Volume 218, number 1, 81–86 FEB 04777 June 1987

The 3′-orf protein of human immunodeficiency virus shows structural homology with the phosphorylation domain of human interleukin-2 receptor and the ATP-binding site of the protein kinase family

Kenneth P. Samuel[+], Arun Seth, Andrzej Konopka[+], James A. Lautenberger[°] and Takis S. Papas[°]

*Program Resources, Inc., *Bionetics Research, Inc., Basic Research Program, [+]Laboratory of Mathematical Biology,* and *[°]Laboratory of Molecular Oncology, Carcinogenesis Regulation Section, National Cancer Institute-Frederick Cancer Research Facility, Frederick, MD 21701-1013, USA*

Received 9 April 1987

The primary amino acid sequence within a stretch of 25 residues (positions 91–116) of the middle portion of the 3-′orf protein ($p27^{3'}$-orf) of the human immunodeficiency virus (HIV) shares structural homology with a highly charged region within the intracytoplasmic phosphorylation domain of human interleukin-2 receptor (IL-2R) and the ATP-binding site of the catalytic subunit of cAMP-dependent protein kinase (cAMP-PK) and other members of the protein kinase family. Comparison of the predicted secondary structure within this region of $p27^{3'}$-orf with the phosphorylation domain of human IL-2R and the ATP-binding region of the phospho-kinase family of protein suggests that the 3′-orf protein could serve homologous function(s).

3′-orf protein; IL-2 receptor; Phosphorylation domain; ATP-binding site; Protein kinase family; Human immunodeficiency virus

1. INTRODUCTION

The genome of HIV, the infectious agent linked with AIDS and ARCs [1–3], encodes viral gag, pol and env genes typical of other retroviruses [4–7].

Correspondence address: K.P. Samuel, Program Resources, Inc., Frederick Cancer Research Facility, Bldg. 469, Frederick, MD 21701-1013, USA

Abbreviations: HIV, human immunodeficiency virus; IL-2R, interleukin-2 receptor; orf, open reading frame; AIDS, acquired immune deficiency syndrome; ARC, AIDS-related complex; cAMP-PK, cyclic AMP-dependent protein kinase; cGMP-PK, cyclic GMP-dependent protein kinase; phos.b-K, γ-subunit of rabbit skeletal muscle phosphorylase *b* kinase; Xaa, any amino acid

However, unlike the typical retrovirus, other genes are present, including the tat, art, sor, and 3′-orf [4–7]. The 3′-orf gene (648 bp) begins in a different reading frame at the end of the gp41 envelope sequence, and extends into the 3′-LTR where it terminates in the U_3 region [4–7]. A premature termination codon (TAG) at position 124 truncates the gene product (124 amino acids) in HIV isolates HX-B2 and BH10 [4,8], whereas most other isolates are predicted to express a protein of 206 amino acids [9,10]. This protein, $p27^{3'\text{-orf}}$ [9], has been localized to the cytoplasmic fraction of Molt-4 infected [11] and H9-infected ([11], and Samuel, K. Showalter S. and Zweig, M., unpublished) T-cells.

As a first step towards delineating a possible function for $p27^{3'\text{-orf}}$ during HIV infection, a

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

search was conducted for structural homology with other cellular or viral oncogene proteins previously localized to the cytoplasm. This report describes the finding of sequence homology between $p27^{3'\text{-orf}}$, the protein kinase C catalyzed phosphorylation domain within the short intracytoplasmic tail of human IL-2R, and the nucleotide-binding site of the protein kinase family of proteins.

2. EXPERIMENTAL AND RESULTS

An initial visual search of the deduced amino acid sequences of T-cell receptors IL-2R and T_4 (CD4) antigen revealed a stretch of highly charged, mainly basic residues of consensus $QRRQ(X_{aa})_6I$, at positions 241–251 of the intracytoplasmic tail of human IL-2R [12,13], showing striking sequence homology (~41%) with a highly charged region (positions 104–114) within the $p27^{3'\text{-orf}}$ protein of HIV isolates, located just prior to the premature termination codon found in the 3'-orf sequence of some infectious HIV isolates [8]. As summarized in fig.1, this sequence is highly conserved, with few amino acid changes, among all of the HIV isolates, and has also been identified among a limited but diverse group of proteins, including N-myc [14] and cytochrome P-450 of *Pseudomonas putida* [15]. The significance of this homology is not yet understood. However, it is known that stimulation of resting T-cells with mitogens and antigens triggers the activation of protein kinase C catalyzed phosphorylation of serine (S) residue 247 and threonine (T)-250 at the

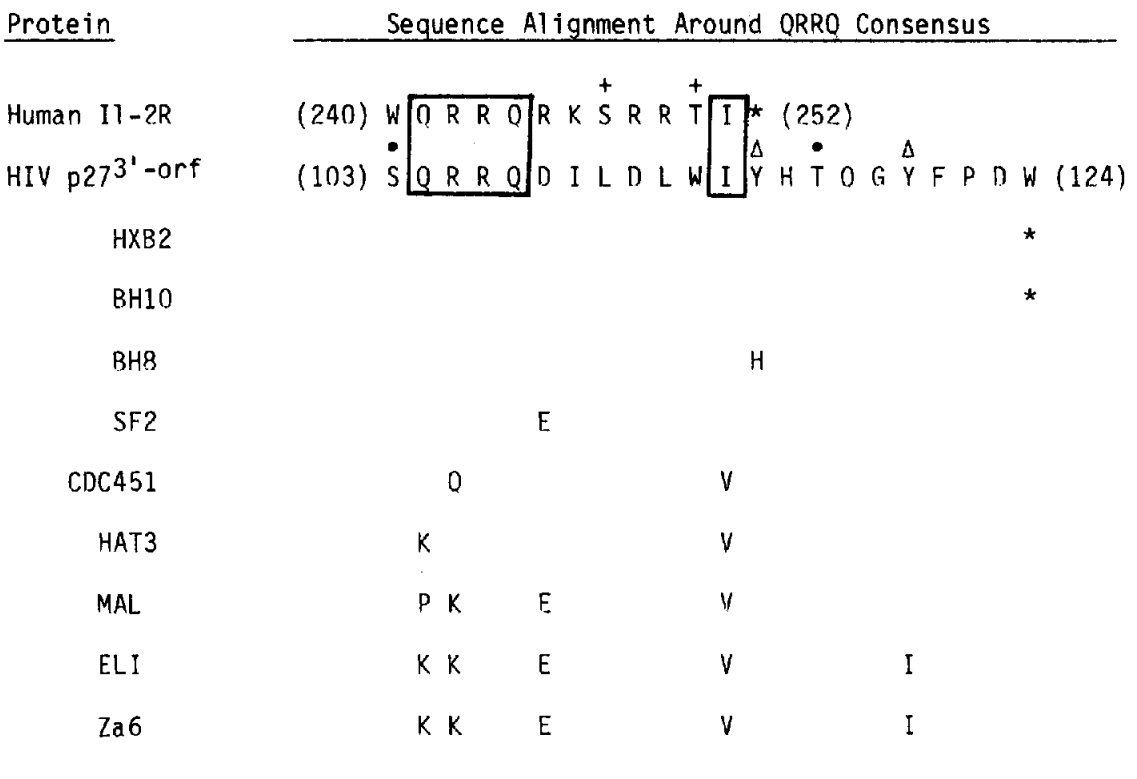

Fig.1. Alignment of amino acid residues around the $QRRQ(X_{aa})_6I$ consensus of human IL-2R and HIV $p27^{3'\text{-orf}}$. Identifying symbols: (+) serine-247 and threonine-250 of human IL-2R; (●) highly conserved serine-103 and threonine-117 of $p27^{3'\text{-orf}}$; (Δ) less conserved tyrosines-115 and -120 of $p27^{3'\text{-orf}}$; (*) termination codons. References cited for HIV isolates: LAV strains BRU, MAL, and ELI [5]; HXB2, HXB3, C15, HAT3, BH10, BH8, PV22, and SF2 [8]; and CDC451 isolate [38]. The one letter amino acid code is used.

intracytoplasmic domain of human IL-2R [16,17]. Located within the corresponding $p27^{3'\text{-orf}}$ sequence are a threonine (T) at position 117 and serine (S) at 103, which are highly conserved among all HIV isolates (fig.1), and are potential targets of protein kinase C catalyzed phosphorylation.

We then looked for structural relationship between $p27^{3'\text{-orf}}$ and conserved sequences at the active site of the protein kinases and proteins that are

Protein		Sequence Alignment at Nucleotide Binding Domain		Ref.
HIV HAT3 $p27^{3'\text{-orf}}$	(93)	E K G - G L D G L V F - 2 - K R Q D I L D	(112)	[8]
Cyclic nucleotide-dependent				
cAMP-PK	(48)	T L G T G S F G R V M - 10 - Y A M K I L D	(75)	[18]
cGMP-PK	(364)	T L G V G G F G R V E - 11 - F A M K I L D	(392)	[19]
Ca^{2+}dependent				
Phos. b-kinase	(24)	I L G R G V S S V V R - 10 - Y A V K I I D	(51)	[19]
PK-C	(344)	V L G K G S F G K V M - 10 - Y A I K I L K	(371)	[36]
GTP-binding				
$p21^{ras}$	(8)	V V G A G G V G K S A - 32 - C L L D I L D	(60)	[37]
Dinucleotide-binding				
LDH	(25)	V V G V G A V G M A C - 12 - V A L V D V M	(55)	[27]
GR	(25)	V I G G G S G G L A S - 10 - A A V V E S H	(52)	
Growth Factor Receptors				
EGF-R	(693)	V L G S G A F G T V Y - 14 - V A I K E L R	(724)	[19]
In-R	(988)	E L G Q G S F G M V Y - 15 - V A V K T V N	(1021)	
Cell Division Protein				
CDC28	(13)	K V G E G T Y G V V Y - 10 - Q R V V A L K	(40)	[28]
Src Family of Viral Oncogenes				
p60 src	(272)	K L G Q G C F G E V W - 9 - V A I K T L K	(298)	[19]
p120 gag-abl	(368)	K L G G G Q Y G E V Y - 10 - V A V K T L K	(395)	
p37 mos	(99)	R L G S G G F G S V Y - 8 - V A I K Q V N	(124)	
CONSENSUS:		G X_{aa}G$X_{aa}$$X_{aa}$G_____ 15-23____ K		

Fig.2. Alignment of amino acid sequences at the ATP-binding site of the protein kinase (PK) family. Closed boxes identify conserved consensus or canonical residues of known and putative members of the PK family. Broken-line boxes enclose residues in $p27^{32'\text{-orf}}$ and subfamilies of nucleotide-binding proteins having a highly conserved DILD or ILD sequence. Numbers in parentheses are amino acid coordinates at end of sequences identified. Sequences were aligned by eye.

themselves targets for protein kinase catalyzed phosphorylation. During this search, a glycine (G)-rich cluster of amino acids, located only four residues N-terminal to the QRRQ consensus in $p27^{3'\text{-orf}}$, was identified (residues 93–112), which showed striking sequence homology with the glycine-rich canonical $GX_{aa}GX_{aa}X_{aa}G$ sequence of the catalytic subunit of bovine cAMP-PK [18], the prototype of a family of homologous protein kinases [19]. As shown in fig.2, only a single gap (-) was introduced between the first and second G-residues of $p27^{3'\text{-orf}}$ to maximize the homology.

This apparent structural homology between the HIV $p27^{3'\text{-orf}}$ protein and the nucleotide-binding region of the protein kinase family [18,19] raises interesting questions about possible structural and functional evolution of this viral protein. A conserved lysine (K) residue proposed to be critical in binding ATP or other nucleotides by the cAMP-PK [20], oncogene encoded tyrosine and serine/threonine protein kinases [19], and growth factor receptor kinases [21,22], is positioned 15–28 residues C-terminal to the glycine-rich cluster (fig.2). However, a corresponding lysine residue is not similarly located in the HIV $p27^{3'\text{-orf}}$ protein, which was replaced by either a glutamic (E) or aspartic (D) acid among different viral isolates (see also fig.1). In some HIV isolates, a lysine is located two amino acids preceding the E or D residues. Moreover, the primary amino acid sequence around the VA_I^VK consensus, which is located 15–28 residues from the glycine-rich cluster, appears to be quite variable among the subfamilies of protein kinases. Thus within this region of a consensus nucleotide-binding sequence, $p27^{3'\text{-orf}}$ more closely resembles the viral oncogene $p21^{ras}$, a guanine nucleotide-binding protein [23], in that they both lack the VA_I^VK consensus (fig.2) and instead have retained the ILD or DILD sequence in the same relative position (broken-line box). Similarly, the dinucleotide-binding proteins, such as lactate dehydrogenase (LDH) and glutathione reductase (GR) (fig.2), also lack the corresponding lysine residue at the VA_I^VK consensus. It is of interest to also note that the regulatory subunit of bovine cAMP-PK, which binds cAMP but lacks kinase activity [24], resembles $p27^{3'\text{-orf}}$ in having $GGX_{aa}X_{aa}G$ as the glycine-rich sequence (GGSFG), while the third glycine residue of phos.b-K is replaced by a serine (fig.2).

A comparison of the predicted secondary structure of $p27^{3'\text{-orf}}$ surrounding the receptor-like and nucleotide-binding consensus sequences was conducted, using the algorithm of Garnier et al. [25]. The results are summarized in fig.3. The program predicts that the QRRQ consensus sequence of human IL-2R and $p27^{3'\text{-orf}}$ of HIV isolates with unmodified QRRQ sequence, favor the beginning of turn (T) regions (fig.3A,B) while the $GX_{aa}GX_{aa}X_{aa}G$ glycine-rich consensus of $p21^{ras}$ (fig.3D) and $GGX_{aa}X_{aa}G$ of $p27^{3'\text{-orf}}$ (fig.3B) are found at or close to the beginning of random coil (C) regions. Amino acid changes at the QRRQ site in other HIV isolates (fig.3C) resulted in a reversal of the configuration at that site. The ATP-binding

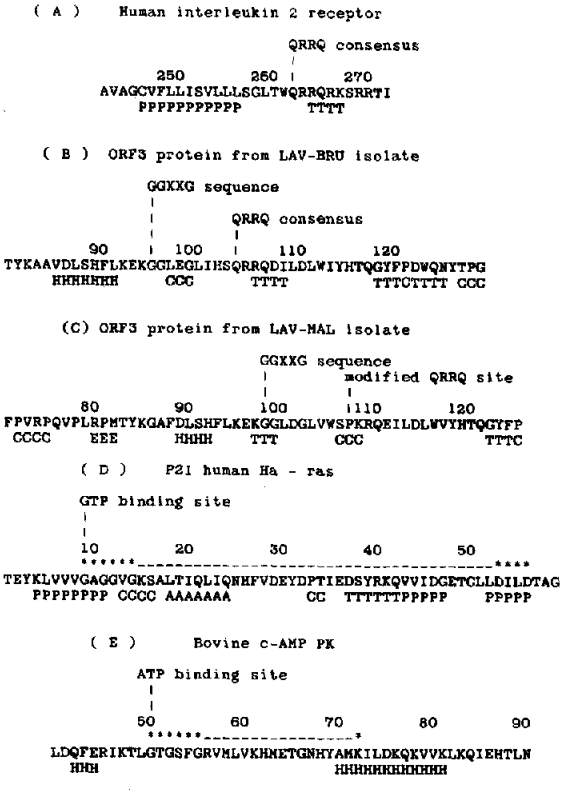

Fig.3. Predicted secondary structure analysis of QRRQ consensus and $GX_{aa}GX_{aa}X_{aa}G$ sequence in $p27^{3'\text{-orf}}$ compared to human IL-2R and kinase related proteins. Symbols in last lines of A–E denote the following secondary structural features: A, antiparallel β-sheet; C, random coil; H, α-helix; P, parallel β-sheet; T, turn. The regions indicated by 'star-dash' lines above the sequence mark the actual GTP (D) and ATP (E) binding sites of $p21^{ras}$ and cAMP-PK proteins, respectively.

region of the cAMP-PK is located between two α-helical (H) regions (fig.3D). The BαB unit structure originally proposed for the configuration at the dinucleotide-binding pocket of dehydrogenase enzymes was adapted as a model structure for the analogous site of the protein kinases [26–28]. The QRRQ consensus of the intracytoplasmic domain of human IL-2R occurs at a turn (T) region, since such turn configuration is located at the junction between transmembrane (TM) and peripheral areas of membrane receptor proteins [25,29]. Thus, the results of computer modeling summarized in fig.3 reveal different configurations around the QRRQ consensus of $p27^{3'\text{-orf}}$ proteins due to amino acid changes (fig.3B,C) within the sequence in some HIV isolates. These structural features therefore suggest, but do not prove, that the $p27^{3'\text{-orf}}$ protein of HIV possesses both a receptor-like phosphorylation domain similar to human IL-2R and a nucleotide-binding domain resembling that of the protein kinase family.

3. DISCUSSION

At present, no known function has as yet been ascribed to the 3′-orf gene of HIV. The protein product of this gene or its truncated analogue may act to repress viral replication [30,31] during HIV infection of T-cells in culture. The protein is relatively immunogenic in the native host [9,12], has been localized to the soluble cytoplasmic fraction of infected T-cells ([11], and Samuel, K., Showalter S. and Zweig, M., unpublished), and was reported to be modified at its N-terminus by myristylation [9], thus suggesting a possible membrane association [32]. The catalytic subunit of cAMP-PK and the protein phosphatase calcineurin are also soluble cellular proteins modified by myristylation [33,34]. It is of great interest that $p27^{3'\text{-orf}}$ more closely resembles the $p21^{ras}$ oncogene protein at its guanine nucleotide-binding site, with conservation of the DILD sequence and loss of the conserved lysine (K) residue in $p21^{ras}$ and 3′-orf protein of the various HIV isolates. Since $p21^{ras}$ is known to bind GTP, autophosphorylates, and possesses an intrinsic GTPase activity [23,35], homologous biochemical properties of $p27^{3'\text{-orf}}$ should be sought, notwithstanding differences in their predicted secondary structure around the DILD motif. The possibility of interaction between $p27^{3'\text{-orf}}$ with viral or cellular targets such as DNA, RNA, and proteins, and phosphorylation by cellular PK-C, should also be examined.

ACKNOWLEDGEMENTS

We thank Sue Toms for typing the manuscript. Research sponsored (in part) by the NCI, DHHS, under contract N01-CO-23910 with Program Resources, Incorporated. The contents of this publication do not necessarily reflect the views or policies of the DHHS nor does mention of trade names, commercial products, or organizations imply endorsements by the United States Government.

ADDENDUM

After submission of this manuscript for publication, the recently identified consensus sequence elements at the GTP-binding domain of guanine nucleotide-binding proteins [39] were also found to be conserved in the 3′-orf protein of HIV. The consensus GTP-binding domain in $p27^{3'\text{-orf}}$ is (6)GXXXXXXGK(14)---(38)DXXG(41)---(167)-NKXE(170). Numbers in brackets represent amino acid positions [5].

REFERENCES

[1] Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Danguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983) Science 220, 868–871.

[2] Popovic, M., Sarngadharan, M.G., Read, E. and Gallo, R.C. (1984) Science 224, 497–500.

[3] Levy, J.A., Hoffman, A.D., Kramer, S.M., Landis, J.A., Shimabukuro, J.M. and Oshiro, L.S. (1984) Science 225, 840–842.

[4] Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehom, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R. jr, Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Staal, F. (1985) Nature 313, 277–284.

[5] Wain-Hobson, W., Sonigo, P., Danos, O., Cole, S. and Alizon, M. (1985) Cell 40, 9–17.

[6] Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D. and Luciw, P.A. (1985) Science 227, 484–492.
[7] Muesing, M.A., Smith, D.H., Cabradilla, C.D., Benton, C.V., Lasky, L.A. and Capon, D.J. (1985) Nature 313, 450–458.
[8] Ratner, L., Starcich, B., Josephs, S.F., Hahn, B.H., Reddy, E.P., Livak, K.J., Petteway, S.R. jr, Pearson, M.L., Haseltine, W.A., Arya, S.K. and Wong-Staal, F. (1985) Nucleic Acids Res. 13, 8219–8229.
[9] Allan, J.S., Coligan, J.E., Lee, T.-H., McLane, M.F., Kanki, P.J., Groopman, J.E. and Essex, M. (1985) Science 230, 810–813.
[10] Franchini, G., Robert-Guroff, M., Wong-Staal, F., Ghrayeb, J., Kato, I., Chang, T.W. and Chang, N.T. (1986) Proc. Natl. Acad. Sci. USA 83, 5282–5285.
[11] Franchini, G., Robert-Guroff, M., Ghrayeb, J., Chang, N.T. and Wong-Staal, F. (1986) Virology 155, 593–599.
[12] Leonard, W.J., Depper, J.M., Crabtree, G.R., Rudikoff, S., Pumphrey, J., Robb, R.J., Kronke, M., Svetlik, P.B., Peffer, N.J., Waldmann, T.A. and Greene, W.C. (1984) Nature 311, 626–631.
[13] Nikaido, T., Shimizu, A., Ishida, N., Sabe, H., Teshigawara, K., Meada, M., Uchiyama, T., Yodoi, J. and Honjo, T. (1984) Nature 311, 631–635.
[14] Kohl, N.E., Legouy, E., DePinho, R.A., Nisen, P.D., Smith, R.K., Gee, C.E. and Alta, F.W. (1986) Nature 319, 73–77.
[15] Haniu, M., Armes, L.G., Yasunobu, K.T., Shastry, B.A. and Gunsalus, I.C. (1982) J. Biol. Chem. 257, 12664–12671.
[16] Shackelford, D.A. and Trowbridge, I.S. (1984) J. Biol. Chem. 259, 11706–11712.
[17] Gallis, B., Lewis, A., Wignall, J., Alpert, A., Moshizuki, D.Y., Cosman, D., Hopp, T. and Urdal, D. (1986) J. Biol. Chem. 261, 5075–5080.
[18] Barker, W.C. and Dayhoff, M.O. (1982) Proc. Natl. Acad. Sci. USA 79, 2836–2839.
[19] Hunter, T. and Copper, J.A. (1986) The Enzymes 17, 191–247.
[20] Zoller, M.J. and Taylor, S.S. (1979) J. Biol. Chem. 254, 8363–8367.
[21] Russo, M.W., Lukas, T.H., Cohen, S. and Staros, J.V. (1985) J. Biol. Chem. 260, 5205–5208.
[22] Chou, C.K., Dull, T.J., Russell, D.S., Gherzi, R., Lebwohl, D., Ulrich, A. and Rosen, O.M. (1987) J. Biol. Chem. 262, 1842–1847.
[23] Hattori, S., Ulsh, L.S., Halliday, K. and Shih, T.Y. (1985) Mol. Cell. Biol. 5, 1449–1455.
[24] Titani, K., Sasagawa, T., Ericsson, L.H., Kumar, S., Smith, S.B., Krebs, E.G. and Walsh, K.A. (1984) Biochemistry 23, 4193–4199.
[25] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[26] Rossmann, M.G., Liljas, A., Branden, C.-I. and Banaszak, L.J. (1975) The Enzymes 11, 61–102.
[27] Wierenga, R.K. and Hol, W.G.J. (1983) Nature 302, 842–844.
[28] Sternberg, M.J.E. and Taylor, W.R. (1984) FEBS Lett. 175, 387–392.
[29] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 222–245.
[30] Fisher, A.G., Ratner, L., Mitsuya, H., Marselle, L.M., Harper, M.E., Broder, S., Gallo, R.C. and Wong-Staal, F. (1986) Science 233, 655–659.
[31] Terwilliger, E., Sodroski, J.G., Rosen, C.A. and Haseltine, W.A. (1986) J. Virol. 60, 754–760.
[32] Garber, E.A., Cross, F.R. and Hanafusa, H. (1985) Mol. Cell. Biol. 5, 2781–2788.
[33] Carr, S.A., Blemann, K., Shoji, S., Parmelee, D.C. and Titani, K. (1982) Proc. Natl. Acad. Sci. USA 79, 6128–6131.
[34] Aitken, A., Cohen, P., Santikarn, S., Williams, D.H., Calder, A.G., Smith, A. and Klee, C.B. (1982) FEBS Lett. 150, 314–318.
[35] Sweet, R.W., Yokoyama, S., Kamata, T., Feramisco, J.R., Rosenberg, M. and Gross, M. (1984) Nature 311, 273–275.
[36] Parker, P.J., Coussens, L., Totty, N., Rhee, L., Young, S., Chen, E., Stabel, S., Waterfield, M.D. and Ullrich, A. (1986) Science 233, 853–859.
[37] Yuasa, Y., Srivastava, S.K., Dunn, C.Y., Rhim, J.S., Reddy, E.P. and Aaronson, S.A. (1983) Nature 303, 775–779.
[38] Desai, Kalyanaraman, V.S., Casey, J.M., Srinivasan, A., Andersen, P.R. and Devare, S.G. (1986) Proc. Natl. Acad. Sci. USA 83, 8380–8384.
[39] Dever, T.E., Glynias, M.J. and Merrick, W.C. (1987) Proc. Natl. Acad. Sci. USA 84, 1814–1818.