Phylogenetically preserved inter-rRNA base pairs: involvement in ribosomal subunit association

T.A.Thanaraj*

Centre for Cellular and Molecular Biology, Hyderabad, India 500 007

Received May 31, 1994; Revised and Accepted August 15, 1994

ABSTRACT

Intermolecular complementary base pairs, that can be formed between the bases from single-stranded and weak stem regions on 16S rRNA and those on 23S rRNA, were located and checked for preservation in a variety of species covering the complete phylogenetic spectrum. Putative base pairs that exhibited two 'compensatory base pair changes' (a requisite as dictated by the approach of 'comparative sequence data analysis') were picked up. Potential base pairs were selected by assessing the frequency and taxonomy specificity of the occurrence of compensatory base pair changes against those of other types of base pair changes. The selected base pairs were classified as universal, non-mitochondrial, and prokaryote-specific. The positions of the proposed base pairs occur near the structurally and functionally important regions on rRNAs.

INTRODUCTION

The association of the 30S initiation complex — comprising the 30S ribosomal subunit loaded with initiator tRNA, mRNA, and initiation factors — with the 50S ribosomal subunit, to form the 70S initiation complex is a crucial step in protein synthesis. Topologically distinct regions, such as the 'decoding region' (on 30S subunit), the 'peptidyl transferase center', and the 'GTPase center' (on 50S subunit) are brought into functional proximity to form the 'translation' and the 'exit' domains on the ribosome. Moazed and Noller (1) have suggested subunit-subunit interaction as one of the molecular bases for tRNA movement on the ribosome; relative movement of the subunits is implicit in their model for the elongation cycle.

It appears that the primary interactions in the formation of the ribosome are those between the 16S and the 23S rRNA. Experiments based on chemical modification studies (2–5), antibiotic interactions (6), and DNA hybridization as well as mutagenic studies (7,8) have indicated the involvement of specific regions or bases from rRNAs in subunit association. It had been suggested that complementary sequences on 16S and 23S rRNAs might form base pairs during subunit association (2,3).

Under specific conditions, free 16S and 23S rRNAs from *E. coli* may form a binary complex that carries out a variety of steps in translation (9). Though the activity reported was of low order, it was suggested that RNA components of the ribosome can form a functional complex and perform most, if not all, ribosomal functions. Since only a few ribosomal proteins are involved in the above experiment, the stability of the complex may mainly be brought about by interactions between rRNA molecules. As a consequence, subunit association may be thought of as a multivalent process involving a large number of hydrogen bonds between specific regions on 16S and 23S rRNAs.

Not much success has been achieved, so far, in identifying the pairs of interacting sites on 16S and 23S rRNAs. The few sites as well as the few bases, that have been reported (7–8,10) are either from 30S or 50S subunit. In this report, using the approach of comparative sequence data analysis, an attempt has been made to identify intermolecular complementary bases on 16S and 23S rRNAs.

Comparative sequence data analysis has been successfully used for arriving at the structures of rRNA (11–14), tRNA (15), snRNA (16–18), and the RNA component of RNase P (19). The approach has also been found useful in identifying the intra-rRNA tertiary contacts (20–24). By this approach (25), a base pair is said to exist in the structure of an RNA molecule only if it undergoes a minimum of two independent compensatory base pair changes (CBC). This criterion was used in our search for inter-rRNA base pairs. The analysis allowed us to pick up the potential inter-rRNA base pairs that might be involved in subunit-subunit interactions.

MATERIALS AND METHODS

Ribosomal RNA molecules fold onto phylogenetically preserved structures (13–14). The bases in the single-stranded regions are found to be highly conserved, those in the stem regions show variability. The rRNA compositional differences are reflected in the bases from the stem regions. Many ribosomal functions have been found to be associated with the single-stranded regions (10,26,27). Ribosomal subunit association, at the first instance, might involve only the single-stranded regions. At subsequent

*Present address: c/o Professor Patrick Argos, European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, Meyerhofstrasse 1, Germany

Table 1. Summary chart for the proposed inter-rRNA base pairs

Positions # and the predominant base pair	Base pair changes and the species [@]			
	CBCs	NWCs	MMs	del
A. Universal base pairs				
A1. 15, 1965 (G-C)	U-A(17) A-U(12;18; 30-33)	G.U(1,3;6-11; 13-14;15) G.A(2;16)	-----	----
A2. 367, 529 (U-A)	G-C(57) C-G(46,52 54,56)	-----	C*A(18,53) C*U(55)	(30-33)
A3. 779, 805 (C-G)	A-U(17-18) U-A(30-31, 33)	U.G(15-16; 26-27)	C*A(32) C*U(73)	-----
A4. 813, 1342 (U-A)	C-G(44) A-U(22)	-----	-----	(17-18; 30-33)
A5. 1359, 1071 (C-G)	A-U(17;26-27) U-A(18;21;24; 31,33)	U.G(1-11; 14;15; 23;28-29) A.G(13,16)	A*A(30,32)	-----
A6. 1397,1091 (C-G)	A-U(18;22) U-A(30-32)	U.G(12) A.G(65)	A*A(21;24) U*U(33) C*U(49)	-----
A7. 1504, 1100 (G-C)	A-U(17;21) U-A(30-31,33)	G.U(24) G.A(22-23;49)	A*A(18) U*U(32)	(15-16)
A8. 1504, 1837 (G-C)	A-U(18;21) U-A(30-33)	G.U(24;67-68)	A*C(17)	(15-16)
A9. 893, 1099 (C-G)	U-A(17;21; 34-42) A-U(18)	U.G(22;28-29; 45;64-67; 78) A.G(15-16;27)	A*A(23) C*A(43-44)	(30-33)
Positions # and the predominant base pair	Base pair changes and the species [@]			
	CBCs	NWCs	MMs	Mitochondrial* species that show base pairs
B. Non-mitochondrial base pairs				
B1. 121, 88 (C-G)	A-U(34-45) U-A(67-68)	U.G(71-74)	G*G(75)	C-G(19-20) A-U(24)
B2. 265, 915 (G-C)	A-U(45;59-63) U-A(66)	U.G(36) G.U(67-68) G.A(64-65)	-----	G-C(19-20; 26-27) A.G(21;24)
B3. 321, 1940 (A-U)	C-G(49;34-35; 37-44) U-A(46,48,51, 53)	U.G(36) A.G(55-57) G.A(47)	C*A(54) G*G(58)	A-U(19-20; 21;25) U-A(22) G-C(26-27)
B4. 652, 2740 (U-A)	A-U(34-42) G-C(44)	-----	G*G(43) A*A(45)	U-A(19-20; 21-25;26-27)
B5. 991, 2058 (U-A)	C-G(34-37, 39-41,44-45; 46-50,52-53, 56,58) A-U(55)	A.G(42-43; 51,54,57)	del(38)	U.G(1-12,14- 17;21-23; 28-29) U-A(13;24; 25-27)
B6. 1322, 500 (C-G)	A-U(34-37,39, 45) G-C(40)	A.G(41,44) G.U(38,42-43)	-----	C-G(19-20)
B7. 1322, 954 (C-G)	A-U(34-37,39, 41,44-45) G-C(43)	G.U(38,40,42) U.G(58)	-----	C-G(19-20; 21;22;28-29) U-A(1) U.G(24,26-27) A.G(2,13,17, 25)
B8. 1381, 2058 (U-A)	C-G(34-45; 46-54,56-58) A-U(19-20)	-----	C*U(55; 72)	C-G(1-16,21, 23) A.G(22,28-29) A-U(19-20)
C. Prokaryote-specific base pair				
C1. 559, 2779 (A-U)	U-A(46-54) G-C(56,58)	G.A(55,57)	A*C(69, 70)	-----

REMARKS

A. Universal base pairs

A1 — The NWCs are mostly confined to specific subclasses in animalia mitochondria.

A2 — Though the MMs and CBCs occur in archaeobacteria, the MMs are due to failures in the changes from the predominant (U–A) to archaeobacteria-specific base pair (C–G). The G–C CBC is independent of the mismatches.

A5 — The mismatches in 30 and 32 can be corrected, if the corresponding sequences on 23S rRNA are aligned as UUA AAAA A/U UA.

30	AUAUGU	AUG	-----	UUA AAAA UU(A)	A–U
31	AUUAGC	AUG	-----	same	U–A
32	AAAAGC(A)UA		-----	same	A–U
33	AUUAUU(A)AAU		-----	same	U–A

A7 — MMs in 18 and 32 can be corrected to A–U and U–A with the following rearrangements in the 23S rRNA structure.

18	AAUUUAAU		AA UUU	AAU		32	UA UUAUAA		U		UA UUA	UAA	
	UA		UA		A		UU		U		UU		U
	AUUU		AUUU		A		UUAU		A		UAAU		A

The rearranged intra-rRNA structure agrees with that from 2 (Rat mitochondria). The six-member loop occurs as a five-member loop in rat, as well.

A9 — The deletions and MMs occur in specific subclasses. The CBCs are coming from different subclasses from eukaryote and also from mitochondria.

B. Non-mitochondrial base pairs

B1 — Except in a few species from prokaryote and eukaryote, the region on 23S rRNA is of length 7 bases. The alignment is as follows:

43	UACG(C)	AGA	71	GGCA(A)	ACA	72	GGCA(A)ACA
82	GUCA(A)	CCG	83	GAUA(A)	CCG	84	GUCA(A)CCA
34	UGCA	GGA	35	UGCA	GGA		

B2 — The position on 23S rRNA is involved in intra-rRNA structure. In those species, where the intra-rRNA base pairing is bad, the inter-rRNA base pairing is good and vice versa.

B3 — In archaeobacteria, the MMs occur in specific subclasses (archaeoglobales and crenarchaeota) that are different the subclasses that show CBCs.

B4 — Though MMs and CBCs occur in eukaryotic protist class, they are from 2 separate subclasses.

B5 — The del16S in 38 can be corrected to C–G by taking the 3'–end base from the helix 34 on 16S rRNA.

B8 — This is included in the non-mitochondrial list, because the protista and subclass in fungi mitochondria show MMs/deletions. The CBC that occurs in plant mitochondria is significant because generally the plant mitochondria show conserved base pairs as those of eubacteria and plastid.

C1 — In both mitochondria and eukaryote, the regions show deletions on 23S rRNA. The 2 MMs in eubacteria are confined to the subclass of proteobacter alpha. Abbreviations: NWC — Non-(Watson–Crick) base pairs; del — deletion on one or both rRNAs. MM — mismatch; CBC — compensatory base pair change. MMs are shown by * (as in G*G); CBCs in bold.

*The positions (on 16S rRNA, on 23S rRNA) are as numbered in *E.coli*. The base pair in *E.coli* is the same as the predominant base pair, except for the site B1 which shows U.G.

*All other species from mitochondria show deletions on one or both rRNAs.

~Mitochondria and eukaryote show deletions on one or both rRNAs.

@The numbers correspond to the species given below. The species with numbers in bold were used in the first stage of analysis and the others were used in the second stage of the analysis. Semicolon (;) separates subclasses. Comma (,) separates the species within a subclass.

I. MITOCHONDRIA (1–33) — **Animalia** — **Mammalia** — **Homo sapiens** (1), *R. norvegicus* (2), *Bos taurus* (3), *Mus musculus* (4), *Antilocapra americana* (5), *Capra hircus* (6), *Cervus unicolor* (7), *Hydropotes intermis* (8), *Muntiacus reevesi* (9), *Odocoileus virginianus* (10), *Tragulus napu* (11); **Aves** — *Gallus gallus* (12); **Amphibia** — *Xenopus laevis* (13), *Rana catesbeiana* (14); **Echinoidea** — *Paracentrotus lividus* (15), *Strongylocentrotus purpuratus* (16); **Arthropoda** — *Drosophila yakuba* (17), *Apis mellifera* (18). **Plantae** — *Zea mays* (19), *Triticum aestivum* (20). **Fungi** — **Hemiascomycetes** — *Saccharomyces cerevisiae* (21); **Euscomycetes** — *Aspergillus nidulans* (22), *Podospora anserina* (23); **Uncertain affiliation** — *Saccharomyces pombe* (24). **Protista** — **Chlorophyta** — *Chlamydomonas reinhardtii* (25); **Ciliophora** — *Paramecium primaurelia* (26), *Paramecium tetraurelia* (27), *Tetrahymena pyriformis* 1 (28), *Tetrahymena pyriformis* 2 (29); **Kinetoplastidea** — *Trypanosoma brucei* (30), *Leishmania tarentolae* (31), *Critidia fasciculata* (32), *Leptomonas* sp. (33).

II. EUKARYOTE (34–45) — **Animalia** — **Mammalia** — **Homo sapiens** (34), *Rattus norvegicus* (35), *Mus musculus* (36); **Amphibia** — *Xenopus laevis* (37); **Arthropoda** — *Drosophila melanogaster* (38); **Nematoda** — *Caenorhabditis elegans* (39). **Plantae** — *Oryza sativa* (40). **Fungi** — *Saccharomyces cerevisiae* (41). **Protista** — **Ciliophora** — *Tetrahymena thermophila* (42); **Erylenophyta** — *Euglena gracilis* (43); **Kinetoplastidea** — *Critidia fasciculata* (44); **Eumycetozoa** — *Physarum polycephalum* (45).

III. ARCHAEBACTERIA (46–58) — **Euryarchaeota** — **Halobacteria** — *Halobacterium halobium* (46), *Halobacterium marismortui* (47), *Halococcus morrhua* (48); **Methanobacteriales** — *Methanobacterium thermoautotrophicum* (49); **Methanococcales** — *Methanococcus vannielii* (50); **Methanomicrobium group** — *Methanospirillum hungatei* (51); **Thermococcales** — *Thermococcus celer* (52); **Thermoplasma** — *Thermoplasma acidophilum* (53); **Archaeoglobales** — *Archaeoglobus fulgidus* (54). **Crenarchaeota** — *Desulfurococcus mobilis* (55), *Thermoproteus tenax* (56), *Sulfolobus solfataricus* (57), *Thermofilum pendens* (58).

IV. PLASTID (59–68) — **Plantae** — **Marchantiopsida** — *Marchantia polymorpha* (59); **Liliopsida** — *Zea mays* (60), *Oryza sativa* (61); **Magnoliopsida** — *Nicotiana tabacum* (62), *Pisum sativum* (63). **Protista** — **Chlorophyta** — *Chlamydomonas reinhardtii* (64), *Chlamydomonas eugametos* (65), *Chlorella ellipsoidea* (66); **Euglenophyta** — *Astasia longa* (67), *Euglena gracilis* (68).

V. EUBACTERIA (69–87) — **Proteobacteria** — **alpha** — *Rhodobacter capsulatus* (69), *Rhodobacter sphaeroides* (70); **beta** — *Pseudomonas cepacia* (71); **gamma** — *Pseudomonas aeruginosa* (72); **delta** — *Ruminobacter amylophilus* (73), *Escherichia coli* (74). **Green sulfur** — *Chlorobium limicola* (75). **Flavobacteria** — *Flavobacterium odoratum* (76), *Flexibacter flexilis* (77). **Spirochetes** — *Leptospira interrogans* (78). **Cyanobacteria** — *Anacystis nidulans* (79). **Gram positives** — **low G+C** — *Bacillus stearothermophilus* (80), *Bacillus subtilis* (81); **high G+C** — *Frankia* sp. (82), *Micrococcus luteus* (83), *Streptomyces ambifaciens* (84). **Radioresistant micrococci** — *Thermus thermophilus* (85). **Thermogales** — *Thermotoga maritima* (86). **Planetomyces** — *Pirallula marina* (87).

stages of translation, rRNA might undergo conformational changes leading to the involvement of stem regions, as well. In our analysis, only single-stranded regions were considered.

Escherichia coli was taken as the model species as it is known that the structures of rRNA molecules from other species fall onto the general pattern shown by *E.coli* (10). Single-stranded regions including bulges and interior loops, of all lengths, were identified on both 16S and 23S rRNA molecules from *E.coli*. In addition, regions from short stems (of length upto 3 base pairs)

as well as from stems that contain too many non-Watson–Crick (NWC) base pairs were considered. The number of such single-stranded regions on 16S and 23S rRNA was 120 and 209, respectively. The number of bases constituting such regions in 16S rRNA was 620 (40% of the length), and in 23S rRNA was 1386 (48% of the length).

A total number of 87 species was considered in the analyses (see footnote to Table 1). In stage 1, 44 species (shown by bold numbers in the footnotes) from all the 5 kingdoms were

considered. Homologous single-stranded regions, corresponding to those of *E. coli*, were picked up from the other 43 species, both in the cases of 16S and 23S rRNAs. The data bases used for this purpose were from Neefs *et al.* (28–29) and Gutell *et al.* (30). The regions were aligned with each other and the data thus generated was entered onto computer.

A program package was developed to check for the formation of a base pair between each of the 620 bases from 16S rRNA and each of the 1386 bases from 23S rRNA in the 44 species. Both the regular Watson–Crick (WC) base pairs and the non-Watson–Crick (NWC) base pairs (U.G, G.U, A.G, and G.A) were allowed. The total number of possible base pairs that were preserved in at least 40 out of 44 species was found to be 51,584. For each such base pair, the following information was stored: (a) Positions of the bases on 16S and 23S rRNA (as numbered in *E. coli*); (b) Score (total number of species in which the base pair is preserved); and (c) number of times the base pair undergoes changes of the types, CBC, mismatch (MM) and NWC.

RESULTS AND DISCUSSION

Putative universal base pairs

The stored data was scanned for entries which satisfied the additional criteria that the base pair undergoes at least two independent CBCs (involving only WC base pairs), and undergoes NWC base pair changes in not more than 4 species. The total number of such entries was found to be 113.

An examination of the positions (on 16S and 23S rRNA) of each of the 113 entries showed that there were many cases of one-to-many combinations (a position in a region on 16S rRNA showed complementarity to positions from more than one region on 23S rRNA and *vice versa*). The number of unique regions on 16S and 23S rRNAs were 38 and 51, respectively. For each of the 113 cases, the alignment was checked and some optimisation was done. The total number of valid entries reduced to 35 (which is in agreement with the above mentioned number of unique regions on 16S rRNA). This reduction was due to (a) CBCs were observed only in mitochondria in most of the 113 cases; (b) since alignment of mitochondrial sequences was difficult, due to the occurrence of insertions/deletions and unstructuredness at a large number of positions, wrong regions had been picked up as homologous regions (we were intentionally liberal in the beginning); and (c) on close scrutiny and examination, such sequences were corrected. The corrections were incorporated in the data and the search was repeated. Only the above mentioned 35 cases emerged again.

Putative non-mitochondrial base pairs

Prompted by the observation that mitochondrial rRNA sequences show deletions at a number of positions, it was decided to search for base pairs that were specific to non-mitochondrial species. The initial data base contained 34 such species. The program package was modified to ignore mitochondria and to select the base pairs that satisfied the criteria that score ≥ 31 ; NWC ≤ 4 ; and CBC (involving only WC base pair) ≥ 2 . Fourteen such base pairs could be picked up. Each of these cases was also checked in mitochondria. If in mitochondria the regions did not show deletions in both the rRNAs, the base pair was expected to be maintained. If this was not the case, then the base pair was rejected. Such a scrutiny led to the rejection of 7 cases. The

remaining 7 cases were considered for further analysis. Such base pair positions (on 16S rRNA, 23S rRNA) are (652,2740), (321,1940), (121,88), (59,767), (1322,500), (1322,75), and (991,2058).

Putative prokaryote-specific base pairs

The initial data base contained 27 species from prokaryotes. The program package was modified to ignore mitochondria and eukaryote and to select base pairs that satisfied the criteria that score ≥ 23 ; NWC ≤ 4 ; and CBC (involving only WC base pair) ≥ 2 . Five cases were picked. Two of these 5 cases were rejected since they were found to be not preserved in eukaryote, though the corresponding regions were not deletions in either of the rRNAs. The remaining 3 cases are (265,915), (559,2779) and (110,748).

Significance of the occurrence of the putative base pairs

A similar search was done using non-eukaryotic species (37 in number). Base pairs that satisfied the criteria that score ≥ 33 ; NWC ≤ 4 ; and CBC (involving only WC base pair) ≥ 2 were located. It was found that 7 cases could be picked up. Further scrutiny revealed that 6 of these were already included in the 35 cases mentioned earlier. In the remaining one, it was found that the base pair was not maintained in eukaryotes though the corresponding sequences on rRNAs were not deletions. Hence, this case was ignored. This observation that no base pair could be picked up from the non-eukaryotic species suggests that the occurrence of non-mitochondrial and prokaryote-specific base pairs was genuine and significant.

Selection of potential base pairs

The total number of putative base pairs for further scrutiny amounted to 45 (35 universal, 7 non-mitochondrial and 3 prokaryote-specific). At this stage, another 43 species were added (making a total of 87 species). The putative base pairs were scrutinised for preservation in the newly added species. Summary charts showing the taxonomic-classification-wise preservation of each of the 45 base pairs in the 87 species were prepared.

The preservation of each base pair was examined by comparing the type and the number of events of the CBCs with those of the other of base pair changes (NWC, deletions and MM). The cases which behaved as below were ignored. (a) MMs and NWCs were in excess of CBCs in the same subclass (as well as in the class, in the case of MMs); (b) not enough support for a CBC was forthcoming from the limited data base used — e.g. if in a subclass having 4 species, two showed CBCs and the other two showed mismatches; (c) the NWCs and CBCs, that occurred in the same subclass were mostly of same origin — e.g. C–G to U–A and to U.G/G.A (rather than to G.U/A.G); (d) the alignment and the structuredness of the regions on rRNAs were doubtful not only in the subclass (in which the CBC occurred) but also in the class; (e) the deletions were not class or subclass specific; and (f) even few classes of a kingdom did not show good preservation.

The above scrutinisation reduced the number of putative base pairs to 17. These potential base pairs showed the following features. (a) the two CBCs came from 2 different kingdoms/classes or at least from 2 subclasses; (b) the set of species showing MMs and NWCs were from a subclass that was different from the one that showed CBCs; and (c) if a CBC is dominated by NWCs in a subclass, the same CBC occurred in another subclass

Figure 1. Locations of the proposed interactions on 16S rRNA (A), and on 23S rRNA (B–C). Domains are marked in Roman letters. A1–A9 are universal, B1–B8 are non-mitochondrial and C1 is prokaryote-specific. Solid lines with arrows indicate the relevant intra-rRNA tertiary contacts (20–24,27,31) and crosslinks (32,33) that are nearby the proposed interactions. The positions, marked by M, are occupied by modified nucleotides (10) in one or more of the five species (*E.coli*, *S.carlsbergensis*, *X.laevis*, *H.sapiens*, Rabbit). The functions (10) of the nearby regions are indicated.

(wherein the base pair is well preserved). The selected base pairs along with comments are given in Table 1. Irrespective of the original selection procedure, the base pairs were newly classified as universal, non-mitochondrial, and prokaryote-specific base pairs. Base pairs showing preservation in one or two classes of mitochondria but showing uniform deletions in the remaining classes of mitochondria were also classified as non-mitochondrial base pairs.

Is the selection procedure missing any potential base pairs?

At this stage, it was decided to check whether there are any base pairs (that are not showing CBCs in the species covered in the initial data base) that may show CBCs as the size of the data base is increased. The data set was scanned to pick up those base pairs with a score ≥ 40 and $NWC \leq 4$. From the set of base pairs thus picked up, those that showed 2 CBCs were ignored. Since the remaining base pairs was large in number, only those that showed one CBC were considered. Such base pairs (45 in number) were scrutinised for preservation in the 43 species (that were added in the 2nd stage of analysis). Only one out of these 45 cases showed the second CBC: the remaining 44 cases did not, despite the observation that a majority of these 44 cases showed the single CBC in a large number of species — very often covering one (or more) complete kingdom(s). This suggested that our search for potential base pairs showing 2 CBCs is extensive and is near complete. The above mentioned base pair that showed the 2nd CBC is shown in Table 1 (as site A9).

Significance of the occurrence of CBCs

All the proposed base pairs show 2 independent events of CBCs. However, it should be noted that the proposed base pairs do not show extensive CBCs (as is the case with intra-rRNA base pairs). This may be because that inter-rRNA base pairs are functional rather than structural entities, and that regions of functional importance are relegated to conserved single-stranded regions on rRNA (10,26,27).

The evidences for the proposed universal base pairs come largely from mitochondria. Though the sequence alignment is unreliable at a number of positions on mitochondrial rRNAs, the proposed interactions occur at regions that are structurally well preserved. Such regions do not contain any deletions/insertions. As far as the proposed non-mitochondrial and prokaryote-specific base pairs are concerned, the evidences come from a complete kingdom as well as from within a kingdom. For the sites B1 and B2, one of the two CBCs occur in complete kingdoms. To rule out the possibility that the occurrence of such a CBC is either due to phylogenetic diversity/affinity or out of chance, the following calculation was done. The number of positions on the regions from 16S rRNA at which all eubacterial species possess a conserved nucleotide (b1) and all eukaryotic species possess a conserved nucleotide (e1, $e1 < > b1$) was estimated, and found to be 37 (= N1). Similarly, the number of positions on regions from 23S rRNA at which all eubacterial species possess a conserved nucleotide (b2) and all eukaryotic species possess a conserved nucleotide (e2, $b2 < > e2$) was found to be 85 (= N2). Out of the ($N1 \times N2 = 3145$) possible intercomplementary base pairs, only 264 (= N3) turned out to be intercomplementary. Thus the fraction that show inter-kingdom CBC by chance is $N3/(N1 \times N2) = 0.084$. Such a low value suggests that the occurrence of the inter-kingdom CBC for the proposed

interactions is neither due to phylogenetic diversity/affinity nor due to chance. This is further substantiated by the value (0.21) obtained for the case $b1 = e1$; $b2 = e2$.

Structural and functional aspects of the proposed sites

The number of unique positions on 16S and 23S rRNAs, forming the proposed interactions, are 16 and 17, respectively. The interactions that share common positions on either 16S or 23S rRNA are [(1322, 500),(1322, 954)], [(1504, 1100),(1504, 1837)], and [(991, 2058),(1381, 2058)]. These may act as 'conformational switches' and bring about the structural rearrangements in the translating ribosome. Figure 1(A–C) show the locations of the proposed sites on the structures of the rRNAs. *Structural aspects.* It is observed that the positions of the interactions (A1,A6,A7[A8],A9,B5,B8) on 16S rRNA and those of the interactions (A2,B5[B8],C1) on 23S rRNA are located at/near (either by primary structure or by higher order structure) the 'domain junctions'. In addition, the positions of the interactions (A2,A3,A6,A7[A8],B1,B5,B8) and of the interactions (A4,A8,B1,B2,B5[B8],C1) on 16S and 23S rRNAs, respectively, occur at/near the bases that have been implicated in intra-rRNA tertiary contacts (20–24,27,31) and crosslinks (32,33). The site A1 is located near a 'pseudoknot' located at the 5'-end of 16S rRNA. These observations suggest that the quaternary structural elements (the proposed inter-rRNA base pairs) are organised over the higher order structural elements, such as 'domain junctions', 'tertiary contacts', 'pseudoknots'.

The switch [B5(991,2058),B8(1381,2058)] requires the regions 991 and 1381 to be nearer on 16S rRNA. It turns out that an intra-rRNA tertiary contact, 994–1380 (22), brings these two regions nearer on the 16S rRNA. This suggests that our results are compatible with those of other structural studies.

Functional aspects. As can be seen from the figures, the positions of the interactions (A2,A3,A5,A6,A7[A8],B1,B3,B5,B8) and of the interactions (A8,B1,B3,B5[B8],C1) on 16S and 23S rRNAs, respectively, occur near (either by primary structure or by higher order structure) positions that are occupied by modified nucleotides (10) in one or more of the species. Earlier studies (34,35) have indicated that a majority of the modified nucleotides on rRNAs are located in close proximity to functional regions. As can be further seen in the figures, a considerable number of the proposed interactions occur in functional regions (10) on rRNAs. They are (A1–A3,A6,A7[A8],A9,B5–B8) on 16S rRNA and (A3,A5–A7,A9,B1,B4,B5[B8],C1) on 23S rRNA. The above sites are largely located in 'decoding region' on 16S rRNA and in 'GTPase center' and 'peptidyl transferase center' on 23S rRNA.

Mismatch base pairs. The total number of mismatch base pair changes, taking into account all the proposed sites, is 24 (column 4 of Table 1). The percentage of such a value is 1.9 (= 24/1240). It is interesting to note that when we considered sets of 10 randomly selected stem regions from 23S rRNA, the value turned out to be 8–10 (data not shown). The distribution of the 24 MM base pair changes is: A*A(7), G*G(3); A*C(1), C*A(6); C*U(5), U*U(2). It may be recalled that two of the tertiary interactions in tRNA involve A*A and G*G base pairs, and protonated A and C have the potential to form a base pair with a geometry similar to that of G.U base pair (36).

CONCLUSION

The results show that through the approach of comparative sequence data analysis potential inter-rRNA base pairs could be identified. All the proposed base pairs show 2 independent events of CBCs. They are further classified as universal, non-mitochondrial and prokaryote-specific base pairs. An examination of the base pairs reveals the existence of a set of three conformational switches. The proposed base pairs occur in regions of structural and functional importance on rRNAs and they seem to be organised over the higher order structural elements on rRNAs. The set of proposed base pairs might serve as good starting points for mutagenic studies and further help in studying the overall structural organisation of rRNAs in the ribosome.

ACKNOWLEDGEMENTS

The kind help of Dr R.De Wachter and Dr M.W.Gray in providing the data bases is gratefully acknowledged. Distributed Informatics Centre at CCMB is thanked for providing computer facilities. The author likes to thank Prof. D.Balasubramanian for his continued support and Dr M.W.Pandit for his encouragement during the course of work. The suggestions and comments of the referees are gratefully acknowledged.

REFERENCES

- Moazed,D. and Noller,H.F. (1989) *Nature*, 342, 142–148.
- Chapman,N.M. and Noller,H.F. (1977) *J. Mol. Biol.*, 109, 131–149.
- Herr,W., Chapman,N.M. and Noller,H.F. (1979) *J. Mol. Biol.*, 130, 433–449.
- Santer,M. and Shane,S. (1977) *J. Bacteriol.*, 130, 900–910.
- Meier,N. and Wagner,R. (1985) *Eur. J. Biochem.*, 146, 83–87.
- Moazed,D. and Noller,H.F. (1987) *Nature*, 327, 389–394.
- Hill,W.E., Tappich,W.E. and Tassanakajohn,A. (1986) In Hardesty,B. and Kramer,G. (ed.), *Structure, Function and Genetics of Ribosomes*. Springer Verlag, Berlin, pp233–252.
- Tappich,W.E., Goss,D.J. and Dahlberg,A.E. (1989) *Proc. Natl. Acad. Sci. USA*, 70, 2350–2352.
- Burma,D.P., Tewari,D.S. and Srivastava,A.K. (1985) *Arch. Biochem. Biophys.*, 239, 427–435.
- Raue,H.A., Musters,W., Rutgers,C.A., Riet,J.V. and Planta, R.J. (1990) In Hill,W.E., Dahlberg,A., Garrett,R.A., Moore,P.B., Schlessinger,D. and Warner,J.R. (ed.), *The Ribosome Structure, Function, and Evolution*. ASM, Washington,D.C., pp217–235.
- Fox,G.E. and Woese,C.R. (1975) *Nature*, 256, 505–507.
- Nishikawa,K. and Takemura,S. (1974) *FEBS Lett.*, 40, 100–109.
- Noller,H.F. and Woese,C.R. (1981) *Science*, 212, 403–411.
- Noller,H.F., Kop,J.A., Wheaton,V., Brosius,J., Gutell,R.R., Kopylov,A.M., Dohme,F. and Herr,W. (1981) *Nucleic Acids Res.*, 9, 6167–6189.
- Holley,R.W., Appgar,J., Everett,G.A., Madison,J.T., Marquise, M., Merrill,S.H., Penswick,J.R. and Zamir,A. (1965) *Science*, 147, 1462–1465.
- Mount,S.M. and Steitz,J.A. (1981) *Nucleic Acids Res.*, 9, 6351–6368.
- Brow,D.A. and Guthrie,C. (1988) *Nature*, 334, 213–218.
- Zucker-Aprison,E., Thomas,J.D., Blumenthal,T. (1988) *Nucleic Acids Res.*, 16, 7188–7188.
- James,B.D., Olsen,G.J., Liu,J. and Pace,N.R. (1988) *Cell*, 52, 19–26.
- Woese,C.R. and Gutell,R.R. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 3119–3122.
- Gutell,R.R., Noller,H.F. and Woese,C.R. (1986) *EMBO J.*, 5, 1111–1113.
- Haselman,T., Camp,D.G. and Fox,G.E. (1989) *Nucleic Acids Res.*, 17, 2215–2221.
- Leffers,H., Kjems,J., Ostergaard,L., Larsen,N. and Garreth, R. (1987) *J. Mol. Biol.*, 195, 43–61.
- Hopfl,P., Ludwig,W., Schleifer,K.H. and Larsen,N. (1989) *Eur. J. Biochem.*, 185, 355–364.
- Haselman,T., Gutell,R.R., Jurka,J. and Fox,G.E. (1989) *J. Biomol. Str. Dyn.*, 7, 181–186.
- Dahlberg,A.E. (1989) *Cell*, 57, 525–529.
- Gutell,R.R., Weiser,B., Woese,C.R. and Noller,H.F. (1985) *Prog. Nucl. Acid Mol. Biol.*, 32, 155–216.
- Neefs,J.M., Van de Peer,Y., De Rijk,P., Chapelle,S. and De Wachter,R. (1993) *Nucleic Acids Res.*, 21, 3025–3049.
- Neefs,J.M., Van de Peer,Y., Hendriks,L. and De Wachter,R. (1990) *Nucleic Acids Res.*, 18 (Supplement), r2237–r2317.
- Gutell,R.R., Schnare,M.N. and Gray,M.W. (1992) *Nucleic Acids Res.*, 20 (Supplement), r2095–r2109.
- Woese,C.R., Gutell,R.R., Gupta,R. and Noller,H.F. (1983) *Microbiol. Rev.*, 47, 621–669.
- Brimacombe,R., Atmadja,J., Kyriatsoulis,A. and Stiege,W. (1986) In Hardesty,B. and Kramer,G. (ed.), *Structure, Function and Genetics of Ribosomes*. Springer Verlag, Berlin, pp184–202.
- Brimacombe,R., Atmadja,J., Stiege,W. and Schuler,D. (1988) *J. Mol. Biol.*, 199, 115–136.
- Brimacombe,R., Mitchell,P., Osswald,M., Stade,K. and Bochkariov,D. (1993) *FASEB J.*, 7, 161–167.
- Lane,B.G., Ofengand,J. and Gray,M.W. (1992) *FEBS Lett.*, 302, 1–4.
- Wyatt,J.R., Puglisi,J.D. and Tinoco,Jr.I. (1989) *Bio Essays*, 11, 100–106.