Protein Science (1998), 7:1250–1254. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Homologues of 26S proteasome subunits are regulators of transcription and translation

L. ARAVIND[1,3] AND CHRIS P. PONTING[2]

[1]Department of Biology–BSBW, Texas A&M University, College Station, Texas 77843
[2]University of Oxford, The Old Observatory, Fibrinolysis Research Unit, South Parks Road, Oxford, OX1 3RH, United Kingdom

(RECEIVED October 27, 1997; ACCEPTED February 11, 1998)

Abstract: Single copies of an α-helical-rich motif are demonstrated to be present within subunits of the large multiprotein 26S proteasome and eukaryotic initiation factor-3 (eIF3) complexes, and within proteins involved in transcriptional regulation. In addition, p40 and p47 subunits of eIF3 are shown to be homologues of the proteasome subunit Mov34, and transcriptional regulators JAB1/pad1. Finally, the proteasome subunit S5a and the p44 subunit of the basal transcription factor IIH (TFIIH) are identified as homologues. The presence of homologous, and sometimes identical, proteins in contrasting functional contexts suggests that the large multisubunit complexes of the 26S proteasome, eIF3 and TFIIH perform overlapping cellular roles.

Keywords: eukaryotic initiation factor-3 subunits; Fus6; Mov34; PINT motif; proteasome subunit S5a; transcription factor IIH subunits

Proteasomes are responsible for the selective degradation of intracellular proteins in eukaryotic cells (Coux et al., 1996; Hilt & Wolf, 1996). Proteasome substrates include metabolic enzymes, cell-cycle control factors, transcriptional regulators and mature forms of antigenic peptides. Many of these are targeted for proteolysis by ubiquitination. Two components contribute to the eukaryotic 26S proteasome (2,000 kDa) (Coux et al., 1996): (a) the 20S (700 kDa) proteasome, thought to resemble in structure and in function the 20S proteasome of *T. acidophilum* (Löwe et al., 1995), and (b) a 19/22S regulator containing at least 18 proteins with molecular weights between 25 and 110 kDa. The regulator complex appears to present ubiquitinated proteins to the 20S complex for digestion following their association with subunit 5a (Deveraux et al., 1994; van Nocker et al., 1996).

Reprint requests to: Christopher P. Ponting, University of Oxford, The Old Observatory, Fibrinolysis Research Unit, South Parks Road, Oxford, OX1 3RH, United Kingdom; e-mail: Ponting@Bioch.ox.ac.uk.

[3]Current address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bldg. 38A, Bethesda, Maryland 20894.

Understanding the structure, function, and evolution of the multisubunit and multifunctional proteasome represents a considerable challenge. One of many approaches that may be used to investigate the proteasome's form and function is the detailed analysis of subunits' amino acid sequences. We have subjected the known sequences of 26S proteasome subunits to local alignment and Hidden Markov model (HMM) analyses and present evidence that homologues of 26S proteasome subunits participate in the regulation of transcription and translation initiation. Three families of domains were found to be represented among regulators of proteasome, transcription, and translation functions. These are: an α-helix-rich domain present in p48 and p110 subunits of eIF3, a Mov34-related domain present in p47 and p40 subunits of eIF3, and the S5a-like domain found in the p44 subunit of TFIIH (summarized in Table I).

Sequence analyses: 26S proteasome subunit sequences were used as queries in Ssearch (Pearson, 1991) and gapped BLAST (Altschul et al., 1997) searches of nonredundant amino acid databases. Putative homologues with significant pairwise similarities ($E < 10^{-4}$) were aligned using ClustalW (Thompson et al., 1994). Hidden Markov models were calculated from these alignments and compared, in an iterative manner, with databases (Eddy et al., 1995). Sequences scoring >28 bits (or >35 bits for the α-helical-rich PINT motif) were considered to be homologues and were added to the query alignment for subsequent iterations. In addition, proteasome sequences were subjected to position-specific iterative BLAST (PSI-BLAST) (Altschul et al., 1997) searches using an *E*-value threshold of 0.005.

PINT: A motif in Proteasome subunits, Int-6, Nip-1, and TRIP-15: Database searches with the human 26S proteasome p44.5 (subunit 9) sequence revealed significant similarities (BLASTP2, $E < 10^{-9}$; Ssearch, $E < 10^{-8}$) with *Caenorhabditis elegans* and *Saccharomyces cerevisiae* hypothetical proteins, and with a putative thyroid receptor interacting protein from *Drosophila*, termed alien (Goubeaud et al., 1996). Additional significant similarity was detected for *Arabidopsis thaliana* Fus6 (also called COP11) (BLASTP2, $E = 7 \times 10^{-4}$; Ssearch $E = 3 \times 10^{-3}$). Reciprocal

Table 1. *Mammalian homologues of 26S proteasome subunits*[a]

	26S Proteasome subunits	Regulation of transcription	Regulation of translation
PINT family	p44.5 (subunit 9), p55, P91A/S3	TRIP15 (thyroid-hormone receptor interacting protein 15)	eIF3p48 (Int-6), eIF3p110
Mov34 family	Mov34 (subunit S12, p40)	JAB1	eIF3p40, eIF3p47
S5a family	S5a	Basal transcription factor IIH p44, S5a	—

[a]These are distinguished between those found by experiment to be regulators of transcription and others found to be regulators of translation. Although no mammalian S5a homologue is known to regulate translation, Ssl1, a yeast member of the family, is known to be essential for translation initiation (Yoon et al., 1992). Alternative names for proteins are given in parentheses. References and additional abbreviations are given in the text.

searches with Fus6-like sequences (human Gps1 and KIAA0107, yeast YPR108w, *C. elegans* F49C12.8, and *Schizosaccharomyces pombe* C19G10.05) provided further evidence that Fus6-like and p44.5-like molecules are homologues (not shown). These proteins exhibit only a single region of significant similarity, of length 80–95 amino acids, as assessed using MACAW- (Schuler et al., 1991) and ClustalW-derived (Thompson et al., 1994) alignments. Four iterations of database searching using HMMer (Eddy et al., 1995) and an HMM derived from this region of similarity was sufficient to detect the majority of the putative homologues shown in Figure 1. Four remaining sequences (*C. elegans* T06D8.8 and K08F11.3, and *S. cerevisiae* YIL071w and YOR427w) were identified using PSI-BLAST searches with Fus6- and p44.5-like query sequences.

The predominantly α-helical PINT motif (Fig. 1) is seen in three mammalian 26S proteasome subunits, namely p44.5 (Hoffman & Rechsteiner, 1997), p55 (T. Watanabe et al., EMBL accession AB003103), and P91A/S3, a tumor transplantation antigen (Lurquin et al., 1989), which is associated with the mammalian 20S proteasome (DeMartino et al., 1994). *S. cerevisiae* Sun2, a P91A orthologue, is known to be a suppressor of NIN1, a component of the 26S proteasome (Kawamura et al., 1996). This suggests proteasomal functions of Sun2 orthologues (Kawamura et al., 1996) rather than previously proposed diphenol oxidase activities (Pentz & Wright, 1991). In addition, a plant Sun2 homologue is localized to the nucleus and shows a cell-cycle dependent variation in levels (Smith et al., 1997). This suggests that these homologues are involved in cell cycle stage specific regulation of proteasome function.

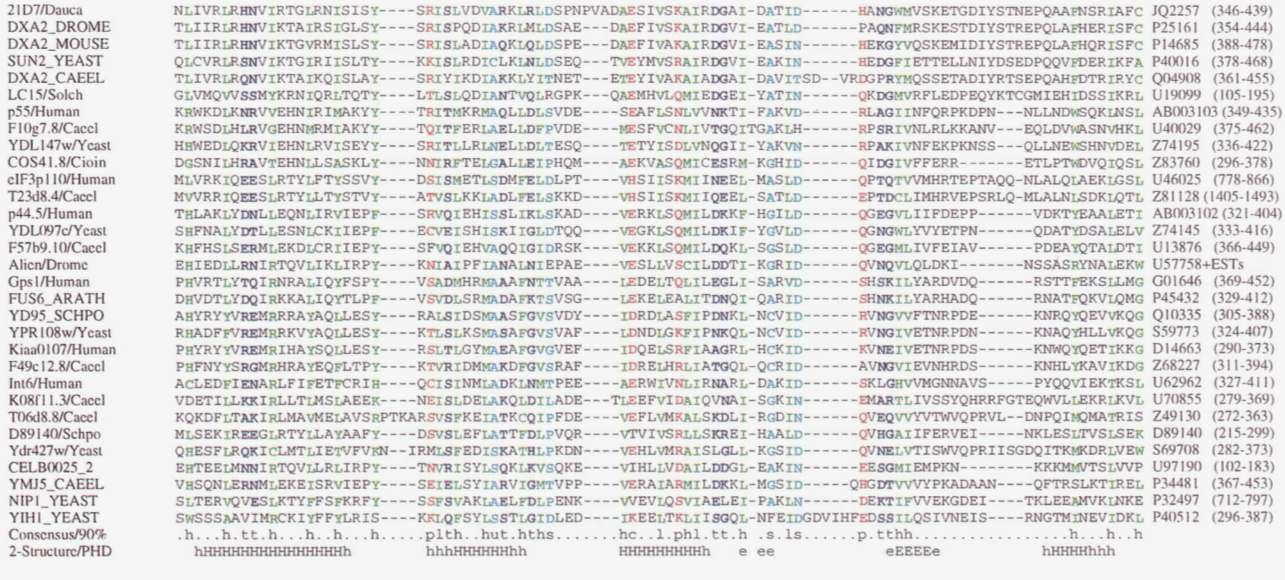

Fig. 1. Multiple alignment of PINT motifs. Amino acids are colored according to a 90% consensus (shown beneath the alignment): a, aromatic (green; FHWY); c, charged (red; DEHKR); h, hydrophobic (green; ACFGHIKLMRTVWY); l (green; ILV); o (magenta; ST); p, polar (red; CDEHKNQRST); s, small (cyan; ACDGNPSTV); t, turn-like (blue; ACDEGHKNQRST); u, tiny (cyan; AGS); +, positively charged (red; HKR); and, − negatively charged (red; DE). Predicted secondary structure (Rost & Sander, 1993) is shown beneath the alignment [H/h denotes an α-helix and E/e a β-strand with an expected accuracy higher than 82% (upper case)/72% (lower case)]. Expressed sequence tags partially encoding PINT motifs have been omitted from the alignment. These are: H24402 and AA233250 (human), W75295 and W54432 (mouse), T02119 (*C. elegans*), W43761 (*A. thaliana*), C27458 and C26812 (rice), and AA520167 (*Toxoplasma gondii*). The sequence of *Drosophila* alien has been extended using overlapping ESTs, including AA391270. Consensus sequences were calculated using all homologous sequences, including ESTs. PIR, EMBL or SwissProt database accession codes and residue numbers are shown following the alignment. A previous proposal of *E. coli* BirA-like helix-turn-helix motifs in Fus6-like proteins (Mushegian & Koonin, 1996) could not be corroborated using methods described in the text. Species: ARATH, *Arabidopsis thaliana*; CAEEL, *Caenorhabditis elegans*; CIOIN, *Ciona intestinalis*; DAUCA: *Daucus carota* (carrot); DROME, *Drosophila melanogaster*; SCHPO, *Schizosaccharomyces pombe*; SOLCH, *Lycopersicon chilense*; and YEAST, *Saccharomyces cerevisiae*.

Unexpectedly, PINT motifs were found in two human eukaryotic initiation factor 3 (eIF3) subunits, eIF3p48 and eIF3p110. eIF3p48, also called Int-6 (Hershey et al., 1996), appears to mediate other functions that are distinct from translation initiation since it has been found as a component of chromatin-associated PML complexes (Everett et al., 1997), unless bound to the HTLV-I Tax oncoprotein when it is redistributed to the cytoplasm (Desbois et al., 1996). The second largest subunits in yeast and human eIF3 (NIP1 and eIF3p110, respectively) (Naranda et al., 1996; Asano et al., 1997) also contain the PINT motif.

PINT motif-containing proteins also function as transcriptional mediators. *Drosophila* alien protein (Goubeaud et al., 1996) is a close homologue of both 26S proteasome subunit p44.5 and human TRIP15, a rat thyroid-hormone receptor-interacting protein that is likely to act as a negative regulator of transcription (Lee et al., 1995). This suggests that alien p44.5 and TRIP15 regulate two distinct cellular functions: (a) transcriptional regulation and (b) 26S proteasome-mediated protein degradation. This would not be unprecedented since Sug1/TRIP1, a thyroid-hormone receptor-interacting protein and 26S proteasome subunit, possesses both such functions (Lee et al., 1995; Swaffield et al., 1995; Rubin et al. 1996).

The remaining PINT motif-containing proteins include Fus6, known to be a component of a multiprotein complex in the nucleus that participates in a plant photomorphogenesis pathway (Castle & Meinke, 1994; Staub et al., 1996), and a human homologue, GPS1, which is seen to suppress lethal G-protein subunit activating mutations in the yeast pheromone response pathway (Spain et al., 1996).

Mov34 is a homologue of both eIF3p40 and eIF3p47: A second homologous domain family was found to be represented among both proteasomal and eIF3 subunits. Ssearch and PSI-BLAST database searches with murine 26S proteasome subunit Mov34 (subunit S12, p40) (Tsurumi et al., 1995) homologues showed significant similarities with two eIF3 subunits, eIF3p40 and eIF3p47 (gapped BLASTP: $E = 4 \times 10^{-10}$ [query: human p40, hit: eIF3p47], and $E = 2 \times 10^{-8}$ [query: *S. pombe* pad1, hit: eIF3p40]), whose functions are unknown. This analysis corroborates similar findings by Hershey et al. (1996) and is included here for completeness (Fig. 2). A third function, adding to those of proteasomal and translational initiation, appears to be mediated by members of this domain family. Mov34 homologues human JAB1 and *S. pombe* pad1 have been shown to selectively potentiate transcription via binding to particular gene regulatory proteins AP-1 (Shimanuki et al., 1995; Claret et al., 1996). A further Mov34 homologue, C6.1A, is fused to the T-cell receptor in pro-lymphocytic T-cell leukemia (PLL) (Fisch et al., 1993), suggesting that disruption of one or more of the three functions of Mov34 homologues could be important in the etiology of PLL.

Proteasomal subunit S5a is a homologue of TFIIH subunit p44: The use of 26S proteasomal subunit homologues in regulating transcription and translation is emphasized further by the finding that the 26S proteasomal subunit S5a and its yeast orthologue Sun1 are homologues of the p44 subunit of the RNA polymerase II basal transcription factor IIH (TFIIH) (e.g., PSI-BLAST $E = 4 \times 10^{-4}$ on pass 2 [query: human S5a]) (Fig. 3). Human TFIIH possesses

Fig. 2. Multiple alignment of Mov34 (subunit S12, p40, SwissProt nomenclature: PRSC) homologues. The Drosophila 26S proteasome subunit sequence (PRSC_DROME) has been modified to account for a double frameshift. Abbreviations, coloring, and calculation of consensus and predicted secondary structures are as given in the legend to Figure 1.

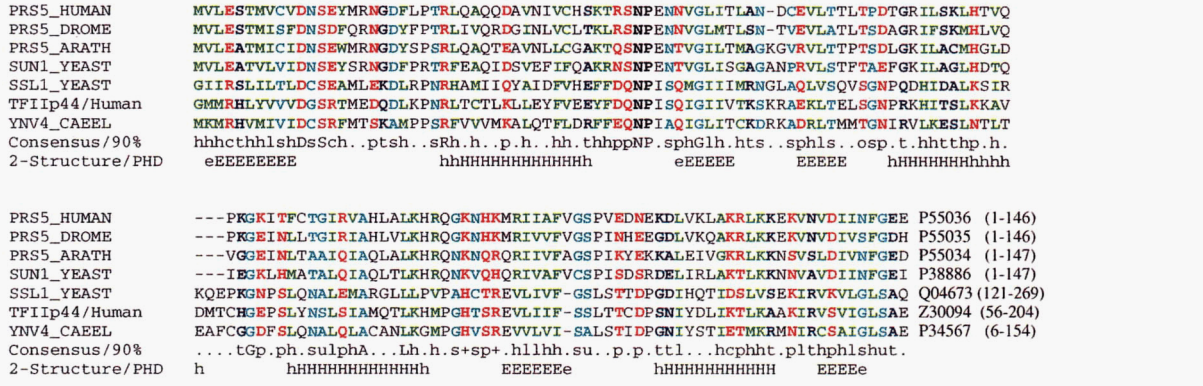

Fig. 3. Multiple alignment of 26S proteasome subunit S5a sequences (SwissProt nomenclature: PRS5) and TFIIH p44 homologues. Abbreviations, coloring, and calculation of consensus and predicted secondary structures are as given in the legend to Figure 1.

multiple roles in transcription and DNA repair mechanisms (reviewed in Svejstrup et al., 1996) and its p44 subunit is thought to associate with several TFIIH components (Iyer et al., 1996). Ssl1, the *S. cerevisiae* p44-orthologue (Humbert et al., 1994), is a component of the yeast TFIIH complex and is essential for translation initiation in yeast possibly by promoting the interaction of ribosomes with mRNA (Yoon et al., 1992).

The proteasome subunit S5a also appears to possess a transcriptional function since it interacts strongly with Id1 (Inhibitor of DNA-binding 1), and less strongly with MyoD and E12; in addition, it restores DNA-binding by Id1-E21 and Id1-MyoD heterodimers and enhances DNA-binding by homodimers of E12 or MyoD (Anand et al., 1997).

Functional similarities among proteasome, eIF3, and transcription associated complexes: There is considerable evidence implicating homologous proteins in the regulation of eukaryotic transcription, protein synthesis, and protein degradation. As described above, proteasome subunit homologues S5a, TRIP15, and JAB1, as well as Sug1/TRIP1 (Swaffield et al., 1995) have all been implicated in transcriptional regulation, and other subunit homologues are implicated in translation regulation. The converse also appears to hold true: a modulator of HIV TAT-dependent transcriptional activation is known to be identical to the proteasome S7 subunit (Dubiel et al., 1995) and protein synthesis elongation factor EF-1α has been shown to be essential for ubiquitin-mediated degradation of certain proteins by the 26S proteasome (Gonen et al., 1994). Each of the three cellular functions in question is mediated by large multimolecular assemblages, and it is possible that homologues in different complexes provide similar core structures upon which the assemblies are built. On the other hand, the known transcription and translation regulatory properties of proteasome subunits (Lee et al., 1995; Shimanuki et al., 1995; Swaffield et al., 1995; Claret et al., 1996; Anand et al., 1997) point to cellular functions of the 26S proteasome that are distinct from ubiquitin-mediated proteolysis. It is concluded that processes regulating transcription, translation, and protein degradation are interdependent and are regulated in part by proteins that share common ancestors.

Note added in proof: It has come to our attention that the PINT motif is identical to the PCI domain discovered independently by K. Hofmann et al. which will be published elsewhere.

Acknowledgments: LA was supported by a Regent's Fellowship. CPP is a Wellcome Trust Career Development Fellow, and is a member of the Oxford Centre for Molecular Sciences.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl Acids Res 25*:3389–3402.

Anand G, Yin X, Shahidi AK, Grove L, Prochownik EV. 1997. Novel regulation of the helix-loop-helix protein Id1 by S5a, a subunit of the 26 S proteasome. *J Biol Chem 272*:19140–19151.

Asano K, Kinzy TG, Merrick WC, Hershey JWB. 1997. Conservation and diversity of eukaryotic translation initiation factor eIF3. *J Biol Chem 272*:1101–1109.

Castle LA, Meinke DW. 1994. A FUSCA gene of Arabidopsis encodes a novel protein essential for plant development. *Plant Cell 6*:25–41.

Claret FX, Hibi M, Dhut S, Toda T, Karin M. 1996. A new group of conserved coactivators that increase the specificity of AP-1 transcription factors. *Nature 383*:453–457.

Coux O, Tanaka K, Goldberg AL. 1996. Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem 65*:801–847.

DeMartino GN, Moomaw CR, Zagnitko OP, Proske RJ, Chu-Ping M, Afendis SJ, Swaffield JC, Slaughter CA. 1994. PA700, an ATP-dependent activator of the 20 S proteasome, is an ATPase containing multiple members of a nucleotide-binding protein family. *J Biol Chem 269*:20878–20884.

Desbois C, Rousset R, Bantignies F, Jalinot P. 1996. Exclusion of Int-6 from PML nuclear bodies by binding to the HTLV-I tax oncoprotein. *Science 273*:951–953.

Deveraux Q, Ustrell V, Pickart C, Rechsteiner M. 1994. A 26S protease subunit that binds ubiquitin conjugates. *J Biol Chem 269*:7059–7061.

Dubiel W, Ferrel K, Rechsteiner M. 1995. Subunits of the regulatory complex of the 26S protease. *Mol Biol Rep 21*:27–34.

Eddy SR, Mitchison G, Durbin RJ. 1995. Maximum discrimination hidden Markov models of sequence comparisons. *J Comput Biol 2*:9–23.

Everett RD, Meredith M, Orr A, Cross A, Kathoria M, Parkinson J. 1997. A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regulatory protein. *EMBO J 16*:1519–1530.

Fisch P, Forster A, Sherrington PD, Dyer MJ, Rabbitts TH. 1993. The chromosomal translocation t(X;14)(q28;q11) in T-cell pro-lymphocytic leukaemia breaks within one gene and activates another. *Oncogene 8*:3271–3276.

Gonen H, Smith CE, Siegel NR, Kahana C, Merrick WC, Chakraburtty K, Schwartz AL, Ciechanover A. 1994. Protein synthesis elongation factor EF-1α is essential for ubiquitin-dependent degradation of certain N$^{\alpha}$-acetylated proteins and may be substituted for by the bacterial elongation factor EF-Tu. *Proc Natl Acad Sci USA 91*:7648–7652.

Goubeaud A, Knirr S, Renkawitz-Pohl R, Paululat A. 1996. The *Drosophila* gene *alien* is expressed in the muscle attachment sites during embryogenesis and encodes a protein highly conserved between plants, *Drosophila* and vertebrates. *Mech Dev 57*:59–68.

Hershey JW, Asano K, Naranda T, Vornlocher HP, Hanachi P, Merrick WC. 1996. Conservation and diversity in the structure of translation initiation factor eIF3 from humans and yeast. *Biochimie 78*:903–907.

Hilt W, Wolf DH. 1996. Proteasomes: Destruction as a programme. *Trends Biochem Sci 21*:96–102.

Hoffman L, Rechsteiner M. 1997. Molecular cloning and expression of subunit 9 of the 26S proteasome. *FEBS Lett 404*:179–184.

Humbert S, van Vuuren H, Lutz Y, Hoeijmakers JH, Egly JM, Moncollin V. 1994. p44 and p34 subunits of the BTF2/TFIIH transcription factor have homologies with SSL1, a yeast protein involved in DNA repair. *EMBO J 13*:2393–2398.

Iyer N, Reagan MS, Wu KJ, Canagarajah B, Friedberg EC. 1996. Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein. *Biochemistry 35*:2157–2167.

Kawamura M, Kominami K, Takeuchi J, Toh-e A. 1996. A multicopy suppressor of *nin1-1* of the yeast *Saccharomyces cerevisiae* is a counterpart of the *Drosophila melanogaster* diphenol oxidase A2 gene, *Dox-A2*. *Mol Gen Genet 251*:146–152.

Lee JW, Ryan F, Swaffield JC, Johnston SA, Moore DD. 1995. Interaction of thyroid-hormone receptor with a conserved transcriptional mediator. *Nature 374*:91–94.

Löwe J, Stock D, Jap B, Zwickel P, Baumeister W, Huber R. 1995. Crystal structure of the 20S proteasome from the Archaeon T. acidophilum at 3.4Å resolution. *Science 268*:533–539.

Lurquin C, Van Pel A, Mariame B, De Plaen E, Szikora JP, Janssens C, Reddehase MJ, Lejeune J, Boon T. 1989. Structure of the gene of tum- transplantation antigen P91A: The mutated exon encodes a peptide recognized with Ld by cytolytic T cells. *Cell 58*:293–303.

Mushegian AR, Koonin EV. 1996. Sequence analysis of eukaryotic developmental proteins: Ancient and novel domains. *Genetics 144*:817–818.

Naranda T, MacMillan SE, Donahue TF, Hershey JWB. 1996. SUI1/p16 is required for the activity of eukaryotic translation initiation factor 3 in *Saccharomyces cerevisiae*. *Mol Cell Biol 16*:2307–2313.

Pearson WR. 1991. Searching protein sequence libraries: Comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics 11*:635–650.

Pentz ES, Wright TR. 1991. *Drosophila melanogaster* diphenol oxidase A2: Gene structure and homology with the mouse mast-cell tum- transplantation antigen, P91A. *Gene 103*:239–242.

Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol 232*:584–599.

Rubin DM, Coux O, Wefes I, Hengartner C, Young RA, Goldberg AL, Finley D. 1996. Identification of the gal4 suppressor Sug1 as a subunit of the yeast 26S proteasome. *Nature 379*:655–657.

Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins Struct Funct Genet 9*:180–190.

Shimanuki M, Saka Y, Yanagida M, Toda T. 1995. A novel essential fission yeast gene pad1+ positively regulates pap1(+)-dependent transcription and is implicated in the maintenance of chromosome structure. *J Cell Sci 108*:569–579.

Smith MW, Ito M, Miyawaki M, Sato S, Yoshikawa Y, Wada S, Maki H, Nakagawa H, Komamine A. 1997. Plant 21D7 protein, a nuclear antigen associated with cell division, is a component of the 26S proteasome. *Plant Physiol 113*:281–291.

Spain BH, Bowdish KS, Pacal AR, Staub SF, Koo D, Chang CY, Xie W, Colicelli J. 1996. Two human cDNAs, including a homolog of Arabidopsis FUS6 (COP11), suppress G-protein- and mitogen-activated protein kinase-mediated signal transduction in yeast and mammalian cells. *Mol Cell Biol 16*:6698–6706.

Staub JM, Wei N, Deng XW. 1996. Evidence for FUS6 as a component of the nuclear-localized COP9 complex in Arabidopsis. *Plant Cell 8*:2047–2056.

Svejstrup JQ, Vichi P, Egly EM. 1996. The multiple roles of transcription/repair factor TFIIH. *Trends Biochem Sci 21*:346–350.

Swaffield JC, Melcher K, Johnston SA. 1995. A highly conserved ATPase protein as a mediator between acidic activation domains and the TATA-binding protein. *Nature 374*:88–91.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL-W: Improving sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucl Acids Res 22*:4673–4680.

Tsurumi C, DeMartino GN, Slaughter CA, Shimbara N, Tanaka K. 1995. cDNA cloning of p40, a regulatory subunit of the human 26S proteasome, and a homologue of the Mov-34 gene product. *Biochem Biophys Res Commun 210*:600–608.

van Nocker S, Deveraux Q, Rechsteiner M, Vierstra RD. 1996. *Arabidopsis MBP1* gene encodes a conserved ubiquitin recognition component of the 26S proteasome. *Proc Natl Acad Sci USA 93*:856–860.

Yoon H, Miller SP, Pabich EK, Donahue TF. 1992. SSL1, a suppressor of a HIS4 5′-UTR stem-loop mutation, is essential for translation initiation and affects UV resistance in yeast. *Genes Dev 6*:2463–2477.