Abstract. Northern blot analysis of rat heart mRNA probed with a cDNA coding for the principal polypeptide of rat liver gap junctions demonstrated a 3.0-kb band. This band was observed only after hybridization and washing using low stringency conditions; high stringency conditions abolished the hybridization. A rat heart cDNA library was screened with the same cDNA probe under the permissive hybridization conditions, and a single positive clone identified and purified. The clone contained a 220-bp insert, which showed 55% homology to the original cDNA probe near the 5' end. The 220-bp cDNA was used to rescreen a heart cDNA library under high stringency conditions, and three additional cDNAs that together spanned 2,768 bp were isolated. This composite cDNA contained a single 1,146-bp open reading frame coding for a predicted polypeptide of 382 amino acids with a molecular mass of 43,036 D. Northern analysis of various rat tissues using this heart cDNA as probe showed hybridization to 3.0-kb bands in RNA isolated from heart, ovary, uterus, kidney, and lens epithelium.

Comparisons of the predicted amino acid sequences for the two gap junction proteins isolated from heart and liver showed two regions of high homology (58 and 42%), and other regions of little or no homology. A model is presented which indicates that the conserved sequences correspond to transmembrane and extracellular regions of the junctional molecules, while the nonconserved sequences correspond to cytoplasmic regions. Since it has been shown previously that the original cDNA isolated from liver recognizes mRNAs in stomach, kidney, and brain, and it is shown here that the cDNA isolated from heart recognizes mRNAs in ovary, uterus, lens epithelium, and kidney, a nomenclature is proposed which avoids categorization by organ of origin. In this nomenclature, the homologous proteins in gap junctions would be called connexins, each distinguished by its predicted molecular mass in kilodaltons. The gap junction protein isolated from liver would then be called connexin32; from heart, connexin43.

Published December 1, 1987
NH₂-terminal sequence of connexin32 by Zervos et al. (45) recognizes immunoreactive peptides in homogenates of liver, heart, and uterus. Taken together, these results suggest that connexin32 and the immunoreactive polypeptide in the heart share some antigenic determinants, but also contain unique structure. Independent evidence for this conclusion was obtained by NH₂-terminal sequence analysis of proteins found in gap junction preparations. Nicholson et al. (30) and Manjunath et al. (27) have shown that a 44-47-kd polypeptide from myocardium is 43% homologous to connexin32 over the first 28 amino acids. Low stringency Northern blot analysis (32) provides additional support for the existence of a molecule in myocardium related to connexin32, and provides the basis for the screening strategy used in this paper to clone a unique connexin cDNA.

Materials and Methods

RNA Isolation and Northern Blots

RNA was isolated by homogenization of freshly dissected rat organs in guanidine isothiocyanate followed by centrifugation through CsCl (6). Uteri were obtained from 20-22-d pregnant rats and ovarian tissue from 21-d-old females primed for two successive days before sacrifice with 10 IU intraperitoneal injections of pregnant mare's serum gonadotropin (PMSG). Sigma Chemical Co., St. Louis, MO) in PBS. Previous studies have shown that gap junction structures are abundant in these two tissues (1, 15). For whole lens mRNA isolation, lenses were dissected from 72 rat eyes and directly homogenized in guanidine. While care was taken to dissect the lenses free from adherent ciliary epithelium, electron microscopy of similarly dissected lenses frequently revealed a tightly adherant layer of ciliary epithelium cells adjacent to the lens capsule, unobservable in the dissecting microscope (data not shown). To remove these adherent cells, a second preparation of mRNA was made from lenses dissected first with 0.1% trypsin. This procedure was used to rescreen this library and the Clontech library by hybridization and washing in the same solutions, but at 65°C.

DNA Sequence Analysis

Lambda clones were purified from plate lysates by DEAE cellulose chromatography (19). The cDNAs were released by Eco RI digestion and subcloned into the Eco RI site of the plasmid Bluescript (Stratagene). Ordered sets of overlapping deletions were constructed by the exonuclease III/S1 nuclease procedure of Henikoff (20) as modified by Lawler and Hynes (25). Single-stranded DNA was isolated by culturing Bluescript transformants in Escherichia coli strains JM101 or XLI-B (Stratagene) with the helper phage R408 (Stratagene) as described by the supplier. All sequencing was performed by the chain termination method of Sanger et al. (36) using this single-stranded template and dideoxy sequencing reagents as described by Williams et al. (42), except that 7-Deaza-dGTP (Boehringer Mannheim Biochemicals, Indianapolis, IN) was used in place of dGTP in equimolar amounts. All clones described were completely sequenced in both directions. Sequence data was analyzed using computer programs from Intelegenetics (Palo Alto, CA) and International Biotechnologies (New Haven, CT). Database searches were conducted by the staff of the Molecular Biology Computer Research Resource of the Dana-Farber Cancer Institute.

Results

Northern Blots and Isolation of cDNA Clones

Previous low stringency Northern blots suggested that the cDNA for connexin32 hybridized to a message of ~1.2 kb digested, dephosphorylated lambda gtlI arms (Promega Biotec, Madison, WI) and packaged using extracts prepared by Stratagene (San Diego, CA). This library contained ~101 recombinants with an average insert size of 1.1 kb. A second rat heart cDNA library was obtained commercially (Clontech, Palo Alto, CA).

The synthesized cDNA library was screened by hybridization of nitrocellulose filter plaque lifts in 5 x SSC (1 SSC is 150 mM NaCl, 15 mM Na citrate, pH 7.0), 1% SDS, 100 μg/ml salmon sperm DNA with the 32P-labeled rat liver gap junction cDNA overnight at 50°C, followed by three washes, 20 min each, in 2X SSC, 1% SDS at 50°C before exposure to Kodak XAR-5 film. The rat heart cDNA clone (D7) isolated by this low stringency procedure was used to rescreen this library and the Clontech library by hybridization and washing in the same solutions, but at 65°C.

RNA samples (10 μg) were subjected to electrophoresis on 1% agarose/formaldehyde gels and capillary blotted onto nylon membranes (Hybond-N; Amersham Corp., Arlington Heights, IL) as described by Davis et al. (8). Hybridization and washing were carried out in Hybrid-Ease chamber (Amersham) (28) over the first 28 amino acids. Low stringency Northern blot analysis (32) provides additional support for the existence of a molecule in myocardium related to connexin32, and provides the basis for the screening strategy used in this paper to clone a unique connexin cDNA.

1. Abbreviation used in this paper: PMSG, pregnant mare's serum gonadotropin.
in rat heart RNA (32). Further studies suggested that such hybridization was due to noncoding portions of the connexin32 cDNA (data not shown). Connexin32 cDNA was digested with Bgl I and Eco RI to isolate a fragment containing bases 1–848, which made up 95% of the coding region and none of the 3' untranslated sequence. This fragment hybridized to a single band of 3.0 kb on Northern blots of rat heart RNA at low stringency (Fig. 1 A).

We used the connexin32 cDNA fragment to screen 150,000 bacteriophage plaques from the rat heart cDNA library under similar low stringency conditions. One consistently positive clone was isolated. This eDNA clone (D7, Fig. 2) contained 220 bp, and its eDNA sequence showed 55% homology to a region near the 5' end of connexin32 cDNA. The amino acid sequence predicted by an open reading frame in this eDNA showed a similar high level of amino acid homology to connexin32. No further clones were isolated from this library. Clone D7 was used as hybridization probe to screen a second rat heart library (Clontech) under high stringency conditions, and three longer clones (G1, 2.5 kb; G2, 1.4 kb; and G3, 1.3 kb) were isolated. The cDNA inserts were oriented and aligned by restriction mapping (Fig. 2). All three of these cDNA inserts hybridized to a single 3.0-kb band on Northern blots of rat heart RNA (at high stringency). Hybridization with the probe G2 is shown in Fig. 1 B. The 3.0-kb band is indistinguishable in mobility from the band identified at low stringency with the connexin32 cDNA fragment.

*Nucleotide and Amino Acid Sequence*

The nucleotide sequences of all clones were determined; they overlapped with no discrepancies. The composite cDNA sequence (Fig. 3) contains 2,768 nucleotides. The first ATG initiation codon occurs at base 202 and is followed by an open reading frame of 1,146 bases and a TAA termination codon at position 1,348. The reading frame is closed by an open reading frame of 1,218 bases of 3' untranslated sequence, which contains multiple termination codons in all three frames, but no polyadenylic acid tail.

The cDNA sequence predicts a polypeptide containing 382 amino acids (Fig. 3) with a calculated molecular mass of 43,036, a size similar to the major polypeptide of 44,000–47,000 which Manjunath et al. (26) have characterized in cardiac gap junction preparations. Confirmation that this represents the sequence of a heart gap junction protein comes from the close match between the published NH2-terminal amino acid sequence of a rat heart gap junction protein determined by Nicholson et al. (30) and Manjunath et al. (27) and residues 2–20 of the predicted sequence shown here. The only difference is that those authors were unable to tell whether residue 2 was glycine, alanine, or histidine. The predicted residue is glycine. The predicted sequence also contains a single methionine residue preceding the first residue in the mature protein. Presumably, this methionine is proteolytically removed posttranslationally to leave a mature protein of 42,867 D. There is no evidence for a longer leader or signal sequence. Following the suggestion on nomenclature presented in the introduction, this protein will be called connexin43.

Connexin43 is predicted to be very basic (pI = 10.19), similar to the pI of 10.88 of connexin32. At neutral pH connexin43 would have 13.9% basic, 9.4% acidic, 34.3% polar, and 42.4% nonpolar residues. With 53 basic residues
The amino acid sequence was analyzed by the procedure of Kyte and Doolittle (24), which predicts the hydropathic nature of local regions in the sequence (Fig. 4 b). There are four major hydrophobic regions (labeled i, iii, v, and vii), alternating with four hydrophilic regions (labeled ii, iv, vi, and viii). Region iv corresponds approximately to residues 98-150 and region viii to residues 237-382. The hydropathy plot appears remarkably similar to that of connexin32 (Fig. 4 a).

**Comparison of Connexin43 and Connexin32**

The predicted amino acid sequence of connexin43 shows areas of homology to the predicted sequence (32) of connexin32 (Fig. 5). Of residues 1-105 in connexin43 (regions i-iii), 61 (58%) are identical to their counterparts in residues 1-104 in connexin32 (regions I-III). Of residues 142-246 in connexin32 (regions v-vii), 44 (42%) are identical to their counterparts in residues 122-227 of connexin32; amino acid matches occur only slightly more frequently than would be predicted by chance. The nucleotide sequences for these unique areas, as well as the 3' untranslated tails, show no homology.

No regions of internal homology within connexin43 were identified. A search of the National Biomedical Research Foundation protein sequence data base and of the predicted translations of identified exons in the GenBank/Los Alamos database identified no other proteins with significant homology to connexin43. In particular, the derived sequence of the lens membrane protein, MP26, which has been suggested to be a structural component of the lens fiber junctions (16), showed no homology to either connexin43 or connexin32.

**Northern Blot Organ Survey**

Total RNAs from various rat organs were examined for homologous messages by Northern blot analysis using G2, a 1.4-kb probe that contains the entire coding region of connexin43 cDNA. Under high stringency conditions of hybridization and washing, a single band of 3.0 kb was seen in RNA from heart, term uterus, PMSG-primed ovary, and kidney. The 3.0-kb band was seen in intact and trypsinized rat lenses; but not in decapsulated lenses, suggesting that it derived from lens epithelial cells but not lens fibers (Fig. 6). No bands were seen in RNA from brain, stomach, spleen, or liver. Further high stringency Northern blots were connexin32 cDNA confirmed the previous finding (32) that a homologous 1.6-kb mRNA is expressed in liver, brain, stomach, and kidney (data not shown). The 1.6-kb band was also seen in whole lenses; however, digestion of the lenses with trypsin abolished this signal, indicating that it derived from adherent ciliary epithelium. We did not find any of the 1.6-kb connexin32 message on Northern blots of RNA from ovary, uterus, spleen, or heart.

**Discussion**

This paper describes the molecular cloning of cDNA for
Connexin mRNAs Are Not Confined to Single Organs

The mRNAs for the two connexin molecules are found in different abundances in different organs. Connexin32 mRNA appears more abundant in liver, stomach, and brain, while connexin43 mRNA is more abundant in heart, term uterus, PMSG-primed ovary, and lens. There is a strong signal for both connexin messages in the kidney, although it is unknown whether they derive from the same or different cells. Experiments suggest that in the lens the connexin43 mRNA is localized to the epithelial cells. Crudely dissected and enzymatically digested lenses show a band by Northern analysis using the connexin43 cDNA (see Fig. 6). This signal is lost when the mRNA is prepared from decapsulated lenses, a procedure which is known to remove the bulk of the lens epithelial cells.

These comparative Northern analyses have limitations. First, while the Northern blots were performed under highly stringent conditions, they do not provide direct evidence that the protein coded by the foreign mRNA is being synthesized and assembled into gap junctions. Functional evidence may be provided by demonstrating that the protein can facilitate communication between cells.

To demonstrate that polypeptides predicted by cDNAs are gap junctional channel proteins, rigorous morphological and functional criteria need to be satisfied. Morphological evidence may be provided by EM immunocytochemical staining of the structure with antibodies either to bacterially expressed fusion protein (32) or to synthetic oligopeptides corresponding to the predicted sequence. Functional evidence may be provided by demonstrating that the protein can facilitate communication between cells.

Three approaches have been used to demonstrate that a protein is capable of forming a gap junction channel. In the first, intercellular communication via gap junctions, as assayed by dye and electrical coupling, is shown to be blocked by intracellular application of an antiserum specific for the putative channel-forming polypeptide (14, 22, 38, 39). One difficulty in this approach is the design of appropriate controls. Cross-linking of nonjunctional membrane proteins by an antibody may trigger a nonspecific intracellular response, resulting secondarily in closure of gap junctional channels. In addition, polyclonal antisera, even if affinity purified, may contain activity against unknown epitopes, or have a toxic activity apparently unrelated to binding of the specific protein.

In the second approach, channels are reconstituted in artificial lipid bilayers (18, 33, 44). A problem with this approach is that gap junctions are double-membrane structures while reconstituted lipid bilayers are single membranes, although an innovative double-membrane reconstitution system is currently being developed (5). Comparison of the reconstituted channel with an in vivo channel is difficult because the properties of gap junction channels in single membranes have not been determined in vivo. In addition, proteins must be very carefully purified so that reconstitution of unwanted channels does not occur. Reconstitution of proteins produced in vitro from cloned cDNAs could alleviate this difficulty.

In the third approach, mRNA coding for the putative junctional polypeptide may be introduced into cells and then the cells assayed for the ability to communicate (7). This type of experiment should include a demonstration that the protein coded by the foreign mRNA is being synthesized and assembled into gap junctions. The foreign channels must also be discriminated from endogenous channels, which requires that the properties of each type be measurably different.

Confirmation that a particular protein is capable of forming a gap junctional channel requires rigorous application of one or more of these methods, with close attention to the shortcomings peculiar to each. As detailed above, many of these structural and functional criteria have been satisfied for connexin32. A similar demonstration for connexin43 will require additional experimentation.

Confirming that a cDNA Codes for a Gap Junction Protein

To demonstrate that polypeptides predicted by cDNAs are gap junctional channel proteins, rigorous morphological and functional criteria need to be satisfied. Morphological evidence may be provided by EM immunocytochemical staining of the structure with antibodies either to bacterially expressed fusion protein (32) or to synthetic oligopeptides corresponding to the predicted sequence. Functional evidence may be provided by demonstrating that the protein can facilitate communication between cells.

Three approaches have been used to demonstrate that a protein is capable of forming a gap junction channel. In the first, intercellular communication via gap junctions, as assayed by dye and electrical coupling, is shown to be blocked by intracellular application of an antiserum specific for the putative channel-forming polypeptide (14, 22, 38, 39). One difficulty in this approach is the design of appropriate controls. Cross-linking of nonjunctional membrane proteins by an antibody may trigger a nonspecific intracellular response, resulting secondarily in closure of gap junctional channels. In addition, polyclonal antisera, even if affinity purified, may contain activity against unknown epitopes, or have a toxic activity apparently unrelated to binding of the specific protein.

In the second approach, channels are reconstituted in artificial lipid bilayers (18, 33, 44). A problem with this approach is that gap junctions are double-membrane structures while reconstituted lipid bilayers are single membranes, although an innovative double-membrane reconstitution system is currently being developed (5). Comparison of the reconstituted channel with an in vivo channel is difficult because the properties of gap junction channels in single membranes have not been determined in vivo. In addition, proteins must be very carefully purified so that reconstitution of unwanted channels does not occur. Reconstitution of proteins produced in vitro from cloned cDNAs could alleviate this difficulty.

In the third approach, mRNA coding for the putative junctional polypeptide may be introduced into cells and then the cells assayed for the ability to communicate (7). This type of experiment should include a demonstration that the protein coded by the foreign mRNA is being synthesized and assembled into gap junctions. The foreign channels must also be discriminated from endogenous channels, which requires that the properties of each type be measurably different.

Confirmation that a particular protein is capable of forming a gap junctional channel requires rigorous application of one or more of these methods, with close attention to the shortcomings peculiar to each. As detailed above, many of these structural and functional criteria have been satisfied for connexin32. A similar demonstration for connexin43 will require additional experimentation.

Figure 6. Hybridization of connexin43 cDNA to total RNAs from various tissues. Blots were probed at high stringency with cDNA clone G2, which contains the entire coding region. Three different blots are shown, each with a lane of heart RNA (lanes A, G, and L) for comparison. In the first blot, lanes A–F, a positive band comigrating with the 3.0-kb heart signal is seen in PMSG-stimulated ovary (B), and kidney (F). At the level of sensitivity of this experiment, no signals are evident in RNA fractions from liver (C), stomach (D), or brain (E). In this blot, 10 μg of RNA were loaded per lane. In lanes G and H, also loaded with 10 μg RNA, heart (G) and term uterus (H) both show co-migrating 3.0-kb bands. In the third blot (lanes I–L), RNA from heart (L) and whole lens (J) were loaded at 7.5 μg per lane, and show the 3.0-kb signal. Trypsin-treated (J) and decapsulated (K) lens RNA fractions were loaded at 3.0 μg per lane. The RNA from the trypsin-treated lens contains the 3.0-kb band (J), which is not detectable in the RNA prepared from the decapsulated lens (K).
stringent conditions, and show similarly sized mRNAs, this
does not guarantee absolute mRNA identity. It is possible
that identically appearing mRNAs might specify proteins
with extensive sequence homology that might contain minor
but key differences. Second, except for the case of the lens,
RNA was prepared from whole organs, containing many di-
verse tissues and cell types. Thus, there is no information
about which cell type expresses a given message. Third,
presence of message does not guarantee that protein is being
translated. Finally, the inability to detect a signal on a North-
ern blot does not mean that a homologous mRNA is absent,
only that its abundance is too low to detect. Our results show
that there is a predominance of different connexin mRNAs
in certain organs, but do not rule out the simultaneous ex-
pression of both connexin messages, as clearly seen in the
kidney.

**Connexin Membrane Topology**

Recent studies of proteolytically treated liver gap junctions
(46) have demonstrated that the COOH-terminal portion of
connexin32 (region VIII) and a Lys-X proteolytic site (region
IV) are located on the cytosolic sides of the junctional
membranes. These data place both regions IV and VIII at the
cytoplasmic surfaces, as drawn by Zimmer et al. (Fig. 14 in
reference 46) and here in Fig. 7. The cytoplasmic localization
of these regions and the assumption that the hydrophobic
portions represent membrane-spanning regions suggest a
topological model of the liver and heart gap junction proteins
with relation to the junctional membrane (Fig. 7). In this
drawing, connexin32 and connexin43 are depicted as parallel
lines; the dark dots between them indicate positions of iden-
tical amino acids. The putative membrane spanning and ex-
tracellular regions of the two connexins are conserved struc-
tures, and the cytoplasmic portions of the molecules are
divergent. Each of the short extracellular loops (regions II
and VI) contains three cysteine residues that are conserved
between the connexin sequences. Neither the single con-
sensus glycosylation site near the beginning of connexin32
nor any of the sites in connexin43 lie in the predicted ex-
tracellular regions. In this context, it is pertinent that no
glycosylation of connexin32 has been reported.

The conserved transmembrane and extracellular structures
in connexin molecules suggest that the structure of the
transmembrane portion of the channels and the mechanisms
of cell-cell interaction may be similar in different tissues.
Previous authors have demonstrated that heterologous cells
can form low resistance communication channels in culture
(13). An interesting question in this regard is whether or not
heterologous cells actually make heteromolecular junctions
in tissue culture, or whether they express different junctional
phenotypes in culture. Flagg-Newton and Loewenstein (13)
have demonstrated an asymmetric physiology of heterolo-
gous cell junctions in culture, consistent with the existence
of heteromolecular structures.

The unique cytoplasmic primary structures of the conn-
exins may confer different physiological behavior. These
unique regions are exposed to intracellular mechanisms of
phosphorylation (2, 4, 35, 40), calmodulin binding (46), and
proteolysis concomitant with protein turnover (10, 43). Phys-
iological studies on pH sensitivity have shown that liver and
heart channels have measurably different pKs of 6.4 and 6.8,
respectively (34, 37, 41), properties which presumably reside
at the junctional cytoplasmic surfaces. The heart junctions
are obligatory conductors of electrical excitation between the
myocardiocytes, a role not shared with the liver, but possibly
shared with uterine myometrium. Whether or not this func-
tional difference can be localized to specific protein struc-
tural domains has yet to be determined.

**Figure 7.** A model depicting topo-
logy of connexin proteins in
relation to the junctional mem-
bane, following the orienta-
tion presented by Zimmer et al.
(1987). Connexin32 (1–283) and
connexin43 (1–382) are drawn as
parallel lines to show their opti-
mal alignments, except in the
middle of the molecules and at
the COOH terminus where con-
 nexin43 has additional mass.
Heavy dots are drawn between
the lines at positions of identical
amino acids. This model was
constructed based on previous
evidence that the COOH termi-
nus and region IV of connexin32
are on the cytoplasmic side of
the membrane and the assumption
that hydrophobic regions span
the membrane. The model reveals
that the putative transmembrane
and extracellular regions of the
connexins are conserved and that
the cytoplasmic portions are di-
vergent.
The Connexin Family of Proteins

We propose a new system of nomenclature, using the term connexin to identify the members of a family of proteins that are related by a high degree of conservation in their predicted amino acid sequences, prototypically shown here between connexin32 and -43. We will not attempt to further define the degree of conservation required, since only two members of the family have been characterized. We think that the connexin family of proteins will be shown to be gap junction proteins. Additional support that connexin43 is a gap junction protein comes from a strong homology with the NH2-terminal of a previously identified heart gap junction protein. However, by the criteria suggested above, connexin43 has not yet been shown to be a gap junction protein.

The family of connexins may contain additional proteins beyond the two reported here. Initial data have been presented (29) demonstrating the presence in rodent livers of an M, 21,000 protein, present in junctional plaques, which is 45% homologous to connexin32 in the NH2-terminal 20 amino acids. In addition, using the same strategy followed in this paper, we have isolated an additional homologous cDNA from rat lens fibers (Paul, D. L., E. C. Beyer, and D. A. Goodenough, manuscript in preparation). All gap junction proteins will not necessarily be members of the connexin family. Thus, the lens fiber protein, MP26, for which the entire sequence has been determined (16), is clearly not a member of the connexin family on the basis of its very different amino acid sequence.

We have proposed using the predicted polypeptide molecular mass in kilodaltons to distinguish between different members of the connexin family (e.g. connexin32, connexin43). In the event that different connexins are identified in the same organism with nearly the same mass, it may be necessary to distinguish them by using the first decimal place. As we have shown, it is likely that the same connexin mRNA is expressed in many different organs; therefore, we believe that any reference to organ (as liver, heart, or uterus) is inappropriate. Connexins have been identified from other species (23); they must be distinguished by specification of species and tissues.

Although there are potentially many biological roles for intercellular communication, currently only a few functions are understood. For this reason, we have no explanation for connexin diversity.

We gratefully acknowledge helpful discussions with Dr. Katherine Swenson and Dr. Bruce Nicholson. Salary support for E. C. Beyer has been generously provided by training grant HL05754 from the National Institutes of Health. E. C. Beyer is currently a recipient of a Clinician-Scientist Award from the American Heart Association. These studies are supported by grants GM37751 (D. L. Paul) and GM18974 (D. A. Goodenough) from the National Institutes of Health. E. C. Beyer is a recipient of a Clinician-Scientist Award from the American Heart Association. These studies are supported by grants GM37751 (D. L. Paul) and GM18974 (D. A. Goodenough) from the National Institutes of Health.

References