Evolution and structural diversification of hyperpolarization-activated cyclic nucleotide-gated channel genes

Heather A. Jackson,1 Christian R. Marshall,2 and Eric A. Accili1

1Department of Cellular and Physiological Sciences, Faculty of Medicine, University of British Columbia, Vancouver; and 2Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

Submitted 3 July 2006; accepted in final form 9 January 2007

Jackson HA, Marshall CR, Accili EA. Evolution and structural diversification of hyperpolarization-activated cyclic nucleotide-gated channel genes. Physiol Genomics 29: 231–245, 2007. First published February 13, 2007; doi:10.1152/physiolgenomics.00142.2006.—Hyperpolarization-activated cyclic nucleotide-gated (HCN) channels are members of the voltage-gated channel superfamily and play a critical role in cellular pace-making. Overall sequence conservation is high throughout the family, and channel functions are similar but not identical. Phylogenetic analyses are imperative to understand how these genes have evolved and to make informed comparisons of HCN structure and function. These have been previously limited, however, by the small number of available sequences, from a minimal number of species unevenly distributed over evolutionary time. We have now identified and annotated 31 novel genes from invertebrates, urochordates, fish, amphibians, birds, and mammals. With increased sequence numbers and a broader species representation, a more precise sequence comparison was performed and an evolutionary history for these genes was constructed. Our data confirm the existence of at least four vertebrate paralogs and suggest that these arose via three duplication and diversification events from a single ancestral gene. Additional lineage-specific duplications appear to have occurred in urochordate and fish genomes. Based on exon boundary conservation and phylogenetic analyses, we hypothesize that mammalian gene structure was established, and duplication events occurred, after the divergence of urochordates and before the divergence of fish from the tetrapod lineage. In addition, we identified highly conserved sequence regions that are likely important for general HCN functions, as well as regions with differences conserved among each of the individual paralogs. The latter may underlie more subtle isoform-specific properties that are otherwise masked by the high identity among mammalian orthologs and/or inaccurate alignments between paralogs.

HCN channels are members of the voltage-gated cation channel superfamily and, based on sequence homology, are most closely related to the cyclic nucleotide-gated (CNG) channel and ether-a-go-go (EAG) potassium channel families. Individual subunits (Fig. 1) are predicted to have six transmembrane segments (S1–S6), with a voltage sensor domain in the S4 segment and an ion conducting pore between S5 and S6. Based on the crystal structures of related potassium channels (14, 27, 32) it is proposed that four HCN subunits come together to form a tetramer around a central pore. The distal termini of each subunit are cytoplasmic and from the crystal structure of the COOH terminus (78), it is now known that an α-helical linker region joins the transmembrane region to an evolutionarily conserved cyclic nucleotide-binding domain (CNBD).

To date, four mammalian isoforms have been cloned (HCN1–HCN4) (26, 35, 57). These paralogs exhibit 80–90% sequence identity between the start of S1 to the end of the CNBD. Differences in both sequence and length between paralogs occur predominantly in the NH2 and COOH termini (36). This high sequence identity suggests that the four mammalian HCN genes arose from duplications of a single ancestral gene prior to the lineage divergence. However, when these duplication and diversification events occurred remain unknown. The functional properties of the individual isoforms, such as their responses to changes in voltage and modulation of channel opening by cAMP, are similar but not identical. Likewise, the four HCN paralogs display partial overlapping tissue distribution in both the heart (26, 45, 46, 63) and the central nervous system (36). Despite what seems like redundant functional and expression profiles, the overall physiological importance of the individual isoforms is clearly evident from the different phenotypes of knockout mouse models. General and forebrain-specific HCN1 knockout mice demonstrate learning and movement dysfunction (49); HCN2-deficient mice exhibit absence epilepsy as well as cardiac sinus dysrhythmia (34); and HCN4-deficient mice die during embryonic development, possibly due to reduced HCN-mediated currents and heart rate (66). Furthermore, the physiological impact of the mammalian HCN4 isoform in the heart has been confirmed by the association of mutations in this isoform to sinus arrhythmias in humans (44, 60, 71). This pattern of
Fig. 1. Schematic representation of the overall topology of hyperpolarization-activated cyclic nucleotide-gated (HCN) channels. Included are the 6 putative transmembrane segments (S1–S6), including an S4 voltage sensor (light gray) and the reentrant P-loop region between S5 and S6, all structural characteristics common to the entire voltage-gated potassium channel superfamily, including ERG, CNG, and HCN channels. HCN and ERG channels also share the potassium selectivity filter GYFG. Based on the crystal structure of mHCN2 (78), the C-linker region in HCN channels contains 6 alpha-helices (A’–E’) and connects the cytoplasmic end of S6 to the start of the cyclic nucleotide binding domain (CNBD). The CNBD consists of 2 alpha-helices (A and B) that flank an 8-stranded beta-roll followed by a 3rd alpha-helix at its COOH-terminal end (C-helix). The cyclic nucleotides bind to a pocket that is formed by the beta-roll and the C-helix. Thick black lines delineate the region that is included in the phylogenetic analyses. In HCN channels, these lines encompass vertebrate exons 2 through 7. Alignments of HCN vertebrate isoform-specific NH2-terminal (exon 1) and COOH-terminal (exon 8) regions include those residues that fall outside of this region.

In addition to the four mammalian isoforms, only a few HCN genes have been cloned from lower vertebrate and invertebrate species. A single HCN1 ortholog has been cloned from the rainbow trout (9), whereas one and two HCN homologs have been cloned from arthropods (20, 21, 30, 43) and the sea urchin (16, 18), respectively. These sequences demonstrate considerable identity with the mammalian homologs, as well as some intriguing differences that provide preliminary clues about the evolutionary relationships among HCN genes. But due to the lack of sequence representation from a diverse sampling of species distributed throughout evolutionary history, previous phylogenetic analyses of the HCN family and its relatives have been limited and have yielded inconsistent patterns of evolution (10, 16). Furthermore, the current sampling of the four vertebrate isoforms is limited primarily to closely related mammalian sequences. The high residue conservation among these sequences, combined with the lack of sequences from more distantly related vertebrates, renders comparative analyses of the orthologs from the individual vertebrate isoforms ineffective.

In this study, we report the first thorough phylogenetic analysis and sequence comparison of the HCN gene family. By performing an extensive search of the currently available protein and whole-genome databases we derived a comprehensive list of known and putative full-length sequences for HCN homologs from a wide variety of species including urochordates and lower vertebrates. The increased number of sequences and broadened species representation provide information about HCN gene structure at critical periods during its evolutionary history. We identified sequences that are conserved and likely important for general HCN functions, as well as regions that may underlie more subtle differences in function among the different isoforms. Exon structure and genomic organization were also determined. These analyses provide insight into the molecular evolution of this protein within different taxa and support the hypothesis that both lineage-specific and ancestral duplication and divergence events of the HCN genes have occurred throughout its history.

MATERIALS AND METHODS

Sequence data. Currently available HCN protein sequences were identified through BLASTP (1) searches of the National Center for Biotechnology Information (NCBI) and UniProt/SwissProt nonredundant protein databases. Human HCN2 (GenBank GI no. 21359848), trout HCN1 (GenBank GI no. 33312350) or drosophila HCN (GenBank GI no. 5326833) amino acid sequences were used as queries and default parameters were applied. To prevent the inclusion of incorrect sequences generated by computer prediction programs, all computer-derived annotations were temporarily removed until confirmed by sequence alignment or database annotation described below. Splice variants, short fragments, and other duplicates were also removed.

The Ensembl Genome Browser (http://www.ensembl.org/) (6, 24) was used to determine the genomic position and distribution of the established HCN genes and either to examine new HCN genes identified by the computer gene-prediction programs and annotation process or to identify novel genes. Sequences classified as HCN genes by Ensembl were examined, and those that spanned the entire length of known sequences were downloaded. Protein annotations that resembled HCN but either lacked regions of the predicted sequence or showed signs of additional exons or exon fragments were not included. However, the genomic DNA underlying these protein predictions were used as further reference to help in the manual reannotation of the protein sequence. TBLASTN (1) searches of the available genome databases using either the low-sensitivity default parameters (optimized for near-exact matches) or medium-sensitivity default parameters (optimized to allow for local mismatch) were conducted to identify any further genes or genomic regions that showed significant sequence identity to the peptide query sequence used. In total, 13 genomes were analyzed, including: zebrfish (Danio rerio) (v. 35.5b), Japanese pufferfish (Fugu rubripes) (v. 29.2e), green pufferfish (Tetraodon) (v. 31.1c), opossum (Monodelphis domestica) (v35.2), dog (Canis familiaris) (v. 35.1d), cow (Bos taurus) (v36.2), chimpanzee (Pan troglodytes) (v.31.2a), clawed frog ( Xenopus tropicalis) (v. 31.1a), chicken (Gallus gallus) (v.35.1k), sea squirt (Ciona intestinalis) (v. 35.195b), Pacific sea squirt (Ciona savignyi) (CSAV2.0), mosquito (Anopheles gambiae) (v. 23.2b.1), and worm (Caenorhabditis elegans) (v.29.130). Human HCN2, trout HCN1, and drosophila HCN sequences were used as query sequences for vertebrate, fish, and insect genomes, respectively. For the urochordate genomes, sea urchin HCN (GenBank GI no. 74136757), trout HCN1, and human HCN2 sequences were used. In general, similar TBLASTN results were found regardless of the query sequence used, reflecting the high degree ofsequence identity found within the core region throughout the HCN family. Full-length putative protein sequences were constructed from the conceptual translation of genomic DNA as previously described (42) with the proposed starting methionine in vertebrates supported by the presence of a consensus start sequence (29). Genome position and intron-exon structure were ex-
amined and recorded. A list of the sequences used in the analysis is shown in Table 1.

**Multiple sequence alignments.** We performed protein sequence alignments using the default parameters of ClustalX (version 1.83) (69). Alignments were subsequently examined and edited in GeneDoc (48). To ensure that the HCN genes identified were full length and nonredundant, those lacking a putative start codon were discarded. Isoform pairs identified in some of the fish genomes were included if they corresponded to a different location in the genome, as they are likely the result of a polyploidization event that is thought to have

Table 1. List of HCN sequences used in analyses

<table>
<thead>
<tr>
<th>Name</th>
<th>Organism</th>
<th>Common Name</th>
<th>Taxonomy</th>
<th>Identifier</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCN_silkmoth (Hvlh)</td>
<td>Heliolus virescens</td>
<td>silkmoth</td>
<td>Intv.</td>
<td>GI:3970750</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN_bee (Amih)</td>
<td>Apis mellifera</td>
<td>honey bee</td>
<td>Intv.</td>
<td>GI:3335927</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN_lobeater (Palh)</td>
<td>Panulirus argus</td>
<td>spiny lobster</td>
<td>Intv.</td>
<td>GI:3335925</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN_fly (Dmih)</td>
<td>Drosophila</td>
<td>fruit fly</td>
<td>Intv.</td>
<td>GI:5326833</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN_urchin (Spih/SpHCN)</td>
<td>Stronglyocentrotus purpuratus</td>
<td>sea urchin</td>
<td>Intv.</td>
<td>GI:47551101</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN2_urchin (spHCN2)</td>
<td>Stronglyocentrotus purpuratus</td>
<td>sea urchin</td>
<td>Intv.</td>
<td>GI:74136757</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN1_mouse (mHCN1)</td>
<td>Mus musculus</td>
<td>mouse</td>
<td>Mam.</td>
<td>GI:6754168</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN1_trout (HHCN1)</td>
<td>Oncorhynchus mykiss</td>
<td>rainbow trout</td>
<td>Vt.</td>
<td>GI:33312350</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN1_human (HHCN1)</td>
<td>Homo sapiens</td>
<td>human</td>
<td>Mam.</td>
<td>GI:32698746</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN1_rabbit (rHCN1)</td>
<td>Oryctolagus cuniculus</td>
<td>rabbit</td>
<td>Mam.</td>
<td>GI:38605639</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN3_mouse (mHCN3)</td>
<td>Mus musculus</td>
<td>mouse</td>
<td>Mam.</td>
<td>GI:6680191</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN3_human (HHCN3)</td>
<td>Homo sapiens</td>
<td>human</td>
<td>Mam.</td>
<td>GI:38327037</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN3_rat (rHCN3)</td>
<td>Rattus norvegicus</td>
<td>Norway rat</td>
<td>Mam.</td>
<td>GI:29840772</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN3_cow</td>
<td>Bos taurus</td>
<td>cow</td>
<td>Mam.</td>
<td>GI:76612489</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN3_dog</td>
<td>Canis familiaris</td>
<td>dog</td>
<td>Mam.</td>
<td>GI:73954256</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN4_human (HHCN4)</td>
<td>Homo sapiens</td>
<td>human</td>
<td>Mam.</td>
<td>GI:4885407</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN4_mouse (mHCN4)</td>
<td>Mus musculus</td>
<td>mouse</td>
<td>Mam.</td>
<td>GI:29840776</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN4_rabbit (rHCN4)</td>
<td>Oryctolagus cuniculus</td>
<td>rabbit</td>
<td>Mam.</td>
<td>GI:38605640</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN4_rat (rHCN4)</td>
<td>Rattus norvegicus</td>
<td>Norwegian rat</td>
<td>Mam.</td>
<td>GI:29840771</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN4_lobster (HHCN4)</td>
<td>Canis familiaris</td>
<td>dog</td>
<td>Mam.</td>
<td>GI:7400965</td>
<td>GenBank</td>
</tr>
<tr>
<td>HCN1_chimpanzee*</td>
<td>Pan troglodytes</td>
<td>chimpanzee</td>
<td>Mam.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4_chimpanzee†</td>
<td>Pan troglodytes</td>
<td>chimpanzee</td>
<td>Mam.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN1_ophossum*</td>
<td>Monodelphis domestica</td>
<td>opossum</td>
<td>Mam.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN3_ophossum*</td>
<td>Monodelphis domestica</td>
<td>opossum</td>
<td>Mam.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4_ophossum*</td>
<td>Monodelphis domestica</td>
<td>opossum</td>
<td>Mam.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN2_chicken†</td>
<td>Gallus gallus</td>
<td>red jungle fowl</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4_chicken†</td>
<td>Gallus gallus</td>
<td>red jungle fowl</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN1_frog†</td>
<td>Xenopus tropicalis</td>
<td>frog</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN2_frog*†</td>
<td>Xenopus tropicalis</td>
<td>frog</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN1_green_puffer*</td>
<td>Tetraodon nigroviridis</td>
<td>spotted green</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4a_green_puffer*</td>
<td>Tetraodon nigroviridis</td>
<td>spotted green</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4b_green_puffer*</td>
<td>Tetraodon nigroviridis</td>
<td>spotted green</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN2a_green_puffer*</td>
<td>Tetraodon nigroviridis</td>
<td>spotted green</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN2b_green_puffer*</td>
<td>Tetraodon nigroviridis</td>
<td>spotted green</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN3_green_puffer*</td>
<td>Tetraodon nigroviridis</td>
<td>spotted green</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN3_zebrafish*</td>
<td>Danio rerio</td>
<td>zebrafish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4_zebrafish*</td>
<td>Danio rerio</td>
<td>zebrafish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN2b_fugu†</td>
<td>Takifugu rubripes</td>
<td>Japanese pufferfish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN2a_fugu*</td>
<td>Takifugu rubripes</td>
<td>Japanese pufferfish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN3a_fugu*</td>
<td>Takifugu rubripes</td>
<td>Japanese pufferfish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN3b_fugu*</td>
<td>Takifugu rubripes</td>
<td>Japanese pufferfish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4a_fugu*</td>
<td>Takifugu rubripes</td>
<td>Japanese pufferfish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN4b_fugu*</td>
<td>Takifugu rubripes</td>
<td>Japanese pufferfish</td>
<td>Vt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN_mosquito*</td>
<td>Anopheles gambiae</td>
<td>mosquito</td>
<td>Invt.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN_c_intestinalis*</td>
<td>Ciona intestinalis</td>
<td>sea squirt</td>
<td>Uro.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCNb_c_intestinalis*</td>
<td>Ciona intestinalis</td>
<td>sea squirt</td>
<td>Uro.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN_c_savignyi*</td>
<td>Ciona savignyi</td>
<td>Pacific sea squirt</td>
<td>Uro.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCNb_c_savignyi*</td>
<td>Ciona savignyi</td>
<td>Pacific sea squirt</td>
<td>Uro.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
<tr>
<td>HCN_c_savignyi*</td>
<td>Ciona savignyi</td>
<td>Pacific sea squirt</td>
<td>Uro.</td>
<td>NA</td>
<td>Ensembl</td>
</tr>
</tbody>
</table>

Naming of sequences: hyperpolarization-activated cyclic nucleotide-gated (HCN) isoform followed by the species. Previous names for sequences are given in parentheses. For duplicates, ‘a’ and ‘b’ labels were assigned arbitrarily. *Those that were reannotated or identified in this study; †sequence missing at least 1 internal exon and therefore removed from phylogenetic analyses. Invt., invertebrate; Mam., mammal; Vt., lower vertebrate; Uro., urochordate. NA, protein identifier was not available for the sequence included in final analysis.
occurred early in their lineage (Ref. 23 and references therein). Partial sequences that lacked a portion of the distal COOH terminus corresponding to the last mammalian exon were included as this region was not used in the phylogenetic analyses due to the high degree of length and sequence variability among the different HCN genes. For the NH2- and COOH-terminal alignments, vertebrate sequences corresponding to the region upstream of S1 or different downstream of the CNBD (see Fig. 1) were individually realigned for HCN1–4 and were shaded based on sequence conservation of 100% (black), 80% (dark gray), and 60% (light gray).

**Phylogenetic analyses.** Sequences were trimmed to produce a core alignment spanning from the start of the transmembrane segment S1 to the end of the CNBD (see Fig. 1). This region exhibits high sequence identity within the HCN family and among the other related sequences that were used. As there is no known bacterial HCN channel, and based on a previous phylogenetic analysis of HCN and CNG channels (10), KAT1 (GenBank GI no. 44888080), human ERG1 (GenBank GI no. 7531135), human CNGA1 and CNGA3 (GenBank GI no. 2506302 and 13959682) sequences were included in the analyses to serve as an out-group for the rooted phylogenetic trees of the HCN family. Six sequences that were missing exons due to gaps in the genome assembly were removed from the alignment prior to running the programs. Neighbor-joining (NJ) trees were generated using ClustalX, followed by tree evaluation with bootstrap resampling of 1,000 times. Additional NJ, maximum parsimony (MP), and maximum likelihood (ML) trees were created using the Seqboot, Protdist, Neighbor, Protpars, Proml, and Consense programs from the PHYLIP package (version 3.65), bootstrapping with 100 replicates with randomized input order and 10 jumbles (15). The TreeView program (version 1.6.6) (51) was used to examine and display all trees.

**RESULTS AND DISCUSSION**

To date, 21 unique full-length HCN coding sequences have been cloned and deposited in GenBank: 4 from arthropods, 2 from sea urchin, 1 from fish, and 14 from mammals, specifically human, rabbit, mouse and rat. Previous phylogenetic analyses of the HCN family and its relatives have been limited and inconsistent, in part due to the lack of available sequences from a diverse sampling of species. For example, the product of the first duplication event has been shown to be either HCN1 (10) or HCN3 (16). Furthermore, the recently cloned spHCN2 (16) sequence was shown to group with the HCN family but was situated at the base of the phylogenetic tree. This is not consistent with species evolution, which places the sea urchin genes with the other deuterostome species, after the arthropod clade division. Therefore, spHCN2 may instead represent either a new family of channels not yet identified in any other species or an HCN homolog that was the product of a duplication event followed by rapid sequence divergence. These inconsistencies underscore the need for more comprehensive analyses. Accordingly, the aim of this study was to expand the breadth of available HCN sequences through the manual annotation of available genomes and in turn, to provide the first comprehensive sequence comparison and phylogenetic analysis of this protein family.

**HCN genes are present in multiple copies across a wide spectrum of species.** Using a BLASTP search at NCBI or by searching the genome databases available at the Ensembl website, we identified 58 nonredundant HCN sequences. Twenty-seven of these sequences were previously identified by cloning or by computer annotation and confirmed by genomic data. The remaining 31 are novel and were completed by data mining the genomes available at Ensembl or the manual reannotation of the predicted protein translations. Several of the Ensembl predictions were inconsistent with known sequences, or the predicted gene did not span the estimated length of the transcript. The inherent problems in gene prediction and computer protein annotation methods have been previously described (Ref. 42 and references therein). We resolved these problems by multispecies comparisons and manual reannotation on a gene-by-gene basis. A complete list of the sequences included in this analysis is shown in Table 1 and their protein sequences can be found in Supplementary Fig. S1. (The online version of this article contains supplementary material.) Included in the final list are: 23 mammalian sequences; including 2 and 3 new sequences from the chimpanzee and opossum genomes, respectively; 22 lower vertebrate sequences; 6 urochordate sequences, including 3 new sequences from both C. intestinalis and C. savignyi genomes; and 7 invertebrate sequences, with a new annotation of the single HCN gene found in the mosquito genome. Of the lower vertebrates, 2 new sequences were from chicken, 3 from frog, 6 new sequences from both the green pufferfish and the Japanese pufferfish genomes, and 2 new sequences from zebrafish. No HCN sequences were identified in C. elegans. With this substantial increase in the number of full-length sequences representing a wide range of species, we reconstructed the evolutionary history of this protein family and probed the relationship of channel structure and function in greater detail than was previously possible.

**High sequence identity among four vertebrate HCN isoforms within the core region.** Due to the length and sequence variability that occurs in the NH2 and COOH termini (see Supplementary Fig. S2), these regions were trimmed and a core alignment was produced corresponding to a region between S1 and the end of the CNBD and approximately to exon 2 through 7 of the mammalian isoforms. Sequence conservation of the four vertebrate isoforms within this region is high, at least 80–90% identity amongst the mammalian sequences and over 90% residue conservation between the newly identified fish sequences and their respective mammalian orthologs (Fig. 2). This indicates that this region has been slow to evolve during the 450 million years (MY) that separate the fish and human lineages. Also of interest is that HCN4 shares the highest sequence identity with all other vertebrate isoforms and that the mammalian HCN3 and fish HCN3 sequences are as similar to the other vertebrate isoforms as they are to each other. This suggests that HCN4 sequences have diverged the least from a common ancestral sequence, as is further evident from the branch lengths in the NJ and ML phylogenetic trees, and that HCN3 sequences have evolved independently within the fish and mammalian lineages to an equal degree but at different sites. Overall, the sequences of the invertebrates and urochordates display a lower conservation with the mammalian sequences. Arthropods share a general sequence identity of ~60–65% with the mammalian homologs, whereas one of the sea urchin sequences, spHCN1 (also known as spH or spHCN), and two of the Ciona homologs, here named HCNb and HCNa, are only 55% identical. The second sea urchin sequence, spHCN2, and the other sequence from the Ciona species (HCNC) are even more diverged.
Three different HCN duplication events occurred prior to the divergence of the fish lineage. To examine the phylogeny of the HCN family, a multiple sequence alignment of the core region was created using 52 of the 58 identified sequences (see MATERIALS AND METHODS). Three different phylogenetic programs, two NJ methods and one MP method, produced similar results. Due to the high sequence conservation among mammalian sequences, ML methods were incapable of analyzing the complete list of sequences. In a subgroup of 41 sequences, ML methods produced similar results to the NJ and MP trees. Figure 3 is a rooted MP consensus tree produced by the PHYLIIP program. The NJ rooted phylogram created by ClustalX and a single representative ML phylogram produced by PHYLIIP are provided in Supplementary Fig. S3. Four sequences were included as an out-group: human CNGA1 and CNGA3, human ERG1 and KAT1. Tree topology did not change when any of these out-group sequences were removed.

HCN sequences did not group according to species but rather within isoform clades, indicating that the HCN1–4 isoforms are paralogs. Genes from fish, birds, and amphibians partitioned within the four mammalian clades, which strongly suggests that the four isoforms were established prior to the origin of the vertebrate clade and were present in the common ancestor of fish and tetrapods. From the tree topology, we also predict that an HCN ancestor underwent three separate gene duplication events prior to the divergence of the fish lineage, over 450 million years ago (MYA). Our data are inconsistent with two rounds of whole genome duplication and suggest that the HCN family evolved through independent gene duplications or chromosomal block duplication (25). However, due to a relatively low sequence resolution, the 2R hypothesis cannot be ruled out. Of the four HCN isoforms, the clade belonging to HCN3 was the first to diverge and represents the product of the first duplication event. This position of HCN3 is supported by all four phylogeny methods and by a high bootstrap value and supports the findings of Galindo and colleagues (16). Functional data will be required to further explain this branching position. The low bootstrap value at the division of the HCN3 clade in the MP and NJ trees indicates that this subdivision remains unresolved. We hypothesize that this clade division may be due to the independent sequence evolution which has occurred in the teleost and tetrapod HCN3 genes following the lineage divergence. Interestingly, the HCN3 clade is not divided in the ML tree for which the Jones-Taylor-Thornton model of evolution was imposed. We cannot rule out the possibility that this difference is due to the decreased number of sequences used in the ML analysis. The resolution of the HCN3 clade will only be improved with the inclusion of more complete sequences from species that evolved following the emergence of the four HCN paralogs but prior to the divergence of mammals. Functional data of the HCN3 clade from lower vertebrates will also be required to conclusively determine whether a functional divergence occurred within this clade.

Our data also suggest that HCN4 is the product of the second duplication event followed by the emergence of HCN1 and HCN2. This proposed evolutionary pattern is evident in all three phylogenetic trees and is further supported by the sequence conservation pattern shown in Fig. 2. On the basis of phylogenetic results, the division between HCN1 and HCN2 remains unresolved. Nevertheless, functional data suggest a
clear difference between these two isoforms. The discrepancy with the predicted order of species evolution within each mammalian clade is likely due to the high sequence conservation seen within this core region. This results in a limited number of informative sites and produces a low phylogenetic signal (17). The lack of sequence divergence in the mammalian clades may be due to insufficient evolutionary time to permit the accumulation of mutations and/or a strong selective pressure to retain the conserved sequence of this channel.

Fig. 3. A maximum parsimony (MP) consensus tree of the HCN family. We included 52 HCN sequences in this analysis. The MP tree was constructed by resampling 100 datasets (bootstrap) with randomized input order of 10 jumbles using Seqboot, Protpars, and Consense programs in the PHYLIP software. To control for length differences between the different isoforms, only the region between the start of S1 and the end of the CNBD was used. The tree is rooted with an out-group of KAT1, hERG1, CNGA1, and CNGA3, 4 sequences that are related to the HCN family. Similar to the previous studies, HCN2_urchin branches before the invertebrate clade, which may be a consequence of the MP method’s sensitivity to unequal rates of evolution. The six urochordate sequences from C. intestinalis and C. savignyi fall according to their evolutionary position, between the invertebrates and vertebrate species, which is supported by high bootstrap values. Because they group within their own species, they are not considered orthologous to any of the mammalian isoforms and are therefore named HCNa, HCNb, and HCNc. HCN3 is the product of the first duplication event. A split between the fish and mammalian sequences is observed but the partition is not well defined. HCN4 is the product of the second duplication event, followed by the emergence of HCN2 and HCN1. Low bootstrap values within isoform clades are most likely a consequence of the high sequence identity within the region used and in turn, the lack of informative sites required by the MP method. Duplicate copies of mammalian paralogs in the fish species are denoted with ‘a’ or ‘b’ in their sequence names. *Sequences annotated in this study. Numbers indicate bootstrap values and represent % support for the respective partitions.

Fish lineage show evidence of duplicate HCN genes. Genome data mining revealed multiple HCN genes in both the F. rubripes and Tetraodon species, in both cases exceeding the number found in the mammalian lineage. From the branching order in the phylogenetic tree shown in Fig. 3, it is clear that fish gene pairs group together within the clades of individual mammalian orthologs. This suggests that the common ancestor of teleost fish and tetrapods had four HCN genes and that these underwent further duplications independently in the fish lineage. These sequences have therefore been named according to their mammalian ortholog and subsequently designated ‘a’ or ‘b’. This pattern of duplication is clearly evident in the HCN2 and HCN4 clades, where the green puffer and fugu ‘a’ and ‘b’ sequences are grouped together and are separated from each other by high bootstrap values. Because the teleost fish are predicted to have undergone a complete genome duplication early in their own lineage (Ref. 23 and references therein), this distribution pattern is not unexpected. However, because some of the genes identified were not full length and some showed evidence of intron insertion, further analyses are required to determine whether these duplicates are expressed and functional or have become pseudogenes (73).

Ciona genes most likely arose through lineage-specific duplication events. In contrast to the observed pattern of the multiple fish genes, the multiple copies of HCN sequences found within the urochordate species do not partition with any of the mammalian isoform clades. However, each gene did show a highly significant pairing between C. intestinalis and C. savignyi, indicating that these duplication events occurred before these two species diverged. The timing of this duplication is unknown, but with the presence of multiple HCN genes previously identified in the sea urchin, two different scenarios are possible. First, one duplication of an ancestral gene may have occurred prior to the divergence of the deuterostomes and given rise to the multiple HCN genes seen in the sea urchin and urochordate species. A lineage-specific duplication then occurred in the urochordates to produce the third Ciona HCN gene. Subsequently, the genes evolved rapidly and independently in the different lineages and thus no longer resemble each other or any specific HCN isoform. The loss of one ancestral gene and the duplication and diversification of the
other would then have given rise to the four isoforms now common to all chordates. A second, and more parsimonious, pattern of evolution is that lineage-specific duplication events of a single HCN ancestor occurred and gave rise to the three HCN homologs within the urochordate lineage. Further ancestral duplication events occurred within the vertebrate lineage, after the divergence of the urochordates and prior to the fish lineage, and produced the four known mammalian isoforms. Lineage-specific gene duplication in *Ciona* has also been shown for the evolution of sodium channel gene family. Two sodium channel genes have been identified, but one possesses a sequence that has diverged considerably both from its paralogous pair and from the other known sodium channel gene sequences. The authors concluded that the duplication events occurred just prior to the fish lineage (33). Similarly, independent lineage-specific duplication was suggested for the ankyrin gene family based on their phylogenetic results and the differences in gene sequences between *Ciona* and the vertebrate homologs (7). The general branching order between the *Ciona* HCN homologs, in which HCNa and HCNc group together and HCNb is independent, is consistent among the different trees produced. Their position within the tree, however, is variable and is most likely a result of the different tree building methods used. It may also reflect the amount of lineage-specific sequence divergence that has occurred in these HCN genes, which has caused them to evolve independently of both the invertebrate and vertebrate clades and has blurred their phylogenetic position.

**Predicted phylogenetic patterns are supported by exon boundary structure.** Our phylogenetic analyses indicate that the four vertebrate isoforms arose via three duplications of an ancestral HCN gene. This is supported by the exon structure of their coding sequences. The four human HCN genes are located on different chromosomes: 1q21.2 (HCN3), 5p12 (HCN1), 15q24–q25 (HCN4) (61), and 19p13 (HCN2) (72). Using the genome databases, we found this pattern of differential localization for all vertebrate HCN genes, from fish through mammals. Furthermore, the overall exon structure of the four isoforms has remained consistent since the duplication and divergence events (Fig. 4). It comprises eight exons, with highly conserved length and sequence in exons 2 through 7. An exception to this is observed in both of the fish HCN3b genes, which are predicted to have an intron of <75 bp inserted in the middle of exon 2. Exons 1 and 8, which directly correspond with the distal NH2 and COOH termini, vary in both length and sequence for all vertebrate genes analyzed. In addition, exon boundary positions are highly conserved throughout the vertebrate lineage. This suggests that the extant vertebrate HCN sequences are derived from a single ancestral gene that had an exon structure similar to current mammalian HCN genes and that the duplication events occurred after the intron positions were fixed in the linear sequence.

Our phylogenetic results suggest that the invertebrate and urochordate HCN genes are homologs of the four vertebrate isoforms, but because they group outside the vertebrate clades, they are not orthologous to any one in particular. The presence...
of multiple genes in the urochordate species, therefore, raised the question as to where and when duplications occurred and what course of evolution led from the invertebrate to vertebrate gene structure. To address these questions, we compared the exon structures of invertebrates and urochordates to the highly conserved vertebrate structure. The exon structure of the single HCN gene in the arthropod lineage is quite different from the mammalian structure. Using the fly and mosquito genes as representatives, we found that the invertebrate genes comprise 13 or 14 shorter exons, separated by short introns. Of the seven boundary positions conserved throughout the vertebrate lineage, only two to four are conserved in these species. The exon structures of the urochordate genes, on the other hand, show similarities to both the invertebrate and vertebrate genes. The HCNb gene contains 15 exons separated by short introns, which is similar to the invertebrate structure. Seven of the exon boundaries parallel those found in the vertebrate structure. The other two urochordate genes, HCNa and HCNc, are different from both the vertebrate and invertebrate genes. HCNa contains 10 exons with only four boundaries conserved with the vertebrate structure and none with the invertebrate genes. The 13 exon boundaries of HCNc are not conserved with any other gene, including paralogs from within the same species. In conjunction with sequence identity, residue conservation, and phylogenetic analyses, these findings further support the hypothesis that HCNb represents the homolog most similar to the ancestral HCN gene of the urochordates and that HCNa and HCNc arose through lineage-specific duplication processes. Furthermore, we suggest that the ancestral gene of the HCN family contained numerous introns that were then lost during the course of evolution. Exons became fused to form the structure now seen within the vertebrate lineage.

Evolution of key residues in the voltage sensing domain and pore region. HCN channels open in response to changes in membrane voltage and allow for the passage of Na\(^+\) and K\(^+\) ions across the plasma membrane. The transmembrane domains of the individual subunits, which form tetramers around a central pore, are primarily responsible for these functions. As might be expected from the natural constraints of the hydrophobic bilayer, sequence conservation is abundant in areas predicted to correspond to the six transmembrane segments. Similar to depolarization-activated K\(^+\) channels, the fourth transmembrane segment (S4) contains positively charged residues and is likely to sense changes in voltage across the cell membrane (Fig. 5A, ‘a’). In contrast to depolarization-activated K\(^+\) channels, however, HCN channels open instead of close in response to hyperpolarization. Interestingly, HCN channels possess an additional four or five charged residues at the NH\(_2\)-terminal end of S4 (Fig. 5A ‘b’). Together, in response to changes in voltage, these charges have been shown to move in the same direction as that in depolarization-activated K\(^+\) channels (3). Therefore, it has been suggested that the movement of the voltage-sensing domain is coupled to channel opening in the opposite way (40). Throughout the HCN family, there is high sequence identity in this region among the vertebrate isoforms and high conservation across all sequences. The more diverged sequences from the sea urchin and urochordate species, spHCN2, Ciona HCNa and Ciona HCNc, which are predicted to be the products of lineage-specific duplication, possess only three of the upstream positive charges. The effect of this loss of charge awaits functional characterization but could be due to relaxed evolutionary constraints in these duplicate genes.

On the basis of their similarity to K\(^+\) channels (77), we predict that the HCN channels are in their lowest energy state when closed and that the voltage sensors work to open the gate upon hyperpolarization of the membrane potential. The S4–S5 linker couples the sensing of membrane voltage by S4 to the opening of the channel gate, which is located at the inner portion of the S6 segment. Channel opening is highly sensitive to mutations in the S4–S5 linker. Within this region (Fig. 5A ‘c’), a dipeptide of a tyrosine followed by an acidic residue is conserved across the vertebrate and urochordate channels but is not present in invertebrates. Mutation of the tyrosine residue to a polar or acidic amino acid in the mammalian HCN2 isoform disrupts channel closure (8, 38). Similar results were shown for the nearby arginine residue located at the end of this linker, which is conserved throughout the HCN family. These experiments confirm that this region is critical for the regulation of channel opening by changes in voltage. The lack of complete conservation within this region therefore suggests that the strength of coupling between the S4 segment and channel opening may be variable among the different HCN genes. For example, spHCN1 channels exhibit an inactivation process that is due to a desensitization of the opening of the intracellular gate to voltage, which may reflect a weaker coupling of the voltage sensor and gate (55).

Based on sequence homology to the crystal structure of a bacterial K\(^+\) channel (14), the region between S5 and the end of S6 is believed to form the ion conduction pore in HCN channels. It contains a pore helix and selectivity filter and is involved in both ion selectivity and transport. In K\(^+\) channels, the selectivity filter exhibits the K\(^+\) signature sequence (GY/FG), which is thought to confer their K\(^+\)-selective nature (14). In the HCN family, this tripeptide has been conserved but the channels allow the passage of significant amounts of both Na\(^+\) and K\(^+\). In all but two HCN genes, the sequence motif that corresponds to the selectivity filter is CIGYG (5) (Fig. 5B ‘a’). The conservation of this sequence only differs from K\(^+\)-selective channels at the cysteine residue, implicating this site’s involvement in the reduced K\(^+\) permeability. The two exceptions are again found in gene duplicates from sea urchin and urochordate species. spHCN2 has a filter sequence of SIGFG (16), which makes it more similar to the filter sequences of channels in the EAG K\(^+\) channel family. Functional data do not exist for the spHCN2 channel, so whether this difference affects ion selectivity is not known. In Ciona HCNc genes, the selectivity filter motif is CIGYS. In mammalian HCN channels, substitution of the second glycine residue by serine (G404S in HCN2) reduced the slow activating current (39). Evidence of channel function is required to determine if this residue difference in the urochordate channel is involved in gene silencing. If Ciona HCNc is functional, a different mutational tolerance at this particular site seems likely and may reflect an adaptive process that has enabled these channels to fill a different functional niche specific to these species.

Recently, an N-linked glycosylation site (NX/T/S) located in the outer turret between the end of S5 and the pore helix (Fig. 5B ‘b’) has been shown to play a role in membrane expression of mammalian isoforms (47). Similar to residues in the S4–S5 linker, this motif emerges with the urochordate sequences and is conserved in two of the three urochordate genes and through...
out all vertebrate sequences. To understand the necessity of this motif and the role it plays in channel function throughout evolution, further studies are needed to determine if glycosylation or some other compensatory posttranslational modification that allows the channel to mature through the ER/Golgi process, occurs in these Ciona channels and other invertebrate sequences.

On the basis of the K⁺ channel crystal structure (14), the S6 segments of HCN likely form the inner vestibule of the pore and the gate. They show high sequence conservation with other protein relatives, such as ERG and CNG, and are almost completely conserved throughout the HCN family. Not surprisingly, the S6 segments of Ciona HCNa and HCNc and of spHCN2, which are predicted to be the result of lineage-

Fig. 5. Sequence comparison of the voltage sensing domain and pore region. A: sequence of voltage sensing domain conserved throughout evolution. The fly, human, and fugu sequences were chosen to represent arthropod, mammal, and fish species, respectively. Sequences are organized into invertebrates, urochordates, and vertebrate groups. A high sequence identity is notable among all vertebrate sequences. Shading indicates 4 levels of sequence conservation (see MATERIALS AND METHODS), excluding the assembly gap regions. Black stars indicate regions of assembly gaps in genome database, and the arrow indicates the conserved exon boundary between mammalian exon 2 and 3. a, Region of charge common to all potassium channel voltage sensing domains; b, region of charge unique to HCN channels; c, region associated with normal channel closure. B: sequence comparison of the HCN pore region. The same representative sequences as in 5A were used, except HCN4_chicken. This was omitted due to a gap in the genome assembly across this region. The selectivity filter sequence CIGYG (a) is conserved throughout the HCN family, with the exception of 2 duplicate genes from sea urchin and urochordate species. The NXT/S glycosylation site (b) is located immediately prior to the pore helix and is conserved in urochordate and vertebrate sequences. Two conserved glycine residues (c and d) are located in S6, one of which was hypothesized to be the hinge that obstructs that pore. More recent evidence suggests that the hinge is located further downstream between the threonine and glutamine residues (e) at the COOH-terminal end of S6 in which a dipeptide is conserved throughout the entire HCN family. f, The residue identified as responsible for the difference in chloride sensitivity in HCN2 vs. HCN1.
specific duplication, are the exceptions. Two glycine residues (Fig. 5B ‘c’ and ‘d’) are completely conserved among the HCN genes. Based on sequence homology with voltage-gated K⁺ channels (62), it has been suggested that one of these may form a glycine hinge involved in the opening of the channel gate in response to the movement of the S4–S5 linker. However, more recent experimental and homology modeling evidence (19, 55, 56) has shown that a threonine residue, positioned after the glycines in the linear sequence and on the intracellular side of the channel, is only accessible in the open state. Furthermore, a glutamine residue at the end of S6 is accessible in both the open and closed positions, suggesting that the putative hinge position is located between these two residues (Fig. 5B ‘e’).

**Evolution of the cyclic nucleotide binding and modulatory domains.** Cyclic nucleotides bind directly to the CNBD (12) located in the intracellular COOH terminus and modify channel opening. The CNBD is an evolutionarily conserved domain that is found in several cyclic nucleotide binding proteins, including the bacterial catabolic activating peptide and the protein kinase A family (4). The crystal structure of the COOH terminus of mouse HCN2 has been solved (78). Despite a low overall sequence conservation, the tertiary structure of its CNBD is highly conserved with the crystal structures of other CNBDs (4). Furthermore, residues identified as being critical to structure and nucleotide binding are conserved throughout the HCN family (Fig. 6). This includes the phosphate binding cassette (PBC), the most conserved feature of the CNBD that makes direct contact with cAMP (13), and the hinge region, important for the capping of cAMP by the C-helix of the CNBD. The middle of the PBC has diverged in one of the sea urchin genes and all of the urochordate genes, but these residues are also variable in other cyclic nucleotide-binding proteins, suggesting that they may not be critical for nucleotide binding. Because the HCN family possesses these conserved key residues, it seems probable that all of their CNBDs have a stable tertiary structure and bind cyclic nucleotides. One exception to this is seen in urochordate HCNc genes. These two sequences are missing a key hydrophobic residue in the PBC (Fig. 6 ‘↑’) that forms a conserved interaction with cAMP (53). In these genes, this residue is threonine, a hydrophilic amino acid that could disrupt this cAMP interaction. This difference is consistent with our hypothesis that the Ciona HCNc genes have undergone diversification following a lineage-specific duplication. Whether this difference also corresponds to a functional change in cAMP binding is an interesting question that will require functional studies to confirm.

The effects of cAMP binding on HCN channel opening are variable among isoforms and depend on the CNBD as well as the C-linker region that connects it to the transmembrane domain. For the mammalian HCN1, HCN2, and HCN4 isoforms, the binding of cAMP relieves a tonic inhibition of the transmembrane domain by the CNBD (75). The extent to which this relief occurs (HCN4 = HCN2 > HCN1) depends in part on the C-linker region that transmits the effect of cAMP onto the gating machinery (76). In contrast, cAMP binding has no effect on mouse HCN3 (45) and enhances the inhibition of channel opening of human HCN3 (67). Whether the sequence differences within the C-linker are responsible for these differences is not known. Further complicating the issue is that the effects of cAMP also depend upon regions in the transmembrane domains (76), which may lead to more complex interactions with the cytoplasmic portions of the gating elements. The complexity of these interactions is especially apparent in the spHCN1 channel in which channel inactivation/desensitization after first opening is eliminated by cAMP binding (64). C-linker sequences show considerable divergence throughout the HCN family, including differences among mammalian paralogs, and even between mammalian
and nonmammalian orthologs. This divergence is consistent with the variability in cAMP-mediated effects on channel opening.

Overall, the role of the C-linker in HCN channels is unusual compared with other regions in the channel. Its sequence and function show variation throughout the family, but it connects two domains that are themselves highly conserved in sequence throughout evolution and confer highly conserved, but distinct, functions in a cooperative manner (transmembrane = voltage-sensing, channel opening, and ion permeation; CNBD = cAMP binding and modulation of channel open by the transmembrane domain). Because of its position between these two domains, the sequence variability in the C-linker may be, in part, the result of an adaptive process that has enabled the diversification of cyclic nucleotide binding and the modulation of channel opening by the CNBD within the HCN family.

Sequence variability and functional divergence of vertebrate HCN paralogs. From our data, we predict that the four vertebrate isoforms are paralogs of each other resulting from three duplication events that occurred before the divergence of the teleost and tetrapod lineages. At the time of their origin, the gene pairs produced by these events would have been functionally redundant. Because the vast majority of duplicate genes are silenced throughout the course of evolution (37), the retention of all four is probably due to the acquisition of unique functional characteristics and/or expression patterns that result from tolerated mutations specific to each paralog (65). Throughout the core region, there is high sequence identity among the four vertebrate HCN isoforms. Some of these invariant residues probably contribute to functional properties common among all vertebrate channels. In contrast, some of the residues that vary among paralogs within this conserved region probably contribute to the more subtle isoform-specific differences in function (31). The search has begun to identify which residues underlie differences in function among the four mammalian isoforms using direct sequence comparisons followed by single site mutagenesis and/or construction of chimeras, and functional assays. Using these approaches, three studies have identified residues or regions responsible for phenotypic variation among the four mammalian HCN channels. First, differences in the rates of channel opening and cyclic nucleotide modulation between the HCN2 and HCN4 isoforms have been localized to a variant residue in the S1 segment (68). Second, a single residue difference in the inner selectivity filter was shown to confer chloride sensitivity to the HCN2 channel (74) (see Fig. 5B). Lastly, differences in the C-linker have been shown to contribute to differences in cAMP efficacy between HCN1 and HCN2, although the specific residues involved were not identified (76). From these few examples, it is clear that the functional consequences of residue variation among the four vertebrate isoforms, which may encompass not only overt differences in channel opening and closing but also differences in permeation, cyclic nucleotide affinity, homo- and heterotetrameric assembly, glycosylation status, protein-channel interactions, and abilities to traffic to the cell surface, cannot be predicted simply by the location within the channel. Differences in function may involve interactions among multiple residues and domains located throughout the channel. Therefore, functional assays in combination with site-directed mutagenesis and/or the construction of chimeras between vertebrate isoforms may not be sufficient to determine all differences.

In this study, we expanded the list of sequences for each of the four HCN paralogs by the addition of sequences from lower vertebrates. Broadening the evolutionary representation of the four vertebrate isoforms improves the sequence signal-to-noise ratio and enhances the identification of residues that are conserved among orthologs but differ among paralogs. However, genes from lower vertebrates (e.g., fish) and mammals have continued to evolve under different selective pressures, since their divergence from a common ancestor, several MYA. Therefore, information derived from this expanded list of sequences is complex. We identified three groups of residues based on similarities among vertebrate orthologs. First are residues conserved among orthologs and could contribute to a function specific to each vertebrate isoform. Second are residues conserved in the mammalian and fish orthologs but differ between these two groups. These residues may contribute to functional differences between the mammalian and fish channels of a particular isoform. Finally are residues conserved in orthologs of mammals or fish, but not both. These could be involved in species-specific or paralog-specific functions within each species. Alternatively, they may be the result of relaxed evolutionary constraints. By analyzing these different sets of conserved sites together with functional characterization of the various channels belonging to each of the vertebrate isoforms, we can more easily identify sites that may contribute to paralog-specific differences in channel function.

An informative subset of sites identified when comparing HCN genes from lower vertebrates and mammals are those that are conserved with a paralog rather than its own ortholog. If we assume that the probability of identical sites mutating to the same residues independently following duplication and species divergence is low, then these probably represent sites that have been retained from ancestral genes. Differential retention of sites from ancestral genes among orthologs suggests that channel phenotypes may not be completely conserved among them. In conjunction with the identification of conserved and nonconserved sites, the functional analysis of HCN channels from lower vertebrates and comparisons of their properties with those of mammalian channels will help to identify residues that contribute to different phenotypes and will also have the potential to shed light on the sequences and functions of ancestral HCN channels.

Vertebrate isoform-specific alignments of sequences spanning 450 MY reveal conserved motifs in the NH2 and COOH termini. The large number of vertebrate HCN sequences assembled in this study has greatly increased the power to identify isoform-specific motifs in the NH2 and COOH termini that may contribute to unique functions and specific patterns of expression within cells and tissues. The high sequence conservation among the four vertebrate isoforms extends beyond the core region analyzed above and includes regions just upstream of the S1 segment and downstream of the CNBD. The more distal NH2 and COOH termini, however, vary in both length and sequence among paralogs and do not align well. The sequences of the mammalian orthologs identified prior to this study were too close in evolutionary time to reveal divergence in the distal NH2 and COOH termini. On the other hand, the four paralogs from a single species were too diverged to align reliably. By expanding the taxa sampling to include species
from different time periods of vertebrate evolution, the signal-to-noise ratio that is inherent in the sequence information was considerably improved (2). Alignments of the distal termini, using this expanded list of vertebrate sequences, revealed several isoform-specific blocks of conserved sequence interspersed with diverged regions of variable length (Fig. 7).

In the distal NH2 and COOH termini, multiple regions specific to each of the four vertebrate paralogs were identified and were interspersed among regions of variable length and sequence. The conservation of these motifs throughout 450 MY of evolution highlights a selective constraint placed upon them and suggests that they confer isoform-specific properties. These properties remain to be identified, but there is some evidence to support a role for these motifs in the regulation of cell surface expression. For example, a single block of five residues is conserved in the NH2 terminus of HCN2 in fish, frog, chicken, and mammals (Fig. 7A ‘a’). This motif is not analogous to any known consensus sequence. The deletion of the first 130 amino acids of the mouse HCN2 NH2 terminus, which includes this conserved block of five amino acids, reduces cell surface expression and has only minor effects on channel opening and closing (70). A second example is a block of residues found in the COOH terminus of HCN1 channels (Fig. 7A ‘a’). This region is responsible for binding of filamin A to HCN1, and not the other isoforms (22); its removal abolishes the effects of filamin on channel cell surface expression. Overall, the motifs conserved within the termini of an individual isoform are more likely involved in roles specific to the function or expression of that paralog or may be involved in protein interactions specific to the tissues where these proteins are expressed. The regions of variable sequence and length observed between these conserved motifs result from DNA insertions and/or deletions. This variability may represent either relaxed constraints and/or the beginnings of species-specific adaptation processes. These adaptations may involve properties such as temperature sensitivity (41, 42) or binding to regulatory proteins with species-specific sequences.

Finally, both the NH2 and COOH termini possess regions of high sequence conservation among the four vertebrate isoforms, in addition to those of the core region used for the phylogenetic analyses. These regions probably confer properties that are not unique to any individual isoform but are important for all vertebrate HCN channels. In the NH2 terminus, a region of ~50 residues immediately upstream of the start of S1 is conserved in all four vertebrate isoforms (Fig. 7A ‘b’, HCN3 not shown). In mouse HCN2, this region interacts with itself and may be involved in intersubunit interactions of tetrameric assembly (70). Furthermore, the removal of this region, along with the rest of the NH2 terminus, prevents the formation of functional channels. Based on the high level of sequence conservation among the four isoforms, it seems probable that this region provides similar interactions for HCN1, 3, and 4. If this is true, the few differences among paralogs within this region may modify intersubunit interactions and thus regulate homomeric and/or heteromeric assembly.

In the COOH terminus, a block of residues conserved among the vertebrate HCN1, 2, and 4 isoforms was identified immediately downstream of the CNBD, which corresponds to the start of the last exon (Fig. 7B ‘b’). Deletion of this region, together with the entire distal COOH terminus and the C-helix of the CNBD, does not affect HCN1, HCN2, or HCN4 channel cell surface expression or function in heterologous systems (54, 75). A function for this conserved block of residues is not known, but whatever this function may be, based on the lack of sequence conservation, it is not likely possessed by HCN3 channels. Finally, a motif found at the distal end of the COOH terminus, SNL, is conserved in HCN1, 2, and 4 (Fig. 7B ‘c’). This motif, which qualifies as a consensus PDZ-binding domain, interacts with PDZ-containing proteins in vitro (28), and also with the TRP8 protein in vitro and in vivo, where it regulates channel cell surface expression (58). The absence of this motif in the HCN3 genes suggests that either this isoform does not interact with PDZ-containing proteins or that this interaction takes place elsewhere within the channel.

Summary and Perspectives

The availability of an increasing number of genome sequences has enabled us to generate a list of putative HCN coding sequences that has doubled the number of those previously known and covers a significantly greater portion of evolutionary time. With improved species representation, we were able to more accurately complete sequence comparisons, phylogenetic analyses and exon structure comparisons of the HCN gene family and put forward a model of its molecular evolution. We propose that the vertebrate isoforms evolved from a single ancestral sequence that had an exon structure similar to current mammalian HCN genes and that the duplication events occurred after the intron positions were fixed in the linear sequence. Duplications appear to have occurred independently in the sea urchin, urochordate, and fish lineages. Increasing the evolutionary distance between the sequences for each HCN isoform provided a good contrast and enabled us to unmask and identify regions putatively important for isoform-specific, as well as species-specific, functions. Together, our
study provides a strong basis from which to refine the proposed model of evolution as more genomes become available and as the functional analysis of HCN genes progresses. Finally, our study provides a valuable tool to aid in the planning of experiments that probe the relationship between structure and function of HCN channels and the functional significance of sequence similarities and differences among them.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Fiona Brinkman (Molecular Biology and Biochemistry, Simon Fraser University) for advice, training, and feedback on the manuscript as well as the Ouellette laboratory at the University of British Columbia Bioinformatics Centre (http://bioinformatics.ubc.ca) for resource advice that has made this work possible.

Current address of C. R. Marshall: The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Department of Molecular and Medical Genetics, University of Toronto, Ontario, Canada.

GRANTS

Supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES


Physiol Genomics • VOL 29 • www.physiolgenomics.org
Marcella CR, Pape HC.

Mannikko R, Elinder F, Larsson HP.


