

Letter to the Editor

Evolution of Ribonuclease Inhibitor by Exon Duplication

Marcia C. Haigis,* Eric S. Haag,* and Ronald T. Raines*†

*Department of Biochemistry and †Department of Chemistry, University of Wisconsin–Madison

The leucine-rich repeat (LRR) is a prevalent structural motif that mediates specific protein-protein interactions (Kobe and Deisenhofer 1995*b*). The LRR is defined by a region of 20–29 residues that contains the consensus sequence: XLXXLXXN, where L, N, and X are leucine, asparagine, and any residue, respectively. Typically, LRRs are displayed in 1–30 tandem copies within a protein.

The first crystalline structure of an LRR-containing protein revealed a new protein fold (Kobe and Deisenhofer 1993) (fig. 1*A* and *B*). In porcine ribonuclease inhibitor (RI [Hofsteenge 1997]), the LRRs are arranged tandemly in a horseshoe shape with 16 α -helices enclosing a parallel β -sheet of 17 strands. Hydrophobic forces between the consensus leucine residues appear to stabilize the β -sheet. RI is able to bind tightly to members of the ribonuclease (RNase) A superfamily, which not only catalyze RNA degradation but also mediate angiogenesis, cytotoxicity, and the host-defense response (D'Alessio and Riordan 1997; Raines 1998; Leland and Raines 2001).

The RI fold is conserved in other proteins containing LRRs. Human U2A' is a small nuclear protein that contains five LRRs and binds to U2B'' to form an active spliceosomal complex. Another protein, Skp2, contains both 10 LRRs that function as a protein-binding domain and an F-box motif that binds to the catalytic core of ubiquitin-protein ligases. The LRRs in U2A' and Skp2 adopt an RI-like fold (Price, Evans, and Nagai 1998; Schulman et al. 2000). The crystalline structures of U2A' bound to U2B'' and those of RI bound to RNase A and angiogenin reveal that the parallel β -sheet formed by the LRRs furnishes the surface for interacting with other proteins (Kobe and Deisenhofer 1995*a*; Papageorgiou, Shapiro, and Acharya 1997; Price, Evans, and Nagai 1998).

The modularity of the LRR is consistent with the evolution of RI by gene duplication. To test this hypothesis, we investigated the genomic structure of human RI. Our findings provide a new insight into the evolution of RI as well as the many proteins containing LRRs or other modules.

Human RI consists of alternating 28- or 29-residue LRR units (Kobe and Deisenhofer 1993; Papageorgiou,

Shapiro, and Acharya 1997; Kajava 1998) (fig. 1*A* and *B*). By using the National Center for Biotechnology Information (NCBI) nr database (<http://www.ncbi.nlm.nih.gov>), we identified three potential transcripts obtained from human placental, HeLa, and testis cDNA libraries (GI = 186260, 35843, and 7328055, respectively). These clones are 1,682, 1,921, and 2,982 bp, respectively. Although alternative splice forms of the 5'-untranslated region have been identified in placenta, Northern blot analysis revealed that the RI gene is expressed as a single transcript in placental and HeLa cells (Lee et al. 1988; Schneider et al. 1988; Crawford, Hagerly, and Beutler 1989). Using these transcripts, we searched the NCBI database for a corresponding genomic sequence and identified a region of DNA on human chromosome 11 that overlaps cosmids cSRL125c1 and cSRL135f4 and contains the RI gene (GI = 2281060).

The genomic locus of human RI spans more than 12 kb (fig. 1*C*). All three transcripts that we identified possess a colinear open reading frame encoded by nine 3'-terminal exons, numbered 2–10, as determined by the presence of consensus splice donor and acceptor sites. The transcripts differ in their 5'-untranslated region and contain zero, one, or two alternatively spliced exons. As all transcripts encode an identical protein sequence, the functional significance of alternative splice forms is unknown. Remarkably, all but the first and last coding exons are exactly 171 bp in length. In addition, all introns are of phase 2, which denotes codon interruption after the second nucleotide. Hence, any of the internal exons could be moved or removed without disrupting the alignment of LRRs encoded by neighboring exons.

A striking relationship exists between the intron-exon structure and the tertiary structure of RI (fig. 1). Each of the internal exons encodes 57 residues. Except for the junction after the N-terminus, which occurs within a longer β -strand, every exon translation product correlates with a pair of β -strands and α -helices without any disruption (fig. 1*A* and *B*). Thus, the LRR motif within RI is defined completely by its intron-exon structure.

The phasing and position of the RI introns are consistent with exon amplification during evolution (de Souza et al. 1997; Gilbert, de Souza, and Long 1997; Fedorov et al. 1998). To test this hypothesis, we investigated the similarity of the coding exons. RI exons were aligned using the PILEUP program from the Wisconsin Package of Genetics Computer Group (Madison, Wis.). Although the terminal exons (2 and 10) have low sequence identity, the internal exons (3–9) can be aligned without gaps and have considerable sequence identity (fig. 2*A*). Pairwise comparisons revealed that exons 3–9 are 50%–60% identical. In addition, translation products corresponding to exons 3–9 were aligned and determined to be 30%–50% identical (fig. 2*C*).

Abbreviations: LRR, leucine-rich repeat; RI, ribonuclease inhibitor.

Key words: angiogenin, exon duplication, leucine-rich repeat, ribonuclease A, ribonuclease inhibitor, RNA.

Address for correspondence and reprints: Ronald T. Raines, Department of Biochemistry, University of Wisconsin–Madison, 433 Babcock Drive, Madison, Wisconsin 53706-1544.
E-mail: raines@biochem.wisc.edu.

Mol. Biol. Evol. 19(6):959–963. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

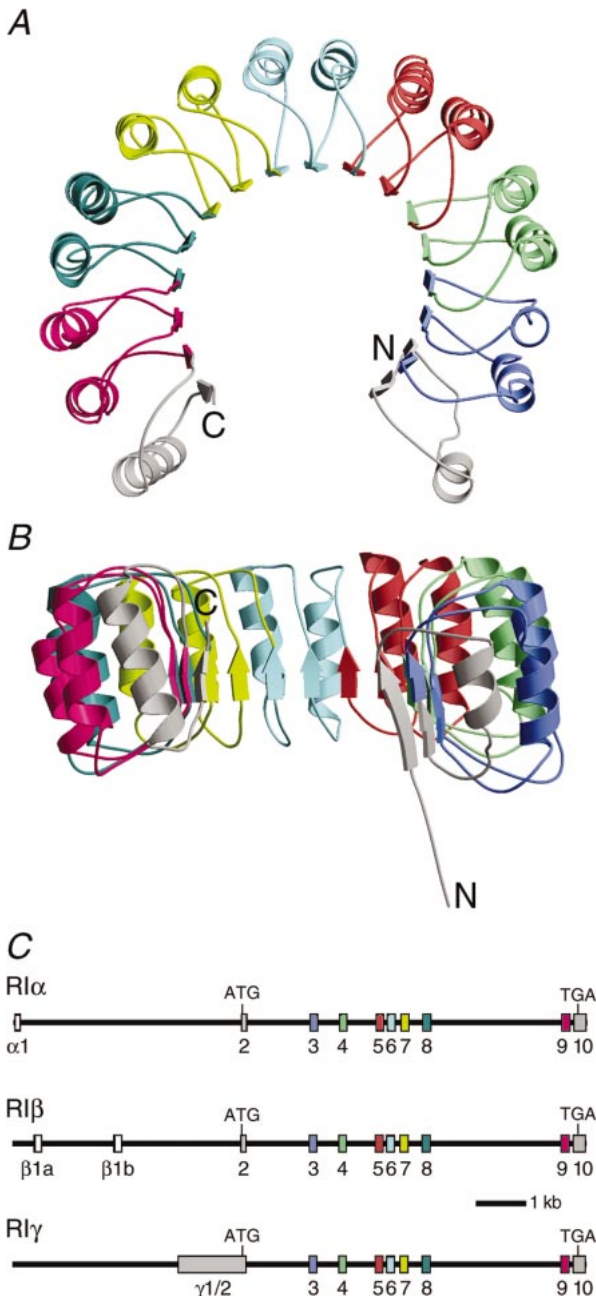


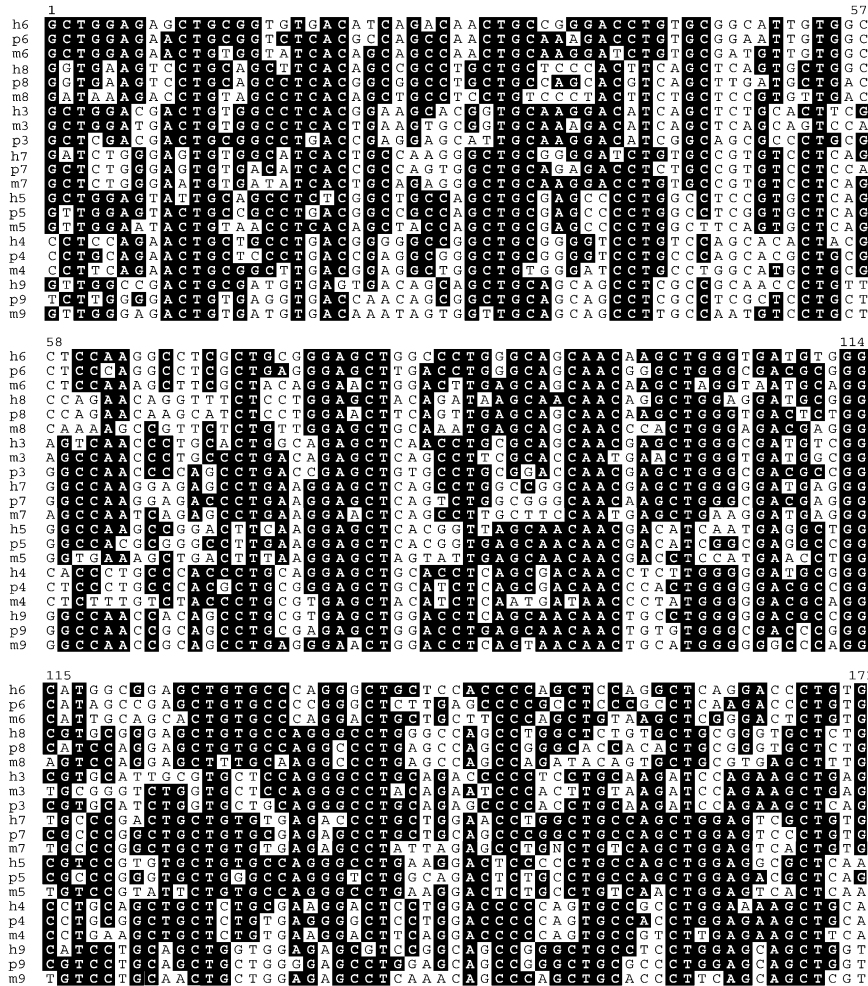
FIG. 1.—Correlation of the exons of human RI with its LRR units. *A*, Ribbon diagram of RI was created using the atomic coordinates derived by X-ray diffraction analysis (Papageorgiou, Shapiro, and Acharya 1997) and the program MOLSCRIPT (Kraulis 1991). Each exon-encoded sequence is in a different color. The N- and C-termini are labeled. *B*, Same as *A*, except rotated by $\sim 90^\circ$ around the horizontal axis. *C*, Physical map of the human RI gene. Alternate forms isolated from placenta, HeLa cells, and testes are depicted as RI α , RI β , and RI γ , respectively. Exons and introns are drawn to scale. Start and stop codons are indicated. Exon colors match those of exon-encoded sequences in *A* and *B*.

The high degree of similarity and their identical length strongly suggests that the internal exons of RI are homologues related by exon duplication. We wished to know whether these duplications had occurred all at once or in distinct phases and whether interexonic conversion events had occurred since different RI-contain-

ing genomes diverged. To distinguish between these distinct mechanisms, we performed phylogenetic analyses on the internal coding exons of human, mouse, and pig RI. The mouse RI exons were identified by a twofold approach: analyzing existing sequences and generating new sequence data. By using the NCBI mouse genome database, we identified genomic sequences that contain several exons of mouse RI. In addition, we generated new sequence data to verify the remaining intron-exon junctions. To obtain the mouse RI locus, we screened a mouse 129 library (Genome Systems, St. Louis, Mo.) by the PCR using the primers: 5' GGTGCAAAGACA-TCAGCTCAGCAGTCCAAGC 3' and 5' CCTGGA-GCACCAGACCCACACC 3'. Like the human gene, mouse RI consists of seven internal exons that are 171 bp in length and are separated by phase 2 introns. The pig exons were inferred from the cDNA for pig RI (GI = M58700) by using the genomic structures of human and mouse RI. The internal exons from human, pig, and mouse were aligned as described previously (fig. 2A). Sequence alignment reveals strong identity between analogous exons. Neighbor-Joining phylogenetic trees built by using a Kimura two-parameter correction have strongly supported clades for analogous exons between species with bootstrap values of 99%–100% (fig. 2B). A maximum parsimony analysis of the data set also strongly supports the existence of clades for each exon with bootstrap values of 98%–100%. Apparently, the tandem repeats within RI are extremely stable and homogenization has not taken place, as that would have resulted in a tree in which the internal exons from each species form a strongly supported clade. A consistent internal branching pattern did not emerge with strong support from either analysis. Furthermore, the branches leading to each exon clade are extremely long, which is consistent with a nearly simultaneous divergence of each clade. Thus, we conclude from this analysis that exons 3–9 were most likely duplicated in a radiation that occurred rapidly early in the history of RI. Members of the RNase A superfamily apparently exist in all tetrapods (D'Alessio and Riordan 1997). Thus, it is likely that RI was assembled after the origin of tetrapods in the late Devonian, approximately 360 MYA, but before the divergence of eutherian mammals.

Our data indicate several unusual features about the evolution of RI. The internal exons of RI are homologous and stable. In addition, the branching pattern suggests that the internal exon duplications occurred rapidly and then diverged over a long period of time. To assess whether positive selection has been a factor in the divergence of the RI exons, we measured the rate of non-synonymous versus synonymous changes in human exons 3–9 (Nei and Gojobori 1986). The ratios of non-synonymous-to-synonymous changes range from 0.31 to 1.00. Although the nonsynonymous-to-synonymous ratio is not always a definitive indicator of positive selection (Wyckoff, Wang, and Wu 2000), the low ratios for RI do not provide support for positive selection. Moreover, these ratios can be overestimated if the GC content in the third codon position is high (Moriyama and Gojobori 1992; Bielawski, Dunn, and Yang 2000). In hu-

A



B

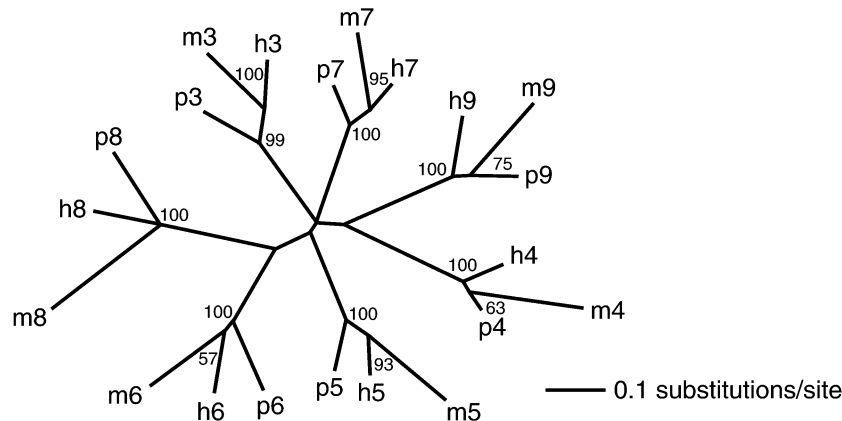


FIG. 2.—Analysis of the internal exon sequences of RI A. An alignment of DNA sequences that corresponds to each internal exon from human (h), pig (p), and mouse (m) was generated with the program PILEUP. Sequences are listed in order of decreasing identity. Nucleotides present in at least 10 of 21 exons are in black boxes. B, A neighbor-joining tree for all codon positions in A was generated with the PAUP program (Swofford 1998) using a Kimura two-parameter correction. Bootstrap values are indicated at the nodes. The bar indicates the distance for 0.1 nucleotide substitutions per site for the 171 positions of RI.

man RI, the GC content is indeed high, being 79%–93%. We conclude that the divergence of the RI exons from one another is unlikely to have been promoted by positive selection.

Why did the RI exons duplicate? There are no known RI homologues that vary in the number of LRR units (Hofsteenge et al. 1988; Lee et al. 1988; Schneider et al. 1988; Kawanomoto et al. 1992). In contrast, other

LRR-containing proteins have a variable number of repeats. For example, the tomato Cf-2 and Cf-5 genes encode homologues that do differ in LRR unit number (Dixon et al. 1998). Modular mutagenesis of RI has demonstrated that not all of its LRRs are required for binding to RNase A (Lee and Vallee 1990). In other words, RI progenitors with fewer than the extant number of LRRs could be functional. The contemporary RI could, however, have an increased affinity for a broad spectrum of RNases. The crystalline structures of human RI complexed with angiogenin and porcine RI complexed with RNase A support this hypothesis (Kobe and Deisenhofer 1995a; Papageorgiou, Shapiro, and Acharya 1997). These structures reveal that only 17 of the 28 RI residues that contact RNase A also contact angiogenin (Papageorgiou, Shapiro, and Acharya 1997). Yet, the affinity of RI for both angiogenin and RNase A is extreme. The K_d values of the RI-angiogenin and RI-RNase A complexes are near 10^{-15} M, among the lowest for known biomolecular interactions (Lee, Shapiro, and Vallee 1989; Vicentini et al. 1990). High LRR unit number could be critical for the evolution of both broad specificity and high affinity by RI.

RI and RNase have coevolved. The amino acid sequence of RI has been determined only for the protein from human, pig, rat, and mouse (Hofsteenge et al. 1988; Lee et al. 1988; Schneider et al. 1988; Kawanomoto et al. 1992) (GI = AF071546). In contrast, the amino acid sequence of RNase A homologues has been determined for >40 species and reveals expansion through gene duplication (Beintema et al. 1997; Beintema 1998). The binding of RI to members of the RNase A superfamily is class specific. For example, porcine RI binds tightly to human RNase A, but no mammalian RI inhibits the activity of RNase A homologues isolated from frogs and vice versa (Roth 1962; Nagano et al. 1976; Tomita et al. 1979; Wu et al. 1993). Likewise, mammalian RI does not inhibit the activity of the RNase A homologue from chicken liver (Hayano et al. 1993; Wu et al. 1993). Uninhibited ribonucleolytic activity can result in cell death by the degradation of cellular RNA (Leland et al. 1998; Leland and Raines 2001; Leland et al. 2001). Coevolution of RI and RNases maintains a safeguard against this circumstance.

The precise exon phasing seen in RI exists in some, but not all, LRR-containing proteins. For example, *Mus musculus* MATER (*Maternal Antigen That Embryos Require*) has similar phasing. Several of its repeats are encoded by 171 bp exons (Tong, Nelson, and Dean 2000). In contrast, the protein FBL2 contains six LRRs, which do not coincide with its exon structure (Ilyin et al. 1999). These examples suggest independent evolution for the LRR motif and that not all LRR-containing proteins have evolved by exon duplication.

RI has no obvious homologues in invertebrates. Yet, RI does share significant amino acid similarity with other proteins having LRRs. Two such proteins are U2A' (vide supra) and rna1p (Hillig et al. 1999). These proteins are not homologues of RI but are extremely similar in LRR arrangement and tertiary structure.

The β - α barrel is a common structural scaffold in soluble proteins. Like RI, β - α barrel proteins are formed by an alternating pattern of parallel β -strands and α -helices, with the α -helices situated peripheral to inner β -strands. Members of this group are often enzymes that bind and catalyze the interconversion of small-molecule metabolites (Reardon and Farber 1995). The versatility of the β - α barrel in enabling protein-small-molecule interactions resembles that of LRRs in enabling protein-protein interactions (Kobe and Deisenhofer 1994). An investigation of the ancestry of β - α barrels suggests that many β - α proteins have evolved from a stable intermediate—a primordial half-barrel (Lang et al. 2000; Miles and Davies 2000). In contrast, our data are not consistent with the existence of such an intermediate during the evolution of RI (fig. 2B).

In summary, the intron-exon junctions of RI reveal a remarkable mechanism of molecular evolution by exon duplication. These junctions are precisely in phase with the tandem LRR modules. RI evolved through rapid duplication of its LRR modules and their subsequent divergence. These modules are highly stable and have not homogenized since the divergence of human, pig, and mouse. RI thus acquired a malleable, modular scaffold with which to mediate protein-protein interactions. We conclude that the accumulation of functional units can be a driving force for exon duplication. Our mechanism for the evolution of RI could also apply to the evolution of the many other proteins containing LRRs, as well as guide the study of modular proteins in general.

Acknowledgments

We thank K. M. Haigis and Drs. N. T. Perna and C. Park for discussion and critical reading of the manuscript, and Dr. A. M. Dean for a helpful suggestion. E.S.H. was supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research and by the Howard Hughes Medical Institute. This work was supported by grants GM44783 and CA73808 (NIH).

LITERATURE CITED

- BEINTEMA, J. J. 1998. The ribonuclease A superfamily. *Cell. Mol. Life Sci.* **54**:763–765.
- BEINTEMA, J. J., H. J. BREUKELMAN, A. CARSANA, and A. FURIA. 1997. Evolution of vertebrate ribonucleases: ribonuclease A superfamily. Pp. 245–269 in G. D'ALESSIO and J. F. RIORDAN, eds. *Ribonucleases: structures and functions*. Academic Press, New York.
- BIELAWSKI, J. P., K. A. DUNN, and Z. YANG. 2000. Rates of nucleotide substitution and mammalian nuclear gene evolution: approximate and maximum-likelihood methods lead to different conclusions. *Genetics* **156**:1299–1308.
- CRAWFORD, D., K. HAGERTY, and B. BEUTLER. 1989. Multiple splice forms of ribonuclease-inhibitor mRNA differ in the 5'-untranslated region. *Gene* **85**:525–531.
- D'ALESSIO, G., and J. F. RIORDAN, eds. 1997. *Ribonucleases: structures and functions*. Academic Press, New York.
- DE SOUZA, S. J., M. LONG, L. SCHOENBACH, S. W. ROY, and W. GILBERT. 1997. The correlation between introns and the three-dimensional structure of proteins. *Gene* **205**:141–144.

- DIXON, M. S., K. HATZIXANTHIS, D. A. JONES, K. HARRISON, and J. D. JONES. 1998. The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell* **10**:1915–1925.
- FEDOROV, A., L. FEDOROVA, V. STARSHENKO, V. FILATOV, and E. GRIGOR'EV. 1998. Influence of exon duplication on intron and exon phase distribution. *J. Mol. Evol.* **46**:263–271.
- GILBERT, W., S. J. DE SOUZA, and M. LONG. 1997. Origin of genes. *Proc. Natl. Acad. Sci. USA* **94**:7698–7703.
- HAYANO, K., M. IWANA, H. SAKAMOTO, H. WATANABE, A. SANDA, K. OHGI, and M. IRIE. 1993. Characterization of poly C preferential ribonuclease from chicken liver. *J. Biochem. (Tokyo)* **114**:156–162.
- HILLIG, R. C., L. RENAULT, I. R. VETTER, T. DRELL, A. WITTINGHOFFER, and J. BECKER. 1999. The crystal structure of rnalp: a new fold for a GTPase-activating protein. *Mol. Cell* **3**:781–791.
- HOFSTEENGE, J. 1997. Ribonuclease inhibitor. Pp. 621–658 in G. D'ALESSIO, and J. F. RIORDAN, eds. *Ribonucleases: structures and functions*. Academic Press, New York.
- HOFSTEENGE, J., B. KIEFFER, R. MATTHIES, B. A. HEMMINGS, and S. R. STONE. 1988. Amino acid sequence of the ribonuclease inhibitor from porcine liver reveals the presence of leucine-rich repeats. *Biochemistry* **27**:8537–8544.
- ILYIN, G. P., M. RIALLAND, D. GLAISE, and C. GUGUEN-GUILLOUZO. 1999. Identification of a novel Skp2-like mammalian protein containing F-box and leucine-rich repeats. *FEBS Lett.* **459**:75–79.
- KAJAVA, A. V. 1998. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **277**:519–527.
- KAWANOMOTO, M., K. MOTOJIMA, M. SASAKI, H. HATTORI, and S. GOTO. 1992. cDNA cloning and sequence of rat ribonuclease inhibitor, and tissue distribution of mRNA. *Biochim. Biophys. Acta* **1129**:335–338.
- KOBE, B., and J. DEISENHOFER. 1993. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**:751–756.
- . 1994. The leucine-rich repeat: a versatile binding motif. *Trends Biochem. Sci.* **19**:415–421.
- . 1995a. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**:183–186.
- . 1995b. Proteins with leucine-rich repeats. *Curr. Opin. Struct. Biol.* **5**:409–416.
- KRAULIS, P. J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**:946–950.
- LANG, D., R. THOMA, M. HENN-SAX, R. STERNER, and M. WILMANN. 2000. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion [see comments]. *Science* **289**:1546–1550.
- LEE, F. S., E. A. FOX, H.-M. ZHOU, D. J. STRYDOM, and B. L. VALLEE. 1988. Primary structure of human placental ribonuclease inhibitor. *Biochemistry* **27**:8545–8553.
- LEE, F. S., R. SHAPIRO, and B. L. VALLEE. 1989. Tight-binding inhibition of angiogenin and ribonuclease A by placental ribonuclease inhibitor. *Biochemistry* **28**:225–230.
- LEE, F. S., and B. L. VALLEE. 1990. Modular mutagenesis of human placental ribonuclease inhibitor, a protein with leucine-rich repeats. *Proc. Natl. Acad. Sci. USA* **87**:1879–1883.
- LELAND, P. A., and R. T. RAINES. 2001. Cancer chemotherapy—ribonucleases to the rescue. *Chem. Biol.* **8**:405–425.
- LELAND, P. A., L. W. SCHULTZ, B.-M. KIM, and R. T. RAINES. 1998. Ribonuclease A variants with potent cytotoxic activity. *Proc. Natl. Acad. Sci. USA* **98**:10407–10412.
- LELAND, P. A., K. E. STANISZEWSKI, B.-M. KIM, and R. T. RAINES. 2001. Endowing human pancreatic ribonuclease with toxicity for cancer cells. *J. Biol. Chem.* **276**:43095–43102.
- MILES, E. W., and D. R. DAVIES. 2000. Protein evolution. On the ancestry of barrels [comment]. *Science* **289**:1490.
- MORIYAMA, E. N., and T. GOJOBORI. 1992. Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. *Genetics* **130**:855–864.
- NAGANO, H., H. KIUCHI, Y. ABE, and R. SHUKUYA. 1976. Purification and properties of an alkaline ribonuclease from the hepatic cytosol fraction of bullfrog, *Rana catesbeiana*. *J. Biochem. (Tokyo)* **80**:19–26.
- NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
- PAPAGEORGIOU, A., R. SHAPIRO, and K. ACHARYA. 1997. Molecular recognition of human angiogenin by placental ribonuclease inhibitor—an X-ray crystallographic study at 2.0 Å resolution. *EMBO J.* **16**:5162–5177.
- PRICE, S. R., P. R. EVANS, and K. NAGAI. 1998. Crystal structure of the spliceosomal U2''–U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**:645–650.
- RAINES, R. T. 1998. Ribonuclease A. *Chem. Rev.* **98**:1045–1065.
- REARDON, D., and G. K. FARBER. 1995. The structure and evolution of alpha/beta barrel proteins. *FASEB J.* **9**:497–503.
- ROTH, J. S. 1962. Ribonuclease IX. Further studies on ribonuclease inhibitor. *Biochim. Biophys. Acta* **61**:903–915.
- SCHNEIDER, R., E. SCHNEIDER-SCHERZER, M. THURNHER, B. AUER, and M. SCHWEIGER. 1988. The primary structure of human ribonuclease/angiogenin inhibitor (RAI) discloses a novel highly diversified protein superfamily with a common repetitive module. *EMBO J.* **7**:4151–4156.
- SCHULMAN, B. A., A. C. CARRANO, P. D. JEFFREY, Z. BOWEN, E. R. KINNUCAN, M. S. FINNIN, S. J. ELLEDGE, J. W. HARPER, M. PAGANO, and N. P. PAVLETICH. 2000. Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex. *Nature* **408**:381–386.
- SWOFFORD, D. L. 1998. PAUP* phylogenetic analysis using parsimony (*and other methods). 4th edition. Sinauer Associates, Sunderland, Mass.
- TOMITA, Y., Y. GOTO, T. OKAZAKI, and R. SHUKUYA. 1979. Liver ribonucleases from bullfrog, *Rana catesbeiana*. Purification, properties and changes in activity during metamorphosis. *Biochim. Biophys. Acta* **562**:504–514.
- TONG, Z. B., L. M. NELSON, and J. DEAN. 2000. Mater encodes a maternal protein in mice with a leucine-rich repeat domain homologous to porcine ribonuclease inhibitor. *Mamm. Genome* **11**:281–287.
- VICENTINI, A. M., B. KIEFFER, R. MATHIES, B. MEYHACK, B. A. HEMMINGS, S. R. STONE, and J. HOFSTEENGE. 1990. Protein chemical and kinetic characterization of recombinant porcine ribonuclease inhibitor expressed in *Saccharomyces cerevisiae*. *Biochemistry* **29**:8827–8834.
- WU, Y., S. M. MIKULSKI, W. ARDELT, S. M. RYBAK, and R. J. YOULE. 1993. A cytotoxic ribonuclease. *J. Biol. Chem.* **268**:10686–10693.
- WYCKOFF, G. J., W. WANT, and C.-I. WU. 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**:304–309.

ANTONY DEAN, reviewing editor

Accepted November 12, 2001