

Structure of the human histamine H1 receptor gene

Marianne D. DE BACKER^{*1}, Inge LOONEN^{*2}, Peter VERHASSELT^{†3}, Jean-Marc NEEFS^{†3} and Walter H. M. L. LUYTEN^{*2}

^{*}Department of Experimental Molecular Biology, Janssen Research Foundation, Turnhoutseweg 30, B2340 Beerse, Belgium and [†]Department of Applied Molecular Biology, Janssen Research Foundation, Turnhoutseweg 30, B2340 Beerse, Belgium

Histamine H1 receptor expression has been reported to change in disorders such as allergic rhinitis, autoimmune myocarditis, rheumatoid arthritis and atherosclerosis. Here we report the isolation and characterization of genomic clones containing the 5' flanking (regulatory) region of the human histamine H1 receptor gene. An intron of approx. 5.8 kb was identified in the 5' untranslated region, which suggests that an entire subfamily of G-protein-coupled receptors may contain an intron immediately upstream of the start codon. The transcription initiation site was mapped by 5' rapid amplification of cDNA ends to a region 6.2 kb upstream of the start codon. Immediately upstream of the transcription start site a fragment of 1.85 kb was identified that showed promoter activity when placed upstream of a luciferase reporter gene and transiently transfected into cells expressing the histamine H1 receptor. The promoter sequence shares a number

of characteristics with the promoter sequences of other G-protein-coupled receptor encoding genes, including binding sites for several transcription factors, and the absence of TATA and CAAT sequences at the appropriate locations. The promoter sequence described here differs from that reported previously [Fukui, Fujimoto, Mizuguchi, Sakamoto, Horio, Takai, Yamada and Ito (1994) *Biochem. Biophys. Res. Commun.* **201**, 894–901] because the reported genomic clone was chimaeric. Furthermore our study provides evidence that the 3' untranslated region of the H1 receptor mRNA is much longer than previously accepted. Together, these findings provide a complete view of the structure of the human histamine H1 receptor gene. Both the coding region of the H1 receptor gene and its promoter region were independently mapped to chromosome 3p25.

INTRODUCTION

Histamine is a major mediator in allergic reactions, and histamine H1 receptor (H1-R) antagonists have a long history of clinical efficacy in a variety of allergic disorders [1]. The H1-R is involved mainly in smooth-muscle contraction and vascular permeability [2,3]. It activates phospholipase C, whose reaction products Ins(1,4,5)P₃ and diacylglycerol respectively induce Ca²⁺ mobilization from intracellular stores and activate protein kinase C. Action through arachidonic acid release has also been reported [4].

Several studies have documented changes in the expression of H1-Rs in pathological situations. Increased expression of H1-R mRNA was observed in the nasal mucosa of patients with allergic rhinitis [5] and in cultured aortic intimal smooth muscle cells of patients suffering from atherosclerosis [6]. The latter finding suggests that up-regulation of the H1-R expression might have a role in the initiation and progression of cardiovascular disease. Increased H1-R expression has been documented in the inflamed joints of rheumatoid arthritis patients [7] and in the heart of mice with autoimmune myocarditis [8]. Furthermore Hamano et al. [9] recently showed that the exacerbation of asthmatic symptoms experienced by some women during pregnancy is due to the increased expression of H1-R mRNA in nasal epithelial cells on induction by female sex hormones.

Given the importance of understanding the transcriptional regulation of the H1-R, we have cloned and analysed its 5' regulatory region.

H1-Rs have been cloned from human [10–12], cow [13], guinea pig [14], mouse [15] and rat [16]. It has been assumed so far that

H1-Rs are intronless, as is true of most G-protein-coupled receptors. Consequently the region immediately 5' to the H1-R-coding region in genomic clones has been analysed for the presence of promoter elements and transcription factor-binding sites [17,18]. For the human H1-R, two different 5' untranslated regions (5' UTRs) have been published [1,18] and both have been reported to contain TATA boxes (although different ones) as well as glucocorticoid response elements, in addition to other regulatory motifs.

Here we report the isolation of genomic clones containing several kilobases upstream of the coding region of the human H1-R. Moreover, by using 5' rapid amplification of cDNA ends (RACE), the presumptive transcription start site was mapped and an intron demonstrated in the 5' UTR. This 5.8 kb intron was sequenced, as was a 1.85 kb fragment upstream of the putative transcription start site. This 1.85 kb fragment could drive expression of a luciferase reporter gene in COS-7 cells, which express endogenous H1-Rs, but not in another cell line devoid of H1-Rs.

We believe that this represents the first report of the complete structure of an H1-R gene, including the promoter region and an extended 3' UTR. Knowledge of the H1-R gene structure might advance our understanding of how H1-Rs receptors are regulated in pathological situations and how their expression can be manipulated.

MATERIALS AND METHODS

Human genomic libraries

Library I: the genomic DNA that was used to clone the main

Abbreviations used: AP, adaptor primer; DAPI, 4,6-diamidino-2-phenylindole; EST, expressed sequence tag; FISH, fluorescent *in situ* hybridization; H1-R, histamine H1 receptor; RACE, rapid amplification of cDNA ends; UTR, untranslated region.

¹ To whom correspondence should be sent at the following address: Department of Functional Genomics, Janssen Research Foundation, Turnhoutseweg 30, B2340 Beerse, Belgium (e-mail mbacker@janbe.jnj.com).

² Present address: Department of Functional Genomics, Janssen Research Foundation, Turnhoutseweg 30, B2340 Beerse, Belgium.

³ Present address: Department of Biotechnology, Janssen Research Foundation, Turnhoutseweg 30, B2340 Beerse, Belgium.

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AJ000742.

part of the H1-R gene (5.9 kb) encoding the 5' UTR (including a 5.8 kb intron) was isolated as described by Miller et al. [19] from the blood of a single healthy donor. High-molecular-mass human genomic DNA was partly digested with *Sau3A* [20]; fragments of 14–23 kb were selected by sucrose-density-gradient centrifugation. Arms of the EMBL3 bacteriophage vector λ (λ EMBL3/*Bam*HI vector kit; Stratagene) and the fragmented human DNA were ligated (DNA ligation kit; Stratagene) to form long concatameric molecules, which were subsequently packaged into bacteriophage λ heads with Gigapack II packaging extract (Stratagene).

Library II: for the cloning of the promoter region of the H1-R, a λ FIXII human placenta genomic library (Stratagene) was used.

Screening of genomic libraries

Specific DNA fragments (bovine H1-R fragment for library I; cDNA fragment located immediately downstream of the putative transcription start site for library II) were labelled with [³²P]dCTP by using the Multiprime[®] random-primed labelling system (Amersham). The human genomic libraries were plated out and replica filters were prepared as described [21]. Duplicate nylon filters were prewashed in 5 × SSC (1 × SSC = 0.15 M NaCl/0.015 M sodium citrate)/0.5% SDS/1 mM EDTA, then hybridized at 37 °C (for library I) and 42 °C (for library II) for 43 h in 5 × SSPE [1 × SSPE = 0.15 M NaCl/10 mM sodium phosphate (pH 7.4)/1 mM EDTA]/50% (v/v) formamide/5 × Denhardt's solution/0.1% SDS/100 µg/ml denatured salmon sperm DNA containing probe (10⁶ c.p.m./ml). Filters were then washed in 2 × SSC/0.5% SDS for 1 h at ambient temperature, then for 1 h at 50 °C. A single intense autoradiographic spot was found for library I and several for library II; the corresponding λ clones were plaque-purified.

Phage analysis and subcloning

Candidate clones were digested with a panel of restriction enzymes, resolved on 0.6% and 1.0% (w/v) agarose gels in 0.5 × TBE (1 × TBE = 45 mM Tris/borate/1 mM EDTA), stained with ethidium bromide and transferred to nylon membranes by vacuum blotting. The blots were probed under the same conditions as the genomic library. Lambda DNA was isolated with the Lambda Maxiprep kit (Qiagen). Fragments covering the entire hybridizing segment were subcloned into pUC18. The promoter region was subsequently subcloned into pGL3basic (Promega).

Preparation of poly(A)⁺ mRNA

Total RNA was extracted as described by Chirgwin et al. [20]. Poly(A)⁺ RNA was isolated with the PolyATtract mRNA isolation system I (Promega) in accordance with the manufacturer's instructions.

5' RACE PCR amplification of cDNA clones

Two internal gene-specific primers GSP1 and GSP2 were designed on the basis of the published human H1-R sequence [11]. Adaptor-ligated internally primed double-stranded cDNA was prepared from human placenta poly(A)⁺ RNA with the Marathon cDNA amplification kit (Clontech). To obtain internally primed cDNA the gene-specific GSP1 primer was used. Adaptor-ligated DNA was used in a first-round hot-start touch-down PCR with primers GSP1 and adaptor primer 1 (AP1, Marathon kit) for 40 cycles, each consisting of (1) 1 min at 94 °C, (2) 1 min at 72 °C, gradually (0.3 °C per cycle) decreasing to 65 °C, and (3)

3 min at 72 °C. In the reaction mixture 2.5 units of TaKaRa Ex *Taq* polymerase (Imtec Diagnostics) were used, and the final concentrations were 0.2 µM of each primer, 5% (v/v) DMSO and 200 µM dNTPs. A 100 µl second-round nested PCR was performed with 1 µl of the first-round sample by using the AP2 (Marathon kit) primer and GSP2. Reaction conditions were the same as in the first-round PCR.

Northern blot analysis

A DNA fragment carrying the entire coding region of the H1-R was labelled with [³²P]dCTP by using the Multiprime[®] random-primed labelling system (Amersham). A multiple tissue Northern blot (Clontech) containing approx. 2 µg of poly(A)⁺ RNA from human heart, brain, placenta, lung, liver, skeletal muscle and kidney was hybridized at 68 °C overnight in ExpressHyb (Clontech) hybridization solution containing probe (10⁶ c.p.m./ml). The blot was then washed in 2 × SSC/0.05% SDS for 40 min at ambient temperature, then in 0.1 × SSC/0.1% SDS for 1 h at 50 °C.

Promoter analysis

Nested deletions were prepared of a 1.85 kb subclone containing the putative promoter region of the H1-R in pGL3basic by using Exonuclease III [21]. These subclones were end-sequenced by fluorescent dideoxy dye terminator automated sequencing (Model 377, Applied Biosystems).

Monkey kidney COS-7 cells were transiently transfected with the pGL3 reporter construct by using the DEAE-dextran transfection method as described by Huylebroeck et al. [22], except that incubations were at 37 °C. Cells were harvested after 48 h, washed with PBS and lysed in 1 × reporter lysis buffer (Promega). Luciferase activity was determined with the luciferase assay system (Promega) in accordance with the manufacturer's instructions. Cells were co-transfected with pSV β -gal to determine the transfection efficiency. Protein concentrations were estimated with the Bio-Rad DC protein assay kit.

Chromosomal localization

Chromosomal localization was determined by SeeDNA Biotech with fluorescent *in situ* hybridization (FISH). Lymphocytes isolated from human blood were cultured in α -minimal essential medium supplemented with 10% (v/v) foetal calf serum and phytohaemagglutinin at 37 °C for 68–71 h. The lymphocyte cultures were treated with bromodeoxyuridine (0.18 mg/ml; Sigma) to synchronize the cell population. The synchronized cells were washed three times with serum-free medium to release the bromodeoxyuridine block and were then recultured at 37 °C for 6 h in α -minimal essential medium with thymidine (2.5 µg/ml; Sigma). Cells were harvested and slides were made by using standard procedures including hypotonic treatment, fixing and air-drying.

Genomic probes were biotinylated with dATP by using the BRL BioNick labelling kit (15 °C, 1 h) [18]. The procedure for FISH detection was performed as described in [23,24]. In brief, slides were baked at 55 °C for 1 h. After treatment with RNase, the slides were denatured in 70% (v/v) formamide in 2 × SSC for 2 min at 70 °C followed by dehydration with ethanol. Probes were denatured at 57 °C for 5 min in a hybridization mix consisting of 50% (v/v) formamide and 10% (v/v) dextran sulphate and human cot I DNA. Probes were loaded on the denatured chromosomal slides after 15 min incubation at 37 °C. After overnight hybridization, slides were washed and detected as well as amplified. FISH signals and the 4,6-diamidino-2-

phenylindole (DAPI) banding pattern were recorded separately by taking photographs, and the assignment of the FISH mapping data to chromosomal bands was achieved by superimposing FISH signals with DAPI-banded chromosomes.

Computer analysis

DNA sequences were analysed using the Findpatterns algorithm of the Genetics Computer Group (GCG) software package (University of Wisconsin). Alignments were obtained with ClustalW (GCG) and further edited in Genedoc (version 2.0001; Karl Nicholas).

RESULTS

Isolation of the putative promoter region of the human H1-R gene

A single strongly hybridizing clone (referred to as G1) was identified in 4.7×10^5 phages (three genome equivalents) from a human genomic DNA library screened with a ^{32}P -labelled bovine H1-R PCR fragment [11]. The isolated clone had a 14.9 kb insert containing the complete coding sequence of the H1-R in addition

to approx. 5.9 kb 5' to the start codon (Figure 1). The putative transcription start site was determined by isolating cDNA clones with 5' RACE. Five independent cDNA clones were isolated and sequenced (Figure 2). Comparison of the genomic and cDNA clones revealed the presence of a 5.8 kb intron in the genomic region encoding the 5' UTR. The isolated genomic clone, G1, did not contain the transcription start site and the region further upstream. Therefore a second library was screened with a 150 bp *TaqI* 5' RACE fragment (designated 'probe' in Figure 1) located immediately downstream of the putative transcription start site; 18 strongly hybridizing plaques were identified and screened by PCR with probe-specific primers. Two clones clearly yielded the expected 150 bp probe fragment and one of these (referred to as G2) was chosen for further analysis. Restriction mapping of this phage was performed as described above. A 2.3 kb *HindIII* fragment hybridized to the above-mentioned 150 bp *TaqI* probe and this putative promoter fragment was subcloned into pUC18. Sequencing revealed that this fragment contained 1.85 kb sequence upstream of the transcription start site. This 1.85 kb region was amplified by PCR and subcloned into pGL3basic (Promega) for further functional analysis.

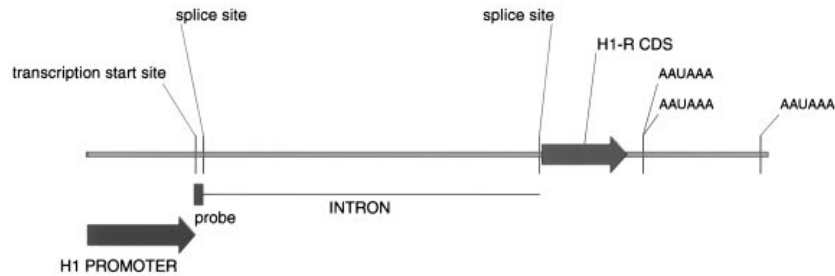


Figure 1 Genomic organization of the 11687 bp human histamine H1-R gene

The region designated 'probe' is a 150 bp *TaqI* fragment used to clone the promoter region by library screening. Potential polyadenylation sites (AAUAAA) are shown. Abbreviation: CDS, coding sequence.

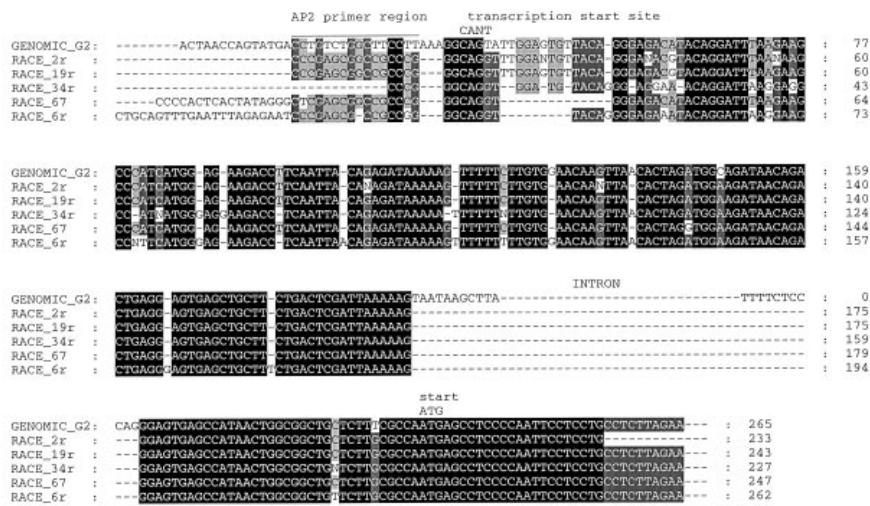


Figure 2 Alignment of cDNA sequences obtained by 5' RACE with the human H1-R genomic clone G2

```

1  cccgggctcg actcgatctg caacttgtct agcaagggTA TAAAttcccta cgcattccata
61  aactctgagg actcgatcac acatagcttg ggggcccccg tgagttccctt gatgtagctg
121 tgacaatacc tcttgcatag agggcttaga cagcatttcc ttcttgtctg gttgtaccoca
181 tcttccctcg tgactttcct tggtctctat ttcttttaac ttttcttttt ctctggggga
241 gaaaacaggg actgtagctg gagggggaag ataaggggtt aggtggaaaa tgggtgctgc
301 ttgagaagag gcaacatggt tagccacttg gtcagctaga ttattccctc agctctcaaaa
361 ggaaagatto ttctagtgc acggaacatg aacaactgct gtctctcttg gcggctgtaa
421 gttctctagt acttggatta tcaagtctct atggactaag ttttggcctt tactgttaat
481 aaggcctcgc tctgcccaaa ttttcccaaa ggcgtggact actccaaaagg catacgtgga
541 gtcagTATAA Aatagttcctt cctgggtttg cagaaatttt aaggcttgat ttagtgtaaa
601 caactcacat gtttgcctag accagtcatt aggtaggggtg tctccgtcta ctgctgagta
661 cctgttatgc cttttccctt ttattacttg agaagaacca tctcaaaaaa ggtgtcttcc
721 ggtttgaaaag ggagttttatc taaacatcta tgcocaaagt cttctggctt ggggcatggg
781 tttttttctg cgtttgaatt tctgtttaa gaggcagcag ggttaaggga atcaatctgta
841 gttcggatta aatcatcttt ttctaacaag atagccttgt attttaaaat tcttggagtca
901 gtaagcaacc tttctgctt ctgatttagg atagtctgtg tctgggtgagg tgtgctcaca
961 atggggtttc ctccaaaagt ttttttcta ctttctctg ttagcaaaagt agttgccgct
1021 acagattgaa tgcatttggg caatccatgg gtactgggtg taagaatttt tgacaggaag
1081 cctacgggtt gggagtgggc tccgtgcttt tgggtaagta ctcccaagggt tacgcccctg
1141 tttacattga cgaaaaagatg gaatggctgc ttaggggaggg taaagctagg acaggggagcag
1201 ttactaatag atgttttaac cttctacact tttggatttc tggtaattgc caaatgatgg
1261 ggtctggctc gtcttctgtg agctttttgt atgagagttc tgttctctag gcataagagt
1321 ctatccatag acggcagtat ctgactaatc ctccttcaat ccattcaagc ccaattttcc
1381 atttgccttt gtaattaaat gccctaaata ttttacttca ggttctacaa attggagttt
1441 gttttctgag gccgttaacc ctctatccca cagaaaaatgt aagacatggg ttgagaaagc
1501 tgetacttct tttctatcat ctccgaaat tagaagatca tccatgtact ggaggggaca
1561 tatgcacgag ggcagggaaa atttctctag gacttgttct aatattgac taagttaaata
1621 tggagactcc gtaaacccccggggcaagac tgtccatcag tattgctgtt ttcaaccgga
1681 gtgagggctc ttcactcaa aggcaagtag gtccctggctg cctctgctc atggacaagc
1741 ccagaaggca tcatttaaat ccactactgt actatctgat taatagctct aaggtcttgc
1801 actaaccagt atgacctgtc tggcttcttt aaaggcagta ttgagtggtt acagggagac
1861 atacaggatt taagaagccc atcatggaga agacctcaa ttacagagat aaaaagtttt
1921 tcttggtaa caagttaaca ctagatggca gataacagac taggagtgta gctgctctcg
1981 actcgattaa aaaggtaata agcttagggt taggtcccac ttctaaattt accaagggct
2041 gttggtggga ctcaagagac aaagatggag cccctgacct cctagtctt ctcaaaaaac

```

Figure 3 Nucleotide sequence of the promoter region of the H1-R gene

Potential AP1-binding sites are dot-boxed, potential AP2-binding sites are dash-boxed, potential Sp1-binding sites are underlined, potential TATA boxes are in bold, GC-rich regions are line-boxed, a potential NF- κ B site is dash-underlined twice, relevant AT-rich regions are underlined twice; the splice site is marked with a vertical arrow and the putative transcription start site is marked with a horizontal arrow.

Sequence analysis of the 5' flanking region and 5' UTR of the human H1-R

To identify the DNA elements that control H1-R gene transcription, the 1.85 kb region immediately upstream of the putative transcription initiation site was sequenced. As shown in Figure 3, neither TATA nor CAAT promoter elements were found close to the transcription initiation site, although two TATA sequences and one CCAAT sequence were present more than 500 bp further upstream. Further computerized analysis of the 5' flanking region revealed potential Sp1-binding, AP1-binding, AP2-binding and NF- κ B-binding sequences as well as two GC-rich regions (Figure 3).

The 100 bp regions immediately upstream of the translation start site of the human, rat, mouse and bovine H1-R genes were compared and were found to be modestly conserved (Figure 4). The same region found upstream of the initiator AUG in the guinea pig H1-R gene differs significantly from the other species (an alignment of these sequences is shown in Figure 4). The consensus splice site that we found in the human gene at nt -37

was also found in mouse and rat H1-R genes. For the bovine H1-R only a cDNA sequence is available, which diverges from the genomic sequence in the other species upstream of this putative splice site, as could be expected if the genomic sequences contained an intron at that point. The bovine cDNA has a rather short 5' UTR (108 bp), which is similar to our finding for the human H1-R cDNA (178 bp).

Functional analysis of the putative promoter region of the human H1-R

To define the promoter region that controls H1-R expression, several 5' deletion mutants were generated of the 1.85 kb fragment that is located immediately upstream of the transcription start site. The deletion constructs, fused 5' to the luciferase gene of a reporter plasmid, were transfected into COS-7 cells for transient expression analysis (Figure 5 and Table 1). Both COS-7 and HeLa cells have been shown to express an H1-R [25-28]. Results of the luciferase assays in COS-7 cells indicate that the highest

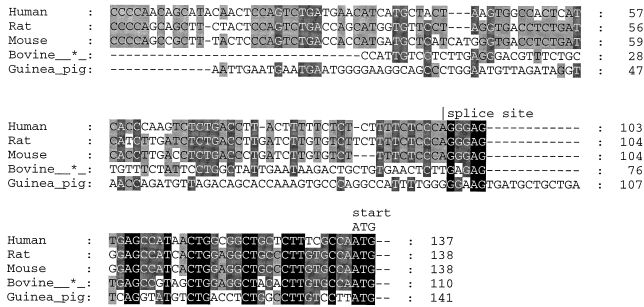


Figure 4 Sequence alignment of human, rat, mouse, bovine and guinea-pig 5' UTR — 100 bp immediately upstream of the initiator AUG codon

The location of the splice site in the 5' UTR of the human H1-R and the translation start site are marked. The bovine sequence (*) is derived from a cDNA clone.

promoter activity was produced with the clone containing a 984 bp region of the H1-R promoter. Two deletion clones showed decreased promoter activity (clone 5.8, 842 bp; clone 5.5, 612 bp), suggesting the unmasking of inhibitory control elements. A deletion clone containing 227 bp of 5' flanking sequence with a GC-rich region, an AT-rich region, an AP1-binding site, an AP2-binding site and an SP1-binding site showed promoter activity, essentially equivalent to the full 1.85 kb fragment, with a luciferase activity 9-fold higher than for a promoterless pGL3-basic vector. As expected, no promoter activity could be detected in Chinese hamster ovary cells, which do not express an endogenous H1-R (results not shown). The promoter activity in COS-7 cells was not affected by dexamethasone (100 nM) (results not shown).

Chromosomal localization of the H1-R gene

A genomic fragment (3.9 kb) containing 1.3 kb of the coding region of the H1-R gene [11] and a genomic *HindIII* fragment (2.3 kb) containing 1.85 kb of the promoter region of the H1-R gene were independently used for FISH analysis and were shown to co-localize on chromosome 3p25 (Figure 6).

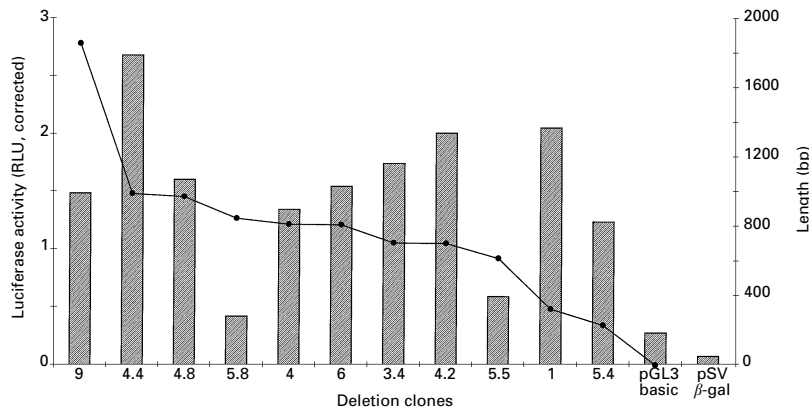


Figure 5 Deletion constructs of the H1-R promoter region and summary of luciferase activity on transfection in COS-7 cells: schematic representation

The lengths (in bp) of the promoter fragments fused to the luciferase reporter gene are shown (line). Luciferase activity (bar chart) is corrected for transfection efficiency by using co-transfection with pSV β -gal and is corrected for protein content. Abbreviation: RLU, relative light units.

Table 1 Deletion constructs of the H1-R promoter region and summary of luciferase activity on transfection in COS-7 cells: Clone ID, length and luciferase activity

Results are means \pm S.D. for three separate experiments. Luciferase activity is corrected for transfection efficiency by using co-transfection with pSV β -gal and is corrected for protein content. Abbreviation: RLU, relative light units.

Deletion clone	Length	Luciferase activity (RLU, corrected)
9	1853	1.48 \pm 0.18
4.4	984	2.68 \pm 0.19
4.8	966	1.60 \pm 0.12
5.8	842	0.41 \pm 0.33
4	807	1.34 \pm 0.37
6	805	1.54 \pm 0.02
3.4	701	1.74 \pm 0.03
4.2	698	2.00 \pm 0.13
5.5	612	0.59 \pm 0.28
1	320	2.05 \pm 0.11
5.4	227	1.23 \pm 0.02
pGL3basic	0	0.28 \pm 0.09

Northern blot analysis

Northern blots (Figure 7) show that a variety of human tissues contain a transcript doublet at approx. 4 and 4.4 kb. No signal was detected in liver and only very low levels in kidney; highest expression levels were found for the 4.4 kb transcript in brain and for both the 4 and 4.4 kb transcripts in placenta. On the basis of the transcription start site (-178) and the length of the coding region (1464 bp) this predicts a 3' UTR of approx. 2 kb, inconsistent with previous reports [12]. However, numerous cDNA clones were identified in a proprietary expressed sequence tag (EST) database (licensed from Incyte Pharmaceuticals, Palo Alto, CA, U.S.A.) (Figure 8) that extend the previously published 3' UTR by at least 2 kb and are co-linear over this range with the genomic sequence of Fukui [17].

An overview of the structure of the entire H1-R gene is presented in Figure 1.

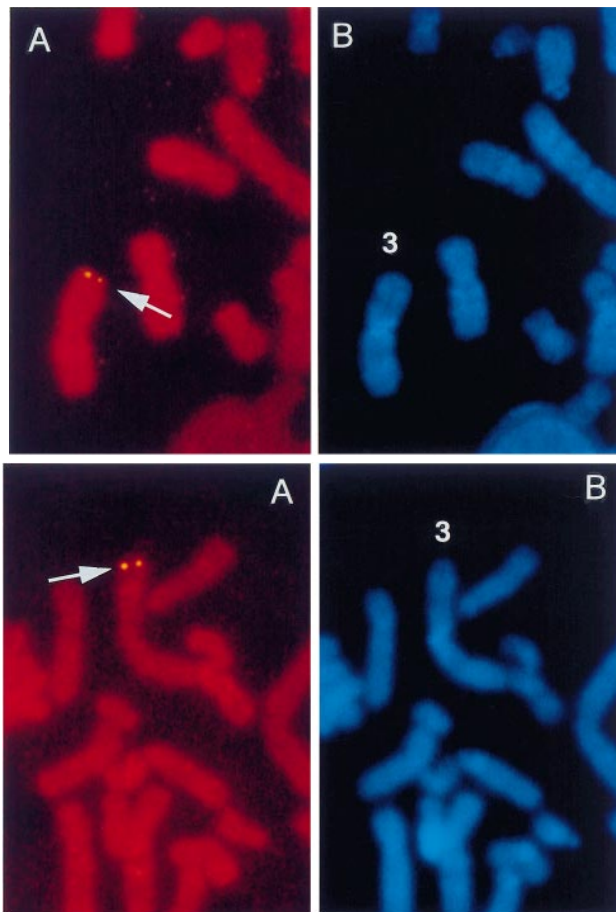


Figure 6 Chromosomal localization of the coding region (top) and the promoter region (bottom) of the H1-R gene, (A) with FISH signals marked with an arrow; the corresponding DAPI-stained chromosomes are shown in (B) with chromosome 3 marked

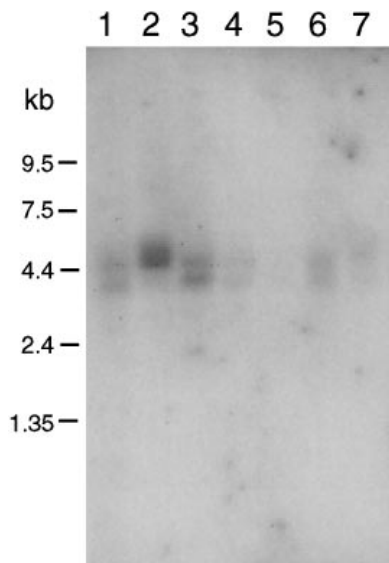


Figure 7 Northern blot analysis of human H1-R mRNA in various tissues

Lanes contained 2 μ g of poly(A)⁺ RNA from heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6) and kidney (lane 7).

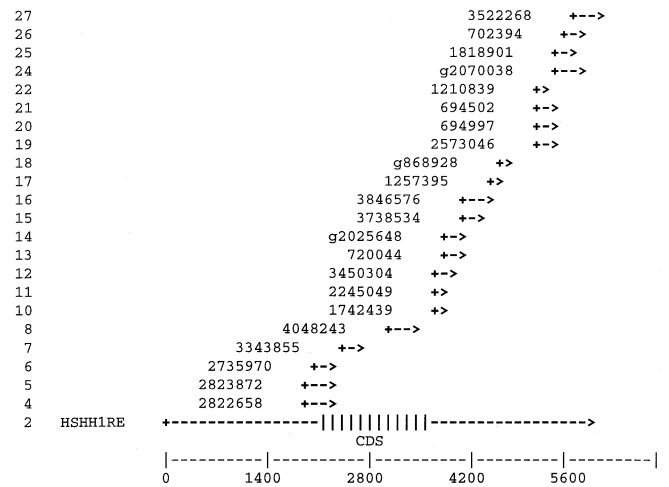


Figure 8 Alignment of cDNA clones in a proprietary EST database with the genomic sequence of Fukui [17] (EMBL-ID HSHH1RE)

This sequence starts at -2165 and extends to $+3691$. 17 cDNA clones are co-linear with the region downstream of the coding sequence. Abbreviation: CDS, coding sequence.

DISCUSSION

5' UTR region of the H1-R

Two different sequences have been reported for the 5' UTR of the H1-R [1,18], diverging at -143 bp from the start codon. Although the sequence of the genomic clone (G1) that we have isolated agrees with the shorter sequence reported independently [18], we have obtained 5.6 kb further upstream sequence and performed a number of analyses that show our sequence to be the correct one. Multiple primer pairs that straddle the point at which the sequences diverge were used for PCR and yielded fragments of identical size when our genomic lambda clone or human genomic DNA was used (results not shown). Moreover we have multiple 5' RACE clones that also straddle the divergence point and that remain co-linear with our genomic clone. In contrast, the sequence described by Fukui [17] was based on a single genomic clone containing a *Sau3AI* restriction site at the point of divergence (this restriction enzyme was used for the construction of their genomic library) and contains a repetitive LINE element upstream of this point. This clone is therefore most probably chimaeric.

Intron in the region encoding the 5' UTR of the H1-R

Comparison of our genomic and 5' RACE clones clearly shows the presence of a 5.8 kb intron in the regions encoding the 5' UTR. This intron is flanked by a consensus 3' splice site sequence (NCAG; here and in 70% of the cases CCAG) located precisely at the location at which our genomic DNA clone deviates from the isolated cDNA clones. Also, just upstream from the 3' splice junction there is a pyrimidine-rich region. Splice sites typically have a pyrimidine-rich region of 15 bases preceding the invariant AG, which is considered to be part of the recognition sequence near the 3' end of the intron-exon boundary [29]. As an 'AG' splice site consensus sequence can be found at the same position in rat, mouse and human 5' UTRs, the chances are high that H1-R genes from other species also contain an intron in their 5' UTR. At the 5' end of the intron a consensus 5' splice site sequence, GT, was found.

An intron in a similar location has been documented in genes for other G-protein-coupled receptors: m1 (rat), m2 (human, chicken), m3 (human, rat), m4 (human) and m5 (human, rat) muscarinic acetylcholine receptors, the human angiotensin II type 1 receptor, the human cannabinoid receptor, the human PAF receptor, the mouse B2 bradykinin receptor and the rat serotonin 1E receptor. In muscarinic acetylcholine receptor genes, this intron is always located close to the translation initiation codon (−6 to −80) which is similar to our finding for the human H1-R gene (−37). Moreover, this is consistent with phylogenetic trees showing histamine receptors grouped most closely with muscarinic and biogenic amine receptors (http://www.sander.embl-heidelberg.de/7tm/seq/001_001/001_001.TREE.html). In fact, all the above-mentioned receptors belong to the 'class A rhodopsin-like' G-protein-coupled receptors, and it is likely that a systematic search in this subfamily would reveal more such introns in a gene family generally considered intronless.

Transcription start site

The precise sequences surrounding transcription start sites seem to be poorly conserved, but a central CA surrounded by several pyrimidines is usually present. This is consistent with what we found (Figure 2).

Although our 5' RACE clones suggest that several transcription start sites close together, as is typical for TATA-less promoters [30], we cannot exclude the presence of additional sites further upstream because such longer RACE products might not have amplified successfully. The sequence differences that we have found for the five 5' RACE clones analysed are most probably due to polymorphisms or errors that occurred during PCR.

3' UTR

Our proposed transcription start site (−178), in combination with the predicted polyadenylation sites from data published in [12] (at +1735 and +1741), would predict a 1.9 kb mRNA, which seems inconsistent with published estimates of mRNA size (4 and 4.4 kb). Our Northern blot analysis confirms the previously published transcript sizes. Although we cannot formally exclude additional upstream transcription start sites, numerous cDNA clones found in a proprietary EST database (licensed from Incyte Pharmaceuticals, Palo Alto, CA, U.S.A.) attest to the presence of a long (more than 2.4 kb) 3' UTR, contrary to that described by Moguilevsky et al. [12]. Presumably their cDNA clone primed off an A-rich region around nt 1730. The 3' UTR cDNA consensus sequence is co-linear over a 2.0 kb range with the genomic sequence of Fukui [17] and contains predicted polyadenylation signals at positions +1735, +1741 and +3746 (Figure 1).

The position of the last polyadenylation site, combined with the position of the transcription start site that we found, predicts a total mRNA length of approx. 4 kb, which fits nicely with the size of the smaller transcript identified by Northern blot analysis. Longer cDNA species that would account for the 4.4 kb mRNA transcript that was predominantly found in brain have not yet been identified, but they could originate from an alternative polyadenylation site further downstream. It is unlikely that the polyadenylation sites predicted at nt 1735 and 1741 are used.

Promoter of the H1-R

Two papers have suggested (different) promoter regions for the H1-R. Although Fukui [17] has described promoter activity for

a fragment flanking the 5' end of the H1-R coding region, we have explained above why we believe the genomic clone to be a chimaera. This does not of course preclude the possibility that the 5' flanking region might exhibit promoter activity, but its relevance to the H1-R is doubtful. Max et al. [18] have presented a theoretical analysis only of potential transcriptional control elements in what we believe to be an intronic sequence.

Sequencing of 1.85 kb of the region upstream of the transcription start site revealed an absence of any appropriately located TATA or CAAT sequences (for a recent review on TATA-less promoters see [30]). This feature is shared by many G-protein-coupled receptor genes [31–45]. It has been suggested that the initiator function of TATA-less genes are replaced by other transcription initiator sequences. In other genes these include the GATA motif at −30 [46], a variety of AT-rich elements [47–49], CpG islands, and transcriptional initiator elements in conjunction with Sp1-binding sites [50]. We also found CpG islands and AT-rich regions in our clone. Several potential general transcription factor-binding sites were identified, including Sp1, AP1, AP2 and NF- κ B. Interestingly, both AP1 and NF- κ B transcription factors have been reported to be important in the regulation of a number of inflammatory genes [51].

The fact that the H1-R is regulated by glucocorticoids such as dexamethasone [52], even though no glucocorticoid response element is present in its promoter, can possibly be explained by the presence of multiple potential AP1-binding sites in that region. AP1 (Jun/c-Fos) and steroid hormone receptors are distinct families of transcription factors that convert extracellular signals into long-term genetic responses. It has been shown that a hormone-activated glucocorticoid receptor can negatively or positively modulate the expression of AP1-regulated promoters lacking glucocorticoid response elements [53]. This type of regulation was shown to be mediated through the DNA-binding domain of Jun.

It has also been shown that the activation of c-Fos/AP1 is essentially important in the destruction of arthritic joints [54], where expression of the H1-R is induced [8].

Chromosomal localization

Although Fukui [17] had already mapped the H1-R gene to chromosome 3p25, we wanted to confirm this result because of the chimaeric nature of their clone and somewhat conflicting reports on the chromosomal localization of the H1-R [1,55]. Moreover, the independent mapping with two different genomic fragments (one coding, one 5' UTR) not only strengthens the confidence in the result but also is consistent with the hypothesis that the two fragments derive from the same gene.

The chromosomal localization in humans fits well with the report of the mouse H1-R's mapping to the central region of mouse chromosome 6, which is known to be syntenic with human chromosome 3p [15].

Conclusion

In the present study we have clarified the organization, genomic structure and promoter function of the human histamine H1-R gene. We have isolated a human histamine H1-R gene region spanning 11.7 kb and have identified a 5.8 kb intron in the 5' UTR of this gene by a comparison of genomic and cDNA clones. This suggests that a whole subfamily belonging to the 'class A rhodopsin-like' G-protein-coupled receptors contains an intron in the 5' flanking region of its genes.

5' RACE revealed the position of the transcription initiation site to be approximately 6 kb upstream of the translation start site.

We also showed that the 1.85 kb regulatory region can drive luciferase expression in transfected COS-7 cells, with 227 bp of the region still able to generate comparable reporter gene expression. Furthermore we provided evidence that the 3' UTR of the H1 receptor mRNA is much longer than previously accepted and therefore we were able to explain the size of the 4 kb mRNA transcript present in various tissues, most abundantly in placenta.

The possible regulation of the H1-R through the transcription factors AP1 and NF- κ B is consistent with increased H1-R expression in various inflammatory conditions in which AP1 and NF- κ B are known to be increased.

Both the coding region of the H1-R gene and the promoter region were independently mapped to chromosome 3p25.

We thank the following personnel at the Janssen Research Foundation: Geert Nobels for sequencing, Paul Van Gompel for help with transfections, Joerg Sprengel and Bart Nelissen for bio-informatics support, and Lambert Leijssen and Gert Verheyen for graphical support.

REFERENCES

- Negro-Alvarez, J. M., Funes, E., Garcia Canovas, A., Hernandez, J., Garcia-Selles, F. J., Pagan, J. A. and Lopez-Sanchez, J. D. (1996) *Allergol. Immunopathol.* **24**, 177–183
- Mitsuhashi, M. and Payan, D. G. (1992) *J. Invest. Dermatol.* **98**, 8S–11S
- Timmerman, H. (1992) *Trends Pharmacol. Sci.* **13**, 6–7
- Leurs, R., Traiffort, E., Arrang, J. M., Tardivel-Lacombe, J., Ruat, M. and Schwartz, J. C. (1994) *J. Neurochem.* **62**, 519–527
- Iryoshi, N., Takeuchi, K., Yuta, A., Ukai, K. and Satakura, Y. (1996) *Clin. Exp. Allergy* **26**, 379–385
- Takagishi, T., Sasaguri, Y., Nakano, R., Arima, N., Tanimoto, A., Fukui, H. and Morimatsu, M. (1995) *Am. J. Pathol.* **146**, 981–988
- Zenmyo, M., Hiraoka, K., Sasaguri, T., Komiya, S., Inoue, A., Morimatsu, M. and Sasaguri, Y. (1996) *Biochem. Mol. Biol. Int.* **40**, 1253–1260
- Goren, N., Perez Leiros, C., Sterin-Borda, L. and Borda, E. (1994) *Int. J. Immunopharmacol.* **16**, 737–745
- Hamano, N., Terada, N., Maesako, K., Ikeda, T., Fukuda, S., Wakita, J., Yamashita, T. and Konno, A. (1998) *Int. Arch. Allergy Immunol.* **115**, 220–227
- Fukui, H., Fujimoto, K., Mizuguchi, H., Sakamoto, K., Horio, Y., Takai, S., Yamada, K. and Ito, S. (1994) *Biochem. Biophys. Res. Commun.* **201**, 894–901
- De Backer, M., Gommeren, W., Moereels, H., Nobels, G., Van Gompel, P., Leysen, J. E. and Luyten, W. H. M. L. (1993) *Biochem. Biophys. Res. Commun.* **197**, 1601–1608
- Moguilevsky, N., Varsalona, F., Guillaume, J. P., Noyer, M., Gillard, M., Daliers, J., Henichart, J. P. and Bollen, A. (1995) *J. Recept. Signal Transduct. Res.* **15**, 91–102
- Yamashita, M., Fukui, H., Sugama, K., Horio, Y., Ito, S., Mizuguchi, H. and Wada, H. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11515–11519
- Horio, Y., Mori, Y., Higuchi, I., Fujimoto, K., Ito, S. and Fukui, H. (1993) *J. Biochem. (Tokyo)* **114**, 408–414
- Inoue, I., Taniuchi, I., Kitamura, D., Jenkins, N. A., Gilbert, D. J. and Copeland, N. G. (1996) *Genomics* **36**, 178–181
- Fujimoto, K., Horio, Y., Sugama, K., Ito, S., Liu, Y. Q. and Fukui, H. (1993) *Biochem. Biophys. Res. Commun.* **190**, 294–301
- Fukui, H. (1996) *Jpn. J. Clin. Med.* **54**, 389–395
- Max, S. I., Chowdbury, B. A. and Fraser, C. M. (1996) *Gene* **171**, 309–310
- Miller, S. A., Dykes, D. D. and Polesky, H. (1988) *Nucleic Acids Res.* **16**, 3
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Current Protocols in Molecular Biology (1990) (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), vol. 1, unit 7.2, Wiley, New York
- Huylebroeck, D., Maertens, G., Verhoeven, M., Lopez, C., Raeymakers, A., Min Jou, W. and Fiers, W. (1988) *Gene* **66**, 163–181
- Heng, H. H. Q., Squire, J. and Tsui, L.-C. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9509–9513
- Heng, H. H. Q. and Tsui, L.-C. (1993) *Chromosoma* **102**, 325–332
- Young, K. W. and Young, J. M. (1994) *Eur. J. Pharmacol.* **269**, 283–292
- Zamani, R. and Bristow, D. R. (1993) *Biochem. Soc. Trans.* **21**, 382S
- Bristow, D. R. and Zamani, M. R. (1993) *Br. J. Pharmacol.* **109**, 353–359
- Arias-Montano, J. A. and Young, J. M. (1993) *Eur. J. Pharmacol.* **245**, 291–295
- Padgett, R. A., Grabowsky, P. J., Kenarska, M. M., Seiber, S. and Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150
- Smale, S. T. (1997) *Biochim. Biophys. Acta* **1351**, 73–88
- Murasawa, S., Matsubara, H., Kijima, K., Mauyama, K., Mori, Y. and Inada, M. (1995) *J. Biol. Chem.* **270**, 20042–20050
- McCuiag, K. A., Clarke, J. C. and White, J. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5051–5055
- Tsai-Morris, C. H., Buczko, E., Geng, Y., Gamboa-Pinto, A. and Dufau, M. L. (1996) *J. Biol. Chem.* **271**, 14519–14525
- Greenwood, T., Robertson, L.-A. and Patel, Y. C. (1995) *Gene* **159**, 291–292
- Wood, I. C., Roopra, A. and Buckley, N. J. (1996) *J. Biol. Chem.* **271**, 14221–14225
- Wood, I. C., Roopra, A., Harrington, C. and Buckley, N. J. (1995) *J. Biol. Chem.* **270**, 30933–30940
- Liu, C. H. and Hla, T. (1997) *Genomics* **43**, 15–24
- Chen, W. M., Chang, A. C., Wang, C. M., Lin, C. C. and Chang, N. C. (1996) *Neurosci. Lett.* **210**, 33–36
- Lee, S. H., Minowa, M. T. and Mouradian, M. M. (1996) *J. Biol. Chem.* **271**, 25292–25299
- Parks, C. L. and Shenk, T. (1996) *J. Biol. Chem.* **271**, 4417–4430
- Dufau, M. L., Tsai-Morris, C. H., Hu, Z. Z. and Buczko, E. (1995) *J. Steroid Biochem. Mol. Biol.* **53**, 283–291
- Augustin, L. B., Felsheim, R. F., Min, B. H., Fuchs, S. M., Fuchs, J. A. and Loh, H. H. (1995) *Biochem. Biophys. Res. Commun.* **207**, 111–119
- Ikuyama, S. and Nawata, H. (1994) *Jpn. J. Clin. Med.* **52**, 962–968
- Minowa, T., Minowa, M. T. and Mouradian, M. M. (1994) *J. Biol. Chem.* **269**, 11656–11662
- Minowa, T., Minowa, M. T. and Mouradian, M. M. (1993) *J. Biol. Chem.* **268**, 23544–23551
- Aird, W. C., Parvin, J. D., Sharp, P. A. and Rosenberg, R. D. (1994) *J. Biol. Chem.* **269**, 883–889
- Hahn, S., Buratowski, S., Sharp, P. A. and Guarente, L. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5718–5722
- Singer, V. L., Wobbe, C. R. and Struhl, K. (1990) *Mol. Cell. Biol.* **10**, 3825–3832
- Wobbe, C. R. and Struhl, K. (1990) *Genes Dev.* **4**, 636–645
- Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S. T. (1994) *Mol. Cell. Biol.* **14**, 116–127
- Adcock, I. M. (1997) *Monaldi Arch. Chest Dis.* **52**, 178–186
- Teurich, S. and Angel, P. (1995) *Chem. Senses* **20**, 251–255
- Hardy, E., Farahani, M. and Hall, I. P. (1996) *Br. J. Pharmacol.* **118**, 1079–1084
- Shiozawa, S., Shimizu, K., Tanaka, K. and Hino, K. (1997) *J. Clin. Invest.* **99**, 1210–1216
- Le Coniat, M., Traiffort, E., Ruat, M., Arrang, J. M. and Berger, R. (1994) *Hum. Genet.* **94**, 186–188