Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6907–6911, September 1986
Evolution

Homologous ribosomal proteins in bacteria, yeast, and humans

(ribosomes/evolution)

I-TSUEN CHEN, APARNA DIXIT, DOUGLAS D. RHOADS, AND DONALD J. ROUFA*

Center for Basic Cancer Research, Division of Biology, Kansas State University, Manhattan, KS 66506

Communicated by Charles D. Michener, June 2, 1986

ABSTRACT **We describe sequences of two human ribosomal proteins, S14 and S17, and messenger RNAs that encode them. cDNAs were used as molecular hybridization probes to recognize complementary genes in rodent, *Drosophila*, and yeast chromosomal DNAs. Human ribosomal protein sequences are compared to analogous Chinese hamster, yeast, and bacterial genes. Our observations suggest that some ribosomal protein genes have been conserved stringently in the several phylogenetic lines examined. These genes apparently were established early in evolution and encode products that are fundamental to the translational apparatus. Other ribosomal protein genes examined, although similar enough to heterologous DNA sequences to indicate their structural relationships, appear to have diverged substantially during evolution, probably reflecting adaptations to different genetic environments.**

Phylogenetic differences among contemporary "housekeeping" genes are likely to reflect pathways followed in successful evolutionary lineages. Ribosomal genes are particularly useful evolutionary markers, because natural selection is difficult to envision prior to establishment of an accurate mechanism for translating nucleic acid sequences into proteins, and because ribosomes from diverse organisms display remarkably similar subunit structure and function. These similarities are likely to reflect fundamental features of the translation mechanism fixed during the earliest moments of evolution.

For lack of sequence data pertaining to ribosomal proteins (rproteins) from most organisms other than the bacterium *Escherichia coli*, structural comparisons of modern ribosomal components usually have focused on ribosomal RNAs (rRNAs) and their genes. Recently recombinant DNA and nucleic acid sequencing techniques have been used to study rprotein genes from a variety of organisms (1–16). We have characterized several Chinese hamster and human rprotein transcripts (cDNAs) and polypeptides (17–20). Now we describe comparisons among two human rproteins, the DNAs that encode them, and homologous rprotein genes in other organisms (Chinese hamster, *Drosophila*, yeast, and *E. coli*).

MATERIALS AND METHODS

Materials. Chinese hamster (*Cricetulus griseus*) fibroblasts (CHO) and human fibrosarcoma (HT-1080) genomic DNAs as well as rprotein cDNA clones used in this study have been described (17–21). *Drosophila melanogaster* (strain Canton S) DNA was a gift from Robb Denell (Kansas State Univ.). Genomic DNAs were purified from *Saccharomyces cerevisiae* (X2180-1B) and *E. coli* (K-12) spheroplasts as described (17). GeneScreen*Plus* filter membranes and [α-^{32}P]dCTP (800–900 Ci/mmol; 1 Ci = 37 GBq) were purchased from DuPont-NEN Products. $(dG)_{10-12}$ was obtained from Pharmacia P-L Biochemicals. Restriction endonucleases and *E. coli* DNA polymerase (Klenow fragment) were from Bethesda Research Laboratories.

Methods. Hybridization probes were electrophoretically purified cDNAs radiolabeled to specific activities of 1–3 × 10^8 cpm/μg by using [^{32}P]dCTP, *E. coli* DNA polymerase (Klenow fragment), and $(dG)_{10-12}$ primer (20). Southern blots of genomic DNAs (22) were prepared essentially as described before (17, 19) with the addition of denatured pBR322 DNA, poly(rA), poly(rG), poly(rC), and poly(rU) competitors (3 μg/ml each) to hybridization solutions. Moderately stringent hybridization conditions were employed: 0.15 M NaCl, 0.015 M sodium citrate, 0.003 M sodium phosphate, 0.05% sodium dodecyl sulfate, pH 7.0 at 65°C.

Two protein and nucleic acid sequence databases were surveyed for entries similar to human rproteins S14 and S17: GenBank† and National Biomedical Research Foundation–Protein Identification Resource‡. Searches for protein similarities employed the FASTP algorithm (23), based upon the PAM250 matrix of amino acid similarities (24). Protein homologies were evaluated by Monte Carlo statistics using the program RDF (23). Results are expressed as differences between mean Monte Carlo similarity indices and test indices (in units of standard deviation). Optimized alignments of protein and nucleic acid sequences were generated by using the computer programs PRTALN and NUCALN (25).

RESULTS

Genomic Sequences Complementary to Mammalian rprotein cDNAs. Because Chinese hamster and human rproteins appear nearly identical in two-dimensional electrophoresis (refs. 26 and 27, and unpublished data), it seemed likely that nucleic acids encoding them also might be similar. Thus we surveyed a panel of genomic DNAs for sequences that cross-hybridized with our mammalian (human and Chinese hamster) rprotein cDNA probes (Fig. 1).

Three cDNA probes were chosen for this analysis. Human rprotein S14 cDNA was selected because somatic mutations affecting this gene have permitted us to recognize and map the functional human S14 locus, *RPS14*, to chromosome 5q (17, 19, 20) and to characterize mutant alleles of Chinese hamster S14 *emt b* genes (18). Human S17 cDNA was studied because it encodes a nearly full-length mRNA sequence and because human and Chinese hamster S17 cDNAs cross-hybridize strongly (19). Chinese hamster rprotein L32 cDNA clone was selected as representative of 60S ribosomal subunit protein probes and because the nucleic acid and polypeptide sequences of murine L32 are known (13).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: rprotein, ribosomal protein; bp, base pair(s).

*To whom reprint requests should be addressed.

†National Institutes of Health (1986) Genetic Sequence Databank: GenBank (Research Systems Div., Bolt, Beranek, and Newman, Inc., Boston), Tape Release No. 42.

‡National Biomedical Research Foundation (1986) Protein Identification Resource Protein Sequence Data Base (Natl. Biomed. Res. Found., Washington, DC 20007), Release No. 9.0.

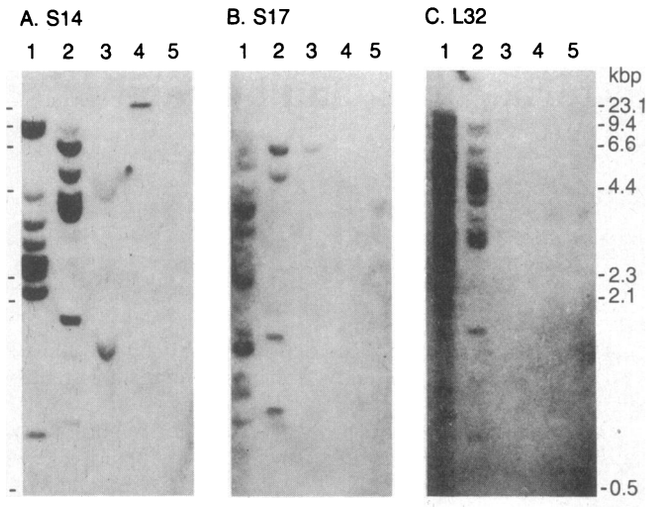

FIG. 1. Sequences complementary to human rprotein cDNAs in several eukaryotic and prokaryotic genomes. Agarose gels contained *Pst* I digests of DNA from human fibrosarcoma cells HT-1080 (lanes 1, 5 μg; CHO Chinese hamster cells (lanes 2, 5 μg); *D. melanogaster* (lanes 3, 1 μg); yeast (lanes 4, 0.05 μg); and *E. coli* (lanes 5, 0.05 μg). Gels were analyzed by hybridization using human S14 (*A*), human S17 (*B*), and Chinese hamster L32 (*C*) cDNA probes. Positions of markers are indicated on the left and right; kbp, kilobase pairs.

The filter blots illustrated in Fig. 1 contained different amounts of genomic DNA from the organisms surveyed to control for differences in genetic complexities. All three mammalian cDNAs recognized similar patterns of fragments in human and hamster DNAs (lanes 1 and 2). S14 and S17 probes both detected homologous sequences in *Drosophila* DNA (Fig. 1 *A* and *B*, respectively), in contrast to the L32 probe, which did not (Fig. 1*C*). Patterns of mammalian rprotein DNA bands (lanes 1 and 2) appear to reflect complex families of rprotein sequences composed of at least one active intron-containing gene and multiple processed pseudogenes (13, 14, 16, 17, 19, 20, 28). Consistent with the lesser complement of repetitive DNA sequences and pseudogenes in *Drosophila* DNA, S14 cDNA detected only two *Pst* I *Drosophila* genomic DNA fragments (4.2 and 1.7 kbp) and S17 cDNA only one (6.6 kbp). S14 cDNA hybridized with a large yeast *Pst* I DNA fragment (Fig. 1*A*, lane 4) whereas S17 and L32 probes detected nothing in yeast DNA (Fig. 1 *B* and *C*, lane 4). None of the probes hybridized with sequences in *E. coli* DNA (Fig. 1, lanes 5).

Although we could not attribute bands in nonmammalian DNAs (Fig. 1) to rprotein genes *per se*, we concluded that sequences complementary to one or more of the mammalian rprotein probes are contained in all of the eukaryotic DNAs examined. A yeast genomic DNA fragment containing a sequence virtually identical to mammalian S14 cDNA has been isolated by others and shown to encode rprotein rp59 (ref. 29; see below). Thus it is likely that the bands observed in nonmammalian DNAs (Fig. 1) reflect heterologous cross-reacting rprotein sequences.

Ribosomal Proteins Homologous to Human S14. When the structure of yeast rprotein rp59 was reported (30), we were impressed by the similarity between the sequence of its carboxyl-terminal amino acids and that of mammalian S14 protein (18). This portion of the mammalian S14 coding sequence includes nucleotides affected by emetine resistance mutations in Chinese hamster cells (18), and for this reason it appears to be important for S14's function. Dot matrix comparison of yeast rp59 and human S14 amino acid sequences (Fig. 2) indicated striking homology. The homology was particularly intriguing, because rp59 is the protein affected by yeast *cryl* (cryptopleurine resistance) mutations (29, 32), and *cryl* mutations often confer cross-resistance to emetine (33). Furthermore, homology between human S14 and yeast rp59 involves the two proteins through virtually their entire lengths, despite the fact that their amino acid sequences differ at several residues (Fig. 3).

Motivated by the similarity between mammalian S14 and yeast rp59, we surveyed protein and DNA databases for other sequences similar to human rprotein S14. We recognized homology between S14 and *E. coli* rprotein S11. These amino acid sequences are aligned in Fig. 3.

Alignments in Fig. 3 are statistically extremely significant (RDF scores of 42.9 SD and 13.1 SD for S14 vs. rp59 and S14 vs. S11, respectively). It should be noted, however, that much of the homology between S14 and S11 reflects conservative amino acid differences (underlined in the figure), not identities. The alignments involve a few proposed gaps: one long one in rp59 and six short ones in S11. The mammalian and yeast rproteins share 109 out of 151 residues, and the bacterial and mammalian proteins are identical in 56 of 130 positions. The great majority of differences that distinguish rp59 and S11 from mammalian S14 are conservative. Proteins S11 and S14 possess very basic amino- and carboxyl-terminal sequences, features of the two proteins that have been noted previously (18, 34). Arginine residues at positions 149 and 150 (Fig. 3), which are mutationally altered in emetine resistant Chinese hamster cells, are present in rp59

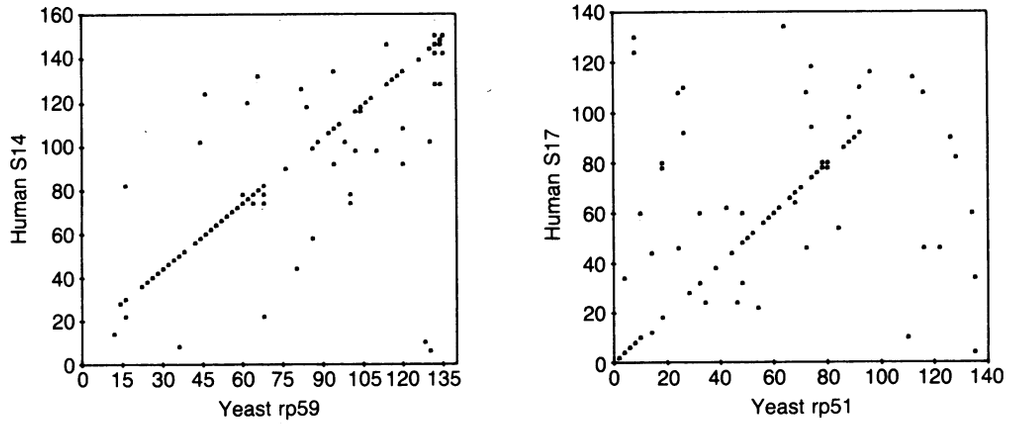

FIG. 2. Comparison of human and yeast rprotein sequences. Amino acid sequences of human S14 (20) and S17 (unpublished) were deduced from cloned cDNA sequences isolated in our laboratory. Yeast rp59 (30) and rp51 (31) amino acid sequences were determined from cloned genomic DNA fragments. To minimize sporadic amino acid similarities expected by chance alone, homology reported in the dot matrix comparisons required that two consecutive residues match.

```
                          10         20         30
Mammalian S14       MAP   RKGKEKKEEQVIS-LGPQVAEGENVFGVCHIFAS
Yeast rp59          .                .!NVV.ARDNSQ....AR.Y..
E. coli S11         AK..IRA..RVR.      -   ..SD.     .A..H..

                     40         50         60         70
Mammalian S14       FNDTFVHVTDLSGK!ETICRVTGGMKVKADRDESSPYAAM
Yeast rp59          .............-...A.....................
E. coli S11         ..N.I.TI..RQ.N-ALGWATA..SGFRGS.K .T.F..Q

                     80         90         100        110
Mammalian S14       LAAQDVAQRCKELGITALHIKLRATGGNR!TKTPGPGAQS
Yeast rp59          .......AK.R.V....V.V.I.....T.-.......G.A
E. coli S11         V..ERC.DAV..Y..KN.EVMV       - .G....RE.

                     120        130        140        150
Mammalian S14       ALRALARSGMKIGRI!EDVTPIPSDSTRRKGGRRGRRL
Yeast rp59          .........LR....-.....V.C....K.........
E. coli S11         TI...NAA.FR.TN.-T......HNGC.PP  KK ..V
```

FIG. 3. Alignment of mammalian S14, yeast rp59, and *E. coli* S11. Amino acid sequences of Chinese hamster and human S14 were deduced from full-length cDNAs. The structure of yeast rp59 was determined from a genomic DNA clone (30). The structure of *E. coli* S11 was deduced both by sequence analysis of the purified protein (34) and from a cloned genomic DNA fragment (15). Amino acid residues are indicated by the following abbreviations: A, alanine; R, arginine; N, asparagine; D, aspartate; C, cysteine; E, glutamate; Q, glutamine; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; and V, valine. Spaces indicate gaps proposed to optimize alignment; dots, matching residues; underscored residues represent conservative amino acid differences compared with mammalian S14 (24). Exclamation points indicate intron positions and hyphens are used to maintain alignment.

and S11. Yeast rp59 differs from mammalian S14 and bacterial S11 primarily in its amino terminus (positions 1–30), apparently reflecting a complex sequence of genetic events.

Nucleic acid sequences of human S14 (20), yeast rp59 (30), and *E. coli* S11 (15) genes are compared in Fig. 4. Mammalian and bacterial ATG initiator codons are aligned with nontranslated sequences 42 bp upstream of the yeast rp59 initiator. Alignments involve one gap in the eukaryotic sequences and four gaps in the bacterial gene. The human and Chinese hamster genes differ in only 41 of 456 bases (91% identity), none of which affect their amino acid sequences (see Fig. 3). The human and yeast genes possess 298 identical base residues (65% identity), reflecting similarities between the amino acid sequences they encode (Figs. 2 and 3). The bacterial and human genes are identical in 193/441 positions (44%). These levels of homology are expected by chance with probabilities $<< 10^{-5}$.

The three eukaryotic ribosomal protein coding sequences each terminate in a single TGA codon (position 454), whereas the bacterial S11 gene terminates in a TAA codon (position 445). The S11 gene, part of the *E. coli* RNA polymerase α operon (15), lacks intervening sequences. The yeast rp59 and mammalian S14 genes contain them. The human gene is composed of five exons, and the yeast gene is composed of two (Fig. 5). Although the sizes of the mammalian and yeast rprotein coding sequences are approximately the same (558 vs. 562 bp), the human gene spans 5.4 kbp of chromosomal DNA (20), whereas the yeast gene is less than 0.9 kbp (30). Intron–exon junctions in the two genes are indicated by exclamation points in Fig. 3.

A Yeast rprotein Homologous to Mammalian S17. In our survey of protein and nucleic acid sequence databases, we recognized substantial similarity between human S17 and yeast rp51. Dot matrix comparison indicates striking homology that involves the amino-terminal two-thirds of their primary sequences (Fig. 2). Fig. 6 depicts an alignment among human and Chinese hamster rproteins S17 and yeast rp51 (31). Human and hamster S17 proteins are identical except for a single proline (P) vs. alanine (A) difference at position 133. Statistical analysis of homology involving the first 92 residues of the human and yeast proteins (Fig. 6)

```
                    10         20         30         40         50                  60         70         80         90         100
Human S14           ATGGCACCTCGAAAGGGGAAGGAAAAGAAGGAAGAACAGGTCATCAGCCTCG          GACCTCAGGTGGCTGAAGGAGAGAATGTATTTGGTGTCTGCCATATCTTTGCAT
CHO S14             ............C.....A.....................................          .......A.....................................C..T.......
Yeast rp59          .attttaaa.cc....a..gta.g..accac......cc.c...GTCTAA..          TTGT...A.CTCG...CAATTCCC.A..T........TGCTAGA..T.AC..T.
E. coli S11         ...                        .C..AGG..CCAAT..GTG.A..TAAACGTGTAA..AAA.AA..CT....C..C                ..GGCT......CA...T.

                     110        120        130        140        150        160        170        180        190        200        210        220
Human S14           CCTTCAATGACACTTTTGTCCATGTCACTGATCTTTCTGGCAAGGAAACCATCTGCCGTGTGACTGGTGGGATGAAGGTAAAGGCAGACCGAGATGAATCCTCACCATATGCTGCTA
CHO S14             .T.....C.....C..........T.....................T........G...........A.............T...A.......G..T..T..T.....A..C.
Yeast rp59          .T.....C..T.....C..T.....T..C...T.A.....T.............GC.A.A..T.........T.........T.....T...A..........T..T.....C.......
E. coli S11         .T.....CA....CA.C..GAC.A.........G.CAG..T..C.CGTTGGGT..GGCAACAG.C.....TTCCGGTT.CCGT.GTTCT..CA    .....A.T..G.T...A...C

                     230        240        250        260        270        280        290        300        310        320        330
Human S14           TGTTGGCTGCCCAGGATGTGGCCCAGAGGTGCAAGGAGCTGGGTATCACCGCCCTACACATCAAACTCCGGGCCACAGGAGGAAATAGGACCAAGACCCCTGGACCTGGGGCCCAGT
CHO S14             .................T.....................C.....T.....G..T...........A..T..G........C.....................A.......
Yeast rp59          .............A.....T...GCT.A...T.G...AG.C.........T...G.T...G.T..GA..A.A..T..C..T..T.C...A........T..A..T..A..T.GT..AG
E. coli S11         A.G.T..A..AG..CG.TGC..TG.CGCCGTG..A..ATAC..C....AGAAT..GG.AG.                              ..T.GTT..AGGT..G..T..A..CCG.G.A.

                     350        360        370        380        390        400        410        420        430        440        450
Human S14           CGGCCCTCAGAGCCCTTGCCCGCTCGGGTATGAAGATCGGGCGGATTGAGGATGTCACCCCCATCCCCCTCTGACAGCACTCGCAGGAAGGGGGGTCGCCGTGGTCGCCGTCTGTGA
CHO S14             .A...........T.....T..T..A..G........T.....................................C..A........T.....T.................
Yeast rp59          .T..TT.G.....TT.G...A.A..T...T...GA..T..C..T..C..A.....T.....AG.T..A.G....TC...CA.A.A......T...A.AA.A...A.gaaga.t...
E. coli S11         .TA.TA.TC.T..T..GAA.GC.G.A...T.CCGC...ACTAAC...ACT.....G..T..G.....TCA.A..G.TTG...TCC.CC.AAAAAA..T..C.TATAA..c..cgtt
```

FIG. 4. Human and Chinese hamster S14, yeast rp59, and *E. coli* S11 DNA sequences. Human and Chinese hamster S14 cDNA and *E. coli* S11 genomic sequences were aligned at their ATG initiator codons. The yeast rp59 initiator codon is located at position 42 in the figure. Terminator codons in each sequence are as follows: human and hamster, TGA at position 454; yeast, TAG at position 444; and *E. coli*, TAA at position 445. Residues identical to the human sequence are indicated by dots. Spaces represent gaps proposed to optimize alignment. Underlined residues are shared by at least two of the four genes. Lowercase indicates flanking, noncoding sequence.

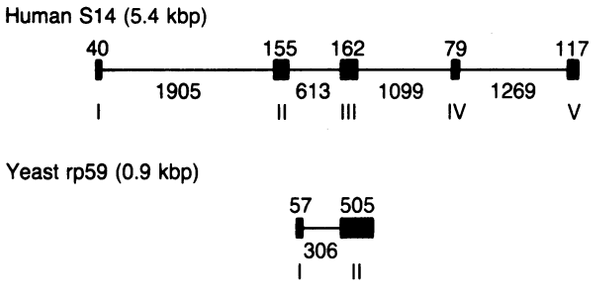

FIG. 5. Exon–intron organization in mammalian S14 and yeast rp59 rprotein genes. Exons are represented by Roman numbered boxes; introns, by lines. Lengths of each segment are indicated in bp.

indicated a high degree of significance (57.2 SD). Virtually all amino acid differences in this portion of the proteins are conservative (underlined). From residue 93 through residue 130 approximately half of the differences are nonconservative.

The nucleic acids encoding mammalian S17 and yeast rp51 can be aligned with two gaps in the mammalian sequence and one in the yeast gene (Fig. 7). That all three gaps are in the 3′ half of the proteins' coding sequences suggests a history of genetic recombinations 3′ to residue 276 in Fig. 7. The 5′ 276 bp of the human and yeast gene sequences are 67% identical, despite the fact that human S17 cDNA probes do not detect a yeast genomic DNA sequence (Fig. 1*B*). In their regions of greatest similarity (Fig. 7, residues 143–186), S17 and rp51 share 36 of 44 (82%) nucleic acid bases. In contrast, S14 and rp59 share 66 of 73 residues (90% identity) between positions 169 and 242 (Fig. 4). This small difference appears to account for the failure of S17 cDNA to detect a yeast genomic DNA sequence under the relatively stringent nucleic acid hybridization conditions employed (Fig. 1*B*).

DISCUSSION

Models accounting for evolution of the translational apparatus have been difficult to formulate (6, 35–37). Substantial heterogeneity among analogous rproteins is responsible in large part for the lack of a universal, systematic rprotein nomenclature. Now, through application of recombinant DNA technology, rprotein gene sequences can be determined rapidly and their structures can be compared directly.

The hypothesis that an efficient, accurate translation apparatus evolved early in biological time, allowing primitive organisms to express phenotypic traits precisely, motivates the search for structural similarities among the ribosomes of

```
                    10        20        30        40
Human S17   MGRVRTKTVKKAARVIIEKYYTRLGNDFHTNKRVCEEIAIIPSKKLR
CHO S17     ...........................................
Yeast rp51  ..........R.SKAL..R..PK.TL..Q....L.D...T.Q..R..

                    60        70        80        90
Human S17   NKIAGYVTHLMKRIQRGPVRGISIKLQEEERERRDNYVPEVSALDQE
CHO S17     ...........................................
Yeast rp51  ......T........K.......F.........K.Q.........LS

                   100       110       120       130
Human S17   IIE    VDPDTKEMLKLLDFGS LSNLQVTQPTVGMNFKTPRGPV
CHO S17     ...    ................................A.
Yeast rp51  RSNGVLN..NQ.SDLV.S.GLKLP..VIN.SAQRD.RYR.RV
```

FIG. 6. Amino acid sequences of mammalian rprotein S17 and yeast rp51. Human and Chinese hamster S17 amino acid sequences were deduced from clones cDNAs (I.-T.C. and D.J.R., unpublished data). The yeast rp51 amino acid sequence was derived from a cloned yeast genomic DNA fragment (31). Other features of the figure are as described in the legend to Fig. 3.

diverse species. If found, these similarities might provide a means to distinguish rproteins that were fixed early in evolution and thus are likely to be fundamental components of the translation process. Woese (37) suggested that differences among ribosomes in contemporary organisms are likely to reflect relatively recent, small evolutionary adaptations of the translation apparatus for specific purposes.

Nucleic acid similarities among rprotein genes in phylogenetically distant organisms support this view of the evolution of the translation apparatus. Mammalian S14 and S17 families of rprotein genes appear to have been conserved stringently during evolution. Nucleic acid homologies among human S14 cDNA, insect, and yeast genomic DNAs are evident on Southern blot filters (Fig. 1*A*). Comparison of yeast rp59 and mammalian S14 nucleic acids confirmed this homology (Figs. 2–4). Human S17 cDNA detected complementary nucleic acid sequences in insect but not yeast DNA (Fig. 1*B*). Despite this, comparison of human S17 and yeast rp51 sequences indicated impressive similarities at both the protein and nucleic acid levels (Figs. 2, 6, and 7). Thus, it is possible that substantial homology among the rprotein genes of these and other species could be detected by careful, stepwise reduction in the stringency of nucleic acid hybridization conditions.

In accord with the model of ribosome evolution described above, S14 and S17 appear to be ribosome components fixed early in biological history. Presumably, because of their fundamental roles in the translation process and interactions with other ribosomal components, these gene products have been maintained stringently during the course of evolution. In sharp contrast, mammalian L32 appears to be a recent addition to the rprotein family or, alternatively, its gene is highly polymorphic. L32 cDNA detected no homology in heterologous genomic DNAs surveyed (Fig. 1*C*). It hybridized to mammalian but not insect, yeast, or bacterial DNA. Failure to detect L32-like genes in heterologous DNAs by nucleic acid hybridization does not preclude their existence. However, in contrast to the results with S14 and S17, a survey of protein databases did not uncover sequences similar to murine L32 (data not shown).

Mammalian S14 and yeast rp59 proteins appear to be involved in the ribosome-catalyzed translocation cycle. Mutations in their genes result in 40S ribosomal subunits that support drug-resistant protein biosynthesis *in vitro* (21, 38) as well as *in vivo* (21). The 40S ribosomal subunits from emetine-resistant Chinese hamster ovary cell mutants are unusually sensitive to high ionic strength ribosome wash buffers. Mutant 40S ribosomal subunits release a large subset of proteins, including S14, when exposed to buffers containing 0.5 M KCl or NH_4Cl (27, 39). This testifies to the role of S14 in maintaining the structural integrity of eukaryotic 40S ribosomal subunits.

Few examples of homologous, transcriptionally active eukaryotic genes that display markedly dissimilar intron–exon structures are known. Data summarized in Figs. 3 and 4 leave little doubt as to the structural homology between mammalian S14 and yeast rp59 genes. Yet, from their amino acid compositions, the two proteins appear to differ significantly in molecular weight and net charge. The calculated molecular weight of S14 is 16,275, and S14 is predicted to display a net charge of +11 at neutral pH. The molecular weight of rp59 is 14,582, and its net charge is +12. As illustrated in Fig. 5, these genes exhibit totally different architectures. None of the S14 introns have a counterpart in the yeast gene. The single rp59 intron is much shorter than the first S14 intron. The rp59 intron interrupts the gene's third codon, whereas the first S14 intron separates 5′ leader sequence from coding sequence (Fig. 3 and ref. 20). The S14 family of rprotein genes therefore provides an unusual example of gene evolution in which the number, length, and distribution

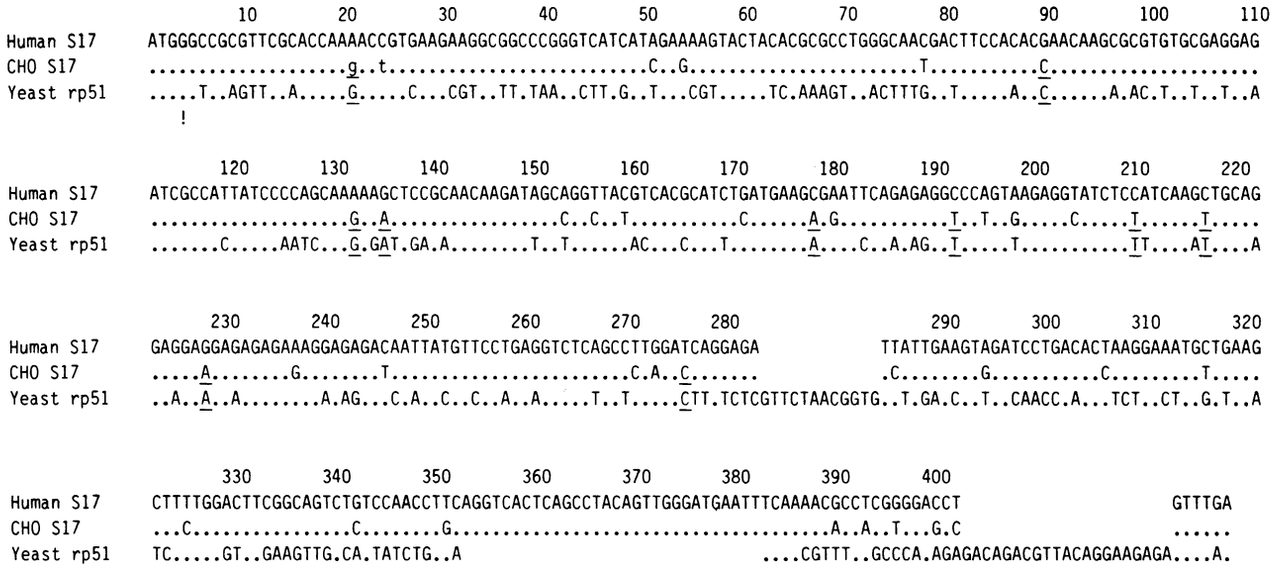

FIG. 7. Mammalian S17 and yeast rp51 rprotein genes. The DNA sequences are aligned at their initiator ATG codons. Both mammalian genes terminate in a TGA codon, whereas the yeast gene terminates in TAA. The exclamation point under position 4 of the yeast sequence indicates the location of a 397-bp intervening sequence (31).

of intronic sequences vary substantially, despite remarkable conservation of polypeptide coding (exon) sequences.

The authors thank Mses. Maggy Weston-Fina and Barbara Van Slyke for their excellent technical assistance. This research was supported by National Institutes of Health Grant GM23013 to D.J.R. and is contribution number 86-238-J of the Kansas State Agricultural Research Station. D.D.R. was supported by National Institutes of Health Training Grant CA09418.

1. Nomura, M., Tissieres, A. & Lengyel, P., eds. (1974) *Ribosomes* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
2. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M., eds. (1980) *Ribosomes: Structure, Function, and Genetics* (University Park Press, Baltimore).
3. Wittmann, H. G. (1982) *Annu. Rev. Biochem.* **52,** 35–65.
4. Thompson, J. F. & Hearst, J. E. (1983) *Cell* **33,** 19–24.
5. Noller, H. F. (1984) *Annu. Rev. Biochem.* **53,** 119–162.
6. Lake, J. A. (1985) *Annu. Rev. Biochem.* **54,** 507–530.
7. Meyuhaus, O. & Perry, R. P. (1980) *Gene* **10,** 113–129.
8. Vaslet, C. A., O'Connell, P., Izquierdo, M. & Rosbash, M. (1980) *Nature (London)* **285,** 674–676.
9. Bollen, G. H. P. M., Cohen, L. H., Mager, W. H., Klassen, A. W. & Planta, R. J. (1981) *Gene* **14,** 279–287.
10. Bozzoni, I., Beccari, E., Luo, Z.-X. & Amaldi, F. (1981) *Nucleic Acids Res.* **9,** 1069–1086.
11. Fried, H. M., Pearson, N. J., Kim, C. H. & Warner, J. R. (1981) *J. Biol. Chem.* **256,** 10176–10183.
12. Burns, D. K., Stark, B. C., Macklin, M. D. & Chooi, W. Y. (1984) *Mol. Cell. Biol.* **4,** 2643–2652.
13. Dudov, K. P. & Perry, R. P. (1984) *Cell* **37,** 457–468.
14. Wiedemann, L. M. & Perry, R. P. (1984) *Mol. Cell. Biol.* **4,** 2518–2528.
15. Bedwell, D., Davis, G., Gosink, M., Post, L., Nomura, M., Kestler, H., Zengel, J. M. & Lindahl, L. (1985) *Nucleic Acids Res.* **13,** 3891–3903.
16. Wagner, M. & Perry, R. (1985) *Mol. Cell. Biol.* **5,** 3560–3576.
17. Nakamichi, N., Rhoads, D. D. & Roufa, D. J. (1983) *J. Biol. Chem.* **258,** 13236–13242.
18. Rhoads, D. D. & Roufa, D. J. (1985) *Mol. Cell. Biol.* **5,** 1655–1659.
19. Nakamichi, N., Kao, F.-T., Wasmuth, J. & Roufa, D. J. (1986) *Somatic Cell Mol. Genet.* **12,** 225–236.
20. Rhoads, D. D., Dixit, A. & Roufa, D. J. (1986) *Mol. Cell. Biol.*, in press.
21. Boersma, D., McGill, S., Mollenkamp, J. & Roufa, D. J. (1979) *J. Biol. Chem.* **254,** 559–567.
22. Southern, E. (1975) *J. Mol. Biol.* **98,** 503–517.
23. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1444.
24. Dayhoff, M. (1978) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, MD), Vol. 5, Suppl. 3.
25. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
26. Madjar, J.-J. & Traut, R. R. (1980) *Mol. Gen. Genet.* **179,** 889–901.
27. Madjar, J.-J., Frahm, M., McGill, S. & Roufa, D. J. (1983) *Mol. Cell. Biol.* **3,** 190–197.
28. Klein, A. & Meyuhaus, O. (1984) *Nucleic Acids Res.* **12,** 3763–3776.
29. Larkin, J. C. & Woolford, J. L., Jr. (1983) *Nucleic Acids Res.* **11,** 403–420.
30. Teem, J. L., Abovich, N., Kaufer, N. F., Schwindinger, W. F., Warner, J. R., Levy, A., Woolford, J., Leer, R. J., van Raamsdonk-Duin, M. M. C., Mager, W. H., Planta, R. J., Schultz, L., Friesen, J. D., Fried, H. & Rosbash, M. (1984) *Nucleic Acids Res.* **12,** 8295–8312.
31. Teem, J. L. & Rosbash, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4403–4407.
32. Himmelfarb, H. J., Vassarotti, A. & Friesen, J. D. (1984) *Mol. Gen. Genet.* **195,** 500–506.
33. Sanchez, L. D., Vasquez, D. & Jimenez, A. (1977) *Mol. Gen. Genet.* **156,** 319–326.
34. Kamp, R. & Wittmann-Liebold, B. (1980) *FEBS Lett.* **121,** 117–122.
35. Matheson, A. T., Moller, W., Amons, R. & Yaguchi, M. (1980) in *Ribosomes: Structure, Function, and Genetics*, eds. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M. (University Park Press, Baltimore), pp. 297–332.
36. Osawa, S. & Hori, H. (1980) in *Ribosomes: Structure, Function, and Genetics*, eds. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M. (University Park Press, Baltimore), pp. 333–355.
37. Woese, C. R. (1980) in *Ribosomes: Structure, Function, and Genetics*, eds. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M. (University Park Press, Baltimore), pp. 357–373.
38. Gupta, R. S. & Siminovitch, L. (1976) *Cell* **9,** 213–219.
39. Wejksnora, P. J. & Warner, J. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5554–5558.