Minireview

Characterization of a novel protein-binding module – the WW domain

Marius Sudol[a,*], Henry I. Chen[a], Cecile Bougeret[a], Aaron Einbond[a], Peer Bork[b]

[a]*Laboratory of Molecular Oncology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA*
[b]*EMBL, 69012 Heidelberg and Max-Delbruck-Center for Molecular Medicine, 13122 Berlin-Buch, Germany*

Received 10 May 1995

Abstract **We have identified, characterized and cloned human, mouse and chicken cDNA of a novel protein that binds to the Src homology domain 3 (SH3) of the Yes proto-oncogene product. We subsequently named it YAP for Yes-associated protein. Analysis of the YAP sequence revealed a protein module that was found in various structural, regulatory and signaling molecules. Because one of the prominent features of this sequence motif is the presence of two conserved tryptophans (W), we named it the WW domain. Using a functional screen of a cDNA expression library, we have identified two putative ligands of the WW domain of YAP which we named WBP-1 and WBP-2. Peptide sequence comparison between the two partial clones revealed a homologous proline-rich region. Binding assays and site-specific mutagenesis have shown that the proline-rich motif binds with relatively high affinity and specificity to the WW domain of YAP, with a preliminary consensus that is different from the SH3-binding PXXP motif. This suggests that the WW domain has a role in mediating protein–protein interactions via proline-rich regions, similar but distinct from Src homology 3 (SH3) domains. Based on this finding, we hypothesize that additional protein modules exist and that they could be isolated using proline-rich peptides as functional probes.**

Key words: Protein–protein interaction; Protein module; Polyproline

1. Background and rationale – Retrospective look

What started as a pilot project ended up being the main focus of an entire lab. Our general aim has been to decipher molecular steps of signaling by the Yes proto-oncogene product which represents a non-receptor type protein-tyrosine kinase of the Src family [1]. The specific goal was to isolate substrates and regulators of the Yes kinase in order to understand at least some aspects of the molecular role it plays in normal and the viral-Yes oncogene transformed cells [2]. Initially, our experimental approaches were descriptive of nature [3,4]. We argued that by characterizing the pattern of Yes expression in various tissues and cells, we would have been able to find a common denominator that would provide a clue regarding the physiological function of the Yes protein [5]. Localization of the Yes proto-oncogene transcript and protein in cerebellar Purkinje cells was an exciting finding which gave us hope for functional clues [6]. Unfortunately, that was immediately followed by the frustration of trying to study the Yes kinase in difficult experimental systems of isolated Purkinje neurons or cerebellar slices. The direction of our research activities shifted swiftly when Hirai and Varmus proposed that the amino termini of Src family kinases form complexes with cellular proteins and that these apparently transient and dynamic complexes constitute a part of the mechanism by which Src, Yes and other kinases signal [7]. The SH2 domain, residing at the amino terminal half of the Src kinases had already been delineated at that time and was a primary candidate for a signaling domain [8]. The proposal of Hirai and Varmus had two attractive elements. (i) It agreed with a number of observations from other labs including that of Hanafusa and colleagues [9]. (ii) Implicitly, it promised a molecular explanation for specificity in action for members of the Src family whose sequences diverged most at the very amino terminal regions (specifically at their 'finger print' [1] or unique domains).

In order to isolate proteins that interact with the Yes kinase, we decided to use anti-idiotypic antibodies directed to the amino terminal domain of the Yes protein expressed in bacteria [10,11]. Although controversial and risky as far as the well-defined molecular probes are considered, the anti-idiotypic antibodies have been known in some instances to mimic, perhaps in most of the cases only partially, the structure of the original epitope [12]. Our original antigen was composed of the unique and SH3 domain of the Yes protein expressed in bacteria. Its epitope seemed to reside mostly in the SH3 domain of Yes (see the discussion in [11]). Using the anti-idiotypic antibodies and lambda phage expression library from cerebellar Purkinje cells, we isolated one cDNA clone that encoded a new proline-rich protein that we named Yes-Associated Protein (YAP) [11]. We showed that YAP interacts with the SH3 domain of Yes and other signaling molecules through a proline rich sequence [11]. It was through the detailed characterization of YAP that we obtained clues to a novel protein module that we named the WW domain.

2. YAP is a proline-rich phosphoprotein that binds to SH3 domains

YAP is of 65 kDa molecular mass, it is phosphorylated in vivo on serine and is particularly rich in proline [11]. Within the YAP sequence we have identified a motif, PVKQPPPLAP that is in agreement with the consensus proposed by Schreiber and

*Corresponding author. Fax: (1) (212) 327 7582.
E-mail: SUDOL@HCX.ROCKEFELLER.EDU

Abbreviations: GTP, guanosine triphosphate; kDa, kiloDalton; MAP, Microtubule-associated protein; P, Proline; PH, pleckstrin homology; PID, Phosphotyrosine-interacting domain; PTB, phosphotyrosine-binding domain; SH, Src homology; W, tryptophan; WBP, WW domain-binding protein; www, world wide web; YAP, Yes-associated protein of 65 kDa.

SSDI 0014-5793(95)00550-1

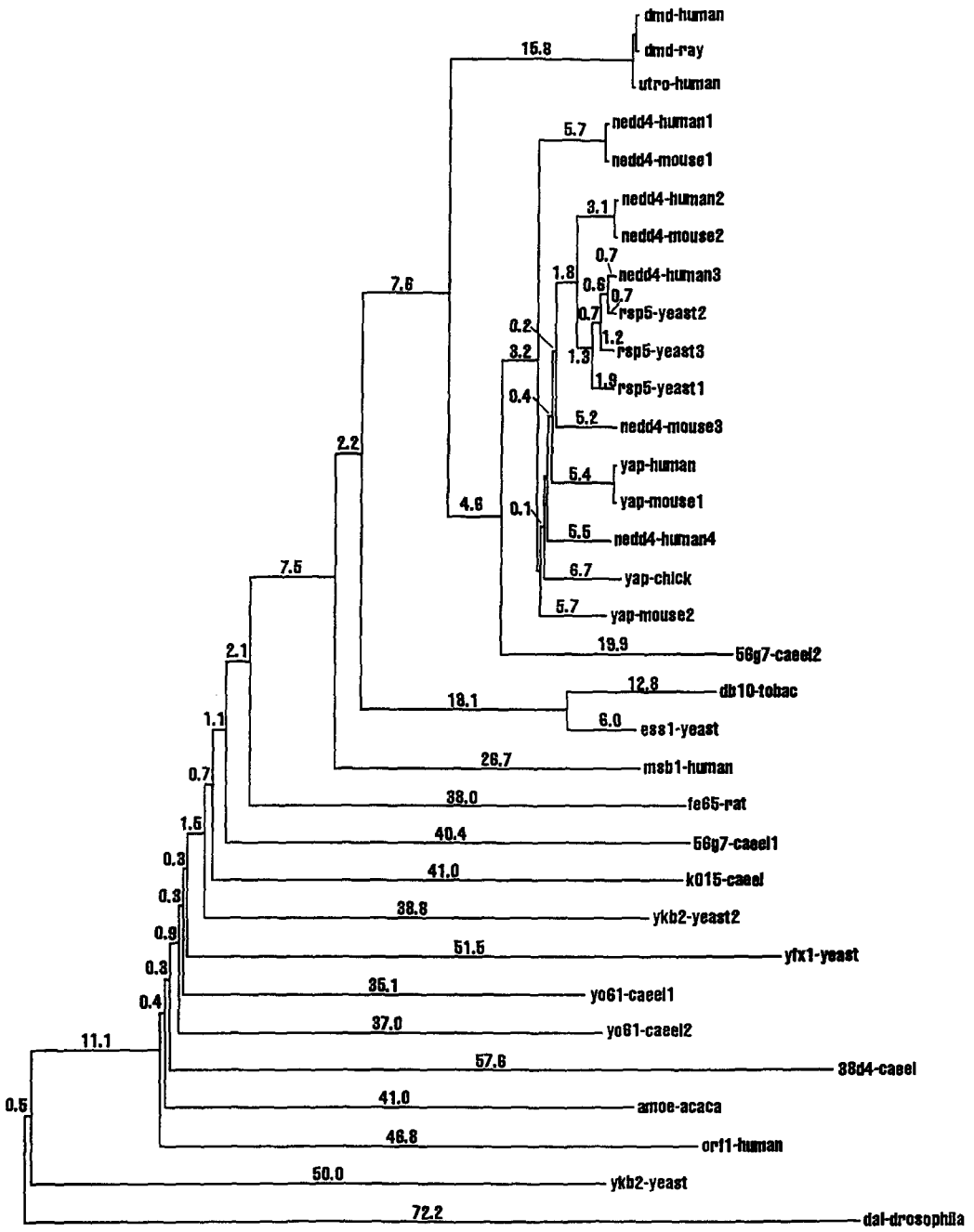

Fig. 1. Phylogenetic tree of the selected WW domains [14,15] constructed using the method of progressive sequence alignment [30]. Instead of comparing the entire domains, variable residues (see the consensus) were omitted to reduce the background. The numbers represent the proportional lengths of the horizontal line segments, or the phylogenetic distance between sequences. Line segments without numbers represent scores of −0.0, or a negligible distance (often sequences with identical consensus).

colleagues for the 'Class I Sequences' of SH3-binding motifs with **PXXP** motif predicted at positions 7 and 10: PVKQPPPLAP [13]. Competition assays with synthetic peptides showed the involvement of the predicted proline-rich sequence in binding between YAP and the Yes kinase. Interestingly, YAP was also shown to bind other signaling molecules that contain SH3 domains including Nck, Crk and Src, and to lesser extent to the SH3 domains of Abl and GAP (guanosine triphosphatase activating protein). The specific sites of serine phosphorylation on YAP are being mapped at present. One of the attractive candidates for the kinase phosphorylating YAP is the MAP kinase. A consensus site for the MAP kinase phosphorylation is present just at the carboxyl end of the proline-rich domain of YAP. We do not know whether YAP phosphorylation could affect its binding to SH3 domains. However, it is tempting to speculate that a negatively-charged phosphate group on serine residue located close to the polyproline domain could in fact modulate (disrupt or enhance) the proline-rich motif-SH3 domain interaction in vivo as a means of regulating the process of signal propagation.

```
Yap/Mouse1    VPLPAGWEMAKTSS-GQRYFLNHNDQTTTWQDPRKAMLS   X80508
Yap/Mouse2    GPLPDGWEQAMTQD-GEVYYINHKNKTTSWLDPRLDPRF   X80508
Dmd/Human     TSVQGPWERAISPN-KVPYYINHETQTTCWDHPKMTELY   P11532
WW consensus    LPtGWE    ttt Gt YYhNH TtTTtW tPt
Orf2/Human    RTVNEPWTMGFSKSFKKKFFYNKKTKDSTFDLPADSIAP   D43949
Yf43/Human    YDSADDWSEHISSS-GKKYYYNCRTEVSQWEKPKEWLER   R16603
Yg46/Human    XXLPPXWEKRMSRSSGRVYYFNHITNASQWERPSGNSSS   R20490
X132/Human    .........VWDENTGCYYYWNTQTNEVTWELPQYLATQ   Z36787
```

Fig. 2. Alignment of the WW domains of mouse YAP, human dystrophin and 4 new entries retrieved from the data base. Database accession numbers are given in the last column. The consensus line shown is based on the alignment of all previously described WW modules (h, hydrophobic position; t, turn-like or polar position). For more information about the new and other WW domains and their protein products see 'The WW Domain Page' available from the www network (the address is in the text).

In chicken embryo fibroblasts, YAP protein was localized by immunofluorescent staining to cytoplasm and nucleus (Sudol, M., unpublished). Interestingly, the cytoplasmic staining with YAP antibodies was reminiscent of the picture obtained with Yes antibodies – the staining patterns suggested colocalization of Yes and YAP proteins to the membranes of the endoplasmic reticulum. Although the nuclear staining of YAP was distinct and was observed with two different preparations of the affinity purified IgG, the final conclusion regarding its localization requires further confirmation by cell fractionation.

Human and mouse orthologues of the chicken YAP were cloned and characterized [14]. The analysis of coding sequences together with the data of Southern blot analysis revealed that the human YAP gene is well conserved among higher Eukaryotes but may not be conserved in yeast [14]. The expression of YAP at the RNA level in adult human tissues is nearly ubiquitous, being relatively high in placenta, prostate, ovary and testis, but is not detectable in peripheral blood leukocytes. Using fluorescence in situ hybridization on human metaphase chromosomes and by analyzing rodent-human hybrids by Southern blot hybridization and PCR amplification, we mapped the human YAP gene to chromosome band 11q13, a region to which the multiple endocrine neoplasia type 1 gene has been mapped [14]. Taken together, the relatively high expression of YAP in tissues regulated predominantly by steroid hormones, the presence of YAP protein in the cytoplasm and in the nucleus, plus the possible involvement of YAP in the multiple endocrine neoplasia, a syndrome affecting polypeptide hormone producing tissues, could suggest that YAP is a part of signaling pathways for steroid and/or polypeptide hormones.

3. Murine YAP provides a clue for the WW domain

Sequence comparison between mouse, human, and chicken YAP proteins showed an inserted sequence in the mouse YAP that represented a non-identical repeat of the upstream sequence [14]. Further analysis of this sequence revealed that it is a new protein module: the WW domain (named after two tryptophans that mostly contributed to the signal in sequence comparisons) [15]. Shortly thereafter, two other groups independently reported the identification of this conserved domain [16,17]; the widespread occurrence in otherwise different intracellular regulatory proteins and the arrangement in tandem arrays of up to 4 copies within a protein (for an overview see [14,15,18]) justify the term module [19]. The domain appears to contain beta-strands grouped around four conserved aromatic positions ([15,17], H. Oschkinat, personal communication). Other important features of the domain are a high content of polar amino acids and the presence of prolines distributed preferentially towards both termini of the linear sequence. One of the C-terminal prolines seems strictly invariant. The length of the WW domain is aproximately 38 residues as suggested by the length of the second WW module (the insert) identified in the mouse ortholog of YAP [14]. As NMR spectra of the WW domain of YAP (collaboration with laboratories of Hartmut Oschkinat and Matti Saraste, EMBL, Heidelberg, Germany) indicate a globular structure, the size of the WW module appears relatively small compared with SH2, SH3 or PH domains.

Examination of the primary sequences indicated that WW domains of YAP [14], Nedd-4 [20], and Rsp5 [21] show more similarity to each other than to WW domains of other proteins (Fig. 1). It is likely that these domains share certain functional features; for example, they could interact with similar ligands or localize the proteins to similar cellular compartments. When the repeats of the WW domains within the same protein are examined, the second or third WW domain does not necessarily show as high a sequence similarity to the first WW domain as one would expect from a recent evolutionary duplication event, but does show a high similarity to WW domains of other proteins. For example, the second domain in mouse YAP is more similar to one of the WW domains of the yeast Rsp5 gene product than to the first domain in mouse YAP (Fig. 1). This suggests that multiple WW domains within the same molecule may not be redundant but could have evolved to carry out subtly diverged functions.

After the initial delineation of the WW domain in animals [15–17], numerous additional sequences appeared in sequence databases that harbored a WW module [14,18]. Our most recent database searches identified yet another 4 distinct proteins that host a WW domain (Fig. 2). Thus, as many as 20 distinct proteins (taken species redundancy aside) have been described so far to contain this conserved module. Anticipating the number of WW sequences to grow rapidly, we have been providing updated information on the WW domain via world wide web (www) (http://www.embl-heidelber.de/~bork/ww1.html) since December 1994, including both the alignment and a diagram with modular structure of the corresponding proteins.

The current list of proteins with the WW motifs confirmed our initial conclusion that like SH2, SH3 and PH domains [22], the WW domain occurs in a variety of structural and signaling molecules with no apparent common functions [15]. Three proteins that contain the WW domain could provide a clue to one of the roles of this module in major signaling pathways. One of the human proteins, named ORF1, [database accession number D29640], contains a WW domain just upstream from the

C-terminal sequence that shows similarity to yeast Ras GTPase activator protein. A nematode (*C. elegans*) protein named 38D4 [Z46241] harbors a WW domain at the amino terminus, followed by a PH domain in the middle and a C-terminal sequence that is conserved in the breakpoint cluster region (BCR), n-chimaerin, and p85 subunit of the phosphoinositol 3-kinase. A gene product named Msb1, with one WW domain, was isolated in a genetic screen in yeast and was implicated in the MAP kinase pathway (K. Matsumoto, personal communication). Taken together these data suggest an involvement of the WW domain in the Ras and/or MAP kinase signaling pathways.

The function of the WW domain in YAP and in other proteins remains to be determined. The occurrence of the domain in yeast proteins provides a powerful genetic system that could be employed to analyze the function of the WW motif in vivo.

The Rsp5 gene was identified as a suppressor of mutations in the SPT3 gene, which encodes a transcription factor that interacts with TATA-binding protein [21]. It is unlikely that Rsp5 and SPT3 proteins interact directly, since Rsp5 mutations suppress a deletion of SPT3. It is more likely that this interaction is indirect because the Rsp5 gene was recently isolated by researchers in several other labs studying various aspects of cytoplasmic signaling in yeast. One of the explanations for this apparent diversity of roles in signaling could come from biochemical studies of Rsp5, showing that its catalytic domain (designated HECT) can form a high energy thio-ester bond with ubiquitin, therefore proving that the Rsp5 protein is indeed a ubiquitin-protein ligase [23,24]. It is likely that Nedd-4, a gene whose expression is modulated during early development of the central nervous system, encodes the similar enzymatic activity as Rsp5. Since ubiquitination is directly related to protein metabolism, and the WW domains in Rsp5 and Nedd-4 could be considered as molecular adhesive to anchor the ligase to the appropriate targets, one could speculate that a ligand for the WW domain, in general, is of proteinaceous nature and that this domain represents a module mediating protein–protein interaction. Our recent data on the isolation of ligands for the WW of YAP confirm this assumption.

4. The WW domain of YAP binds a novel proline-rich ligand

Using a functional screen of a cDNA expression library, we have identified two putative ligands of the WW domain of YAP which we named WBP-1 and WBP-2 [25]. Peptide sequence comparison between the two partial clones revealed a homologous region consisting of a proline-rich domain. Interestingly, binding assays and site-specific mutagenesis have shown that the 10 amino acid long proline-rich domain binds with relatively high affinity and specificity to the WW domain of YAP, with the preliminary consensus that clearly differs from the PXXP motif [25]. Accordingly, WBP-1 does not bind to the SH3 domains of arbitrarily chosen proteins including SH3 domain of Yes, Fyn, Abl and GAP. We have also shown that bacterially expressed WW domain of YAP is able to precipitate one or two low-molecular-weight proteins from organ or cell culture lysates [25]. Using similar approaches, our data also suggests a protein that specifically interacts with the WW domain of dystrophin. Judging from the molecular-weight determination, the ligand for the WW of dystrophin differs from the WBP-1 and WBP-2 proteins (Bougeret, C., Chen, H. and Sudol, M., unpublished). The data stronly supports a role of the WW domain in mediating specific protein–protein interactions and indicates a binding activity of proline-rich ligands distinct from Src homology (SH3) domains. One has to await structural data to compare the binding mode with the SH3-PXXP complex and the poly-proline interaction described recently for profilin [26].

5. Significance and possible ramifications for medicine

A new group of modular tyrosine kinase-associated proteins with multiple protein-binding domains and no apparent catalytic activity has gained considerable interest. These proteins, including Crk, Shc, Nck and Grb2, act as adaptor molecules that are charged with the responsibility of bringing together various components of a pathway by virtue of protein-binding domains such as SH2 and SH3 in order to propagate the signal. YAP, although still unknown in function, possesses not only a WW domain but also a proline-rich domain that binds to the SH3 domain of Yes. The FE65 factor cloned and characterized by Tommaso Russo laboratory at the University of Naples [27] contains one WW domain and two PID/PTB (phosphotyrosine-interacting/-binding) domains [28]. YAP and FE65 may in fact represent examples of a new family of adaptor molecules [25,28]. Nonetheless, at this time the biological significance of the interaction between the WW domain and the proline-rich motif of the WBP-1 is not clear. The genetic manipulation of the WW module in dystrophin, a molecule that has been implicated in a specific disease phenotype, (Duchenne's and Becker's muscular dystrophy) and other genetic approaches to analyze the WW domain of the yeast protein Rsp5 should provide us with useful biological analogies and perhaps with strategies to control certain muscular diseases in humans.

6. Future trends

Since the WW domain and the recently characterized PID/PTB domain [28] represent fairly large modular families, it is likely that many more of such modules and their families will be found in the future [29,19]. The history of the WW domain shows that the characterization of a module by a combination of sensitive sequence analysis methods and various experimental approaches is extremely powerful and can guide further exploration of the involved proteins and pathways; one obvious approach that is apparent from our data is to use various proline- rich polypeptides as functional probes in search of new, biologically relevant, protein modules.

Acknowledgements: We thank all members of the Laboratory of Molecular Oncology for stimulating discussions and Drs. Hidesaburo Hanafusa, Matti Saraste and Hartmut Oschkinat for valuable advice. The work from our laboratory that was reviewed in this paper was supported by grants from: The National Cancer Institute (CA01605), The Council for Tobacco Research USA (3035), The Klingenstein Award, and by the Human Frontier Science Program Grant (to M.S.) and by NIH Grant (5T32GM07739-16) (to H.I.C.).

References

[1] Sudol, M. (1993) in: The Molecular Basis of Human Cancer (Neel, B. and Kumar, R., Eds.) Futura, New York, pp. 203–224.
[2] Sudol, M. (1990) Exp. Med. 8, 94–100.
[3] Sudol, M. and Hanafusa, H. (1986) Mol. Cell. Biol. 6, 2839–2846.

[4] Sudol, M., Alvarez-Buylla, A. and Hanafusa, H. (1988) Oncogene Res. 2, 345–355.
[5] Zhao, Y-H., Krueger, J. and Sudol, M. (1990) Oncogene 5, 1629–1635.
[6] Sudol, M., Kuo, F.C., Shigemitsu, L. and Alvarez-Buylla, A. (1989) Mol. Cell. Biol. 9, 4545–4549.
[7] Hirai, H. and Varmus, H. (1990) Mol. Cell. Biol. 10, 1307–1318
[8] Sadowski, I., Stone, J.C. and Pawson, T. (1986) Mol. Cell. Biol. 6, 4396–4408.
[9] Hanafusa, H. (1989) in: Proceedings of XXXIII Conference on Membrane Proteins: Targeting and Transduction, Houston, TX, pp. 237–245.
[10] Jerne, N.K. (1974) Ann. Immunol. (Inst. Pasteur) 125C, 373–389.
[11] Sudol, M. (1994) Oncogene 9, 2145–2152.
[12] Strosberg, A.D. (1989) Methods Enzymol. 178, 265–275
[13] Feng, S., Chen, J.K., Yu, H., Simon, J.A. and Schreiber, S.L. (1994) Science 266, 1241–1247.
[14] Sudol, M., Bork, P., Einbond, A., Kastury, K., Druck, T., Negrini, M., Huebner, K. and Lehman, D. (1995) J. Biol. Chem. 270, June issue.
[15] Bork, P. and Sudol, M (1994) Trends Biochem. Sci. 19, 531–533.
[16] André, B. and Springael, J.-Y. (1994) Biochem. Biophys. Res. Commun. 205, 1201–1205.
[17] Hofmann, K. and Bucher, P. (1995) FEBS Lett. 358, 153–157.
[18] Bork, P. and Sudol, M. (1995) Trends Biochem. Sci. 20, 54.
[19] Bork, P. and Bairoch, A. (1995) Trends Biochem. Sci. 20, poster March issue.
[20] Kumar, S., Tomooka, Y. and Noda, M. (1992) Biochem. Biophys. Res. Commun. 185, 1155–1161.
[21] Eisenmann, M.D., Arndt, K.M., Ricupero, S.L., Rooney, J.W. and Winston, F. (1992) Genes Dev. 6, 1319–1332.
[22] Musacchio, A., Wilmanns, M. and Saraste, M. (1994) Prog. Biophys. Mol. Biol. 61, 283–297.
[23] Scheffner, M., Huibregtse, J.M., Vierstra, R.D. and Howley, P. (1993) Cell 75, 495–505.
[24] Huibregtse, J.M., Scheffner, M., Beaudenon, S. and Howley, P. (1995) Proc. Natl. Acad. Sci. USA, in press.
[25] Chen, H.I. and Sudol, M. (1995) Proc. Natl. Acad. Sci. USA, in press.
[26] Cedergren-Zeppezauer, E.S., Goonesekere, N.C.W., Rozycki, M.D., Myslik, J.C., Dauter, Z., Lindberg, U. and Schutt, C.E. (1994) J. Mol. Biol. 240, 459–475.
[27] Duilio, A., Zambrano, N., Mogavero, A.R., Ammendola, R., Cimino, F. and Russo, T. (1991) Nucleic Acid. Res. 19, 5269–5274.
[28] Bork, P. and Margolis, B. (1995) Cell 80, 693–694.
[29] Cohen, G.B., Ren, R. and Baltimore, D. (1995) Cell 80, 237–248.
[30] Feng, D.-F. and Doolittle, R.F. (1987) J. Mol. Evol. 25, 351–360.