J Mol Evol (2003) 56:69–76
DOI: 10.1007/s00239-002-2381-6

JOURNAL OF MOLECULAR EVOLUTION

Common 5S rRNA Variants Are Likely To Be Accepted in Many Sequence Contexts

Zhengdong Zhang, Lisa M. D'Souza,* Youn-Hyung Lee, George E. Fox

Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA

Received: 12 March 2002 / Accepted: 3 August 2002

Abstract. Over evolutionary time RNA sequences which are successfully fixed in a population are selected from among those that satisfy the structural and chemical requirements imposed by the function of the RNA. These sequences together comprise the structure space of the RNA. In principle, a comprehensive understanding of RNA structure and function would make it possible to enumerate which specific RNA sequences belong to a particular structure space and which do not. We are using bacterial 5S rRNA as a model system to attempt to identify principles that can be used to predict which sequences do or do not belong to the 5S rRNA structure space. One promising idea is the very intuitive notion that frequently seen sequence changes in an aligned data set of naturally occurring 5S rRNAs would be widely accepted in many other 5S rRNA sequence contexts. To test this hypothesis, we first developed well-defined operational definitions for a *Vibrio* region of the 5S rRNA structure space and what is meant by a highly variable position. Fourteen sequence variants (10 point changes and 4 base-pair changes) were identified in this way, which, by the hypothesis, would be expected to incorporate successfully in any of the known sequences in the *Vibrio* region. All 14 of these changes were constructed and separately introduced into the *Vibrio proteolyticus* 5S rRNA sequence where they are not normally found. Each variant was evaluated for its ability to function as a valid 5S rRNA in an *E. coli* cellular context. It was found that 93% (13/14) of the variants tested are likely valid 5S rRNAs in this context. In addition, seven variants were constructed that, although present in the *Vibrio* region, did not meet the stringent criteria for a highly variable position. In this case, 86% (6/7) are likely valid. As a control we also examined seven variants that are seldom or never seen in the *Vibrio* region of 5S rRNA sequence space. In this case only two of seven were found to be potentially valid. The results demonstrate that changes that occur multiple times in a local region of RNA sequence space in fact usually will be accepted in any sequence context in that same local region.

Key words: Sequence space — RNA evolution — 5S rRNA — Site-directed mutagenesis

Introduction

RNA evolution can usefully be considered to be an exploration of a sequence space (Ninio 1983; Eigen et al. 1988). Typically there are many RNAs that differ in primary sequence which can perform the same biological function. Such functionally equivalent RNAs are frequently similar in three-dimensional structure because it is the positioning of specific chemical groups that is essential for proper function. Therefore, it is useful to consider that there is a subset of RNA sequences that comprise a "structure space" or "shape space" for any func-

**Present address:* Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030–3498, USA
Correspondence to: Dr. George E. Fox; *email*: fox@uh.edu

tional RNA. Over evolutionary time, the set of sequences that comprise such a structure space can be explored with little or no change in function. When this occurs, the individual changes are essentially neutral. On rare occasions, circumstances may exist that result in a significant change in biological function. Such an event would lead to the population of a new structure space with different sequence constraints imposed by the new function. If one fully understands how structure and function are interrelated, it should be possible to predict whether any specific sequence does or does not belong to a particular structure space. In practice, because many changes are not truly neutral, there will be fussiness associated with membership in real structure spaces. For example, certain sequences might work at certain temperatures, pH levels, etc., but not at others. Likewise, membership may be effected by the necessity to interact with another macromolecule and therefore dependent to some extent on the current sequence in that second structure space.

The ability to explicitly predict or determine whether or not particular sequences belong to a structure space would provide considerable insight to the evolutionary process. For example, by examining extant sequences it would be possible to determine the extent to which the space has been explored. In phylogenetic reconstructions it would be possible to determine the reasonableness of predicted nodal sequences and possibly to determine relative probabilities of alternative paths through the sequence space itself. The latter would be a powerful tool in evaluating alternative phylogenetic trees.

It is not currently feasible to determine explicitly for every sequence whether it does or does not belong to a particular stmcture space. Therefore, predictive principles are needed. A particularly promising source of such insight is atomic-resolution structures of sequences known to belong to a particular structure space. Structural information defines key hydrogen bonding interactions, which are essential for proper three-dimensional folding and therefore likely to be preserved in most sequences that belong to a particular structure space. By imposing the requirement that most, or all, of the known structure constraints be maintained, one obtains explicit predictions of whether or not particular primary sequences will belong to the structure space.

Another source of insight that is available for naturally occurring molecules is to examine the sequence variation seen in functionally equivalent sequences isolated from phylogenetically diverse organisms. By their very occurrence in Nature, such sequences are known to be valid members of the structure space of interest. Therefore, patterns of sequence variation reflect the types of change that can be tolerated in members of the structure space. When used in combination, sequence and structure comparisons may be very effective in defining practical rules for predicting inclusion/exclusion of particular sequences in a structure space. Experimental techniques such as *in vitro* evolution and directed mutagenesis can then be used to test the accuracy of the predictions. Using the bacterial 5S rRNA structure space as a model system, we are seeking to understand the extent to which patterns of sequence variation seen in comparative sequence sets can be used as reliable predictors.

An experimental system was previously described that can be used to determine whether or not a particular sequence is a potentially valid member of the 5S rRNA structure space (Hedenstierna et al. 1993). In this system, a test sequence is placed in a plasmid expression vector, which provides proper recognition sequences for transcription and processing in the *E. coli* cellular milieu. An initial assay is performed to determine the extent of accumulation of the test RNA in the total cellular RNA pool, 50S ribosomal subunits, 70S ribosomes, and polysomes. It is expected that if the test RNA accumulates in the 50S subunits, 70S ribosomes, and polysomes in large amounts, it would be very deleterious unless it were behaving as a functional 5S rRNA. The second assay, then, is to look at the effect of the test RNA on the cell growth rate relative to the effect of an identical plasmid carrying a wild-type 5S rRNA. A sequence is considered to be a potentially valid member of the 5S rRNA structure space if it accumulates in the ribosomal fractions without a significant effect on the growth rate.

Initially, this experimental system was used to test the likely validity of wild-type 5S rRNA sequences from 11 species of the genus *Vibrio* 5S rRNA in the *E. coli* context (Lee et al. 1993). These sequences were constructed by successive site-directed mutagenesis from a chemically synthesized *Vibrio proteolyticus* 5S rRNA gene that had been cloned in the expression plasmid. The sequences used were those that had been reported previously (MacDonell and Colwell 1985). It was found that these RNAs all accumulated to high levels without an obvious detrimental effect on the growth rate. In contrast, mutant 5S rRNAs were found that either failed to accumulate or simply were not stable in the *E. coli* cellular environment. Thus, the experimental system clearly distinguished known members of the 5S rRNA structure space from sequences that were not likely to be members. A second important finding from this initial study was that membership in the 5S rRNA structure space is apparently not strongly dependent on the host environment, as the wild-type sequences were examined in an *E. coli* cellular context rather than their native context. Subsequently, this experimental system was employed to examine all the intermediate sequences

on apparently equally parsimonious trajectories through sequence space between two pairs of 5S rRNAs. It was found that for some trajectories all the intermediate sequences belong to the structure space, whereas in other cases they do not (Lee et al. 1997). Since the intermediates on the latter trajectories would probably not be fixed in a population, it would seem that these trajectories would be far less likely paths.

When examining numbers of closely related 5S rRNA sequences, e.g., the *Vibrio* type, it is typically observed that certain positions, or pairs of positions (when known secondary structure base pairs are involved), frequently are changed at the same site relative to the consensus sequence. If one examines closely related sequences from a different region of the 5S rRNA structure space, e.g., *Bacillus*, similar variable positions exist but their location is different. From a structure/function perspective, it can be argued that single sequence positions that can be changed in a particular 5S rRNA without functional consequence are unconstrained and hence unimportant to 5S rRNA function. It is an obvious extension of this idea that the same positions will be unconstrained in other highly similar 5S rRNA sequences. Thus, it seems likely that for any local region of the structure space there is a set of positions and associated changes that are currently "in play" in that region of the structure space. It is reasonable, then, to expect that the variations seen at these positions are changes which could be readily incorporated into any other 5S rRNA sequence in the same neighborhood of the 5S rRNA structure space without deleterious effect. The critical issue is whether that neighborhood is of sufficient size that information obtained from changes to one sequence can be usefully applied to other sequences. Hence, we herein hypothesize that any change that is commonly seen in the *Vibrio* region of 5S rRNA sequence space would be accepted with a high probability in any other sequence in the *Vibrio* region of sequence space. To test this hypothesis, we examine the validity of the most variable positions in the *Vibrio* region of the 5S rRNA structure space in the *E. coli* cellular context. Since *E. coli* 5S rRNA is actually farther away (18 positions) from the *Vibrio* consensus sequence than any of the individual *Vibrio* sequences (all less than 13 positions), this is a strong test of the hypothesis.

Materials and Methods

Bacterial Strains and Plasmids. *Escherichia coli* strain JM109 was used in conjunction with bacteriophage M13. *E. coli* strain HB101 was used for ribosome preparation and in growth rate studies. The competing strain in growth rate studies was ML401. This strain is a lac^+ isogenic derivative of HB 101. Plasmid pKK5-1 (Brosius 1984; Szeberenyi and Apirion 1984) was kindly provided by Dr. H.F. Noller (University of California, Santa Cruz). The source of pure Vibrio proteolyticus 5S rRNA was ATCC strain 15338.

The Vibrio *Region of 5S rRNA Sequence Space and Identification of Common Sequence Changes.* 5S rRNA sequences are known for a number of Gram-negative bacteria that are more or less related to *V. proteolyticus*. In the work conducted here, we have operationally defined the *Vibrio* region of 5S rRNA structure space to consist of all known 5S rRNA sequences which include *V. proteolyticus* 5S rRNA, with the provison that no member of the set differs from any other member in more than 15 positions. Using this definition the vast majority of the 5S rRNAs in the cluster are from those organisms generally considered to belong to the genus *Vibrio*. Including *V. proteolyticus*, are 36 5S rRNA sequences that belong to the *Vibrio* cluster are available (MacDonell and Colwell 1984, 1985).

The 5S rRNA sequences belonging to the *Vibro* region of sequence space were aligned and the numbers of total base changes at every position relative to the *V. proteolyticus* sequence were tabulated as shown in Table 1. Any sequence variation that occurred at least five times was considered to be a potentially universally accepted change in the *Vibrio* region of the 5S rRNA structure space. The distribution of the changes encountered was next evaluated in conjunction with a maximum parsimony tree to assess whether or not the particular change had been fixed at least twice in the history of the *Vibrio* cluster. All 14 of the variable sites met this second criterion as well.

Control Variants. To understand the results obtained with variants at variable positions it is necessary to have a context in which they can be evaluated. We therefore identified and included seven variants that are not seen in any of the *Vibrio* sequences. These include the previously reported A10C (base at position 10, which is normally A changed to C) and A39C (Lee et al. 1993) and the new variants C37G, C43U, G56C, A58C, and A59G. Position A10 frequently changes to G in 5S rRNA and in two cases in the *Vibrio* cluster. However, the change to C is only occasionally seen among 5S rRNAs in general and never in the *Vibrio* cluster. Position C37 changes occasionally to A or U outside the *Vibrio* cluster, with one example in which the C37G change is seen. Position A39 is sometimes a U but there are no examples of C39. Changes of any type at positions 43 and 58 are extremely rare or nonexistent. Position G56 frequently changes to U outside the *Vibrio* cluster but changes to C are rare. Finally, the change A59G is common outside the *Vibrio* cluster.

Parsirnony Analysis. A parsimony algorithm (Felsenstein 1988), DNAPARS, in PHYLIP version 3.5 was used to construct phylogenetic trees for all *Vibrio* 5S rRNA sequences. The 5S rRNA sequence of *Bacillus subtilus* was used as an outgroup.

Construction of Sequence Variants. All variants were constructed by *in vitro* mutagenesis of the *V. proteolyticus* 5S rRNA sequence using the protocol provided with the Sculptor *in vitro* mutagenesis kit (Amersham Life Science Corp.). The mutant *V. proteolyticus* 5S rDNA was cloned into plasmid pKK5-1 and subsequently transformed into *E. coli* strain JM109.

RNA Preparation and Analysis. Total cellular RNA was isolated by low-pH phenol extraction (Wallace 1987). Ribosomes and 50S subunits were separated by sucrose gradient centrifugation (Godson and Sinsheimer 1967; Hedenstierna et al.. 1993). The RNA was analyzed by electrophoresis on a low-pH, 13% polyacrylamide gel as described in detail previously (Hedenstierna et al. 1993).

Table 1. All variable positions in *V. proteolyticus* 5S rRNA and their frequency of changes[a]

Position in *V. Proteolyticus* 5S rRNA	Number of total base changes	Number of specific base changes				
		A	C	G	U	Y[b]
U1	1		1			
C3	2				2	
C4	15				15	
C8	2				2	
C11	2				2	
C12	2			1	1	
C17	5				5	
G18	1	1				
A19	31		2		29	
U21	13	5		8		
A25	2				2	
C26	3	1			2	
U35	9	2	7			
U36	6		5			1
C37	1					1
C38	3				2	1
A39	2				2	
G41	13				13	
U52	28	28				
G61	3	3				
A62	14		6	1	7	
U64	30	27		3		
U65	11		11			
G67	1	1				
C70	9				9	
U87	6		6			
U88	11		11			
G106	8	8				
A107	1				1	
C108	1				1	
C114	2				2	
G117	1				1	
A119	4		1		3	

[a] Comparison is made within 5S rRNA sequences from 36 *Vibrio* strains. Any cell in which the number is zero is left blank. The 15 most variable positions are highlighted.
[b] Y = C or U.

RNA Quantitation. The proportion of *Vibrio* derived 5S rRNA relative to total 5S rRNA in either whole cells, 70S ribosomes, polysomes, or 50S ribosomal subunits was measured by hybridizing filter-bound RNA to a probe complementary to *V. proteolyticus* 5S rRNA, stripping the filter and hybridizing a second time with a probe complimentary to *E. coli* 5S rRNA. Oligonucleotide HV2 (5′-GTCCAAATCGCTATGGTCGC-3′) and HE2 (5′-GACCACCGCGCTACTGCCGC-3′), complementary to bases 7 to 26 in *V. proteolyticus* and *E. coli* 5S rRNA, respectively, were end-labeled with ^{32}P and used as probes. The density of *V. proteolyticus* (ATCC 15338) and *E. coli* 5S rRNA on the membrane was quantified by using a phosphorimager. A standard curve was constructed by plotting the ratios of known amounts of *V. proteolyticus* 5S rRNA to *E. coli* 5S rRNA against the ratios of the density of *V. Proteolyticus* 5S rRNA bands to that of *E. coli* 5S rRNA bands on the phosphorimager. Incorporation levels are typically reproducible to within 20%.

Growth Rate Assays. The growth rate of the strains carrying each of the variants was measured relative to a control strain carrying the *V. proteolyticus* 5S rRNA by monitoring the ratio of cell numbers in actively growing mixed cultures. The ratio of the two cell types was determined by spreading samples on McConkey agar plates, which discriminate between lac $^-$ and lac $^+$ strains by colony color (Dykhuizen and Hartl 1983; Hartl et al. 1983; Lenski 1988a, b). The growth rate difference coefficient, S, was determined as the slope of a line fitted to data points plotted as t against $\ln[N_p(t)/N_r(t)]$ (Chao and McBroom 1985). N_p and N_r represent the relative number of the two competing strains at time t (h) and S is a measure of the differential growth rate per unit time between the two strains. The strain designated N_p is favored, neutral, or disfavored relative to the strain designated by N_r as $S > 0$, $S = 0$, or $S < 0$. Two sets of experiments were done in each case such that the variant was in one strain and the control in the other. The values reported are thus an average of two independent determinations. The results presented here contrast the growth rates of each variant with cells carrying a plasmid that expresses *Vibrio proteolyticus* 5S rRNA. Cells carrying the *V. proteolyticus* plasmid have previously been shown to differ in growth rate by $-0.28\ h^{-1}$ from cells carrying the identical plasmid with an *E. coli* wild-type sequence (Lee et al. 1993).

Results

A total of 33 positions (Table 1) exhibits sequence variability among the 36 known sequences that

comprise the *Vibrio* 5S rRNA cluster. Of these variable positions, 15 differ in at least 5 cases relative to the consensus sequence. The bases at positions 19, 52, and 64 are actually largely conserved in the *Vibrio* cluster as a whole but differ specifically in *V. proteolyticus* and a small number of other species relative to the consensus sequence. In several cases, the variable positions are involved in Watson–Crick base pairs according to widely accepted secondary structure models. The positions that are involved in base pairs are 19/64, 21/62, and 70/106. In each case, there is a strong pattern of covariation between the correlated positions such that the potential for base pairing is preserved.

The 14 most common changes were introduced into the *V. proteolyticus* sequence as point mutations. Appropriate double mutants were also made for the positions involved in base pairs. In addition, seven less common sequence variants were constructed. In total, 27 sequence variants of *V. proteolyticus* were constructed and evaluated in the *E. coli* expression system. The results for five of these (A19U, U52A, U64A insertion U34.1, and the double variant A19U/U64A) were reported previously (Lee et al.. 1997). Each of these variants was assayed for accumulation in the total 5S rRNA pool as well as for incorporation into 50S ribosomal subunits, 70S ribosomes, and polysomes. In addition, the effect of the plasmid carrying each variant on the growth rate of the *E. coli* host was compared to that of the same plasmid carrying the wild-type *V. proteolyticus* gene.

Variant 5S rRNAs tested in this experimental system typically are distinguished primarily by the extent to which the variant RNA accumulates in the cell and is or is not incorporated into ribosomal fractions rather than growth rate changes. Most variants exhibit one of three characteristic phenotypes (Lee et al.. 1997): Type I, potentially valid variants in which the RNA accumulates in large amounts in both the total RNA and the 50S subunit fractions; Type II, unstable variants in which the amount of variant RNA in the total 5S rRNA fraction is greatly reduced; and Type III variants, in which the variant RNA accumulates in large amounts in the total 5S rRNA fraction but is not incorporated into 50S subunits to a significant extent. Using this experimental system it was possible to assess for each variant whether or not the expressed sequence was behaving as a 5S rRNA in the *E. coli* cellular context and hence could potentially be a member of the *Vibrio* 5S rRNA structure space. The results of these assays are detailed for each sequence variant in Table 2.

Five of the sequence variants produced an RNA that accumulated at levels below 10% of the total RNA. The A19U variant, which disrupts a base pair, was the least stable and was not detected at all in the ribosome fractions. Two other variants, U21A and C70U, produced an RNA that was virtually undetectable in the ribosome fraction. Both of these variants disrupt base pairs. The other two low accumulation variants, C17U and C70U (which disrupts a base pair), accumulated in the ribosome fractions at levels similar to those seen in the total 5S rRNA pool. One variant, U64A, was encountered which accumulated at high levels without significant incorporation into ribosomal fractions. This variant also disrupted a base pair. Two variants, U35A and U88C, accumulated in the ribosomal fractions, but at levels noticeably below those seen in the total RNA pool. The remaining variants, including all the double mutants, accumulated to significant levels in both the total pool and the ribosomal fractions. In terms of growth rate, none of the variant sequences were significantly deleterious. In fact, many were slightly advantageous relative to the *V. proteolyticus* construct itself.

The control changes (Table 2) consisted of seven variants that are not seen in the *Vibrio* region of sequence space. Four of these variants, C37G, A39C, A58C, and A59G accumulate to significant levels but are not incorporated into ribosomes at significant levels. Two changes, C43U and G56C, accumulate and incorporate at significant levels and therefore may be producing a usable 5S rRNA. Finally, A10C is a Type III variant that accumulates only to low levels. It was not detected in ribosomes to any significant extent and therefore is unlikely to be functional.

Discussion

There are two distinct types of variable positions in the *Vibrio* region of the 5S rRNA structure space. The first type is idiosyncratic changes. These are positions that have been uniquely changed relative to the consensus sequence in one particular organism or in a set of organisms that are specifically related to one another in a phylogenetic reconstruction. In either case, such positions do not appear to represent more than one evolutionary event. In contrast, a second category consists of highly variable positions, which likely represent multiple fixation events. The distinction is potentially important because if a change has been accepted only once, this may reflect some unique aspect of the sequence at the time the change occurred. In contrast, changes that have been successfully incorporated several times in a local region of the structure space would appear more likely to be compatible with essentially all the sequences in the local region.

The purpose of the work described here was to test the intuitively reasonable hypothesis that changes which have been accepted multiple independent times

Table 2. Characterization of point mutations in the *V. proteolyticus* 5S rRNA

Mutation	Growth rate difference (h^{-1})	% Incorporation				Phenotype[c]
		∑RNA[a]	70S	50 S	∑Poly[b]	
Controls						
A10C	0.047	5	tr[d]	tr	tr	III
C37G	−0.050	>50	tr	tr	tr	II
A39C	−0.028	>50	tr	tr	tr	II
C43U	−0.036	>50	37	38	37	I
G56C	−0.080	36	28	28	27	I
A58C	−0.038	34	tr	tr	tr	II
A59G	ne[d]	>50	tr	7	tr	II
Single and double mutations at the 15 variable positions						
C4U	0.024	21	21	29	23	I
C17U	0.062	6	7	7	8	III
U35C	0.041	45	47	47	46	I
U35A	0.015	>50	15	16	21	I
U36C	0.019	>50	32	29	26	I
G41U	ne	>50	>50	>50	>50	I
U52A	ne	>50	>50	>50	>50	I
U65C	0.025	>50	>50	>50	>50	I
U87C	ne	>50	46	46	48	I
U88C	0.025	>50	23	22	21	I
A19U;U64A	0.017	>50	48	43	41	I
U21G;A62C	−0.014	33	23	25	22	I
U21A;A62U	ne	>50	>50	>50	>50	I
C70U;G106A	0.026	23	21	22	24	I
Single mutations at the six covariant positions						
A19U	0.021	tr	nd[d]	nd	nd	III
U21A	−0.024	5	tr	tr	tr	III
A62U	0.069	22	21	20	19	I
U64A	0.047	23	tr	tr	tr	II
C70U	0.045	9	8	6	10	III
G106A	0.042	6	tr	tr	5	III

[a] Total cellular RNA (cytosolic RNA and ribosomal RNA).
[b] Polysomal pool (i.e., 2×, 3×, 4×).
[c] Major phenotypes: I, potentially valid; II, not incorporated; III, not stable.
[d] ne, no effect; nd, not detected; tr, trace (<5%).

in a group of related sequences would likely be accepted in any of the sequences in the group. To this end, the naturally occurring alternative versions of the 15 most variable positions in the *Vibrio* 5S rRNA sequence cluster were introduced into *V. proteolyticus* 5S rRNA and the resulting variant sequences were tested for their ability to function as 5S rRNAs in the *E. coli* cellular context. Six of the variable positions were involved in known secondary structure interactions and these positions covary in the *Vibrio* data set so as to produce alternative base pairs. These six positions were changed both individually and as double mutations in which the naturally occurring alternative pair was introduced. Because some of the positions had more than one common alternative nine individual changes, four base-pair variants, and six half-base-pair changes were constructed and assayed.

It was found that in essentially every case the individual variants produced a 5S rRNA that was apparently functional and stable in *E. coli*. The major exception was C17U, which produces an RNA with a greatly reduced stability and is therefore classified as not valid. The C17U RNA actually accumulates to levels in the ribosomes that are similar to those seen in the total RNA pool. Therefore the product RNA is likely functional to the extent it is able to accumulate. Two other variants, U35A and U88C, accumulate to a lesser extent in the ribosome fractions than they do in the total RNA pool, suggesting that they may be slightly impaired in their ability to participate in 50S subunit assembly. Both of these variants are considered to be potentially valid 5S rRNAs. All of the base-pair variants produce an RNA that appears to be functional. In contrast, all but one of the RNAs containing half-base-pair changes produce an RNA that is unstable and/or not incorporated into ribosomal fractions.

To provide a proper context in which to evaluate these results, it was necessary to consider what happens when one changes positions that are not variable in the *Vibrio* region of sequence space. One expects

that many, but not necessarily all, of such variants would be deleterious in some way. We therefore examined a sampling of such variants (Table 2). Of the seven changes, one is a Type III variant, in which the RNA product has lost its stability and no longer accumulates to a significant extent. Four other changes are Type II variants, which accumulate to high levels but are not incorporated into ribosomes. Our experience is that variants of this type usually have a negative effect on growth rate, presumably due to the synthesis of large quantities of an RNA that is not useful. This effect seldom exceeds $-0.050\ h^{-1}$ and, in the case of these four variants, averages $-0.029\ h^{-1}$. The remaining two, C43U and G56C, are both expressed and incorporated at significant levels. One of these, G56C, has a much greater negative impact on growth rate than is typically seen with the Type II variants and is therefore likely to be considerably less effective as a 5S rRNA than typical wild-type 5S rRNAs, which do not negatively impact growth (Lee et al.. 1993).

The results demonstrate that sequence variations, which arise independently in a number of related sequences, have a very high probability of being accepted in other sequences in the same general region of sequence space. Although we have only constructed the complete set of variants in 1 of the 26 known *Vibrio* sequence contexts, e.g., *V. proteolyticus*, we have done so under stringent conditions, in that the resulting RNAs were tested for validity in the *E. coli* cellular context. This is a strong test because *E. coli* does not belong to the *Vibrio* cluster, and at this increased phylogenetic distance there may be changes in other macromolecules that interact with the 5S rRNA that would preclude a variant from being functional that would work in the *Vibrio* cellular milieu. Those changes involving base pairs that covary in the known sequences typically must be introduced as a double change to produce a viable RNA. In contrast, single changes involving bases at basepaired positions (e.g., U65C, which converts a G/U wobblepair to a Watson–Crick pair) that do not exhibit covariation in the sequence set are likely to be accepted as single mutations in other sequence contexts as well. Finally, it should be noted that the positions with multiple independent fixation events likely represent only a subset of the positions that are actually readily changed in the *Vibrio* region of the structure space. Many of the known idiosyncratic changes as well as some changes that have not been seen at all might appear to be highly variable with multiple independent fixation events if a larger data set were available.

Although a complete high-resolution structure is not yet available for *E. coli* 5S rRNA itself, a considerable number of relevant high-resolution structure data are available. The recently determined structure of the 50S ribosomal subunit of *Haloarcula marismortui* (Ban et al. 2000) provides the atomic-resolution structure of an archaebacterial 5S rRNA as it exists in the ribosome. Positions 1–68 of this 5S rRNA are likely to be structurally homologous with *E. coli* 5S rRNA (Fox 1985). In addition, there is an atomic-resolution structure for a fragment of *E. coli* 5S rRNA covering positions 69–106 (Dallas et al. 1997). It is apparent from examination of this structural data that none of the variable positions are likely to be involved in base–base tertiary interactions in 5S rRNA. Although some isosteric base substitutions exist which allow for variation at only one site (e.g., an AC reverse Hoogsteen can be replaced with an AU reverse Hoogsteen pair), this is clearly rare, Thus, the ability readily and reliably to identify many of the freely changing positions in any local regions of a sequence space from sequence data alone makes it easier to find positions that are interdependent by simply narrowing the field of possibilities. When a change which is not generally allowed is accepted in one sequence, there likely will be a second change that has facilitated it. The responsible position is not likely to be one of the freely changing positions.

Acknowledgments. We would like to thank Ms. Qing Lu for constructing the phylogenetic tree for the 5S rRNAs belonging to the *Vibrio* cluster. This work was supported in part by NASA Grant NAG-5-4004 and its successor, NAG5-8140.

References

Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 289:905–920

Brosius J (1984) Toxicity of an overproduced foreign gene in *Escherichia coli* and its use in plasmid vectors for the selection of transcription terminators. Gene 27:161–172

Chao L, McBroom SM (1985) Evolution of transposable elements: An IS10 insertion increases fitness in *Escherichia coli*. Mol Biol Evol 2:359–369

Dallas A, Moore PB (1997) The loop E-loop D region of *Escherichia coli* 5S rRNA solution structure reveals an unusual loop that may be important for binding ribosomal proteins. Structure 5:1639–1653

Dykhuizen DE, Hartl DL (1983) Selection in chemostats. Microbiol Rev 47:150–168

Eigen M, Winkler-Oswalitsch R, Dress A (1988) Statistical geometry in sequence space: A method of quantitative comparative sequence analysis. Proc Natl Acad Sci USA 85:5913–5917

Felsenstein J (1988) Phylogenies from molecular sequences: Inferences and reliability. Annu Rev Genet 22:521–565

Fox GE (1985) The structure and evolution of archaebacterial ribosomal RNA. In: Woese CR, Wolfe RS (eds) The bacteria, Vol VIII. Academic Press, Orlando, FL, pp 257–310

Godson GN, Sinsheimer RL (1967) Use of Brij lysis as a general method to prepare polysomes from *Escherichia coli*. Biochim Biophys Acta 159:489–495

Hartl DL, Dykhuizen DE, Miller RD, Green L, de Framond J (1983) Transposable element IS50 improves growth rate of *E. coli* cells without transposition. Cell 188:503–510

Hedenstierna KOF, Lee Y-H, Yang Y, Fox GE (1993) A prototype stable RNA Identification cassette for monitoring plasmids of genetically engineered microorganisms. Sys Appl Microbiol 16:280–286

Lee Y-H, D'Souza LM, Fox GE (1993) Experimental investigation of an RNA sequence space. Orig Life Evol Biosphere 23:365–372

Lee Y-H, D'Souza LM, Fox GE (1997) Equally parsimonious pathways through an RNA sequence space are not equally likely. J Mol Evol 45:278–284

Lenski RE (1988a) Experimental study of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. Evolution 42:425–432

Lenski RE (1988b) Experimental study of pleiotropy and epistasis in *Escherichia coli*. II. Compensation for maladaptive effects associated with resistant to virus T4. Evolution 42:433–440

MacDonell MT, Colwell RR (1984) Identical 5S rRNA nucleotide sequence of *Vibrio cholerae* strains representing temporal, geographical, and ecological diversity. Appl Environ Microbiol 48:119–121

MacDonell MT, Colwell RR (1985) Phylogeny of the *Vibrionaceae* and recommendation for two new genera, *Listonella* and *Shewanella*. Syst Appl Microbiol 6:171–182

Ninio J (1983) Molecular approach to evolution. Princeton University, Princeton, NJ

Szeberenyi J, Aprion D (1984) Synthesis and processing of 5S rRNA from an rrnB minigene in a plasmid. Biochim Biophy Acta 783:15–25

Wallace DM (1987) Large- and small-scale phenol extraction. Methods Enzymol 152:33–41