Vol. 170, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 16, 1990 Pages 187-193

PRIMARY STRUCTURE OF THE NON-TRANSCRIBED SPACER REGION AND FLANKING SEQUENCES OF THE RIBOSOMAL DNA OF *NEUROSPORA CRASSA* AND COMPARISON WITH OTHER ORGANISMS

S.K. Dutta and Mukesh Verma

Molecular Genetics Laboratory, Department of Botany
Howard University, Washington, DC 20059

Received May 24, 1990

The non-transcribed spacer (NTS) region of the rDNA of *Neurospora crassa* contains the transcription regulatory sequences. We isolated a 3.4 kb EcoRI fragment from wild type *N.crassa* rDNA and cloned in the plasmid pBR325 at the EcoRI site. The insert contains the entire NTS region along with the flanking sequences. Nucleotide sequencing of 3592 nt shows many interesting features like: the NTS region is rich in G+C content (65% G+C); it contains the conserved rRNA processing site 6 (with the nucleotide sequence motif GGTGCGAGAACCCGG, from nt residue 226 to 240, a characteristic feature of most eukaryotic rDNA nontranscribed spacer region); and the NTS region also contains the transcription termination site with the representative Sal I box (from nt residue 1469 to 1477). The potential sequences of transcription termination site are located 288 nt downstream from the end of 26S rRNA gene, and another sequence motif CTTCCT (from nt residue 512 to 517) shows similarity with the human transcription termination site T-2 of its pre-rRNA. Nucleotide sequence homology matrix analysis suggests its relatedness to *Saccharomyces cerevisiae* and not to human, mouse, rat, Drosophila, Xenopus, wheat, rice and cucumber NTS region. The phylogenetic implication of the NTS region and exploitation of *N.crassa* NTS rDNA clone to correlate the otherwise indistinguishable species of Neurospora and the correlation with other organisms has been discussed. To the best of our knowledge this is the first report where the nucleotide sequence of the entire NTS region of a filamentous fungus has been determined.

Determination of the primary structure of rRNA molecules and the respective gene is important for understanding ribosome structure, function, biogenesis and evolution. Studies on molecular characterization and regulation of rRNA gene expression are greatly facilitated by isolation of rDNA clones. In eukaryotes, the 17S, 5.8S, and 25S rRNA genes are encoded in multiple copies and are transcribed as a large precursor (pre-rRNA) which is processed to give mature rRNAs (1-4). In *Saccharomyces cerevisiae* there is close linkage of 5S rRNA genes with the rRNA transcriptional unit, whereas in *N. crassa* and *Schizosaccharomyces pombe*, the 5S rRNA genes are unlinked to the major rRNA transcription unit (1). In yeast and Neurospora, as is true in most eukaryotes, more than a hundred copies of the rRNA genes are repeated in a tandem array (2). The repeat units are reasonably homogeneous within a given species. For example, the repeat unit is 9.2 kb in yeast and 8.7 kb in *N.crassa* (3, 4). In *N.crassa* each rRNA repeat unit codes for a 35S rRNA precursor molecule which,

0006-291X/90 $1.50

during maturation, gives rise to 17S, 5.8S, and 26S rRNAs. The organization of rRNA genes in *N.crassa* is as follows: external transcribed spacer (ETS), 17S coding region, internal transcribed spacer region-1 (ITS-1), 5.8S coding region, internal transcribed spacer region-2 (ITS-2), 26S coding region, terminal external transcribed spacer region (TETES), and NTS region. Thus in the transcription unit the coding regions are separated by NTS. Since the transcriptional regulatory sequences are present in the NTS region, we constructed a clone (namely pCC3400) containing the NTS and flanking sequences and determined its sequence. A comparison of the nucleotide sequence of the NTS region of *N. crassa* with those in other organisms has also been investigated. The phylogenetic implication of the NTS clone in identifying Neurospora species has been elucidated.

MATERIALS AND METHODS

Restriction Enzymes, rDNA Probe, Radioisotopes, and Other Chemicals

All restriction enzymes, T4 DNA ligase, DNase I, DNA polymerase I, and polynucleotide kinase were the products of Bethesda Research Laboratory, Bethesda, MD. *N. crassa* rDNA probe containing the coding sequences has been described previously (3). For nick-translation alpha-^{32}P-dCTP (3000 Ci/m mol) and for terminal labeling during nucleotide sequencing gamma-^{35}S-thio-ATP (1000 Ci/m mol) were purchased from the New England Nuclear, Wilimington, DE. Nitrocellulose membrane for Southern hybridization was from Schleicher and Schull, Keene, NH. Ultra pure agarose, low melting point agarose, urea, acrylamide, bisacrylamide, ammonium persulfate, Tris, boric acid, EDTA, ethidium bromide and other chemicals were purchased from Bethesda Research Laboratory, Bethesda, MD, and Sigma Chemical Company, St. Louis, MO. Nucleotide sequencing kit (Sequanase Version 2) was the product of United States Biochemicals, Cleveland, OH.

Cloning of the NTS Region of rDNA of *N. crassa*

The wild strain *N. crassa* 74A was obtained from the Fungal Genetics Stock Center, Kansas City, KS. Its DNA was isolated as described below. Mycelia were grown to stationay phase in liquid Vogel's minimal media (4) at 25°C for 3 days. Mycelia were collected by passing through cheesecloth and were ground in a mortar and pestle with acid washed sea sand using buffer A (25 mM triethylamine ethylene diamine tetra acetic acid, pH 8, 0.5% (v/v) Triton-X 100 and 50 mM NaCl). The suspension was centrifuged at low speed (2,000 g for 20 min at 4°C) and the supernatant was collected. The nuclear pellet was isolated from the supernatant by centrifugation (14,000 g for 60 min at 4°C) and was suspended in a buffer containing sorbitol, ficoll-400, glycerol, magnesium chloride, and Triton-X100. It was then treated with lysozyme (50 ug/ml) at 0°C for 5 min. The reaction was stopped by adding 0.05 M EDTA and 0.5M NaCl. The suspension was treated with RNase (40 uig/ml) at 37°C for 60 min followed by treatment with protease (100 ug/ml) at 37°C for 60 min. DNA was then isolated by phenol extraction (three times with equal volume of phenol saturated with 0.1M Tris-HCl, pH 9) and once with chloroform: isoamylalcohol (24:1) followed by precipitation with a double volume of cold ethanol (at -20°C for 4 hr). The DNA was centrifuged down and suspended in TE (5). After digestion of 20 ug DNA with EcoRI it was electrophoresed on 0.7% agarose gel. Two such gels with EcoRI digested sample were made. One gel was used for Southern hybridization to locate the rDNA region using *N.crassa* rDNA coding region containing clone (6) as probe and the other gel was used to elute a 3.4 kb EcoRI fragment which contains complete NTS region along with a part of the 3' end of 26S rRNA and 5' end of the 17S rRNA. This EcoRI fragment was ligated to EcoRI digested plasmid pBR325 and the construct was transformed in *Escherichia coli* HB101 (7). This plasmid (pCC3400) is resistant to ampicillin and tetracyclin but sensitive to chloramphenicol.

Subcloning and Nucleotide Sequencing

The restriction map of the NTS region of *N. crassa* rDNA is shown in Fig. 1. All the fragments shown in the figure were isolated from pCC3400 by digesting it with appropriate enzymes and

electroeluted as described (7). These fragments were subcloned in filamentous phage M13mp18 and M13mp19. Using M13 universal primer and modified T7 DNA polymerase (Sequanase 2 from United States Biochemicals, Cleveland, Ohio) the nucleotide sequencing was done by chain termination method (8, 9) and both strands were sequenced.

Analysis of Sequencing Data

The overlapping sequences were aligned by using the Pustell Sequence Analysis Program, Version 2 (International Biotechnologies, Inc., New Haven, CT) and dot matrix analysis was done according to Harshey et al. (10).

RESULTS AND DISCUSSION

The NTS and flanking sequences of rDNA are the most significant parts of the ribosomal RNA transcription unit as they contain the transcription regulatory sequences. To understand the regulation of transcription of rRNA genes of *N. crassa* it is reasonable to isolate a clone which contains these regions. We therefore cloned 3592 nt long EcoRI-EcoRI fragment isolated from wild type *N. crassa* rDNA into a plasmid pBR325 and the clone was named pCC3400. This clone was analysed extensively with several restriction enzymes and the sites of few enzymes are shown in Fig. 1. The nucleotide sequence of the insert of pCC3400 was determined for both strands (being published elsewhere) and its homology with other eukaryotic sequences in the same region was performed (Fig. 2 and Fig. 3). These sequences exhibit many interesting features and the most significant features of this analysis will be discussed in the following section.

The nucleotide sequence of the NTS region is known only for few rDNAs, viz., Xenopus, human, rat, mouse, yeast, wheat, rice, Drosophila (11-15). The common feature of the nucleotide sequence in this region among eukaryotes is that the NTS region is highly rich in its G+C content. In *N.crassa* the G+C content of the NTS region is 65%. As we mentioned earlier that pCC3400 contains some part of the 3' end of the 26S region also, it is expected that the sequences of the processing site 6 are present in this region. When we examined the *N.crassa* sequence we found that *N.crassa* contains the sequenbce motif GGTGCGAGAACCCGG (from nucleotide residue 226 to 240) which is similar to the

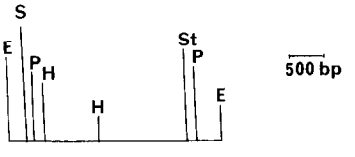

Fig. 1. Restriction sites on the insert of NTS rDNA clone pCC3400. The 3.4 kb EcoRI-EcoRI fragment of the 8.7 kb rDNA repeat unit of *N.crassa* was cloned in pBR325. Restriction analysis was done for several enzymes but in this figure only those restriction sites have been shown which are used for the nucleotide sequencing. The restriction fragments were cloned in M13mp18 and/or M13mp19 and sequenced in both directions. E, EcoRI; S, SmaI; H, Hind III; P, Pst I; St, Sst I.

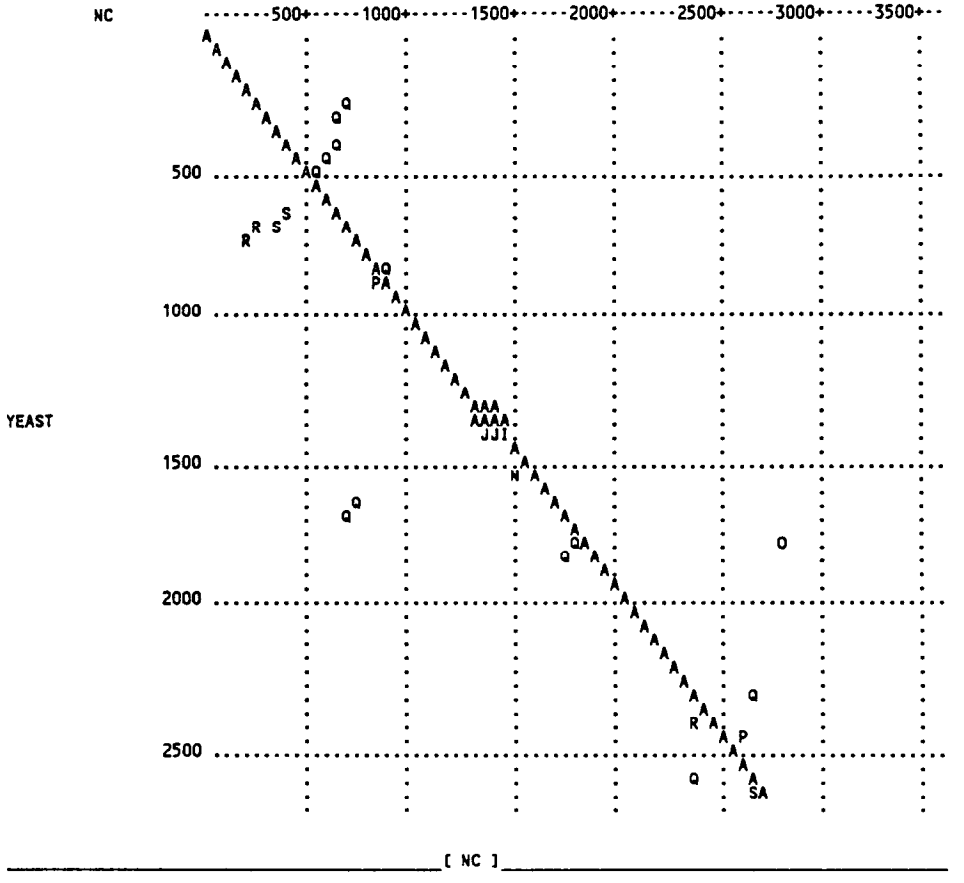

Fig. 2. Dot-matrix nucleotide homology analysis of *N. crassa* and *S. cerevisiae* NTS region. The nucleotide sequence software package distributed by International Biotechnologies, Inc. was used in the sequencing data. Numbers on both axes denote the nucleotide residues. We choose a minimal value of homology to be 50% as described (10). Letters within the plot (more clear with a magnifying glass) indicate the minimum homology found between the two sequences (A=100%, B=98-99%, C=96-97% etc. with Z=50-51%).

rRNA processing site 6 identified in yeast, Xenopus, mouse and human (13 and references therein). An exception in this category is Drosophila where the transcription of the tandem array of ribosomal DNA does not terminate at any fixed point (13). Another interesting feature of the NTS region of *N. crassa*

GTCGTCGACTC	I
AGGTCGACCAGTANTCCG	II
TCCGCACGGTCGACCAG	III
TCCG---GGTCGACGC	IV
TCCG---GGTCGACCGGCGG	V
TCCG--GGGTCGACCAG	VI
TCCG--GGGTCGACCAG	VII

Fig. 3. Sequence comparison of the Sal I boxes of *N. crassa* (I), mouse (II) (24-25), and human (III-VII) rDNA spacers. Conserved nucleotides are boxed. Note that the motif TCGAC is common to all sequences shown above.

is that it contains long stretches of pyrimidines all over, and this type of structural organization is present in maize (16), wheat (17), *Raphanus sativus* (13), *Vicia faba* (18), pea (19), cucumber (20), and rice (21).

The nucleotide sequence homology of *N. crassa* in the spacer regions was compared with other known sequences in the same region. Although the overall homology of *N. crassa* NTS is observed with *S. cerevisiae* only (Fig. 2), streches of similarity exist in other cases also. For example, the sequence in the region of transcription termination site (T1) for human shows 80% homology with *N.crassa* rDNA (in a span of 70 nt) (22). S1 nuclease mapping (7) will confirm the exact location of the transcription termination site in the case of *N. crassa.*

Compared to 8 sal I boxes in mouse rDNA in the NTS region, only one Sal I box (from nucleotide residue 1469 to 1477) is present in *N. crassa.* Sequence comparison of the Sal I box and neighbouring sequences from different organisms show that the motif TCGAC is common to all (Fig. 3). Here we would like to mention that the transcription termination starts near the Sal I box in the spacer region. The analysis of the sequencing data indicates that in *N. crassa* the end of the 26S rRNA is located at nucleotide residue 183 and 288 nt downstream of the 26S rRNA is located the transcription termination sequence motif CTCCC (from nt residue 471 to 475) which is similar to the human termination site T-1 (23). In human rRNA, the most abundant termination site is located 360 nt downstream of the 28S gene. In mouse the transcription terminates 565 nt downstream from the 3' end of 28S rRNA coding region and 24 nt upstream from the first Sal I box (24, 25). Although only one major site exists in human rDNA, however, transcription partially passed through this site and three additional 3' termini were detected inside the first repititive unit of the heterogeneous spacer region of human rDNA, about 932, 1032, and 1108 nt downstream from the 28S gene (23). It will be interesting to know whether *N.crassa* also contains more than one termination sites. However, the motif CTTCCT of *N. crassa* (from nt residue 512 to 517) is similar to the second termination site of human rRNA precursor (site T-2). S-1 nuclease mapping experiments (7) will be performed in *N.crassa* will be performed to confirm this site. It has been suggested that the neighbouring sequences highly influence the termination capacity of the Sal I box in most eukaryotes (23). In mouse the Sal I box is preceeded by long T runs and Py-rich sequences. In *Xenopus laevis* consecutive T residues were detected in front of the termination signal, while an identical nucleotide sequence, preceeded by a Py-rich region caused only a rapid processing reaction (26). In rat the inefficiently functioning termination signals (Sal I boxes)

contain point mutations (27) but these boxes are not preceeded by T runs or by Py-rich sequences. In human the T-1 site preceeded the first Sal I box of the 3' terminal spacer in a Py-rich region. Additional Py-rich sequences and T tracks appeared only downstream from the fifth box and the T-2, T-3, and T-4 potential termination sites were detected in this region. In *N. crassa* also the Sal I box is preceeded by Py-rich region. In yeast the termination site is extremely T-rich (15 out of 18 nt are T-residues) and an extensive dyad symmetry is present in the sequence downstream from the termination site which plays a significant role in the regulation of transcription (28 and references therein). These data support the idea that Py-rich sequences, and mainly the T-tracks, are important for the efficient termination and/or the processing of the pre-rRNA transcript.

In *N. crassa* the sequence motif CTGGCAAGCCATCAGC (from nt residue 2435 to 2450) shows similarity with the transcription initiation conserved sequence for mammalian ribosomal DNAs (12, 29). Another interesting feature of the NTS region of *N. crassa* is that certain sequences like TCTC, TTTT, and TTGC reitrate several times. Similar repetition of sequences has been reported for human rDNA (30) although the functional significance of these sequences is not yet established.

The implication of the present studies was done by us as follows. The genus Neurospora has homothallic, heterothallic, and pseudohomothallic species. These three incompatible groups of Neurospora are distinctly different from each other based on the presence or lack of conidial or ascospore phases (31). Several investigators have expressed concern as to whether Neurospora homothallic species should belong to genus *Gelacinospora* or genus *Sordaria*, since those genera comprise only homothallic species (31, 32). Using pCC3400 as probe and by following the restriction fragment length polymorphism (RFLP) we distinguished not only the three species of Neurospora (33) but also the members of these species. It is suggested that the differentiation (detected by RFLP) which occurs in different members of a species lies mainly in the NTS region. Similar evidence has been reported for rice also (13). As a further step involving the phylogenetic implication of the clone characterized in the present investigation (pCC3400), we constructed the most parsimonous evolutionary tree (based on the number of phylogenetically singnificant events) of the members of fungal and related and unrelated groups (26).

Determination of the *N. crassa* transcription initiation and termination sites *in vivo* and *in vitro* using clone pCC3400 characterized in the present investigation await the results of the appropriate experimental design and as such are a continuing focus of our laboratory.

ACKNOWLEDGMENTS

This study was supported partly by grants received from USDA and from an institutional grant from the NIGMS to SKD.

REFERENCES

1. Mao, J. Apple, B., Schnack, J., Sharp, S., Yamada, H., and Soll, D. (1982). Nuc. Acids Res. 10:487-495.
2. Dutta, S.K., Williams, N.P., and Mukhopadhyay, D.K. (1983). Mol. Gen. Genet. 189:207-210.
3. Verma, M., Ghosh, B., Chambers, C., Crouch, R.J., and Dutta, S.K. (1985). Curr. Sci. 54:1092-1097.
4. Chambers, C., Dutta, S.K., and Crouch, R.J. (1986). Gene 44:159-164.
5. Maniatis, T., Fritisch, E., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.
6. Dutta, S.K., Chambers, C., Ghosh, B., and Verma, M. (1986). The Nucleus 29:9-20.
7. Sambrook, J., Fritisch, E., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
8. Messing, J., Crea, R., and Coulson, A.R. (1981). Nuc. Acids Res. 9:309-321.
9. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 94:5463-5467.
10. Harshey, R.M., Estepa, G., and Yangi, H. (1989). Gene 79:1-8.
11. Garkavtsev, I.V., Tsvetkova, T.G., Yegolina, N.A., and Gudkov, A.V. (1988). Human Genetics 81:31-37.
12. Harrington, C.A., and Chikaraishi, D.M. (1983). Nuc. Acids. Res. 11:3317-3322.
13. Nandabalan, K., and Padayatty, J.D. (1988). Biochem. Biophys. Res. Commn. 157:68-74.
14. Tautz, D., and Dover, G.A. (1986). EMBO J. 5:1267-1273.
15. Takaiwa, F., Oono, K., and Sugiura, M. (1985). Plant Mol. Biol. 4:355-364.
16. Toloczyki, C., and Feix, G. (1986). Nuc. Acids Res. 14:4969-4985.
17. Apples, R., and Dvorak, J. (1982). Theor. Appl. Genet. 63:337-348.
18. Delcasso-Tremousaygue, D., Griller, F., Panabieres, F., Ananiev, E.D., and Delseny, M. (1988). Eur. J. Biochem. 172:767-776.
19. Yakuro, K., Kato, A., and Tanifuji, S. (1984). Mol. Gen. Genet. 193:400-405.
20. Ganal, M., Torres, R., and Hemleben, V. (1988). Mol. Gen. Genet. 212:548-554.
21. Botstein, D., White, R., Skolnick, M., and Davis, R.W. (1980). Am. J. Hum. Genet. 32:314-331.
22. Financsek, I., Mizumoto, K., Mishima, Y., and Muramatsu, M. (1982). Proc. Natl. Acad. Sci. USA 79:3092-3095.
23. Safrany, G., Kominami, R., Muramatsu, M., and Hidvegi, E.J. (1989). Gene 79:299-307.
24. Grummt, I., Maier, U., Ohrlein, A., Hassouna, N., and Bachellerie, J.P. (1985). Cell 43:801-810.
25. Miwa, T., Yoshikura, H., Sudo, K., Kominaimi, R., and Muramatsu, M. (1987). Nuc. Acids Res. 15:2043-2058.
26. Labhart, P., and Reeder, R.M. (1988). Cell 45:431-443.
27. Kuhn, A., Norman, A., Bartsch, I., and Grummt, I. (1988). EMBO J. 7:1497-1502.
28. Veldman, G., Klootwijk, J., Jonge, P., Leer, R.J., and Planta, R.J. (1980). Nucl. Acids Res. 8:5179-5192.
29. Dickson, K.R., Braaten, D.C., and Schessinger, D. (1989). Gene 84:197-200.
30. Sylvester, J.E., Peterson, R., and Schmickel, R.D. (1989). Gene 84:193-196.
31. Dutta, S.K., Frederick, L., and Austin, W.L. (1981). J. Elis. Mit. Sci. Soc. 97:126-135.
32. Dutta, S.K. (1976). Mycologia LXCIII:388-401.
33. Verma, M., and Dutta, S.K. (1987). Curr. Genet. 11:309-314.
34. Dutta, S.K., and Verma, M. (1990). Lower Eukaryotic Genome (Sobti, R.C., ed.). Panjub University Publication, India (In press).