Hardy et al., 1981)] arose by transduction of feline c-fes proto-oncogene sequences to feline leukemia viruses (Frankel et al., 1979; Franchini et al., 1981; Hardy et al., 1981). Several avian sarcoma viruses (ASV) [Fujinami sarcoma virus (FSV), the PRC viruses, URI virus and 16L virus (reviewed by Bishop and Varmus, 1982; Bishop, 1983)] arose from transduction of similar sequences of the chicken counterpart c-fps to avian leukemia viruses (Shibuya et al., 1980; Shibuya and Hanafusa, 1982; Groffen et al., 1983). To emphasize this relationship between c-fes and c-fps the proto-oncogene is generally referred to as c-fes/fps.

Transduction of proto-oncogene sequences to a retrovirus is generally accompanied by loss of viral genetic sequences essential for virus replication and, consequently, acutely transforming retroviruses need a helper virus for replication. In case of the FeSV and ASV viruses, the v-fes and v-fps sequences replace parts of the viral gag and pol gene sequences. In the sarcoma viruses mentioned above, the translation product of the viral transforming gene consists of a feline or avian leukemia virus gag gene-encoded amino-terminal component fused to a v-fes/fps-encoded carboxy-terminal component. The fusion protein possesses tyrosine-specific protein kinase activity \textit{in vitro} 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). This kinase activity, which is located in the v-fes/fps-encoded carboxy-terminal region of the polyproteins (Weinmaster et al., 1983), is responsible for the transformation as analysis of mutants has shown (Donner et al., 1980; Pawson et al., 1980; Reynolds et al., 1981; Hanafusa et al., 1981; Lee et al., 1981).

The protein encoded by the c-fes/fps proto-oncogene also shows tyrosine-specific protein kinase activity (Mathey-Prevot et al., 1982; Feldman et al., 1985), but the function of this protein, a 92,000 molecular-weight protein (NCP92) in cat (Barbacid et al., 1980) and man (Feldman et al., 1985; McDonald et al., 1985) and a 98,000 molecular weight protein (NCP98) in chicken (Mathey-Prevot et al., 1982; Samarut et al.,
1985) is not yet known. The restriction of its expression to mainly haematopoietic tissue points to a possible function of the c-fes/fps proto-oncogene in haematopoietic proliferation and differentiation. The reported size of the c-fes/fps mRNA species varied between 2.6 and 3.2 kilobases (Slamon et al., 1984; Huang et al., 1985; Samarut et al., 1985; Roebroek et al., this thesis, chapter 3).

3. Protein kinases

3.1 Perspectives

The function of the protein encoded by the c-fes/fps proto-oncogene is not yet known. Yet, because of its tyrosine-specific protein kinase activity (Mathey-Prevott et al., 1982; Feldman et al., 1985) its classification as a transducer of signals seems obvious. In view of the importance of protein kinases in general in signal pathways between cells and inside cells, I present a brief survey of the several types of protein kinases and an outline on how these kinases may interact and form complex signal transducing systems. For more extensive reviews on protein kinase and their subclasses see Flockhart and Corbin (1982), Berndge (1985), Hunter and Cooper (1985), Nairn et al. (1985) and Nishizuka (1986).

Protein kinases in general display a relative broad specificity, often resulting in phosphorylation of several different proteins by a single kinase. They react with a number of distinct sites in target proteins. Especially by phosphorylation of enzymes, resulting in modification of their activity, protein kinases regulate several physiological processes. Based upon the specificity for amino acid residues phosphorylated by them, protein kinases may be divided into two subgroups of protein kinases: those which phosphorylate serine and/or threonine residues and those which phosphorylate tyrosine residues. In the next sections the two classes will be discussed separately, followed by a brief outline on how several signal pathways involving protein kinase activity may interact.

3.2 Serine/threonine-specific protein kinases

The serine- and/or threonine-specific protein kinases can be classified into three types according to the modifiers which regulate their activity, namely cAMP-dependent protein kinase (protein kinase A), cGMP-dependent
protein kinase (protein kinase G) and Ca\(^{2+}\)-dependent protein kinase (e.g. Ca\(^{2+}\)/CaM-dependent protein kinase and protein kinase C).

Protein kinase A belongs to the first discovered signal transducing system. Upon an external signal (first messenger: e.g. hormones or neurotransmitters) to a receptor on the outer surface of the cell membrane, stimulatory GTP-binding proteins (G\(_s\)) or inhibitory GTP-binding proteins (G\(_i\)) inside the membrane stimulate or inhibit, respectively, the activity of the adenylate cyclase (AC) on the inner surface of the membrane. Stimulation or inhibition of the conversion of ATP in cAMP by AC results through increasing or decreasing levels of cAMP in the cytoplasm in activation or inactivation of protein kinase A, which is responsible for further cellular responses (see figure 1) (reviewed by Berridge, 1985). In terms of signal pathway, the G proteins (e.g. c-ras gene-encoded proteins) are also referred to as transducers, whereas AC, cAMP and protein kinase A are indicated as amplifier, second messenger and internal effector, respectively.

Regulation of activity of protein kinase G, the cGMP-dependent kinase, seems to be different from that of protein kinase A. Although external signals as hormones or neurotransmitters activate protein kinase G through activation of guanylate cyclase (GC) and increasing cGMP levels, GC is not directly linked to receptors and G proteins (see figure 1). Ca\(^{2+}\), diacylglycerol (see below) and cAMP play a role in regulation of GC and protein kinase G, as they can increase cGMP levels in certain cell types (reviewed by Flockhart and Corbin, 1982; Nairn et al., 1985). Thus protein kinase G may be regulated by the action of other protein kinases as protein kinase A and protein kinase C (see below).

The signal pathway involving Ca\(^{2+}\)-dependent protein kinases (reviewed by Berridge, 1984, 1985) starts with an external signal intercepted by a receptor. Stimulatory GTP-binding proteins (G\(_s\)) subsequently transduce the signal to a phosphatidyl inositol 4,5-biphosphate (PIP\(_2\))-specific phosphodiesterase (PDE) located in the membrane, which hydrolyses PIP\(_2\) into the second messengers inositol triphosphate (IP\(_3\)) and diacylglycerol (DAG). Increased IP\(_3\) concentrations in the cytoplasm trigger the rapid release of Ca\(^{2+}\) from the endoplasmatic reticulum. This released Ca\(^{2+}\) then binds to a protein, called calmodulin (CaM) and this complex activates Ca\(^{2+}\)/CaM-dependent protein kinases as e.g. myosin light chain kinase and phosphorylase kinase (see figure 1). The other second messenger released by the hydrolysis of PIP\(_2\) is the membrane bound DAG which activates protein kinase C located on the inner surface of the membrane (see figure 1). This enzyme
Figure 1. Schematic representation of mode of activation of the several different protein kinases. For detailed explanation see text. Abbreviations: AC, adenylate cyclase; CaM, calmodulin; DAG, diacylglycerol; IP₃, inositol triphosphate; GC, guanylate cyclase; G₁ or s, inhibitory or stimulatory GTP-binding protein; PDE, PIP₂-specific phosphodiesterase; PIP₂, phosphatidyl inositol 4,5-biphosphate; R₁ or s, inhibitory or stimulatory receptor.
is also dependent on Ca$^{2+}$ and phospholipids for its activation (reviewed by Nishizuka, 1986). DAG dramatically increases the affinity of this enzyme for Ca$^{2+}$ and, thereby, renders it fully active even without a net increase in Ca$^{2+}$ concentration. Thus, the enzyme can be activated by synergetic action of a increase in Ca$^{2+}$ concentration and the formation of DAG.

3.3 Tyrosine-specific protein kinases

The protein kinases with specificity for tyrosine residues can roughly be divided into two groups on the basis of the absence or presence of an inherent growth factor receptor function. Generally, kinases of the latter group consist of a ligand binding extracellular domain which, through a transmembrane domain, is linked to an intracellular domain carrying protein kinase activity and protruding in the cytoplasm (see figure 1). To this receptor group belong e.g. the insuline receptor (Ullrich et al., 1985), the receptor for insuline-like growth factor I (Ullrich et al., 1986) and the PDGF receptor (Yarden et al., 1986). Recently, the v-fms, v-erbB and v-ros viral oncogene encoded proteins appeared to be modified or truncated forms of such growth factor receptors. The proto-oncogenes c-fms and c-erbB encode the receptor for the colony stimulating factor-1 (CSF-1) (Sherr et al., 1985) and the epidermal growth factor (EGF) receptor (Downward et al., 1984) respectively, whereas the growth factor binding to the receptor encoded by c-ros-1 (Matsushine et al., 1986) still has to be identified.

Tyrosine-specific protein kinases without a known growth factor receptor function include proto-oncogenes encoded proteins such as c-fes/fps, c-abl, c-yes and c-src and they are mostly found in the cytoplasm (see figure 1). In general, little is known about the mode of activation of this type of tyrosine-specific protein kinases.

3.4 Network of signal pathways

In the previous sections on protein kinases, little attention was paid to the phosphate acceptor proteins, phosphorylated by these kinases and the ultimate cellular responses. It would be beyond the scope of this brief survey on protein kinases to give an extensive overview of phosphate acceptor proteins and the subsequent cellular responses. Especially our knowledge on the action of serine/threonine-specific protein kinases (e.g. protein kinase A, protein kinase C and Ca$^{2+}$/CaM-dependent protein kinase)
expanded the last decade extensively. Control of cellular activity (e.g. muscle contraction, glycogen metabolism) by several different signal pathways involving protein kinases and their interaction upon external stimuli is described in detail in reviews by Adelstein and Eisenberg (1980) and Cohen (1982).

The serine/threonine-specific protein kinases are of particular interest for short-term responses on external signals, by virtue of the second messengers regulating their activity. Upon stimulation of signal pathways, the levels of these second messengers (e.g. cAMP, cGMP, Ca$^{2+}$ and DAG) increase rapidly and fall almost immediately upon termination of the stimulus by high turn-over. Interaction between the different signal pathways occurs at different levels. By phosphorylation of different residues of a single protein e.g. an enzyme by two different protein kinases, these kinases may act synergistically or asynergistically on the activity and/or affinity of the enzyme. Proteins with binding affinity for certain substrates and/or enzymatic activity belonging to a particular signal pathway can be regulated by phosphorylation of these proteins by action of protein kinases of another signal pathway. For example, protein kinase C regulates adenylate cyclase activity, whereas protein kinase A regulates the affinity of calmodulin for Ca$^{2+}$. Regulation of turn-over of second messengers and induction or down regulation of receptors contributes also to the interaction between signal pathways. In general, the interactions between different signal pathways may be divided into two modes of control systems: bidirectional and monodirectional control systems (see review by Nishizuka, 1986). In a bidirectional control system a particular signal pathway is counteracted by the activity of another signal pathway, whereas in monodirectional control systems a signal pathway may potentiate another signal pathway.

Tyrosine-specific protein kinases, especially the growth factor receptors among them, are thought to play a role particularly in long-term responses as cell differentiation and proliferation (Weinberg, 1985; Goussin et al., 1986). Naturally, the above mentioned modes of control systems for interactions between signal pathways involving serine/threonine-specific protein kinases are also important for interaction with or within signal pathways involving tyrosine-specific protein kinases. The activity of growth factor receptors is subject to modulations by protein kinases. Protein kinase C phosphorylates the EGF receptor resulting in decrease in both its growth factor binding and tyrosine-specific protein kinase activities. Binding of EGF evokes autophosphorylation in vitro of the EGF recep-
tor and a concomitant increase of its kinase activity (reviewed by Hunter and Cooper, 1985). The same phenomenon is shown by the insulin receptor in vitro upon binding of insulin (reviewed by Gammeltoft and Van Obberghen, 1986). Autophosphorylation in vitro by tyrosine-specific protein kinases and its concomitant increase of kinase activity points to a mode of regulation of the kinase activity of tyrosine-specific protein kinases in vivo. Apart from this, the insulin receptor complex contains an unidentified serine/threonine-specific protein kinase non-covalently bound to the insulin receptor. Putatively, this serine/threonine-specific protein kinase relays the stimulus of insulin to a different signal pathway as does the tyrosine-specific protein kinase activity, suggesting some interaction between the protein kinases.

Of the tyrosine-specific protein kinases that are not a growth factor receptor, the best studied is the kinase encoded by the proto-oncogene c-src. The protein encoded by this proto-oncogene, pp60c-src, is in vivo phosphorylated at tyrosine 527, a residue not present in pp60v-src (its transforming homolog) (Cooper et al., 1986). While both pp60c-src and pp60v-src autophosphorylate tyrosine 416 in vitro, this residue is only phosphorylated in vivo in case of pp60v-src (Smart et al., 1981; Patschinsky et al., 1982). Analysis of protein kinase activity of mutant pp60c-src encoded by mutated c-src genes revealed that loss of the C-terminal phosphorylation site (tyrosine 527) increases the kinase activity of pp60c-src and even gives it oncogenic potential (Cartwright et al., 1987; Kmiecik and Shalloway, 1987; Piwnica-Worms et al., 1987). Thus, dephosphorylation at tyrosine 527 activates the c-src encoded kinase, whereas phosphorylation at tyrosine 416 also may be responsible for increased kinase activity (Piwnica-Worms et al., 1987).

In general, the function of proto-oncogenes encoding a tyrosine-specific protein kinase that are not growth factor receptors is poorly understood: it is not known how their activity is primarily regulated. Nor are the identity and function of their phosphate acceptor proteins known, except that some of the tyrosine-specific protein kinases may be their own substrates. The search and characterization of cellular substrates of the viral transforming tyrosine-specific protein kinases and growth factor receptors (EGF and PDGF receptor) was focused on two prominent proteins among many candidates (Hunter and Cooper, 1985). These proteins, P35 and P36, also referred to as calpain II and calpain I respectively, are postulated to function in linking membrane phospholipids and proteins of
the cytoskeleton as actin and spectrin. This conclusion is based upon the 
$\text{Ca}^{2+}$-dependent association of these proteins with the plasma membrane, 
phospholipid vesicles and actin and spectrin (Glenny, 1986a, 1986b). Recent­
ly, calpactin I and calpactin II appeared to be identical to the pro­
teins lipocortin II and lipocortin I, respectively (Huang et al., 1986, 
Saris et al., 1986). These lipocortins inhibit phospholipase $A_2$ activity 
(Pepinski et al., 1986; Huang et al., 1986). Thus binding of EGF to its re­
ceptor modifies eventually the breakdown of phospholipids.

A matter not addressed to yet is the role of phosphatases, counteracting 
the activity of the protein kinases. Kinases and phosphatases do not exhi­
bbit a one-to-one correspondence in their reactions. For example residues at 
different positions may be phosphorylated by a single kinase, yet dephos­
phorylated by different individual phosphatases. In general, protein ki­
nases are more or less strictly tuned to amino acid sequence surrounding 
the residue to be phosphorylated, whereas phosphatases recognize higher 
order structures (reviewed by Sparks and Brautigon, 1986). Thus, commu­
iceation within cells will prove to be far more complex than outlined here for 
signal pathways involving protein kinases.

4. Outline of the investigations described in this thesis.

This thesis describes the genetic organization of the human and feline 
c-fes/fps proto-oncogenes and their upstream sequences.

Chapter 2 presents the nucleotide sequence analysis of a 13 kbp EcoRI 
restriction fragment containing the entire human v-fes/fps cellular homolog 
[moleculary cloned by Groffen et al. (1982)]. Comparison with v-fes/fps 
(Hampe et al., 1982; Shibuya and Hanafusa, 1982) and chicken c-fes/fps 
(Huang et al., 1985) revealed almost completely the distribution of exons 
and introns of the human gene. The sequence data made it possible to 
predict the primary translation product of the human gene.

However, the localization of the first exon(s) of the human gene 
remained unclear. Based upon the assumption that exon sequences and regula­
tory sequences as promoters of proto-oncogenes are conserved between 
species to a higher extend than intron sequences, human sequences upstream 
of known c-fes/fps exon sequences were compared with homologous feline se­
quences [moleculary cloned by Verbeek et al. (1985)]. These analyses in­
cluded Southern blot hybridization techniques, heteroduplex analysis and 
sequence analysis. These studies led to the discovery of the 3'-end of the
putative first c-fes/fps exon. Together with the use of Northern blot analysis and cDNA cloning they revealed the existence of a different transcription unit immediately upstream of known c-fes/fps proto-oncogene sequences, which was called fur (for fes/fps upstream region). These results are described in chapter 3.

In chapter 4, further studies on the human fur gene are presented. The nucleotide sequence analyses of a cloned fur cDNA and fur genomic sequences are described. Despite the fact that the fur cDNA contained only 3.1 kb of 3'-end sequences of the 4.5 kb fur mRNA, it was possible to predict a part of the putative fur translation product, furin. The fur gene probably encodes a transmembrane protein with a receptor function.

Chapter 5 describes the nucleotide sequence analysis of the feline c-fes/fps proto-oncogene and its comparison with the human gene, especially with respect to 5'-end sequences in order to localize the putative promoter region. As the feline sarcoma viruses transduced their v-fes/fps sequences from the feline proto-oncogene, analysis of feline sequences could provide insight in the recombination event resulting in the transduction. Based upon comparison of FeLV, v-fes/fps and feline c-fes/fps sequences a model is proposed for the genesis of the retroviral v-fes/fps oncogenes.
REFERENCES


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

EMBO J. 4, 2897-2903, 1985
The structure of the human \textit{c-fes/fps} proto-oncogene

Anton J.M. Roebroek, Jack A. Schalken, 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, De Toonderlaan 21, 6555 RE, Nijmegen, and Department of Tumour Surgery, St Radboud University 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 \textit{c-fes/fps} cellular homolog. The \textit{v-fes/fps} homologous sequences were dispersed over 11 kb 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 \textit{v-fes} of the Gardner-Arnstein (GA) and Snyder-Theilen (ST) strains of feline sarcoma virus and \textit{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 \textit{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 NC92, the recently described human \textit{c-fes/fps} product. The topography of human \textit{c-fes/fps} appeared to resemble that of chicken \textit{c-fps}.

\textbf{Key words} human \textit{c-fes/fps} proto-oncogene/nucleotide sequence

\section*{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 H71 (Hardy et al., 1981)] have captured sequences from the feline \textit{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 L6 virus (reviewed by Bishop and Varmus, 1982, Bishop, 1983)] have acquired similar sequences from the avian counterpart \textit{c-fps} (Shibuya et al., 1980, Shibuya and Hanafusa, 1982, Groffen et al., 1983). The translational products of these viral transforming genes are polypeptides which possess tyrosine-specific protein kinase activity in vitro and are capable of autophosphorylation as well as phosphorylation of exogenous protein substrates (Russetti 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 polypeptides (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 \textit{c-fes/fps} proto-oncogene has been identified in a number of species. In chicken myeloblasts, a 98 000 mol wt protein (NC92) (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 (NC92). NC92 was also shown to exhibit associated protein kinase activity (Mathey-Prevot et al., 1982). The murine and human \textit{c-fes/fps} proto-oncogene products have recently been identified in myeloid cells as NC92 and these proteins were found to be <g-cAMP-dependent 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 \textit{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 \textit{fes/fps} proto-oncogene, we have determined the complete nucleotide sequence of a molecular clone of human \textit{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 \textit{c-fes/fps} gene with those deduced from sequence data of the \textit{fes} gene of GA-FeSV and ST-FeSV (Hampe et al., 1982), the \textit{fps} gene of FSV (Shibuya and Hanafusa, 1982) and the chicken \textit{c-fps} proto-oncogene (Huang et al., 1985). In addition, the phosphokinase domain of human \textit{c-fes/fps} was compared with those of other members of the tyrosine kinase multigene family.

\section*{Results and discussion}

\textbf{Topography and nucleotide sequence of human \textit{c-fes/fps}}

A human DNA fragment of ~13 kbp, which was shown by Southern blot analysis to contain the entire \textit{v-fes/fps} cellular homolog (Groffen et al., 1982, Franchini et al., 1982, Trus et al., 1983, 1985) was used as a donor of 7-phosphate (Feldman et al., 1985). In addition, the \textit{fes/fps} homologous region A schematic restriction map of the 13 kbp EcoRI DNA fragment is presented. Black boxes represent the \textit{human fes/fps homologous segments}. These putative exons are numbered similarly to the chicken locus (Huang et al., 1985)

\begin{figure}
\centering
\includegraphics[width=\textwidth]{map.png}
\caption{Topography of the human \textit{fes/fps} homologous region. A schematic restriction map of the 13 kbp EcoRI DNA fragment is presented. Black boxes represent the human \textit{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 \textit{v-fes} ST \textit{v-fes} and FSV \textit{v-fps} is indicated by lines. B BamHI, E EcoRI, H HindIII, K KpnI, P PstI, P* cluster of Pst site. S Sfll, X XbaI, Xh XhoI.}
\end{figure}
A.J.M. Roobroeck et al.

GCACCCAGCAGCCCCTCACCAAGAAGAGTGGTGTTGTCCTGCACAGGGCTGTGCCCAAGLgagcctgcaecLageuLggcccatgrcacctgtggcagggcttggggagtgtgggtca
gggctcacgccccctcagaatggaggctgctgaccccgggtccctgccctgcagAACCTGTACCGKTGGAAtKGGMGGCTTTCCTAGCATTCCTTTCCTCATCGACCACCTACTGA
GAAGCCCCTGCATGAGCAGCTCTGGTACCACGGG(KXATCCCGAGGGCAGAGGTtXXTCAGCTTCTCCTO
atcgccccagtgtBctaaagggaccagcaacctcgactattccatggctrtccctgcttcaggagcggttgggggcctgtggcctggaggaggaggcaccagettggtttggggtcttcc
actgggatgLcctagagaggaggctctgcccaggctgcttgtattgggaagttcctctcttccctgggattccaggctgcagatlcctccccagaccctgcccctgtgacccctccctttcc
GAAGAGGAGAACACCCACCCCCGGGAGCGgtgagtgggcccctgcctgcagcagcctcctgggcctccctccctcctacctaccctaactgctgctggcLacccgccgcagaccgagccc
agacaaggatagcagttcatttatttatttaittatttatttatttatttalttagagatBtagtçtçattçtttçgççagBçtggagtgçaBtggçgçgatçttBgçtça
agggatctttcgtgttagtggagtgaggatgtaggagcactaagagccatggagaaaaataaaecaagagaagtggatcgggacctgegagcacggaggcaagggaggaggtgacagtt
GGCCCCTCAGCOttTGAGCCTTCTCATCCGreAACGGCAGClCeTTCGC«^CCCATCAGTCAGgtgggtctctatgggactctggtggglgctggcgtatctgecttctccttcctctcctgBgggccctctggggcagtggctgg^gatciggedggteaaigcfBgga
GAGGGCATGAGAAAGTGGA1t»XCCAGCGGGTCAAGACTCACAGGGAGTATGCAGGACTGCTTCACCACATGTCCCTGCAWACAGTGGGGGCCAGAGC(^^
tcctcccctggtggggcagcaggatgtcatgtgccatcagatgecatcttttctggaggtctctctgcccctggtcctgggcaggccctttct.cccctgctgctctccctttccccetce
ggggggagagagagacccccggctgcccccacggcctcttcaacaaggtggttaagtgactcctcctcgatcctcccttgcccagCTCCCTCCACCGCTGCAGCTCATTCCGGAGGTGCA
tgcctgggcttcccttcccagctctgcccagcgtgagcctgggccagtccaatgcccactccaggggcctgtggaLggctctgcatgccactccatggttgtaagBgctgagggcatata
CCOGGCAAGCACACCCACGCTCGAGATt^TTAAGAGCCACATCTCAGGUTtrrTOMMXCCAAGTT^tgaaattgtgggggtctttgttaaaagattgctaggaatttccaggtggcaataatggaBaatgaaaccaagcacagggcccttetacatgLggagccccgtgtgactgcacaggcegtg
tgcttttctgtgcctgggcaaagtgctgggagtgtaaggatgagtgaccggtcacgtgectggg^ga
gctgccattgtgcccccctccctgccLcccecatctgtgcLgtataBTTCTGGGCTCAGATCKCAGWAAACTGAGGGCCTGACXXMTTGCTGCCCCAGCACGCAGAGGATCTGAACTCAG
tgaaattgtgggggtctttgttaaaagattgctaggaatttccaggtggcaataatggaBaatgaaaccaagcacagggcccttetacatgLggagccccgtgtgactgcacaggcegtg
tgcttttctgtgcctgggcaaagtgctgggagtgtaaggatgagtgaccggtcacgtgectggg^ga
ggcattgtgcccccctccctgccLcccecatctgtgcLgtataBTTCTGGGCTCAGATCKCAGWAAACTGAGGGCCTGACXXMTTGCTGCCCCAGCACGCAGAGGATCTGAACTCAG
tgaaattgtgggggtctttgttaaaagattgctaggaatttccaggtggcaataatggaBaatgaaaccaagcacagggcccttetacatgLggagccccgtgtgactgcacaggcegtg
tgcttttctgtgcctgggcaaagtgctgggagtgtaaggatgagtgaccggtcacgtgectggg^ga

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

---

**Fig. 2.** Nucleotide sequence of the human v-fes/fps homologous region. Sequence data are presented from the KpnI site just upstream of the putative exon 2 to the PstI site downstream of exon 19. Segments of the sequence that are homologous to v-fes/fps as well as the 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 utilized as alternative splice junction for exon 19 is underlined.

---

*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 segments over the DNA region of ~11 kb. It should be noted that in the comparative analysis with the two v-fps 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 kb 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,
homologous segments were flanked by the AG and GT splice sites between the human and chicken fes/fps coding sequences. The virus PRC1I which as a result of a deletion lacks exon 2 at -50 nucleotides from the 3' end of exon 2. In feline fps and are missing in the two feline viral fes genes were also a potential splice site at a position where sequence homology between human and chicken fes sequences corresponded to the end of exon 2 to somewhere in the middle of exon 2. The only other difference in the middle of exon 9 of fes/fps gene was presented As can be seen all the 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 was observed. The same was also true for chicken fes/fps. Present in exon 2 and the first 57 bp of exon 2 of chicken fes remained unaccounted for. As indicated in Figure 1, homologous genetic segments of some of the putative exons (exons 2-9) were completely present in FSV v fps and 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 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 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 cosmids 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 PRC1I 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 was observed. The same was also true for chicken fes/fps. Present in exon 2 and the first 57 bp of exon 2 of chicken fes remained unaccounted for. As indicated in Figure 1, homologous genetic segments of some of the putative exons (exons 2-9) were completely present in FSV v fps and 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 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 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 cosmids 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 PRC1I 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 was observed. The same was also true for chicken fes/fps. Present in exon 2 and the first 57 bp of exon 2 of chicken fes remained unaccounted for. As indicated in Figure 1, homologous genetic segments of some of the putative exons (exons 2-9) were completely present in FSV v fps and 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 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 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 cosmids 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 PRC1I 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).
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 EcoRI 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 non-coding 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-fes (Bojály et al., 1983), chicken c-fes (Takeya and Hanafusa, 1983), v-fps (Hampe et al., 1984), human epidermal growth factor receptor gene (HFR) (Ullrich et al., 1984) and human insulin receptor gene (HIR) (Ullrich et al., 1988). Boxes, common residues among at least four of the six proteins, asterisk (*), lysine residue specifying the ATP binding site, open circle (.), possible phosphotyrosine site.

**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 [1] direct 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, Levison et al 1981, Ween master 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 do-
mam (Barker and Dayhoff 1982) that is a growth factor receptor (Yamada et al 1985) and the insulin receptor kinase (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 et al., 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 (Groth et al. 1984). A 13 kbp EcoRl DNA fragment which contains all v-fes/fps homologous genetic sequences and a 9.2 kbp HindlII 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 (Schaal et al. 1984). The cDNA inserts were sequenced using the dideoxy sequencing method as described by Sanger et al. 1977. All parts of the reported DNA sequence were obtained by both strands of the cloned DNA. The gel readings were recorded and compared using the Staden programs (Staden 1982)

Acknowledgements

We thank J R Stephenson and J Groitsch for the human v-fes/fps cellular homolog. We acknowledge the assistance of J Leuvenich with the computer analysis and of E Oostensen and H Thijssen for their help with the DNA sequencing. This work was supported by the Netherlands Cancer Foundation Contract Nos NL KC 80.011 and by the Foundation for Medical Research FNGO which is subsidised by the Netherlands Organisation for Advancement of Pure Research (ZWO).

References

Barker M, Beecken K and Devare S G (1980) Proc Natl Acad Set USA 77 5158-5162


Barker W C and Dayhoff M O (1982) Proc Natl Acad Set USA 79 2836-2839


Botel M, Srinivasan M and Hanafusa H (1982) Proc Natl Acad Set USA 82 2379-2383


Grolle J, Heyrman T, Grosveld F, van de Ven W J M and Stephen-

Verbeek et al. 1985, Science (Wash.) 216 1136-1138


Levison A D, Courtendride S A and Bishop J M (1984) Proc Natl Acad Set USA 78 1624-1628


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


Sibitz D E, deKornwall E, Verna M and Clene M O (1976) Science (Wash.) 214 256-262

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


Received on 19 June 1985; revised on 29 July 1985
Characterization of human c-fes/fps reveals a new transcription unit (fur) in the immediately upstream region of the proto-oncogene.

Characterization of human c-fes/fps reveals a new transcription unit (fur) in the immediately upstream region of the proto-oncogene

A J M Roebroek¹, J A Schalken², M J G Bussemakers¹, H van Heerikhuizen¹, C Onnekink¹, H M Debruyne², J P J Bloemers¹ and W J M Van de Ven¹

¹Molecular Oncology Section, Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands
²Department of Urology, St Radboud University Hospital, Nijmegen, The Netherlands

Abstract

Comparison of nucleotide sequence data of the 5' region of a fes/fps viral oncogene with those of the v-fes/v-fps homologous regions of man and cat revealed the position of the 3' portion of an as yet unidentified c-fes/fps exon. Comparative Southern blot and heteroduplex analysis of human and feline DNA immediately upstream of the v-fes/v-fps homologous regions showed extensive but discontinuous homology over a 9 kbp DNA stretch, which we have designated as fur. Northern blot analysis of mRNA from KG-1 myeloid cells with fes/fps- or fur-specific probes revealed a 3.0 kb fes/fps and a 4.5 kb fur transcript. Analysis of a number of tissues of an adult Wistar Lewis rat for the presence of fur transcripts revealed its differential expression pattern. An 0.95 kb fes/fps related and a 2.2 kb fur related cDNA recombinant clone were isolated from an oligo(dT)-primed KG-1 cDNA library. Comparative nucleotide sequence analysis of the fes/fps cDNA and its human genomic counterpart indicated that the cDNA contained genetic sequences that were identical to and colinear with exon 15-19 and, furthermore, that the poly(A) addition signal near the 3' end of exon 19 was functional. Similar analysis of the 2.2 kb fur cDNA indicated that the poly(A) addition signal of the fur transcript was in close proximity of the newly discovered fes/fps exon. The region in between contained a CATT sequence but no 'TATA' box. The fur transcript was characterized by a long noncoding region at its 3' end.

Introduction

Proto oncogenes can be defined as cellular genes which under normal physiological conditions are thought to exhibit proliferative and developmental functions (13, 26, 27) but which exhibit transforming activity upon transduction by a retrovirus (3, 4, 10). From data gathered to date, the consensus has emerged that retroviral mobilization of the transforming potential of proto-oncogenes involves alterations in regulation of their expression, alteration of their primary gene product, or both (5).

Transduced c-fes/fps proto-oncogene sequences of cat and chicken are present in the genomes of a number of independent feline and avian sarcoma virus isolates (4, 12, 18, 32). Mobilization of its transforming potential was in all these cases characterized by fusion of c-fes/fps proto-oncogene sequences to retroviral gag gene sequences.

The function of the c-fes/fps proto-oncogene is not yet known but the proto oncogene might be of particular interest in hematopoietic proliferation since expression seems mainly restricted to hematopoietic tissue. Expression of the c-fes/fps proto oncogene has been found in normal and leukemic myeloid cells and cell lines of a number of species (2, 9, 22, 23, 31). As its main translational product, a tyrosine specific protein kinase of about 92000 mol wt has been described. The c-fes/fps mRNA species that have been reported varied in size between 2.6 and 3.2 kb (19, 29, 31). However,
the genetic organization of the fes fps proto-
oncogene is not yet completely defined (11, 14, 19, 28, 35, 37). In previous studies, we have described the isolation and genetic organization of fes fps proto-oncogene sequences of man (28) and cat (17) and their malignant potential was established by fusing them to gag gene sequences of feline leukemia virus origin (33, 37). Recently, we have derived the topographical distribution of almost all human fes fps homologous sequences from the complete nucleotide sequence of a human 13 kbp EcoRI/EcoRI restriction fragment (28). A sequence homologous to 12 bp at the 5' end of the v fes/vps homologous sequences of the Gardner-Arnstein strain of feline sarcoma virus (GASV) (16) and 140 bp at the 5' end of v fps of Fujinami sarcoma virus (FSV) (19) were still missing. In the present study, we have identified the position of the missing 12 bp in the human and feline proto-oncogene. We also present characteristics of the 5' and 3' end of the human c fes/cps transcription unit. Furthermore, we describe the identification and characterization of a new transcription unit, designated fur, in the immediately upstream region of the v fes/vps homologous sequences.

Materials and methods

Cell line and cosmid clones

The permanent KG-1 human cell line, which was established from bone marrow cells of a patient with erythroleukemia (21), was kindly provided by Dr H P Koeffler. Isolation of the human v fes/vps cellular homolog was described previously (14). Subclones in pSP64 of a 13 kbp EcoRI/EcoRI DNA fragment, which contains all v fes/vps homologous sequences, and a 9.2 kbp Hpal/EcoRI DNA fragment flanking the former at its 5' end were described (28). The feline v fes/vps cellular homolog was isolated from a feline cosmid library (17).

DNA probes and hybridization

Probes were isolated and labeled as described before (36). Hybridization experiments on nitrocellulose membranes were performed as described previously (36), and for hybridizations on nylon membranes (Hybond N, Amersham) the method of Church and Gilbert (7) was used. The nylon membranes were dehybridized by incubation in 5 mM Tris-HCl (pH 8.0), 2 mM EDTA and 0.1 × Denhardt's solution (1 × Denhardt's solution contains 0.02% (w/v) bovine serum albumin, 0.02% (w/v) polyvinylpyrrolidone and 0.02% (w/v) lactalbumin) at 65°C for 2 h. If necessary, the procedure was repeated for up to six times. Dehybridization was always checked by autoradiography. Blots could be used for up to six times without significant loss of signal.

Heteroduplex analysis

Electron microscopic analysis of duplexes between the 9.2 kbp human Hpal/EcoRI DNA fragment and the 14 kbp feline EcoRI/EcoRI DNA fragment was performed as described by Davis et al (8). Grids were examined in a Zeiss EM109 electron microscope at 40 kV. Electron micrographs were taken at a magnification of 6000 ×. Measurements were made on 25 heteroduplex structures.

mRNA isolation and Northern blot analysis

Total cellular RNA was isolated using the lithium-urea procedure described by Auffray and Rougeon (1). Ten μg of oligo(dT)-cellulose purified mRNA was glyoxalated and size fractionated on 10% agarose gels and transferred to Hybond-N (procedure as recommended by Amersham)

Construction and screening of cDNA library

An oligo(dT) primed cDNA library was constructed in λgt11 as described by Huynh et al (20). The cDNA reaction was modified according to Gubler and Hoffman (15). About 250 000 plaques obtained upon infection of λ col E.coli Y1090 (20) were screened as described by Hanahan and Meselson (17).

DNA sequence analysis

DNA fragments were inserted into the polylinker region of M13mp8-11 (24). All of the DNA sequences were determined by the dideoxy-sequencing method (30). 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 (34).

Results and discussion

Comparison of human and feline DNA immediately upstream of the v-fes/fps homologous regions

The topography of most of the human v-fes/fps homologous genetic sequences has recently been described. Upon nucleotide sequence analysis of a 13 kbp EcoRI/EcoRI DNA fragment, they appeared to be split up in at least 19 exon segments and, from these, exons 2–19 were characterized (28). However, genetic sequences at the 5' end of v-fps of FSV or v-fes of GA-FeSV remained unaccounted for. In KG-1 cells, which were shown to exhibit a high level of expression of the fes/fps proto-oncogene (9), no such transcripts could be detected in a Northern blot analysis with the 1 kbp DNA region upstream of exon 2 as a probe (data not shown). In an attempt to identify the missing sequences in the human genome and to investigate whether the region immediately upstream of the v-fes/fps homologous sequences contained additional sequences that were not represented in any of the known v-fes/fps isolates but did belong to the c-

![Diagram](image_url)

Fig. 1. Comparison of human and feline DNA upstream of the v-fes/fps homologous region. (A) A schematic representation of the v-fes/fps cellular homolog and its upstream region in man and cat. The v-fes/fps homologous regions indicated in the figure are described elsewhere in more detail (28, 37). The upstream regions are depicted in more detail in the middle of this part of the figure. The heavy bars represent the DNA regions in which man and cat share common genetic sequences. B. BamHI; E. EcoRI; HpaI; K, KpnI; P. PstI; Xb, XhoI; Xh, Xhol. (B) Electron micrograph of a heteroduplex between the 9.2 kbp human HpaI/EcoRI and 14 kbp feline EcoRI/EcoRI DNA fragment. (C) Interpretive drawing of heteroduplex shown in part B. (D) Schematic representation of the results of measurements on 25 heteroduplexes. The loops that are shown were observed in all heteroduplexes. Arrows indicate the positions and frequencies of loops found in a part of the duplex structures.
fes/fps protooncogene sequence compared with human DNA upstream of the v-fes/fps homologous region (9.2 kbp Hpal/EcoRI DNA fragment) with a 14 kbp EcoRI/EcoRI DNA fragment representing the corresponding feline region by Southern blot analysis. The 9.2 kbp Hpal/EcoRI human DNA fragment (Fig. 1A) was digested with restriction endonuclease EcoRI, KpnI, PstI, XbaI, and XhoI of various combinations of these in double digestions, and hybridizations were performed under conditions of high stringency with the 14 kbp EcoRI/EcoRI feline DNA fragment (Fig 1A) as a molecular probe. It appeared that an extensive and strong homology existed between the two fragments (data not shown). In a reciprocal experiment, similar restriction endonuclease digestions of the feline 14 kbp EcoRI/EcoRI DNA fragment and hybridization analysis with the human 9.2 kbp Hpal/EcoRI DNA fragment as a probe revealed similar results (data not shown). As is schematically indicated in Figure 1A, homology between the two fragments started almost immediately upstream of the v-fes/fps homologous regions and extended over a DNA region of about 9 kbp. We propose to designate this fes/fps upstream region as fur.

To further characterize the fur region in man and cat, heteroduplex analysis was performed with the human 9.2 kbp Hpal/EcoRI and the feline 14 kbp EcoRI/EcoRI DNA fragments (Fig 1B and 1C). This experiment confirmed the observations of the Southern blot experiments and, furthermore, revealed that the homologous genetic sequences were interspersed with at least four non-homologous regions. In a number of the 25 heteroduplex structures that were analyzed, three regions of apparent reduced homology were observed. Their positions and frequencies are indicated by arrows in Figure 1D.

Human DNA upstream of the v-fes/fps homologous region is transcriptionally active

To establish whether DNA immediately upstream of the v-fes/fps homologous region was a part of the protooncogene, we studied c-fes/fps transcription in the human KG 1 cell line, which was derived from a patient suffering from acute myelogenous leukemia (21). RNA was isolated from these cells poly(A) selected, size fractionated on a 1% agarose gel and blotted onto Hybond N (Amersham) as described under Materials and Methods. Expression of c-fes/fps protooncogene sequences could readily be detected with the S

Analysis of fur expression was also studied in a number of tissues of an adult Wistar Lewis rat (Fig 2B). These tissues included brain, heart muscle, kidney, lung, testis, and thymus. High levels of the 4.5 kb fur transcript were found in brain, kidney, and thymus. No detectable levels of fur transcripts were present in heart muscle, lung, and testis. These data indicate a differential expression pattern of the fur transcription unit. To get more insight into the genetic organization of the fur and fes/fps transcription units, nucleotide sequence data of the transcripts and their genomic counterparts should be compared.

Molecular cloning and DNA sequence analysis of human c-fes/fps and fur-related cDNA

To isolate c-fes/fps and fur-related cDNA, we constructed an oligo(d1) primed cDNA library in λgt11 using KG 1 mRNA as template. Upon screening of 25 x 10^6 plaques with the v-fes specific S probe (II), we isolated a cDNA clone containing a c-fes/fps specific insert of 0.95 kbp. The same library was also screened with a human fur specific probe (probe 3, see lower part of Fig 42.
Fig. 2. (A) Expression of the human fes/fps proto-oncogene and its immediately upstream region in KG-1 cells. Poly(A)-selected mRNA isolated from KG-1 cells was glyoxylated, size fractionated by electrophoresis using a 1% agarose gel, immobilized on Hybond-N and subjected to hybridization analysis as described under Materials and methods. DNA probes included v-fes-S, (lane 1), probe 1 (lane 2), probe 2 (lane 3), and probe 3 (lane 4). Probe 1, 2, and 3 are defined in the lower part of the figure. As molecular weight markers λ DNA digested with restriction endonuclease HindIII was used. B, BamHI; Hp, HpaI; P, PstI; Xh, XhoI. (B) Expression of the c-fes/fps upstream region in a number of rat tissues. Poly(A)-selected mRNA was isolated from thymus (lane 1), lung (lane 2), heart muscle (lane 3), testis (lane 4), kidney (lane 5), or brain (lane 6) of an adult Wistar Lewis rat and subjected to Northern blot analysis. As molecular probe, a mixture of probe 2 and 3 was used. As molecular weight markers λDNA digested with restriction endonuclease HindIII was used.

Fig. 3. Nucleotide sequence of the 0.95 kbp fes/fps-related cDNA and its predicted amino acid sequence in the conventional one letter code. The consensus sequence for the polyadenylation signal is underlined. The poly(A)-stretch of the cDNA clone is indicated by (A). Arrows indicate the positions of the three possible poly(A) addition sites. Open box (□), possible phosphoacceptor tyrosine. Asterisk (*), stop codon.
Fig 4 Nucleotide sequence of the 5' portion of the human fes/lps cellular homolog and its immediately upstream region. The nucleotide sequence of the 4 16 kbp BamHI/BglII human DNA fragment includes fur exon 2 and c. fes/lps exons 1 and 2. Established exon sequences are represented by capitals. Established exon borders are indicated by arrows. Relevant nucleotide sequence data of the genomic DNA are given between square brackets in positions below the corresponding human genomic sequence data. The consensus sequence for the polyadenylation signal of the fur transcription unit is underlined. A potential CATT box ( ) is represented in a box Triangle ( ) putative start codon (28)
Previously, we have determined the topography of the exon sequences in the human c-fes/fps proto-oncogene on the basis of homology with v-fes/fps (28). The 3' border was tentatively placed at a position about 200 bp downstream from a stop codon in exon 19 where a potential polyadenylation signal was present. The availability of nucleotide sequence data of the 0.95 kbp human fes/fps-related cDNA (Fig 3) and previously published sequence data of genomic fes/fps DNA (28) enabled a comparative analysis. The 5' end of the cDNA clone appeared to start 5 nucleotides downstream of the 5' end of exon 15 and at the 3' end, the cDNA clone contained a poly(A) stretch representing the poly(A) tail. Therefore, the 3' border of the fes/fps transcript in KG-1 cells could be determined rather precisely. The data confirmed the previous suggestions of Roebroek et al. (28) that the poly(A) addition signal at a position about 200 bp downstream from the stop codon in exon 19 was functional. Furthermore, it could be established that the poly(A) addition site was 17, 18 or 19 bp downstream from the polyadenylation signal. There were no differences found between the nucleotide sequence of the human c-fes/fps related cDNA described here and the corresponding genomic region of the human fes/fps proto-oncogene. The exon/intron arrangement from exon 15 to 19, as proposed by Roebroek et al. (28), also proved to be correct.

Although the c-fes/fps-related cDNA was only one third of the observed mRNA length, some additional characteristics of the c-fes/fps transcription unit could be derived from analysis of the fes/fps fragment. Comparative analysis of nucleotide sequence data of the 2.2 kbp fes/fps related cDNA and its corresponding and flanking region in genomic DNA revealed that the nucleotide sequence of the cDNA insert was identical to that of the genomic DNA fragment (Fig 4). These sequences are indicated by capitals in Figure 4. At its 3' end, the fes/fps cDNA had a poly(A)-stretch marking the end of the fes/fps gene. At precisely 21 bp upstream of the poly(A) addition site, there was an ATTAAA consensus sequence of a poly(A) addition signal (38). Comparative sequence analysis further indicated that the 2.2 kbp fes/fps cDNA was contained in a single exon, which we have designated 'Z'. To predict a potential fes/fps translation product, we compared the three reading frames in exon Z. In all three reading frames, numerous stop codons were present. The first stop codon in each of the three different reading frames was encountered within the first 420 bp from the 5' end of the cDNA clone. We therefore concluded that the fes/fps mRNA is characterized by a long non-coding region at its 3' end.

The 3' end of the fes/fps transcription unit was located in close proximity of the v-fes/fps homologous region in man and cat. However, it should be noted that in front of exon 2 of c-fes/fps, 12 bp at the 5' end of v-fes, 12 bp at the 5' end of v-fes of Ga FeSV (16) and 140 bp at the 5' end of v-fps of FSV (19) were unaccounted for (28). In case of the 140 bp of v-fps, we realized that in human c-fes/fps possibly no sequences homologous to these 140 bp could be identified because of divergence of non-coding exon sequences. In case of the 12 bp of v-fes, it could not be excluded that homologous sequences were not recognized because comparison was limited to a DNA stretch of only 12 nucleotides. In order to resolve this matter, we have sequenced the feline c-fes/fps proto-oncogene region that contained exon 2 and about 1 kbp of upstream flanking sequences (data not shown). These sequence data supported the boundary assignment of human exon 2 as suggested by Roebroek et al. (28). The homology between the genetic sequences in exon 2 of man and cat was about 85% upstream of exon 2, the homology was about 65%. In this region in the feline DNA, a stretch of 12 bp was found, which was identical to the 12 bp at the 5' end of the feline v-fes oncogene. In the human DNA, an almost identical stretch was found at a similar position. From these bases, 11 were shared by the viral oncogene and the feline proto-oncogene. An insertion of one nucleotide was observed in the human DNA stretch. The 3' end of both the feline and the human DNA stretch (see Fig 4) was followed by a sequence characteristic for a splice donor site (25). These data suggested that these stretches overlapped the 3' region of a new human and feline c-fes/fps exon. The 3' boundary of exon 1 and the 5' end of exon 2 in human and feline DNA are compared in Figure 4. The 5' boundary of exon 1 remains to be established, but, based upon the size of the polyadenylated mRNA detected in KG-1 cells, one can roughly esti-
mate the first c-fes/fps exon as about 200 nucleotides.

The location of the 3' end of the fur transcription unit in relative close proximity of fes/fps exon 1 (see Fig. 5) raised also the question as to whether the promoter of the fes/fps transcription unit was located between fur exon Z and fes/fps exon 1. Searching for sequences typical for promoter regions revealed a sequence that is in good agreement with a CATT box (6), namely the GGCCATTCT sequence at position 3093-3101 in Figure 4. Limited sequence data available from feline c-fes/fps indicated that the GGCCATTCT sequence was also present at a similar position in the feline DNA. However, no sequences resembling the consensus sequence of a TATA box (6) could be identified. The potential CATT box would be in an appropriate position relative to the 5' end of the transcription unit to accommodate a c-fes/fps transcript of 3 kb as found in KG-1 cells. More data are required to locate the 5' end of the protooncogene and to identify the DNA sequences that control its expression. The observation that in man and cat the linkage of the fur and fes/fps transcription units has been conserved during evolution is interesting because it could have functional implications and, therefore, be instrumental in elucidating regulation of expression of c-fes/fps.

Acknowledgements

We acknowledge the excellent technical assistance of J. J. M. van Groningen. We thank G. C. M. Selten, A. M. W. van den Ouweland, and J. S. Verbeek for helpful discussions. This work was supported by the Netherlands Cancer Foundation, Contract No. NUKC Bioch. 80-1 and NUKC Bioch. 85-9, and by the Foundation for Medical Research FUNGO which is subsidized by the Netherlands Organization for the Advancement of Pure Research (ZWO).

References

1. Auffray, Ch & Rougeon, I., 1980 Eur J Biochem 107 303-314
Received 24 Jan 1986 and in revised form 25 Feb 1986
Evolutionary conserved close linkage of the c-fes/fps proto-oncogene and genetic sequences encoding a receptor-like protein.
Evolutionary conserved close linkage of the c-fes/fps proto-oncogene and genetic sequences encoding a receptor-like protein

Anton M Roebroek, Jack A Schalken1, Jack A M Leunissen Carla Omnekink, Henk P J Bloemers and Wim J M Van de Ven

Molecular Biol. Cell 1 (4) 1986

Introduction

In a recent study which was directed towards the elucidation of the genetic organization of the 5' end of the c-fes/fps proto-oncogene, we found that the genomic DNA immediately upstream of the human and feline v-fes/fps homologous regions showed extensive homology over a DNA stretch of ~9 kbp (Roebroek et al 1986). This 9 kbp DNA was designated fur (for fes/fps upstream region) and encoded a 4.5 kb mRNA in both species. From the data presented here we deduce that fur may encode a membrane-associated protein with a recognition function.

Key words: fur gene/c-fes/fps proto-oncogene/nucleotide sequence/receptor

Results

Topography of the fur transcription unit

In a recent study (Roebroek et al 1986) we found that the DNA region immediately upstream of the c-fes/fps proto-oncogene in man and cat shared homologous genetic sequences and that this region was transcriptionally active in human KG 1 myeloid cells. We reported that in human DNA the transcribed genetic sequences were dispersed over at least the entire 9.2 kbp Hpal/EcoRI DNA fragment that is depicted in Figure 1. To investigate whether additional human sequences further upstream would hybridize with topographically corresponding feline sequences, we have now used an EcoRI/EcoRI DNA fragment of ~20 kbp. This fragment was located immediately upstream of the 5' end of the 13 kbp EcoRI/EcoRI DNA fragment that contained the human v-fes/fps cellular homologue (Figure 1) (R Stephenson personal communication). Comparative Southern blot analysis with the insert of the previously described feline cosmid clone 1 (Verbusk et al 1985) indicated that homology between man and feline fes/fps upstream region extended ~1 kbp further upstream beyond the Hpal restriction endonuclease cleavage site (Figure 1) (data not shown). The observed divergence of human and feline genetic sequences upstream of the PstI cleavage site could mark the positions of the 5' borders of the fur transcription unit. Support for this assumption could be derived from the preliminary finding that within the 10 kbp fur region the 2.4 kbp Hpal/BamHI DNA fragment showed promoter ac...
tivity (70% of the activity of the promoter in the LTR of Rous sarcoma virus) in a chloramphenicol acetyl transferase assay (data not shown).

To characterize further the genetic organization of fur we constructed a cDNA library in λgt11 with poly(A) selected RNA from KG 1 cells. Using the 0.7-kbp Xhol/Xhol DNA fragment as a molecular probe (Figure 1) a 3.1-kbp fur specific cDNA was isolated. The nucleotide sequence of the 3.1-kbp cDNA clone was determined and from this it appeared that up to the poly(A) addition site the cDNA insert was 3078 bp long and included furthermore a short poly(A) stretch at its 3' end (Figure 2). When the 3.1-kbp cDNA clone was used as a molecular probe in Northern blot analysis, the previously described 4.5-kbp fur mRNA was detected as the only transcript (data not shown).

To elucidate the genomic organization of the sequences corresponding to the molecularly cloned fur specific cDNA, the nucleotide sequence of the genomic DNA regions that hybridized to the 3.1-kbp cDNA probe was also determined (Figure 2). Comparison of the cDNA and the genomic nucleotide sequences revealed that the 3.1-kbp transcribed sequences were distributed over eight exons within 5.4-kbp of genomic DNA (for a schematic representation see also Figure 1). We have labeled these exons from S to Z. All exons were flanked by splice junction consensus sequences. In the nucleotide sequence of the 3.1-kbp cDNA clone we looked for the presence of a large open reading frame. In a study on a previously described 2.2-kbp fur specific cDNA clone we found numerous stop codons in all three reading frames (Roebrock et al. 1986). However, in the 3.1-kbp cDNA clone we found an open reading frame of 1498 bp starting at the 5' end of the clone (Figure 3). The identification of the open reading frame in the transcribed genetic sequences in the fur region was important since it provided information about the possible function of its putative translation product.

Computer analysis of the putative fur translational product revealed receptor like characteristics.

The amino acid sequence was deduced from the nucleotide sequence of the 1498 bp open reading frame (Figure 3A). It should be emphasized that this sequence represents only a carboxy terminal portion of the putative fur translational product whose complete size and amino acid sequence remains still to be established. For reasons of convenience, we have dubbed the putative fur translational product furin. A computer program using the algorithm described by Kyte and Doolittle (1982) was used to display the hydrophy of the carboxy terminal portion of furin (Figure 3B). Such analysis revealed a highly hydrophobic domain (amino acid residues 418-450), indicating that it might function as a transmembrane domain. This putative transmembrane domain was located ~50 amino acid residues from the carboxy terminus of furin. Furthermore, screening of the PIR and PIGtrans databases showed that this domain was highly homologous to the transmembrane domain of class II HLA antigens (Schumning et al. 1984). The homology was 42% in the stretch extending from residues 422 to 445. Conservative amino acid substitutions not included. When these were taken into account, the homology was 67% (Figure 3C). Thus, alignment data supported the possibility that the amino acid stretch 422-445 might function as a transmembrane domain.

A further characteristic feature of furin was the presence of a cysteine rich region (extending from amino acid residue 261 to 410). Recent studies on the structure of receptors indicated that cysteine rich domains seem characteristic for certain types of receptors (Coussens et al. 1986). Screening of the PIR and PIGtrans databases with the deduced furin sequence revealed a strong homology between its cysteine rich region and that of the human insulin receptor (Ullrich et al. 1985) and those of the human epidermal growth factor receptor (Ullrich et al. 1984). The overall homology between furin and the human insulin receptor in this stretch was only 20% but the cysteine topography was very similar. The position of nine out of 12 cysteine residues of furin in that region appeared to be the same as in the human insulin receptor. The same topographical conservation of cysteine residues was also found in an alignment between furin and both of the cysteine rich regions of the human epidermal growth factor receptor. The overall homology between the cysteine rich regions of these two proteins was lower than between furin and the human insulin receptor. In Figure 4 the alignments of the cysteine rich regions of furin, the human insulin receptor and the amino terminal cysteine rich region of the human epidermal growth factor receptor are depicted with cysteine residues.

![Figure 1](image1.png)

Fig 1: Topography of the human fur transcription unit. The relative position of fur genetic sequences with respect to the 14.5 kbp probe is shown, as indicated in the upper part of the figure. In the lower part a more detailed restriction map of the human fur transcription unit is depicted. Exons (filled black boxes) are represented by boxes. The position of the polyadenylation signal is indicated by an arrow. An asterisk (*) marks the position of the stop codon (B BamHI). F, RRI, HpaI, KpnI, P, EcoRI, Sm, HindIII, XbaI, XhoI, XhoII.
Fig. 2 Nucleotide sequence of a > 4 kbp for genomic DNA fragment extending from a BamH I restriction endonuclease cleavage site to an FokI cleavage site. The positions of the exons, as based upon comparison with the 31 kbp DNA sequence, are indicated by arrows. Sequences depicted in capital letters are those that are present in the cDNA sequence. A stop codon is indicated by a triangle (△).
Fig. 3 Characterization of the large open reading frame on the AMV-N RNA. The deduced amino acid sequence is shown in (A). The cysteine-rich non-coding potential V region of the AMV-N RNA is indicated by a dotted line. The transmembrane domain is assigned according to the basis of multiple alignments (Fig. 4). (B) shows the alignment of the V region sequences among the human insulin receptor and (C) the overall homology between the cysteine-rich regions of the human insulin receptor and the epidermal growth factor receptor (Fig. 4).

Discussion

Here we provide evidence that the fur genetic sequences in the immediately upstream region of the fur gene proto-oncogene encode a protein with receptor-like features. This is based upon a cysteine-rich region and the presence of a transmembrane domain in the putative translation product of fur. Such domains appear widely distributed among receptor proteins.
transmembrane domain is concerned furin resembles the LDL receptor in that its transmembrane domain is located at the carboxy terminal portion of the protein. In this respect it is of interest to note that the LDL receptor belongs to a class of receptors that carry their ligands into cells after clustering in clathrin coated pits (Sudhof et al. 1985).

An intriguing feature of fur is its conserved close proximity to fes/fps proto oncogene sequences. From nucleotide sequence data we now know that the first fur transcription unit is situated closely to s fes/fps as in the human locus less than 1 kbp (data not shown) (Pfaff et al. 1985) have shown that in chicken cells the s fes/fps upstream region encoded a 4.9 kbp mRNA. It is possible that this 4.9 kbp transcript is related to fur. It has been postulated that a further argument in support of a topographical conservation of fur and fes/fps during evolution. It is not yet clear whether fur and fes/fps belong to the same class of receptors with domains shared with different proteins, or whether a functional relationship exists between furin and the human epidermal growth factor receptor. In such a model the fes/fps expression patterns of fes/fps derived ligand binding domain. In the human epidermal growth factor receptor these domains are present in a single polypeptide and in the human insulin receptor in two polypeptides derived from a single precursor molecule. The differentiation pattern in the function of furin could resolve the question as to whether a functional relationship exists between fur and fes/fps. Since truncated forms of the fes/fps protein possess transforming capabilities elucidation of the function of furin may also shed light on the mechanism of cell transformation by s fes/fps derived oncogenes.

Materials and methods

Cell lines and constellations

The primary KG1A human cell line, which was established from bone marrow cells of a patient with chronic myelogenous leukemia (Dobbie et al. 1978) was kindly provided by Dr. H. P. Koller. Isolation of the human s fes/fps cellular homologues has been described previously (Graff et al. 1982). The subclones in pSV40.4 of the 14 kbp Fori Forl Forl DNA fragment which contains all s fes/fps homologous sequences has been described previously (Roebroek et al. 1985). A 20 kbp Forl Forl Forl Forl Forl DNA fragment which flanked the 13 kbp Forl Forl Forl Forl Forl DNA fragment at S 5 and 3 Flank kbp was subcloned in pl. 18 (data not shown). The kane s fes/fps cellular homologues were isolated from KG1A cellular libraries (Roebroek et al. 1985) DNA probes and hybridization probes were isolated and labeled as described before (Van den Ouweland et al. 1985). Hybridization experiments on nuclear membranes were performed as described previously (Van den Ouweland et al. 1985) and for hybridizations on plasma membranes (Hybridon N. Amsterdam) the method of Church and Gilbert (1984) was used. The membranes were hybridized as described by Roebroek et al. (1986).
Abstract and sequence cloning of DNA libraries

Whole genomic RNA from a library of SFV 1 or SFV 2 was cloned into a library using the lambda-clone DNA library system. The library was then used to isolate DNA sequences encoding RNA polymerase. The DNA sequence was then sequenced using the Maxam and Gilbert method. The DNA sequence was then compared to the genomic sequence of SFV 1 and SFV 2.

Competitor analysis

Phage lambda DNA was isolated and used as a competitor in a series of hybridization experiments. The DNA was then hybridized to the probe DNA and the resulting bands were visualized on a gel.

Acknowledgements

We acknowledge the excellent technical assistance of A van Bochove and M de Jong and wish to thank P Pellen for her work in this manuscript. This work was supported by the Netherlands Cancer Foundation Contract No NKC Book 85-9 and by the Netherlands Klinische Foundation Contract No C 85-351.

References


Hunableman D and Mesciach M (1986) Gene 10 61-67


Koenirat H P and Dobie D W (1978) Science 200 1151-1154


Received 3 Jan 1986

56
CHAPTER 5

Structure of the feline c-fes/fps proto-oncogene:
genesis of a retroviral oncogene

J. Virol. 61, 2009-2016, 1987
Structure of the Feline c-fes/fps Proto-Oncogene: Genesis of a Retroviral Oncogene

ANTON J M ROEBROEK,1 JACK A SCHALKEN,2 CARLA ONNEKINK,1 HENRI P J BLOEMI RS,1 AND WIM J M VAN DL VEN1*

Molecular Oncology Section, Department of Biochemistry,1 and Department of Urology,2 University of Nijmegen 6525 EZ Nijmegen The Netherlands

Received 24 December 1986/Accepted 12 March 1987

The nucleotide sequence of the feline c-fes/fps proto-oncogene was analyzed. Comparison with v-fes and v-fps revealed that all v-fes/fps homologous sequences were dispersed over 11 kilobase pairs in 19 interspersed segments. All segments, numbered exon 1 to exon 19 as in the chicken and human loci, were flanked by consensus splice junctions. The putative promoter region contained a CATI sequence and three CCGCCC motifs which were also found in the human locus at similar positions. About 200 nucleotides downstream of a translational stop codon in exon 19, a putative poly(A) addition signal was identified. Sing the putative translation initiation codon in exon 2, a 93,000-molecular-weight protein could be deduced. This protein resembled very well the putative protein of the human c-fes/fps proto-oncogene (94% overall homology) and, although less well, the putative protein of the chicken c-fes/fps proto-oncogene (78% overall homology). As far as the feline c-fes/fps proto-oncogene sequences are concerned, homology in deduced amino acid sequences between the GA- and ST-v-fes viral oncogenes and the proto-oncogene was 99%. Analysis of the recombination junctions between feline leukemia virus and v-fes sequences in GA- and ST-FeSV proviral DNA revealed for the left-hand junction the involvement of homologous recombination, presumably at the DNA level. The right-hand junction, which appeared identical in the GA-FeSV and ST-FeSV genomes, could have been the result of a site-specific recombination at the RNA level.

Retroviruses capable of acute transformation have acquired proto-oncogene sequences from their natural hosts as an integral part of their genome (4, 5). Three independently derived feline sarcoma virus (FeSV) isolates (Gardner-Arnstein [GA] [15], H21 [45], and Snyder-Theilen [ST] [46]) appear to have acquired sequences from the same feline proto-oncogene (c-fes) (13, 14, 18, 45). Several avian sarcoma viruses (Fujinami sarcoma virus, the PRC viruses, URI virus, and 16L virus [reviewed in references 4 and 5]) have captured similar sequences from the avian counterpart (c-fes) (17, 42, 43). To recognize this relationship, the proto-oncogene is generally referred to as c-fes/fps. In GA- and ST-FeSV, the translational product of the viral transforming gene consists of a feline leukemia virus (FeLV) gag gene-encoded amino-terminal component fused to a v-fes-encoded carboxy-terminal component (2, 39, 51) which was shown to exhibit tyrosine-specific kinase activity (3, 52). This enzyme activity appeared to be required for malignant transformation (8, 34). The kinase domain is encoded by the 3' end of the v-fes sequences that both viral oncogenes share (18). At their 5' ends GA- and ST-FeSV both have unique v-fes sequences.

The precise function of the c-fes/fps proto-oncogene is not yet known. Its translation product, a 92,000-molecular-weight protein (NCP92) in cats (1) and humans (11, 28) and a 98,000-molecular-weight protein (NCP98) in chickens (29, 40), exhibits tyrosine-specific kinase activity. It is mainly found in hematopoietic tissue (11, 28, 29, 40). This restriction to hematopoietic tissue points to a possible function of the c-fes/fps proto-oncogene in hematopoietic proliferation. The c-fes/fps mRNA species that have been reported varied in size between 2.6 and 3.2 kilobases (19, 35, 40, 44).

The molecular cloning of the chicken (19), feline (53), and human (46) c-fes/fps proto-oncogenes has been described previously. Only the chicken (19) and the human (35, 37) c-fes/fps proto-oncogenes have been analyzed by nucleotide sequence analyses. To elucidate some previously unknown features of the c-fes/fps proto-oncogene, we have analyzed the feline c-fes/fps proto-oncogene by sequence analysis. Comparison with the human and chicken data revealed some interesting features concerning the putative promoter region and the protein-encoding sequences. Based on nucleotide sequence data, a molecular model for the transduction of feline c-fes/fps sequences to GA- and ST-FeSV is proposed.

MATERIALS AND METHODS

**Molecular cloning.** Isolation of the feline v-fes/fps cellular homolog from a cosmid library has been described previously (53). The v-fes/fps homologous sequences were subcloned in pUC18.

**DNA sequence analysis.** DNA fragments were inserted into the polylinker region of M13mp8-19. All of the DNA sequences were determined by the dye deoxy sequencing method (41). The gel readings were recorded, edited, and compared using the Staden programs (47).

**RESULTS AND DISCUSSION**

Topography and nucleotide sequence of feline c-fes/fps. A DNA fragment of 13 kilobase pairs (kb) containing the feline v-fes/fps cellular homolog (53) was analyzed by DNA nucleotide sequencing. Comparison of these feline nucleotide sequences with those of the GA- and ST-v-fes (18), v-fps (42), and the human (35, 37) and chicken (19) c-fes/fps
proto-oncogene revealed the distribution of the feline v-fes/fps homologous sequences over a DNA region of 11 kbp. Figure 1 shows a restriction map of the feline c-fes/fps proto-oncogene with a schematic distribution of 19 putative exons. In Fig. 2 the nucleotide sequence of the feline c-fes/fps proto-oncogene is presented. Major portions of the intervening sequences (introns) are not presented because their sequences were in most cases obtained only by sequencing one strand whereas the sequences homologous to v-fes/fps (exons) were accurately obtained by sequencing both strands. As can be seen all v-fes/fps homologous fragments were flanked by the AG and GT splice junction consensus sequences and in most cases good agreement with the complete consensus splice junction sequences (31) was observed.

The size and distribution of feline and human exons appeared highly similar. Overall homology between human and feline sequences within the exons was 88%. In fact, the 5' boundary assignment of human exon 2 and identification of human exon 1 sequences was made possible by comparison of human and feline c-fes/fps proto-oncogene sequences (32). Compared with the chicken locus, conservation of genetic sequences is lost upstream of a putative translation initiation codon in exon 2 which is conserved in feline, human and chicken sequences. Chicken exon 2 extended 57 nucleotides further upstream of the putative start A10 than did the feline and human counterparts. Downstream of this A10 homology between the protein coding sequences of chicken and feline (or human) DNA was 74%. Divergence upstream of this A10 is probably due to genetic drift outside the protein coding region of c-fes/fps.

At their 3' ends the c-fes/fps loci in feline and chicken DNAs diverge downstream of the TGA stop codon in exon 19. At this point divergence between feline and human sequences is less striking. Overall homology between feline and human genes dropped from 92% in the protein-coding sequences of exon 19 to 68% in the noncoding sequences of exon 19 whereas in the region of the putative poly(A) addition signal and poly(A) addition site and 30 nucleotides further downstream the homology was found to be 85% supporting the idea that the putative poly(A) addition signal and poly(A) addition site in the feline locus are functional as has been proven for the human locus (35).

The 5' boundary of the c-fes/fps proto-oncogene requires further elucidation. Comparison of feline c-fes/fps nucleotide sequences with those of GA v-fes revealed only 18 nucleotides of the 3' end of exon 1 (see Fig. 2). Recently the discovery was described of a new transcription unit (tta) in the immediately upstream region of the human and feline c-fes/fps proto-oncogene (15), apparently encoding a receptor-like protein (36). The 5' boundary of this transcription unit is in close proximity to c-fes/fps exon 1 11 kbp upstream of known c-fes/fps exon 1 sequences in both human and feline DNA. Searching in this narrow region for sequences typical for promoter regions revealed a sequence that is in good agreement with a CATT box (see Fig. 2 and 3). This sequence was also present at a similar position (285 nucleotides) upstream of the 5' boundary of exon 1 in human DNA (35) (see Fig. 3). However as was the case in the human locus, no sequences resembling the consensus sequence of a TATA box (6) could be identified in the feline locus. Other interesting features of this region between 5' and c-fes/fps were three conserved CA motifs (in human and feline loci at similar positions) (see Fig. 2 and 3). These CA motifs and the sequence GGCGCCC is found in multiple copies in promoter regions of a number of genes such as the gene that encodes the receptor for the human epidermal growth factor (23) the human Harvey ras proto-oncogene (see 22) and the mouse (30) and human (31) hypoxanthine guanine phosphoribosyltransferase genes. It has the potential to bind specifically transcription factor SP1 (9). Although the above mentioned promoter characteristics could point to promoter activity of this region there is still no proof of actual promoter activity. Further studies to localize the transcription initiation sites by S1 nuclease and primer extension experiments in feline as well as human genes are in preparation in our laboratory.

The feline c-fes/fps protein. The feline c-fes/fps coding region probably starts at an ATG near the 5' end in exon 2 (see Fig. 2). At a similar position an ATG was found in the human and chicken c-fes/fps and it was postulated that they represented translation start sites (39-41). Several indications argue in favor of the assignment of the translation start to this ATG. First the surrounding nucleotides in feline and human as well as chicken loci are consistent with the consensus translation initiation sequence 5'ATAATGG (37, 26). Second downstream of the ATG significant homology in nucleotide and deduced amino acid sequences is found between feline, human and chicken genes. Upstream of the ATG homology is limited to the human and feline nucleotide sequences (data not shown). The amino acid sequences corresponding to the codons in this presumably untranscribed

---

**Figure 1** Topographs of the feline c-fes/fps proto-oncogene. A schematic restriction map of the 11 kbp DNA region containing the v-fes/fps proto-oncogene is depicted in the upper part of the figure. The boxes representing the c-fes/fps exons are numbered as in the chicken (39) and human loci. **Exon 1** is indicated by an open box with a dashed line since its 5' border is not yet defined. The asterisks above exon 19 indicate a position of the stop codon. Homologous segments within GA and ST v-fes are indicated by boxes in the lower part of the figure. B BamHI Bg Il F EcoRI H HindIII K KpnI S SacI S Sall X XhoI.
sequence are not conserved. Furthermore, starting at the proposed initiation ATG, the molecular weights of the deduced proteins of the feline, human, and chicken c-cps are 93,000, 93,000, and 94,000, respectively. These values are in good agreement with the reported molecular weights of the c-cps proteins (see Introduction). For chicken c-cps, it was shown that the same postulated start codon indeed functioned as a translation start ATG in retroviral
FIG. 3. Comparison of the putative promoter region of the feline and human proto-oncogene c-fes/fps. The 3' end of exon 1 is indicated; nucleotides of exon 1 identified so far are given in capitals. Nucleotide positions not represented in one of the DNA stretches are indicated by hyphens. The potential CATT box is outlined with a box. CCGCCC motifs are underlined.

constructs, encoding a protein electrophoretically indistinguishable from the chicken NCP98 protein (12).

Counting downstream of the presumed initiation ATG, the feline c-fes/fps proto-oncogene contains a large open reading frame of 2,460 nucleotides, whereas in the human and chicken genes open reading frames of 2,466 and 2,472 nucleotides, respectively, were found. At approximately 50 nucleotides from the 3' end of exon 2, feline c-fes/fps lacked

FIG. 4. Comparison of the deduced amino acid sequences of the gene products of feline c-fes/fps, GA- and ST-v-fes (18), human c-fes/fps (35, 37), and chicken c-fes/fps (19). The amino acid sequence of feline c-fes/fps is shown in the conventional one-letter code. The position of the putative start methionine (marked with an asterisk) is considered as position 1. The amino acid substitutions in GA- and ST-v-fes and human and chicken c-fes/fps relative to the feline proto-oncogene product are indicated. Identical amino acids are represented by hyphens.

Positions not represented are left blank. O, Possible phosphoacceptor tyrosine. The amino acid sequence of ST-v-fes is incomplete, because it was deduced from nucleotide sequence data of a deletion mutant of ST-v-NSV (18).
Comparison of the feline c-fesfps protein and the FeSV-encoded transforming protein. The genomes of GA and SI FeSV encode a transforming fusion protein consisting of a gag gene-derived amino terminal part and a v-fes-derived carboxy-terminal part (18). GA-FeSV shows six nucleotides homologous to feline c-fesfps exon 1 sequences and ends with sequences homologous to sequences within the noncoding region of exon 19 (see Fig 1 and below). As the left-hand junction between the gag-related and the v-fes-related sequence is located upstream of the putative translation initiation ATG of c-fesfps, the fusion protein encoded by GA-FeSV contains a short stretch of amino acids presumably not present in c-fesfps protein. The GA-FeSV-encoded fusion protein does not contain all the c-fesfps-encoded amino acid sequences. During or after the generation of GA-FeSV sequences homologous to feline c-fesfps exon 1 to exon 9 sequences are deleted (see Fig 1 and next section).

The v-fes gene of the other virus strain SI-FeSV starts at its 5' end with sequences homologous to feline c-fesfps exon 8, whereas the left-hand junction of v-fes to FeLV is identical to the one found in GA-FeSV (see Fig 1 and below). Compared with feline c-fesfps, the nucleotide substitutions were observed in GA-v-fes which gave rise to three amino acid substitutions (Fig 4) In SI-FeSV eight point mutations resulted in four amino acid substitutions (Fig 4).

Sites of recombination between feline c-fesfps and FeLV. To gain more insight into the events that led to the transduction of feline c-fesfps sequences by FeLV we studied the feline c-fesfps and FeLV sequences in the regions in which the c-fesfps and FeLV DNA sequences are present in GA and SI FeSV were fused together. We assumed that the left-hand and right hand recombination junctions are present in GA and ST-FeSV reflected the initial recombination event and not some subsequent viral rearrangement feature. Figure 5 (A and B) shows the alignments of feline c-fesfps and FeLV sequences with regard to the left-hand junctions. GA-FeSV shares six nucleotides at the left hand recombination site with c-fesfps exon 1 and FeLV sequences (Fig 5A). This alignment allowed the identification of as much as 16 nucleotides of c-fesfps exon 1. It should be noted that the comparison was made with gag gene sequences of the natural helper (and presumed progenitor) FeLV strain of ST-FeSV (FeLV [ST-strain]) because no gag gene sequence data of FeLV (GA strain) were available. This comparison must be considered as permissible because the overall homology of nucleotide sequences between the two FeLV strains is expected to be about 95%. This is based upon the strong homology between gag sequences of both FeLV and FeSV strains and SI-strain (18) and a homology of about 95% between the c-myc genes of both FeLV and FeSV strains (10). SI-FeSV shares eight of nine nucleotides at the left-hand recombination site with c-fesfps exon 8 and FeLV sequences (Fig 5B). Although the finding of shared nucleotides at the recombination site by itself gives no information about a possible mechanism for homologous recombination in short regions of homologous DNA sequences to have been involved in the left hand junction in GA and SI FeSV especially because it is in good agreement with the presumed first step in the general model of transduction of cellular protooncogenes by retroviruses (4) the left-hand junction is thought to be the result of recombination between cellular and proviral DNA. Upon integration of a provirus upstream of a protooncogene subsequent deletion of 3' end proviral sequences could link viral to protooncogene sequences and generate an oncogene. The role of short stretches of homologous sequences in the left hand recombination is not clear. Previous studies on other retroviral oncogenes argue in favor of a role (54) or against (19, 24, 48, 49) the importance of homologous sequences at the left hand recombination junctions. Apparently they are helpful although not necessary. Probably a second homologous recombination event in short stretches of homologous DNA sequences was involved in generation of GA-FeSV. GA-FeSV lacks sequences homologous to sequences of feline c-fesfps exon 5 to exon 9 due to a presumed deletion event in a progenitor FeSV. At the junction, GA-FeSV

63
The second step in the right hand recombination is thought to occur by copy choice during reverse transcription of a heterodimer molecule formed between the chimeric gene and the retrovirus (7). Support is found in the location of all known right hand recombination junctions within exons of proto-oncogenes (19, 24, 48-50). Recently Huang et al. (20) provided direct experimental support for this view by showing that the right hand end of the chimeric gene-Prib II was formed by recombination within the pol(A) tract at the 3' end of the mRNA of the chicken proto-oncogene *fev* (28). In the model described above, the right hand recombination sites in GA and S1 FeSV are also likely to be the result of recombinations at the RNA level. The common right hand recombination site in the two FeSV strains (Fig. 5C) is inferred from the observation that the FeSVs share seven nucleotides at their common recombination site. The FeSVs are expressed from the same copy of the FeLV genomic RNA. The common left hand end of the FeLV genomic RNA to the *fev* mRNA resulting in the right hand recombination junction. A similar heterodimer structure was proposed by Tagawa et al. (21) to explain the variation in the sequences of sequences in FeLV. It has been suggested that the heterodimer structure could have been involved in the copy switch in a similar way as postulated for recombinant between RNA genomes of polioviruses (51). However, additional complexities in these recombinant molecules have not been observed in experiments, although homologous sequences were present and these sequences could have been involved in the copy switch in a similar way as postulated for recombinant between RNA genomes of polioviruses (51). The first step in the right hand recombination is thought to occur by copy choice during reverse transcription of a heterodimer molecule formed between the chimeric gene and the retrovirus (7). Support is found in the location of all known right hand recombination junctions within exons of proto-oncogenes (19, 24, 48-50). Recently Huang et al. (20) provided direct experimental support for this view by showing that the right hand end of the chimeric gene-Prib II was formed by recombination within the pol(A) tract at the 3' end of the mRNA of the chicken proto-oncogene *fev* (28). In the model described above, the right hand recombination sites in GA and S1 FeSV are also likely to be the result of recombinations at the RNA level. The common right hand recombination site in the two FeSV strains (Fig. 5C) is inferred from the observation that the FeSVs share seven nucleotides at their common recombination site. The FeSVs are expressed from the same copy of the FeLV genomic RNA. The common left hand end of the FeLV genomic RNA to the *fev* mRNA resulting in the right hand recombination junction. A similar heterodimer structure was proposed by Tagawa et al. (21) to explain the variation in the sequences of sequences in FeLV. It has been suggested that the heterodimer structure could have been involved in the copy switch in a similar way as postulated for recombinant between RNA genomes of polioviruses (51). However, additional complexities in these recombinant molecules have not been observed in experiments, although homologous sequences were present and these sequences could have been involved in the copy switch in a similar way as postulated for recombinant between RNA genomes of polioviruses (51).
17. Groffen J N Heisterkamp M Shihubva H Hanafusa and J R Stephenson 1983: Transforming genes of F-Mu (v-fis) and mamillin (v-fes) retroviruses correspond to a common cellular locus Virology 125 480-486
20. Huang C C and H van and J M Bishop 1986: The role of RNA molecules in transformation of the protooncogene c-fes Cell 44 915-940
22. Ish o T G Merlino and I Pastan 1985: Promoter region of the human Harvey sarcoma virus oncogene similarity to the c-c-fos provirus oncogene promoter Science 220 1776-1818
23. Ish o T Y Su R H Stratton B A Rose G Merlino and I Pastan 1985: Characterization and sequence of the murine c-MuLV cell transforming gene Natl Acad Sci USA 82 4924-4928
25. Kozak M 1984: Computation and analysis of sequences up stream from the translational start site in cellular mRNAs Nuc acids Res 12 857-872
29. Mathes Prevot B H Hanafusa and S Kawai 1982: A cellular protein immunologically cross reactive with and functionally homologous to the Fujinami sarcoma virus transforming protein Cell 28 897-906
40. Sanger F S Nicklen and A R Coulson 1977: DNA sequencing with chain termination inhibitors Proc Natl Acad Sci 74 5463-5467
42. Shihubva M J Hanafusa H Hanafusa and J R Stephenson 1980: Homology exists among the transforming sequences of avian feline sarcoma viruses Proc Natl Acad Sci USA 77 6546-6540
44. Snyder W H M C Singhal E E Zuckerman and W D Hardy 1984: Isolation of a new feline sarcoma virus (FI-FeMr) biochemically and immunologicall characterization of its translation product Virology 132 205-210
47. Swanson R R C Parker H F Varanv and J M Bishop 1983: Transformation of a cellular oncogene the genome of Rous sarcoma virus Proc Natl Acad Sci USA 80 7519-7523

65


52 Van de Ven, W. J. M., F. H. Reynolds, Jr., and J. R. Stephenson. 1980 The nonstructural components of polyproteins encoded by replication-defective mammalian transforming retroviruses are phosphorylated and have associated protein kinase activity. Virology 101:185-197


The discovery of retroviral oncogenes was an important step to better understand tumorigenesis. Today many so-called proto-oncogenes are known and malignant activation may turn them into oncogenes involved in tumorigenesis. The gene, subject of investigation in this thesis, is the \textit{c-fes/fps} proto-oncogene.

In chapter 1, a summary is given on proto-oncogenes in general and how they can be activated malignantly. The proto-oncogene \textit{c-fes/fps} is discussed in more detail. The function of this gene is not known yet, however its translation product belongs to the cytoplasmatic tyrosine-specific protein kinases. To give some insight in the type of function the \textit{c-fes/fps} proto-oncogene may fulfill, the protein kinases, all of which are involved in signal transduction systems, are described in general terms.

Chapter 2 defines the human \textit{c-fes/fps} proto-oncogene by nucleotide sequence analysis. Comparison with viral \textit{v-fes/fps} and chicken \textit{c-fes/fps} nucleotide sequence data revealed the distribution of 18 exons over 11 kbp. These 18 exons, together about 2.7 kbp, encode the complete putative protein, a tyrosine-specific protein kinase. The most probable start codon was found in the first identified exon at the 5′-end.

The localization of the 5′-end of the human \textit{c-fes/fps} proto-oncogene, however, remained unclear. Further studies to resolve this problem led to the discovery of the 3′-end of a non-coding \textit{c-fes/fps} exon upstream of the exon containing the presumed initiator ATG, and to the discovery of a transcription unit immediately upstream of the \textit{c-fes/fps} proto-oncogene, in man as well as in cat. This transcription unit is called \textit{fur} (for \textit{fes/fps} upstream region). The \textit{c-fes/fps} promoter region is probably located between the known \textit{c-fes/fps} exon sequences and the \textit{fur} gene, a region of only 1.1 kbp. The investigations mentioned above are presented in chapter 3.

The human \textit{fur} gene is described in more detail in chapter 4. Analysis of a 3.1 kbp cDNA covering about 65% of the \textit{fur} mRNA revealed the C-terminal part of the putative \textit{fur} translation product, furin. Analysis of the predicted amino acid sequence pointed to a possible function as receptor located in the cell membrane.

Chapter 5 describes the feline \textit{c-fes/fps} proto-oncogene. It appears highly similar with the human gene. Based upon comparison of nucleotide sequences of the feline and human genes, further indications are presented to
indentify the putative promoter region. Comparison with FeLV, GA- and ST-
v-fes sequences made it possible to propose a model to explain the
transduction of feline c-fes/fps sequences to FeLV resulting in GA- and
ST-FeSV. This model explains why both viruses, though independent iso-
lates, have identical left-hand recombination sites.
SAMENVATTING

De ontdekkking van virale oncogenen was een belangrijke stap voor een beter begrip van het ontstaan van tumoren. Heden ten dage zijn er vele zogenaamde proto-oncogenen bekend en door maligne activering veranderen zij in oncogenen betrokken bij het ontstaan van tumoren. Het gen onderwerp van onderzoek in dit proefschrift, is het proto-oncogen c-fes/fps.

In hoofdstuk 1 wordt een samenvatting gegeven over oncogenen in het algemeen en hoe zij maligne geactiveerd kunnen worden. Het proto-oncogen c-fes/fps wordt in meer detail besproken. De functie van dit gen is nog niet bekend, hoewel het protein, waarvoor het codeert, behoort tot de cytoplasmatiscne tyrosine-specifiche proteïne-kinases. Om enig inzicht te geven in het soort functie, dat het proto-oncogen c-fes/fps zou kunnen vervullen, worden proteïnez-kinases, allen betrokken bij signaal-transductie-systemen, in het algemeen beschreven.


De locatie van het 5'-uiteinde van het proto-oncogen c-fes/fps was vooralsnog onduidelijk. Verdere studies om dit probleem op te lossen leidden tot de ontdekking van het 3'-uiteinde van een onbekend, aan het 5'-uiteinde gelegen exon van c-fes/fps, en tot de ontdekking van een transcriptie-eenheid onmiddellijk stroomopwaarts van het proto-oncogen c-fes/fps, zowel bij de mens als bij de kat. Deze transcriptie-eenheid wordt fur genoemd (voor fes/fps upstream region). De promoter-regio van c-fes/fps zou gelegen kunnen zijn tussen de bekende c-fes/fps-exonen en het fur-gen, een regio van slechts 1,1 kbp. Het hierboven vermelde onderzoek wordt gepresenteerd in hoofdstuk 3.

Het menselijke fur-gen wordt in meer detail in hoofdstuk 4 beschreven. Analyse van een 3,1 kbp cDNA, dat ongeveer 65% van het fur mRNA representeren, onthulde het C-terminale deel van het vermoedelijke translatie-produkt van fur, furine. Analyse van de voorspelde aminozuur-gevolgende wijst op een mogelijke functie als receptor gelocaliseerd in de cellmembraan.
LIST OF PUBLICATIONS


Sinds 1 maart 1983 is hij verbonden als wetenschappelijk assistent aan de afdeling Biochemie van de Faculteit der Wiskunde en Natuurwetenschappen van de Katholieke Universiteit te Nijmegen, tot 1 november 1985 in het kader van een FUNGO-project, gesubsidieerd door de stichting voor Zuiver Wetenschappelijk Onderzoek (Z.W.O.), en daarna in het kader van een door het Koningin Wilhelmina Fonds gesubsidieerd project. Gedurende de periode tot januari 1987 werd ondermeer onderzoek verricht aan de mense- en katte-c-fes/fps-proto-oncogenen en het menselijke fur-gen in de werkgroep van Prof. Dr. H.P.J. Bloemers en Dr. W.J.M. Van de Ven. Momenteel richt hij zijn onderzoek op longtumoren in een samenwerkingsverband tussen de werkgroep Moleculaire Oncologie (Dr. W.J.M. Van de Ven) van de afdeling Biochemie en de afdeling Pathologische Anatomie (Prof. Dr. G.P. Vooyks).
I

De overgang van gag-gen-nucleotide-sequenties naar v-fes-oncogen-nucleotide-sequenties, zoals door Hampe et al. door vergelijking van FeLV-sequentie-gegevens met ST- en GA-FeSV-sequentie-gegevens bepaald, is voor ST-FeSV en GA-FeSV respectievelijk een en twee nucleotiden te ver stroomopwaarts gelegd.


II

Immunohistochemie en in situ hybridisatie zouden uitsluitse kunnen geven over de identiteit van het celtype, dat verantwoordelijk is voor de verhoogde c-sis-transcriptie in artherosclerotische lesies in vergelijking met normaal arteriëel weefsel.


III

Een internationale wetenschappelijke commissie dient afspraken te maken over het gebruik van een eenduidige nomenclatuur voor groeifaktoren en hun receptoren.

IV

Pfaff et al. betrachten onvoldoende terughoudendheid ten aanzien van het gebruik van zgn. consensus-sequenties bij het toekennen van potentiële functies aan een DNA-sequentie.


V

Detectie in de NIH-3T3-transfectie-assay van een geactiveerd oncogen, dat niet in de primaire tumor kan worden aangetoond, duidt waarschijnlijk niet op activering tijdens de transfectie, maar op het voorkomen van dit geactiveerde gen in een klein deel van de tumorcellen als gevolg van tumortoprogressie.

De suggestie van Fukui et al. dat in alle drie de primaire transformanten verkregen door transfectie van glioblastoma GL-5-JCK DNA, een menselijke homoloog van v-raf verantwoordelijk is voor de transformatie, wordt in een latere publicatie tegengesproken.


Katalytische antilichamen zijn belangrijke potentiële bouwstenen voor zgn. protein engineering.

Tramontano et al. (1986) Science 234, 1566-1570.

Toepassing van een differentiële hybrïdisatie-assay biedt betere perspectieven voor detectie en verdere karakterisering van genen, die specifiek in een tumor tot expressie komen, dan een benadering die uitgaat van het opwekken van monoklonale antilichamen tegen de genprodukten van dergelijke genen en deze vervolgens te gebruiken voor isolatie van de desbetreffende genen.

Bij de realisatie van de Europese Economische Gemeenschap ligt de nadruk te veel op de kwantiteit en te weinig op de kwaliteit.

Vanwege het te grote aantal honden in Nederland dient de overheid een ontmoedigingsbeleid te voeren ten aanzien van het houden van honden als huisdier.

De overeenkomsten tussen de derde dinsdag in september en een andere, vooral in het zuiden van het land zeer gewaardeerde dinsdag zijn legio.

Nijmegen, 19 november 1987

A.J.M. Roebroek