recruited from other Pediatric Oncology Group (POG) institutions before treatment began. Informed consent was obtained, and procedures approved by the Committee on Human Research at the University of Vermont and other POG institutions were followed.
10. Peripheral blood was separated and the mononuclear cell fraction was obtained for the T cell cloning assay within 12 to 24 hours of its collection. The T cell cloning assay and analysis have been described [see B. A. Finette *et al.*, *Mutat. Res.* **308**, 223 (1994)].
11. B. A. Finette *et al.*, *Mutat. Res.* **308**, 223 (1994).
12. Mutant frequencies in subjects at diagnoses, remission, and relapse were compared with normal controls by the nonparametric Kruskal–Wallis test. Pairwise differences between groups were assessed by Mann–Whitney tests, with a Bonferroni adjustment for multiple comparisons. Linear regression analysis was used to examine the relationship between the logarithm of Mf (lnMf) and months since diagnosis.
13. H. Hirota *et al.*, *Mutat. Res.* **315**, 95 (1994).
14. S. Koishi *et al.*, *Mutat. Res.* **422**, 213 (1998).
15. R. A. Gibbs, P.-N. Nguyen, A. Edwards, A. B. Civitello, C. T. Caskey, *Genomics* **7**, 235 (1990).
16. J. C. Fuscoe *et al.*, *Cancer Res.* **51**, 6001 (1991).
17. B. A. Finette, J. P. O'Neill, P. M. Vacek, R. J. Albertini, *Nat. Med.* **4**, 1144 (1998).
18. N. F. Cariello *et al.*, *Nucleic Acids Res.* **26**, 198 (1998).
19. Because the *HPRT* gene is located on the X chromosome, molecular analysis at the DNA/RNA level was performed in a different way for mutant isolates from males and females [B. A. Finette, J. P. O'Neill, P. M. Vacek, R. J. Albertini, *Nat. Med.* **4**, 1144 (1998)].
20. S. Tonegawa, *Nature* **302**, 575 (1983).
21. R. J. Albertini, J. A. Nicklas, T. R. Skopeck, L. Recio, J. P. O'Neill, *Mutat. Res.* **400**, 381 (1998).
22. To determine the clonality of peripheral *HPRT* mutant T cell clones we performed a two-step polymerase chain reaction (PCR) with about 10^4 cells from each isolate to perform a reverse transcriptase–PCR amplification followed by a TCRβ gene hot start PCR with a V_β and C_β consensus primer mix. The resulting PCR product was purified and the highly polymorphic CDR3/variable regions of TCRβ were sequenced.
23. Supplementary material is available at www.sciencemag.org/feature/data/1047822.shl
24. Single-letter abbreviations for amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
25. R. F. Branda, J. P. O'Neil, L. M. Sullivan, R. J. Albertini, *Cancer Res.* **51**, 6603 (1991).
26. B. S. Strauss, *Carcinogenesis* **18**, 1445 (1997).
27. B. S. Strauss, *Semin. Cancer Biol.* **6**, 431 (1998).
28. J. H. Miller, *Annu. Rev. Microbiol.* **50**, 625 (1996).
29. E. F. Mao, L. Lane, J. Lee, J. H. Miller, *J. Bacteriol.* **179**, 417 (1997).
30. C. R. Hunt *et al.*, *Cancer Res.* **58**, 3986 (1998).
31. Supported by National Institute of Child Health & Human Development grant 1R29HD35309, National Cancer Institute (NCI) grant 1K01CA77737, grant 6103-98 from the Leukemia Society of America, the Lake Champlain Cancer Research Organization, the Vermont Chapter of the American Cancer Society, the Art Ehrmann Cancer Fund and NCI grant P30CA22435 to the University of Vermont Cancer Center DNA Analysis Facility. We thank C-C. Duriez-Lu, H. Kendall, and J. Rivers for technical assistance; S. Billado for assistance with blood samples; and P. Vacek for statistical analyses.

9 December 1999; accepted 28 February 2000

Homologs of Small Nucleolar RNAs in Archaea

Arina D. Omer,[1] Todd M. Lowe,[2]* Anthony G. Russell,[1] Holger Ebhardt,[1] Sean R. Eddy,[2] Patrick P. Dennis[1]†

In eukaryotes, dozens of posttranscriptional modifications are directed to specific nucleotides in ribosomal RNAs (rRNAs) by small nucleolar RNAs (snoRNAs). We identified homologs of snoRNA genes in both branches of the Archaea. Eighteen small sno-like RNAs (sRNAs) were cloned from the archaeon *Sulfolobus acidocaldarius* by coimmunoprecipitation with archaeal fibrillarin and NOP56, the homologs of eukaryotic snoRNA-associated proteins. We trained a probabilistic model on these sRNAs to search for more sRNAs in archaeal genomic sequences. Over 200 additional sRNAs were identified in seven archaeal genomes representing both the Crenarchaeota and the Euryarchaeota. snoRNA-based rRNA processing was therefore probably present in the last common ancestor of Archaea and Eukarya, predating the evolution of a morphologically distinct nucleolus.

Ribosome biogenesis in Eukarya occurs in the nucleolus. Several nucleolar proteins (NOPs), including fibrillarin, Nop56, and Nop58, and dozens of snoRNAs are involved in this process (*1*). The snoRNAs fall into two major classes: C/D box and H/ACA box RNAs. The C/D box snoRNAs are efficiently precipitated with antibodies against fibrillarin. Most C/D box snoRNAs target specific ribose methylations within rRNA, whereas most H/ACA box RNAs target specific conversions of uridine to pseudouridine within rRNA (*2*).

The general mechanism of C/D box snoRNA-targeted ribose methylation has been well established. Each snoRNA contains a 9- to 21-nucleotide (nt)–long sequence, located 5′ to the D or D′ box motif, that is complementary to an rRNA target sequence. Methylation is directed to the rRNA nucleotide that participates in the base pair 5 nt upstream from the start of the D or D′ box. It is likely that most, if not all, eukaryotic rRNA ribose methylations are guided by snoRNAs. In the yeast *Saccharomyces cerevisiae*, methylation guide snoRNAs have been assigned to all but four of the 55 rRNA ribose methylation sites (*3*).

SnoRNAs, which are apparently ubiquitous in Eukarya, have not been found in Bacteria or Archaea. However, the rRNA of the archaeon *Sulfolobus solfataricus* (*Sso*) has been shown to contain 67 ribose methylation sites, a number similar to that found in eukaryotes (*4*). Even though Archaea are unicellular prokaryotic organisms that lack a nucleolus, their genomes encode homologs to the essential eukaryotic nucleolar proteins, fibrillarin and NOP56/58 (*5*, *6*). On the basis of these observations, we decided to examine Archaea for the presence of sno-like RNAs (sRNAs).

To isolate sRNAs from the archaeon *Sulfolobus acidocaldarius* (*Sac*), we cloned the *S. acidocaldarius* homologs of the eukaryotic fibrillarin and NOP56/58 proteins, designated aFIB and aNOP56, using sequence information from a related species, *S. solfataricus* (*7*). The cloned genes were expressed in *Escherichia coli*, and the recombinant proteins were purified and used to raise polyclonal antibodies in rabbits. The two antibody preparations were each highly specific and recognize single polypeptides of the predicted size in total *S. acidocaldarius* cell extracts (Fig. 1A). The antibodies were used to monitor the size distribution of particles containing aFIB and aNOP56 in a glycerol gradient fractionation of partially purified cell lysate (Fig. 1A) (*8*). Both aFIB and aNOP56 sedimented as a large heterogeneous complex.

To detect RNAs that associate with aFIB- and aNOP56-containing complexes, we immunoprecipitated aliquots from gradient fractions with either antibody to aFIB or antibody to aNOP56. Total RNA was extracted with phenol from the supernatants and the pellets, and a portion from each was 3′ end-labeled with ^{32}P-cytidine-5′,3′-bis-phosphate (pCp) and displayed by denaturing polyacrylamide gel electrophoresis (Fig. 1, B and C). The most abundant RNAs that were coimmunoprecipitated appear as a family of bands ranging in length from about 50 to 70 nt. This size class of RNAs, which is substantially shorter than eukaryotic C/D box snoRNAs, was invisible when total cellular RNA was labeled with pCp. To obtain cDNA clones, we gel-purified the RNAs precipitated from fraction 5 with antibody to aFIB and from

[1]Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada. [2]Department of Genetics, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA.

*Present address: Department of Genetics, M322, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305–5120, USA.
†To whom correspondence should be addressed. E-mail: patrick.p.dennis@ubc.ca

Table 1. The sequences of *S. acidocaldarius* (*Sac*) cDNA clones are aligned with the C, D′, C′, and D boxes as anchors. Dashes are gaps in the alignment. The GenBank accession numbers for the sRNA sequences are AF195095 through AF195112.

		C box		D′ box		C′ box			D box	
Sac										
sR1	CAG	UUGAUGA	--GAAGUUAAAAAA	GCGA	-----	UGGAUGA	------GCUUAACUCCCAUGGU	CUGA	UAAC
sR2	GA	GUGAUGA	--GACGAGCGCUAA	CAGA	-GAGA	GUGAAGA	----------GGUCACUGCGAA	CUGA	AGAAA
sR3	AGG	AUGACGA	---GACCCAAAAUA	UUGA	-----	ACGAUGA	-------UAUAACCUGUCUCGG	CUGA	UCAGU
sR4	G	UUGAUGA	--GCACAUUCUUUU	CUGA	-UUUA	AUGAAGA	----------AAGUGGCCAGGU	CUGA	GGUAG
sR5	GAA	AUGAUGA	-AUGGUCGACGGAA	CGGA	--CCU	AUGAAGA	----------AUUGUUGCCGGA	CUGA	CAAAC
sR6	GG	AUGAUGA	-----CCAAAUAGA	CUGA	--AAG	AUGAAGA	----------AAUGCACCUCAA	CUGA	CUAAA
sR7	G	AUGAUGA	--CAAAGAGCCGAA	UGGA	-----	UUAGUGA	CAUCUAAUUUUGUGGGCAGCCA	CUGA	UAGAG
sR8	G	AUGAUGA	-AGCCCGCCAUCAA	CAGA	--UAA	GUGAAGA	----------GGGAACCCGAGG	CUGA	GAAU
sR9	GUUAAAAUA	AUGAUGA	--CUAACUCCAAUA	CUGA	--CCA	AUGAUGU	----------CGUAACCCGAAA	CUGA	AUAAA
sR10	GA	AUGAUGU	--GGAAUCCGGGAU	CUGA	---GA	AUGAUGA	-------CAAAAAGCGCGAGCG	CUGA	UUAUA
sR11	GAAU	GUGAUGA	-UGGGUCGAUGUUA	CUGA	-UUAG	UUGAUGA	----------GAUUAUCUCCGG	CUGA	GAAU
sR12	GA	AUGAAGA	--ACCCAACCUUAU	CUGA	-GGUU	AUGAUGA	----------CAGGUUGUUCGU	CAGA	UCGAUGUGAG
sR13	AGG	AUGAUGU	-ACUUUCACCCUCA	CUGA	--AAG	GUGAGGA	----------UGAGUCCGACUA	CUGA	CGCAA
sR14	GCU	GUGAAGA	-CGCUAGACUUAGA	CUGA	--CUC	AUGAUGA	----------AGGGCCAAAGCU	CAGA	GCAAAC
sR15	A	GUGAUGA	GGAACCAACGAGAG	CUAG	----U	UUGAUGG	-------CUUCGACGCUCUGCU	CUGA	AA
sR16	GA	AUGAAGA	--CGUUCCACCCGA	GCGA	-----	GUGAUGA	------GCGAAACGGUUAAUA	CUGA	UGAUG
sR17	AGAA	AUGAAGA	--CUAAAAAACCGG	CUGA	GAUAA	GUGAUGA	----------CGACGUCUCGCA	CUGA	UC
sR18	AA	GUGAUGA	--CAGAACCCCGGC	UUGA	--AAG	AUGAUAG	---------AGCCGUGUGAGAA	CUGA	UCAAU

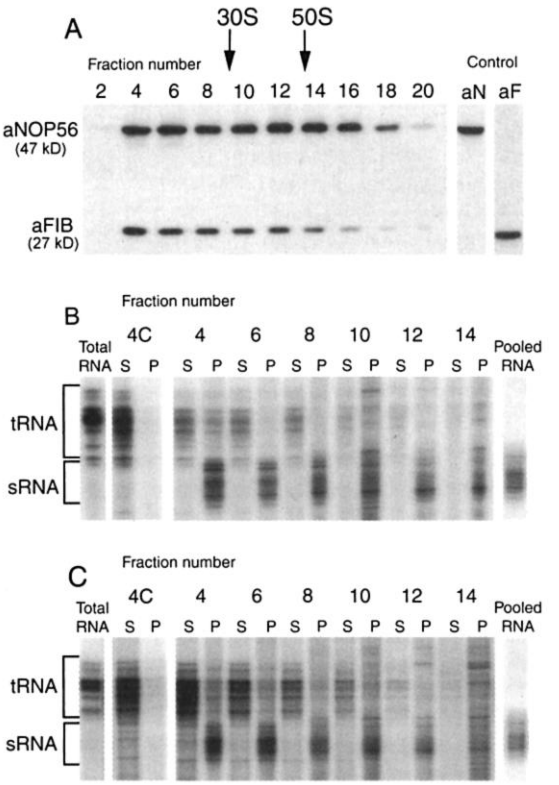

Fig. 1. Glycerol gradient sedimentation of aFIB- and aNOP56-containing particles present in *S. acidocaldarius* cell-free extracts. A sonicated cell extract was precipitated by addition of 35% ammonium sulfate, redissolved in buffer (50 mM tris, pH 8), layered onto a 35-ml 10 to 30% glycerol gradient in the same buffer, and sedimented in an SW27 rotor (10°C, 17 K, 16 hours). Fractions (1.5 ml) were collected. (**A**) Aliquots of every second fraction between 2 and 20 were simultaneously analyzed by Western blotting for the presence of aFIB and aNOP56 with the two antibodies prepared against the recombinant proteins expressed and purified from *E. coli*. The positions of 30*S* and 50*S* ribosomal subunits in the gradient are indicated. In the control, the aFIB and aNOP56 antibodies were shown to be highly specific for single polypeptides of the expected size (27 kD and 47 kD, respectively) in *S. acidocaldarius* crude cell extract (right). (**B**) Aliquots from every other gradient fraction between 4 and 14 were immunoprecipitated with antibody to aFIB (*8*), and RNA was recovered by phenol extraction from the precipitates (P) and the supernatants (S). Only about 0.1% of the RNA in each fraction was coprecipitated with the antibody; the bulk of the RNA was retained in the supernatant. As a control (4C), an aliquot of fraction 4 was immunoprecipitated with preimmune serum. To visualize the precipitated RNAs, we pCp end-labeled aliquots (0.005% and 2.5% of the total RNAs recovered from the supernatant and pellets, respectively) with RNA ligase and displayed them on an 8% denaturing polyacrylamide gel. The positions of tRNA and sRNA are indicated on the left. The precipitated RNA recovered from fraction 5 was separated on an 8% denaturing polyacrylamide gel, recovered by electroelution, and used as a template for RT-PCR cloning (*9*). An aliquot of the RNA recovered after electroelution was end-labeled and displayed on an 8% denaturing polyacrylamide gel (right). (**C**) Aliquots from every other gradient fraction between 4 and 14 were immunoprecipitated with antibody to aNOP56. Other details are as described above, except that recovered RNAs from fractions 6 to 8 and 10 to 13 were pooled and used for cDNA cloning. An aliquot of the pooled RNA was end-labeled and displayed on an 8% denaturing polyacrylamide gel (right).

fractions 6 to 8 and 10 to 13 with antibody to aNOP56, ligated them to the oligonucleotide AO30, used them as template for reverse transcription polymerase chain reaction (RT-PCR), and cloned them (*9*).

A total of 104 clones from the two immunoprecipitated RNA pools were sequenced. From these, one or more representatives of 18 different sequences that exhibited features characteristic of eukaryotic C/D box snoRNAs were recovered (Table 1) (*2*). Other clones contained small fragments of *S. acidocaldarius* 16*S*, 23*S*, and 5*S* rRNAs. The snoRNA-like clones contained well-defined C and D box motifs located near their 5′ and 3′ ends, respectively, and recognizable internal C′ and D′ box motifs, giving the RNAs a dyad repeat structure characteristic of eukaryotic methylation guide snoRNAs (*10*).

Primer extension analysis was used to confirm the presence of sRNAs within total RNA extracted from *S. acidocaldarius*. Each sRNA primer was designed to overlap the D box motif, the adjacent guide region, and a portion of the C′ box motif. Extension products were obtained for sR1 to sR17 (clone sR18 was identified later and not tested); a subset of these is illustrated in Fig. 2. The lengths of the products for all sRNAs were within 2 nt of the 5′ ends of the cDNA, except for sR3, 4, 6, and 8, which were between 3 and 5 nt longer than the respective cDNA clones (*11*).

To find additional homologs of our cloned *S. acidocaldarius* sRNA genes, we ran BLASTN on each cDNA clone against the nonredundant nucleotide database (*12*) and recovered two weak hits against sequences in other *Sulfolobus* species: *Sac*-sR3 had a hit near the *Sulfolobus shibatae* top6B topoisomerase II gene (score = 40.1 bits, expectation value = 0.038), and *Sac*-sR1

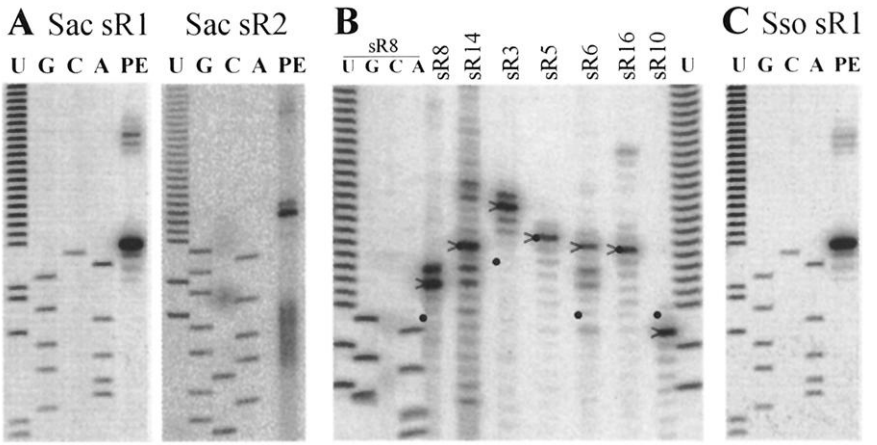

Fig. 2. Detection and 5′ end mapping of sRNAs from *S. acidocaldarius* and *S. solfataricus*. Primers specific for the D box guide region of *Sac* sR1 to sR17 were 5′ end-labeled with γ^{32}P-ATP and polynucleotide kinase and used in extension reactions with total RNA (10 μg) isolated from *S. acidocaldarius* as template. **(A)** The extension products obtained with *Sac* sR1 and sR2 specific oligonucleotide primers were run alongside a ^{33}P-DNA sequence ladder generated with the same primers and *Sac* sR1 or sR2 cDNAs as template. **(B)** The extension products obtained with *Sac* sR8, sR14, sR3, sR5, sR6, sR16, and sR10 specific primers and run with the DNA sequence ladder generated with *Sac* sR8 cDNA clone. The main sR8 extension product is 3 nt longer than the 5′ end of the sR8 cDNA clone. For each extension reaction, the major extension product (>) and the approximate positions of the 5′ terminal nucleotide in the corresponding cDNA clone (•) are indicated beside the lane. **(C)** The primer extension reaction was as in (A), except that the primer was specific to *Sso* sR1 and total RNA from *S. solfataricus* was used as template. The DNA ladder was generated with *Sac* sR1 primer and the *Sac* sR1 cDNA clone as template. The *Sac* and *Sso* primers are complementary to the same region but differ at two internal positions.

had a hit that partially overlapped with the *S. solfataricus* aspartate aminotransferase gene (score = 38.2 bits, expectation value = 0.15). Although these candidates contained canonical C and D boxes, their authenticity as true sRNAs remained questionable because of their low scores. We tested for the presence of the *Sac*-sR1 homolog by primer extension analysis using *S. solfataricus* RNA as a template. A product with a length similar to that of *Sac*-sR1 was detected (Fig. 2C) and was designated *Sso*-sR1. Primer extension products for cloned *S. acidocaldarius* RNAs sR1 to sR17 and the apparent *S. solfataricus* sR1 homolog demonstrate the existence of archaeal snoRNA-like C/D box sRNAs.

To determine if these sRNAs might guide ribose methylation as in eukaryotes, we examined the sRNAs for potential guide sequences by comparison with *S. acidocaldarius* rRNA (*13*). Regions complementary to rRNA and adjacent to the D or D′ boxes were identified for 14 of the sRNAs (Table 2). Using the D/D′ box plus 5 nt rule, we predicted the locations of potential ribose methyl modifications in rRNA and experimentally tested for some of these sites using the deoxyribonucleotide triphosphate (dNTP) concentration-dependent primer extension assay (*3*, *14*). In this assay, ribose 2′-*O*-methyl sites cause characteristic pauses that are displayed in the reverse transcriptase reactions at low but not at high dNTP concentrations. We identified characteristic pauses at six predicted sites of methylation in *S. acidocaldarius* rRNA (Table 2). Several examples are shown in Fig. 3. Both *Sac*-sR1 and *Sso*-sR1 were predicted to target methylation to position U52 in the respective 16*S* rRNAs; pause sites were detected at this position in both rRNAs. Two of the sRNAs, sR10 and sR14, exhibit strong complementarities to *S. solfataricus* tRNAs (Table 2). The target nucleotide for sR14 is C34, the anticodon "wobble" base, which is commonly ribose methylated in eukaryotes (*15*). Not all eukaryotic C/D box snoRNAs containing complementary regions participate in ribose methylation (i.e., U3 and U8), so methylation guide function should not be assumed for all archaeal C/D box sRNAs. Gene disruption systems for *S. acidocaldarius* and most other Archaea are currently not available; consequently, we were not able to verify loss of predicted methylation sites upon disruption of sRNA genes. However, our evidence suggests that many of

Table 2. Annotations of *S. acidocaldarius* sRNAs.

Sac	Ab*	PE Conf†	Guide	Target‡	Match§	Notes‖
sR1	F	+	D	16*S* U52	11/0	CDP, Mod
sR2	F	+	D	23*S* C1914	11/0	CDP
sR3	F	+	D	23*S* G2739	10/0	No pause
sR4	N	+	D	23*S* G1995	10/0	
sR5	F, N	+	D	16*S* G1056	12/0	CDP, Mod
sR6	F	+	D	23*S* G2666	9/1	
sR7	F	+	D′	23*S* G2649	9/0	CDP
			D	23*S* U2692	10/0	No pause
sR8	N	+	D′	23*S* U2972	9/1	
			D	23*S* G334	12/1	
sR9	F	+	D′	16*S* G926	8/0	
sR10	F	+	D′	tRNA Gly-CCC C50	12/0	
			D	23*S* C2539	9/0	
sR11	F	+	D′	23*S* A2618	10/2	
			D	23*S* A724	11/1	
sR12	F	+	D′	23*S* G1114	11/0	No pause
			D	23*S* A1134	10/1	CDP
sR13	N	+	D′	23*S* G385	10/1	
			D′	23*S* G2999	11/1	
			D	23*S* C2746	10/0	CDP
sR14	F, N	+	D	tRNA Gln-UUG U34	10/0	
sR15	F	+		??		
sR16	N	+		??		
sR17	F	+		??		
sR18	F, N		D′	23*S* G140	9/1	

*The precipitations from which the respective sRNAs were recovered: F, antibody to aFIB; N, antibody to aNOP56. †The presence of the sRNA in total cellular RNA was verified by primer extension. ‡The position of the guide within the sRNA (D or D′ box associated) and the predicted site of methylation in the target RNA are indicated. "??," no strong prediction found. The D′ guide in sR10 is also predicted to methylate numerous other tRNAs including tRNA Pro-CGG and Pro-GGG at the homologous C position in the stem of the TψC arm. The D box guide in sR14 is predicted to methylate the wobble base in tRNA Gln. All tRNA sequences are from *S. solfataricus*. tRNA coordinates are in canonical numbering (e.g., anticodon is N34 to N36) (*15*). §The number of matches and mismatches in the complementarity between the guide and target sequence are indicated. Proposed complementarity to RNA targets is based on the following criteria: Watson-Crick base pair at position −5 (site of methylation); a minimum of eight base pairs with no more than two G:U base pairs; one mismatch permitted at positions other than −5. ‖"CDP," dNTP concentration-dependent primer extension pause observed at predicted site of methylation, indicating likely ribose 2′*O*-methyl; "No Pause," no pause was detected at either high or low dNTP concentrations; "Mod," known site of nucleotide modification of unknown type in *S. acidocaldarius* 16*S* rRNA (*13*). Guides without notation were not experimentally examined.

these sRNAs function as guides for ribose methylation, as in eukaryotes.

We next asked whether sRNAs are found in other Archaea. We retrained a previously developed eukaryotic snoRNA search program with the verified *S. acidocaldarius* sRNA genes (*3*, *16*) and used it to screen the available archaeal genome sequences. We first searched the genome sequence of the closely related archaeon *S. solfataricus* (*17*). The program identified dozens of sRNA candidates, each of which had the potential to target a modification to a particular position in rRNA. We designed primers complementary to the 20 top-scoring candidate sRNAs and performed primer extensions on *S. solfataricus* total RNA to detect stable RNAs. Ten candidates (*Sso* sR1 to sR10), all ranking within the top 13 candidates by score, generated products of the anticipated size, 2 to 6 nt upstream of the predicted C box. An alignment of the 10 verified *S. solfataricus* sRNAs, plus three high-scoring, untested sRNA predictions (sR11 to sR13), is available (*18*). Six predicted target ribose methylation sites were assayed with the dNTP concentration-dependent primer extension assay (Fig. 3), and four showed reverse transcription pauses characteristic of ribose methylation (*18*). Three additional target site predictions are known to be modified at the homologous position in *S. acidocaldarius* 16*S* rRNA (*13*, *18*).

Sulfolobus is a member of the Crenarchaeota, one of the two main phyla of Archaea; the other phylum, the Euryarchaeota, is evolutionarily distant. Complete genome sequences are available from archaeal species covering a wide range of genera, including both the Crenarchaea (*Aeropyrum pernix*) and the Euryarchaea (*Methanococcus jannaschii*, *Archaeoglobus fulgidus*, *Methanobacterium thermoautotrophicum*, *Pyrococcus horikoshii*, *Pyrococcus abyssi*, and *Pyrococcus furiosus*) (*5*, *17*). In searching these genomes for guide sRNAs, we found strong candidates in six of the seven species (*18*).

The searches of the *M. jannaschii* (*Mja*) and *A. fulgidus* (*Afu*) genomes gave eight and four strong sRNA hits, respectively; guide regions in most of these candidates exhibit complementarity to rRNA (*18*). The presence of all eight *Mja* sRNAs was confirmed by primer extension analysis on *M. jannaschii* total RNA (*18*). We attempted to verify seven of the ribose methylation sites predicted by the *Mja* sRNAs. Five sites showed concentration-dependent pauses indicative of ribose methylation, and the two other sites showed concentration-independent pauses, inconclusive for ribose methylation (*18*). An example pause site predicted by *Mja*-sR6 at position C2034 in 23*S* rRNA is shown (Fig. 3D). The D box guide region of *Mja*-sR8 predicts methylation of the anticodon wobble base for the intron-containing precursor of tRNA-Met. We did not test any tRNAs for ribose modifications, although the wobble base in tRNA-Met is known to be ribose-methylated within another hyperthermophilic crenarchaeon (*19*). The search of the *Aeropyrum pernix* (*Ape*) genome produced 23 candidate sRNAs (*18*). There were no strong sRNA hits in the genome of *M. thermoautotrophicum* (*20*).

The genomes of three *Pyrococcus* species have been sequenced: *P. horikoshii* (*Pho*), *P. furiosus* (*Pfu*), and *P. abyssi* (*Pab*) (*17*). These related sequences enabled us to infer support for sRNA predictions using comparative sequence analysis. From separate genome searches followed by comparative analysis, we identified 57 groups of homologous *Pyrococcus* sRNA genes (*21*). Forty-seven groups were found in all three species, eight were found in only two species, and two were unique to single species. Examples of two of these groups, sR3 and sR4, are illustrated (Fig. 4), and the complete set is available online, as are the alignment, annotation, and genomic distribution of the candidate sRNAs found in *P. horikoshii* (*18*).

We asked whether predicted rRNA methylation sites occurred at homologous rRNA positions in different archaeal genera. We view our site predictions with caution, as the sRNA complementarities are short and few have been experimentally tested. Nonetheless, on the basis of an rRNA multiple alignment (*22*), a total of 19 predicted methylation sites were conserved between two or more genera. Figure 4A shows 16*S* Um52, a confirmed modification in *Sulfolobus*, which we predict is guided by sR1 in *Sulfolobus* and by sR4 in *Pyrococcus*. However, *Sulfolobus* sR1 and *Pyrococcus* sR4 also have dissimilar D′ associated guide sequences that are predicted to target methylation to nonhomologous positions (16*S* Um33 in *S. solfataricus* and 16*S* Am361 in *Pyrococcus*). Figure 4B shows that the predicted guide sequences for a site in 23*S* rRNA (*Sac* U2692, *Ape* U2714, and *Pho* U2673) contain four separate nucleotide substitutions that are matched by compensatory substitutions in

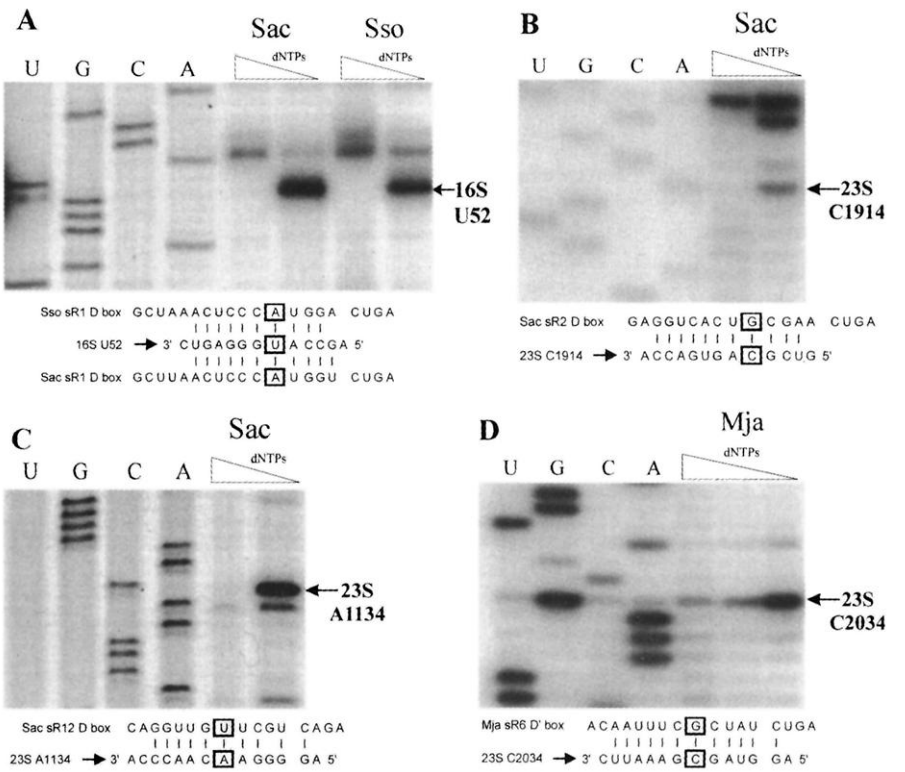

Fig. 3. Detection of 2′-*O*-ribose methylation sites in rRNA. Positions of ribose methylation in rRNA were detected with the dNTP concentration-dependent primer extension pause assay (*3*, *14*). Total RNA from *S. acidocaldarius* (**A** to **C**), *S. solfataricus* (A), or *M. jannaschii* (**D**) was used as template. The sequence ladders were generated from either DNA or RNA templates with the same primers used in the pause reactions. The position of pausing is indicated on the right along with the position of the methylated nucleotide; the pause characteristically occurs 1 nt upstream of the modification. When a sequence ladder is generated from DNA template, as in (A), the pause may occur 1 nt upstream of the modification or directly at the modification. The sequence of the sRNA guide and the complementary rRNA target are shown below each panel; the site of methylation in rRNA is in the base pair (boxed) positioned 5 nt upstream of the start of the D or D′ box.

23*S* rRNA, strong evidence that this sRNA/rRNA interaction is evolutionarily conserved. In nearly all cases, the intergenera sequence similarity between sRNAs that predict methylation at a homologous site is limited to the interacting guide region. In only one instance, we detected some end-to-end sequence similarity between two sRNAs from different archaeal genera: *Pho*-sR39 and *Mja*-sR6 (Fig. 4C). Moreover, the guide sequences can be either both in the same position (i.e., both D box associated) or in different positions (i.e., one D′ and the other D box associated; see Fig. 4B). Therefore, simple relationships of homologous sRNAs with homologous methylation sites are not obvious, and it remains uncertain whether sRNA guide sequences directing methylation to a homologous site are related to each other by common ancestry or by sequence convergence.

In general, all the archaeal sRNAs we identified are small, usually 50 to 60 nt in length, whereas human and yeast methylation guide snoRNAs average roughly 75 and 100 nt, respectively (*3*, *10*). A much larger proportion of archaeal sRNAs appear to have the ability to guide methylation from both D′ and D boxes as "double guides." On the basis of program predictions and comparative sequence analysis among *Pyrococcus* groups, we estimate that the majority of verified and putative archaeal sRNAs have two guide regions, whereas only 20% of human and yeast snoRNAs have been reported to be double guides (*3*, *10*). Often, the predicted target sites of double-guide sRNAs are within the same RNA molecule, and often, they are closely linked. For example, *Sso*-sR1 appears to direct methylation with D′ and D box guides to positions U33 and U52 in 16*3S* rRNA (Fig. 4A). This is in contrast to yeast snoRNA double guides, in which there is no apparent correlation between molecules targeted by the same snoRNA.

The number of sRNAs revealed by the search program seems to correlate with the optimum growth temperature of the organism: *Pyrococcus* species (95°C) have more than 50 putative sRNAs, whereas *M. thermoautotrophicum* (65°C) has no easily recognizable sRNAs. This may imply that a larger number of methylation modifications in rRNA might be required to fold or stabilize rRNA at high temperature (*4*) or that sRNAs are easier to recognize in hyperthermophiles because their gene features are more canonical.

In eukaryotes, snoRNAs do not act solely on rRNA. A number of cellular and viral RNAs transit through the nucleolus during maturation and at least one of these, the spliceosomal snRNA U6, is a substrate for snoRNA guide-directed methylation (*23*). Three cloned, verified *Sac* sRNAs (Table 2) do not appear to target any known stable RNAs (*18*), and several archaeal sRNAs exhibit complementarity to various tRNAs. Four of the sRNAs we identified (the *Pyrococcus* sR40 genes and *Afu* sR3) reside within the intron of the genes encoding tRNA-Trp. Our program detected these putative intronic sRNAs because they appeared to be capable of targeting methylation to sites within rRNA (*18*). However, Daniels and co-workers (*24*) have independently identified these sRNAs and suggest that the D′ and D box guides are targeting methylations to positions C34 and C39 within the intron-containing precursor tRNA. These observations suggest that both ribosomal and nonribosomal RNAs may be substrates for sRNA guide-directed methylation in Archaea.

Thus, it appears that an RNA-based guide mechanism for directing specific RNA 2′-*O*-ribose methylations was an established feature in the common ancestor of Archaea and Eukarya (*5*). In Bacteria, there is a low abundance of 2′-*O*-methylation and pseudouridylation in rRNA, and neither a fibrillarin homolog nor C/D box sRNAs have been described. Nonetheless, the existence of sRNA-directed modifications in bacterial stable RNAs remains a possibility.

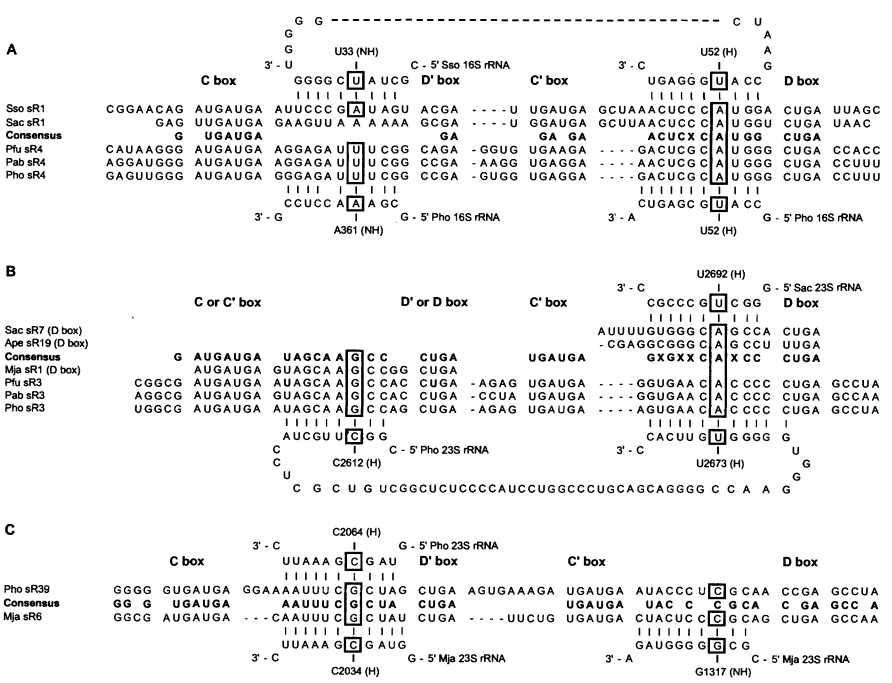

Fig. 4. Guide region sequence similarity in archaeal sRNA. Three sets of sRNAs that are predicted to direct 2′-*O*-ribose methylation to homologous sites in 16*S* or 23*S* rRNA are aligned either over their entire length or over their D′ or D box guide regions. The guide complementarities to regions of 16*S* or 23*S* rRNA are shown above and below the guide regions; predicted sites for methylation within the rRNAs that are homologous (H) or nonhomologous (NH) are boxed. An archaeal consensus is included with each set of aligned sequences in bold; X in the consensus indicates the presence of compensatory nucleotide substitutions in the sRNA guide and rRNA target region. For other details, see text.

References and Notes

1. E. S. Maxwell and M. J. Fournier, *Annu. Rev. Biochem.* **64**, 897 (1995); A. G. Balakin, L. Smith, M. J. Fournier, *Cell* **86**, 823 (1996); D. Tollervey, H. Lehtonen, M. Carmo-Fonseca, E. C. Hurt, *EMBO J.* **10**, 573 (1991); T. Gautier, T. Berges, D. Tollervey, E. C. Hurt, *Mol. Cell. Biol.* **17**, 7088 (1997).
2. J. Cavaille, M. Nicoloso, J.-P. Bachellerie, *Nature* **383**, 732 (1996); Z. Kiss-Laszlo, Y. Henry, J.-P. Bachellerie, M. Caizergues-Ferrer, T. Kiss, *Cell* **85**, 1077 (1996); C. Bousquet-Antonelli, Y. Henry, J.-P. Gélugne, M. Caizergues-Ferrer, T. Kiss, *EMBO J.* **16**, 4770 (1997); J. Ni, A. L. Tien, M. J. Fournier, *Cell* **89**, 565 (1997); P. Ganot, M. L. Bortolin, T. Kiss, *Cell* **89**, 799 (1997); B. Michot, N. Joseph, S. Mazan, J. P. Bachellerie, *Nucleic Acids Res.* **27**, 2271 (1999); K. T. Tycowski, C. M. Smith, M. D. Shu, J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14480 (1996); L. B. Weinstein and J. A. Steitz, *Curr. Opin. Cell Biol.* **11**, 378 (1999).
3. T. M. Lowe and S. R. Eddy, *Science* **283**, 1168 (1999).
4. K. R. Noon, E. Bruenger, J. A. McCloskey, *J. Bacteriol.* **180**, 2883 (1998).
5. C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576 (1990).
6. K. A. Amiri, *J. Bacteriol.* **176**, 2124 (1994); D. L. Lafontaine and D. Tollervey, *Trends Biochem. Sci.* **23**, 383 (1998).
7. A clone (*Sso* cosmid number 33) containing the aFIB and aNOP56 genes from *S. solfataricus* was provided by M. A. Ragan and C. W. Sensen. We used Southern hybridization to identify a 5-kb Xba I restriction fragment containing the corresponding genes from *S. acidocaldarius*; the fragment was cloned (pPD 1238), and its sequence was determined. The 16*S* rRNA sequences from *S. acidocaldarius* and *S. solfataricus* are 90% identical; the aFIB and aNOP56 predicted amino acid sequences are 76% and 65% identical, respectively. The *S. acidocaldarius* aFIB protein is 45% and 46% identical to the yeast and human proteins, and aNOP56 protein is 32% identical to the yeast and human proteins. Our choice of the name aNOP56 is arbitrary, because the aNOP56 sequence is also similar to eukaryotic NOP5/NOP58; the eukaryotic SIK1/NOP56 and NOP5/NOP58

proteins appear to be paralogs that arose by a gene duplication in Eukarya. The GenBank accession numbers for the aNOP56 and aFIB protein sequences from *S. acidocaldarius* are AF201092 and AF201093, respectively.

8. J. R. Chamberlain, Y. Lee, W. S. Lane, D. R. Engelke, *Genes Dev.* **12**, 1678 (1998).
9. P. Wu, J. S. Brockenbrough, A. C. Metcalfe, S. Chen, J. P. Aris, *J. Biol. Chem.* **273**, 16453 (1998). Briefly, the primer AO30 (5′-CTCGAGATCTGGATCCGGG-3′) was 5′ phosphorylated with T4 polynucleotide kinase and adenosine triphosphate (ATP) and blocked at the 3′ end with terminal deoxynucleotidyl transferase (Gibco BRL) and dideoxycytidine triphosphate (ddCTP). The modified oligo was then ligated to gel-purified sRNA for 16 hours at 4°C. The ligation products were reverse transcribed with Thermoscript RT (Gibco BRL) at 55°C for 30 min, with AO31 (5′-CCCGGATCCAGATCTCGAG-3′) as primer. The RNA template was hydrolyzed with ribonuclease H, and the cDNA strand was extended with deoxy-ATP with terminal deoxynucleotidyl transferase. The extended cDNA strand was used as template for PCR (95°C denaturation, 65°C hybridization, 72°C extension, 30 cycles) with AO31 and AO32 [5′-GCGAATTCTGCAG$(T)_{30}$-3′] as primers. The DNA products were cleaved with Pst I and Xho I, ligated between the Pst I and Xho I sites of plasmid pSP72, and transformed into *E. coli*. Plasmids were isolated and their sequences were determined.
10. Z. Kiss-Laszlo, Y. Henry, T. Kiss, *EMBO J.* **17**, 797 (1998).
11. Southern hybridizations confirmed the existence of single-copy genes for sR1, sR2, sR5, and sR13 in *S. acidocaldarius* genomic DNA. Genomic clones of the *S. acidocaldarius* sR1 and *S. solfataricus* sR1 were isolated and sequenced. In both cases, the sRNA genes overlap the 3′ end of the corresponding aspartate aminotransferase genes. The translation termination codons UAA for *S. acidocaldarius* and UAG for *S. solfataricus* fall within the D′ box guide regions in the two sRNAs (Fig. 4A).
12. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997). The BLAST searches against the nonredundant nucleotide database were performed at www.ncbi.nlm.nih.gov/BLAST/ on 15 June 1999.
13. G. J. Olsen *et al.*, *J. Mol. Evol.* **22**, 301 (1985); D. A. Stahl, K. R. Luehrsen, C. R. Woese, N. R. Pace, *Nucleic Acids Res.* **9**, 6129 (1981); E. Dams, P. Londei, P. Cammarano, A. Vandenberghe, R. De Wachter, *Nucleic Acids Res.* **11**, 4667 (1983); P. Durovic and P. P. Dennis, *Mol. Microbiol.* **13**, 229 (1994). As a result of strain misidentification, the 16*S* RNA sequence published in the first paper is now recognized to be from *S. acidocaldarius* and not *S. solfataricus* as originally indicated. The two originally published 5*S* sequences are identical and both are derived from *S. solfataricus*. The *S. acidocaldarius* 5*S* sequence has been published more recently along with a brief documentation of the strain misidentification problem [P. Durovic, U. Kutay, C. Schleper, P. P. Dennis, *J. Bacteriol.* **176**, 514 (1994)].
14. B. E. H. Maden, M. E. Corbett, P. A. Heeney, K. Pugh, P. M. Ajuh, *Biochimie* **77**, 22 (1995). This assay is not able to identify unambiguously all ribose methylations, presumably because of interference from other neighboring modifications (e.g., base methylations) and strong secondary structure features. Cases of false negatives have also been observed in which no pause was detected at sites of known ribose methylation. For unexplained reasons, we experienced a far lower primer extension success rate with archaeal rRNA than previously with yeast rRNA.
15. S. Steinberg, A. Misch, M. Sprinzl, *Nucleic Acids Res.* **21**, 3011 (1993).
16. The search algorithm and general snoRNA gene model remained as originally described (*3*), although yeast/human snoRNA training data were replaced by *S. acidocaldarius* training data. Alignments of box features (C, D′, and D) from *S. acidocaldarius* sRNAs sR1 to sR17 (Table 1) were used to create log odds weight matrices reflecting the frequency of each nucleotide at each position in each box feature. The lengths of the rRNA complementary region and the gaps between box features were scored with binned length distributions. Overall, training data for the nucleotide content of the box features did not change substantially between the yeast/human and archaeal models, but the distribution of lengths between features did vary; archaeal sRNAs appear to be much more compact than those in eukaryotes, and the rRNA complementary regions are shorter, commonly 8 to 11 nt long compared with 10 to 14 nt complementarities most often found in *S. cerevisiae* snoRNAs. We used final snoRNA scores to rank candidates within each genome and inspected hits manually by descending score. Hits that completely overlapped >600 nt open reading frames called by Glimmer 1.04 [S. Salzberg, A. Delcher, S. Kasif, O. White, *Nucleic Acids Res.* **26**, 544 (1998)] were discarded. No absolute score cutoff was used because each species' candidates matched the *S. acidocaldarius*–trained model to a different degree, and we had no estimate of the total number of genes to be found or the variability of sequence features for each different species.
17. Two-thirds of the *S. solfataricus* genome was completed at the time of our searches. The Web sites for accessing the archaeal genomes are as follows: *S. solfataricus*, niji.imb.nrc.ca/sulfolobus/; *M. jannaschii* and *A. fulgidus*, www.tigr.org/tdb/; *A. pernix*, www.mild.nite.go.jp/APEK1/; *M. thermoautotrophicum*, www.biosci.ohio-state.edu/~genomes/mthermo/; *P. abyssi*, www.genoscope.cns.fr/Pab/; *P. horikoshii*, www.bio.nite.go.jp/ot3db_index.html; and *P. furiosus*, www.genome.utah.edu/sequence.html. We gratefully acknowledge the *Sulfolobus solfataricus* P2 Genome Project (niji.imb.nrc.ca/sulfhome), Genoscope (www.genoscope.cns.fr), and the Utah Genome Center, Department of Human Genetics, University of Utah (www.genome.utah.edu) for access to unpublished genome sequence data.
18. Supplemental information is available at www.sciencemag.org/feature/data/1047007.shl. All newly identified archaeal snoRNAs and annotation are also available at rna.wustl.edu/snoRNAdb/.
19. R. Gupta, *J. Biol. Chem.* **259**, 9461 (1984).
20. Numerous candidate sRNAs that had one or more imperfect features were found in the genomes of the *S. solfataricus*, *M. jannaschii*, *A. fulgidus*, *A. pernix*, and *M. thermoautotrophicum*, but in the absence of identified ribose methylation sites or other confirmatory information, their authenticity remains uncertain. In particular, no strong candidates were identified within *M. thermoautotrophicum*, although we believe at least some of the search hits are legitimate (broader sRNA training data and/or the opportunity to test several dozen candidates experimentally are needed).
21. For each *Pyrococcus* sRNA homology group, the sequence identity for end-to-end alignments of interspecies members was 80 to 98%.
22. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994). For our alignments, we anchored the 5′ ends of the 16*S* rRNA with the conserved sequence GGUUGAUCCU; this block occupies positions 16 to 25 in all the 16*S* sequences in our alignment. The 23*S* sequences were anchored with the conserved sequence GGAUGGCUCG. In the respective 23*S* sequences in our alignment, this block occupies positions 21 to 30 in *Afu*, 22 to 31 in *Mja*, 33 to 42 in *Pho*, 20 to 29 in *Ape*, 28 to 37 in *Sso*, and 33 to 42 in *Sac*.
23. K. T. Tycowski, Z.-H. You, P. G. Graham, J. A. Steitz, *Mol. Cell* **2**, 629 (1998); T. Pederson, *Nucleic Acids Res.* **26**, 3871 (1998).
24. C. J. Daniels, personal communication; D. W. Armbruster and C. J. Daniels, *J. Biol. Chem.* **272**, 19758 (1997).
25. We thank J. W. Brown for generously providing *M. jannaschii* total RNA used in verifying sRNA predictions. Work in P.P.D.'s laboratory was supported by the Medical Research Council of Canada, grant MT6340; work in S.R.E.'s laboratory was supported by NIH National Human Genome Research Institute grant HG01363.

11 November 1999; accepted 24 February 2000

Transport of Peptide–MHC Class II Complexes in Developing Dendritic Cells

Shannon J. Turley,[1] Kayo Inaba,[2] Wendy S. Garrett,[1] Melanie Ebersold,[1] Julia Unternaehrer,[1] Ralph M. Steinman,[3] Ira Mellman[1]*

Major histocompatibility complex class II (MHC II) molecules capture peptides within the endocytic pathway to generate T cell receptor (TCR) ligands. Immature dendritic cells (DCs) sequester intact antigens in lysosomes, processing and converting antigens into peptide–MHC II complexes upon induction of DC maturation. The complexes then accumulate in distinctive, nonlysosomal MHC II$^+$ vesicles that appear to migrate to the cell surface. Although the vesicles exclude soluble lysosomal contents and antigen-processing machinery, many contain MHC I and B7 costimulatory molecules. After arrival at the cell surface, the MHC and costimulatory molecules remain clustered. Thus, transport of peptide–MHC II complexes by DCs not only accomplishes transfer from late endocytic compartments to the plasma membrane, but does so in a manner that selectively concentrates TCR ligands and costimulatory molecules for T cell contact.

A pivotal step in the initiation of T cell immunity is the presentation of antigenic peptides by MHC products expressed on DCs. In general, MHC II molecules bind peptides formed in endocytic organelles (*1*). In antigen-presenting cells (APCs) such as B lymphocytes, MHC II accumulates in late endosomal and lysosomal compartments (collectively termed MIICs) together with other components required for antigen processing. These include the invariant (Ii) chain that targets MHC II from the Golgi to