MOLECULAR AND CELLULAR BIOLOGY, Aug. 1988, p. 3542–3545
0270-7306/88/083542-04$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 8, No. 8

NOTES

Structural and Functional Analysis of the Rat Malic Enzyme Gene Promoter

H. MORIOKA, G. E. TENNYSON, AND V. M. NIKODEM*

Clinical Endocrinology Branch, National Institute of Diabetes, Digestive, and Kidney Diseases, Bethesda, Maryland 20892

Received 28 October 1987/Accepted 30 March 1988

We have characterized sequences of genomic DNA 5′ to the coding region of the rat malic enzyme gene. This sequence possesses neither TATA nor CCAAT sequences in their usual positions but is rich in GC residues. Sequences similar to those found in the regulatory regions of other genes are discussed. Deletion analyses have revealed that sequences +1 to −41 are sufficient to initiate expression, although inclusion of information up to −177 is necessary for maximal promoter activity.

Rat liver malic enzyme (ME) (EC 1.1.1.40) plays an important role in lipogenesis. We have previously reported cloning (22) and the sequence of ME mRNA (21) and shown its regulation by thyroid hormone and a high carbohydrate diet (4–6). To investigate regulation of the ME gene, we have identified and sequenced its 5′-flanking region.

The sequence shown in Fig. 1 extends from the *Eco*RI site through the transcription start site and includes 109 base pairs (bp) of the first exon with a portion of the first intron. The structural organization of the ME promoter differs from that found in tissue-specific promoters, which are usually rapidly regulated and often switched off, but it resembles more closely promoter regions of several eucaryotic constitutive or "housekeeping" genes (23–28, 30, 31), the simian virus 40 late promoter (9), and some recently characterized proto-oncogenes (16, 17). A TATA box (2), usually located 20 to 30 bp upstream from cap sites, lies at −622 (Fig. 1). The sequence CCGAT, between −144 and −140, resembles the canonical CCAAT consensus sequence often found at a position 80 bp from transcription start sites in many eucaryotic promoters (7). The most striking motif is the nine hexanucleotides CCGCCC. Six of these GC boxes are located upstream from the major cap site from −376 to −10. One GC box is present in an untranslated region, while the last two GC boxes are found within the first intron. Noteworthy is the number of nucleotides, (i.e., 65, 63, and 61) separating the GC motifs at −46, −111, −174, and −235, respectively. This rather equal spacing, also seen in the 5′-flanking region of the adenosine deaminase gene (31), might play a role in the chromatin assembly required for expression. The 6-bp sequence of the GC motif is the same sequence that is repeated six times within the simian virus 40 promoter (9) and has been shown to be a core element in the decanucleotide sequence, exhibiting high affinity binding for Sp1 (1, 8). Judging from the adjacent sequences, the GC box at −174 and the first one in the intron are the most likely targets for strong interactions with Sp1.

A 10-bp direct repeat of the sequence 5′-CTCGCCACCC-3′ was noted at positions −73 and −51, respectively. A 26-bp sequence from −110 to −84 is similar (14 of 26 nucleotides with one deletion) to the consensus sequence of the 3′ region of a 48-bp repeated element in the mouse dihydrofolate reductase gene (25).

Farther downstream from −124 to −119 is the hexanucleotide sequence 5′-CGCTTC-3′. Three copies of this sequence occur within the boundaries of the Moloney murine sarcoma virus long terminal repeat distal transcription signal (11), immediately downstream from the CCAAT box. The position of this element relative to a CAT-like box (−144) found in the ME gene 5′-flanking region imitates the position identified in the viral long terminal repeat. We also note that a purine-rich region at the position from −250 to −241 is closely homologous (9 of 10) to enhancerlike sequences present in the adenovirus 5 E1a −200 region (15).

To determine the location of functionally essential sequences required for maximal promoter activity, we utilized a 5′ deletional analysis, using the restriction sites *Nar*I, *Hae*III, *Sca*I, and *Pvu*II or exonuclease *Bal* 31. These constructs, containing various fragments of the putative ME promoter in sequence with the coding region of the bacterial gene for chloramphenicol acetyltransferase (CAT), were tested in transient transfection assays for their ability to express CAT activity using Chinese hamster ovary (CHO), mouse hepatoma (Hepa I) and human cervical carcinoma (HeLa) cells. Recombinant plasmid DNA (5 or 15 μg) (20) and pCH110 (3.75 μg) (13) (Pharmacia) were added to a 10 cm dish containing 1×10^6 to 2×10^6 cells. Each experiment was repeated three to five times, and for each, the results were done in triplicate. The CAT enzyme and β-galactosidase assays (19) were performed 40 h later. The CAT assay was run for 3 h, with the addition of 3 μl of 4 mM acetyl coenzyme A at hourly intervals to maintain linearity of CAT enzyme activity and 5 μl (≈50 mCi/mmol) of ^{14}C-labeled chloramphenicol. The β-galactosidase assay was run for 20 to 90 min.

Figure 2 shows the endpoints of the 5′ deletion mutants and their relative promoter activities. The deletional mutant pME 22 CAT with the 5′ endpoint at −177 was most effective in transcriptional activation and was given a value of 100. A 20% reduction in transcriptional efficiency (pME B 42 CAT) was observed when 32 nucleotides were removed from the 5′ end of pME 22 CAT, suggesting that the deleted

* Corresponding author.

GAATTCGCAT AGCCCAGAAG CTATAGCTGT ACTGATGGGC TCAAGTAAAA TAATTAGAAA TTATTTCTCA GGTATCTAGG CAATATTTAA CCCCCAAATT −783

GTTCCGCAGT GTCTAGATGA ACACCATAGA ATTTGGCCGT GCGACTTAAC TGAAAAGAAA GGGCTTTGTT GTCTGAAGGC TGCTTGGCTG TATTGTTTTG −683

TTTTAATCAG ACATCCTTGG GAGACATAGG ATTTATTTCT CCAGTCCTTG GATCTTCAAG TATAAATATC AATAATACAA CCACTGGGTT TCAGTACTGG −583

AAGACCTGTT ATTCTGACCC TCTGTCATCA GAGAAGAAAC CATACATCAT CTTGCAAAAA TTAACATCTT GGTTTCCAGA ACGCTCAGGA AAATTGTTCT −483

TAAGCTCAAT AGGACTGGCC ACTGGACCTG TGCCCTCTAA CACCTTTTTC TTACCACGTT CGAACACAAT TCCCTCAGAT ACTATTCAGA AACAGGCGAG −383

GAGTCCCCGC CCTATCGCCC AGTGCCATCG AGGCCTGGGC ATTCTGGGTC AAAGTTGATC CCCTCCTGCA TCAGGCCCCT GGGGCATGGC TGGCATCCAG −283

GACGTTGGGG TTAGGGAGGA CATGGACGAG CGGAGGAAGC GAGCGGCCCG CCCCTCACCC GTCGGTGCCC AGGTCGCACG CTCGGCGCTC ACCAGCTTGG −183

CCGGCGCCCC GCCCCCGCCT CCTCGCACGG CGGCTCGGCC GATGCCGCCG TGACTCAGCG CTTCTCGCGG CCCGCCCGCG CGGCCGCGGC TAGGCCGGGC −83

TCCTCCCGCC TCGCCACCCC CTCTCGCCAC CCACGCCCGC CCCCGGCCGC GGGGCCTTCC GTCGCACGGC CGCCGCCCGC CGCACTCCCG TCCGCCCCGC

+10

CACGGTGCTG GCCATGGATC CCCGAGCCCC CCGCCGCCGA CACACCCACC AGCGCGGCTA CCTGCTGACG CGGGACCCGC ATCTCAACAA GGTGAGCCCC

GCCCCGAGAG CCGCCCTGGG CCCGCCGCTG GGCTCGGGCA CCCGCGTCCC ACCGAGGGGA CGGTCCCACC CGGGAGGCCA CTGCGGAGCC GGCGCCAACG

FIG. 1. Nucleotide sequence of the ME 5′-flanking region, including the first exon and adjacent intronic sequences. The sequence was obtained from the genomic subclone λg ME 29, isolated in our laboratory. Restriction fragments were subcloned into M 13 vectors using *Escherichia coli* JM103 as the host and were sequenced in both directions using the dideoxy-chain termination method (29). The major transcription start site is designated by +1. The TATA box at −622 and ATG at +32 are underlined by solid lines, the repeated sequences CCGCCC are enclosed by boxes, and direct repeats are denoted by horizontal arrows. Nucleotide sequences similar to those of the mouse dihydrofolate reductase promoter region (− • − •), Moloney murine sarcoma virus long terminal repeat (• • • •), and adenovirus 5 Ela enhancer (----) are indicated, as are the CCGAT sequence (underline), the exon-intron junction (|), and minor sites determined by S1 nuclease mapping (unpublished data) (○).

nucleotide sequence containing GCCCCGCCCC, a potential high-affinity binding site for Spl (18), contributes little to maximal promoter activity. It has been suggested that two possible Spl-binding sites found in the promoter region of the mouse hypoxanthine phosphoribosyltransferase gene (HPRT) are not essential for transcription; however, some reduction in activation was observed when the sites were deleted (28). The most striking decrease in transcriptional activation in the ME promoter resulted from removal of an additional 22 nucleotides from the deletional mutant with the −145 endpoint. This reduction in expression was more profound when this plasmid was transfected into CHO cells.

A 5′ DELETION MUTANTS

Eco RI, Sca I, Hae III, Nae I, Bam HI, Hind III

5′ −882 −500 −400 −200 +1 CAT 3′

5′ endpoint	Construct
−882	8
−586	18
−348	29
−243	B 40
−177	22
−145	B 42
−122	EPS 25
−113	B 30
−101	EPS 26
−41	EPS 28

B PROMOTER ACTIVITY (% of pME 22 CAT)

20 40 60 80 100

FIG. 2. Structure and activity of ME promoter-CAT chimeric gene (A) Organization of deletion mutants. The *Eco*RI-*Bam*HI DNA fragment used to construct the 5′ deletion mutants is shown at the top. The 3′ end contains the pUC18 polylinker (*Bam*HI-*Hin*dIII) followed by the structural portion of the CAT gene obtained from pSVO CAT. Solid lines illustrate the 5′ endpoints of the ME sequences, determined by sequencing with a reverse primer (14). The nomenclature of the recombinants has been abbreviated such that the designation pME CAT was omitted. (B) Relative promoter activities of the ME CAT chimeric plasmids. Transfections with pME CAT recombinant plasmids and pCH110 plasmid were performed as described elsewhere (10) with Hepa I (■), CHO (□), and HeLa (▨) cells. The percentage of conversion of chloramphenicol was then divided by the A_{420} of the corresponding β-galactosidase assay, giving a value normalized for transfection efficiency.

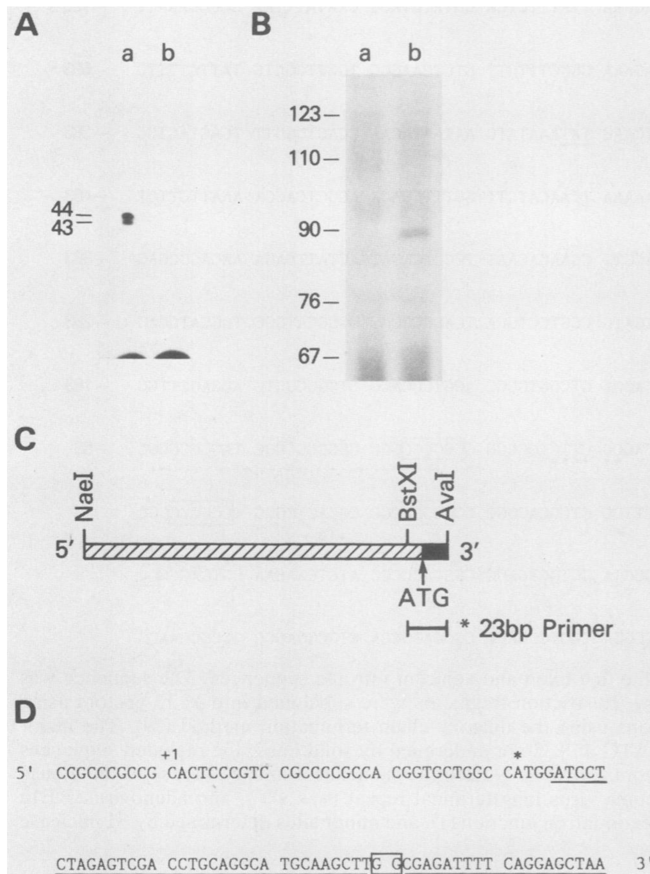

FIG. 3. Primer extension analysis of transcripts from rat liver (A and C) and Hepa I cells (B and D). (A) Rat liver ME transcript. The primer used was derived from a genomic subclone after digestion with *Eco*RI and *Ava*I and was then dephosphorylated and end labeled with polynucleotide kinase. Digestion with *Bst*X1 yielded the 5'-end-labeled 23-bp fragment. The *Bst*X1 and *Nae*I sites are located upstream from the translation initiation codon (ATG), while *Ava*I is present in the coding region (▬). The asterisk indicates the 5' end label of the noncoding strand (C). For a primer extension assay, 2×10^5 cpm of *Bst*X1-*Ava*I fragment was incubated with 25 μg of poly(A^+) RNA (3) from liver of thyroid hormone-treated rats (lane a) or yeast tRNA (lane b) and with 5 U of reverse transcriptase (27). Numbers indicate the sizes of the probe and extended products, determined from a sequencing reaction of the cDNA fragments. (B) Hepa I ME transcript. The primer was a synthetic oligonucleotide (5'-TTAGCTCCTGAAA-3') complementary to 19 nucleotides present in the most 5' end of the CAT sequence (D, heavy underline). The pUC18 linker sequence is indicated by a thin underline, and the cap site of rat liver mRNA is designated +1, representing the 5' end of the elongated product as determined on rat liver. The A in ATG is denoted by the asterisk. The presence of GG residues (boxed) is unexplained. For a primer extension assay, 100 μg of total RNA from control Hepa I cells and cells transfected with pME B 40 CAT (−243) was hybridized to the 5'-end-labeled primer (B, lanes a and b). Numbers indicate the molecular sizes of ^{32}P-end-labeled pBR322 DNA digested with *Msp*I.

The sequence residing within this 22-nucleotide segment harbors the pentanucleotide CCGAT, which somewhat resembles CCAAT. The mutation of the third base in this sequence from A to G could represent a down mutation and could thereby maintain a low level of expression observed for the ME gene in vivo. Indeed, specific point mutations within or adjacent to the CCAAT box modulated the expression efficiency of both the Moloney murine sarcoma virus long terminal repeat and the herpes simplex virus thymidine kinase genes (12).

Internal repeats have been identified in several housekeeping genes. Inclusion of the direct repeats present in the ME 5'-flanking region at positions −73 and −51 in the deletion mutant pME EPS 26 (−101) increased the efficiency of expression by only 20%. The 8-bp direct repeat was found in the promoter region of the human 3-phosphoglycerate kinase gene (30). The human adenosine deaminase promoter region contained three inverted repeats (31), and the mouse dihydrofolate reductase gene contained three copies of the 48-bp repeat (25). More extensive repeats were located in the promoter region of the mouse HPRT gene (27). Deletional analysis of the mouse dihydrofolate reductase promoter region showed that at least one copy of the repeat is necessary for expression (25). Similar analysis of the mouse HPRT promoter region suggests that removal of the first copy of the 18- and 12-bp repeats has no effect on expression, while disruption of the second 12-bp repeat reduces transcription (27). To determine whether transcripts from pME CAT chimeric genes initiated at the same position as those from rat liver, RNA from liver and transfected Hepa I cells and CHO cells (data not shown) was used in primer extension analysis. Results (Fig. 3) indicate that the ME transcripts of the pME B40 CAT initiated at the same sites as those detected for the rat liver transcript.

Housekeeping gene promoters are complex and distinct from tissue-specific promoters. While a dramatic increase in transcription rate may occur with the addition of a relatively short sequence in the 5' region of tissue-specific promoters, only gradual increases are usually observed with the addition of a 5' sequence of housekeeping genes. The ME gene expression clearly demonstrates this effect. Moderate differences in the level of gene expression may therefore be representative of housekeeping promoter systems, such as the threefold change in expression observed within 200 bp of the HPRT promoter (27). We found a similar magnitude of alteration in promoter activity in the ME system when we compared deletions from −177 to −41.

A repeatable reduction of approximately 40% in the expression of CAT activity was detected with deletion mutants pME CAT 29, 18, or 8 when each was transfected into CHO cells (Fig. 3). We did not pursue this observation further. However, a similar finding was reported for the mouse HPRT minigene when promoter mutants were transfected into RJ k 88, the HPRT-deficient Chinese hamster line (27).

In conclusion, it appears that the ME promoter encompasses at least three domains. One domain contains sequences up to −41 and is required for the initiation of transcription. The second element is rather diffuse, with its 5' boundary at position −122, containing sequences similar to those found in promoters of other genes. The precise role of these domains in the ME promoter function is not understood at this time. The third region, from −122 to −145/−177, is required for maximal promoter activity, since deletion mutants containing more 5' nucleotides did not change the level of CAT expression when they were transfected into HepaI or HeLa cells.

We are grateful to J. E. Rall for valuable discussions and encouragement throughout this work. We also thank D. P. Nikodem for preparations of plasmid DNA and Valerie Robertson for preparation of the manuscript.

LITERATURE CITED

1. **Baty, D., H. A. Barrera-Saldana, R. D. Everett, M. Vigneron, and P. Chambon.** 1984. Mutational dissection of the 21 bp repeat region of the SV40 early promoter reveals that it contains overlapping elements of the early-early and late-early promoters. Nucleic Acids Res. **12:**915–932.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:**349–383.
3. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18:**5294–5299.
4. **Dozin, B., M. A. Magnuson, and V. M. Nikodem.** 1985. Tissue-specific regulation of two functional malic enzyme mRNAs by triiodothyronine. Biochemistry **24:**5581–5586.
5. **Dozin, B., M. A. Magnuson, and V. M. Nikodem.** 1986. Thyroid hormone regulation of malic enzyme synthesis. Dual tissue-specific control. J. Biol. Chem. **261:**10290–10292.
6. **Dozin, B., J. E. Rall, and V. M. Nikodem.** 1986. Tissue-specific control of rat malic enzyme activity and messenger RNA levels by a high carbohydrate diet. Proc. Natl. Acad. Sci. USA **83:** 4705–4709.
7. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechi, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human beta-globulin gene family. Cell **21:**653–668.
8. **Everett, R. D., D. Baty, and P. Chambon.** 1983. The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucleic Acids Res. **11:** 2447–2464.
9. **Fromm, M., and P. Berg.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function in vivo. J. Mol. Appl. Genet. **1:**457–481.
10. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2:**1044–1051.
11. **Graves, B. J., R. N. Eisenman, and S. L. McKnight.** 1985. Delineation of transcriptional control signals within the Moloney murine sarcoma virus long terminal repeat. Mol. Cell. Biol. **5:**1948–1958.
12. **Graves, B. J., P. F. Johnson, and S. L. McKnight.** 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell **44:**565–576.
13. **Hall, C. V., P. E. Jacob, G. M. Ringold, and F. Lee.** 1983. Expression and regulation of Escherichia coli lacZ gene fusions in mammalian cells. J. Mol. Appl. Genet. **2:**101–109.
14. **Hattori, M., and Y. Sakaki.** 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152:**232–238.
15. **Hearing, P., and T. Shenk.** 1983. Functional analysis of the nucleotide sequence surrounding the cap site for adenovirus type 5 region E1A messenger RNAs. J. Mol. Biol. **167:**809–822.
16. **Ishii, S., G. T. Merlino, and I. Pastan.** 1985. Promoter region of the human Harvey ras proto-oncogene: similarity to the EGF receptor protooncogene promoter. Science **230:**1378–1381.
17. **Ishii, S., Y. H. Xu, R. H. Stratton, B. A. Roe, G. T. Merlino, and I. Pastan.** 1985. Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. USA **82:**4920–4924.
18. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. Biochem. Sci. **11:**20–23.
19. **Kimura, A., A. Israel, O. LeBail, and P. Kourilsky.** 1986. Detailed analysis of the mouse H-2Kb promoter: enhancer-like sequences and their role in the regulation of class I gene expression. Cell **44:**261–272.
20. **Labarca, C., and K. Paigen.** 1980. A simple, rapid, and sensitive DNA assay procedure. Anal. Biochem. **102:**344–352.
21. **Magnuson, M. A., H. Morioka, M. F. Tecce, and V. M. Nikodem.** 1986. Coding nucleotide sequence of rat liver malic enzyme mRNA. J. Biol. Chem. **261:**1183–1186.
22. **Magnuson, M. A., and V. M. Nikodem.** 1983. Molecular cloning of a cDNA sequence for rat malic enzyme. Direct evidence for induction in vivo of rat liver malic enzyme mRNA by thyroid hormone. J. Biol. Chem. **258:**12712–12717.
23. **Maguire, D. J., A. R. Day, I. A. Borthwick, G. Srivastava, P. L. Wigley, B. K. May, and W. H. Elliott.** 1986. Nucleotide sequence of the chicken 5-aminolevulinate synthase gene. Nucleic Acids Res. **14:**1379–1391.
24. **Martini, G., D. Toniolo, T. Vulliamy, L. Luzzatto, R. Dono, G. Viglietto, G. Paonessa, M. D'Urso, and M. G. Persico.** 1986. Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. EMBO J. **5:**1849–1855.
25. **McGrogan, M., C. C. Simonsen, D. T. Smouse, P. J. Farnham, and R. T. Schimke.** 1985. Heterogeneity at the 5′ termini of mouse dihydrofolate reductase mRNAs. Evidence for multiple promoter regions. J. Biol. Chem. **260:**2307–2314.
26. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217:** 316–324.
27. **Melton, D. W., C. McEwan, A. B. McKie, and A. M. Reid.** 1986. Expression of the mouse HPRT gene: deletional analysis of the promoter region of an X-chromosome linked housekeeping gene. Cell **44:**319–328.
28. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38:**275–285.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
30. **Singer-Sam, J., D. H. Keith, K. Tani, R. L. Simmer, L. Shively, S. Lindsay, A. Yoshida, and A. D. Riggs.** 1984. Sequence of the promoter region of the gene for human X-linked 3-phosphoglycerate kinase. Gene **32:**409–417.
31. **Valerio, D., M. G. Duyvesteyn, B. M. Dekker, G. Weeda, T. M. Berkvens, L. van der Voorn, H. van Ormondt, and A. J. van der Eb.** 1985. Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. **4:**437–443.