Isolation and characterization of the rat *huntingtin* promoter

Carsten HOLZMANN*, Winfried MÄUELER*, Dirk PETERSOHN†¹, Thorsten SCHMIDT*, Gerald THIEL†², Jörg T. EPPLEN* and Olaf RIESS*³

*Department of Molecular Human Genetics, Ruhr-University Bochum, D-44780 Bochum, Germany, and †Institute for Genetics, University of Cologne, D-50674 Cologne, Germany

Huntington's disease (HD) is a neurodegenerative disorder caused by a (CAG)_{>37} repeat expansion in a novel gene of unknown function. Although the *huntingtin* gene is expressed in neuronal and non-neuronal tissues, the disease affects nerve cells of selected regional areas of the central nervous system. To gain insight into the regulation of the *HD* gene we analysed 1348 bp of the rat *huntingtin* promoter region. This region lacks a TATA and a CAAT box, is rich in GC content and has several consensus sequences for binding sites for SP1, PEA3, Sif and H2A. The stretch between nucleotides –56 and –206 relative to the first ATG is highly conserved between human and rodents and it harbours several potential binding sites for transcription

factors. We analysed deletion mutants fused with the chloramphenicol acetyltransferase reporter gene in transfected, HD-expressing neuronal (NS20Y, NG108-15) and non-neuronal Chinese hamster ovary cell lines. Hence these cells should contain the required *trans*-acting factors necessary for *HD* gene expression. Partial deletion of the evolutionarily conserved part of the promoter significantly decreases the activity in both neuronal and non-neuronal cells, indicating that the core promoter activity is located between nucleotides –332 and –15. DNase I footprinting and electrophoretic mobility-shift assays were used to define the nucleotide positions and binding affinity of DNA–protein interactions.

INTRODUCTION

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder usually affecting human individuals in their fourth decade. A (CAG)_n repeat has been identified in a novel gene of unknown function that is expanded and unstable in HD-affected individuals [1]. The mechanisms underlying region-specific neuronal cell death in HD are still obscure. This is in part due to the complexity of the nervous system, which comprises various cell types that express different sets of genes. However, no information exists on the normal function of the *HD* gene product, huntingtin. However, on the basis of immunohistochemical data it has been suggested that huntingtin might have a role in vesicle trafficking [2]. Whatever the physiological role of HD might be, its normal function is critical to the embryonal development of higher organisms. This is suggested by its extreme conservation across animal species and by the finding that the inactivation of the *HD* gene in mice by gene targeting causes fetal death until embryonal day 8.5 [3–5]. Together these results indicate that HD is caused by a gain-of-function mutation rather than by loss-of-function [6].

Protein studies provide evidence that the (CAG)_n repeat is indeed translated and results in a polyglutamine stretch that is expanded in HD patients [7]. Recently it has been shown that expanded polyglutamine stretches form amyloid-like aggregates *in vitro* [8]. Finally, these aggregates have also been found in a transgenic mouse model for HD [9] and in the brain of HD patients [10]. It is now generally accepted that neuronal intranuclear inclusions are a pathological hallmark of polyglutamine disorders [11], as these inclusions have also been found in spinocerebellar ataxias of types 1 and 3 [12–14]. The expression

of the mutated genes in each disease is widespread [15–17] and does not necessarily reflect the sites of neuronal cell death [6]. Furthermore, the level of expression seems to be developmentally regulated, as we have previously shown for the rat *HD* mRNA [18]. The *HD* mRNA in rodent brain is readily detected at embryonal day 12.5 at similar levels to that found during postnatal development [18]. In postnatal non-neuronal tissues, however, the *HD* mRNA decreases markedly except in testis.

To gain insight into the functional regulation of the *HD* gene, we cloned and sequenced the putative promoter region of the rat *HD* gene and studied its capacity to drive the expression of a chloramphenicol acetyltransferase (*CAT*) reporter gene. We identified a highly conserved region between positions –206 and –56, showing 76 % nucleotide identity between human and rat, and 92 % identity between mouse and rat [19]. Sequence analyses, protein mobility-shift and footprint assays identified putative transcription-factor-binding sites, most of which resemble conserved sequences found in the human *HD* promoter [20]. In contrast with functional analyses of the human *HD* gene promoter [20], our results did not reveal differences in the control of reporter gene expression between neuronal and non-neuronal cell lines. This study adds to our understanding of expression regulation of a gene which is, in its mutated form, causative for neurodegeneration in humans.

MATERIALS AND METHODS

Isolation of genomic clones for the 5' region of the *rh*d gene

To clone the promoter region of the *rh*d gene, a genomic cosmid library (male Wistar spleen, subcloned in pWE15, provided by

Abbreviations used: CAT, chloramphenicol acetyltransferase; CHO, Chinese hamster ovary; CMV, cytomegalovirus; CRE, cAMP response element; DTT, dithiothreitol; EMSA, electrophoretic mobility-shift assay; HD, Huntington's disease; RSV, Rous sarcoma virus.

¹ Present address: Department of Dermatology, University of Cologne, Joseph-Stelzmann Strasse 9, D-50931 Cologne, Germany.

² Present address: Medizinische Biochemie, Universität des Saarlandes, Geb. 44, D-66421 Homburg, Germany.

³ To whom correspondence should be addressed (e-mail olaf.riess@ruhr-uni-bochum.de).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AJ224997.

Dr. F. Laccone) was screened with a mixture of two oligonucleotides detecting the first exon of the known *rhod* cDNA sequence [18]. A 2368 bp *Hind*III fragment of cosmid RGC15.2.1.8 was cloned into pBluescript (GRHD1) and then sequenced.

Reporter gene constructs

PCR products of GRHD1 containing the sequences upstream of the putative translation start codon were subcloned upstream of a promoterless *CAT* gene in pCAT-basic (Promega). Deletions were made by digesting the full-length construct with *Sph*I, *Sac*I, *Eag*I, *Pst*I and *Apa*I cutting inside the *rhod* promoter region followed by the generation of blunt ends and religation. As control plasmids, a construct containing the *CAT* gene driven by the 3' long terminal repeat of the Rous sarcoma virus (RSV 3' LTR) [21] and pSyICAT-10 containing the *synapsin I* 5' regulatory region (−2308 to +47) [22] were used.

RNA isolation

Total cellular RNA was isolated from rat tissue by the method of Chirgwin et al. [23]. Poly(A)⁺ RNA was obtained by oligo(dT)–cellulose affinity chromatography with the Fast Track 2.0 Kit (Invitrogen).

Transcription start site mapping

The following primers were used for transcription start site mapping: RHD10 T3.1, TTGCTGAACTCGCTGAGCTTTC-GGAAGTAGTCAAAAAG; GRHD1 T3.2, GCTTGGAGCC-TATTCGCACTACGCAGCGCC; GRHD1 T3.3, GCCTCCT-GCCCGACAGGACAGACCCTGAAG; and GRHD1 T3.4, CGTCTCTTTTCTCAACGTTTGGTCGTGGGC.

Primer extension analyses were performed with 20 µg of total RNA from brain by the method of Sambrook et al. [24]. A ³²P-end-labelled oligomer (0.5 pmol) (RHD10 T3.1, GRHD1 T3.2, GRHD1 T3.3, GRHD1 T3.4) corresponding to the *rhod* gene in anti-sense orientation was hybridized to 20 µg of total RNA from rat brain, liver, testis, muscle, heart, kidney, lung and thymus by incubation at 30–38 °C in 30 µl of 40 mM Pipes (pH 6.4)/1 mM EDTA/0.4 M NaCl/80 % formamide for at least 16 h. After precipitation with ethanol, reverse transcription was done with 50 units of 'Super-Script' reverse transcriptase (Gibco-BRL) in accordance with the manufacturer's instructions. After digestion of the RNA with 1 µl of DNase-free RNase A, phenol/chloroform extraction was performed, followed by precipitation with ethanol. The nucleic acids were then run on denaturing 8 M urea/6 % (w/v) polyacrylamide gels.

For S1-nuclease protection assay the Ambion S1-Assay kit was used with poly(A)⁺ RNA prepared with the Fast Track 2.0 Kit (Invitrogen). To synthesize a single-strand DNA probe complementary to the murine *hd* mRNA, a *Bam*HI/*Xho*I fragment of the genomic clone GRHD1 was inserted into M13 mp18. By using the primers RHD10 T3.1 and GRHD1 T3.3, a primer extension reaction was performed in the presence of radiolabelled dATP. The extension product was digested with *Sac*I and purified on a denaturing polyacrylamide gel. Approx. 1.2×10^5 c.p.m. of probe with a specific radioactivity of 5×10^9 c.p.m./µg (approx. 250 pg or 2.4 fmol) was co-precipitated with 0.5 µg of poly(A)⁺ RNA from rat brain and 9.5 µg of yeast RNA, then hybridized overnight at 42 °C. Nucleic acids were then digested with 50–500 units of S1 nuclease (Ambion) for 1 h at 37 °C, and after ethanol precipitation they were subjected to electrophoresis on denaturing 8 M urea/6 % (w/v) polyacrylamide gels.

Cell culture

Chinese hamster ovary cells (CHO-K1), the neuroblastoma/glioma fusion cell line NG108-15 and the neuroblastoma cells NS20Y were cultured as described [25,26].

Transfections and reporter gene expression studies

NG108-15, NS20Y and CHO cells were transfected by the calcium phosphate co-precipitation procedure [25]. Cells were incubated with calcium phosphate/DNA precipitate containing 8 µg of CAT-plasmid and 2 µg of pCMV-β (in which CMV stands for cytomegalovirus) (Clontech) for 6 h, followed by a glycerol shock of 2 min [15 % (v/v) glycerol in Dulbecco's modified Eagle's medium]. Cells were harvested after 48–60 h and lysed by three cycles of freeze–thawing. The cellular debris was removed by centrifugation, and the supernatant was used to measure β-galactosidase activity as described [24]. The remaining cell extract was heated for 10 min at 57 °C and then used to measure CAT activity. Cell extract (0.1–56 µl) was incubated for 1 h at 37 °C with 20 µl of 4 mM butyryl-CoA and 0.1 µCi of [¹⁴C]chloramphenicol in a total volume of 130 µl. The reaction products were extracted three times with xylene and the incorporated radioactivity was quantified. To normalize for variations of transfection efficiencies, the CAT activity was divided for the β-galactosidase activity. Every construct was tested twice (totalling eight transfections) in four independent transfection experiments for every cell line.

Preparation of nuclear extracts

This was performed by the method of Altschmied et al. [27] containing all total (NH₄)₂SO₄-precipitable proteins (total nuclear extract). Monolayer cells were washed and harvested in 2 ml of PBS per 20 cm dish. After centrifugation for 5 min at 200 g and 0 °C the cell pellets were resuspended in 50 ml of PBS counted and centrifuged again as above. The pellet was resuspended in 5 packed cell vol. of buffer A [10 mM Hepes/KOH (pH 7.6)/10 mM KCl/0.15 mM spermine/HCl/0.5 mM spermidine/1 mM EDTA/0.5 mM dithiothreitol (DTT)] and incubated on ice for 5–10 min. Cells were harvested by centrifugation at 800 g for 5 min at 0 °C, resuspended in 2 packed cell vol. of buffer A and broken by four to ten strokes with a Dounce homogenizer with an S pestle. When approx. 90 % of the cells had been broken, 0.1 vol. of buffer A-iso [10 mM Hepes/KOH (pH 7.6)/100 mM KCl/0.15 mM spermine/HCl/0.5 mM spermidine/1 mM EDTA/0.5 mM DTT] were added and mixed by gentle inversion. The nuclei of 10⁹ cells were pelleted at 1100 g for 10 min at 0 °C, resuspended in 2.5 ml in nuclear lysis buffer [10 mM Hepes/KOH (pH 7.6)/100 mM KCl/3 mM MgCl₂/0.1 mM EDTA/10 % (v/v) glycerol/1 mM DTT/0.1 mM PMSF] and homogenized by 5–10 strokes in a Dounce homogenizer with an L (loose) pestle. Nuclei were placed in a ice–water bath on a magnetic stirrer, lysed by slowly adding (dropwise) 0.1 vol. of 4 M (NH₄)₂SO₄ and stirred for 30 min. Proteins were harvested by ultracentrifugation at 90000 g for 40 min at 0 °C, resuspended in 0.1 vol. of dialysis buffer [25 mM Hepes/KOH (pH 7.6)/40 mM KCl/0.1 mM EDTA/1 % (v/v) glycerol/1 mM DTT] and dialysed in dialysis tubing for 30 min, 1 h and 1.5 h successively against 300 ml of fresh dialysis buffer. Protein concentrations were determined by the method of Bradford [28].

DNase I footprinting

For footprint analyses, fragments (GRHD-F1, nt −580 to −15; GRHD-F2, nt −329 to −15; GRHD-F3, nt −580 to

–330; GRHD-F4, nt –103 to 176) of the *rh*d promoter were subcloned into pUC19, so that the double-stranded probe could be labelled either at the *Eco*RI site or at the *Hind*III site of the polylinker. Assays with total CHO or NS20Y nuclear extracts were performed essentially as described in Altschmied et al. [27] by using 30000 c.p.m. of labelled target DNA per lane and 0.04–0.16 unit of DNase I (RNase-free; Boehringer Mannheim) without protein and 0.5–2.5 units of DNase I with 30–50 μ g of total CHO or NS20Y nuclear extract. DNA fragments were separated in denaturing 8 M urea/6 % (w/v) polyacrylamide gels. Gels were fixed [10 % (v/v) acetic acid/12 % (v/v) methanol], dried and exposed to Fuji RX X-ray films.

Electrophoretic mobility-shift assay (EMSA)

For EMSA, 1 μ g of poly(dI-dC) was mixed with 6–7.5 μ g of CHO or NS20Y total nuclear extract, 4 μ l of 5 $\times$ incubation buffer [1 $\times$ incubation buffer is 5 % (v/v) glycerol/20 mM Hepes/KOH (pH 7.6)/20 mM KCl/5 mM MgCl₂] and 8.5–10.5 μ l of TE buffer [1 mM Tris/HCl (pH 8.0)/10 mM EDTA] and pre-incubated for 20 min on ice. Labelled target DNA (2–4 μ l) (–210 to –139 and –138 to –15, 20 μ l total reaction volume) was added and the mixture was again incubated for 20 min on ice. After the addition of 10 $\times$ loading buffer [0.4 % Bromophenol Blue/0.4 % Xylene Cyanol/25 % (v/v) Ficoll], probes were loaded directly on precooled and pre-run native 5 % (w/v) polyacrylamide gels with TG buffer [25 mM Tris/HCl (pH 8.3)/192 mM glycine]. Gels were run for 3–4 h at 350 V, fixed, dried and exposed overnight to Fuji RX X-ray films.

Competition experiments were performed with constant amounts of labelled target DNA (approx. 10 fmol; 30000 c.p.m.) and 5 μ g of total CHO nuclear extract by including up to 250-fold molar excess of unlabelled competitor DNA. Radioactive bands were cut out from the gels and scintillation-counted. The steady-state affinities were evaluated by the method of Scatchard [29].

Preparation of protein extracts from cultured cells

Cells were washed with PBS, scraped from the culture dish and incubated for 15 min in TNES-APLV [50 mM Tris/HCl (pH 7.5)/1 % (v/v) Nonidet P40/2 mM EDTA/100 mM NaCl/10 mg/ml aprotinin/0.1 M PMSF/10 mg/ml leupeptin/0.1 M sodium orthovanadate]. Debris was removed by centrifugation and the supernatant was adjusted to 10 % (v/v) glycerol, then stored at –70 °C.

SDS/PAGE and immunoblot analysis

Proteins were assayed with a Bio-Rad protein assay by the method of Bradford [28] and subjected to SDS/PAGE [4 % (w/v) gel] (30 μ g of protein in each lane) with the discontinuous buffer system described by Laemmli [30]. Prestained high-molecular-mass markers (Amersham) were run in adjacent lanes. Proteins were transferred electrophoretically to nitrocellulose membranes (Schleicher & Schuell) as described [31]. Membranes were blocked overnight at 4 °C in TBST [10 mM Tris/HCl (pH 7.5)/150 mM NaCl/0.1 % (v/v) Tween 20] containing 5 % (w/v) non-fat dried milk. Antibody AP 78 (kindly provided by C. Ross), recognizing residues 1–17 of the human and rodent huntingtin protein, was used at a dilution of 1:2000 in PBS/3 % (w/v) BSA. After incubation for 2 h at room temperature, membranes were washed once for 15 min and three times for 5 min in PBS/5 % (w/v) dried milk. Blots were incubated with secondary antibody (goat anti-rabbit conjugated with horse-

radish peroxidase; Amersham) at 1:1000 dilution in PBS/3 % (w/v) BSA for 75 min at room temperature. Subsequently, membranes were washed once for 15 min and twice for 5 min in PBS/5 % (w/v) dried milk, then twice for 5 min in PBS/0.1 % (v/v) Tween 20. Antibodies were detected with an enhanced chemiluminescence kit as directed by the manufacturer (Amersham) and the blots were exposed to Hyperfilm (Amersham).

RESULTS

DNA sequence of the upstream *rh*d gene region

Screening a rat cosmid genomic library yielded one clone of approx. 35 kb that contained exon 1 of the *rh*d gene. A 2368 bp *Hind*III fragment (GRHD1) containing exon 1 was subsequently subcloned and sequenced. The rat DNA sequences, numbered relative to the translation start codon, encompassed 1348 bp upstream of the first ATG. A striking feature of the upstream sequence was a pyrimidine stretch 44 bp in length from positions –1147 to –1104. Further inspection of the upstream region revealed DNA sequence motifs that were putative binding sites for transcription factors (Figure 1). For instance, four GC boxes (GGGCGG) indicative of SP1-binding sites were found between positions –297 and –406. A cAMP response element (CRE; TGACGTCA) is located at position –175 bp, which the respective human sequence lacks. However, the CRE consensus sequence is part of a 150 bp segment (located between positions –56 and –206) that is otherwise highly conserved in human, mouse and rat [19,20]. Furthermore several binding sites for

Figure 1 Location of putative transcription-factor-binding sites and DNase I-protected sites in the *rh*d promoter

DNase I-protected sites are shaded and weakly protected sites are lightly shaded. The region highly conserved between rat and human is displayed in italics, and hypersensitive (HS) sites are shown in bold. Putative transcription-factor-binding sites are boxed, and the putative translation start codon is underlined.

Figure 2 Western-blot analysis of neuronal (NG108-15 and NS20Y) and non-neuronal (CHO) cells by using huntingtin antibody AP78

Huntingtin protein of approx. 320 kDa was detected in each of the cell lines. The different sizes of the protein in the cell lines are due to length differences in the polymorphic translated CAG repeat.

Nkx-2.5, NF κ B, H2A, PEA3 and CdxA were present in the rat and human sequences (Figure 1).

Dissection of regulatory elements in the promoter region

To test for the presence of *cis*-regulatory elements, we analysed the transcriptional activity of the *rhdl* 5' region in various cell lines. First, we investigated whether the cell lines express huntingtin and are therefore capable of supporting HD transcription. All of the analysed neuronal (mouse neuro-

blastoma cell line NS20Y and mouse neuroblastoma $\times$ rat glioma hybrid cell line NG108-15) and non-neuronal (CHO) cell lines expressed huntingtin protein (Figure 2). We then generated a series of deletion constructs from the *rhdl* 5' sequence, linked them to the *CAT* reporter gene and transfected them into the NS20Y, NG108-15 and CHO cells. A construct containing the 3' long terminal repeat of RSV (RSV 3' LTR) [21] was used as the reference construct. In addition, a *CAT* construct containing the human *synapsin I* promoter (pSyICAT-10) [22] was used as control for neuronal expression. Transfections of the two controls were invariably performed in parallel with the deletion constructs. The functional activity of the constructs was assessed 48 h after transfection by determining the levels of *CAT* activity. To correct for differences due to transfection efficiency, *CAT* activity was normalized to co-transfected β -galactosidase activity.

Figure 3 shows a histogram outlining the relative *CAT* activities of the deletion constructs relative to the activity of RSV-*CAT* [21]. All *rhdl* promoter deletion constructs revealed significantly higher expression activities than the *synapsin I* construct pSyICAT-10 [22]. Comparison of the expression of the pSyICAT-10 construct between neuronal (NG108-15 and NS20Y) and non-neuronal (CHO) cells revealed, as expected, low expression in CHO cells, confirming the neuron specificity of the human *synapsin I* promoter. The *CAT* gene driven by the *rhdl* gene promoter, however, was expressed at similar levels in CHO cells and in neuronal cells. Careful exploitation of the *rhdl* *CAT* deletion constructs in CHO cells revealed similar expression levels in constructs containing a region between -215 bp (R-CAT5) and -1348 bp (R-CAT1). Deletion of part of the highly conserved region (R-CAT6) decreased the expression of the reporter gene in neuronal as well as in non-neuronal cells. We therefore conclude that the highly conserved region in its entirety functions as a *cis*-acting promoter element necessary to drive the basic expression of the *rhdl* gene.

Figure 3 Activity of the *rhdl* gene promoter in transiently transfected cells

Two neuronal cell lines (NG108-15 and NS20Y) and one non-neuronal cell line (CHO) were transfected with chimaeric *rhdl* promoter constructs. Relative *CAT* activity is expressed after correcting for the transfection efficiency with co-transfected pCMV- β . Each result is the mean $\pm$ S.E.M. for four independent transfection experiments. Every cell line was assayed twice with every construct. Activities are displayed as percentages of pRSV-*CAT* activity. R-CAT1 contained the PCR-cloned promoter region of the rat *huntingtin* gene (*rhdl*) in front of the *Escherichia coli* *CAT* gene. All the other *CAT* constructs were deletion derivatives of R-CAT1. The highly conserved region is shaded and some of the potential transcription-factor-binding sites are shown as ellipses. Filled ellipses mark sites that are conserved in human, rat and mouse. The (Py)₄₄ stretch is shown as a darkly shaded region.

Figure 4 DNase I footprint analysis of the highly conserved region of the *rhd* gene promoter

Labelled target fragment (100000 c.p.m.) and variable amounts of total CHO nuclear extract were incubated with DNase I in each sample. (A) Constructs used in DNase I footprinting. The *EcoRI* and *HindIII* restriction sites were used for end labelling. (B) DNase I-protected site of the *rhd* gene promoter, which includes the consensus sequence of a CRE as determined by analysing construct GRHD-F1 (sense strand). (C) Analysis of fragment GRHD-F2 (anti-sense strand); several DNase I-protected sites and hypersensitive sites in a region containing two possible SP1-binding sites and binding sites for CdxA and Nkx-2.5 were detected. All detected DNase I-protected sequences of the *rhd* gene promoter are combined in Figure 1.

Figure 5 CHO and NS20Y nuclear proteins bind to rat *huntingtin* promoter sequences with high affinity

(A) The two restriction fragments of the putative promoter region and restriction fragments used in EMSA and competition experiments. (B) Competition for protein binding of total CHO nuclear extract to a 123 bp target DNA (fragment II, positions -15 to -138) from the highly conserved region of the *rhdt* gene promoter by using the same fragment and unrelated fragment I (positions -139 to -211) as unlabelled competitor DNA. (C) The same experiment as in (B) but with NS20Y nuclear extract instead of CHO extract. (D) Mobility shift and competition assays analysing fragment I by specific competition with unlabelled fragment I and by cross-competition with unlabelled fragment II by using nuclear extract of NS20Y cells.

Multiple protein-binding domains in the 5'-flanking region of the *rhdt* gene

To identify potential regulatory elements, we systematically performed DNase I footprinting analysis of four DNA fragments of the *rhdt* 5'-upstream region by using CHO nuclear extracts (Figure 4A). Seven major protein-binding domains were identified in the *rhdt* 5'-flanking region (Figures 1, 4B and 4C). The boundaries of the domains on both strands are marked in Figure 1. Several of the DNase I-protected sites do overlap with potential

transcription factor-binding sites [CdxA, EFII, SRY, SP1 and Nkx-2.5 (Figure 1)]. One of the sites predominantly protected is the putative binding site for the CRE (Figure 4B).

To characterize the binding of nuclear proteins to the transcription-factor-binding motifs we used EMSAs with two DNA probes of 72 bp and 123 bp in length (fragments I and II) (Figure 5 and Table 1) covering the evolutionarily conserved region. We identified two major retarded bands corresponding to various DNA-protein complexes by analysing 32 P-labelled fragment II by gel retardation assay with CHO nuclear extract (Figure 5B,

Table 1 Binding constants of fragment I (72 bp) and fragment II (123 bp) in EMSAs with CHO and NS20Y extractsResults are presented as means $\pm$ S.D.

Extract	Shift	Fragment	Binding constant (specific competition) (nM)	Fragment	Binding constant (cross-competition) (nM)
NS20Y	1	I	7.9 $\pm$ 1.2	II	340 $\pm$ 200
	2		5.4 $\pm$ 0.44		310 $\pm$ 150
	3		34 $\pm$ 12		31 $\pm$ 0.4
	4		7.4 $\pm$ 2.2		71 $\pm$ 16
NS20Y	1	II	6.7 $\pm$ 0.11	I	31 $\pm$ 1.3
CHO	1	II	7.5 $\pm$ 0.4	I	24 $\pm$ 1.3
	2		11 $\pm$ 4.0		26 $\pm$ 6.0
	3		4.4 $\pm$ 0.79		20 $\pm$ 2.5
	4		5.7 $\pm$ 1.0		18 $\pm$ 1.8
	5		12 $\pm$ 2.2		71 $\pm$ 29

band shifts 3 and 4; Table 1). Retarded bands were quantified and analysed by the method of Scatchard [29]. Competition with increasing amounts of the respective unlabelled probe (fragment II) was able to displace the gel mobility shift (Figure 5B), suggesting that these bands represent the specific binding of CHO nuclear proteins to the *rhdl* promoter. Binding constants were found to be 4.44 ± 0.795 nM for band shift 3 and 5.66 ± 1.03 nM for band shift 4 (means $\pm$ S.D.). Competition experiments with fragment I as unlabelled cross-competitor provided further evidence for the specificity of binding of CHO nuclear proteins (Table 1). However, the use of nuclear extract of NS20Y cells did reveal a different pattern of band shifts with one major shifted band (Figure 5C), suggesting the binding of a different protein to the analysed fragment in neuronal cells. The intensity of this latter band was efficiently decreased by competition experiments with unlabelled fragment II as the specific competitor (binding constant 6.7 ± 0.107 nM), but not by cross-competition with fragment I (binding constant 14 ± 1.3 mM (Figure 5C and Table 1). Similar specific binding properties were observed for fragment I by gel retardation assays. Four major shifted bands were observed by using nuclear extract of NS20Y cells (Figure 5D and Table 1), indicating the binding of various proteins to this promoter region. Three of the band shifts represented specific binding of NS20Y nuclear proteins to the *rhdl* promoter, as demonstrated by competition experiments with fragment I as unlabelled competitor or, alternatively, with fragment II as unlabelled cross-competitor (Table 1). However, the binding constants of band shift 3 (34 ± 11.7 nM) and cross-competition (31 ± 0.41 nM) provide evidence for rather non-specific protein–DNA interaction of this electrophoretic mobility-shift band.

DISCUSSION

We initiated the characterization of the *rhdl* promoter region by determining the DNA sequence immediately upstream of exon 1, by expression activity assays and by DNA–protein-binding studies. Despite using several oligonucleotides for primer extension experiments as well as employing S1 and mung bean nuclease protection assays we failed to define specific transcription initiation sites (results not shown). Although the existence of untranslated exon(s) in the 5'-flanking region of the *rhdl* gene cannot be excluded completely, detailed screening of several cDNA libraries did not provide any indication of additional 5' exons [18]. By using the *CAT* reporter gene system,

we show that the candidate promoter was capable of initiating the synthesis of the reporter mRNA and protein. We predict that the transcription initiation site(s) of the *rhdl* gene might lie within the region -215 to -141 , because the R-CAT5 (-215) deletion construct is still active as a promoter, but CAT expression activity falls practically to zero when the shorter R-CAT6 (-141) construct is used. Indeed, Lin et al. [19] suggested multiple transcription sites for the human *HD* gene at positions -135 and -145 , and for the mouse homologue at positions -146 and -157 . In addition, functional analysis of the human *HD* promoter by the luciferase reporter gene [20] revealed a significant activity loss on deletion of a similar region (-242 to -145), which further supports the position of the transcription initiation site(s).

The promoter regions of the *rhdl* and the human *HD* genes lack TATA and CAAT boxes in standard positions [32,33], and both are rich in G+C content [18,19]. Several Sp1 sites are located within the 5' upstream region. As Sp1 protein can mediate looping out, these sites might be important for interactions between the distal and proximal promoter regions [34–36]. In fact the functional importance of two Sp1 sites arranged in a tandem repeat and acting synergistically has recently been shown for the human *HD* promoter [20]. Additionally, potential PEA3 [37], SRY [38], CRE [39], CdxA [40], Nkx-2.5 [41], TTR and H2A transcription-factor-binding sites were identified in the rat 5' upstream region of the *HD* gene. These transcription factors might operate during organogenesis (PEA3, Nkx-2.5 and CdxA). Gene-targeting experiments in mice indicate that huntingtin is indeed involved in organogenesis of the brain [3–5,42]. As mouse embryos lacking huntingtin die early during embryonal development between days 7.5 and 9.5 [3–5], these experiments do not allow us to draw conclusions on its function in the development of other organs. For instance, in adult rats *rhdl* is weakly expressed in non-neuronal tissues such as placenta, skeletal muscle, lung, liver and ovary [15] but not in testis, where *rhdl* is found at a similar level to that in brain [18]. In this respect it was interesting to identify a potential SRY binding site at position -510 of the *rhdl* gene by using CHO nuclear extract. The DNA-binding activity of SRY is essential for testis development [43,44]. The function of *rhdl* in testis, however, remains to be determined. Other transcription factors such as CREs might be involved in the regulation of *rhdl* transcription. Data supporting this hypothesis are derived not only from computer searches but also from our DNase I footprinting analysis indicating the DNA–protein interaction of this sequence. CREs have been shown to interact with CRE-binding complexes. So far, ten CRE-binding proteins have been cloned (reviewed in [39]), of which CREB2 has been implicated in signal transduction in the brain [45]. However, on the basis of our preliminary experiments we cannot differentiate between the different isoforms of CREs that might specifically bind to the *rhdl* promoter region. Deletion of the segment containing the putative CRE-binding site and of putative transcription initiation sites significantly decreases expression activity in CAT assays in neuronal and non-neuronal cells (Figure 3, construct R-CAT6), further implying its importance for *rhdl* transcription.

Several DNA motifs have been identified here by sequence comparisons. For instance, a polypyrimidine tract of 44 nt is located between positions -1147 and -1104 , which might represent a potential transcription-factor-binding site. For instance, the multivalent zinc-finger repressor CTCF is known to bind repetitive CCCTC elements [46], and single-stranded polypyrimidine sequences have been shown to bind NDP kinase B [47], a putative transcription factor directly involved in developmental processes. However, evidence for a potential func-

tion of NDP kinase B in the regulation of transcription of neuronal cells has not yet been demonstrated. In addition, this might account for only part of the neuronal silencer function, because neuronal expression activity is further increased by the deletion of an additional 570 bp. Coles et al. [20] observed a similar decrease in the transcription activity on deletion of the region between positions -1032 and -324 of the human HD 5' upstream region. The existence of 5' sequences inhibiting the expression of *huntingtin* has also been suggested in transgenic mice and in cell-culture experiments [48]. However, in these experiments -120 bp and -319 bp 5' flanking sequences were used under the additional control of the CMV promoter. This might have caused interference between interacting transcription factors.

The homology of the *rhd* promoter region to the corresponding human and mouse sequences [19] was especially highly conserved in a region of 150 bp starting at -56 bp upstream of the translation initiating ATG. This region was revealed to be crucial for expression activity as determined by CAT assays in neuronal NS20Y and NG108-15 cells and non-neuronal CHO cell lines; it therefore represents the core promoter of the *rhd* gene. We demonstrated by DNase I footprinting and by EMSA that this DNA segment does indeed bind CHO and NS20Y nuclear protein extract, giving further support to its functional importance in the regulation of this gene. Competition experiments demonstrate the specificity of the crude protein extract's binding to this segment. Overall, significant differences between neuronal and non-neuronal cell lines have not been found in the CAT analysis or in DNase I footprint analysis. This was, however, unexpected as the *HD* gene is also expressed in ovary, although at a much lower level than in brain [20].

In summary, we have cloned, sequenced and analysed the *rhd* gene promoter by CAT expression assays, DNase I footprinting and band-shift experiments. The upstream 1348 bp contains several positive and negative regulatory elements, but not the domains conferring complete tissue-specific expression of the gene. Further investigations, including searching for intronic and/or 3' regulatory sites, will be necessary before the regulation of the *rhd* gene expression can be fully elucidated.

We thank the members of the laboratory of O.R. for their support and helpful discussions; Dr. Ross for the AP78 antibody; Dr. Laccone for the rat genomic cosmid library; and C. Plehn for photography. This work was supported by the Deutsche Forschungsgemeinschaft (Ri682/1-2).

REFERENCES

- 1 The Huntington's Disease Collaborative Research Group (1993) *Cell* **72**, 971-983
- 2 DiFiglia, M., Sapp, E., Chase, K., Schwarz, C., Meloni, A., Young, C., Martin, E., Vonsattel, J.-P., Carraway, R., Reeves, S. A. et al. (1995) *Neuron* **14**, 1075-1081
- 3 Nasir, J., Floresco, S. B., O'Kusky, J. R., Diewert, V. M., Richman, J. M., Zeisler, J., Borowski, A., Marth, J. D., Phillips, A. G. and Hayden, M. R. (1995) *Cell* **81**, 811-823
- 4 Duyao, M. P., Auerbach, A. B., Ryan, A., Persichetti, F., Barnes, G. T., McNeil, S. M., Ge, P., Vonsattel, J. P., Gusella, J. F., Joyner, A. L. and MacDonald, M. E. (1995) *Science* **296**, 407-410
- 5 Zeitlin, S., Liu, J.-P., Chapman, D. L., Papaioannou, V. E. and Efstratiadis, A. (1995) *Nature Genet.* **11**, 155-163
- 6 Ross, C. A. (1995) *Neuron* **15**, 493-496
- 7 Jou, Y.-S. and Myers, R. M. (1995) *Hum. Mol. Genet.* **4**, 465-469
- 8 Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H. and Wanker, E. E. (1997) *Cell* **90**, 549-558
- 9 Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L. and Bates, G. P. (1997) *Cell* **90**, 537-548
- 10 Di Figlia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P. and Aronin, N. (1997) *Science* **277**, 1990-1993
- 11 Ross, C. A. (1997) *Neuron* **19**, 1147-1150
- 12 Skinner, P. J., Koshy, B. T., Cummings, C. J., Klement, I. A., Helin, K., Servadio, A., Zoghbi, H. Y. and Orr, H. T. (1997) *Nature (London)* **389**, 971-974
- 13 Paulson, H. L., Perez, M. K., Trotter, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. L., Fischbeck, K. H. and Pittman, R. N. (1997) *Neuron* **19**, 333-344
- 14 Schmidt, T., Landwehrmeyer, G. B., Schmitt, I., Trotter, Y., Auburger, G., Laccone, F., Klockgether, T., Völpel, M., Epplen, J. T., Schöls, L. and Riess, O. (1998) *Brain Path.*, in the press
- 15 Li, S.-H., Schilling, G., Young, III, W. S., Li, X.-J., Margolis, R. L., Stine, O. C., Wagster, M. V., Abbott, M. H., Franz, M. L., Ranen, N. G. et al. (1993) *Neuron* **11**, 985-993
- 16 Gossen, M., Schmitt, I., Obst, K., Wahle, P., Epplen, J. T. and Riess, O. (1996) *Hum. Mol. Genet.* **5**, 381-389
- 17 Schmitt, I., Brattig, T., Gossen, M. and Riess, O. (1997) *Neurogenetics* **1**, 103-112
- 18 Schmitt, I., Bächner, D., Megow, D., Henklein, P., Hameister, H., Epplen, J. T. and Riess, O. (1995) *Hum. Mol. Genet.* **4**, 1173-1182
- 19 Lin, B., Nasir, J., MacDonald, H., Hutchinson, G., Graham, R. K., Rommens, J. M. and Hayden, M. R. (1995) *Genomics* **25**, 707-715
- 20 Coles, R., Caswell, R. and Rubinsztein, D. C. (1998) *Hum. Mol. Genet.* **7**, 791-800
- 21 Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. and Howard, B. H. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6777-6781
- 22 Jüngling, S., Cibelli, G., Czadybon, M., Gerdes, H.-H. and Thiel, G. (1994) *Eur. J. Biochem.* **226**, 925-935
- 23 Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, J. W. (1979) *Biochemistry* **18**, 5294-5299
- 24 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- 25 Thiel, G., Greengard, P. and Südhof, T. C. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3431-3435
- 26 Thiel, G., Schoch, S. and Petersohn, D. (1994) *J. Biol. Chem.* **269**, 15294-15301
- 27 Altschmid, J., Müller, M., Banihmad, A., Steiner, C. and Renkawitz, R. (1989) *Nucleic Acids Res.* **17**, 4975-4991
- 28 Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248-254
- 29 Scatchard, G. (1949) *Ann. N. Y. Acad. Sci.* **51**, 660-672
- 30 Laemmli, U. K. (1970) *Nature (London)* **227**, 680-685
- 31 Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350-4354
- 32 Smale, S. T. and Baltimore, D. (1989) *Cell* **57**, 103-113
- 33 Pal, S. K., Zinkel, S. S., Kiessling, A. A. and Cooper, G. M. (1991) *Mol. Cell. Biol.* **11**, 5190-5196
- 34 Li, R., Knight, J. D., Jackson, S. P., Tijan, R. and Botchan, M. R. (1991) *Cell* **65**, 493-505
- 35 Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P. and Hough, P. V. C. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5670-5674
- 36 Su, W., Jackson, S., Tijan, R. and Echols, H. (1991) *Genes Dev.* **5**, 820-826
- 37 de Launoit, Y., Baert, J.-L., Chotteau, A., Monte, D., Defossez, P.-A., Coutte, L., Pelczar, H. and Leenders, F. (1997) *Biochem. Mol. Med.* **61**, 127-135
- 38 Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R. and Bianchi, M. E. (1994) *EMBO J.* **13**, 6115-6124
- 39 Benbrook, D. M. and Jones, N. C. (1994) *Nucleic Acids Res.* **22**, 1463-1469
- 40 Margalit, N., Yarus, S., Shapira, E., Gruenbaum, Y. and Fainsod, A. (1993) *Nucleic Acids Res.* **21**, 4915-4922
- 41 Chen, C. Y. and Schwartz, R. J. (1995) *J. Biol. Chem.* **270**, 15628-15633
- 42 White, J. K., Auerbach, W., Duyao, M. P., Vonsattel, J.-P., Gusella, J. F., Loyner, A. L. and MacDonald, M. E. (1997) *Nature Genet.* **17**, 404-410
- 43 Nasrin, N., Buggs, C., Fu Kong, X., Carnazza, J., Goebel, M. and Alexander-Bridges, M. (1991) *Nature (London)* **354**, 317-320
- 44 Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P. N. (1992) *Science* **255**, 453-456
- 45 Takeda, J., Maekawa, T., Sudo, T., Seino, Y., Imura, H., Saito, N., Tanaka, S. and Ishii, S. (1991) *Oncogene* **6**, 1009-1014
- 46 Burcin, M., Arnold, R., Lutz, M., Kaiser, B., Runge, D., Lottspeich, F., Filippova, G. N., Lobanenko, V. V. and Renkawitz, R. (1997) *Mol. Cell. Biol.* **17**, 1281-1288
- 47 Hildebrandt, M., Lacombe, M. L., Mesnilidrey, S. and Veron, M. (1995) *Nucleic Acids Res.* **23**, 3858-3864
- 48 Goldberg, Y. P., Nicholson, D. W., Rasper, D. M., Kalchman, M. A., Koide, H. B., Graham, R. K., Bromm, M., Kazemi-Esfarjani, P., Thornberry, N. A., Vaillancourt, J. P. and Hayden, M. R. (1996) *Nature Genet.* **13**, 442-449