Human amiloride-sensitive epithelial Na⁺ channel γ subunit promoter: functional analysis and identification of a polypurine-polypyrimidine tract with the potential for triplex DNA formation

Scott D. AUERBACH^{*1}, Randy W. LOFTUS^{*}, Omar A. ITANI^{*} and Christie P. THOMAS^{*†2}

^{*}Department of Internal Medicine, University of Iowa College of Medicine, Iowa City, IA 52242, U.S.A., and [†]Department of Veterans Affairs Medical Center, Iowa City, IA 52246, U.S.A.

The mRNA for the epithelial Na⁺ channel γ subunit (γ ENaC) is regulated developmentally in the lung, colon and distal nephron and in response to Na⁺ deprivation and systemic corticosteroids in the distal colon. Because such regulation is likely to be at the level of gene transcription, we examined the function of the promoter and other 5' flanking elements of the human γ ENaC gene. The proximal 5' flanking region contains two GC boxes but does not contain a TATA box. A 450 bp human γ ENaC fragment (−459 to +40) directed the expression of luciferase in H441 cells and primer extension analysis in transfected cells confirmed the correct initiation of human γ ENaC–luciferase chimaeric transcripts. By deletional analysis, GC boxes at −21 and −52 were found to be critical for this promoter activity. To begin to identify transcription factors that bind to the core promoter, a double-stranded oligonucleotide that corresponded to this region was synthesized and tested in a gel mobility-shift assay. Incubation of this radiolabelled oligonucleotide with nuclear extracts from H441 and FRTL5 cells resulted in the formation of

four specific and distinct DNA–protein complexes. On the basis of antibody ‘supershift’ assays, one of these factors corresponds to Sp1, whereas the other three correspond to Sp3. Further upstream, an approx. 300 nt (−1143 to −839) polypurine-polypyrimidine tract (PPy tract) containing internal mirror repeats was identified. When contained in a supercoiled plasmid, the approx. 1200 nt 5' flanking region was sensitive to S1 endonuclease, which was consistent with the formation of an intramolecular triplex DNA (‘H-DNA’) structure with an unpaired single strand. High-resolution mapping with S1 endonuclease and sequencing of S1-generated clones confirmed that all S1-sensitive sites were within the PPy tract. Finally, a negative regulatory element was identified between −1525 and −1296 that functioned in lung, colon and collecting duct cell lines.

Key words: ENaC, gene transcription, gel-mobility-shift assay, H-DNA, S1-nuclease-sensitivity.

INTRODUCTION

Na⁺ reabsorption in the terminal nephron occurs via epithelial Na⁺ channels (ENaCs), which provide the final regulatory step in Na⁺ homeostasis and the control of extracellular fluid volume. These channels are a major rate-limiting step for Na⁺ and fluid absorption in several other epithelia such as the distal colon, the airway epithelia and sweat ducts. The ENaC complex is composed of at least three distinct subunits, α , β and γ [1,2]. Rat and human cDNA species encoding these subunits have been cloned and the contributions of individual subunits to channel function have been studied by expression in *Xenopus* oocytes. Although α ENaC expressed alone has a small but measurable Na⁺ current, β and γ subunits do not elicit any current when expressed singly. However, when all three subunits are co-expressed a 100-fold potentiation of current is seen [3–5]. These results imply that regulation of expression of individual subunits might be an important factor regulating the magnitude of Na⁺ transport *in vivo*.

Members of the ENaC complex are also expressed in sites that do not normally participate in Na⁺ reabsorption, such as the liver, placenta and urothelium [5–7]. Although their role in

these sites is unknown, the expression of some but not all components of the ENaC complex in certain tissues suggests that these subunits might also serve other functions. Homologues of these proteins, such as ASIC (acid-sensing ion channel) and DRASIC (dorsal root acid-sensing ion channel) serve as pH sensors in the vertebrate central nervous system; the degenerins, MEC-4 and MEC-10 as mechanosensors in *Caenorhabditis elegans* and RPK (ripped pocket) and PPK (pickpocket) may serve a similar function in *Drosophila* [8–12]. The identification of the γ subunit of ENaC in nodose ganglia of mammalian baroreceptors raises the intriguing possibility that ENaC members might also function as mechanosensors [13].

Little is known about the factors that regulate tissue-specific expression of the ENaC subunits and those that regulate its expression in response to physiological stimuli. However, there is some evidence that transcriptional control of gene expression might be the basis for regulation by some hormones. For example, arginine vasopressin increases γ ENaC expression and Na⁺ transport in a rat collecting duct cell line, a process that can be inhibited by actinomycin D [14]. Transcription of α ENaC is down-regulated by the protein-kinase-C/ERK (extracellular signal-regulated protein kinase) pathways in parotid cells [15].

Abbreviations used: ADR1, alcohol dehydrogenase gene regulator 1; AP-2, activator protein 2; ENaC, epithelial Na⁺ channel; PPy, polypurine-polypyrimidine; STRE, stress response element; SV40, simian virus 40.

¹ Present address: Department of Cellular and Molecular Biology, University of Texas, Austin, TX 78712, U.S.A.

² To whom correspondence should be sent, at the following address: Division of Nephrology, E300GH, Department of Internal Medicine, University of Iowa College of Medicine, 200 Hawkins Drive, Iowa City, IA 52242-1081, U.S.A. (e-mail christie-thomas@uiowa.edu).

The nucleotide sequence data reported will appear in DDBJ, EMBL, GenBank[®] and GSDB Nucleotide Sequence Databases under the accession number U48937.

Glucocorticoids increase α ENaC expression in cortical collecting duct and in airway and parotid epithelial cells via an increase in transcription of the α ENaC gene [16,17]. However, in the distal colon the corticosteroid-mediated increase in Na^+ transport is accompanied by the increased expression of the mRNA for β ENaC and γ ENaC, but not for α ENaC [4,18–20]. These studies indicate that distinct transcription factors regulate the basal and hormone-stimulated expression of each ENaC subunit, in a cell-specific manner.

We have previously characterized the gene structure for the human and rat γ ENaC subunits [21,22]. Here we determine transcriptional regulatory regions of the human γ ENaC promoter and characterize transcription factors that regulate basal expression. In addition, we identify an S1 nuclease-sensitive polypurine-polypyrimidine (PPy) tract in the 5' flanking region of human γ ENaC that could adopt a triplex DNA ('H-DNA') structure *in vivo* and affect chromatin organization at this locus.

Part of this work has been published in abstract form at the American Society of Nephrology meeting and at the Central Society for Clinical Research in 1996 and 1998.

EXPERIMENTAL

Mapping and sequencing of the 5' flanking region of human γ ENaC

Screening and isolation of cosmid clones from a human genomic library has been described previously [21]. Clones that included intron 1 and the 5' flanking region were directly sequenced by using dye-terminator-cycle sequencing chemistry with Amplitaq DNA polymerase and FS enzyme and analysed on a 373A Stretch Fluorescent Automated Sequencer (Applied Biosystems, Foster City, CA, U.S.A.). Primary DNA sequence analysis was performed with MacDNASIS (Hitachi, San Bruno, CA, U.S.A.). Transcription-factor-binding sites were also analysed with the TFSEARCH program (at <http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html>) and the sequence was screened for interspersed repeats using the CENSOR program (at <http://www.girinst.org/Censor-Server.html>).

To map cosmid clones, these were digested first with *NotI* to release the insert from pSupercos-1 and then with various restriction enzymes. Fragments were resolved by PAGE and transferred to nylon membranes (Zetaprobe GT; Bio-Rad, Hercules, CA, U.S.A.). An oligonucleotide primer, 4g (5'-GCT-GCTCCTCTGCTCGCTCTCG-3'), corresponding to exon 1 was end-labelled with [γ - ^{32}P]ATP and T4 polynucleotide kinase and hybridized overnight with nylon membranes in a hybridization solution that contained 5 $\times$ standard sodium citrate (SSC; 1 $\times$ SSC is 0.15 M NaCl/0.015 M sodium citrate), 20 mM NaPO_4 , pH 7.2, 7% (w/v) SDS, 1 $\times$ Denhardt's solution (0.02% Ficoll 400/0.02% polyvinylpyrrolidone/0.002% BSA) and 100 $\mu\text{g}/\text{ml}$ denatured herring sperm DNA at 50 $^\circ\text{C}$. The membranes were washed at 50 $^\circ\text{C}$ with a solution that contained 3 $\times$ SSC, 1 $\times$ Denhardt's solution, 5% (w/v) SDS and 20 mM sodium phosphate, pH 7.2, and subjected to autoradiography.

Promoter-reporter constructs

Varying lengths of the 5' flanking region were amplified by PCR from human genomic DNA, cloned into pCRII or pCR2.1 (Invitrogen, Carlsbad, CA, U.S.A.) and then directionally subcloned into pGL3basic upstream of the luciferase-encoding region. All constructs were verified by sequencing and included 40 nt of the 5' untranslated region of human γ ENaC. Deletional mutants of some constructs were made by using convenient

restriction enzyme sites within genomic fragments. The –121 to +40 sequence was also cloned into pGL3enhancer, which contains a simian virus 40 (SV40) enhancer sequence downstream of the luciferase-encoding region but lacks a promoter sequence to drive transcription.

Cell culture and transfections

H441, HT29 and FRTL5 cells, obtained from American Type Culture Collection (A.T.C.C., Manassas, VA, U.S.A.), and M-1 cells (a gift from Geza Fejes-Toth) were cultured as described previously [16,22,23]. Initially, H441 cells grown in 12 wells were transfected with 1 μg of luciferase construct and 4 μg of pSV- β -gal (Promega, Madison, WI, U.S.A.) complexed with Lipofectin[®] (Life Technologies, Gaithersburg, MD, U.S.A.). At 48 h after transfection, cytoplasmic lysates were made and luciferase activity was determined with a luciferase assay kit (Promega) in a Beckman liquid-scintillation counter; β -galactosidase activity was determined with a standard assay kit (Promega) in a Bio-kinetics EL312e ELISA plate reader [22]. Subsequent H441 transfections were done with 1 μg each of luciferase and β -galactosidase plasmids and reporter gene activities were determined in a Monolight 2010 luminometer. M-1 cells were transfected with Lipofectamine Plus[®] (Life Technologies) with pSV- β -gal as the control plasmid; luciferase and β -galactosidase activities were measured as described for H441 cells. HT29 cells were transfected with Lipofectamine Plus[®] by using pRL-SV40 (Promega) as the control plasmid, and firefly and sea pansy luciferase activity were measured sequentially with the Dual Luciferase Reporter Assay kit (Promega) [16].

Primer extension analysis

Total RNA was extracted from H441 cells transfected with the –121 +40/pGL3enhancer construct. An oligonucleotide primer, GL2 (5'-CTTTATGTTTTGGCGTCTTC-3'), complementary to a portion of the luciferase-encoding region was end-labelled and gel-purified. Primer extension analysis was performed with yeast tRNA or H441 RNA as described [21] and then subjected to PAGE alongside a labelled 1 kb ladder.

Preparation of nuclear extracts and gel mobility-shift assay

H441 and FRTL5 cells were grown until nearly confluent, scraped up into PBS with a rubber 'policeman' and washed twice with PBS. Nuclear extracts were made as described, with some modifications [24]. In brief, the cell pellet was resuspended in 10 mM Hepes (pH 8)/1.5 mM MgCl_2 /10 mM KCl/1 mM dithiothreitol for 15 min on ice and then sheared through a 25-gauge needle. The cell pellet recovered by centrifugation was resuspended in 20 mM Hepes (pH 8)/1.5 mM MgCl_2 /420 mM NaCl/0.2 mM EDTA/1 mM dithiothreitol/25% (v/v) glycerol/0.5 mM PMSF for 30 min on ice. After another centrifugation, the supernatant was mixed with 2 vol. of a solution that contained 20 mM Hepes, pH 8, 100 mM KCl, 0.2 mM EDTA, 1 mM dithiothreitol, 20% (v/v) glycerol and 0.5 mM PMSF and stored in aliquots at –70 $^\circ\text{C}$.

Single-stranded oligonucleotides that corresponded to the proximal GC box and overlapping putative activator protein 2 (AP-2) site were synthesized, and then annealed together by heat denaturation and slow cooling:

GC: ctaGGGGGCGGGGCTGGCCGGCCG

CCCCGCCCCGACCGGCCGGCctag

Internally truncated (GC-int) and 3'-end-truncated (GC-end) versions of the GC sequence and a non-specific (NS) double-stranded oligonucleotide were also synthesized:

GC-int: ctaGGGGGCTGGCCGGCC

CCCCGACCGGCCGctag

GC-end: ctaGGGGGCGGGGCTGGC

CCCCGCCCGGACCGctag

NS: gatcCACTAGTGTGTCCACAGTGTCTGCCTAG

GTGATCACACAGGTGTACAGGACGGATCctag

Double-stranded oligonucleotides were end-labelled with [γ -³²P]ATP and incubated with or without H441 or FRTL5 nuclear extracts in a reaction mixture that contained 20 mM Hepes, pH 7.9, 60 mM KCl, 5 mM MgCl₂, 2 mM dithiothreitol, 10 ng/ μ l poly(dI-dC), 5 μ g/ μ l BSA and 10% (v/v) glycerol. A 50-fold excess of non-radioactive oligonucleotides was used for competition experiments. For supershift assays, H441 nuclear extracts were preincubated with 1 μ l of anti-Sp1 antibody (Promega) or 1 μ l of anti-Sp3 antibody (Santa Cruz Biotechnology, Santa Cruz, CA, U.S.A.) and then added to labelled probe. Samples were analysed on a non-denaturing polyacrylamide gel in 0.5 $\times$ Tris/borate/EDTA buffer run at 150 V; the gel was dried and subjected to autoradiography.

Low-resolution S1 nuclease mapping

Supercoiled plasmid DNA containing -1248 bp of 5' flanking human γ ENaC sequence was digested first with 40 units/ μ g S1 nuclease (Ambion, Austin, TX, U.S.A.) for 15 min at 37 °C and then with 4 units/ μ g restriction enzyme for 60 min at 37 °C. In some cases, S1 nuclease digestion followed restriction enzyme digestion. To repair single-stranded overhangs, S1 nuclease-digested samples were treated with the Klenow fragment of *Escherichia coli* DNA polymerase in the presence of dNTPs. These samples were then recircularized with T4 DNA ligase, transformed into *E. coli* and several clones sequenced.

High-resolution S1 nuclease mapping

Supercoiled plasmid DNA was digested with S1 nuclease (0.1, 0.25 or 0.5 unit/ μ g) for 15 min at 37 °C. The samples were extracted with phenol/chloroform (1:1, v/v), precipitated with ethanol and then linearized with *Mlu*I or *Eco*RI, distal to the S1 nuclease-sensitive site, to map the anti-sense and sense strands respectively. Primers 14g (5'-CACCGCGCCTGGGCTGTCT-3') and 17g (5'-TGAGACGCTGAAGTGGGAGG-3') were end-labelled with [γ -³²P]ATP and annealed to samples digested with *Mlu*I and *Eco*RI respectively. The primers were then extended along the DNA template with *Taq* DNA polymerase at 72 °C for 10 min in a reaction mixture that included 50 mM KCl, 10 mM Tris/HCl, pH 9.0, 1% (v/v) Triton X-100, 3–3.5 mM MgCl₂ and 400 μ M dNTPs. Primer 17g was used separately to sequence the supercoiled plasmid DNA by the dideoxy chain-termination method (Sequenase 2.0; Stratagene, La Jolla, CA, U.S.A.) to generate a 1 bp ladder. Primer-extended products were run alongside sequenced products on a sequencing gel, to localize the S1 nuclease-sensitive sites accurately.

RESULTS

We have cloned the human γ ENaC gene and reported the primary sequence of portions of the 5' flanking region [21]. The proximal 5' flanking region does not contain a TATA or a CCAAT box, but contains two GC boxes at -21 and -52. To

begin to evaluate transcriptional regulatory regions of the human γ ENaC gene we have now isolated two additional cosmid clones that contain larger portions of the 5' flanking region. One of these clones, 365G3, includes 4 kb of the 5' flanking region, whereas clone 385E5 includes 20 kb of 5' flanking region and the proximal portion of the first intron (Figure 1A). We have determined the primary structure of approx. 3000 bp of 5' flanking sequence and identified several putative transcription-factor-binding sites (Figure 1B). We also analysed the 5' flanking region for repeat sequences by using the CENSOR program [25]. The 5' flanking region contains four *Alu* sequences, three of which are nearly full-length (Figure 1C), whereas the other contains the free right arm only [26]. We also identified a long PPy tract with regions of perfect mirror symmetry, suggesting that this region might be able to adopt a triplex DNA conformation (Figure 1C). This tract contained imperfect tandem tetranucleotide repeats that were consistent with a microsatellite sequence.

We first examined the proximal 5' flanking region for transcriptional regulatory activity by transfection and analysis of expressed luciferase gene activity in the H441 cell line, in which there is regulated expression of human γ ENaC [22]. A 459 bp construct had significant luciferase activity in H441 cells that was comparable to the transcriptional strength of an SV40 promoter (Figure 2). To delineate the core promoter, progressive 5' deletions in this region were made. A -121 bp construct that includes both GC boxes was almost as active as the -459 bp construct, whereas a -26 bp construct that included the more proximal GC box led to a decrease in activity although it was still significantly greater than that detected with the promoterless plasmid. These findings suggest that the GC boxes are essential for the core promoter activity and that the GC boxes together had a synergistic effect on transcription. To confirm the critical role of these motifs, a 60 nt deletion, selectively removing both GC boxes, was made within the -121 nt construct; this decreased luciferase activity to background levels (Figure 2).

Because we did not see evidence for an enhancer sequence within the first 459 nt of the 5' flanking region we wondered whether the human γ ENaC promoter could respond to a heterologous enhancer. To test this possibility, the -121 bp construct was tested with an SV40 enhancer sequence inserted downstream of the luciferase-encoding region. Compared with either the promoter or enhancer activity alone there was synergistic activation of reporter gene activity when both elements were combined in *cis* (Figure 3A). This indicates that human γ ENaC gene transcription can be up-regulated by positive transcriptional elements.

We next examined whether the measured luciferase activities from transfected γ ENaC constructs arise from transcripts that initiate correctly from the human γ ENaC promoter. The transcription start sites of human γ ENaC-luciferase chimaeric transcripts were determined by primer extension analysis of RNA extracted from H441 cells transfected with the -121+40/SV40 enhancer construct. A radiolabelled anti-sense primer from within the luciferase-encoding region was hybridized to H441 or yeast RNA and extended with reverse transcriptase; the products were analysed by PAGE. Two discrete bands 120 and 65 nt in length were seen in H441 RNA but not in yeast RNA (Figure 3B). The larger product was correlated with the transcription start site *in vivo* for human γ ENaC mRNA in lung and kidney [21]. The smaller product ended within plasmid 'filler' DNA and this probably represented the premature termination of reverse transcription, possibly owing to the secondary structure in this region. However, we cannot exclude the possibility that some luciferase constructs were aberrantly initiated.

Figure 1 Structure of the 5' end of the human γ ENaC gene, including 5' flanking sequences

(A) Restriction map of the 5' end of the human γ ENaC gene. Exons 1 and 2 and 17 kb of 5' flanking DNA are shown. The approximate extent of four overlapping cosmid clones, 385E5, 365G3, 320H2 and 404D2, are shown beneath. (B) The 5' flanking region of human γ ENaC is shown with putative transcription-factor-binding sites, including cAMP-response elements (CRE), AP-2 sites, a glucocorticoid response element (GRE) and two GC boxes. The 5' extent of various genomic fragments used to test transcriptional activity is shown. These fragments were created from the -2926 bp fragment either by using restriction enzyme sites (named) or by PCR-amplification (5' primer shown as arrow). (C) Alu repeats and the PPy tract in the 5' flanking region of human γ ENaC are shown. The sequences of the PPy tract with regions of mirror symmetry are indicated in bold.

To identify transcription factors that bind at the core promoter, a double-stranded oligonucleotide sequence that corresponded to the proximal GC box and an overlapping putative AP-2 site was synthesized, annealed together and examined by gel mobility-shift assays with nuclear extracts from H441 and FRTL5 cells. The FRTL5 cell line is a rat thyroid epithelial cell line in which the rat γ ENaC promoter is active [22]. When they are incubated with end-labelled oligonucleotides, several proteins are seen to retard their electrophoretic mobility in both H441 and FRTL5 cells (Figure 4A). The binding of four of these proteins, labelled p1 to p4, is specific because it can be competed for by an excess

of non-radioactive oligonucleotides but not by a non-specific competitor. When analysed with the TFSEARCH program and the TRANSFAC database [27], a stress response element (STRE), an alcohol dehydrogenase gene regulator 1 (ADRI)-binding site and a PacC-binding site were also noted within the 24 bp sequence. The Sp1, ADRI and STRE sites are also conserved in the rat γ ENaC promoter (Figure 4B). To determine the identity of these DNA-binding proteins, we tested the ability of two truncated versions of the GC oligonucleotide to compete for factors p1 to p4 (Figure 4B). GC-int, which had an internal 5 nt deletion that disrupts the GC box, did not compete for any of

Figure 2 Analysis of human γ ENaC 5' flanking sequences

Various γ ENaC genomic DNA constructs were coupled to luciferase (luc) and transfected into H441 cells along with pSV- β -gal, a control plasmid. Activity is compared with SV40 promoter and with the empty vector. Results are means $\pm$ S.E.M. for four separate determinations and are expressed as c.p.m. of luciferase/ μ -unit of β -galactosidase. The GC boxes in the 5' flanking region are shown as small open squares. * P < 0.001 compared with empty vector (Student's t test).

Figure 3 Further analysis of the transcriptional effect of the human γ ENaC promoter sequence

(A) Effect of a heterologous enhancer on human γ ENaC promoter activity. A 121 bp 5' flanking DNA fragment was cloned upstream of luciferase (luc) and then tested in the presence and the absence of an SV40 enhancer. Results are means $\pm$ S.E.M. for four separate determinations and are expressed as c.p.m. of luciferase/ μ -unit of β -galactosidase. * P < 0.01; # P < 0.005 compared with empty vector (Student's t test). (B) Primer extension analysis of luciferase mRNA in transfected cells. A diagram of a chimaeric construct of human γ ENaC (open bar) and luciferase (filled bar) is shown. The hatched bar (+40 to +70) is filler plasmid DNA. The starting position of primer used is shown relative to the brain human γ ENaC transcription start. Primer was 5' end labelled and hybridized to yeast RNA (lane 1) or H441 RNA (lane 2). Products of primer extension are shown. Two bands (arrowed) at 120 and 65 bp are seen with H441 RNA. The molecular mass markers (indicated at the right in bp) are a radiolabelled 1 kb DNA ladder (Life Technologies).

Figure 4 Gel mobility-shift analysis

(A) A gel mobility-shift assay is shown that used either H441 or FRTL5 nuclear extracts. In comparison with free probe (lanes 1 and 7), multiple proteins p1 to p4 that retard DNA mobility are seen with increasing nuclear extract. This is competitively displaced by an excess of non-radioactive specific DNA (GC) but not non-specific DNA (NS). A non-specific protein (N) is also seen. (B) Proximal GC boxes in human and rat γ ENaC promoters are compared and the extent of the GC oligonucleotide used in gel mobility assays is indicated with a dotted line. Below are shown two truncated versions of the GC oligonucleotide. Putative transcription-factor-binding sites are shown for each DNA sequence. (C) Gel mobility-shift assay with competition and with specific antibodies. Labelled oligonucleotides GC (lanes 1–7) or GC-end (lanes 8–14) were incubated with H441 nuclear extracts (except lanes 1 and 8) in the presence or the absence of an excess of the indicated non-radioactive oligonucleotides. In some cases nuclear extracts were preincubated with either anti-Sp1 or anti-Sp3 antibody.

these factors, whereas GC-end, which had a 3' 5 nt deletion, competed for all transcription factors (Figure 4C, lanes 4 and 5). Consistent with the notion that the 3' 5 nt are not required for transcription factor binding was the observation that the labelled GC-end oligonucleotide, when bound to H441 nuclear extracts, was also shifted by transcription factors p1 to p4 (Figure 4C, lane 9). Oligonucleotide GC-end is predicted to abolish the binding of AP-2 and PacC, whereas oligonucleotide GC-int should still be able to bind STRE and ADR1 (Figure 4B). Taken together, the results indicate that the DNA-protein complexes probably do not contain AP-2, PacC, STRE or ADR1. To determine the identity of the transcription factors that bound oligonucleotide GC, the effects of anti-Sp1, anti-Sp3 and anti-(AP-2) antibodies on these complexes were studied. The protein p1 bound to GC or GC-end could be 'supershifted' with an anti-Sp1 antibody, confirming its identity as Sp1 (Figure 4C, lanes 6 and 13). The binding of proteins p2 to p4 was supershifted or abolished by

anti-Sp3, confirming that Sp3 or a closely related transcription factor was contained in each of these three DNA-protein complexes (Figure 4C, lanes 7 and 14). None of the bands were altered by anti-(AP-2) antibody (results not shown). These results suggest that Sp1 and Sp3, but not AP-2, are part of the transcription complex that binds at the core promoter.

We next extended our analysis of transcriptional regulatory activity to an additional 2.5 kb of 5' flanking sequence. First we examined the entire region (–2926 to +40) for glucocorticoid responsiveness in H441 cells. Despite the presence of a hexameric sequence corresponding to a glucocorticoid response element half-site, no functional glucocorticoid-responsive sites were seen (results not shown). Deletional analysis demonstrated that a negative regulatory region was present between –2000 and –1259 bp that, when compared with the –1259 bp construct, decreased luciferase expression in H441 cells (Figure 5A). To determine whether this response was specific to lung epithelial

Figure 5 Analysis of human γ ENaC 5' flanking sequences

Various γ ENaC genomic DNA constructs coupled to luciferase and transfected into H441 (A), M-1 (B) and HT29 (C) cells along with pSV- β -gal (H441 and M-1) or pRLSV40 (HT29) as a control plasmid. Results are means $\pm$ S.E.M. for three separate determinations in duplicate. A negative transcriptional element maps between -2004 and -1248.

cells we tested these sequences in M-1 cells (a cortical collecting duct cell line) and HT29 cells (a colonic epithelial cell line). In both these cell lines the basal promoter was active, as was the negative regulatory element (Figures 5B and 5C). An *Afl*III site at -1525 was used to localize the negative regulatory element further. The -1525 construct also showed negative regulatory activity in H441, M-1 and HT29 cells, confirming that the regulatory element included sequences between -1525 and -1296 (Figure 6).

The identification of a PPy tract in the 5' flanking region of human γ ENaC prompted us to examine whether this region could adopt a triplex DNA structure. Such conformations, called H-DNA, are increasingly being recognized *in vitro* and *in vivo* and are thought to have a role in transcriptional regulation, genomic stability and meiotic recombination [28–30]. In this form of non-B DNA, one half of the polypurine strand forms a Watson–Crick duplex with its corresponding strand and the second half of the polypurine strand folds back into the major groove of the other half-duplex via Höögsteen base pairing. This leaves the corresponding strand of the second half unpaired and susceptible to digestion by S1 nuclease, a single-strand-specific endonuclease. The PPy strand in the human γ ENaC 5' flanking sequence has two regions of mirror symmetry (Figure 1C) that make it particularly likely to form triplex DNA structures. The more proximal region is a perfect mirror repeat, whereas the

Figure 6 Analysis of human γ ENaC 5' flanking sequences

As in Figure 5, various γ ENaC genomic DNA constructs were coupled to luciferase and transfected into H441 (A), M-1 (B) and HT29 (C) cells. Results are means $\pm$ S.E.M. for three separate determinations in duplicate. A negative transcriptional element maps between -1525 and -1248.

distal region has one incongruous base. To test our hypothesis, supercoiled plasmid DNA containing 1248 bp of 5' flanking sequence was digested with S1 nuclease and the nuclease-sensitive sites were mapped by restriction enzyme digestion. Treatment with S1-nuclease-linearized plasmid DNA and subsequent digestion with *Bgl*II, *Pvu*II and *Mlu*I identified three separate S1-nuclease-sensitive sites within the PPy tract (Figure 7, left panel, lanes 2–5 and 7). This S1 nuclease sensitivity was seen only with supercoiled plasmid because DNA linearized first with *Bgl*II was no longer S1-nuclease-sensitive (Figure 7, left panel, lanes 1 and 6).

To identify the linear extent of triplex DNA formation after digestion with S1 nuclease, the plasmid DNA ends were filled in with Klenow fragment, re-circularized and rescued by transformation of bacteria. Direct sequencing of several clones revealed that the deletions varied from 1 to 154 bases; in 10 of 11 clones the deletions included all or part of the downstream mirror repeat (Figure 7, right panel). The deleted region in one clone did not include either repeat sequence.

To map the S1 nuclease sensitive sites precisely and to identify the unpaired strand that did not participate in the triplex DNA formation, high-resolution S1 nuclease mapping was performed. Plasmid DNA was first digested with low concentrations of S1 nuclease and then a labelled primer corresponding to each strand was annealed to the S1-digested DNA and extended along this

Figure 8 High-resolution mapping of S1-nuclease-sensitive sites

The finding that an anti-Sp3 antibody can change the mobility of three distinct DNA-protein complexes on a PAGE gel suggests that an antibody may recognize more than one protein, that each complex represents a different Sp3 protein homo-oligomer, or

We also found evidence for a long PPy tract in the proximal 5' flanking region of human γ ENaC flanked by *Alu* repeats. The PPy tract consists of imperfect tetranucleotide repeats, (AAAG)_n.

and contains two regions of mirror symmetry. *Alu* repeats can be associated with tetranucleotide repeats and it has been noted that AAAG repeats preferentially associate with the oldest *Alu* class, *AluJ* [40]. We confirmed that the PPY tract can adopt an H-DNA conformation, as evidenced by its S1 nuclease sensitivity, and show two possible models for this sequence (Figure 9). There are several interesting and potentially important features of PPY tracts that can form triplex DNA. First, these sites can be transcriptionally active *in vitro*, and triple-stranded and single-stranded DNA binding proteins have been characterized [41–43]. Secondly, these elements can be polymorphic [44] and the sequence heterogeneity could account for variation in transcriptional control within a given species [45]. It is therefore possible that allelic variation in this region of human γ ENaC, by affecting local chromatin structure or by variation in transcription-factor-binding specificities, could alter basal or inducible levels of human γ ENaC RNA. We have no evidence so far that this site is transcriptionally active in the cell lines tested, although we have found evidence for polymorphism in this region (results not shown). Thirdly, expanding repeats are now recognized as a cause of several inherited neurodegenerative diseases [46,47]. At least one of these, Friedreich's ataxia, involves a triplet repeat sequence in the first intron that directly affects the transcriptional regulation of the frataxin gene [48]. The exact consequence of these repeat elements in the promoter region of human γ ENaC has yet to be determined.

We thank Norman Doggett for the gift of a human genomic library, and acknowledge the DNA synthesis and sequencing services provided by the University of Iowa DNA core facility. This work was supported by a grant from March of Dimes Birth Defects Foundation Research Grant #6-FY99-444. C.P.T. is an Established Investigator of the American Heart Association.

REFERENCES

- Garty, H. and Palmer, L. G. (1997) *Physiol. Rev.* **77**, 359–396
- Barbry, P. and Hofman, P. (1997) *Am. J. Physiol.* **273**, G571–G585
- Canessa, C. M., Schild, L., Buell, G., Thorens, B., Gautschi, I., Horisberger, J. D. and Rossier, B. C. (1994) *Nature (London)* **367**, 463–467
- Lingueglia, E., Renard, S., Waldmann, R., Voilley, N., Champigny, G., Plass, H., Lazdunski, M. and Barbry, P. (1994) *J. Biol. Chem.* **269**, 13736–13739
- McDonald, F. J., Price, M. P., Snyder, P. M. and Welsh, M. J. (1995) *Am. J. Physiol.* **267**, C1157–C1163
- Smith, P., Mackler, S., Weiser, P., Brooker, D., Ahn, Y., Harte, B., McNulty, K. and Kleyman, T. (1998) *Am. J. Physiol.* **274**, F91–F96
- Kopp, U. C., Matsushita, K., Sigmund, R. D., Smith, L. A., Watanabe, S. and Stokes, J. B. (1998) *Am. J. Physiol.* **275**, R1780–R1792
- Driscoll, M. and Chalfie, M. (1991) *Nature (London)* **349**, 588–593
- Huang, M. and Chalfie, M. (1994) *Nature (London)* **367**, 467–473
- Waldmann, R., Champigny, G., Bassilana, F., Heurteaux, C. and Lazdunski, M. (1997) *Nature (London)* **386**, 173–177
- Waldmann, R., Bassilana, F., Weille, J. D., Champigny, G., Heurteaux, C. and Lazdunski, M. (1997) *J. Biol. Chem.* **272**, 20975–20978
- Adams, C. M., Anderson, M. G., Motto, D. G., Price, M. P., Johnson, W. A. and Welsh, M. J. (1998) *J. Cell. Biol.* **140**, 143–152
- Drummond, H. A., Price, M. P., Welsh, M. J. and Abboud, F. M. (1998) *Neuron* **21**, 1435–1441
- Djelidi, S., Fay, M., Cluzeaud, F., Escoubert, B., Eugene, E., Capurro, C., Bonvalet, J.-P., Farman, N. and Blot-Chabaud, M. (1997) *J. Biol. Chem.* **272**, 32919–32924
- Zentner, M. D., Lin, H. H., Wen, X., Kim, K. J. and Ann, D. K. (1998) *J. Biol. Chem.* **273**, 30770–30776
- Sayegh, R., Auerbach, S. D., Li, X., Loftus, R., Husted, R., Stokes, J. B. and Thomas, C. P. (1999) *J. Biol. Chem.* **274**, 12431–12437
- Lin, H. H., Zentner, M. D., Ho, H.-L. L., Kim, K.-J. and Ann, D. K. (1999) *J. Biol. Chem.* **274**, 21544–21554
- Renard, S., Voilley, N., Bassilana, F., Lazdunski, M. and Barbry, P. (1995) *Pflügers Arch. Eur. J. Physiol.* **430**, 299–307
- Asher, C., Wald, H., Rossier, B. C. and Garty, H. (1996) *Am. J. Physiol.* **271**, C605–C611
- Stokes, J. B. and Sigmund, R. D. (1998) *Am. J. Physiol.* **274**, C1699–C1707
- Thomas, C. P., Doggett, N. A., Fisher, R. and Stokes, J. B. (1996) *J. Biol. Chem.* **271**, 26062–26066
- Thomas, C. P., Auerbach, S. D., Zhang, C. and Stokes, J. B. (1999) *Gene* **228**, 111–122
- Thomas, C. P., Auerbach, S. D., Stokes, J. B. and Volk, K. A. (1998) *Am. J. Physiol.* **274**, C1312–C1323
- Lee, K. and Green, M. (1990) *Methods Enzymol.* **181**, 20–31
- Jurka, J., Klonowski, P., Dagman, V. and Pelton, P. (1996) *Computers Chem.* **20**, 119–122
- Jurka, J. (1998) *Curr. Opin. Struct. Biol.* **8**, 333–337
- Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Kolpakov, F. A. et al. (1998) *Nucleic Acids Res.* **26**, 364–370
- Frank-Kamenetskii, M. D. and Mirkin, S. M. (1995) *Annu. Rev. Biochem.* **64**, 65–95
- Wells, R. D. (1996) *J. Biol. Chem.* **271**, 2875–2878
- Brahmachari, S. K., Sarkar, P. S., Raghavan, S., Narayan, M. and Maiti, A. K. (1997) *Gene* **190**, 17–26
- Phillipsen, S. and Suske, G. (1999) *Nucleic Acids Res.* **27**, 2991–3000
- Hagen, G., Muller, S., Beato, M. and Suske, G. (1994) *EMBO J.* **13**, 3843–3851
- Kwon, H. S., Kim, M. S., Edenberg, H. J. and Hur, M. W. (1999) *J. Biol. Chem.* **274**, 20–28
- Kennett, S. B., Udvadia, A. J. and Horowitz, J. M. (1997) *Nucleic Acids Res.* **25**, 3110–3117
- Kapitonov, V. and Jurka, J. (1996) *J. Mol. Evol.* **42**, 59–65
- Vansant, G. and Reynolds, W. F. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8229–8233
- Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., Iglehart, J. D., Deininger, P. L. and McDonnell, D. P. (1995) *J. Biol. Chem.* **270**, 22777–22782
- Hewitt, S. M., Fraizer, G. C. and Saunders, G. F. (1995) *J. Biol. Chem.* **270**, 17908–17912
- Hu, X. and Worton, R. G. (1992) *Hum. Mutat.* **1**, 3–12
- Yandava, C. N., Gastier, J. M., Pulido, J. C., Brody, T., Sheffield, V., Murray, J., Buetow, K. and Duyk, G. M. (1997) *Genome Res.* **7**, 716–724
- Brunel, F., Zakin, M. M., Buc, H. and Buckle, M. (1996) *Nucleic Acids Res.* **24**, 1608–1615
- Bayarsaihan, D. and Lukens, L. N. (1996) *Biochem. J.* **314**, 293–296
- Guieysse, A. L., Praseuth, D. and Helene, C. (1997) *J. Mol. Biol.* **267**, 289–298
- Sunden, S. L., Businga, T., Beck, J., McClain, A., Gastier, J. M., Pulido, J. C., Yandave, C. N., Brody, T., Ghazizadeh, J., Weber, J. L. et al. (1996) *Genomics* **32**, 15–20
- Kennedy, G. C., German, M. S. and Rutter, W. J. (1995) *Nat. Genet.* **9**, 293–298
- Reddy, P. S. and Housman, D. E. (1997) *Curr. Opin. Cell Biol.* **9**, 364–372
- Paulson, H. L. and Fischbeck, K. H. (1996) *Annu. Rev. Neurosci.* **19**, 79–107
- Ohshima, K., Montermini, L., Wells, R. D. and Pandolfo, M. (1998) *J. Biol. Chem.* **273**, 14588–14595

Received 15 October 1999/6 December 1999; accepted 12 January 2000