The bidirectional promoter of two genes for the mitochondrial translational apparatus in mouse is regulated by an array of CCAAT boxes interacting with the transcription factor NF-Y

Ernesto Zanotto¹, Zahid H. Shah¹ and Howard T. Jacobs¹,²,*

¹Institute of Medical Technology & Tampere University Hospital, FI-33014 University of Tampere, Finland and ²Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, UK

Received September 12, 2006; Revised and Accepted November 7, 2006

ABSTRACT

The genes for mitoribosomal protein S12 (Mrps12) and mitochondrial seryl-tRNA ligase (Sarsm and Sars2) are oppositely transcribed from a conserved promoter region of <200 bp in both human and mouse. Using a dual reporter vector we identified an array of 4 CCAAT box elements required for efficient transcription of the two genes in cultured mouse 3T3 cells, and for enforcing directionality in favour of Mrps12. Electrophoretic mobility shift assay (EMSA) and in vivo footprinting confirmed the importance of these promoter elements as sites of protein-binding, and EMSA supershift and chromatin immunoprecipitation (ChIP) assays identified NF-Y as the key transcription factor involved, revealing a common pattern of protein–DNA interactions in all tissues tested (liver, brain, heart, kidney and 3T3 cells). The inherently bidirectional activity of NF-Y makes it an especially suitable factor to govern promoters of this class, whose expression is linked to cell proliferation.

INTRODUCTION

The machinery of mitochondrial protein synthesis is partly encoded by the nuclear genome and partly by mitochondrial DNA (mtDNA). In vertebrates, the contribution of the latter is exclusively the ribosomal and transfer RNAs. All mitoribosomal proteins, aminoacyl-tRNA ligases, translation factors and chaperones are nuclear coded [for review see Ref. (1)]. In mammals, two genes for components of the mitochondrial translational apparatus, encoding mitoribosomal protein S12 (Mrps12) and the mitochondriallocally localized isoform of seryl-tRNA ligase (Sarsm or Sars2), are encoded by oppositely transcribed genes that share a common, bidirectional promoter region of <500 bp (2, 3). The genes are located on chromosome 19 in human and chromosome 7 in mouse (2, 4).

Bidirectional promoters are a common feature of mammalian genomes (5,6), and microarray studies indicate that the majority of such gene pairs are co-regulated. However, relatively few studies have addressed the mechanisms of transcriptional regulation that govern such promoters, and ensure that both genes are efficiently and stably transcribed. In the case of human MRPS12 mRNA, previous studies by ourselves (7,8) and others (9) indicated that three classes of transcript are produced by alternate splicing within the 5′-untranslated region (5′-UTR). One splice variant is susceptible to translational regulation in response to cellular growth status (8), even though it lacks the 5′-terminal oligopyrimidine tract characteristic of cytoplasmic ribosomal protein mRNAs that are regulated in this manner. However, the different splice variants of MRPS12 mRNA appear to be expressed at similar relative levels in all tissues analysed, suggesting that they are not selectively transcribed. The SARS mRNA has not been studied in detail, but it appears to comprise a single species with a very short 5′-UTR (27 nt). The two genes show a very similar organization in mouse, although one Mrps12 splice variant appears to be absent, and the Mrps12 first intron is slightly shorter.

Elucidating the ways in which genes of the apparatus of mitochondrial protein synthesis are regulated also has a biomedical dimension. Mutations affecting the mitochondrial translational apparatus, including the seryl tRNAs (10) and at least one mitoribosomal protein, MRPS16 (11) have been implicated in disease, and mutations in the Drosophila homologue of MRPS12 also produce a mitochondrial-disease-like phenotype (12).

In mammals, nuclear genes for components of the mitochondrial oxidative phosphorylation (OXPHOS) system are regulated by a variety of transcription factors, most notably by nuclear respiratory factors NRF-1 and NRF-2, which have single or multiple binding sites in the promoters of many nuclear genes for subunits of the OXPHOS
complexes or components of the apparatus of mitochondrial gene expression [for review see Ref. (13)]. Possible binding sites for NRF-1 and NRF-2 were noted in the SARS/MRPS12 promoter region by Johnson et al. (9), although the former lie within the SARS coding sequence and first intron. NRF-2 and especially NRF-1 also regulate a large number of other genes which need to be highly expressed in proliferating cells. In yeast there are no known homologues of NRF-1 or NRF-2, but the CCAAT box-binding transcription factor Hap2/3/5 appears to be the main activator of transcription of nuclear genes involved in mitochondrial biogenesis and OXPHOS (14–16).

In order to establish the functional elements involved in transcription from the bidirectional promoter of MRPS12 and SARS we decided initially to focus strictly on the intergenic region and 5′-UTR segments, i.e. up to the start codon of each gene, and to study the mouse rather than the human promoter, with a view to using the knowledge later to manipulate expression in the whole organism. We initially aligned and compared the promoter region from mouse and human, in order to identify conserved blocks of sequence. We then cloned the mouse sequence into a dual luciferase reporter and carried out a deletion analysis to localize the main regulatory elements. This was followed up by electrophoretic mobility shift assay (EMSA), in vivo footprinting and chromatin immunoprecipitation (ChIP) assays, to test for evidence of protein-binding within the putative regulatory elements, and by further reporter assays in which such sites were mutated. These studies revealed the importance of an array of four CCAAT box elements interacting with the transcription factor NF-Y, the mammalian homologue of yeast Hap2/3/5. The evidence for involvement of NRF-2 and other factors was less clear, and we propose that these act most likely as accessory factors to boost transcription.

NF-Y exhibits an inherently bidirectional activity, capable of recognizing its core binding site sequence in either orientation, i.e. as CCAAT or as ATTGG on the coding strand. We propose that this makes it an especially suitable factor to govern bidirectional promoters of this class, especially when their expression is linked, like that of many genes involved in mitochondrial biogenesis, to cell proliferation.

**MATERIALS AND METHODS**

**Cells and cell culture**

NIH 3T3 cells were cultured in DMEM (Cambrex) supplemented with 10% FBS, 50 U/ml penicillin (Cambrex) and 50 μg/ml streptomycin (Cambrex). All cells were maintained at 37°C in 5% CO₂.

**RNA extraction**

RNA was extracted from cells at 80% confluence using TRIzol reagent (Invitrogen) according to the manufacturer’s recommended conditions, with resuspension in 30 μl nuclease-free water per 10 cm plate of cells. Treatment with 15 U RNase-free DNase I (Amersham Biosciences) was carried out for 1 h at 37°C in a final volume of 50 μl of manufacturer’s recommended buffer. RNA was recovered by phenol–chloroform extraction, ethanol precipitation and final resuspension in 30 μl nuclease-free water, and its purity and integrity were checked electrophoretically.

**Reporter constructs**

A dual luciferase reporter vector was constructed by recloning the entire coding sequence of firefly luciferase, including the SV-40 late poly(A) region, from the vector pGL3-basic (Promega) into the vector phRL-null (Promega), as an NcoI–SalI fragment. The resulting vector (pFRL) contains firefly and Renilla luciferase reporter genes oriented such that transcription is in opposite directions, with a unique NcoI site (CCATGG) between them, which provides the in-frame start codon in both directions. The Mrps12-Sarsm intergenic region was amplified with the following chimeric primers, using a mouse genomic DNA clone as template (NcoI restriction sites in bold, non-templated nucleotides underlined, all sequences 5′–3′), Mrps12: CAGTGCATG-GCTGGCGCTGTCAGGCTCC, Sarsm_F: GCTGGCAGG-CTTTGGAGGTGAAAGAAGTGCA. The PCR product was digested with NcoI and then cloned into NcoI-cut pFRL, generating clones in both orientations. To create a deletion series from the wild-type template, either of the same terminal primers was used together with the deletion primers indicated in Supplementary Figure 1b (Mrps12 with the primers designated Sarsm_F7, Sarsm_F2 etc. to create deletions S7, S2 etc. Sarsm_F with primers Mrps_PF1 and Mrps_PF2 to create deletions M1 and M2, respectively), with each product cloned into the NcoI site of pFRL using a similar strategy as for the wild-type promoter. To create point mutants of the dual reporter construct, a two-round mutagenic PCR strategy was employed, except for those cases where the mutations to be introduced were within 40 bp of the end of the intergenic region, and could thus be created in a single mutagenic step using modified terminal primers. Full details of primers are available in Supplementary Table 1. In each case the final PCR product, following digestion with NcoI, was cloned into the NcoI site of pFRL. Every construct was verified by sequencing in both directions. A β-galactosidase reporter construct (17) was kindly provided by Dr H. Renkema.

**Transfections and reporter assays**

Cells grown to 80% confluence were transfected in 6-well plates using TransFectin Lipid Reagent (Bio-Rad) according to the manufacturer’s instructions, with 0.25 μg of DNA from each luciferase construct together with 0.25 μg of DNA of the β-galactosidase control vector. After 24 h firefly and Renilla luciferase activities were assayed in cell lysates using the Dual Luciferase Reporter Assay System (Promega) and a BioORBIT 1254 luminova luminometer. β-galactosidase activity was determined in 50 μl aliquots of the same lysates by the addition of 50 μl ONPG solution (2-nitrophenyl -D-galactopyranoside) and 10 μl of 500 mM NaCl, 100 mM MgCl₂, 100 mM β-mercaptoethanol, with incubation at 37°C for 1 h and measurement of A₄₂₀. Luciferase activities were normalized for transfection efficiency using β-galactosidase activity. See legend to Figure 2 for further details.
q-PCR

First strand cDNA was synthesized from 5 μg cellular RNA using M-MuLV reverse Transcriptase (Fermentas) under manufacturer's recommended conditions, using 0.5 μg oligo(dT) primer. After treatment with boiled RNase (Roche) the cDNA was used in PCR in a 5-fold dilution series in the presence of SYBR Green (Qiagen) for fluorescent quantitation of the product, using the following gene-specific primers (all sequences shown as 5'→3'). For Sarsm, SARS M 8: CTTTCCGGGACCTCCAGTCA and SARS 4: CGGTCAAGATCTCACCAGTC; for Hprt, mHPRT_R1: CACAGGACTGAAACCTGC and mHPRT_F1: GTTGGATACAGGCGGAGATCTTGTT; for Mrps12, Mrps55: TTCCATGGCCACCTAGACCCAG and Mrps32: CAGCACCCTGCCGTTGGCGGAG. Cycle parameters were: denaturation at 94°C for 15 s (initial denaturation at 95°C for 10 min), annealing for 10 s at 57°C (for Hprt) or 55°C for Sarsm and Mrps12, extension at 72°C for 20 s, 55 cycles (Light cycler, Roche).

EMSA

Nuclear extracts were prepared from NIH 3T3 cells essentially as described by Wadman et al. (18), except that KCl was substituted for NaCl in Buffer C, and Pepstatin was omitted. Nuclear extracts were prepared from mouse liver as described by Sun et al. (19). EMSA probes were synthesized by PCR, using as template the wild-type or mutated reporter constructs of the Mrps12-Sarsm intergenic region, plus primers as indicated in Supplementary Table 2. In each case the identity of the product was verified by sequencing before proceeding. After dephosphorylation with calf intestinal alkaline phosphatase (Fermentas) under manufacturer's recommended conditions, 0.6 ng of probe DNA was labelled using 8 U of T4 Polynucleotide Kinase (Fermentas) under manufacture before proceeding. After dephosphorylation with calf intestinal alkaline phosphatase (Fermentas) under manufac-

ChIP assays

NIH 3T3 cells were grown to 80–90% confluence in 10 cm culture plates. After cross-linking for 10 min with 1% formaldehyde in serum-free medium, phosphate-glycine buffer was added to a final concentration of 0.125 M and cells were washed twice with ice-cold phosphate-buffered saline (PBS). Nuclei were isolated as for EMSA and lysed essentially as described by Spencer et al. (26), except that an SDS concentration of 0.1% instead of 1% was used. The chromatin lysate was sonicated on ice to an average DNA length of 600 bp. Chromatin was pre-cleared with blocked Sepharose A and ChIP assays were performed as described by Spencer et al. (26), using 8 μg of either the anti-NF-Y mouse monoclonal antibody described above or, as negative control, M2 anti-FLAG® mouse monoclonal (Stratagene). The final PCR step used primers GAGGTCGACCAACAAGAATCTGTTCACTCAT and CTGAGTACGCCACGGAGGACC (both shown 5'→3') and the fragment spanning np 4→349 of the sequence shown in Supplementary Figure 1b. Following an initial 5 min denaturation at 95°C, PCR involved 32 cycles of the following: denaturation for 1 min at 94°C, annealing for 30 s at 59°C, extension for 30 s at 72°C. Reaction products were analysed on a 1.5% agarose-TBE gel stained with ethidium bromide and visualized under ultraviolet (UV) light.

RESULTS

Organization of the mouse Sarsm-Mrps12 intergenic region

The intergenic region between the mouse genes Sarsm and Mrps12 was compared with the homologous region of the human genome. The predominant 5' ends of transcripts of the two genes were also compared between the species, using publicly available EST data. As summarized in Figure 1 (see also Supplementary Figure 1), the overall organization of the region is highly preserved between the
two species. As in human, the 5'-UTR of Sarsm in mouse is extremely short (26 nt) and largely identical with its human counterpart. From ESTs there is no support in either species for a longer 5'-UTR, as reported for human SARS by Yokogawa et al. (2), NCBI database entry NM_017827, reconstructed from multiple EST clones and in-house RT–PCR. We thus take the latter to be a very minor species. The non-transcribed portion of the intergenic sequence is 184 bp in mouse (7 bp longer in human). The first part of this sequence (i.e. nearest to Sarsm) is dissimilar between the two species, except for the presence in both cases of two copies of the CCAAT box sequence, with slightly different spacing. This is followed by a 147 bp sequence that is 69% identical between human and mouse, which extends to the region of the ‘downstream start site’ for Mrps12 (see below). This region contains two additional CCAAT box elements which are identically spaced in the two organisms, as well as the NRF-2 binding site identified previously by Yokogawa et al. (2), which is conserved in mouse.

Two clusters of start sites for Mrps12 are found in virtually identical positions in the two organisms. The first overlaps (human) or immediately follows (mouse) the putative NRF-2-binding site, and gives rise to transcripts which use a splice donor ~40 nt downstream, removing a 211 nt intron [277 nt in humans, but also with an alternative upstream splice acceptor, see Ref. (8)]. The intron is less similar in the two species apart from a 60 bp segment located towards the 3’ side in mouse, but more centrally placed in human, which shows 77% identity, plus the terminal polypyrrimidine tract and terminal splice acceptor region. A second cluster of Mrps12 transcriptional start sites coincides approximately with the start of the intron, which is thus preserved in these mRNAs as a longer 5'-UTR. Previous data (8) indicated that the shorter (spliced) mRNAs are regulated translationally in response to growth signals in human cells. Overall, the sequence comparison reveals a very similar organization of the bidirectional promoter region of these two genes in mouse and human genomic DNA, and a high degree of sequence conservation of the region lying between their major transcriptional start sites.

Relative expression levels of Sarsm and Mrps12
Quantitative RT–PCR (q-PCR) was used to estimate the steady-state levels of Sarsm and Mrps12 mRNAs in mouse 3T3 cells, relative to that of Hprt as a standard. Based on parallel reactions at different dilutions, the level of Mrps12 mRNA was estimated as 96 + 11% that of Hprt mRNA, whereas Sarsm mRNA was 78 + 16% as abundant as Hprt mRNA. The steady-state levels of the two mRNAs are thus very similar.

Reporter analysis of the Sarsm-Mrps12 intergenic region
In order to test whether the intergenic region contains elements essential for transcription in the two directions, we constructed a dual reporter vector (see Materials and Methods) in which the transcriptional activities in the two opposite directions could be tested simultaneously via the readouts of firefly and Renilla luciferase. We initially inserted the full intergenic region (all 480 bp lying between the two start codons) into the reporter vector in both orientations, in order to allow us to determine the relative activities of the promoter in each direction. As indicated in Figure 2, as a result of comparing the expression of each reporter in the two oppositely oriented constructs, transcriptional activity
in the Mrps12 direction was inferred to be ~3.5-fold greater than in the Sarsm direction, when tested by transfection into exponentially growing 3T3 cells. A series of deletion constructs was then made, in order to test which portions of the intergenic region were important for expression globally or directionally. The first two constructs tested (M1 and S7, see Figure 2) involved very short deletions of the regions upstream of the two start codons, which were initially intended to filter out any effects due to loss of translation-regulatory sequences. Although neither of these deletions produced dramatic effects, gene expression in both directions was affected, suggesting that transcriptional regulatory elements encompass the entire region.

Deleting further upstream from the Sarsm start codon produced a complex series of effects. Removal of the Sarsm mRNA start site and CCAAT box I resulted in a drop in expression in the Mrps12 direction and an increase in the Sarsm direction, but Mrps12 was affected only mildly (Figure 2, constructs S7, S2 and S3) until the boundary of CCAAT box IV was reached (Figure 2, construct S4). Removal of both of the major transcriptional start sites (for Sarsm plus the upstream start for Mrps12), as well as the entire region between them containing all four CCAAT boxes, almost completely abolished expression (Figure 2, construct S5). The construct (S4) in which the putative NRF-2 binding site was preserved was only slightly more transcriptionally active than the one in which it was removed (S5), with almost no transcription in the Sarsm direction in either case, and only 18% of the wild-type level of expression in the Mrps12 direction for construct S4, compared with 6% for S5. However, this does not exclude the possibility that the NRF-2 site contributes to transcriptional activity when the CCAAT boxes or other upstream elements are still present.

Surprisingly, deletion of the region of the splice acceptor for the optional first intron of Mrps12 mRNA led to a significant increase in expression in both directions, indicating the presence of a negative regulatory element. Deletion analysis thus identified two kinds of regulatory element in the bidirectional promoter region: one or more (positively acting)
elements in the region between the two transcriptional start sites, which also influence the relative directionality of transcription, as well as a negatively acting element within the first intron of \textit{Mrps12}.

Analysis of protein-binding sites by EMSA

Reporter studies having indicated the presence of various \textit{cis}-acting elements in the \textit{Sarsm-Mrps12} intergenic region, we then investigated protein-binding to such sites using EMSA. Overlapping probes were synthesized and reacted with 3T3 cell nuclear protein extracts in the presence of each of two different polynucleotide competitors. Clear signatures of specific protein-binding were inferred for three of the four fragments (Figure 3: see also Supplementary Figures 2–5 for full gel images). The specific complexes (denoted A1 and B1 in Figure 3) formed by fragments A and B, each of which contains two of the CCAAT boxes, were reciprocally competed, indicating that they most likely involved the same protein(s). Fragment C gave one clear complex (denoted C1 in Figure 3), which was competed only by fragment C. Fragment D gave no signal indicative of sequence-specific protein-binding.

To localize the sequences critical for binding, we carried out EMSA competition experiments using an overlapping series of unlabelled, double-stranded oligonucleotide competitors. As summarized in Figure 3, (see also the primary data on which Figure 3 is based, i.e. Supplementary Figure 6), oligonucleotide 12 alone was able to compete with fragment C for the formation of complex C1, localizing the region critical for binding to the sequence spanning the first splice acceptor and downstream transcriptional start of \textit{Mrps12}. The CCAAT box containing fragments A and B showed an unexpectedly complex pattern of oligonucleotide competition. The formation of complex B1 was strongly inhibited by oligonucleotide 9, containing CCAAT box IV, and also, though less strongly, by oligonucleotide 7, containing CCAAT box III. The formation of complex A1 was also inhibited strongly by oligonucleotide 9 and moderately by oligonucleotides 5 and 2, containing, respectively, CCAAT boxes II and I. No competition was evident from any oligonucleotide other than those containing the CCAAT box elements. In all cases we were able convincingly to detect only a single, specific complex. The simplest interpretation is that the strongest binding is to the region of CCAAT box IV, but that the binding activity does not reside in single CCAAT boxes, rather in the array as a whole.

Figure 3. EMSA analysis of protein-binding to the bidirectional promoter region. The main features of the promoter region are reproduced from Figure 1, aligned with the four fragments tested in the assays (for fragment details Supplementary Table 2). The right hand panels show summarized data for each fragment, abstracted from Supplementary Figures 2–5, where full details of the gels are shown. In each case, only the major complexes resistant to both of the non-specific competitors, i.e. \(\alpha\) = poly(dA–dT)·(dA–dT) and \(\mathrm{i}\) = poly(dI–dC)·(dI–dC), are shown, designated A1, B1 and C1 as in Supplementary Figures 2–4. No specific EMSA signal was seen for fragment D (see Supplementary Figure 5). Competition from unlabelled DNA fragments in each series was 1-, 25- and 80-fold mass excess. The results of competition from oligonucleotides 1–16 (see Supplementary Table 2) are summarized at bottom left, based on EMSA data shown in Supplementary Figure 6. White boxes indicate oligonucleotides, which showed no competition (even at \(>100\)-fold molar excess), grey boxes those that showed weak competition, and black boxes those that showed strong competition at \(>25\)-fold molar excess.
When CCAAT box III was destroyed by mutation, almost no change was detected by EMSA in protein-binding to fragment B (Figure 4b). Destruction of CCAAT box IV did, however, reduce the efficiency of binding, and the complex migrated slightly faster. Destruction of both CCAAT boxes III and IV completely abolished specific protein-binding to fragment B. Similarly, destruction of CCAAT box II had almost no effect on the complex formed by fragment A (Figure 4b), although in this case, destruction of CCAAT box I abolished specific binding, suggesting that CCAAT box II is at most an accessory element.

In support of this, fragments containing CCAAT boxes I, III or IV individually formed similar complexes (Supplementary Figure 7b), whereas the fragment containing only CCAAT box II formed a faster-migrating complex. CCAAT box I was able to compete efficiently against CCAAT box II for the formation of this complex, whereas competition was much weaker in the reciprocal reaction. The complex formed by the fragment containing CCAAT box IV was resistant to competition from the other CCAAT box containing fragments, whereas that formed by CCAAT box III was moderately competed by CCAAT box IV (Supplementary Figure 7b).

An EMSA probe (fragment E) containing all four CCAAT boxes formed a single major complex (Figure 4c, Supplementary Figure 7c). Destruction of the CCAAT boxes in various permutations confirmed a hierarchy of binding strengths. CCAAT box I or IV on their own yielded complexes of similar but slightly more discrete mobility than the full fragment (Figure 4c). The CCAAT box I complex also migrated slightly faster, consistent with the possibility that the mobility of the complexes might reflect different degrees of distortion by DNA bending, dependent on the location of the binding site within the fragment (although other explanations are possible). CCAAT box III, the most centrally located within the fragment, was, on its own, weakly able to support the formation of an even more retarded complex. CCAAT box II alone was unable to support the formation of a discrete complex, although a residual smear was always seen. The presence of CCAAT box II alone had little effect on the mobility of the complexes formed by the other CCAAT boxes. The combinations of CCAAT box I and IV (or I and III) yielded less discrete mixtures of complexes similar to those produced by the full fragment.

All the above complexes are inferred to represent mutually exclusive occupancy of only one of the CCAAT boxes. In lower percentage gels a more retarded complex was sometimes seen (e.g. Figure 4b), although antibody supershift experiments suggest that it contains a distinct protein composition rather than multiple occupancy of the CCAAT boxes.

Antibody supershift experiments demonstrated that the factor binding to the CCAAT box array was NF-Y (Figure 4d, Supplementary Figure 7d).
Supplementary Figure 7d, e and f). An antibody against NF-Y supershifted the complex formed by the fragment containing all four CCAAT boxes (Figure 4d) or by any of CCAAT boxes I, III or IV individually (Supplementary Figure 7d and e). No supershift was observed with an antibody against c-Jun (Figure 4d and e, Supplementary Figure 7f), despite the presence of two regions of partial homology to the AP-1 binding site consensus in the promoter region (see Supplementary Figure 1b).

With an antibody against C/EBP a weak supershift was seen in at least three separate experiments. However, the signals were clearly seen only for large fragments containing CCAAT boxes III and IV. The supershift was most convincingly seen when CCAAT box IV was destroyed by mutagenesis (Figure 4d). The band which was supershifted by the antibody against C/EBP appeared to be the same one which was supershifted by the antibody against NF-Y. We tentatively conclude that, in vitro, C/EBP interacts weakly or transiently with this complex, perhaps via a second CCAAT box unoccupied by NF-Y.

Using nuclear extract from mouse liver we were able to obtain a similar pattern of protein-binding to fragments containing the CCAAT boxes (Figure 4e). Once again, a clear supershift was obtained using an antibody against NF-Y. However, the antibody against C/EBP also produced a supershift of the complex formed by the fragment containing CCAAT boxes III and IV even when CCAAT box IV remained intact.

The results of EMSA experiments can be summarized as follows: sequence-specific protein-binding was detected to the CCAAT box array and to a sequence element within the Mrps12 first intron. The protein-binding to the CCAAT box array in vitro was identified as NF-Y. At least in nuclear extracts from some tissues, C/EBP also appeared to be present within the same complex.

**Analysis of protein-binding in vivo**

*In vivo* footprinting was used to confirm occupancy of the CCAAT boxes and to probe protein–DNA interactions in vivo in the surrounding region, in both cultured cells and solid tissues. Signals indicating occupancy of all four CCAAT boxes (protection of both guanines in the core sequence) were detected (Figure 5a and b). In addition, we detected protected guanines in the regions between CCAAT boxes I and II, and between CCAAT boxes III and IV.

![Figure 5](https://academic.oup.com/nar/article-abstract/35/2/664/2400181/fig5)
In vivo were free from contamination. The chromatin was omitted also gave no product (data not used under the same conditions. Mock reactions in which was included in the reaction) or only a very weak product, controls either gave no detectable product (when no antibody

to the adjacent region. A number of hyperaccessible guanines were detected, in particular near the transcriptional start site of Sarsm, where they were closely interspersed with protected nucleotides, indicative of a distorted DNA structure. Hyperaccessible sites were found also between some of the CCAAT boxes. The two regions where there was inter-

spersed occupancy of the CCAAT boxes (by NRF-2 binding in vitro, even though the site was not accessible in naked DNA in vitro, possibly because of NF-Y binding to the nearby region of fragment C were slightly hyperaccessible to methylation.

To confirm the binding of NF-Y to the bidirectional promoter region in vivo, we carried out ChIP assays using the antibody against NF-Y. The procedure consistently yielded a strong final PCR product, using primers specific for the Mrps12-Sarsm promoter (Figure 4f), whereas negative controls either gave no detectable product (when no antibody was included in the reaction) or only a very weak product, when a different antibody (against the FLAG epitope) was used under the same conditions. Mock reactions in which the chromatin was omitted also gave no product (data not shown), confirming that the antibody and other reagents were free from contamination.

The analysis of protein-binding to the bidirectional promoter region in vivo, based on in vivo footprinting and ChIP assays, thus indicated occupancy of the CCAAT boxes (by NF-Y), and also of the NRF-2 binding site, as well as possible protein-binding at other sites.

Effects of CCAAT boxes and other sequence elements on promoter activity

Our initial promoter deletion analysis indicated the importance of the CCAAT box region for transcriptional activity, as well as the existence of a putative repressor-binding site in the region of the Mrps12 splice acceptor/intron1. To analyse the effects of these and other sequence elements in finer detail, we constructed a series of point mutants of the promoter region, using the dual luciferase reporter vector, and tested their effects on luciferase expression (Figure 6, Supplementary Figure 9, Supplementary Table 1).

We replaced the central CA of each CCAAT box, both individually and in all combinations, by TG. Destruction of either CCAAT box I or IV resulted in a substantial change in the directionality of transcription, whereas loss of CCAAT boxes II and/or III caused only a modest drop in the total amount of expression. Loss of CCAAT box I, whether alone or in combination with CCAAT boxes II or III, resulted in a large relative upregulation of transcription in the Sarsm direction (Figure 6 and Table 1), whereas loss of CCAAT box IV had the opposite effect. Destruction of all four CCAAT boxes reduced the total amount of expression to about 15% of the wild-type level, but the effect was more marked on Mrps12 than on Sarsm. In other words, the CCAAT box array as a whole seems to enforce directionality in favour of Mrps12.

CCAAT box I alone gave substantial transcriptional activity favouring the Mrps12 direction, which was enhanced by the presence of CCAAT box II. CCAAT box IV alone gave substantial transcriptional activity favouring the Sarsm direction, and the additional presence of CCAAT boxes II and/or III did not affect the directional bias. When only CCAAT box II or III was left intact, transcription was supported at ~30–40% of wild-type levels, with a residual bias in favour of Mrps12.

We also analysed the effects of making point mutations at other sites identified from in vivo footprinting as having potential roles in transcriptional regulation: abolition of the putative NRF-2 binding site reduced total expression <50% of wild-type, without affecting the directional bias (Figure 6). Abolition of the AP-1 binding site similarity in either or both of the two regions where protected and hyperaccessible sites were adjacent resulted in only a small (<30%) drop in transcription and again no change in its directionality (Figure 6). Mutations introduced at many other sites in the promoter region, which were identified bioinformatically as potential binding sites for various transcription factors, had similarly modest effects (Supplementary Figure 9), with one notable exception. A cluster of mutations (Mu3), introduced within the end of the phylogenetically conserved core of the Mrps12 first intron, adjacent to the site of the deletion (M2) that up-regulated both genes, also up-regulated both genes (Figure 6). As in the case of deletion M2 (Figure 2), upregulation of Mrps12 was the more dramatic, in this case by a factor of over 3. When CCAAT boxes II, III and IV were absent, this effect was completely masked (compare constructs ScatP, Mu3 and ScatMu in Figure 6). It was also abolished almost completely when only CCAAT box IV was destroyed at the same time (compare constructs 175 and 175Mu in Figure 6) Destruction of CCAAT boxes I, II or III individually had much milder effects on the upregulation caused by loss of Mu3, at least in the Mrps12 direction. We infer that the Mu3 element causes transcriptional repression in combination with the CCAAT box array, most potently with CCAAT box IV. The Mu3 element contains putative binding sites for a number of transcription factors, including NF-kB, the progesterone receptor and ELK. Note, however, that we did not detect high-affinity protein-binding to this region by EMSA, nor by in vivo footprinting (Supplementary Figure 8).

The mutant constructs tested in the reporter analysis included three carrying point mutations in the region of
oligo-12, where protein-binding was detected in vitro by EMSA [Figure 3 but not in vivo by footprinting (Supplementary Figure 8)]. These mutations destroyed putative binding sites for GATA-1, MYB, MTF-1 and PPAR/RXR heterodimers. None of the mutant constructs showed substantial alteration in reporter expression.

Table 1. Altered ratio of expression in the Mrps12 and Sarsm directions in response to destruction of one or multiple CCAAT box elements

<table>
<thead>
<tr>
<th>Construct</th>
<th>CCAAT box destroyed</th>
<th>Expression ratioa</th>
<th>Total expressionb</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>None</td>
<td>3.5</td>
<td>129</td>
</tr>
<tr>
<td>38</td>
<td>I</td>
<td>0.8</td>
<td>131</td>
</tr>
<tr>
<td>86</td>
<td>II</td>
<td>2.7</td>
<td>74</td>
</tr>
<tr>
<td>126</td>
<td>III</td>
<td>2.6</td>
<td>71</td>
</tr>
<tr>
<td>175</td>
<td>IV</td>
<td>9.3</td>
<td>211</td>
</tr>
<tr>
<td>ScatA</td>
<td>I + II</td>
<td>0.5</td>
<td>106</td>
</tr>
<tr>
<td>ScatG</td>
<td>I + III</td>
<td>0.4</td>
<td>101</td>
</tr>
<tr>
<td>ScatD</td>
<td>I + IV</td>
<td>1.6</td>
<td>134</td>
</tr>
<tr>
<td>ScatB</td>
<td>II + III</td>
<td>2.0</td>
<td>85</td>
</tr>
<tr>
<td>ScatC</td>
<td>II + IV</td>
<td>5.6</td>
<td>48</td>
</tr>
<tr>
<td>ScatE</td>
<td>III + IV</td>
<td>11.0</td>
<td>85</td>
</tr>
<tr>
<td>ScatN</td>
<td>II + III + IV</td>
<td>4.2</td>
<td>66</td>
</tr>
<tr>
<td>ScatF</td>
<td>I + III + IV</td>
<td>1.0</td>
<td>35</td>
</tr>
<tr>
<td>ScatV</td>
<td>I + II + IV</td>
<td>1.7</td>
<td>38</td>
</tr>
<tr>
<td>ScatP</td>
<td>I + II + III</td>
<td>0.5</td>
<td>188</td>
</tr>
<tr>
<td>ScatH</td>
<td>I + II + III + IV</td>
<td>0.9</td>
<td>19</td>
</tr>
</tbody>
</table>

aRatio of expression in the Mrps12 direction to that in the Sarsm direction.
bArbitrary units, as defined in legend to Figure 2, where wild-type expression in the Mrps12 direction represents 100 U of expression.

Figure 6. Reporter analysis of mutated versions of the Mrps12-Sarsm bidirectional promoter. The main features of the promoter region are reproduced from Figure 1. The intact promoter elements remaining in each reporter construct are indicated by open circles (CCAAT boxes) and filled diamonds (NRF-2 site), aligned with the schematic map. The locations of other, specific sets of point mutations introduced at other sites are shown by crosses. For details of mutations and all constructions see Supplementary Table 1 and Supplementary Figure 9b. Expression in the two directions was calculated and plotted as in Figure 2. Note that for the point mutation series, all constructs were oriented in the same manner.

DISCUSSION

In this study, we used reporter analysis, EMSA and in vivo footprinting to identify an array of four CCAAT box elements governing the extent and directionality of transcription of the oppositely transcribed genes Mrps12 and Sarsm from their common intergenic promoter region in mouse cells and
tissues. All four CCAAT boxes are oriented on the same strand, but those at each end appear to direct the transcriptional apparatus to initiate at the opposite end of the intergenic region. The two CCAAT boxes in the middle of the array appear to have weaker binding affinity and are suggested to function as accessory elements. The main transcription factor implicated in this regulation is NF-Y. Weaker or inconclusive evidence was obtained for the involvement of other factors (C/EBP, NRF-2, AP-1, plus two unidentified factors), but their role appears to be minor. The pattern of protein-binding to the promoter, inferred from in vivo footprinting, appears similar in cultured fibroblasts and solid tissues.

**Transcriptional versus post-transcriptional regulation of Mrps12 and Sarsm**

Quantitative RT–PCR showed that the steady-state levels of Sarsm and Mrps12 mRNAs are similar in cultured 3T3 cells, but reporter assays consistently showed a 4-fold greater expression in the Mrps12 direction, in constructs completely lacking the coding sequences and 3’-UTR segments of the mRNAs. The relative mRNA levels from the two genes transcribed in the natural setting is, thus, dramatically different from the relative levels of the two luciferase reporters directed by the cloned promoter. In principle, this could reflect an important contribution from post-transcriptional regulation, e.g. at the level of mRNA stability or, alternatively, an influence of transcription-regulatory elements located far from the proximal promoter, perhaps up to tens of kilobases away. Obviously, the latter cannot be ruled out other than by an exhaustive study of the entire chromosomal region using reporter methodology. NRF-1-like recognition elements within the introns of Sarsm would be one candidate for such activity worthy of special attention.

Nevertheless, we favour the idea that the relative levels of the two mRNAs are mainly adjusted post-transcriptionally. In previous studies we obtained evidence for regulation of Mrps12 at the levels of translation and alternative splicing (8). Previous studies also showed that Sarsm was ubiquitously expressed, although at different levels depending on the tissue (4). Whilst transcription may favour the expression of Mrps12, post-transcriptional regulation, probably at the level of mRNA stability, would adjust the relative levels of the two mRNAs according to physiological needs. The need for the synthesis of the two gene products should differ between growing and quiescent cells. Whereas Sarsm is a soluble enzyme, Mrps12 is a component of a stable particulate structure, the small mitochondrial subunit. In rapidly proliferating cells, Mrps12 synthesis is clearly translationally up-regulated (8), and mRNA stability may also play a role in this. Both Mrps12 and Sarsm mRNAs vary in abundance between tissues (4,8), and the patterns are distinct, although this might also reflect species differences. Conversely, in the present study, in vivo footprinting revealed similar patterns of protein occupancy of the promoter region in different tissues. This supports the idea that tissue-differences in the levels of these two mRNAs are effected post-transcriptionally.

**Transcriptional regulation by NF-Y**

Nuclear genes for mitochondrial proteins involved OXPHOS or mitochondrial gene expression are governed by ubiquitous transcription factors able to gear expression to the demands of cellular growth or respiration. In mammals, the most prominent of these are nuclear respiratory factors 1 and 2 (Ref. (13)). In yeast, however, the heterotrimeric CCAAT box-binding factor composed of the proteins Hap2p, Hap3p and Hap5p (27,28), augmented by the Hap4p activator protein (29,30) is the most important co-regulator of nuclear genes involved in mitochondrial biogenesis and respiratory function (14–16). The present study provides a bridge between these observations, by identifying the mammalian homologue of the Hap2/3/5 complex, NF-Y (31), as a key regulator of two genes of the mitochondrial translational machinery in mouse cells.

CCAAT boxes I and IV of the Mrps12-Sarsm promoter show a good match to the consensus binding site for NF-Y (32), whereas CCAAT boxes II and III of the array, which were inferred to show weaker binding and to function most likely as accessory elements, show a weaker match to the consensus (Supplementary Figure 10). ChIP analysis also confirmed that NF-Y is bound to the promoter in vivo. In vivo footprinting showed evidence for occupancy of all four CCAAT boxes of the array, although the results of EMSA suggest that protein-binding to each site may be mutually exclusive. Moreover, reporter analysis confirmed that binding at each site confers a different transcriptional selectivity. Since all four CCAAT boxes are oriented on the same strand, we propose that NF-Y binding is able to recruit the transcriptional apparatus at this promoter region in two oppositely oriented ways. NF-Y is known to have this property more generally: approximately half of all NF-Y binding sites in vertebrates are oriented with ATTTG on the coding strand, especially when TATA is absent, as here (32).

Based on our findings, NF-Y binding at CCAAT box I, especially when present together with its accessory site CCAAT box II, appears to favour transcription in the Mrps12 direction, whereas binding at CCAAT box IV, in conjunction with either of CCAAT boxes III or II, favours transcription in the opposing, Sarsm direction. In human (Supplementary Figures 1 and 10), CCAAT box II shows a better match to the NF-Y consensus than CCAAT box I, so it will be interesting to compare the regulatory properties of these elements in human cells, using similar assays.

Regulation of transcription by an array of precisely spaced CCAAT elements interacting with NF-Y has previously been reported for genes linked to the cell division cycle, such as cyclin B2 (33), cdc25C (34) and thymidine kinase (35). However, this is, to our knowledge, the first report of such an array as the key component of a bidirectional promoter, although the mouse promoter for O-sialoglycoprotein endopeptidase and APEX endonuclease may represent a second example (36). The inherent bidirectionality of the action of NF-Y makes it an ideal transcription factor to govern bidirectional promoters. The presence of multiple recognition elements favouring transcription in each direction should constitute an additional failsafe mechanism to ensure that both genes are transcribed.

Distortion of promoter DNA appears to be a common mechanism by which NF-Y activates transcription (37). The subtly variable mobility shifts seen in vitro (Figure 4b), plus the detection of adjacent hyperaccessible and protected
sites within and adjacent to the CCAAT box array (Figure 5) suggest that such bending occurs also within the Mrps12-
Sarsm promoter.

Possible roles of other regulatory factors

The Mrps12-Sarsm promoter region contains putative binding sites for several other transcription factors, including AP-1 and NRF-2. Occupancy of these sites in vivo was suggested by in vivo footprinting, and reporter analysis confirmed that they contribute, albeit modestly, to the total transcriptional activity in both directions, despite our failure to detect protein-binding at these sites by EMSA.

As well as facilitating assembly of the basal transcription machinery, NF-Y can recruit other transcription factors (38–40). Recruitment of SREBP-1a to the rat farnesyl diphosphate synthase promoter is dependent on NF-Y, but SREBP-1a binding does not require continued binding of NF-Y to the DNA (40). Transient NF-Y binding, resulting in the recruitment of other factors, may be one way to reconcile our EMSA and in vivo data. It is also possible that NF-Y binding in the intergenic region may facilitate transcription factor binding to more distant binding sites, e.g. the putative NRF-1 binding sites within Sarsm coding or intronic DNA.

In several experiments, we found evidence for an interaction between the NF-Y bound complex and C/EBP, especially in liver (Figure 4e) or when CCAAT box IV was disrupted (Figure 4d). Mrps12-Sarsm may thus be co-regulated by these factors, as previously inferred for mouse adiponectin (41), human microsomal epoxide hydrolase, EPHX1 (42) and mouse amelogenin (43).

In other genes studied, e.g. rat cytochrome c oxidase subunit IV (44), loss of NRF-2 binding usually causes a dramatic drop in transcription. In contrast, destruction of the NRF-2 binding site in the Mrps12-Sarsm promoter entrains only a 50% decrease in transcription (Figure 6). However, NRF-2 might still be important for fine-tuning expression to physiological conditions, since it is regulated by reactive oxygen species (45,46), and is also translocated to the nucleus in neurons under stimulation (47). Another possibility is that, in this promoter, NRF-2 is redundant with other transcription factors (e.g. C/EBP, NRF-1, AP-1), any or all of which may be recruited by NF-Y bound to the CCAAT box array.

We found evidence for two other protein–DNA interactions within the optional Mrps12 first intron that may influence transcription, although in vivo and in vitro findings were not fully consistent. A clear EMSA signal was obtained that was specifically competed by an oligonucleotide covering the splice donor site. However, we found no evidence from in vivo footprinting for occupancy of this region in vivo, nor any substantial effect of point mutations in it on the expression of reporter genes. The possible role of a factor binding at this site in vivo remains open.

Further downstream, in the phylogenetically conserved portion of the intron, we found evidence from reporter analysis for a negatively acting element. The identity of any protein(s) interacting with this site remains unclear. No sequence-specific binding to this region was revealed by EMSA, but the effect of mutating the site was abolished by concomitant disruption of the CCAAT box array, most notably in a construct in which CCAAT box IV was destroyed (compare constructs 175 and 175 Mu in Figure 6). Therefore, whatever protein(s) are involved most likely act only in combination with NF-Y and possibly other factors. If sequence-specific contacts are involved, they do not appear to involve any guanines.

In yeast, the glucose-responsiveness of genes controlled by the Hap2p/Hap3p/Hap5p complex is conferred by the differentially transcribed activator protein Hap4p (48). In mammals, there appears to be no homologue of Hap4p, and instead NF-Y interacts directly with the transcriptional apparatus and/or other transcription factors, as part of more diverse regulatory programmes than simply the switching on of mitochondrial biogenesis during the yeast diauxic shift. However, as we document here, ensuring the supply of at least two key proteins of the machinery of mitochondrial protein synthesis is also one of the tasks of NF-Y in mammals.

Mrps12-Sarsm as a general paradigm for bidirectional promoters

Approximately 2000 bidirectional promoters are found in mammalian genomes, accounting for about 10% of all genes (5,6). As in the majority of such cases, the bidirectional promoter governing Mrps12 and Sarsm is GC-rich, constitutes a CpG island and is TATA-less. This organization is also conserved phylogenetically, at least amongst mammals. Like many such gene pairs, e.g. chaperonins HSP60 and HSP10 (49) the two gene products are functionally related.

The main regulatory elements have been identified in only a handful of examples of such bidirectional promoters. This study therefore represents a useful paradigm, both methodologically and in terms of its major findings, for analyzing the transcriptional regulation of an important class of promoters. A complete understanding of this regulation is a prerequisite of any attempt to manipulate the expression of just one member of such gene pairs, with biotechnological or biomedical applications in view.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

The authors thank Outi Kurronen, Merja Jokela and Anja Rovio for technical assistance, Anne Hyvärinen, Esko Kemppainen, Hans Spellbrink and Herma Renkema for technical advice and help, and our many colleagues in IMT and FinMIT, especially Anu Wartiovaara, for useful discussions. This work was supported by Academy of Finland, Juselius Foundation, Tampere University Hospital Medical Research Fund and the European Union. Funding to pay the Open Access publication charges for this article was provided by University of Tampere and Academy of Finland.

Conflict of interest statement. None declared.
REFERENCES


