
Structure and function of the nontranscribed spacer regions of yeast rDNA

K.G.Skryabin¹, M.A.Eldarov¹, V.L.Larionov¹, A.A.Bayev¹, J.Klootwijk^{2*}, V.C.H.F.de Regt², G.M.Veldman², R.J.Planta², O.I.Georgiev³ and A.A.Hadjilov³

¹Inst. Molecular Biology, Academy of Sciences of the USSR, Vavilovstr. 32, Moscow B-312, USSR, ²Biochemisch Laboratorium, Vrije Universiteit, de Boelelaan 1083, 1081 HV Amsterdam, The Netherlands, and ³Inst. Molecular Biology, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria

Received 22 December 1983; Revised and Accepted 24 February 1984

ABSTRACT

The sequences of the nontranscribed spacers (NTS) of cloned ribosomal DNA (rDNA) units from both Saccharomyces cerevisiae and Saccharomyces carlsbergensis were determined. The NTS sequences of both species were found to be 93% homologous. The major disparities comprise different frequencies of reiteration of short tracts of six to sixteen basepairs. Most of these reiterations are found within the 1100 basepairs long NTS between the 3'-ends of 26S and 5S rRNA (NTS1). The NTS between the starts of 5S rRNA and 37S pre-rRNA (NTS2) comprises about 1250 basepairs. The first 800 basepairs of NTS NTS2 (adjacent to the 5S rRNA gene) are virtually identical in both strains whereas a variable region is present at about 250 basepairs upstream of the RNA polymerase A transcription start. In contrast to the situation in Drosophila and Xenopus no reiterations of the putative RNA polymerase A promoter are present within the yeast NTS. The strands of the yeast NTS reveal a remarkable bias of G and C-residues. Yeast rDNA was previously shown to contain a sequence capable of autonomous replication (ARS) (Szostak, J.W. and Wu, R (1979), Plasmid 2, 536-554). This ARS, which may correspond to a chromosomal origin of replication, was located on a fragment of 570 basepairs within NTS2.

INTRODUCTION

Yeast ribosomal RNA genes are present in more than 100 copies per haploid genome clustered in a tandem array on chromosome XII (1,2). Each rDNA unit comprises about 9000 bp and encodes two primary transcripts, one for 5S rRNA and one for 37S pre-rRNA (3), separated by two nontranscribed spacers both of which are roughly 1000 bp long (NTS1 and NTS2). Restriction analysis of yeast genomic DNA indicates that the transcription units as well as the spacers are homogeneous in length (4,5). This is in contrast with the length heterogeneity observed for the NTS of rDNA in Vertebrata and Arthropoda (6), which is caused by their internally repetitious nature. For Xenopus laevis and Drosophila melanogaster rDNA it has been established that the putative promoter region of the large transcription unit has been duplicated within the NTS (7-9). Transcription initiation at these reiterated sequences has been observed (10,11). In addition, Moss (10) noticed an enhancing effect of

the NTS of Xenopus rDNA on transcription initiation. For these reasons the NTS of rDNA has been implicated in the regulation of transcription of the rRNA operon. We have studied the NTS of yeast rDNA with regard to its primary structure and homogeneity. Earlier we have reported the sequences of part of an rDNA clone from both S. carlsbergensis and S. cerevisiae. These sequences covered the transcription units and the regions immediately adjacent (12-17). Part of the NTS sequence of another rDNA clone of S. cerevisiae was determined earlier by Valenzuela et al. (14). Comparison of this sequence with the available data derived from the other clones (12,13) reveals some heterogeneity in the region downstream of the 26S rRNA gene. Further sequence analysis of different clones may therefore also shed some light upon the mechanisms involved in the generation of structural variations in the NTS. Inspection of the NTS sequences of the rDNAs in the two yeast strains reveals some general features of spacers of tandemly repeated units. Only variations in the number of reiterations of short tracts of six to sixteen bp are observed. In contrast to the situation in Xenopus and Drosophila no longer periodicities were detected. Finally, we determined that a portion of NTS2 can cause replication if present on a plasmid in a recipient yeast cell. This confirms and extends previous reports (18,19) on the presence of an autonomously replicating sequence (ARS) within yeast rDNA.

MATERIALS AND METHODS

Plasmids

Plasmid pMY57 was used for the sequencing of the NTS of S. carlsbergensis (NCYC74) rDNA. It contains the HindIII-generated fragment B cloned in pBR322 (20). Fragments of S. cerevisiae (+ D4) rDNA were isolated from pY1rA3 (21) and combined with YIp1, a vector comprising the yeast chromosomal LEU2 gene inserted in the PstI site of pBR325; the LEU2-containing PstI-fragment was isolated from YEpl3 (22). The structure of these plasmids is shown in Fig. 1. The construction of pCR(A) and pCR(B+G) was described previously (19). Sau3A-generated fragments were inserted in the BamHI-site of YIp1; an AluI- and an AluI/BsuI-generated fragment were cloned in the HindIII-site of YIp1 using synthetic HindIII-linkers. The CEN3 segment was isolated as a BamHI/HindIII-fragment from the plasmid pYe(CEN3)41 constructed by Clarke and Carbon (23), and combined with YIp1, after digestion with HindIII plus BamHI.

Strains and media

Plasmids were propagated in E. coli strain HB101. Transformation of S. cerevisiae DC5 (leu2⁻) was carried out according to Tschumper and Carbon

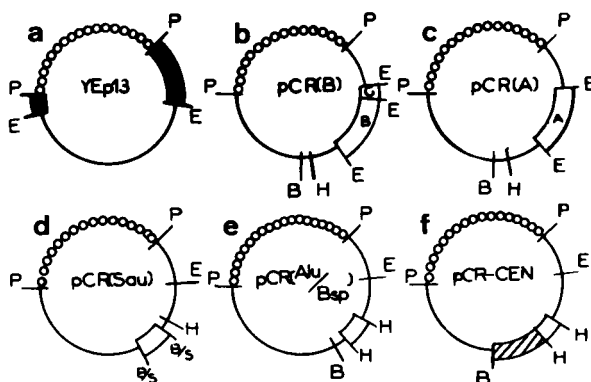


Fig. 1. Schematic representation of plasmids used for transformation. All plasmids contain a 4.1 kb PstI-L&E;U2 fragment (small circles) from YEpl13 (a). The thin line indicates either pBR322(a) or the pBR325 sequence (b-f). The solid bars represent sequences of the 2 μ yeast plasmid. (b) and (c) represent the plasmids containing the EcoRI-fragments A and B+G of yeast rDNA, respectively, inserted into the EcoRI-site of YIp1. (d) represents the general structure of Sau3A-generated fragments of the NTS (see Fig. 4) cloned into the BamHI-site of YIp1. (e) shows the AluI- and AluI plus BspI-generated fragments cloned into the HindIII-site of YIp1. (f) shows the plasmids containing the CEN3 segment (shaded box) derived from pYe(CEN3)41 (23) and the yeast rDNA fragments as shown in (e). EcoRI = E, PstI = P, BamHI = B, HindIII = H, Sau3AI = S.

(24) with the following modifications. Spheroplasts were obtained by glucanase treatment, washed and suspended in an equal volume of 1.2 M sorbitol, 10 mM CaCl₂ and 10 mM Tris-HCl (pH 7.4), 0.2 ml of this suspension was mixed with 10-20 μ g of plasmid DNA and kept for 15 min at room temperature. Then 2 ml of 30% (w/v) PEG-4000 was added and the incubation was continued for 45 min. After low-speed centrifugation the pellet was suspended in 0.5 ml of 1.2 M sorbitol and 67% (v/v) YPD (24) and incubated for 45 min at 30°C. Aliquots of 0.1 ml were mixed with 10 ml of regeneration agar (0.67% yeast nitrogen base without amino acids, 1 M sorbitol, 2% glucose, 3% agar) and plated onto selective agar medium. Colonies became visible after two days. To determine the mitotic stability of the yeast transformants, colonies of primary transformants were transferred to selective agar plates lacking leucine. After 24 h of incubation the cells were cloned on YPD-plates. Single colonies from these YPD-plates were replated on minimal plates and the percentage of leucine-prototrophic colonies was determined.

DNA, DNA fragments and enzymes

Plasmid DNA was isolated according to standard procedures (25,26). DNA fragments were purified either by sucrose density gradient centrifugation or

by electrophoresis in 4-5% polyacrylamide gels (15). Restriction enzymes and T4 DNA ligase were either purchased from Boehringer (Mannheim) and New England Biolabs or were kindly provided by dr. V. Yanulaitis (Vilnius).

DNA sequencing

DNA sequencing was performed according to Maxam and Gilbert (27). All details of the strategy are available (from the authors) upon request.

RESULTS AND DISCUSSION

Sequence of the NTS of yeast rDNAs

We sequenced a 2000 bp HindIII/SmaI fragment (see Fig. 4) of an S. carlsbergensis rDNA clone comprising the 5S rRNA gene and most of the NTS. The data complete the sequence of the NTS, parts of which had been determined earlier (13). The complete sequence of the NTS is presented in Fig. 2. We also sequenced 1700 bp of the NTS of an S. cerevisiae rDNA clone. About half of this sequence has already been published for another S. cerevisiae rDNA clone (14). All clones contain form I of yeast rDNA (21). Alignment of the two NTS sequences (Fig. 2) reveals a high degree (about 93%) of homology. The NTS of yeast rDNA is rather rich in A- and T-residues, viz. 62.5% for NTS1 and 59.4% for NTS2. These figures are close to the value of 58.2% inferred from the melting temperature of the EcoRI-generated fragment B of S. cerevisiae genomic rDNA (29). This fragment corresponds to the sequence from position 97 to 2554 in Fig. 2. It is very remarkable that the two strands of the NTS display a large bias of G- and C-residues, e.g. the sequence of S. carlsbergensis NTS2 depicted in Fig. 2 contains 11.0% C- and 29.6% G-residues; for NTS1 these figures are 11.8% and 25.7% respectively. This bias should cause a significant difference in buoyant density of both strands as has indeed been observed for the two strands of genomic rDNA of S. cerevisiae (0.015 g/cm³, Ref. 30). Inspection of the sequence in Fig. 2 reveals several homopolymeric tracts, in particular of A- and T-residues in NTS1. The previously documented tract of A-residues adjacent to the 3' end of the 5S rRNA gene (14) consists of as much as 29 consecutive A's in our S. cerevisiae rDNA clone. The next most prominent tracts in the NTS1 of S. cerevisiae rDNA comprise 16, 24 and 17 T's respectively. The lengths of these tracts are polymorphic for different yeast rDNA clones (see below). Homopolymeric tracts in NTS2 do not exceed a length of 10. Furthermore, the NTS contains numerous direct repeats of 4 to 16 bp in length. Smith (31) has proposed that duplications (and deletions) easily evolve within a DNA region that has no sequence-specific function, by random unequal crossing over

between sister chromatids. This is supposed to occur at sequences having at least a minimal degree of homology, that has arisen by chance as a result of mutation. Additional cycles of unequal crossing over at the resulting duplicated sequences will lead to periodicities. When these duplicated (or multiplied) sequences are contained within a longer unit which is tandemly repeated (e.g. an rDNA unit), the duplication (or deletion) has a finite chance to become fixed within the population of units by unequal crossing over between the units. Unequal crossing over between rDNA units in yeast has been experimentally documented (32,33). Unequal crossing over within an rDNA unit will generate a polymorphism in the frequency of reiteration of a certain direct repeat. This is indeed found in the NTS of yeast rDNA (see below). No duplications of tracts longer than 16 bp are observed. This is consistent with the observed, apparent length homogeneity of the yeast rDNA unit. Duplication of longer tracts should have given rise to a detectable length polymorphism as observed for Xenopus and Drosophila (6).

Sequence differences in the NTS of different yeast rDNA clones

The sequences of the NTS of S. cerevisiae and S. carlsbergensis rDNA are 93% homologous. This figure is very close to the 96% homology reported for S. cerevisiae and S. carlsbergensis total DNA on the basis of DNA-DNA hybridization (49). The 7% non-homology comprises about 2.5% single base substitutions, deletions or insertions, 1.5% differences in the lengths of homopolymeric tracts while 3% is due to a different number of reiterations of specific tracts. A contribution of cloning artefacts to the latter two categories cannot rigorously be excluded. The differences between both sequences are of the same order of magnitude as those between our S. cerevisiae rDNA clone and that studied by Valenzuela et al. (14). It is remarkable that many of the single base differences in Fig. 2 (viz. 16 out of the 18 in NTS1 that can be compared) are also observed upon comparison of our S. cerevisiae sequence and that of Valenzuela et al. (14), whereas the S. carlsbergensis sequence is identical with the S. cerevisiae sequence determined by Valenzuela et al. at these positions. The two S. cerevisiae clones also differ in the lengths of the homopolymeric tracts. Finally, Fig. 3 illustrates two highly polymorphous regions. One of them, present at about 220 bp downstream of the 26S rRNA gene, has also been sequenced by Swanson and Holland in a third S. cerevisiae rDNA clone (28). In this case again the deviation of the S. carlsbergensis sequence from those of the S. cerevisiae clones does not exceed the differences between the individual S. cerevisiae sequences. We are unable to identify any significant species-specific sequence difference between the

TTTTATTTC TTCTAAATG GGTACTGGCA GGAGCCGGGG CCTAGTTTAG AGAGAATAG ACTGAACAAG TCCTATAAAA TTTTATTTC CTTAAGAATT 100
 TTTTATTTC TTCTAAATG GGTACTGGCA GGAGCCGGGG CCT-CTTTTAG AGAGAATAG * ACTGAACAAG TCCTATAAAA TTTTATTTC CTTAAGAATT 99
 CTATGATCCG GGTAAAACA TGTATTGAT ATATCTATTA TAATATACGA TGAGGATGAT AGTGTGAAG AGTGTACCAT TTACTAATGT ATGTAAGTTA 200
 CTATGATCCG GGTAAAACA TGTATTGAT ATATCTATTA TAGTATACGA TGAGGATGAT AGTGTGAAG AGTGTACCAT TTACTAATGT ATGTAAGTTA 199
 CTAATTACTA TTTGGTCTTT TTAATTTTAA TTTTFTTTTT TTTTFTCGTT GCAAAGATGG GTTGAAGAG AAGGCTTTT ACAAACTTC CCGAGCCGTGA 300
 CTAT- - - - -TTGGTCTTT TTAATTT- - - - -TTT TTTTFTCGTT GCAAAGATGG * GTTAAAAGAG AA-GCTTTT ACAAACTTC CCGAGCCGTGA 280
 AAGATTTGC CCGACAGTT TGCTTCATGG AGCATTTTT TCCGCACCAT CAGAGCGCA AACATGAGTG CTTCATPAAG TTTAGAGAA TCGAAAAAGC 400
 * AAGATTTGC CCGACAGTT TGCTTCATGG AGCATTTTT TCCGCACCAT CAGAGCGCA AACATGAGTG CTTCATPAAG TTTAGAGAA TCGAAAAAGC 380
 TCAATTTCTA TAGTAAACAG GGCATGCCTT TGAGATGGAA AAAAATAGCG ACCAATPACG GTTTCGCCAA GGAAGAATGA AGAGACAAT GAACAGTGA 500
 TCAATTTCTA TAGTAAACAG GACATGCCTT TGATATG-AA AAAAATA-CT ACCAATPACG ATTTTACCAG AAAAGATGTA AGAGACAAT GAACAGTGA 478
 CAGTGAACAG T- - - - -GGG GACATTTTT TTTTFTTTTT TTTTFTTTTT GTAAATGGCA GTTTCAGCG AATGATGAT GCAAATTCOA GAGAGCCAGC 594
 CAGTGAACAG T- - - - -GGG GACATTTTT TTTTFTTTTT TTTTFTTTTT TTTTFTTTTT GTAAATGGCA GTTTCAGCG AATGATGAT GCAAATTCOA GAGAGCCAGC 565
 GTAAAAGAA CAGGCTACTG GGAAGAAGAA AGAGGAAAAG TCGAAGATGA ATAGCCAGTG CAATATAT- - - - -ACAGTATC TTAACGGATA TGGCATGGTT 592
 GTAAAAGGAT GAGGCTACTG GGAAGAAGAA AGAGGAAAAG TCGAAGATGA ATAGCCAGTG CAATATAT * * * * * ACAGTATC TTAACGGATA TGGCATGGTT 565
 GGCAGAAATA 2 TGGTTGCGG AAGTAAATTT TTGGCCAGCG GGCAGCGCG CGACCGGGTA TCCGGATTTG TA-GATGTAC TACGATCATA TAGTGTATG 791
 GGCAGAAATA 2 TGGTTGCGG AAGTAAATTT TTGGCCAGCG GGCAGCGCG CGACCGGGTA TCCGGATTTG TA-GATGTAC TACGATCATA TAGTGTATG 733
 CGGCAAAATG TTAGTGCAGG AAAGCGGAA GGAATAAGAA GAAATAAGAA GAAATAAGAA GAAATAAGAA TCCGAGGATA TACGATCATA TAGTGTATG * 891
 CGGCAAAATG TTAGTGCAGG AAAGCGGAA GGAATAAGAA GAAATAAGAA GAAATAAGAA GAAATAAGAA TCCGAGGATA TACGATCATA TAGTGTATG * 833
 AAATGGAGG AAGAGAACA AACCGGAGT GTTTTTTTTT TTTTFTTTT AGGAACATCG GAGGAGAATA T- - - - -TGT 865
 AAATGGAGG AAGAGAACA AACCGG-AGT G- - - -TTTT TTTTFTTTT AGGAATATCG GAGGAGAATA TCCGAGGATA ATATCGGAGG AGAATATTTG 928
 * * * * * TTTTFTTTT AGGAATATCG GAGGAGAATA TCCGAGGATA ATATCGGAGG AGAATATTTG * * * * *
 TTGTACATPAT CAAGTAGTAG CAACCAATG AGCATAATGG AGTGTAAAC TCCTCAGAAG AAGAGTGCAG CTGGATAGTG CG-GTTTGC CGAATCGAAT 1084
 TTGTACATPAT CAAGTAGTAG CAACCAATG AGCATAATGG AGTGTAAAC TCCTCAGAAG AAGAGTGCAG CTGGATAGTG CGAATTTTTC TGAATCGAAT * 1028
 GGTAGGTTAG TTAGTGGATT TAGCATAGGA AGCCAAGAAA CTAGAAAAAA AAAAAAAA AAAAAAAA AAGAAAATAA AGATTGCAGC ACCTCAGTTT 1164
 GGTAGGTTAG TTAGTGGATT TAGCATAGGA AGCCAAGAAA CTAG- - - - -AAAAA AAAAAAAA AAAAAAAA AAGAAAATAA AGATTGCAGC ACCTCAGTTT 1117
 * * * * * TTTTFTTTT AGGAATATCG GAGGAGAATA T- - - - -TGT 865
 CCGGATGGT CACCCACTAC ACTACTCGGT CAGGCTCTTA CCAGCTTAA CACAGTTGAT CCGACGGGAA ACGGTGTCTT CTGGTAGATA TGGCCGCAAC 1264
 CCGGATGGT CACCCACTAC ACTACTCGGT CAGGCTCTTA CCAGCTTAA CACAGTTGAT CCGACGGGAA ACGGTGTCTT CTGGTAGATA TGGCCGCAAC 1217
 CTATAGTTTA ACGGAAACG AGGTGATATG AGGCAGGGT CCAGACATGT TCAGTAGGTC GGAGTGAGAG GTGTTATGGG TGGAGGACAA TATTTATTAT 1364
 CTATAGTTTA ACGGAAACG AGGTGATATG AGGCAGGGT CCAGACATGT TCAGTAGGTC GGAGTGAGAG GTGTTATGGG TGGAGGACAA TATTTATTAT 1317
 ATTTCACTA ATAGCAATAG GATATGACAG GTGAAAAGC AAAAGCAATA GTGCATTTG ATGTGGACAA TAAGTGCAT ACCATGAAA AGGTGATTTG 1464
 ATTTCACTA ATAGCAATAG GATATGACAG GTGAAAAGC AAAAGCAATA GTGCATTTG ATGTGGACAA TAAGTGCAT ACCATGAAA AGGTGATTTG 1417

TCATTACAA GAGGTAGTTC GAACAGAAC ATGAAAGTTG CTCGGTAGGT GGCATGCAGA GGTAGTTTCA AGGTGACAGG TTATGAAGAT ATGGTGCAA 1564
 TCATTACAA GAGGTAGTTC GAACAGAAC ATGAAAGTTG CTCGGTAGGT GGCATGCAGA GGTAGTTTCA AGGTGACAGG TTATGAAGAT ATGGTGCAA 1517
 AAGCAAAATG GATGGTGGCA GGCATAGTAA AATGATGGTC TGGAAAGCAT AGATGGTATT TGTTTTGCAT TTACGGCAAC GGATCGGGC GATAATGACG 1664
 A-GACAAATG GATGGTGGCA GGCATAGTAA AATGATGGTC TGGAAAGCAT AGATGGTATT TGTTTTGCAT TTACGGCAAC GGATCGGGC GATAATGACG 1616
 GGAAGAGATT TAGTATGTTG GGCAGAAATG TCGCGGGCAG TAITTGAGACC ATGAGAGTAG CAACCGTAG TCTAAAGGTT GTTTTATAGT AGTTAGGATG 1764
 GGAAGAGATT TAGTATGTTG GGCAGAAATG TCGCGGGCAG TAITTGAGACC ATGAGAGTAG CAACCGTAG TCTAAAGGTT GTTTTATAGT AGTTAGGATG 1715
 TAGAAAATGT ATTCCGATAG CCCATTTTAC ATTTGGAGGG ACGGTTGAAA GTGGACAGAG GAAAAGGTGC GAAAATGGCT GATTTTGATT GTTTATGTTT 1864
 TAGAAAATGT ATTCCGATAG CCCATTTTAC ATTTGGAGGG ACGGTTGAAA GTGGACAGAG GAAAAGGTGC GAAAATGGCT GATTTTGATT GTTTATGTTT 1815
 TGTGTGATG TTTTACATTT TGGCATACTA TTAGTATGTC AGATGAAAGA TGAATAGACC TAGGAGTAG AMAATATAGA ATAGTTACCG TTATTGGTAG 1964
 TGTGTGATG TTTTACATTT TGGCATACTA TTAGTATGTC AGATGAAAGA TGAATAGACC TAGGAGTAG AMAATATAGA ATAGTTACCG TTATTGGTAG 1915
 GAGTGTGTG GGTGTGTATA GTCCGCATTG GGAATGTTACT TTCCGTGTAT GGCATGAATT TCCCTTTAGG GTCTCTGAAG CGTATTTCCG TCACC-ATAT 2063
 GAGTGTGTG GGTGTGTATA GTCCGCATTG GGAATGTTACT TTCCGTGTAT GGCATGAATT TCCCTTTAGG GTCTCTGAAG CGTATTTCCG TCACC-ATAT 2015
 -GGCAGATAT AGGAAACTG AAGGAGGAT AGTAGTAAAG TTTCGAATGGT GGTAGTGTAT ATGTATGTTA CCGTTTGGT TTGGTTTCCG GTTGTGAAA 2162
 AGCCAGAAA AGGAAAGTG AAGGAGGAT AGTAGTAAAG TTTCGAATGGT GGTAGTGTAT ATGTAGTATA TCCGTTGGT TTGGTTTCCG-6 GTTGTGAAA 2113
 GTTTTTGGT ATGATATTTT GCAAATAGCA TATATTTCTT CTGTGAGAAA GGTATATTTT GTATGTTTTG TATGTTCCCG CCGGTTTCCG TAITTTCCCG 2262
 GTTTTTGGT ATGATATTTT GCAAATAGCA TATATTTCTT CTGTGAGAAA GGTATATTTT GTATGTTTTG TATGTTCCCG CCGGTTTCCG TAITTTCCCG 2213
 -----TTCC GCTTCCGCAG TAAAAATAG TGAGGAACCTG GGTACCOCGG GGCACCTGTC ACTTTGGAAA AAAAA--TAT AGGCTAAGAT TTTTGGAGAA 2354
 TCCGCTTCC GCTTCCGCAG TAAAAATAG CGAGG-ACTG GGTACCOCGG GGCACCTGTC ACTTTGGAAA AAAAAATAT ACGCTAAGAT TTTTGGAGAA 2312

 TAGCTTAAAT TGAAGTTTT CTCCGGCAGA AATACCTAGT TAAGCCAGAG CGACAGAGAG GGCAAAAGAA AATAAAGTA AGATTTTATG TTGTAATGGG 2454
 TAGCTTAAAT TGAAGTTTT CTCCGGCAGA AATACCTAGT TAAGCCAGAG CGACAGAGAG GGCAAAAGAA AATAAAGTA AGATTTTATG TTGTAATGGG 2412
 AGGGGGGTT TAGTCATGGA GTACAAGTGT GAGGAAAGT AGTTGGGAGG TACTTTCTCC GAAGCAGTT GAAGACAAGT TCGAAAAGAG TTTGGAACG 2554
 AGGGGGGTT TAGTCATGGA GTACAAGTGT GAGGAAAGT AGTTGGGAGG TACTTTCTCC GAAGCAGTT GAAGACAAGT TCGAAAAGAG TTTGGAACG 2512

Fig. 2. Nucleotide sequence of the NTS of *S. cerevisiae* (upper) and *S. carlsbergensis* (lower) rDNA. The nucleotide adjacent to the 3'-terminal T of the 26S rRNA gene is numbered 1. For *S. cerevisiae* the sequences 1-406 and 2139-2554 are taken from Ref. 12 and 15, whereas the sequences 97-470, 859-1808 and 2238-2554 have been reported for another clone of *S. cerevisiae* rDNA (14) as well as the sequence 97-287 for still another one (28). For *S. carlsbergensis* the terminal stretches 1-264 and 2260-2512 are taken from other papers (13, Klootwijk et al., in preparation). Sequence differences are marked by an asterisk and deletions by a hyphen. The 5S rRNA gene is represented by the box from 1265 to 1145 (*S. cerevisiae*) and the beginning of the large rRNA operon is indicated by the boxed sequences starting at position 2510 (*S. cerevisiae*). Some of the direct repeats, marked by overlining, are numbered (1 to 7).

A.

1. CCATTACTA ATGTATGTAA GTTACTA--- ---TTTGGT CTTTTATTI -----T TTTTTTTTCG (176-229)
2. CCATTACTA ATGTATGTAA GTTACTATTI ACTATTGGT CTTTTATTI TTTATTI-----TTTT TTTTTTTTCG (177-248)
3. CCATTAACT A----- ---TTTGGT CTTTTATTI TTTATTITC-----TTTT TTTTTTTTCG
4. CCATTACTA ----- ---TTTGGT CTTTTATTI TTTATTITGT TATTITTTI TTTTTTTTCG

B.

1. CGCGTTCCG TATTTCCGC -----T TCCGCTCCG CTCCGCAGT (2194-2234)
2. CGCGTTCCG TATTTCCGC -----TTCCG CTCCGCAGT (2243-2277)
3. CGCGTTCCG TATTTCCGC TTCCGTATT TCCGCTCCG TATTTCCGC TTCCGTATT TCCGCTCCG CTCCGCAGT

Fig. 3. Two highly variable regions in the NTS of yeast rDNA. Sequence 1 (*S. carlsbergensis*) and 2 (*S. cerevisiae*) are taken from Fig. 2, sequence 3 from Valenzuela et al. (14) and sequence 4 from Swanson and Holland (28).

NTS of *S. carlsbergensis* and that of *S. cerevisiae*. Since the NTS is the most rapidly evolving part of the yeast rDNA unit (34) we expect that different species will differ in their NTS sequence as was observed for instance in sibling species of the *Drosophila melanogaster* subgroup (35). Also, examination of other genes of *S. cerevisiae* and *S. carlsbergensis*, which are interfertile, only reveals small differences between both strains e.g. the actin genes (51) and the genes for ribosomal proteins (C.M.T. Molenaar, this laboratory, and N.J. Pearson, University of Maryland, personal communication). So far, significant differences between both strains have only been described for parts of chromosome III (52). The extended homology in Fig. 2 and the apparent lack of any significant species-specific sequence in the rDNA is consistent with recent revisions in yeast taxonomy, where *S. carlsbergensis* is now classified as a synonym of *S. cerevisiae* (50). Therefore, we shall discuss the *S. carlsbergensis* (NCYC74) rDNA unit in Fig. 2 as an *S. cerevisiae* allele.

The polymorphic sequence in Fig. 3A has previously been proposed to play a role in transcription termination (14). However, the observed variability of this sequence makes this assignment highly unlikely. Recently, this sequence was also suggested to contain a transcription initiation site (28). More detailed studies have shown, however, that the sequence directly adjacent to this region (position 252 to 283 of the *S. cerevisiae* sequence) can promote RNA polymerase A transcription *in vitro* (M.J. Holland, personal communication).

The polymorphic region depicted in Fig. 3B shows that the sequence TTCCGTATTTCCGC, present once in two of the three clones, is tandemly repeated four times in the third one. A terminal part of this 15 bp sequence, TTCCGC, is found two or three times directly adjacent to the full 15 bp sequence. These reiterations can be explained by assuming several successive cycles of unequal crossing over to have occurred at this site. The inherent polymorphism is seen also at a number of other sites, in particular in the NTS1 (see Fig. 2, the direct repeats

1 to 4 at position 500, 700, 740 and 960 resp., of the S. cerevisiae sequence).

Function of the NTS of yeast rDNA

A) Is the NTS involved in the 'concerted' evolution of the yeast rDNA unit?

Unequal crossing over, gene conversion and transposition can all account for the maintenance of sequence homogeneity between the rDNA units of an organism. Both unequal crossing over and gene conversion have been shown to occur in the rDNA cluster of yeast (32,33,36). It has been hypothesized that the NTS of rDNA in higher eukaryotes served as a 'hot spot' for recombination (37) because of long periodicities. However, similar long periodicities are absent in the NTS of yeast rDNA.

B) Is the NTS involved in the regulation of rRNA synthesis?

In both Xenopus and Drosophila the putative RNA polymerase A promoter is reiterated within the NTS (7-9). These repeats have been suggested to play a role in the regulation of transcription by sequestering RNA polymerase A molecules or factors. The NTS of yeast rDNA does not contain similar duplications of the promoter region and therefore a contribution of the NTS to the regulation of rRNA synthesis in yeast, similar to that in Xenopus or Drosophila, is not apparent from the sequence.

C) Does the NTS contain any other transcription unit than the ones for the rRNAs?

The rRNA operons of Escherichia coli (38) and chloroplasts (39) contain tRNA genes. A computer search for the presence of tRNA genes however gave negative results. Further inspection of the sequences did reveal some open reading frames. For instance, a polypeptide of 71 amino acids is encoded by the sequence from position 1841 to 1629 (S. carlsbergensis). However, no specific transcripts of the NTS could be detected by R-looping of cloned NTS-DNA with total S. carlsbergensis RNA (H. van Heerikhuizen, personal communication). Therefore, we consider these open reading frames to be without any significance.

D) Does the NTS contain an origin of replication?

Szostak and Wu (18) previously reported the ability of the BglII-generated fragment B of yeast rDNA, after insertion into a plasmid, to transform yeast with a relatively high frequency. The transformants appeared to be very unstable. Apparently, yeast rDNA contains an ARS (autonomously replicating sequence), which may correspond to a normal chromosomal origin of replication or 'replicator' (18). The high mitotic instability is most likely due to the low copy number of the plasmid and its lack of a centromere. Non-random assortment to the daughter cells will frequently result in cells devoid of plasmid. Recent studies of Larionov and Shubochkina (19) limited the ARS of yeast rDNA to the EcoRI-B fragment (see Fig. 4) and confirmed the suggestion

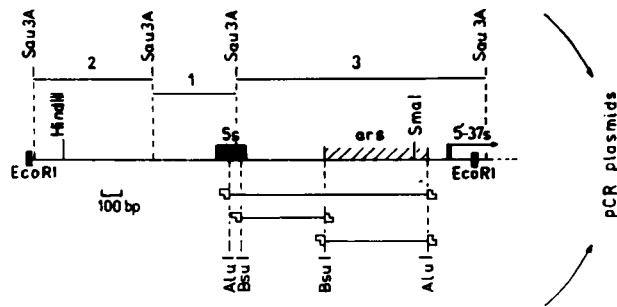


Fig. 4. Map of the NTS of yeast rDNA as contained with the EcoRI fragment B. The 5S rRNA gene and the start of 37S pre-rRNA are indicated as well as the smallest fragment (hatched) which sustained autonomous replication in yeast (ARS). Sau3A fragments 1 (775-1221, see Fig. 2), 2 (105-774) and 3 (1222-2674, containing a small part of the adjacent EcoRI fragment G) were cloned into YIp1 and designated as pCR (Sau) 1, 2 and 3 respectively. AluI-fragment (1209-2357) and AluI/BsuI fragment (1786-2357) were cloned into YIp1 with synthetic HindIII-linkers and designated as pCR (Alu) and pCR (Alu/Bsu) respectively.

of Szostak and Wu (18) that the ARS should be located within the NTS. We inserted subfragments of the EcoRI-fragment B generated by Sau3A (see Fig. 4) into YIp1 which contains the LEU2 gene as a genetic marker. *S. cerevisiae* DC5 (leu2⁻) was transformed by the pCR (Sau) plasmids 1, 2 and 3 and the previously tested plasmids pCR(A) and pCR(B+G), which contain the EcoRI-fragments A and B+G of yeast rDNA respectively. YEp13 containing the 2 μ replicator was used as a reference. The frequency of transformation and the stability of the various plasmids are listed in Table I. Of the pCR(Sau) series, only pCR(Sau)3 was able to transform DC5 with a frequency comparable to YEp13 and pCR(B+G). In addition, pCR(Sau)3 transformants lose the LEU⁺ phenotype with a very high frequency, whereas pCR(Sau)1 and 2 yielded very few transformants, most likely because of integration into the chromosome as inferred from their mitotic stability. These data indicate that the ARS of yeast rDNA resides within NTS2. This was confirmed by the transformation characteristics of pCR(Alu), which contains almost completely the NTS2 region (position 1209 to 2357 for *S. cerevisiae*, Fig. 2). pCR(Alu) also displays a high transformation frequency and a very low mitotic stability (see Table I). These figures are indistinguishable from that of pCR(Alu/Bsu) although this plasmid contains only a part of the NTS2 (position 1786 to 2357). Therefore the ARS is located within the latter portion of NTS2. The mitotic stabilities of the plasmids containing the ARS of rDNA are more than one order of magnitude lower than those reported for plasmids containing the ARS1 or ARS2 segment (24,40),

TABLE I

Transformation of *S. cerevisiae* DC5 with hybrid plasmids carrying various yeast DNA fragments (see Fig. 1 and 4).

Plasmid	Frequency of transformation ^{a)}	Percentage of transformants that retained the plasmid after growth on selective medium
pCR(A)	0.1%	100% ^{b)}
pCR(B+G)	80-100%	1-2%
pCR(Sau) 1	0.1%	100%
pCR(Sau) 2	0.1%	100%
pCR(Sau) 3	80-100%	< 1%
pCR(Alu)	60-80%	< 1%
pCR(Alu/Bsu)	60-80%	< 1%
pCR(B+G) CEN	100%	10-20%
pCR(Alu) CEN	100%	10-20%
pCR(Alu/Bsu) CEN	100%	10-20%
pCR-CEN	< 1%	100%

a) The frequency of transformation was determined as the number of colonies relative to the number transformed by YEp13 (generally about 3000 colonies per μ g) for the pCR series and relative to the number transformed by pYe(CEN3)41 (generally about 2000 colonies per μ g) for the pCR-CEN series.

b) Frequency of marker loss after integration into the chromosome is 10^{-3} .

indicating that the putative rDNA replicator is a rather weak one. This was confirmed by the transformation characteristics of derivatives of pCR(Alu) and pCR(Alu/Bsu) containing the centromeric segment CEN3. This segment, isolated from pYe(CEN3)41, has been shown to ensure proper plasmid segregation (23). However, in pCR(Alu) and pCR(Alu/Bsu) the CEN3 region has a much less profound effect on the plasmid stability (see Table I) than reported for other chromosomal replicators (23,41). Therefore, we have to conclude that the ARS of yeast rDNA is rather inefficient.

Our mapping of the ARS of rDNA within the NTS2 is consistent with the position of small replication bubbles observed in actively transcribed rDNA of synchronized *S. cerevisiae* cdc7-1 cells. Their location suggests a specific origin of replication in or near the promoter proximal site of the NTS (L. Saffer, personal communication). Specific origins of replication have also been attributed to the NTS regions of amplified rDNA genes in *Tetrahymena* (42)

```

A / T T T T A T P u T T T A / T  ARS consensus sequence (48)
G T T T A T G T T T T          (position 1855-1865)
T T C T A T A T T T T          ( "      1945-1935)
T T G T A T G T T T T          ( "      2221-2231)
    
```

Fig. 5. Sequences within NTS2 of *S. cerevisiae* that resemble the ARS consensus sequence.

and *Physarum* (43) and to chromosomal rDNA of *Lytechinus* (44). Our data show that neither the transcription startpoint of 5S rRNA nor that of 37S pre-rRNA are present within the yeast rDNA fragment containing the ARS. If the yeast rDNA-ARS represents a legitimate replicator it must be considered as a potential one, since the rDNA units do not form separate replicons (L. Saffer, personal communication). This rDNA replicator may also be responsible for the propagation of the extrachromosomal copies of the rDNA unit, that have been found in small amounts in various yeast strains (45-47). There is a strain-to-strain variation in the ability to maintain these extrachromosomal genes (Larionov *et al.*, in preparation). This suggests that the activity of the rDNA replicator varies in different strains. This observation may be accounted for by sequence heterogeneity of the yeast rDNA-ARS similar to that observed at other positions of the NTS. Inspection of the "ARS-containing region" in *S. cerevisiae* reveals at least three sequences (cf. Fig. 5) that nearly fit the ARS consensus (48). Further functional analysis will have to be carried out to ascertain the possible significance of these sequences.

ACKNOWLEDGEMENTS

The authors are grateful to their colleagues for their comments and discussions. Part of this work was supported by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

*To whom requests for reprints should be sent

REFERENCES

1. Petes, T.D. (1979) Proc. Natl. Acad. Sci. USA 76, 410-414.
2. Zamb, T.J. and Petes, T.D. (1982) Cell 28, 355-364.
3. Bell, G.I., DeGennaro, L.J., Gelfand, D.H., Bishop, R.J., Valenzuela, P. and Rutter, W.J. (1977) J. Biol. Chem. 252, 8118-8125.
4. Nath, K. and Bollon, A.P. (1977) J. Biol. Chem. 252, 6562-6571.
5. Meyerink, J.H. and Retèl, J. (1976) Nucl. Acids Res. 3, 2697-2707.
6. Long, E.O. and Dawid, I.B. (1980) Ann. Rev. Biochem. 49, 727-764.

7. Moss, T. and Birnstiel, M.L. (1979) *Nucl. Acids Res.* 6, 3733-3743.
8. Simeone, A., DeFalco, A., Macino, G. and Boncinelli, E. (1982) *Nucl. Acids Res.* 10, 8263-8272.
9. Coen, E.S. and Dover, G.A. (1982) *Nucl. Acids Res.* 10, 7017-7026.
10. Moss, T. (1983) *Nature* 302, 221-228.
11. Miller, J.R., Hayward, D.C. and Glover, D.M. (1983) *Nucl. Acids Res.* 11, 11-19.
12. Bayev, A.A., Georgiev, O.I., Hadjiolov, A.A., Nikolaev, N., Skryabin, K.G. and Zakharyev, V.M. (1981) *Nucl. Acids Res.* 9, 789-799.
13. Veldman, G.M., Klootwijk, J., De Jonge, P., Leer, R.J. and Planta, R.J. *Nucl. Acids Res.* 8, 5179-5192.
14. Valenzuela, P., Bell, G.I., Venegas, A., Sewell, E.T., Masiarz, F.R., DeGennaro, L.J., Weinberg, F. and Rutter, W.J. (1977) *J. Biol. Chem.* 252, 8126-8135.
15. Bayev, A.A., Georgiev, O.I., Hadjiolov, A.A., Kermekchiev, M.B., Nikolaev, N., Skryabin, K.G. and Zakharyev, V.M. (1980) *Nucl. Acids Res.* 8, 4919-4926.
16. Veldman, G.M., Klootwijk, J., De Regt, V.C.H.F., Planta, R.J., Branlant, C., Krol, A. and Ebel, J-P. (1981) *Nucl. Acids Res.* 9, 6935-6952.
17. Georgiev, O.I., Nikolaev, N., Hadjiolov, A.A., Skryabin, K.G., Zakharyev, V.M. and Bayev, A.A. (1981) *Nucl. Acids Res.* 9, 6953-6958.
18. Szostak, J.W. and Wu, R. (1979) *Plasmid* 2, 536-554.
19. Larionov, L. and Shubochkina, E.A. (1982) *Molekularnaya Biologiya* 948-960.
20. Meyerink, J.H. (1979) Ph.D. Thesis, Free University, Amsterdam.
21. Petes, T.D., Hereford, L.M. and Skryabin, K.G. (1978) *J. Bacteriol.* 134, 295-305.
22. Broach, J.R., Strathern, J.N. and Hicks, J.B. (1979) *Gene* 8, 121-133.
23. Clarke, L. and Carbon, J. (1980) *Nature* 287, 504-509.
24. Tschumper, G. and Carbon, J. (1980) *Gene* 10, 157-166.
25. Humphreys, G.O., Willshaw, G.A. and Anderson, E.S. (1975) *Biochim. Biophys. Acta* 383, 457-463.
26. Clewell, D.B. and Helinski, D.R. (1970) *Biochemistry* 9, 4428-4440.
27. Maxam, A. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
28. Swanson, M.E. and Holland, M.J. (1983) *J. Biol. Chem.* 258, 3242-3250.
29. Cramer, J.H. and Rownd, R.H. (1980) *Mol. Gen. Genet.* 177, 199-205.
30. Cramer, J.H., Bhargava, M.M. and Halvorson, H.O. (1972) *J. Mol. Biol.* 71, 11-20.
31. Smith, G.P. (1976) *Science* 191, 528-535.
32. Petes, T.D. (1980) *Cell* 19, 765-774.
33. Szostak, J.W. and Wu, R. (1980) *Nature* 284, 426-430.
34. Verbeet, M.P., Klootwijk, J., Van Heerikhuizen, H., Fontijn, R., Vreugdenhil, E. and Planta, R.J. (1983) *Gene* 23, 53-63.
35. Coen, E., Strachan, T. and Dover, G. (1982) *J. Mol. Biol.* 158, 17-35.
36. Klein, H.L. and Petes, T.D. (1981) *Nature* 289, 144-148.
37. Treco, D., Brownell, E.R. and Arnheim, N. (1982) in: *The Cell Nucleus, Vol. XIIC* (ed. Busch, H. and Rothblum, L.) pp. 101-126, Academic Press.
38. Lund, E., Dahlberg, J.E., Lindahl, L., Jaskunas, S.R., Dennis, P.P. and Nomura, M. (1976) *Cell* 7, 165-177.
39. Koch, W., Edwards, K. and Kössel, H. (1981) *Cell* 25, 203-213.
40. Hsiao, C-L. and Carbon, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3829-3833.
41. Hsiao, C-L. and Carbon, J. (1980) *Gene* 15, 107-117.
42. Cech, T. and Brehn, S.L. (1981) *Nucl. Acids Res.* 9, 3531-3543.
43. Truett, M.A. and Gall, J.G. (1977) *Chromosoma* 64, 295-303.
44. Botchan, P.M. and Dayton, A.I. (1982) *Nature* 299, 453-456.
45. Meyerink, J.H., Klootwijk, J., Planta, R.J., Van der Ende, A. and

- Van Bruggen, E.F.J. (1979) *Nucl. Acids Res.* 7, 69-76.
46. Clark-Walker, G.D. and Azad, A.A. (1980) *Nucl. Acids Res.* 8, 1009-1022.
47. Larionov, V.L., Grishin, A.V. and Smirnov, M.N. (1980) *Gene* 12, 41-49.
48. Broach, J.R., Li, Y-Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A. and Hicks, J.B. (1982) *Cold Spring Harbor Symp. Quant. Biol.* 47, 1165-1173.
49. Bicknell, J.N. and Douglas, H.C. (1970) *J. Bacteriol.* 101, 505-512.
50. Barnett, J.A., Payne, R.W. and Yarrow, D. (1983) *Yeasts: Characteristics and identification*, Cambridge University Press, pp. 467-469.
51. Nellen, W., Donath, C., Moos, M. and Gallwitz, D. (1981) *J. Mol. Appl. Gen.* 1, 239-244.
52. Holmberg, S. (1982) *Carlsb. Res. Comm.* 47, 233-244.