Gene expression of the dibasic-pair cleaving enzyme NRD convertase (*N*-arginine dibasic convertase) is differentially regulated in the GH3 pituitary and Mat-Lu prostate cell lines

Andrew G. WINTER and Adrian R. PIEROTTI¹

School of Biological and Biomedical Sciences, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, Scotland, U.K.

NRD convertase (*N*-arginine dibasic convertase, NRD-C) is a dibasic selective metalloprotease which cleaves on the N-terminal side of an arginine residue in a dibasic pair. Abundant in endocrine tissues, the highest levels are found in testis. The mechanism whereby NRD-C expression is regulated at the transcriptional level has been examined by reporter-gene assay and electrophoretic-mobility-shift assays. Analysis of the rat and human promoters show that they are highly conserved, containing a number of motifs which may correspond to transcription-factor binding sites. The rat promoter has been cloned into a luciferase reporter vector and analysed in a number of cell lines. Full functionality of the promoter is observed with 5'

deletions to 411 bp upstream of the transcriptional start site in spermatid, prostate and pituitary cell lines. Further deletion to 101 bp causes a complete loss of activity in spermatid and prostate lines. By contrast, GH3 pituitary cells display no reduction in promoter activity with deletion to 101 bp of upstream sequence. A number of transcription-factor binding sites have been identified by electrophoretic-mobility-shift assays in the region 411–101; however, no differences in binding between the cell lines were observed.

Key words: luciferase, M16 family, metalloprotease, promoter.

INTRODUCTION

Limited proteolysis is a post-translational modification required for the release of bioactive peptide hormones and neuropeptides from larger, inactive polypeptide precursors [1,2]. Within the precursor, peptides are usually flanked by pairs of basic residues or, less commonly, by single arginine residues [1] or non-basic residues [3]. These events mostly begin in the Golgi apparatus [4,5] and continue into the secretory vesicles [6], but also may modulate hormone levels extracellularly [7,8].

The endoproteases responsible for this processing in mammals are known as convertases or maturases, and a family of such enzymes related to the yeast KEX2 gene product [9,10], a subtilisin-like serine protease, has been isolated. These proteases are very specific, cleaving only at certain dibasic sites on the C-terminal side of an arginine residue. To date, seven such enzymes have been identified (for reviews, see [3,11]). Some of these enzymes are generally localized and are involved in processing of constitutively secreted proproteins, whereas others have a restricted distribution, such as the testis-specific convertase PC4 [12].

N-Arginine dibasic convertase (NRD-C; nardilysin, EC 3.4.24.61) is a 140 kDa metalloendopeptidase, originally purified from rat testis and brain, which cleaves a number of neuro-endocrine peptides *in vitro*, including dynorphin A, atrial natriuretic factor and somatostatin-28, on the N-terminal side of an arginine residue at the dibasic processing site [13]. Despite the identification of several *in vitro* substrates for the enzyme, its function *in vivo* remains unknown.

NRD-C activity is present in predominantly endocrine tissues, including brain, pituitary and adrenal, but it is highly abundant in the testis, with activity levels ten times higher than that of the brain cortex [13]. Immunofluorescence studies in the testis show that the enzyme is localized within the seminiferous tubules, distributed in the immature round spermatids and, to a greater extent, in the more mature elongating spermatids [14]. *In-situ*-hybridization studies show NRD-C mRNA is present in all round and elongating spermatids, whereas no enzyme can be detected in spermatogonia and spermatocytes [14]. A study of the subcellular localization in elongated spermatids showed that most of the enzyme was found associated with the microtubular structures of the axoneme and the manchette [15]. There is evidence to suggest that NRD-C activity within the brain is localized to the secretory vesicles [16], and it has been shown to be associated with secretory-granule membranes in rat brain cortex and hypothalamus [17].

Rat, human and mouse cDNA sequences encoding NRD-C have been isolated [14,18–20], and the deduced amino acid sequences have revealed that the enzyme is part of a large family of metalloendopeptidases – the M16 family (using the nomenclature of Rawlings and Barrett [21,22]). Many of the enzymes in this family have been implicated in prohormone or proprotein processing events.

Insulin-degrading enzyme (insulysin, EC 3.4.24.56) is involved in the degradation of a number of peptide hormones [23–29]. It has been implicated as having a processing function with the discovery that it is identical with γ -endorphin-generating enzyme [30], which converts β -endorphin to γ -endorphin. Like NRD-C,

Abbreviations used: NRD-C, NRD (*N*-arginine dibasic) convertase; DMEM, Dulbecco's modified Eagle's medium; FBS, foetal-bovine serum; DTT, dithiothreitol; 5'-UTR, 5'-untranslated region; RACE, rapid amplification of cDNA ends; SV40, simian virus 40.

¹ To whom correspondence should be addressed (e-mail a.pierotti@gcal.ac.uk).

The sequences of the rat and human NRD-C promoters have been submitted to the EMBL, GenBank®, DDBJ and GSDB Nucleotide Sequence Databases under the accession numbers AJ000349 and AJ000350 respectively.

insulysin is also present at elevated levels in the testis [31]. The closely related human metalloendoprotease-1 cleaves a prodynorphin-derived peptide, leumorphin, on the N-terminus of a single arginine residue *in vivo* [32] and is thought to play a broad role in general cellular regulation. Also related to NRD-C are mitochondrial processing peptidases and Axllp, a yeast protein shown to be involved in pro- α -factor maturation [33]. The fact that several members of the protease family to which NRD-C belongs are involved in maturation of precursors or subsequent modification of bioactive peptides clearly strengthens the hypothesis that NRD-C is a processing enzyme.

Currently, little is known about the regulation of either the activity or expression of NRD-C. An unusual feature of NRD-C is the presence of a stretch of acidic amino acids, 71-residues long in the rat sequence region, which is absent from every other member of the family [14]. It has been proposed that this acidic stretch might be a target for polyamines, which have been shown to inhibit the binding of rat, mouse and human NRD-C to its substrates and to control the activity of the enzyme [18,34]. It has been reported that expression of the enzyme is modulated by retinoic acid in human neuroblastoma cell lines and that this retinoic acid-responsiveness might be mediated by a retinoic acid-responsive element within the promoter of the NRD-C gene [35]. The purpose of the present study was to investigate the mechanism whereby NRD-C expression is regulated at the transcriptional level.

We have cloned and sequenced the 5' upstream region of both the human and rat NRD-C genes. Computer analysis shows that the two sequences are well conserved and contain many potential regulatory-element binding sites. Using luciferase reporter technology we have shown that the rat NRD-C 5' upstream region drives high levels of luciferase expression in pituitary and prostate cell lines shown to contain NRD-C protein. 5' Deletions of the rat promoter indicate that the gene is differentially regulated in these two cell lines. Electrophoretic-mobility-shift analysis on a number of potential regulatory-element binding sites confirms that nuclear proteins extracted from these cell lines can bind to the NRD-C promoter.

MATERIALS AND METHODS

PCR amplification of genomic DNA upstream of the human and rat NRD-C gene coding region

The region of genomic DNA upstream of the coding region of the human and rat NRD-C genes was amplified using rat and human GenomeWalker kits from Clontech (Basingstoke, Hants., U.K.). The rat NRD-C promoter was amplified using the gene-specific primers rPF1 (5'-AAAGACTGCACCAACCGCGACTCTC-3') and rPF2 (5'-GCATTCACCAGCAAGTCGTAGCGCTGGACA-3') based on the rat cDNA sequence [14]. Since the sequence of the human orthologue had not at that time been published, sequence information for the design of the human-specific NRD-C PCR primers hPF1 (5'-GATGTATCGGTAT-TGCTTGG-3') and hPF2 (5'-GGAGACTTGACGATCTCAGG-3') was obtained from an EST (expressed sequence tag) clone (GenBank accession no. T25336) with a high level of similarity (77.8%) to the rat cDNA sequence. These oligonucleotides were used to amplify the human NRD-C promoter.

Primary PCR reactions were carried out according to the manufacturer's instructions using a mixture of Tth DNA polymerase (Amersham Pharmacia Biotech) and Vent DNA polymerase (New England Biolabs) at a Tth/Vent ratio of 20:1. A 1/5000th portion of the primary PCR reaction mixture was used in a secondary PCR reaction with the same conditions as the primary reaction using a nested primer pair.

Cloning and sequencing of PCR products

PCR products were ligated into a TA cloning vector (pTAG; R & D Systems, Abingdon, Oxon., U.K.). Resulting ligations were used to transform DH5 α cells, and individual clones were sequenced using the Sequenase version 2.0 DNA Sequencing Kit (Amersham Pharmacia Biotech) to check the identity of the insert. To allow sequencing of the entire insert, a series of 5' deletions was made of each clone using the Erase-a-Base system from Promega. Sequencing of these deletions was carried out using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer) in conjunction with an ABI 373 Automated Sequencer. Inserts were sequenced on both strands. The plasmids containing the rat and human NRD-C promoter fragments were named pRNRD and pHNRD respectively.

Computer analysis of sequences

The human and rat NRD-C 5' upstream sequences were aligned using the program ALIGN [36]. Potential regulatory sequences were identified using MatInspector version 2.2 [37]. CpG islands were identified using XGRAIL [38].

5'-Rapid-amplification-of-cDNA-ends (5'-RACE) analysis

RNA was isolated from rat brain cortex and rat testis using TRI reagent as recommended by the supplier (Sigma). Human whole brain and testis RNA was purchased from Clontech. 5'-RACE analysis was performed using the Life Technologies (Paisley, Renfrewshire, Scotland, U.K.) system according to the manufacturer's instructions. First-strand cDNA synthesis was carried out using 1 μ g of RNA and rGSP1(5'-GCATTCCTGTTTT-ACCCTCC-3') or hPF1. The cDNA was dC-tailed and PCR-amplified using 5'-RACE Anchor Primer (Life Technologies) and either rGSP2 (5'-ATCAAAAGAGCCTGCAAGCC-3') or hPF2. PCR products were gel-purified, ligated into a TA cloning vector and sequenced using the T7 Sequenase version 2.0 DNA Sequencing Kit (Amersham Pharmacia Biotech).

Cloning of the rat NRD-C 5' upstream sequence into luciferase reporter vectors

The rat NRD-C promoter region was cloned between the *Mlu*I and *Xma*I sites of the pGL3-basic vector as follows: pRNRD was digested with *Mlu*I and *Bst*ZI to produce a promoter fragment of 1446 bp. The *Bst*ZI 'sticky end' was partially filled in with dGTP in a Klenow DNA polymerase reaction to leave an overhang of two G residues. The plasmid pGL3-basic was digested with *Psp*AI, phenol/chloroform-extracted and ethanol-precipitated. The ends were partially filled in with dCTP in a Klenow DNA polymerase reaction to leave an overhang of two C residues. After digestion with *Mlu*I, the plasmid was ligated to the promoter fragment, generating the construct pRPb-1376 which was confirmed by sequencing.

Construction of deletion mutants of the rat NRD-C promoter

pRPb-1376 was digested with an appropriate enzyme (*Nhe*I, *Bst*XI, *Afl*II or *Pvu*II) plus *Mlu*I. The restriction fragment containing the pGL3-basic vector was gel-purified and treated with Klenow DNA polymerase. After re-ligation, the 5' end of the insert was sequenced to confirm its identity. These clones were named according to the length of the remaining promoter fragment: pRPb-1049, pRPb-550, pRPb-411 and pRPb-101.

Cell culture

Cell lines were obtained from the American Type Culture Collection or the European Collection of Cell Cultures. All media and supplements were obtained from Life Technologies. Cells were grown at 37 °C in an atmosphere of 5% CO₂. COS-1 cells were grown in Dulbecco's modified Eagle's medium (DMEM), Mat-Lu cells were grown in RPMI 1640/250 nM dexamethasone, GH3 cells were grown in Ham's F10, and GC-2spd(ts) cells were grown in DMEM/1mM non-essential amino acids. In addition, all media contained 10% (v/v) fetal-bovine serum (FBS), 100 units/ml penicillin, 100 mg/ml streptomycin and 2 mM L-glutamine.

Transfection assays

Cells were plated on to a 96-well plate at a density of 10⁵ cells/ml. The day after plating the cells were transfected using Tfx-20 reagent at a Tfx-20/DNA ratio of 24:1 as described by the supplier (Promega). Each construct was transfected six times. Each well contained 150 ng of test plasmid, 15 ng of pRL-CMV plasmid (Promega) in a total volume of 40 µl of serum-free medium. After a 2 h incubation, this mixture was replaced with medium and cells were grown for a further 48 h. Expression of firefly and *Renilla* luciferases was determined using the Dual Luciferase Assay System (Promega) in a MLX Microplate Luminometer (Dynex Technologies, Ashford, Kent, U.K.), which expresses luminescence in terms of relative light units.

Preparation of nuclear protein extracts

Nuclear proteins were extracted from cultured cells as follows. Growing cells (10⁶) were trypsin-treated and centrifuged at 1000 g for 5 min. The pellet was gently resuspended in 500 µl of ice-cold extraction buffer [10 mM Hepes (pH 7.9)/10 mM KCl/1.5 mM MgCl₂/0.5 mM dithiothreitol (DTT)/0.4 mM PMSF] and incubated on ice for 15 min. A 30 µl portion of Nonidet P40 was added and the material vortex-mixed for 10 s. This was micro-centrifuged at 12000 g for 30 s, and the pellet was resuspended in 50 µl of extraction buffer containing 420 mM NaCl and 0.1% Nonidet P40. This mixture was incubated at 4 °C for 20 min. Debris was removed with a 5 min centrifugation at 12000 g, and the supernatant, containing nuclear extract, was snap-frozen in liquid nitrogen and stored at -70 °C. Protein concentration was determined using the Bradford [39] assay.

Electrophoretic-mobility-shift assays

Double-stranded oligonucleotides were made by mixing 8.75 pmol of two complementary single-stranded oligonucleotides in 60 mM NaCl. This was heated to 65 °C for 2 min and cooled to room temperature over 1 h and then 5' end-labelled with [γ -³²P]ATP (10 mCi/ml; > 5000 Ci/mmol; Amersham Pharmacia Biotech) using T4 polynucleotide kinase. Binding reactions were carried out in a volume of 20 µl at 4 °C and contained 20 mM Hepes, pH 7.9, 20% (v/v) glycerol, 0.5 mM DTT, 1 mM MgCl₂ and 40000 c.p.m. of labelled oligonucleotide. Those binding reaction mixtures that contained nuclear extract (2.5 µg) also contained 1 µg of poly(dI-dC)·poly(dI-dC) DNA (Amersham Pharmacia Biotech). Some reaction mixtures also contained a 10-fold molar excess of unlabelled double-stranded oligonucleotide, and those were preincubated for 30 min at 4 °C before addition of the labelled oligonucleotide. After addition of probe, the binding reactions were allowed to proceed for 30 min. Reaction mixtures were then run on a 5%-polyacrylamide gel in 0.5 × TBE

at 4 °C and 200 V for 2 h. Oligonucleotide fragments were revealed by autoradiography.

Oligonucleotides

The oligonucleotides used were as follows: rPF1, 5'-AAAGAC-TGCACCAACCGCGACTCTC-3'; rPF2, 5'-GCATTCACCA-GCAAGTCGTAGCGCTGGACA-3'; hPF1, 5'-GATGTATC-GGTATTGCTTGG-3'; hPF2, 5'-GGAGACTTGACGATCT-CAGG-3'; rGSP1, 5'-GCATTTCTGTTTTACCCTCC-3'; rGSP2, 5'-ATCAAAAGAGCCTGCAAGCC-3'; Oct1, 5'-CC-ACTGGGAATATTTTCTCA-3'; IK-2, 5'-AAACTTCTCC-CGTCGTCCT-3'; SP1, 5'-CTTCGCCCGCCACGTCAG-3'; C/EBP, 5'-TCAGTTTTCGAAATGCTCTC-3'; NFY, 5'-CTCCTATTGGTTGTTT-3'; N_{-249/-226}, 5'-AAACTTCTCC-GTCGTCCTCCGCC-3'; N_{-236/-221}, 5'-CGTCGTCCTCCGC-CTCTTT-3'.

RESULTS

Cloning and sequencing of the human and rat NRD-C 5' upstream regions and identification of their transcriptional start sites

Using primer sets for rat and human NRD-C, a total of 1526 bp of genomic DNA upstream of the translational start codon of the rat NRD-C gene and 1493 bp of genomic DNA (including 316 bp downstream of the start codon) from human was amplified using a 'gene-walking' strategy. The sequences of these 5' upstream regions are shown in Figure 1.

To identify the site of initiation of transcription of the rat and human NRD-C genes, 5'-RACE was performed on RNA from brain and testes. PCR products obtained from this experiment were cloned into a TA vector, and multiple individual colonies were sequenced. An analysis of the 5' ends of the clones is shown in Figure 2. It can be seen that although the gene initiates transcription from multiple sites, a major transcription start site can be identified. This major start site is conserved between rat and human, and in each species there appears to be no significant difference in transcriptional start site between brain and testis. The start site is 150 bp upstream of the putative translational initiation codon in the rat NRD-C gene and 114 bp upstream of the putative translational initiation codon in the human NRD-C gene. For both species all of the 5'-RACE clones end within a 30 bp region.

Computer analysis of the rat and human NRD-C promoters

The two sequences, upstream of their putative translation start codons, were aligned using the sequence-alignment program ALIGN [36]. The complete human upstream sequence aligns with the rat sequence from -1074 in the rat sequence and -1058 in the human sequence to the start codon. The region upstream of the major transcription start site can be divided into two sections on the basis of how well conserved they are. The most 5' section of the human promoter (-1058 to -670) aligns with the rat promoter from -1074 to -642 and is 50.1% identical. The 3' section, from -641 to -1 in the rat and -669 to -1 in the human, is better conserved, being 62.1% identical. The alignment of this section can be seen in Figure 3. The 5'-untranslated region (5'-UTR) is extremely well conserved, except for a 29 bp region in the rat 5'-UTR (from +80 to +108) that is not present in the human 5'-UTR. When the 29 bp section is ignored, the identity of the two 5'-UTRs is 81.8%. The promoter sequences were analysed for potential transcription-factor binding sites using MatInspector version 2.2 [37]. Potential binding

A

```

-1376 ACTAGAGTACAGAGGAAGCTACTTTGAATGAGGACTCCAGTAGAGTTGCTTAAAACTGGTAAACCTTTAAAAGTGACT 80
-1296 GAAACTAGAAATATGTTAATGCTTTCCAAATACTTCTGTCCCAATCTATTATCCTCCTCAATACTTGGCTTTTCATCCT 160
-1216 GTTCTGTAGTCTTAGCATTTTTAAATTTTCTTTGAATGTACCATTATGGTCTGTTTCTCCCATAGAAAGGAACTT 240
-1136 CAAGAAGGCAAGGGTTTTTAAATTTTAAATTAATTAATAAATTTTTTAAATTTTAAATTTTGAAGACAGGGTCTTAGTAGGTA 320
-1056 AACCAGGCTAGCCTTGAACCTCTGTCTCCCTGAGTGCTCAGACTCAAGACGTAGACTACCACACCGGCAATATTGTGT 400
-976 TTTGTTTATGGCATAACTGAGGTTGAGAGGCATAACTACTAATGTATCTACTGGTGCTGCTTTTTTGTGTTGTTTATT 480
-896 CATATTCATTTTCGTTTATTCTTCCACATGGTACGATCTCATCCAATGGTGTCTATCAGCTTTTATTGTATCTGGTCAA 560
-816 CCGCTTTTATTTTATATTGTGTCTGTTCCAGTCGCAACAAAGGCACTGGCAACAACAAAACCAAAAGCTTGGGAAAG 640
-736 TATCTTGAGGGTAGTAGCCTAACATGTTTTTGTCTTTTCAGTAAAAGTTGGCTAAACTTAAAGGTAGGCTGAGGACAACT 720
-656 CCAGTCTGTGGGAAAGCTTAAAAAAAATCGCATACAAGAGTGAGCATATTGAAACCGATATTTTAGGCACTCTTTTGGGA 800
-576 TTTGGAGCGCAAAAGCACCATGGTTTGGCAGCTTACAGGATGGAAGGTGCAATTGTTATTTCCGGTGCTTAGTGACC 880
-496 CCAAAGAACAGCATCCATCTCAGGATAGGGTAAGAGAAGACTTCCGGTGTTCTGCGTTTTTAAAGAGACTCGCTGAGT 960
-416 CACTCTTAAAGACAGAATGCAAAAAATACACTGCAAGTGGGTGAAGAAGATCAGCATCCACTATTGCGACTCGTTCTGCG 1040
-336 GTGCTCCCACTGGGGAATATTTTCTCAGACCCGAGCACTACGCGGTTTTCCAGAAACAGTACGGACTCACGTACAAT 1120
-256 CGCGGCAAACTTCTCCCGTCGCTCCGCTCTTTTCTTCCGCCCGCCACGTCAGTTTTCGAAATGCTCTCTGGGTG 1200
-176 CGCCCGGAATCCAGCCTATTCTATTGGTCTCTAGCCTTCCCTGCTCTATTGGTTGTTTCTGCAACGTGACAGCTGTT 1280
-96 GTGCGTTGTGATTGGCTAGGATCCATAGCGTCATCGGCTGTGAGGAGGTGGGATGAAGAAAGTCGGTGTTGAGCGAAGC 1360
-16 CGGGTTGGAGGAGGGGTTTCAGGCCTGTCTCCCGAGCGGCTGGGGCAGCAACAACGCCGGCCGCCGCTGCCCTTGGG 1440
+65 ACGCGGAGGAGGAGGAGTGGCTGGAGGAGTAGGAGGAGAGGTTGGGCTGGGGAGCGGGATGTCCAGCGCTACGACTTGC 1520
+145 TGGTGAATGC 1530

```

B

```

-1058 ACTAGGATTATAGGTGTGAGCTATCACTATCACGCCAGCCTAAGGCAAGTTTTGTTTTGTTTCATGGTTGTATTCTTGC 80
-978 ATTTGGAACCTCAGAAGAGGCCCTCGAAAACCTTGAAGATACAGTAACATATCAACACATCTCCCTGCTTCACGTCTTTCCC 160
-898 GCATCATTCGTATGTATGAGCTATGATCTCTTACCCCATGGTGTGAGTATAGTGCATCTTACTAATTTAGCTGCTTTTTTA 240
-818 GCTTTCATTTATGTTTGTATTTCCGTCACATTTTACATTGCTAACCAAGACCACTGCTAGCGGAAAACTGGGAAAG 320
-738 AGACTTGAAGGCAGCAGACGCTGCAGCTGCTCGATAAAAATTTGGCCGAATTTAACAGGAGGCTGGAGGAGCTAAAAACG 400
-658 TTTTACACAGAAGAGTTAGCACATCCAGACTGGTCTTTTATAGCACTTTCTTAGGGCTTGAGGATGAACCCCACTG 480
-578 TCGGGCCGACTGTAGGCTAGAGAAGAGCCAATTAGTATTTGCGGCGCTCGGGAGCAGCGTCCATCCGGCGGGTGATAGGAA 560
-498 AATACTTCCGGCGTTCTGGTTTTTATGTGCTGAGTCACAGCCGAAGGCTCGCCCCGGGCTGAGGAGTAGGGTACCCAA 640
-418 ACTATGACACGAATGGGAAAATTTAAACTTTGGAATTAATTTCTCTGCACTCGTTTCTCAGTTTCTCATTAGGGGAATA 720
-338 TTTTCTTTTACACGCGGAGCACTACTCAGTGGTTTCCACAAAACACACTCGTACGTATAAACGCGCCGGGGGCCCTTC 800
-258 TCCCGTCGTCCTCCGCGCACTCTCTCCGCTTCTTCCCTCGCCCCGCCCACTTTGGGTTTCGAAATGTTGCTGGCTG 880
-178 TTCCGAGAATCAGTCCGGGTCACGTTGGTGCCTCTCCTTTCTGTCTGATTGGTTGCTTACGTGAGCGTGACTACTTTC 960
-98 TTGCCTCATGATTGGCTGAAATCAGTAGCGTCATCGCGCGCGGGAGGGCGGAGGACGAGGGAAGTCGCGGAGAGGGAG 1040
-18 ACTGGGTTGGGGGAGGGGTTTCAGGCCTGTCTCCCGCGGCTGCGGCAGCACCAGGGCCGGCCGCCCGCTCTAGAACGC 1120
+63 GGAGGAGGTGGGTCCTGGGAAGCGGATGTCTTTCGCTCCAGCTTGGTGGTGAATGCTGAGGAGAGTCACTGTTGCTGCA 1200
+143 GTCTGTGCCACCCGAGGAAGTTGTGTGAGGCCGGGCGGAGCTCGCGGCGCTCGGGGAATCGAAACGCGGGGTCGGTG 1280
+223 CGAAGACTCTGCTGCTGCCAGACCCTTTCCTATTCTGGCCATGCCTGGAAGGAACAAGGCGAAGTCTACCTGCAGTGCC 1360
+303 CTGACCTGCAGCCCAATGGACAGGATCTGGGCAGAACAGCCGGGTGCCCCGTAGGAGCGGATGAATCTGAGGAAGAG 1440
+383 GGACGGAGGGGCTCTCTCAGTAATGCTGGGACCCCTGAGATCGTCAAGTCTCC 1493

```

Figure 1 Promoter sequences of rat and human NRD-Cs

Rat (**A**) and human (**B**) NRD-C 5' upstream sequences are shown. Putative start codons are shown in bold, the major transcriptional start site determined by 5'-RACE is double underlined. Numbers on the right indicate the position in the GenomeWalker clones pNRD and pHNRD. Numbers on the left indicate the position relative to the transcription start site which is numbered +1. CpG islands are found in each sequence (boxed).

sites were identified in the same location in both sequences for a number of transcription factors and are shown in Figure 3. Several potential transcription-factor binding sites were also found in both species which are not conserved. Studies were

concentrated on conserved sequence stretches, since these are more likely to have significant function. A CpG island was identified in both the rat and human NRD-C promoter using the program XGAIL [38] and shown in Figure 1.

Figure 2 Comparison of transcriptional start sites of rat and human NRD convertase

The Figure shows multiple transcription initiation sites identified by sequencing PCR products obtained from 5'-RACE. Numbers at the sides of the sequences indicate the position relative to the promoter clones, pNRD and pHNRD. Arrows show the position of the 5' nucleotide of the PCR product along with the number of individual sequences which begin at that position.

The rat NRD-C 5' upstream region is a highly efficient promoter in GH3 and Mat-Lu cells

In order to functionally characterize the NRD-C 5' upstream region, the rat sequence was cloned into a firefly luciferase reporter vector (pGL3-basic; Promega) and the resulting construct named pRPb-1376. This was transfected into four cell lines, GH3 (ECACC# 87012603), a rat cell line derived from a pituitary tumour, Mat-Lu (ECACC# 94102735), a rat cell line derived from a dorsal prostate adenocarcinoma, GC-2spd(ts) (ATCC# CRL-2196), a mouse cell line derived from spermatocytes and COS-1 (ECACC# 88031701) which were subsequently assayed for luciferase activity. Compared with that of pGL3-control, a vector containing the simian-virus-40 (SV40) promoter and enhancer, high levels of expression from pRPb-1376 can be seen in GH3 and Mat-Lu cells and very low levels in GC-2spd(ts) and COS-1 cells (Figure 4). This correlates well with the NRD-C content of these cell lines determined by Western blotting (results not shown).

Using unique restriction-enzyme sites within the rat NRD-C promoter, a series of 5' deletions were constructed, containing NRD-C promoter sequence 1049 bp, 550 bp, 411 bp and 101 bp upstream of the transcriptional start site respectively. Figure 5 shows the effects of these gross 5' deletions on expression from the rat NRD-C promoter. On removal of the section of the promoter 5' to nucleotide -1049 in GH3 cells, expression significantly increases approximately twofold and remains at this level with deletion of nucleotides to -101 without any significant changes. In Mat-Lu cells, high expression levels equal to that of the full-length promoter are seen when sequence 5' of nucleotide -411 is removed, without any significant changes. Further removal of nucleotides up to -101 reduces expression fivefold, to a level approximately half that of the pGL3-control plasmid. In GC-2spd(ts) cells expression was very low, the full-length promoter producing less than 25% of the expression level of the pGL3-control plasmid, but significantly greater than the pGL3-basic plasmid ($P < 0.05$). No significant change in expression levels can be seen on deletion of sequence upstream of nucleotide -411. On further deletion up to -101, expression decreases to a level close to, and not significantly different from, that of the pGL3-basic plasmid. In COS-1 cells, expression of the full-length NRD-C promoter is extremely low compared with the pGL3-

control plasmid (less than 1%), due in part to abnormally high expression of the control vector. Despite this, expression from the full-length promoter is low in absolute terms compared with that in other cell lines. No significant difference in expression is observed on deletion of sequence 5' to nucleotide -550. However, deletion of sequence to nucleotide -411 causes a significant twofold increase in expression compared with the full-length promoter. Further deletion to nucleotide -101 decreases the expression tenfold to a level similar to that of pGL3-basic plasmid.

Nuclear proteins bind to conserved motifs in the NRD-C promoter

Since the results of the transfection assays above suggest that the -411 to -101 region of the rat NRD-C promoter plays a crucial role in determining cell-type-specific expression, the possibility that transcription factors bind to specific sequences within this region was investigated. Computer analysis (Figure 3) predicts several conserved potential transcription-factor binding sites within this region. To determine whether any of these sequences are targets for nuclear proteins, electrophoretic-mobility-shift assays were performed on double-stranded oligonucleotides corresponding to these sites, using nuclear protein extract from GH3 and Mat-Lu cells.

The results of this analysis are summarized in Figure 6a. Of the five potential binding sites, only three, SP-1, C/EBP and NF-Y, show a reproducible shift in mobility which can be blocked by addition of an excess of the corresponding unlabelled oligonucleotide. No difference was observed in the binding pattern of the two nuclear extracts. The extended conserved region immediately 3' to the putative IK-2 binding site (-246/-227) was examined for its ability to bind nuclear proteins. Both GH3 and Mat-Lu nuclear extract produced a strong shift in mobility when the oligonucleotide N_{-249/-226} was used (Figure 6b). This can be successfully competed for by an excess of non-radioactively labelled oligonucleotide or by N_{-236/-221}, which contains the conserved 3' portion, but not the IK-2 consensus. The oligonucleotide corresponding to the putative IK-2 site, which does not show a mobility shift (Figure 6a) was unable to compete for binding. N_{-236/-221} was able to produce a shift with the same mobility as N_{-249/-226} (results not shown).

DISCUSSION

Sequencing of the NRD-C promoter and 5'-UTR reveals that they are well conserved between rat and human. With the exception of an extra 29 bp segment in the rat sequence, the 5'-UTR of the NRD-C gene is 81.1% identical, almost as well conserved as the coding region of NRD-C, which has 87.7% identity in nucleotide sequence between the two species. This suggests that this sequence might be involved in regulating expression of this protein, either at the level of transcription, or, more probably, at the level of translation. Both the rat and human 5'-UTRs are extremely G+C-rich (71 and 64% respectively), which makes formation of secondary structures in mRNA more likely. It has been speculated that sequences that form stable secondary structures inhibit translation by providing a barrier to ribosomes [40]. The region around the transcriptional start site is also very well conserved. In a region of 26 bp around the start site (−15 to +11), there is only one nucleotide difference between the human and rat sequences. It is possible that this conservation is due to specific binding of a component of the basal transcription machinery.

Neither a TATAA box, nor sequence resembling a TATAA box, are present in the NRD-C promoter in the region one might be expected. However, a sequence very similar to a consensus for initiator elements as defined by Kollmar and Farnham [41] can be found overlapping the major transcriptional start site of the

Figure 4 Relative expression of the rat NRD-C promoter in four cell lines

The Figure shows expression from the plasmid pRPb-1376, containing the rat NRD-C promoter fragment to 1376 bp upstream of the transcription start site, as a percentage of that from the pGL3-control plasmid (set to 100%) in four cell lines determined by transient transfection followed by luciferase assay. The *Renilla*-luciferase-expressing plasmid pRL-CMV was co-transfected at a 1:10 (*Renilla*/test plasmid) ratio. Values were calculated from the means $\pm$ S.E.M. for six separate assays and are expressed as the ratio firefly luciferase/*Renilla* luciferase. The value for pRPb-1376 expressed in COS-1 cells was 0.85% of that of pGL3-control in this cell line.

Figure 5 Effect of deletions in the NRD-C promoter upon luciferase expression in four cell lines

The Figure shows luciferase expression from the NRD-C promoter/firefly luciferase constructs in GH3, Mat-Lu, GC-2spd(ts) and COS-1 cells determined by transient transfection followed by luciferase assay. The *Renilla*-luciferase-expressing plasmid pRL-CMV was co-transfected at a 1:10 (*Renilla*/test plasmid) ratio. Bars represent the means $\pm$ S.E.M. for six separate assays and are expressed as the ratio firefly luciferase/*Renilla* luciferase. Comparisons were made using unpaired two-tailed *t*-tests with significance set at $P < 0.05$. Asterisks indicate where expression is significantly different to pRPb-1376 expression. The *x*-axis (firefly luciferase/*Renilla* luciferase ratio) has been scaled appropriately to reflect the differing levels of expression seen in the four cell lines.

Figure 6 Comparison of electrophoretic-mobility-shift-assay patterns in the region -411/-101 of the rat NRD-C promoter

The Figure shows the results of an electrophoretic-mobility-shift assay on conserved motifs within the rat and human NRD-C promoter using nuclear protein extracts from Mat-Lu and GH3 cells. (A) Double-stranded oligonucleotides corresponding to predicted transcription-factor binding sites were ³²P labelled and analysed. —, No shift was observed; +, a mobility shift which could be competed for by an excess of the corresponding unlabelled double-stranded oligonucleotide was observed. (B) Analysis of the double-stranded oligonucleotide N_{-249/-226} (5'-AAACTTCTCCCGTCGTCCTCCGCC-3') using both GH3 and Mat-Lu nuclear extract. +, Addition of a 10-fold excess of the appropriate unlabelled double-stranded oligonucleotide. An arrow indicates the position of a retarded band.

NRD-C gene. In common with most other promoters that lack a TATA box, initiation of transcription in the NRD-C gene occurs at a number of different positions. According to the 5'-RACE data, the NRD-C gene initiates transcription imprecisely in a region spanning approx. 30 bp, and the location and spacing of initiation is conserved between the two species. In both species there appear to be no significant differences between the brain and testis in terms of transcriptional initiation site, despite the fact that, in both rat and human testes, the mRNA levels are the highest for all tissues examined ([14,19]; A. G. Winter and A. R. Pierotti, unpublished work). This is in contrast with many genes which are up-regulated or which have variants in the testis. For example, the angiotensin-converting-enzyme gene [42], the carbonic anhydrase II gene [43] and the *Drosophila* dihydro-orotate dehydrogenase gene [44] all use alternative testis-specific promoters.

Although a number of potential binding sites for transcription factors have been predicted, there are no consensus sequences for spermatid specific factors such as Tet-1 [45] and Zfy-1 [46] in this region of the promoter. Germ-cell-specific regulation may be in a region of the promoter further upstream of that investigated. mRNA stability may also play a role in maintaining the high

levels of expression seen in this tissue. Additionally, no retinoic acid regulatory elements, which Draoui et al. suggested may have been present [35], are predicted.

Four cell lines were chosen in which to examine the NRD-C promoter/luciferase constructs. As a model for germ-cell expression, the GC-2spd(ts) line, one of the few germ-cell lines available, was chosen. Human prostate expresses high levels of NRD-C mRNA (A. G. Winter and A. R. Pierotti, unpublished work), suggesting the use of the Mat-Lu prostate line. GH3 pituitary cells have previously been shown to contain high levels of NRD-C mRNA [20]. As a cell line in which low levels of expression would be expected, COS-1 cells were chosen [20].

The pRPb-1376 plasmid expresses luciferase at approximately equal levels in Mat-Lu and GH3 cells. This expression level is very high if compared with that of the pGL3-control plasmid (2.5–3.5-fold greater) in which luciferase expression is driven by the SV40 promoter and enhancer, which is known to give strong expression in many cell types. The expression from pGL3-control is approximately equal in the four cell lines tested (firefly/*Renilla* ≈ 0.5), except for COS-1, in which the expression level is artificially elevated due to replication of the plasmid by action of large T antigen on the SV40 origin of replication in the

SV40 promoter. Expression of pRPb-1376 in GC-2spd(ts) and COS-1 cells is very low. Given the low levels of expression observed in GC-2spd(ts) cells, this line may not be the ideal model in which to study germ-cell expression of NRD-C.

Expression from the NRD-C promoter increases twofold on removal of nucleotides –1376 to –1049 in GH3 cells, but not in other cell lines, suggesting that, in GH3 cells, full transcription from the NRD-C promoter is repressed by a *cis*-acting element within the –1376/–1049 region. Further removal of nucleotides from –1049 to –411 does not significantly change expression in any cell line. Removal of nucleotides from –411 to –101 massively reduces expression in Mat-Lu cells and nearly abolishes expression in COS-1 and GC-2spd(ts). However, this does not affect expression in GH3 cells in which it is still at a maximum. Although maximal expression levels from the NRD-C promoter is much less in GC-2spd(ts) cells and COS-1 cells than in Mat-Lu cells, these three cell lines have similar expression patterns on deletion from the 5' end of the promoter. GH3 cells have a qualitative difference in expression, retaining maximal levels on removal of nucleotides up to –101. This makes it likely that the all the *cis*-elements required for assembly of the general transcription complex in all cell types are located downstream of –101. It also suggests that the elements required for GH3-cell-specific up-regulation are located within this region. For full expression in Mat-Lu, GC-2spd(ts) and COS-1 cells, (a) *cis*-element(s) is (are) required within nucleotides –411 to –101.

The section of the rat NRD-C promoter from –411 to –101 was targeted for further study in order to determine which factors are involved in the difference in expression between Mat-Lu and GH3 cells. This section contains short stretches of 100% identity between the rat and human promoters, which correspond to potential transcription-factor binding sites for Oct-1, IK-2, SP1, C/EBP and NF-Y. Binding studies were performed to determine if any of these sequences are targets for transcription factors. Additional stretches of 100% identity in this region, for which no potential transcription-factor binding sites were predicted, may also be important functional features of this promoter. The sequence N_{–249/–226} was also tested for its ability to bind nuclear proteins.

The mobility-shift-assay experiments demonstrate that nuclear proteins from both GH3 and Mat-Lu cells bind to the predicted SP1, C/EBP and NF-Y sites, suggesting that these transcription factors might be important for expression of NRD-C in the cell lines tested. Supershift experiments using antibodies against these proteins will be necessary to determine if these are the transcription factors that bind to these sites. Nuclear protein binding to the conserved sequence N_{–249/–226} shows that there are also transcription-factor binding sites within this region that are not predicted by MatInspector. These binding studies do not reveal any cell-line-specific differences that might account for the expression pattern seen in the reporter-gene assay. An additional series of deletion constructs and further analysis of protein binding by mobility-shift and DNase I-footprinting analysis will pinpoint this difference.

This work was supported by the Glasgow Caledonian University Research Development Fund.

REFERENCES

- Docherty, K. and Steiner, D. F. (1982) Post-translational proteolysis in polypeptide hormone biosynthesis. *Annu. Rev. Physiol.* **44**, 625–638
- Darby, N. J. and Smyth, D. G. (1990) Endopeptidases and prohormone processing. *Biosci. Rep.* **10**, 1–13
- Seidah, N. G. and Chretien, M. (1999) Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive peptides. *Brain Res.* **848**, 45–62
- Bourdais, J., Devilleirs, G., Girard, R., Morel, A., Benedetti, L. and Cohen, P. (1990) Prosomatostatin II processing is initiated in the *trans*-Golgi network of anglerfish pancreatic cells. *Biochem. Biophys. Res. Commun.* **170**, 1263–1272
- Schnabel, E., Mains, R. E. and Farquhar, M. G. (1989) Proteolytic processing of pro-ACTH/endorphin begins in the Golgi complex of pituitary corticotropes and AT-20 cells. *Mol. Endocrinol.* **3**, 1223–1235
- Orci, L., Ravazzola, M., Storch, M.-J., Anderson, R. G. W., Vassalli, J.-D. and Perrelet, A. (1987) Proteolytic maturation of insulin is a post-Golgi event which occurs in acidifying clathrin-coated secretory vesicles. *Cell* **49**, 865–868
- Thibault, G., Lazure, C., Schiffrin, E. L., Gutkowska, J., Chartier, L., Garcia, R., Seidah, N. G., Chretien, M., Genest, J. and Cantin, M. (1985) Identification of a biologically active circulating form of rat atrial natriuretic factor. *Biochem. Biophys. Res. Commun.* **130**, 981–986
- Yanagisawa, M., Kurihara, H., Kimura, S., Tomobe, Y., Kobayashi, M., Mitsui, Y., Yazaki, Y., Goto, K. and Masaki, T. (1988) A novel potent vasoconstrictor peptide produced by vascular endothelial cells. *Nature (London)* **332**, 411–415
- Julius, D., Brake, A., Blair, L., Kunisawa, R. and Thorner, J. (1984) Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast pre-pro- α -factor. *Cell* **37**, 1075–1089
- Mizuno, K., Nakamura, T., Ohshima, T., Tanaka, S. and Matsuo, H. (1988) Yeast *kex2* gene encodes an endopeptidase homologous to subtilisin-like serine proteases. *Biochem. Biophys. Res. Commun.* **156**, 246–254
- Steiner, D. F. (1998) The proprotein convertases. *Curr. Opin. Chem. Biol.* **2**, 31–39
- Seidah, N. G., Day, R., Hamelin, J., Gaspar, A., Collard, M. W. and Chretien, M. (1992) Testicular expression of PC4 in the rat: molecular diversity of a novel germ cell-specific *kex2*/subtilisin-like proprotein convertase. *Mol. Endocrinol.* **6**, 1559–1570
- Chesneau, V., Pierotti, A. R., Barre, N., Creminon, C., Tougaard, C. and Cohen, P. (1994) Isolation and characterization of a dibasic selective metalloendopeptidase from rat testes that cleaves at the amino terminus of arginine residues. *J. Biol. Chem.* **269**, 2056–2061
- Pierotti, A. R., Prat, A., Chesneau, V., Gaudoux, F., Leseney, A.-M., Foulon, T. and Cohen, P. (1994) *N*-Arginine dibasic convertase, a metalloendopeptidase as a prototype of a class of processing enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6078–6082
- Chesneau, V., Prat, A., Segretain, D., Hospital, V., Dupaix, A., Foulon, T., Jegou, B. and Cohen, P. (1996) NRD convertase: a putative processing endoprotease associated with the axoneme and the manchette in late spermatids. *J. Cell Sci.* **109**, 2737–2745
- Gluschkof, P., Morel, A., Gomez, S., Nicolas, P., Fahy, C. and Cohen, P. (1984) Enzymes processing somatostatin precursors: an Arg-Lys esteropeptidase from the rat brain cortex converting somatostatin-28 into somatostatin-14. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6662–6666
- Gomez, S., Gluschkof, P., Morel, A. and Cohen, P. (1985) The somatostatin-28 convertase of rat brain cortex is associated with secretory granule membranes. *J. Biol. Chem.* **260**, 10541–10545
- Csuhai, E., Chen, G. and Hersch, L. B. (1998) Regulation of *N*-arginine dibasic convertase activity by amines: putative role of a novel acidic domain as an amine binding site. *Biochemistry* **37**, 3787–3794
- Fumagalli, P., Accarino, P., Egeo, A., Scartezzini, P., Rappazzo, G., Pizzuto, A., Avantaggiato, V., Simeone, A., Arrigo, G. and Zuffardi, O. et al. (1998) Human NRD convertase: A highly conserved metalloendopeptidase expressed at specific sites during development and in adult tissues. *Genomics* **47**, 238–245
- Hospital, V., Prat, A., Joulie, C., Cherif, D., Day, R. and Cohen, P. (1997) Human and rat testis express two mRNA species encoding variants of NRD convertase, a metalloendoprotease of the insulinase family. *Biochem. J.* **327**, 773–779
- Rawlings, N. D. and Barrett, A. J. (1993) Evolutionary families of peptidases. *Biochem. J.* **290**, 205–218
- Rawlings, N. D. and Barrett, A. J. (1995) *Methods Enzymol.* **248**, 183–228
- Duckworth, W. C., Hamel, F. G., Peavy, D. E., Liepnieks, J. J., Ryan, M. P., Hermanson, M. A. and Frank, B. F. (1988) Degradation products of insulin generated by hepatocytes and by insulin protease. *J. Biol. Chem.* **263**, 1826–1833
- Hamel, F. G., Posner, B. I., Bergeron, J. J. M., Frank, B. H. and Duckworth, W. C. (1988) Isolation of insulin degradation products from endosomes derived from intact rat liver. *J. Biol. Chem.* **263**, 6703–6708
- Shii, K., Yokono, K., Baba, S. and Roth, R. A. (1986) Purification and characterization of insulin-degrading enzyme from human erythrocytes. *Diabetes* **35**, 675–683
- Roth, R. A., Mesirow, M. L., Yokono, K. and Baba, S. (1984) Degradation of insulin-like growth factors I and II by a human insulin degrading enzyme. *Endocr. Res.* **10**, 101–112
- Garcia, J. V., Gehm, B. D. and Rosner, M. R. (1989) An evolutionary conserved enzyme degrades transforming growth factor- α as well as insulin. *J. Cell Biol.* **109**, 1301–1307

- 28 Kirschner, R. J. and Goldberg, A. L. (1983) A high molecular weight metalloendoprotease from the cytosol of mammalian cells. *J. Biol. Chem.* **258**, 967–976
- 29 Muller, D., Schulze, C., Baumeister, H., Buck, F. and Richter, D. (1992) Rat insulin-degrading enzyme: cleavage pattern of the natriuretic peptide hormones ANP, BNP, and CNP revealed by HPLC and mass spectrometry. *Biochemistry* **31**, 11138–11143
- 30 Safavi, A., Miller, B. C., Cottam, L. and Hersh, L. B. (1996) Identification of gamma-endorphin-generating enzyme as insulin-degrading enzyme. *Biochemistry* **35**, 14318–14325
- 31 Kuo, W.-L., Montag, A. G. and Rosnes, M. R. (1993) Insulin-degrading enzyme is differentially expressed and developmentally regulated in various rat tissues. *Endocrinology (Baltimore)* **132**, 604–611
- 32 Mzhavia, N., Berman, Y. L., Qian, Y., Yan, L. and Devi, L. A. (1999) Cloning, expression, and characterization of human metalloprotease 1: a novel member of the pitrilysin family of metalloproteases. *DNA Cell Biol.* **18**, 369–380
- 33 Adames, N., Blundell, K., Ashby, M. N. and Boone, C. (1995) Role of yeast insulin-degrading enzyme homologs in propheromone processing and bud site selection. *Science* **270**, 464–466
- 34 Cshui, E., Juliano, M. A., Juliano, L. and Hersh, L. B. (1999) Kinetic analysis of spermine binding to NRD convertase. *Arch. Biochem. Biophys.* **362**, 291–300
- 35 Draoui, M., Bellincampi, L., Hospital, V., Cadel, S., Foulon, T., Prat, A., Barre, N., Reichart, U., Melino, G. and Cohen, P. (1997) Expression and retinoid modulation of N-arginine dibasic convertase and an aminopeptidase-B in human neuroblastoma cell lines. *J. Neuro-Oncol.* **31**, 99–106
- 36 Myers, E. W. and Miller, W. (1988) Optimal alignments in linear space. *Comput. Appl. Biosci.* **4**, 11–17
- 37 Quandt, K., Frech, K. and Karas, H. (1995) MatInd and MatInspector-New fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23**, 4878–4884
- 38 Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282
- 39 Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72**, 248–254
- 40 Sonenberg, N. (1996) mRNA cap-binding protein eIF4E and control of cell growth. In *Translational Control* (Hershey, J. W. B., Mathews, M. B. and Sonenberg, N., eds.), pp. 245–269, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- 41 Kollmar, R. and Farnham, P. J. (1993) Site-specific initiation of transcription by RNA polymerase-II. *Proc. Soc. Exp. Biol. Med.* **203**, 127–139
- 42 Howard, T. E., Shai, S.-Y., Langford, K. G., Martin, B. M. and Bernstein, K. E. (1990) Transcription of testicular angiotensin-converting enzyme (ACE) is initiated within the 12th intron of the somatic ACE gene. *Mol. Cell. Biol.* **10**, 4294–4302
- 43 Mezquita, J., Pau, M. and Mezquita, C. (1994) A novel carbonic anhydrase-II messenger RNA isolated from mature chicken testis displays a TATA box and other promoter sequences in a leader 5′ untranslated region not present in somatic tissues. *Gene* **147**, 231–235
- 44 Rawls, J., Kirkpatrick, R., Yang, J. and Lacy, L. (1993) The *dhod* gene and deduced structure of mitochondrial dihydroorotate dehydrogenase in *Drosophila melanogaster*. *Gene* **124**, 191–197
- 45 Tamura, T., Makino, Y., Mikoshiba, K. and Muramatsu, M. (1992) Demonstration of a testis-specific *trans*-acting factor Tet-1 *in vitro* that binds to the promoter of the mouse protamine 1 gene. *J. Biol. Chem.* **267**, 4327–4332
- 46 Zambrowicz, B. P., Zimmermann, J. W., Harendza, C. J., Simpson, E. M., Page, D. C., Brinster, R. L. and Palmiter, R. D. (1994) Expression of a mouse Zfy-1/lacZ transgene in the somatic cells of the embryonal gonad and germ cells of the adult testis. *Development* **120**, 1549–1559

Received 31 May 2000/4 August 2000; accepted 22 August 2000