Studies on the Structure and Molecular Diversity of the Gap Junction

Thesis by

Jan Håkan Hoh

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1991

(Submitted May 23, 1991)

Dedication

For Friends and Family

ACKNOWLEDGEMENTS

There are of course a great number of people who have influenced this thesis, and it is impossible to thank them all. I must however, mention some. My parents, Fang and Birgitta Hoh for their love and the opportunity of life. Jean-Paul Revel, my mentor, let me pave my own way at Caltech allowing me to sample the breadth of science, and to become a scientist. The many others who I would like to thank include Scott (John), a good friend; Rolf, who discovered an American within himself and would even drink the beer; Jim, who helped me find my way around inside of a rat, and the lab; Ross, whose straight face was never to be believed; Bruce, for laying the ground work; Moshe, for making the world a more interesting place; Barbara (Yancey), with whom I often disagreed but learned more from her than she will know; Eric and Erling, for those early years; John, for too many margaritas and a few uphill jaunts; the NIH, for three square meals; Lal, whose penchant for arguments always kept the coffee room a lively place; Morton, who footed the bill for a new path; Doug, for always reminding me of the world outside; Jim and Ray, for taking my money in the back of the van before we got to Vegas; Danielle, though she probably still owes me a lunch at the Fairway House; Howard, for serving on my committee; Barbara (Wold), a friend of all science; Scott (Emr), who's energy is contagious; Kasturi, who learned to use a large spoon with great skill; Fiona, for more than a few great meals; Dale, even though he is Canadian; Mark, who always had good advice; Mathew, for introducing me to science and getting me into Caltech; Mike, for revealing the mind of a geneticist; Micki, for always being well stocked late at night; Susan, for believing in me; Janice, even though she would never listen to me; Stuart and George, for being great cyclists and roommates; Chris, for teaching me to pour a sequencing gel; Roger, for interesting discussions during Letterman; Marit, for so many great times together; and the many others that just don't come to mind at 3 a.m on a Sunday.

ABSTRACT

An improved method for the isolation of hepatic gap junctions that substantially shortens preparation time and improves the yield of previous methods is described. The topology of the 28 kD protein component (connexin-32, Cx32) of gap junctions isolated with this method is examined using proteases and antibodies against specific peptides. These experiments are consistent with the current model for the organization of the protein in the membrane, but reveal that an unexpectedly large part of the carboxy-terminus is protected from proteolytic attack. Together with data from comparisons of the Cx32 protein sequence with other channel proteins, a modified topological model is proposed.

The structure of the gap junction is further studied by atomic force microscopy. Using this new technology, high resolution images of a gap junction in phosphate buffered saline are obtained, and after "force dissection," which removes half the plaque, the extracellular domains of individual connexons in a hexagonal array with lattice constant of 9.1 nm are revealed. These are the first images of an ion channel by atomic force microscopy, and the observations open the door for a variety of new experiments not previously possible.

Low stringency screening of a rat genomic library produced genomic clones for Cx32 and a new member of the gene family, connexin-31 (Cx31) or β3. Cx31 has a unique distribution and is found in the eye, Harderian gland, skin, and placenta. Comparison of the Cx31 with the other known connexins, reveals unique and conserved domains in the protein sequences. This comparison is extended to a phylogenetic analysis of the entire gene family that shows two major branches of connexins that diverged 1.3-1.9 billion years ago. Comparison with other ion channels reveals a short sequence similarity between the connexins and channels such as the voltage activated K^+ channel. In K^+ channels the sequence has been shown to line the aqueous pore, and the model for connexin organization is modified to account for this possibility. The similarity also suggests that gap junctions are part of a superfamily of ion channels.

TABLE OF CONTENTS

Introduction

Gap junctions are close appositions of the plasma membranes from two cells that contain an array of densely packed cell to cell channels (Revel and Karnovsky, 1967). The structure and organization of this plasma membrane specialization are highly conserved in all organisms and tissues that have been examined, and phylogenetically the gap junction is widely distributed and has been found in every multicellular animal that has been examined with few exceptions (for review see Peracchia, 1980). The wide distribution and high degree of structural conservation suggest that the gap junction carries out a fundamental biological function. Many functions have been suggested. However, to date most evidence remains circumstantial and highly correlative in nature. The functions that have been proposed include transmission of synaptic signals (Furshpan and Potter, 1959), excitation and conduction of myocardial cell contractions (Karrer and Cox, 1960), coordination of smooth muscle contraction (Karrer and Cox, 1960), regulation of growth control (Loewenstein and Kanno, 1967), transmission of developmental signals (Potter et al., 1966), and maintenance of metabolic homeostasis (Subak-Sharpe et al., 1969).

Biological Function of Gap Junctions

Furshpan and Potter (1959) are usually credited with the first observation of direct electrical coupling between cells. Their work on the crayfish giant axon lead to the definition of the electrotonic synapse or electrotonic junction. Electrotonic junctions have been described in many different systems. Specific characteristics of these junctions have been reviewed in some detail (Bennett and Goodenough, 1978). The advantages of direct electrical transmission are not well understood, but in some cases may be that it is faster and more synchronous than transmission through chemical synapses. These characteristics have lead to the suggestion that one function of electrotonic junctions is rapid synchronization of events such as escape reflexes.

The role of gap junctions in the conduction and excitation of myocardial cells was first suggested by Karrer and Cox (1960) who described a "quintuple-layered cell

interconnection." Subsequently Barr et al. (1965) demonstrated a physiological basis for the involvement of direct cell to cell connections in electrical coupling of myocardial cells. Using isolated cardiac fibers they demonstrated that an electrical signal could be transmitted along the length of the fiber even when the extracellular fluid had been replaced by isotonic sucrose. This showed that the signal must be propagated between cells by direct cell to cell connections, and not by extracellular signaling molecules such as neurotransmitters. Subsequently, several investigators have demonstrated physiologically and biochemically that cardiac cells are coupled via gap junctions (for recent example see Yancey et al., 1989). Theoretical cable analysis further suggests that the passive properties such as junctional coupling is an important factor in the proper contraction of myocardial cells (Arnsdorf, 1990). However, it should be noted that the fact that the cells are coupled and that gap junctions can provide an electrical pathway for synchronization of the myocardial contraction does not mean that they do. There is no doubt that cardiac cells are coupled. However, the major lines of evidence for the involvement of gap junctions in myocardial contraction summarized here are not particularly strong, and provide no direct evidence for the proposed function. Another model has been proposed in which extracellular currents adequately account for the synchronous contraction and gap junctions are not involved (Sperelakis and Mann, 1977).

Coordination of muscle contraction by gap junctional coupling has also been proposed for other organs (Karrer and Cox, 1960; Garfield et al., 1977). The most dramatic example of this proposed function is uterine contraction during parturition. It has been demonstrated morphologically and molecularly that the number of gap junctions between uterine smooth muscle cells increases dramatically just before birth, and quickly return to normal levels afterward (Garfield et al., 1977; Risek et al., 1990). This rise in gap junctions correlates closely with increased contractions, and they are thought to pass the electrical signals that coordinate the contractions.

Gap junctions were first described in non-excitable cells by Loewenstein and Kanno (1964). This observation was followed by the description of communication defects in many cells of neoplastic origin, which lead to the proposal that gap junctions might be involved in the regulation of growth control (Loewenstein and Kanno, 1967). Several models have been put forth for the involvement of gap junctions in cell and organ growth. For example, it has been suggested that diffusional gradients generated through gap junctions could regulate organ size (Wolpert, 1978). Further correlative evidence for the involvement of gap junctions in the regulation of growth comes from the regenerating liver, where it has been shown that mitotic cells do not have gap junctions with their neighbors, while non-dividing cells do (Dermietzel et al., 1987). More direct evidence is provided by the experiments of Mehta et al. (1986). They used methylcholanthrene to transform a fibroblast cell line, which results in clones with three types of cells: cells that communicate with themselves and with the parent cells, cells that communicate with themselves but not with the parent cells, and cells that do not communicate with themselves or with the parent cells. The growth rate of these clones of cells on a layer of parent cells correlates inversely with their ability to communicate with the parents. Clones of cells that do communicate with the parent cells grow significantly slower than ones that do not. Treatments of the cells that up or down regulate communication, result in a corresponding up or down regulation of growth. Further evidence for gap junctional involvement in growth control originates in the work of Atkinson et al. (1981, 1986). They found that a temperature sensitive mutant of Rous sarcoma virus caused a rapid uncoupling of cells when shifted to the permissive temperature, and proposed that this may be due to phosphorylation of the gap junction protein by $pp60^{src}$. It has now been shown that a gap junction protein is phosphorylated on a tyrosine in $pp60^{src}$ dependent fashion, and that if the tyrosine involved is removed by site-directed mutagenesis the virus fails to affect coupling (Swenson et al., 1990).

The involvement of gap junctions in the transmission of developmental signals was first suggested by Potter et al. (1966). Since then gap junctions have been reported to have

a role in several developmental processes such as induction and the formation of developmental compartments (for reviews see Revel et al., 1985; Caveney, 1985). More recently a series of papers have used a polyclonal antisera against the major protein component from rat liver gap junctions (connexin-32), in attempts to directly interfere with intercellular communication *in vivo* and thereby cause developmental defects. In the first paper, antisera was injected into one cell of an eight cell *Xenopus* embryo, which resulted in defects in eye development on one side of the tadpole (Warner et al., 1984). The same antisera has also been used to prevent compaction of mouse embryos (Lee et al., 1987), cause respecification during chicken limb bud development (Allen et al., 1990), and block the head inhibitor gradient in Hydra (Fraser et al., 1987). Although there are more than 100 different antibodies against different gap junction isoforms, these papers all use the same polyclonal serum for which no full characterization has been published. Until other antibodies are used, or convincing evidence of the specificity of the antisera in question is presented, these experiments remain highly suspect and this author does not take them to provide evidence for the involvement of gap junctions in development.

Metabolic cooperation via gap junctions, results in the sharing of small molecules and metabolites between cells (see review Hooper and Subak-Sharpe, 1981). Such sharing has been well demonstrated in cell culture where hypoxanthine guanine phosphoribosyl transferase (HGPRT) deficient cells can be rescued from the effects of azaserine, which blocks *de novo* purine synthesis, by co-culture with wild type cells (Fujimoto et al., 1971). There are several instances in which metabolic cooperation may be biologically significant. The lens of the eye for example, is a large non-vascularized organ with gap junctions between its fiber cells. It has been proposed that these gap junctions actually provide a pathway for metabolites to move in and out of the interior of the lens (Goodenough et al., 1980). Another possible example is in females heterozygous for the HGPRT deficiency known as Lesch-Nyhan disease. These individuals are largely asymptomatic, which may be due to a rescue of the abnormal process leading to the disease by clones of normal and

abnormal cells (resulting from random X-inactivation) sharing substrates and products of HGPRT. Metabolic cooperation through gap junctions might also play a role in the prevention or masking of cancer (Subak-Sharpe et al., 1969). Cell growth and division involves changes in concentrations of second messengers. Any abnormalities in single cells that lead to such changes could be compensated for by a "buffering" effect created by gap junctions.

In addition to those described above, a variety of other functions have been proposed for the gap junction. However, to date the evidence for any particular biological function is inconclusive. It is a curious contradiction that a structure so widely distributed and highly conserved in its general features, is implicated in such a diverse set of biological functions.

Function versus activity

An important distinction that is often lost in the gap junction literature, is that between biological function and activity. Biological function is the way in which a molecule or structure contributes to the life of an organism. Activity is the chemical or physical process that the same molecule or structure carries out to serve its biological function. Therefore, while the activity and biological function are inextricably linked, they are different. For example, the dye Lucifer yellow can pass through gap junctions and is often used to assay for the presence of gap junctions. But this is an assay for gap junction activity, since it never has been, or never will be the biological function of a gap junction to transfer Lucifer yellow between cells.

Gap junction activity

By the distinction between function and activity described above, the study of gap junction physiology is largely the study of gap junction activity. This activity is usually assayed by electrical coupling, dye coupling, or metabolic coupling.

Metabolic coupling is the ability of cells to transfer metabolites or other small molecules such as second messengers. The early metabolic coupling experiments,

pioneered by Subak-Sharpe, Bürk and Pitts (1966, 1969), involved co-culturing $HGPRT^-$ cells with $HGPRT^+$ cells in the presence of H^3-hypoxanthine. Tritiated precursors of guanine nucleotides are transfered from the $HGPRT^+$ cells to the $HGPRT^-$ cells and incorporated into the DNA, thereby fixing the label inside the cells. Autoradiography is then used to determine if transfer occurred and therefore if gap junctions between the cells were active. In some cases metabolic coupling can also be measured biologically. For example Lawrence et al. (1978) co-cultured ovarian granulosa cells with cardiac myocytes and demonstrated that when the granulosa cells were stimulated with follicle stimulating hormone (FSH), which is known to raise cyclic adenosine monophosphate (cAMP) levels, the rate at which the myocytes beat, which is also cAMP dependent, increased in a dose dependent fashion. Since FSH does not have a direct effect on cardiac myocytes, these observations are interpreted as indicating transfer of cAMP from the granulosa cells to the myocytes through gap junctions. This biological assay for gap junction activity is a good example of the distinction between biological function and activity, since biologically the communication between ovarian granulosa cells and cardiac myocytes is likely to be meaningless.

While metabolic coupling assay is non-invasive, it is fairly cumbersome and time consuming. A similar but more common approach to determining gap junction activity is dye transfer, first described by Potter et al. (1966). This involves loading a cell with a fluorescent dye such as Lucifer yellow, usually by microinjection, and following the dye movement with an appropriately equipped microscope. Dye transfer has the advantage that it is rapid and can be applied to large collections of cells or whole organs. However, it is by nature not very quantitative and is usually used to qualitatively establish gap junctional coupling.

Electrophysiology is the method of choice for obtaining quantitative, time resolved information about gap junction activity. This is performed by impaling two adjacent cells

with microelectrodes and measuring the current flow through the cell-cell channels under different experimental conditions (for review see Bennett and Goodenough, 1978).

The most fundamental aspect of any measure of gap junction activity is the question: is the gap junction responsible for cell-cell coupling? For many years there was only correlative evidence that gap junctions actually were composed of cell-cell channels. That is wherever coupling was found, gap junctions were seen, often in the absence of any other junctional structure, and treatments that disrupted gap junction structure uncoupled cells. The isolation of the major protein components from gap junctions, and the subsequent isolation of cDNAs for these proteins has now allowed a more direct test of the role of gap junctions in the formation of cell-cell channels by two different approaches. The first involves introducing mRNAs for gap junctions into communication deficient cells, and measuring the resulting cell-cell coupling. Such experiments have been carried out in a number of systems like *Xenopus* oocytes, two cell mouse embryos, and cultured mammalian cells (Dahl et al., 1987; Swenson et al., 1989; Lash et al., 1990; Eghbali et al., 1990). The second approach uses antibodies against the protein components of the gap junctions to interfere with cell-cell communication (Hertzberg et al., 1985; Yancey et al., 1989). Together the results from these experiments clearly show that the major gap junction proteins are necessary to form cell-cell channels, and support the now dogmatic view that gap junctions are a (the) site of direct cell-cell communication.

While the major gap junction proteins are necessary for coupling, the question of sufficiency is still controversial. In principle the major gap junction proteins cannot by themselves form a cell-cell channel, since there must be a membrane and therefore lipid components. But given a lipid membrane, can the protein components that have been identified by themselves form channels, or are there other requirements? This question is most obviously addressed by reconstitution of the channel in artificial membranes, and several papers have now been published that claim to have accomplished this (Young et al., 1987; Spray et al., 1986). These groups use two approaches to "reconstitute" gap junction

activity. The first involves using whole gap junction plaques and incorporating these in lipid membranes, with or without any one of several detergents (Triton, CHAPS, octylglucoside). These membranes have channels with unitary conductances of 140 and 280 pS, close to that reported for hepatic gap junctions. However, incorporating channels already in a membrane into another membrane is not truly reconstitution, and at no point is it demonstrated that the detergents actually dissolve any part of the membrane. Even if it is truly gap junctional channels that are responsible for the conductances, this experiment does nothing to address the issue of sufficiency. In another experiment, the 28 kD gap junction protein is excised from SDS-gels and incorporated in lipid vesicles. These vesicles again exhibit conductances of 140 and 280 pS. The main evidence that this conductance is from a gap junction channel is that antibodies against the 28 kD protein prevent it. However, the antibodies used are the same omnipotent antibodies used to perturb development as discussed above (see Warner et al., 1984). Again, no characterization of these antibodies has been published. In addition to the questionable antibody, the issue of how the channel is formed in a single membrane is left unresolved. There is no evidence that a single connexon can form an open channel, in fact this is rather unlikely. This author believes that the observed channels are not gap junctional, but some other channel such as the 30 kD voltage dependent anion selective channel of the mitochondrial outer membrane.

Assuming that the gap junction is composed of cell-cell channels, many of its physiological characteristics have now been studied, including exclusion limits, channel conductance, pharmacology and regulation of gating. The commonly cited number for the exclusion limit of gap junctions is <1000 Da for vertebrates and <2000 Da for invertebrates, with little or no selectivity based on charge, measured by injecting cells with labeled molecules of different sizes (Flagg-Newton et al., 1979). However, these exclusion limits were determined before it was known that gap junctions in different cell types are composed of different proteins. Since this discovery, there have been no studies showing that gap junctions composed of different proteins all have the same exclusion

characteristics. So, while it is one of the most basic characteristics of gap junctions, the chemical and physical characteristics of molecules that can be passed by gap junctions are still poorly defined.

Electrophysiologically, the macroscopic currents of the gap junctions have been well studied (Bennett and Goodenough, 1978). More recently the single channel properties have also been measured (Burt and Spray, 1988, Eghbali et al., 1990). Measurement of single channel properties of gap junction channels is complicated by the presence of several gap junction proteins in many cell types. Because of this it has not been possible to determine conductance properties of channels made up of specific connexins. These reveal channels with different conductances in different tissues, c. 150 pS for liver gap junction and c. 50 pS for cardiac myocytes. Both these types of channels are dependent on transjunctional voltage (Spray and Bennett, 1985; R. Lal, personal communication).

A major difficulty in the study of gap junction biology and physiology has been the lack of any specific drugs that affect the channel. However, there are several general agents that do affect gap junction activity. The most prominent are calcium, pH, octanol, retinoic acid and cAMP. Calcium is known to close hepatic gap junctions when present in the mM concentration range (Rose and Loewenstein, 1975). This effect is seen well above intracellular levels, which are in the μM range, but is close to extracellular calcium concentrations. It was suggested that this calcium induced closure is involved in a cellular response to injury, in which a large influx of calcium, due to damage of the cell membrane, causes the gap junctions to close and seal the injured cell from healthy ones. It is also possible that there is a second level, in the μm range, of calcium mediated gating that is connected to phosphorylation of the protein. This is discussed further below. The effect of pH on junctional communication was first described by Turin and Warner (1977). The pKa for opening and closing the channel varies depending on the cell type, but is usually close to physiological. For example pH dependent gating of liver cell junctions has a pKa of 6.3 (Spray et al., 1984), while purkinje fibers have a pKa of 6.8 (Reber and Weingart, 1982).

Octanol has been shown to close gap junctions in the 0.1 mM concentration range (Spray and Bennett, 1985). However, octanol has also been shown to close other channels such as the ryanodine receptor and its specificity for gap junctions is not obvious. Retinoic acid is another molecule that can affect coupling (Pitts, 1986). This, as the other reagents, is a very non-specific molecule with a variety of biological effects, and the mechanism by which it upregulates gap junctional communication is not known.

The second messenger cAMP is also a molecule that has a variety of biological effects, and has been shown to affect gap junctional activity in a number of systems (Saez et al., 1986; Murray and Gainer, 1989; Voorter and Kistler, 1989). One way in which cAMP may affect gap junctions is through phosphorylation by a cAMP dependent protein kinase. In a recent report Arellano et al. (1990) shows a dramatic effect of cAMP dependent protein kinase in squid axons. They first perfuse the axon and replace the entire cytoplasm with buffer. Subsequently they introduce the catalytic subunit of cAMP dependent protein kinase into the axon and assay the junctional coupling. The kinase does not have a direct effect on coupling, but it does potentiate the effect of calcium, which after kinase treatment gates the channel in the physiological μM range. This is a remarkable finding that mechanistically brings together two very disparate potential regulators of gap junctional activity.

Gap Junction Structure

The gap junction was first seen by electron microscopy as a close apposition between two plasma membranes of smooth muscle cells (Karrer and Cox, 1960; Dewey and Barr, 1962). Subsequently, Robertson discovered a quasicrystalline array of particles in close membrane appositions in the Goldfish Mauthner cell (Robertson, 1963). The gap between the membranes was delineated by Revel and Karnovsky (1967) who used colloidal lanthanum as a marker for the extracellular space. This allowed gap junctions to be distinguished from tight junctions, and demonstrated that the particles in the array spanned the gap, suggesting that they could provide the structural basis for the cell-cell coupling that

had been observed physiologically (Furshpan and Potter, 1959). These and other early electron microscopic observations on the gap junction defined its general structural features of two plasma membranes separated by a 2-3 nm gap, with a hexagonal array of subunits spaced approximately 9 nm center to center.

To further pursue the structure of the gap junction, several investigators have developed isolation methods for this plasma membrane specialization (Henderson et al., 1979; Finbow et al., 1980; Nicholson and Revel, 1983; Hertzberg, 1984). All of these procedures stem from the work of Benedetti and Emmelot (1968) who first identified cell fractions enriched in gap junctions by negative staining. In principle, plasma membrane fractions of varying purity are first isolated based on sedimentation characteristics or density. These fractions are then treated with detergent or alkali to remove most non-junctional membranes, and gap junction enriched fractions are separated from the remaining membrane by, for example, centrifugation on sucrose gradients.

Isolated gap junctions have now been studied extensively by electron microscopy, X-ray diffraction and electron diffraction (Caspar et al., 1977; Unwin and Zampighi, 1980). Based on these data a general model of the gap junction, first proposed by Makowski and his coworkers (Makowski et al., 1977), has gained wide acceptance (Figure 1). This model has been refined and extended by a series of papers (Makowski et al., 1982; Baker et al., 1983; Makowski et al., 1984; Makowski et al., 1984; Baker et al., 1985; Sosinsky et al., 1988). These papers provide a great deal of additional detail on the gap junction structure and the effect of isolation protocol, staining, and radiation. Overall, the orginal model is still widely used as a general reference point.

A major disagreement with regard to the gap junction structure is the mechanism of gating. The Makowski model has only one major conformational state and based on sucrose exclusion data the gate for the pore is near the cytoplasmic ends of the channel. This is in contrast to a model proposed by Unwin and Ennis (1984) that has two major conformations related by a twisting of the channel, resulting in a 7.5° tilting of the protein

subunits. Further structural information at resolutions better than the 1.8-2.5 nm that have been achieved, and better physiological and biochemical data, will be necessary to understand the structural basis for gating of the gap junction channel.

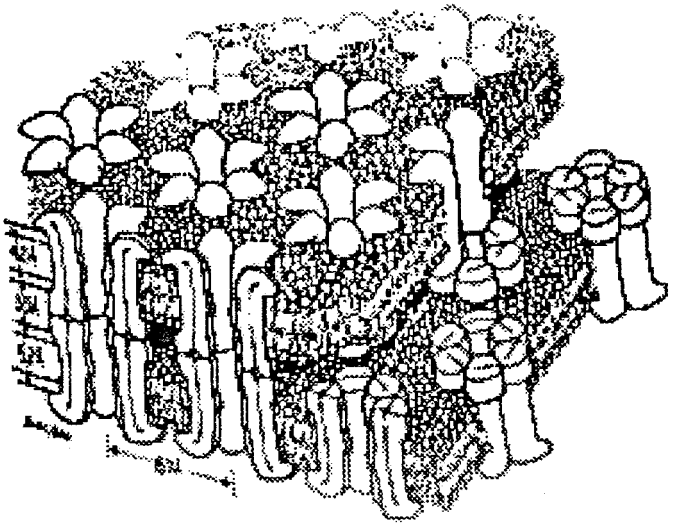


Figure 1. Model of gap junction structure taken from Makowski et al. (1977). Shows the two membrane separated by 2-3 nm with a hexagonal array of channel spaced at 9.1 nm. Each channel is composed of two hemi-channel called connexons, that are roughly cylindrical structures 7 nm in diameter. Each connexon exhibits six fold symmetry, and is thought to be composed of six or possibly a multiple of six protein subunits (connexins).

Isolated gap junctions have also been used to determine the biochemical composition of this structure. The lipid composition of gap junction membranes has been reviewed recently (Malewicz et al., 1990). Direct biochemical analysis of lipids has been restricted to isolated gap junctions, that have been detergent treated. These analyses show that isolated gap junctions have several different lipids, including cholesterol, phosphotidylcholine, phosphotidylethanolamine, phosphotidylserine, and phosphtidylinositol. The presence of cholesterol in isolated gap junctions is contradicted by labeling with the antibiotic filipin. This cholesterol probe does not label gap junctions, though this may be due to the physical nature of the gap junction membrane. Since

detergent treatment is likely to distort the composition, reported contents and ratios of lipids found in isolated gap junctions are not likely to represent the *in vivo* composition.

While proteins of many different molecular masses were initially described as components of gap junctions, there is now a consensus for at least three major protein components, the 28 kD and 21 kD proteins from liver, and the 45 kD protein from heart. These proteins were shown by amino-terminal sequence analysis to be homologous and probably members of a gene family (Nicholson et al., 1985; Nicholson et al., 1987). The isolation of a cDNA for the 28 kD protein (Paul, 1986) opened the door for identifying several new members of the gap junction gene family (see Chapter 3). Based on the molecular weight predicted from the cDNA (32,000), the 28 kD protein is now refered to connexin-32 (Cx32) and other connexins have been renamed accordingly (see Table 1, Chapter 3 for a current list).

There are also two other proteins that have been suggested as gap junction proteins, the 16 kD protein described by Finbow et al. (1984) and MP26 from lens (Broekuyse et al., 1976; Gorin et al., 1985). The 16 kD protein has now been shown to be a part of a H^+ ATPase and is generally no longer regarded as a candidate for the gap junction protein. The role of MP26 in the formation of gap junctions remains unclear. However, homologues to MP26 such as the *big brain* protein from *Drosphila* (Rao et al., 1990), soybean nodulin 26, and *E. coli* glycerol facilitator (Baker and Saier, 1990), and an erythrocyte Mr 28,000 multimeric transmembrane protein (Smith and Agre, 1991) have been isolated from several different systems that suggest a role for the molecule in transport. MP26 may form a channel in a single cell membrane, or form a cell-cell channel distinct from the gap junction. However, based on sequence it is clearly not a member of the connexin family of proteins, and will therefore not be discussed further.

Prior to the cloning of Cx32, the topology of the protein had been studied by proteases (Nicholson, 1983). These experiments demonstrated that there were two large domains of 10 kD that appeared to be protected by the membrane. Interpreted in the context

of the protein sequence, the membrane protection experiments suggested a model (Figure 2) with 4 transmembrane segments (TM1-TM4), cytoplasmic amino- and carboxy-termini, two extracellular loops, one separating TM1 from TM2 and the other TM3 from TM4, and a cytoplasmic loop separating TM3 from TM4 (Paul, 1986). This model has now been extensively tested (see Chapter 1).

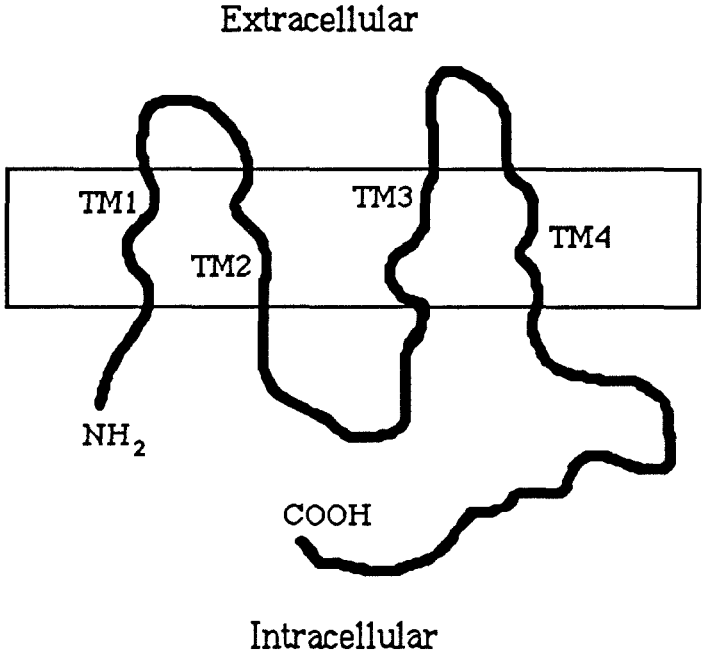


Figure 2. Schematic of Cx32 organization in the membrane (Milks et al., 1988; Beyer et al., 1990).

Summary of this Thesis

When the research for this thesis began, hepatic gap junctions could be isolated and the major hepatic gap junction protein (Cx32) had been determined to be 28,000 Da by gel electrophoresis, and the amino terminal protein sequence had been determined. While peptide maps suggested that gap junction proteins from other organs might be different (Nicholson, 1983), many people in the field thought that there was only a single protein that comprised all gap junctions (Hertzberg and Skibbens, 1984; Zervos et al., 1985; Finbow et al., 1984).

The work on a new isolation procedure for rat liver gap junctions described in Chapter 1, began as an effort to gather more information about the major protein component, particularly more protein sequence, to facilitate the isolation of cDNA clones for Cx32. The method developed is much more rapid and has improved yields over previous procedures, and gap junctions isolated by this method are highly purified based on morphological and biochemical criteria.

Chapter 1 also describes the use of isolated gap junctions to examine the organization of Cx32 in the membrane using proteases and antibodies against specific peptides, following the publication of the cDNA sequence and a topological model for the protein by others (Paul, 1986; Kumar and Gilula, 1986). Antibody binding and membrane protection experiments confirm the predicted protein sequence for Cx32 and agree in general with the proposed model. However, the protease sites detected in the carboxy-terminus, reveal that much of this region is unexpectedly protected, suggesting that it may be more organized than previously thought.

The structure of the isolated gap junction is probed further with an atomic force microscope (AFM). This is a new type of instrument that can image surfaces at high resolution under near physiological conditions. Using the AFM, the first high resolution images of a gap junction plaque in phosphate buffered saline were obtained. When examining the compressibility of the gap junction membrane with the AFM, a remarkable discovery was made. When a high force (about 10 nanoNewton) was applied to the tip of the microscope, the gap junction membrane could be "dissected" in half, leaving one membrane of the double membrane structure adsorbed to the substrate with the extracellular domains of connexons exposed. This provides direct experimental access to these domains. Force dissection could also be carried out on gap junctions that had been fixed with glutaraldehyde, suggesting that no crosslinks across the gap were formed. This in turn provides constraints on the model of Cx32, and provides an assay to further explore the nature of the interactions between the connexons across the gap.

The extracellular domains of the gap junction were imaged at a horizontal resolution of 1-2 nm, revealing the hexagonal array of channels. These images are the first direct visualizations of the gap junction protein as it protrudes from the membrane, and are in fact the first images of any ion channel by atomic force microscopy. They confirm some gap junction dimensions measured under less physiological conditions, but also provide new measurements not previously available. These observations open the door for a variety of experiments that were not previously possible, and begin to fulfil the promise of a new level of understanding of biological structures held by this technology.

While isolation of the Cx32 cDNA substantially improved our understanding of the structure of the gap junction, it also provided a new opportunity to study biological function. The approach attempted was to isolate clones for gap junction proteins from *Drosophila* by low stringency screening, and then utilize the powerful genetics of this organism to understand function. All attempts to isolate gap junction clones from *Drosophila* were unsuccessful. However, in parallel with the *Drosophila* libraries a rat genomic library was screened for the purpose of identifying new connexin homologues (Chapter 3). This screen produced genomic clones for the *Cx32* gene, and a new connexin, Cx31. The Cx32 gene was partially characterized and a nearby restriction fragment length polymorphism was identified. Cx31 is described in Chapter 4. This new connexin encodes a predicted protein of 30,960 Dalton and has a unique distribution when compared with other connexins.

Comparison of the Cx31 sequence with the other known connexins, reveals unique and conserved domains. This comparison is extended to a phylogenetic analysis of the entire gene family known to date that shows two major branches of connexins that diverged 1.3-1.9 billion years ago. Comparison with other ion channels reveals a short sequence similarity between the connexins and channels such as the voltage activated K^+ channel. In K^+ channels the sequence has been shown to line the aqueous pore, and the model for connexin organization is modified to account for this possibility (Chapter 1). The

similarity also suggests that gap junctions may be related and part of a superfamily of ion channels.

References

Allen, F., C. Tickle and A. Warner (1990). "The role of gap junctions in patterning of the chick limb bud." Development **108**: 623-634.

Arellano, R. O., A. Rivera and F. Ramon (1990). "Protein phosphorylation and hydrogen ions modulate calcium-induced closure of gap junction channels." Biophys. J. **57**: 363-367.

Arnsdorf, M. F. (1990). The cellular basis of cardiac arrhythmias. Electrocardiography: Past and future. New York, New York Academy of Sciences. 263-280.

Atkinson, M. M., et al. (1981). "Rapid and reversible reduction of junctional permeability in cells infected with a temperature-sensitive mutant of avian sarcoma virus." J. Cell Biol. **91**: 573-578.

Atkinson, M. M., S. K. Anderson and J. D. Sheridan (1986). "Modification of gap junctions in cells transformed by a temperature-sensitive mutant of Rous sarcoma virus." J. Membr. Biol. **91**: 53-64.

Baker, T. S., et al. (1985). "Gap junction structures. VII. Analysis of connexon images obtained with cationic and anionic negative stains." J. Mol. Biol. **184**: 81-98.

Baker, T. S., D. L. D. Caspar, C. J. Hollingshead and D. A. Goodenough (1983). "Gap junction structures. IV. Asymmetric features revealed by low-irradiation microscopy." J. Cell Biol. **96**: 204-216.

Baker, M. E. and M. H. Saier (1990). "A common ancestor for bovine lens fiber major intrinsic protein, soybean nodulin 26, and *Escherichia-coli* glycerol facilitator." Cell **60**(2): 185-186.

Barr, L., F. F. Dewey and W. Berger (1965). "Propagation of action potentials and the structure of the nexus in cardiac muscle." J. Gen. Physiol. **48**: 797-823.

Benedetti, E. L. and P. Emmelot (1968). "Hexagonal array of subunits in tight junctions separated from isolated rat liver plasma membranes." J. Cell Biol. **38**: 15-24.

Bennett, M. V. L. and D. A. Goodenough (1978). "Gap junctions, electrotonic coupling, and intercellular communication." Neurosci. Res. Prog. Bull. **16**(3): 371-486.

Beyer, E. C., D. L. Paul and D. A. Goodenough (1990). "Connexin family of gap junction proteins." J. Membr. Biol. **116**: 187-194.

Burt, J. M. and D. C. Spray (1988). "Single-channel events and gating behavior of the cardiac gap junction channel." Proc. Natl. Acad. Sci. USA **85**: 3431-3434.

Broekuyse, R. M., E. D. Kuhlman and A. L. Stols (1976). "Lens membranes. II. Isolation and characterization of the main intrinsic polypeptide (MIP) of bovine lens fiber membranes." Exp. Eye Res. **29**: 303-313.

Caspar, D. L. D., D. A. Goodenough, L. Makowski and W. C. Phillips (1977). "Gap junction structures. I. Correlated electron microscopy and X-ray diffraction." J. Cell Biol **74**: 605-628.

Caveney, S. (1985). "The role of gap junctions in development." Ann. Rev. Physiol. **47**: 319-335.

Dahl, G., et al. (1987). "Expression of functional cell-cell channels from cloned rat liver gap junction complementary DNA." Science **236**: 1290-1293.

Dermietzel, R., et al. (1987). "Major loss of the 28,000 Dalton protein of gap junctions in proliferating hepatocytes." J. Cell Biol. **105**(4): 1925-1934.

Dewey, M. M. and L. Barr (1962). "Intercellular connections between smooth muscle cells: The nexus." Science **137**: 670-672.

Eghbali, B., J. A. Kessler and D. C. Spray (1990). "Expression of gap junction channels in communication-incompetent cells after transfection with the cDNA encoding connexin 32." Proc. Nat. Acad. Sci. USA **87**: 1328-1331.

Finbow, M. E., J. Shuttleworth, A. E. Hamilton and J. D. Pitts (1984). "Analysis of the vertebrate gap junction protein." EMBO J. **2**: 1479-1486.

Finbow, M. E., S. B. Yancey, R. G. Johnson and J. Revel (1980). "Independent lines of evidence suggesting a major gap junctional protein with a molecular weight of 26,000." Proc. Nat. Acad. Sci. U.S.A. **77**: 970-974.

Flagg-Newton, J., I. Simpson and W. R. Loewenstein (1979). "Permeability of the cell-to-cell membrane channels in mammalian cell junction." Science **205**: 404-407.

Fraser, S. E., C. Green, H. R. Bode and N. B. Gilula (1987). "Selective disruption of gap junctional communication interferes with a patterning process in hydra." Science **237**: 49-55.

Fujimoto, W. Y., H. Subak-Sharpe and J. E. Seegmiller (1971). "Hypoxanthine-guanine phosphoribosyltransferase deficiency: chemical agents selective for mutant or normal cultured fibroblasts in mixed and heterozygote cultures." Proc. Nat. Acad. Sci. U.S.A. **68**: 1516-1519.

Furshpan, E. J. and D. D. Potter (1959). "Transmission at the giant motor synapses of the crayfish." J. Physiol. **145**: 289-325.

Garfield, R. E., S. M. Sims and E. E. Daniel (1977). "Gap junctions: their presence and necessity in myometrium during parturition." Science **198**: 958-959.

Goodenough, D. A., J. S. B. Dick and J. E. Lyons (1980). "Lens metabolic cooperation: a study of mouse lens permeability visualized with freeze-substitution autoradiography and electron microscopy." J. Cell Biol. **82**(2): 576-589.

Gorin, M. B., et al. (1984). "The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning." Cell **39**: 49-59.

Henderson, D., H. Eibl and K. Weber (1979). "Structure and biochemistry of mouse hepatic gap junctions." J. Mol. Biol. **132**: 193-218.

Hertzberg, E. L. (1984). "A detergent-independent procedure for the isolation of gap junctions from rat liver." J. Biol. Chem. **259**: 9936-9943.

Hertzberg, E. L., et al. (1988). "Topology of the Mr 27,000 liver gap junction protein." J. Biol. Chem. **263**(35): 19105-19111.

Hertzberg, E. L., D. C. Spray and M. V. L. Bennett (1985). "Reduction of gap junctional communication by microinjection of antibodies against the 27 kD liver gap junction protein." Proc. Nat. Acad. Sci. U.S.A. **82**: 2412-2416.

Hooper, M. L. and J. H. Subak-Sharpe (1981). "Metabolic cooperation between cells." Int. Rev. Cytol. **69**: 45-104.

Karrer, H. E. and J. Cox (1960). "The striated musculature of blood vessels. II. Cell interconnections and cell surface." J. Biophys. Biochem. Cytol. **8**: 135-150.

Lash, J. A., E. S. Critser and M. L. Pressler (1990). "Cloning of a gap junctional protein from vascular smooth muscle and expression in two-cell mouse embryos." J. Biol. Chem. **265**(22): 13113-13117.

Lawrence, T. S., W. H. Beers and N. B. Gilula (1978). "Transmission of hormonal stimulation by cell-cell communication." Nature **272**: 501-506.

Lee, S., N. B. Gilula and A. E. Warner (1987). "Gap junctional communication and compaction during preimplantation stages of mouse development." Cell **51**: 851-860.

Loewenstein, W. R. and Y. Kanno (1964). "Studies on an epithelial (gland) cell junction I. Modifications of surface membrane permeability." J. Cell Biol. **22**: 565-586.

Loewenstein, W. R. and Y. Kanno (1967). "Intercellular communication and tissue growth. I. Cancerous growth." J. Cell Biol. **33**: 225-234.

Makowski, L., et al. (1984b). "Gap junction structures. VI. Variation and conservation in connexon conformation and packing." Biophys. J. **45**: 208-218.

Makowski, L., D. L. D. Caspar, D. A. Goodenough and W. C. Phillips (1982). "Gap junction structures. III. The effect of variations in the isolation procedure." Biophys. J. **37**: 189-191.

Makowski, L., D. L. D. Caspar, W. C. Phillips and D. A. Goodenough (1977). "Gap junction structures. II. Analysis of the X-Ray diffraction data." J. Cell Biol. **74**: 629-645.

Makowski, L., D. L. D. Caspar, W. C. Phillips and D. A. Goodenough (1984a). "Gap junction structures. V. Structural chemistry inferred from X-ray diffraction measurements on sucrose accessibility and trypsin susceptibility." J. Mol. Biol. **174**: 449-481.

Malewicz, B., V. V. Kumar, R. G. Johnson and W. J. Baumann (1990). "Lipids in gap junction assembly and function." Lipids **25**(8): 419-427.

Mehta, P., J. Bertram and W. R. Loewenstein (1986). "Growth inhibition of transformed cells correlates with their junctional communication with normal cells." Cell **44**: 187-196.

Murray, A. and H. S. C. Gainer (1989). Regulation of gap junctional communication by protein kinases. Cell interactions and gap junctions. Boca Raton, CRC Press. 97-107.

Nicholson, B. J., et al. (1985). "The Mr 28,000 gap junction proteins from rat heart and liver are different but related." J. Biol. Chem. **260**(11): 6514-6517.

Nicholson, B., et al. (1987). "Two Homologous Protein Components of Hepatic Gap Junctions." Nature **329**(6141): 732-734.

Nicholson, B. J. and J.-P. Revel (1983). "Gap junctions in liver: Isolation, morphological analysis, and quantitation." Meth. Enzymol. **98**: 519-537.

Paul, D. L. (1986). "Molecular cloning of cDNA for rat liver gap junction protein." J. Cell Biol. **103**: 123-134.

Peracchia, C. (1980). "Structural correlates of gap junction permeation." Int. Rev. Cytol. **66**: 81-146.

Pitts, J. D. (1986). "Retinoic acid inhibits junctional communication between animal cells." Carcinogenesis **7**: 1003-1008.

Potter, D. D., E. J. Furshpan and E. S. Lennox (1966). "Connections between cells of the developing squid as revealed by electrophysiological methods." Proc. Nat. Acad. Sci. **55**: 328-366.

Rao, Y., L. Y. Jan and Y. N. Jan (1990). "Similarity of the product of the *Drosophila* neurogenic gene *big brain* to transmembrane channel proteins." Nature **345**: 163-167.

Reber, W. R. and R. Weingart (1982). "Ungulate cardiac purkinje fibers: the influence of intracellular pH on the electrical cell to cell coupling." J. Physiol. **328**: 87-104.

Revel, J.-P., S. B. Yancey and B. J. Nicholson (1985). Biology of Gap Junction Molecules and Development. The Cell in Contact: Adhesion Junctions as Morphogenetic Determinant. 411-425.

Revel, J.-P. and M. Karnovsky (1967). "Hexagonal array of subunits in intercellular junctions of the mouse heart and liver." J. Cell Biol. **33**: C7-C12.

Robertson, J. D. (1963). "The occurrence of a subunit pattern in the unit membrane of club endings in the Mauthner cell synapse in the goldfish brain." J. Cell Biol. **19**: 201-221.

Rose, B. and W. R. Loewenstein (1975). "Permeability of cell junctions depends on local cytoplasmic calcium activity." Nature **254**: 250-252.

Saez, J., et al. (1986). "cAMP increases junctional conductance and stimulates phosphorylation of the principal gap junction polypeptide." Proc. Nat. Acad. Sci. U.S.A. **83**: 2473.

Smith, B. L. and P. Agre (1991). "Erythrocyte Mr-28,000 transmembrane protein exists as a multisubunit oligomer similar to channel proteins." J. Biol. Chem. **266**: 6407-6415.

Sosinsky, G. E., J. C. Jesior, D. L. D. Caspar and D. A. Goodenough (1988). "Gap junction structures. VIII. Membrane cross-section." Biophys. J. **53**: 709-722.

Sperelakis, N. and J. E. Mann (1977). "Evaluation of electric field changes in the cleft between excitable cells." J. Theor. Biol. **64**: 71-96.

Spray, D. C., et al. (1984). "Physiological and pharmacolgical properties of gap junctions between dissociated pairs of rat hepatocytes." J. Cell Biol. **99**: 344a.

Spray, D. C., et al. (1986). "Isolated liver gap junctions: Gating of transjunctional current is similar to that in intact pairs of rat hepatocytes." Proc. Nat. Acad. Sci. U.S.A. **83**: 5494-5497.

Spray, D. C. and M. V. L. Bennett (1985). "Physiology and pharmacology of gap junctions." Ann. Rev. Physiol. **47**: 281-303.

Subak-Sharpe, H., R. R. Bürk and J. D. Pitts (1966). "Metabolic co-operation by cell-cell transfer between genetically different mammalian cells in tissue culture." Heredity **21**: 342-343.

Subak-Sharpe, H., R. R. Bürk and J. D. Pitts (1969). "Metabolic co-operation between biochemically marked cells in tissue culture." J. Cell Sci. **4**: 353-367.

Swenson, K. I., J. R. Jordan, E. C. Beyer and D. L. Paul (1989). "Formation of gap junctions by expression of connexins in *Xenopus* oocyte pairs." Cell **57**: 145-155.

Swenson, K. I., H. Piwnica-Worms, H. McNamee and D. L. Paul (1990). "Tyrosine phosphorylation of the gap junction protein connexin43 is required for the pp60$^{v\text{-}src}$-induced inhibition of communication." Cell Regul. **1**: 989-1002.

Turin, L. and A. Warner (1977). "Carbon dioxide reversibly abolishes ionic communication between cells of early amphibian embryo." Nature **270**: 56-57.

Unwin, P. N. T. and G. Zampighi (1980). "Structure of the junction between communicating cells." Nature **283**: 545-549.

Unwin, P. N. T. and P. D. Ennis (1984). "Two configurations of a channel-forming membrane protein." Nature **307**: 609-613.

Voorter, C. E. M. and J. Kistler (1989). "cAMP-dependent protein kinase phosphorylates gap junction protein in lens cortex but not in lens nucleus." Biochem. Biophys. Acta **986**: 8-10.

Warner, A. E., S. C. Guthrie and N. B. Gilula (1984). "Antibodies to gap junctional protein selectively disrupt junctional communication in the early amphibian embryo." Nature **311**: 127-131.

Wolpert, L. (1978). Gap junctions: channels for communication in development. Intercellular junctions and synapses. London, Chapman and Hall. 83-94.

Yancey, S. B., et al. (1989). "The 43-kD Polypeptide of Heart Gap Junctions: Immunolocalization (I), Topology (II), and Functional Domains (III)." J. Cell Biol. **108**: 2241-2254.

Young, J. D., Z. A. Cohn and N. B. Gilula (1987). "Functional assembly of the gap junction conductance in lipid bilayers: demonstration that the major 27 kd protein forms the junctional channel." Cell **48**: 733-743.

Zervos, A. S., J. Hope and H. W. Evans (1985). "Preparation of a gap junction fraction from uteri of pregnant rats: the 28-kD polypeptides of uterus, liver and heart gap junctions are homologous." J. Cell Biol. **101**: 1363-1370.

Chapter 1

Isolation and Biochemical Analysis of Hepatic Gap Junctions

Introduction

Understanding the composition and organization of gap junctions is fundamental to understanding their biological activity and function. Cell fractions enriched in gap junctions were first identified morphologically by Benedetti and Emmelot (1965; 1968), though at the time they thought they had isolated tight junctions. Two features of that work have remained the center piece of most isolation efforts since, the relative detergent resistance of the junctional membrane to separate it from non-junctional membrane, and the use of electron microscopy and negative stains such as uranyl acetate and phosphotungstic acid to assay the purity of preparations. Several isolation methods based on these principles have been described (Henderson et al., 1979; Finbow, 1980; Nicholson and Revel, 1983; Hertzberg, 1984). Gap junctions isolated from the rodent liver are the preparation of choice for studying the biochemical composition, structure, and topology of the cell to cell channel and its protein components.

Based on X-ray diffraction and electron microscopy of isolated gap junctions, two models of the gap junction channel have been suggested (Makowski et al., 1977; Unwin and Zamphigi, 1980; Makowski et al., 1985). These models, at a resolution of approximately 2.5 nm, show the gap junction channel as a roughly cylindrical structure 15 nm tall and 7-8 nm in diameter, with a 2-3 nm aqueous pore in the middle. Each channel is composed of two half channels, often refered to as connexons, that meet head to head in the gap between the two membranes. Each connexon exhibits six fold symmetry and is thought to be composed of six, or a multiple of six protein subunits.

The major protein component of the rodent hepatic gap junction was initially identified as a 28 kiloDalton (kD) molecule (Henderson et al., 1979; Hertzberg and Gilula, 1979; Finbow et al., 1980). A minor component of 21 kD was also present is some preparations, particularly from mouse liver (Henderson et al., 1979). These two proteins, and the 45 kD cardiac gap junction protein, were shown by amino-terminal (N-terminal)

sequencing to be members of a gene family (Nicholson et al., 1985; Nicholson et al., 1987). Subsequently cDNAs for these proteins were isolated (Paul, 1986; Beyer et al., 1987; Zhang et al., 1989), and based on the predicted molecular mass of the 21 kD, 28 kD, and 45 kD proteins, they were renamed connexin-26 (Cx26)[1], connexin-32 (Cx32) and connexin-43 (Cx43) respectively. The sequences of these and several other connexins that have now been identified are discussed in Chapter 3.

Proteolytic analysis of Cx32 indicated the presence of two 10 kD membrane protected domains, one of them contained the N-terminus (Nicholson, 1983). While this data did not reveal if the carboxy-terminus (C-terminus) was accessible to proteases, it did define a cytoplasmic protease hypersensitive region separating the two protected domains, and provided a biochemical context in which to interpret the Cx32 protein sequence. Standard Kyte and Doolittle (1982) hydropathy analysis of that sequence revealed four highly hydrophobic transmembrane domains, and several similar models for the organization of the protein have now been proposed (Paul, 1986; Hertzberg et al., 1988; Milks et al., 1988). These models predict that the N-terminus and C-terminus of Cx32 are cytoplasmic. There are four transmembrane segments (TM1-TM4), two extracellular loops, the first separating TM1 from TM2, and the second TM3 from TM4, and a cytoplasmic intracellular loop (IL) separating TM3 from TM4. This model has now been tested in a variety of ways, including with proteases (Hertzberg et al., 1988) and antibodies against specific peptides (Milks et al., 1988; Goodenough et al., 1988). This model has also been tested for Cx43 (Yancey et al., 1990; Laird and Revel, 1990), and based on sequence homology, is thought to hold for all connexins.

This chapter describes an improved isolation procedure for hepatic gap junctions, characterization of the isolated junctions, and the use of isolated gap junctions to confirm and improve the current topological model for Cx32. The characterization includes

---

[1] The hepatic gap junction proteins will be refered to as the 28 kD protein or Cx32, and 21 kD or Cx26 interchangeably. See Table 1 in Chapter 4 for naming of other connexins.

morphological purity based on electron microscopy, biochemical purity based on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and N-terminal sequence analysis. Membrane protection experiments using a variety of proteases provide further insight into the organization and topology of this gap junction protein. In addition antibodies against specific peptides from the predicted Cx32 sequence were produced and used to examine the arrangement of the protein. The results are discussed in relationship to a modified model for the organization of Cx32.

Materials and Methods

Isolation of Hepatic Gap Junctions [2]

Twelve Sprague-Dawley rats (150-200 g, Simonsen Laboratories, Gilroy California or Harlan Sprague-Dawley, Indianapolis Indiana) were routinely used. Rats were killed by cervical dislocation and the livers were perfused with normal saline through the spleen after cutting the renal vein. The livers homogenized in ice cold 1 mM $NaHCO_3$ buffer (BB), pH 8.2, with 1 mM phenylmethylsulfonyl fluoride (PMSF) with a tissuemizer (Tekmar Ultra Turrax, SDT-182 EN) for two bursts of 10 seconds. Homogenization was carried out after the removal of every three livers to minimize proteolytic degradation. The homogenate was then diluted to 1200 ml with BB/PMSF and filtered through 16 layers of cheesecloth. The filtrate was centrifuged at 11,000 rpm (16,000 x g_{av}) in a Sorvall GSA rotor for 50 minutes at 4° C. The supernatants were aspirated and about 50 ml of 5 mM Tris-(hydroxymethyl)-aminomethane pH 10 (Tris10) with 1 mM PMSF was added to each bottle. The loose fluffy portion of the pellet was removed by gently swirling the bottles, being careful not to disturb the harder core of the pellet. These fluffy parts of the pellets were pooled and diluted to 600 ml with Tris10/PMSF and homogenized for a few seconds with the tissuemizer. At room temperature, 600 ml of 1.1% of n-lauryl sarcosine (sarkosyl)

---

[2] The isolation procedure for gap junctions was developed together with Dr. James D. Hatton while he was a postdoctoral fellow in the laboratory, and parts of this chapter describing the isolation method were adapted from a manuscript prepared by Dr. Hatton and myself. That manuscript was not published.

in Tris10/PMSF was stirred in slowly. The solution was stirred for 10 minutes, after which it was centrifuged at 11,000 rpm in a Sorvall GSA rotor for 50 minutes at 18° C. The supernatants were gently aspirated and the top parts of the pellets were resuspended in 40 ml of 0.3% sodium deoxycholate (DOC) in Tris10 by 4-5 strokes of a Dounce homogenizer (Wheaton, pestle B). Two discontinuous sucrose gradients were prepared by successively layering 8 ml 49% (w/v) sucrose in Tris10/DOC, 10 ml 35% (w/v) sucrose in Tris10/DOC and 20 ml sample. The gradients were centrifuged at 25,000 rpm ($81,000 \times g_{av}$) in a Beckman SW27 or SW28 rotor for 1 hour at 18° C. Gap junctions were collected at the 35/49 interface, diluted with 2-3 volumes of Na_2CO_3 (pH 11) and centrifuged at 25,000 rpm for 20 minutes. The pellets were collected in 0.5 ml distilled water or phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4, and 4.3 mM Na_2HPO_4 at pH 7.2) and stored at 4° C for several days or -20° C for longer periods.

Electron Microscopy

Isolated gap junctions were routinely stained for electron microscopy with uranyl acetate (UA) or phosphotungstic acid (PTA). A drop of isolated gap junctions was pipetted onto a standard carbon coated grid and allowed to adsorb for a minute. Saturated UA in water was filtered through a 0.2 μm membrane and added a fraction of a drop at a time to approximately a volume equal to the gap junction volume. The sample was allowed to stain for 30-60 seconds after which the liquid was drawn away from the edge with a tissue paper. The grid was then air dried for several minutes and examined in a Phillips 201 electron microscope operating at 80 kV. PTA (2% (w/v), pH 7.2) was used in the same way as UA except that 1-2 minutes was allowed for the staining.

PAGE and N-terminal Sequence Analysis

Gap junction proteins and proteolytic fragments were separated by SDS-PAGE on 12% or 15% gels as described (Laemmli, 1970). Gels were usually stained with coomassie blue, though occasionally silver staining was employed (Merril et al., 1984). Molecular

mass was estimated by linear regression analysis against the standards phosphorylase B (97.4 kD), bovine serum albumin (66.2 kD), ovalbumin (45.0 kD), carbonic anhydrase (31.0 kD), soy bean trypsin inhibitor (21.5 kD) and lysozyme (14.4 kD). Electro-elution of the Cx32 protein from gel slices was performed as previously described (Nicholson, 1983). N-terminal sequencing of the eluted protein was carried out at the Caltech microchemical facility.

Proteolytic Analysis and Membrane Protection

Isolated liver gap junctions were digested with trypsin, V8 protease (specific for Glu or Glu/Asp), endoproteinase Arg-C (EndoArg-C), or endoproteinase Lys-C (EndoLys-C) (Boeringer Mannheim, Indianapolis, Indiana). All digestions, except Glu specific V8 protease (V8-Glu), were carried out in 0.1 M NaH_2PO_4 (pH 8) at an enzyme to substrate ratio of between 10:1 and 20:1 for 3 hours at 37° C. Proteolysed gap junctions were separated from proteases by centrifugation at c. 14,000 x g in a microfuge for 30 minutes. The supernatant with the proteases was aspirated off, and the pellet of gap junctions was solubilized and separated by SDS-PAGE. The Glu specific V8 protease was used in the same conditions except that the buffer was 0.1 M $(NH_4)CO_3$ or $NH_4(CH_3COO)$. In addition to the different enzymes, a series of digestions with V8-Glu at pH's ranging from 5 to 8 were performed. Bovine serum albumin was used as a control for enzyme activity throughout the pH range used.

Production of Antibodies

Three peptides corresponding to amino acids 222-231 (CT222), 238-247 (CT238), and 274-283 (CT274) from the C-terminal tail of Cx32 were synthesized as amides by Biosearch (San Rafael, California). These peptides were analyzed by reverse phase high pressure liquid chromatography (HPLC) to judge their purity, in addition to the HPLC analyses provided by the manufacturer. The peptides were determined to be sufficiently pure for antibody production and used without further purification. The actual sequence of these peptides is shown in Table 1.

Name	Position	Length	Peptide Sequence	Source
CT222	222-231	10	QRRSNPPSRK	Biosearch
CT238	238-247	10	RLSPEYKQNE	Biosearch
CT274	274-283	10	AEKSDRCSAC	Biosearch
IL100	100-129	31	KKMLRLEGHGDPLHLEEVKKHKVHISGTWC	MCF
IL107	107-129	24	GHGDPLHLEEVKKHKVHISGTWC	MCF
IL112	112-129	19	LHLEEVKKHKVHISGTWC	MCF

Table 1. Peptides from the predicted Cx32 sequence used for antibody production. The CT peptides were synthesized as amides by BioSearch (San Rafael, California). The nested series of IL peptides were synthesized at the Caltech microchemical facility (MCF). These peptides had a cysteine not in the actually Cx32 sequence added to their C-termini, for the possible use in coupling to a carrier protein.

A nested series of peptides corresponding to amino acids 100-129 (IL100), 107-129 (IL107), and 112-129 (IL112) from the intracellular loop region of Cx32 was synthesized at the Caltech microchemical facility. These peptides all had a cysteine (not in the actual protein sequence) added to their N-termini for the possible use in coupling to carrier proteins. All the IL peptides were purified by reverse phase HPLC at the Caltech microchemical facility before use.

Peptides were coupled to Key Hole Limpet Hemocyanin (KLH) using difluorodinitrobenzene (DFDNB) as described by Tager (1976), or to ovalbumin by glutaraldehyde (GA) based cross linking. KLH coupling to peptides, 1-2 mg peptide was dissolved in 200 µl 0.1 M KH_2PO_4 and 7 M guanidine-HCl, pH 7.2, and 1 ml DFDNB (30 mg/ml in methanol) was added. The mixture was allowed to react for 15 minutes at room temperature in a 15 ml Corex tube. Unreacted DFDNB was then removed by several ether extractions, which also caused the FDNB-peptide to precipitate. The precipitate was vacuum dried briefly, resuspended in 0.1-0.2 ml 0.1 M KH_2PO_4 and 7 M guanidine-HCl, pH 7.2, and 0.3 ml KLH (in saturated sodium borate, pH 10) was added. Coupling was allowed to proceed for 24 hours at room temperature, after which the mixture was dialysed against several changes of PBS. The final KLH-peptide was diluted to 4.5 ml with PBS and stored at -20° C. For GA coupling, 7.5 mg ovalbumin (OVA) and 7.5 mg peptide were dissolved in 1 ml PBS and briefly centrifuged to remove debris. Glutaraldehyde (25 % v/v)

was added to a final concentration of 0.25 % and the mixture was incubated at room temperature for 15 minutes. An additional aliquott of GA was added, to give a total concentration of 0.375 % (v/v), and the mixture was incubated for an additional 15 minutes after which 100 µl 1 M glycine (pH 6.0) was added to quench excess aldehyde groups.

New Zealand White Rabbits were injected with CT222-KLH, CT238-KLH, or CT274-KLH. To minimize discomfort for the rabbits an alternative to Freund's adjuvant, RIBI MPL+TDM+CWS emulsion (RIBI ImmunoChem, Inc., Hamilton, Montana), was used. Two ml of the coupling mixture was used to resuspend the RIBI as per manufacturers instructions. Injections were 2 x 0.2 ml intramuscular in the leg, 6 x 0.05 ml intradermal on the back, and 3 x 0.1 ml subcutaneous behind the neck. Rabbits were boosted after 4 weeks, and bled on the 7th, 10th, and 14th day after the boost. Subsequently rabbits were boosted every 3-4 weeks. After several months the CT222-KLH and CT238-KLH did not produce any response. Therefore these peptides, and the IL100 peptide were coupled to ovalbumin using glutaraldehyde and used to raise antibodies as above. The IL100 peptide was also used without coupling to a carrier.

Antibody activity was assayed by enzyme linked immuno-sorbant assay (ELISA) or western blots against isolated gap junctions or the specific peptides. Controls for the specificity of the antibodies were peptides, rat liver plasma membranes, or crude membrane fractions from other organs.

ELISAs and Western Blots

ELISAs performed by dissolving peptides or isolated gap junctions in 0.5 M Na_2CO_3, pH 9.0, and adsorbing the peptides or junctions to an uncoated flat bottom microtiter dish for 1-4 hours. The wells were then aspirated and washed with TBS-T (10 mM Tris, 140 mM NaCl, 0.1% (v/v) Tween-20). Non-specific binding was prevented by blocking the wells with 1% BSA in TBS (same as TBS-T without Tween-20) for 1-2 hours. Primary antibodies were diluted to the appropriate concentration in TBS with 1% BSA, applied to the wells, and incubated for 1-4 hours. After washing with TBS-T, the

secondary antibody, anti-rabbit IgG coupled to either horse radish peroxidase (HRP) or alkaline phosphatase (AP), was added to the wells at a concentration of 0.2-1.0 ng/ml. Second antibody was incubated for 1-2 hours, and plates were washed and developed with either chloronaphthol for HRP, or 5-bromo-4-chloro-3-indoyl-phosphate together with nitro blue tetrazolium for AP. Absorbance was determined at 650 nm on a Vmax microplate reader (Molecular Devices, Menlo Park, California).

Western blots of proteins separated by SDS-PAGE were produced using a custom built semi-dry apparatus (Kyhse-Andersen, 1984). Gels were soaked briefly in 10 mM Caps (3-(cyclohexylamino)-propanesulfonic acid) with 10 % (v/v) methanol, sandwiched between 3 sheets of 3MM paper (on each side) soaked in caps/methanol, and transfered to nitrocellulose for 30-45 minutes at 1-2 mA/cm^2. The blots were air dried prior to further use, and cut into strips if necessary. To prevent non-specific binding to the blot, it was first incubated in 1% BSA in TBS for 1-2 hours. The primary antibody was then added directly to the blocking solution at the desired concentration and incubated on a rocker for 1-2 hours. The blots were washed at least 3 times 5 minutes with TBS-T before being incubated with the secondary antibody (anti-IgG-HRP or anti-IgG-AP) for 1-2 hours, and finally developed with the same chromogens as the ELISAs.

Results and Discussion

Gap Junction Isolation and Characterization

Isolation of gap junctions (Figure 1) from 12 rat livers was routinely completed in less than 5 hours, a significant advance over the two days required by previous methods (Henderson et al., 1979; Finbow et al., 1980; Nicholson and Revel, 1983; Hertzberg, 1984). Low magnification electron micrographs of PTA stained preparations show flat pieces of gap junction membrane, often referred to as plaques, about 0.5 to 1.5 μm large (Figure 2). These membranes sometimes form large aggregates on the grid, but generally have the same distribution and over all morphology as previously described.

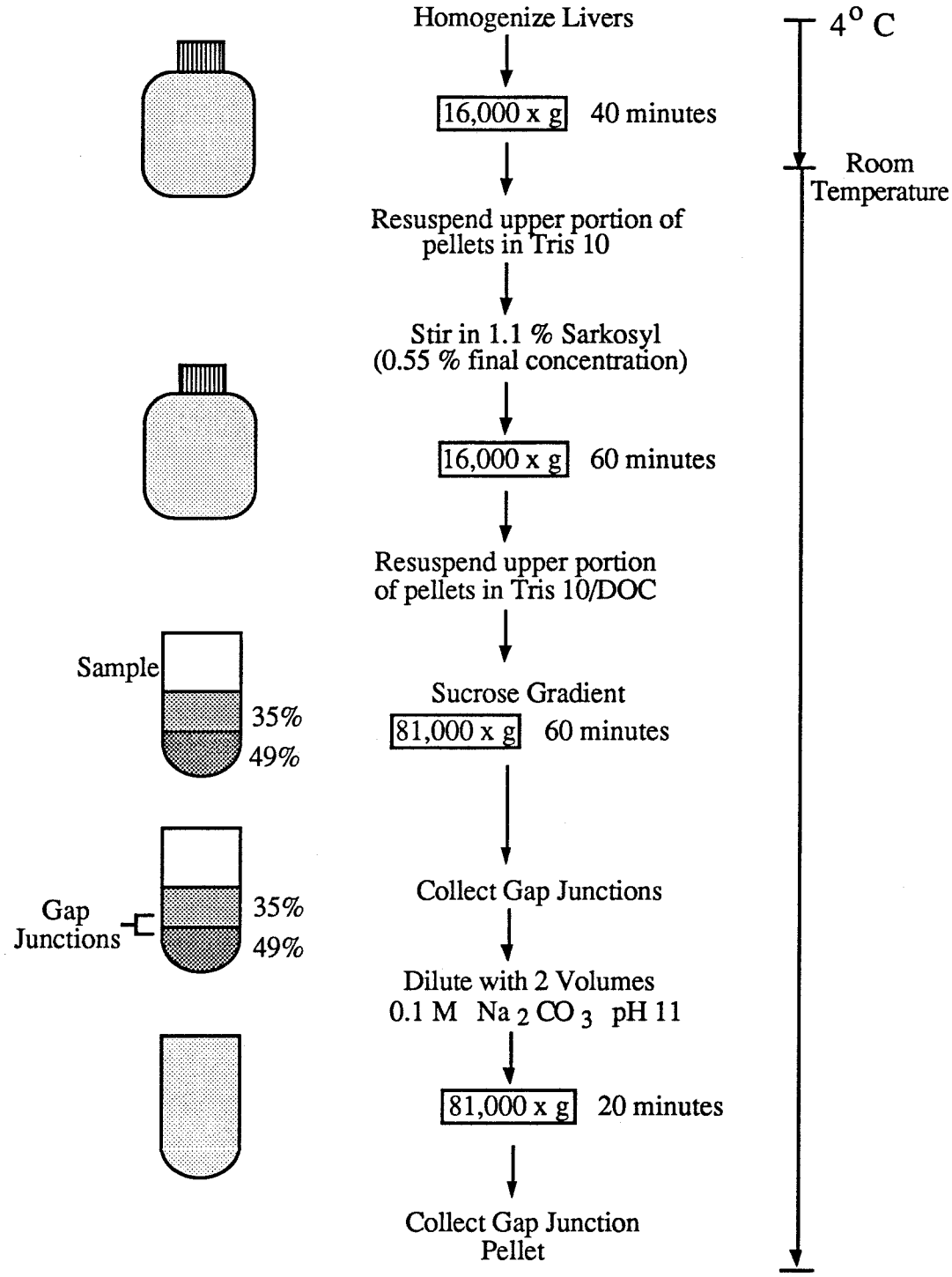
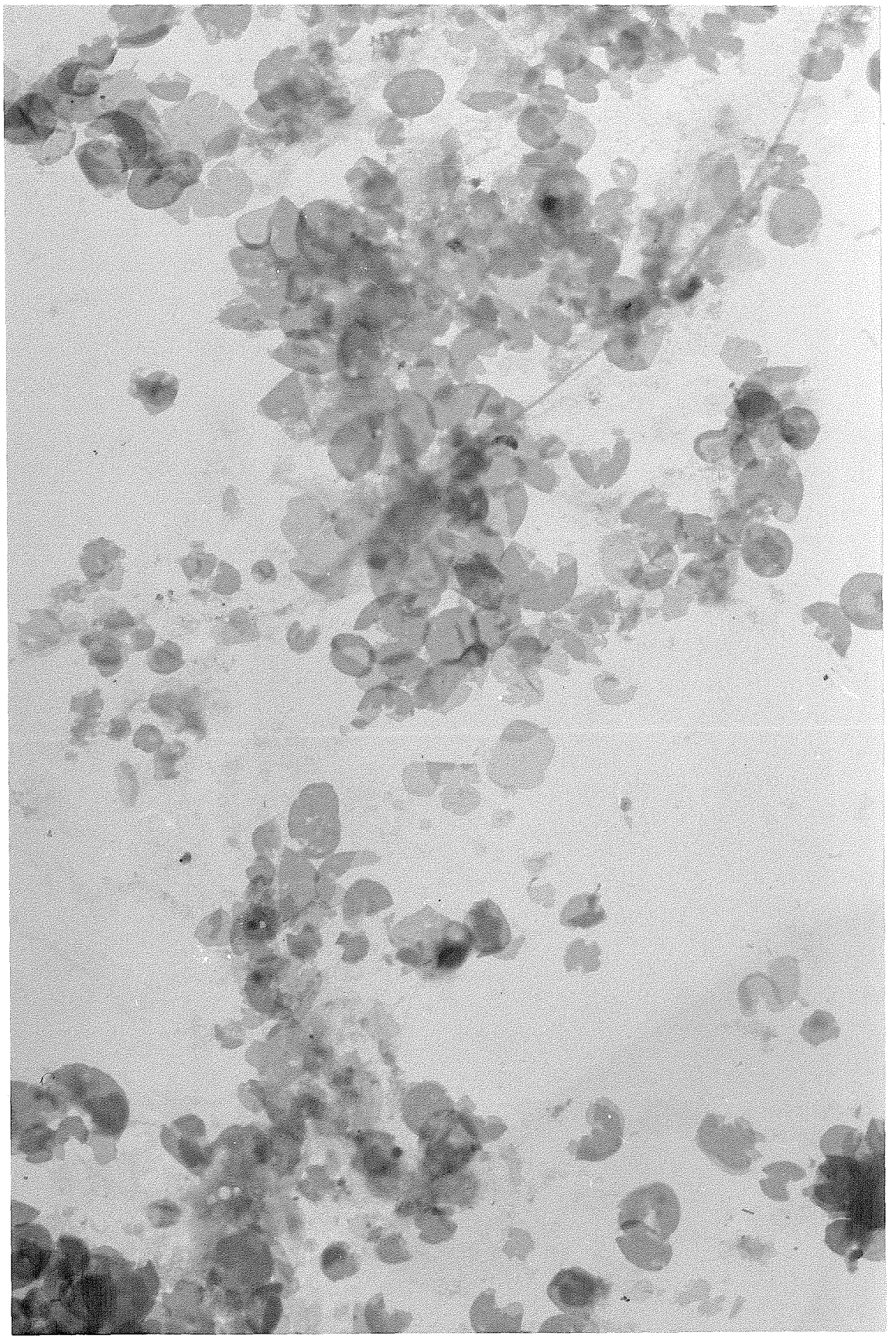


Figure 1. Flow chart for the isolation of rat liver gap junction. This procedure was routinely used on 12 or 24 rat livers, and carried out in less than 5 hours. It produced 2.8 μg of Cx32 per gram of liver.

Figure 2. Low magnification (11,800x) electron micrograph of isolated gap junctions stained with phosphotungstic acid. Plaque shaped structures typical of isolated gap junctions are seen over the entire field. Few contaminants are seen, however a single fiber of what is probably collagen is present.



The high magnification micrograph of a UA stained gap junction plaque reveals the hexagonal array of cell to cell channels at a center to center spacing of 7.7 nm (Figure 3). The image clearly shows a dot in the middle of the connexon, which only seen with UA, that presumably represents the pore. PTA would routinely reveal the hexagonal array but the pore was not visible. Exactly where in the pore the UA is binding cannot be determined from this image. However, it is known that gap junctions isolated by methods similar to the one used here exclude sucrose from the pore, suggesting that the channels are closed at the ends and that the connexon-connexon interactions are tight (Makowski et al., 1984). This means UA binding is probably on the intracellular side of the channel near the mouth of the pore. Since the uranium ion is positively charged, and the PTA is negatively charged, this in turn may indicate that there are negative charges near the pore opening. These observations are similar to differential staining of the pore by uranyl and tungstate based stains described by Baker et al. (1985). These investigators believe that the pore is made leaky by the staining process or the electron radiation, and that uranyl stains actually penetrate the channel and bind to negative residues within the pore. Additional evidence for a negative charge in or near the pore is provided by the Brink and Dewey (1980), who show that injecting an axon with aminofluorescein slows subsequent transfer of carboxyfluorescein, whereas the converse is not true. They propose that this is due to the binding of aminofluorescein to the gap junction pore, which prevents the subsequent transfer of carboxyfluorescein.

Isolated gap junctions would often fold onto themselves revealing the profile of the gap junction (Figure 4). These folds are about 35 nm thick and consist of the adjacent pentalaminar gap junction profiles. The profiles result from the accumulation of stain on the cytoplasmic surfaces, and in the gap. It is possible to see the distinctive hexagonal array as it folds into the typical double membrane structure, similar to the classic lanthanum stained preparations of Revel and Karnovsky (1967). Some profiles also show the channels crossing the gap (Figure 5).

Figure 3. High magnification (445,000x) electron micrograph of a gap junction plaque stained with uranyl acetate. It shows the individual channels at a center to center spacing of 7.7 nm. A dot of stain in the middle of each connexon probably represents the pore, or the mouth of the pore.

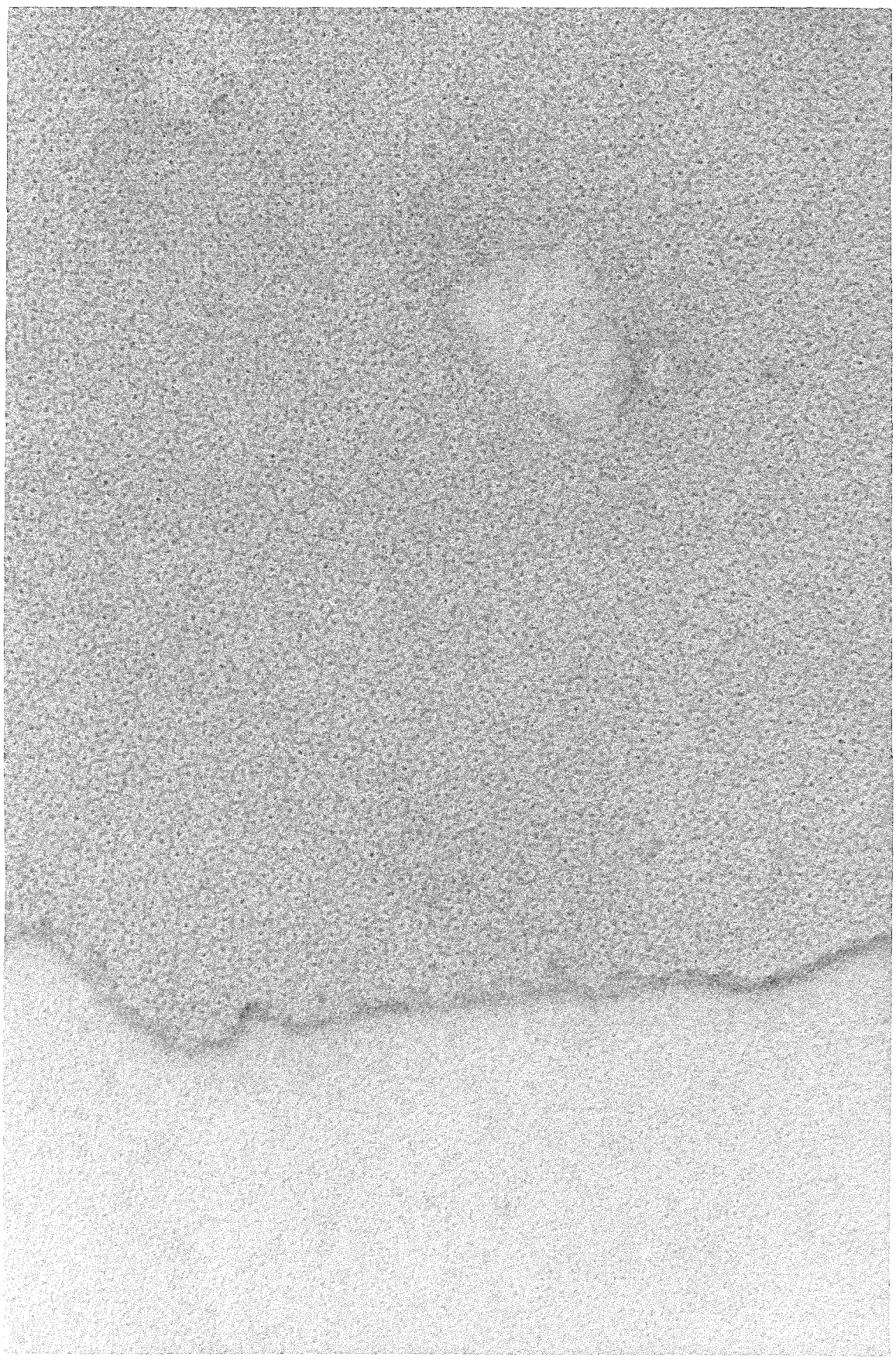


Figure 4. Electron micrograph of negatively stained isolated gap junction (231,000x). The hexagonal array of channels is seen, and a 34.6 nm thick fold in the structure reveals the gap junction in profile.

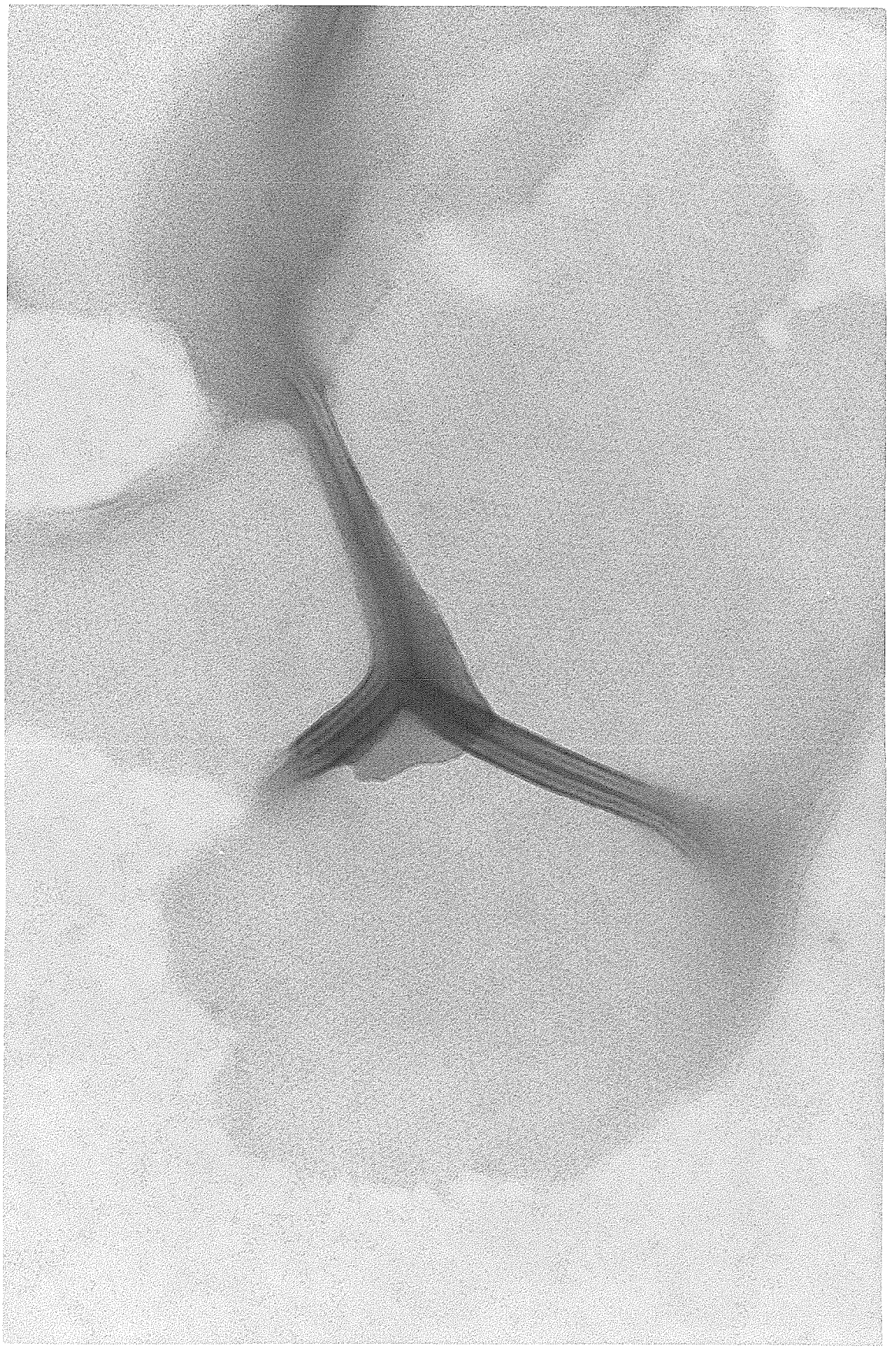
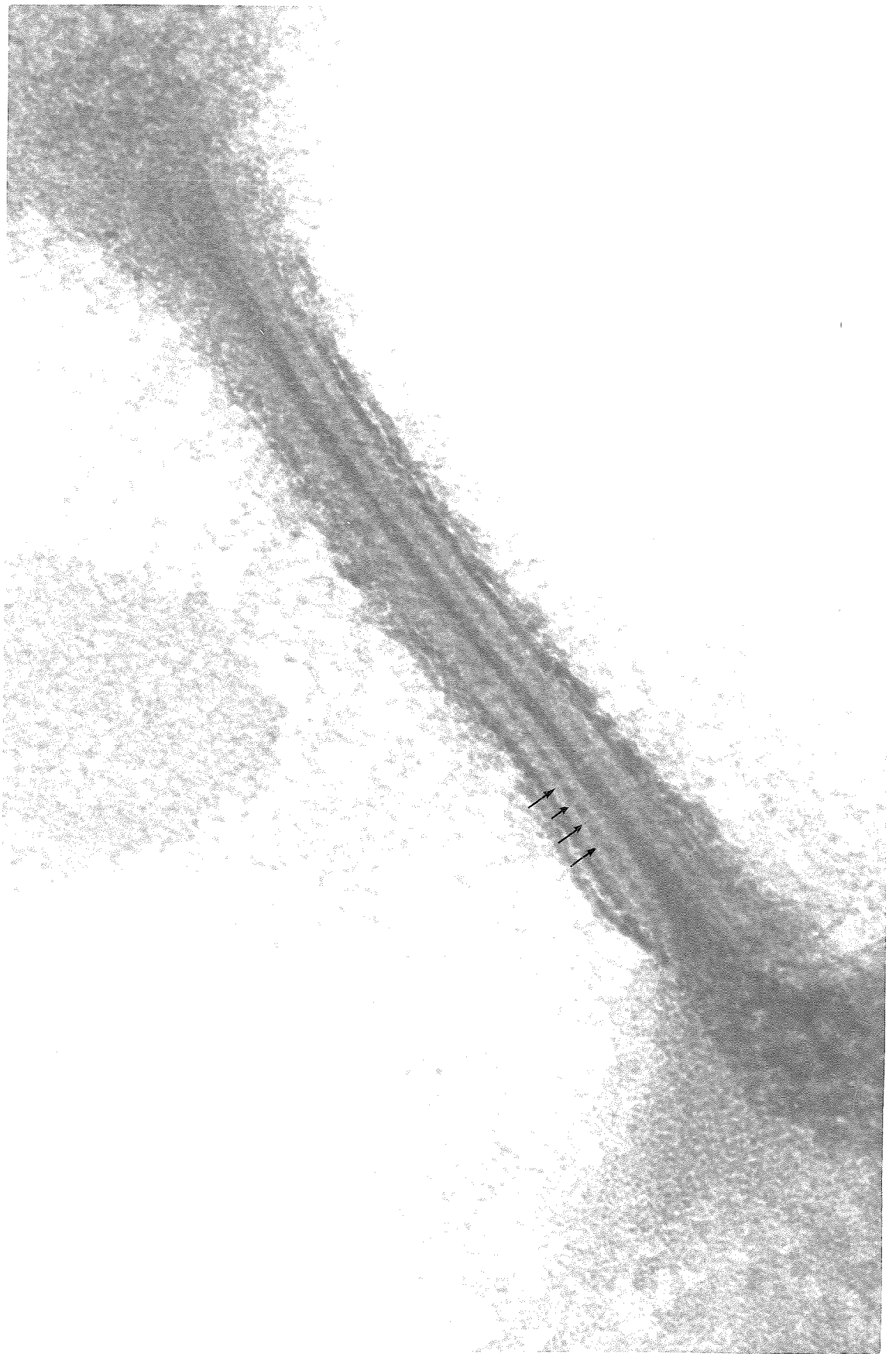


Figure 5. Electron micrograph of negatively stained isolated gap junction folded onto itself. The two adjacent profiles clearly show the channels (arrows) spanning the gap. The fold is 36 nm thick, and the channels are spaced 8.5 nm apart.



A typical preparation of isolated gap junctions has few discernable impurities. However, in some preparations a significant amount of a non-protein impurity is observed. Based on its physical behavior and morphology, the impurity is most likely glycogen. The glycogen is usually removed in the first centrifugation step, as it tends to form the hard core of the pellet that is discarded. If it was not, the glycogen could be detected in the sucrose gradient, as it made the gradient appear cloudy, and removed by extending the centrifugation of the gradient by several hours.

Isolated gap junctions produced a single major band with an estimated molecular mass of 28 kD when separated by SDS-PAGE (Figure 6, lane C). N-terminal protein sequence analysis of the 28 kD protein electro-eluted from the gel was identical with that determined previously for the major hepatic gap junction protein, confirming the identity of the protein as Cx32. The SDS-PAGE shows four minor bands at 54 kD, 26 kD and 24 kD, and some preparations also show the presence of a 21 kD component. The 54 kD is a dimer of the 28 kD protein and the 26/24 kD bands have been shown by peptide mapping to be break down products (Nicholson and Revel, 1983). This is consistent with the observation that the 28 kD band decreased and the 26/24 kD bands increased in quantity if the protease inhibitor PMSF was omitted from the preparation (Figure 6, lane B). In addition, on western blots the 26/24 kD bands react with the αCT238 antibody, but not the αCT274, described below, further corroborating the peptide mapping data. The 21 kD band was initially thought to be another breakdown product, but has now been shown to be a homologue of Cx32, Cx26 (Nicholson et al., 1987). The yield of protein from this isolation procedure was 2.8 μg per gram of starting material (liver). This is substantially more than most previously used methods, though less than half of the alkaline method of Hertzberg (Hertzberg, 1984). In our hands, however, the alkaline methods failed to produce the yields reported in the literature.

The protein composition found here for rat liver gap junctions is consistent with previous suggestions that hepatic gap junctions are composed of a single major

polypeptide. It is unlikely that there are any other integral membrane protein that is stoichiometrically represented in the connexons, since it would have to represent at least 15% (on a molar basis) of the protein isolated, relative to Cx32. No such protein is detected. Cx26 is present in gap junctions isolated from rat liver in varying amounts, from barely detectable to 15% (Nicholson et al., 1987), while in mouse liver, it composes approximately 30% of the protein (Henderson et al., 1979). It is known that Cx32 and Cx26 usually if not always co-localize in the liver by antibody labeling and immunofluorescence (Nicholson et al., 1987). However, it is not known if Cx26 can form heterologous connexons with Cx32, or if connexons of only Cx26 exist.

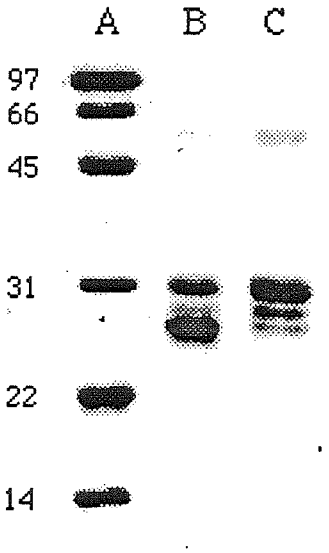


Figure 6. Isolated gap junctions separated on a 15% SDS polyacrylamide gel. A. Standards with approximate molecular masses on left. B. Gap junctions isolated from rat liver without the protease inhibitor PMSF, show the major 24 kD breakdown product generated by endogenous proteases. C. Gap junctions isolated in the presence of PMSF. The 28 kD hepatic gap junction protein, Cx32, is prominent, and minor amounts of breakdown products at 26 and 24 kD are also present. A weak band at 54 kD represents a dimer of the 28 kD protein.

No proteins peripherally associated with the gap junction have been identified, and the isolation procedure used here would likely remove any proteins not tightly associated with the membrane. If any such proteins exist, they will have to be identified by other

means. The morphological evidence of Hirokawa and Heuser (1982) would suggest that there are no associated proteins. They show by freeze fracture and deep etching that the cytoplasmic surface of the gap junction is smooth, and similar results have now been obtained by atomic force microscopy (Hoh et al., 1991; Chapter 2).

Membrane Protection Experiments

Isolated gap junction plaques were used in membrane protection experiments, in which the two membranes of the gap junction protect parts of protein from proteolytic attack (Figure 7). Several investigators have demonstrated that proteolytic digestion does not detectably change the morphological appearance of isolated gap junctions by electron microscopy (Henderson et al., 1979; Finbow et al., 1980; Makowski et al., 1982; Makowski et al., 1984). It has also been shown that the 2-3 nm gap between the gap junction membranes and the 1.5-2.0 nm pore exclude proteases, so that the only the

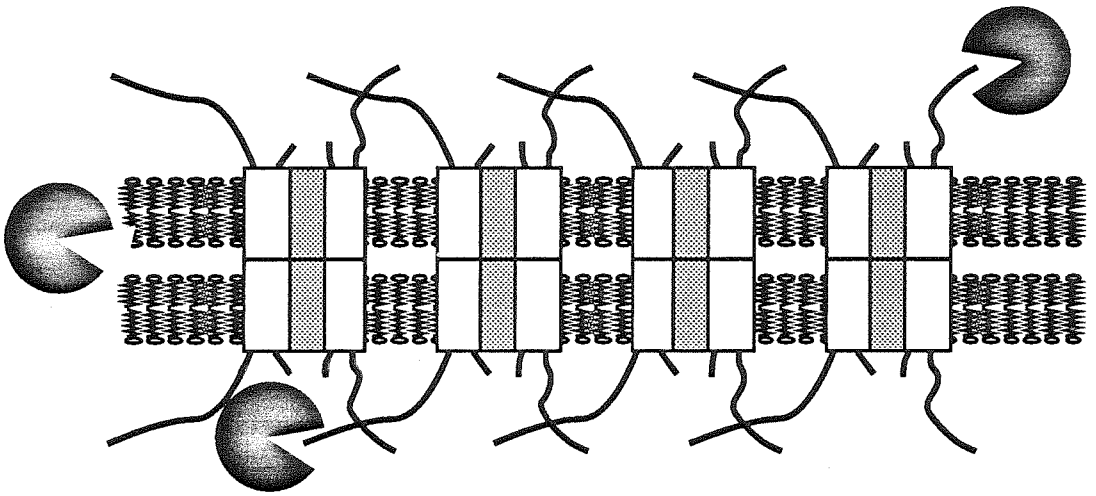


Figure 7. Cartoon showing the basis for the membrane protection experiments. The N-terminus and C-terminus of Cx32, predicted to be cytoplasmic, are shown as protruding from the membrane and therefore exposed to proteases (round figures). The membrane protects the transmembrane domains of the protein. The 2-3 nm gap between the two membranes and the 2 nm pore are too small for the protease to penetrate, protecting the extracellular domains. Dimensions are not to scale.

cytoplasmic domains of the protein are exposed (Goodenough and Revel, 1971; Nicholson, 1983). Data from these experiments suggests there are two large protected domains of about 10 kD, one of which contains the N-terminus of the protein, that are linked by a proteolytically sensitive region. The proteolytic data together with hydropathy analysis and other predictive methods have been used to produce a model for Cx32 topology (Paul, 1986; Hertzberg et al., 1988; Milks, 1988). That model predicts 4 transmembrane domains with both the N-terminus and C-terminus in the cytoplasm. The two proteolytically resistant domains represent, approximately, amino acids 1-95 and 130-220 (see proposed model in Figure 12 for sequence positions).

These two protected domains are linked by a cytoplasmic loop, which is cleaved by a number of proteases including trypsin, chymotrypsin, V8 protease (pH 4.0), and papain (Nicholson, 1983). These enzymes all cleave the loop, and presumably the cytoplasmic tail. The enzymes used here, EndoArg-C, EndoLys-C, V8-Glu (pH 8.0) and trypsin cleave proteins at arginines, lysines, glutamic acids, and arginine/lysine respectively. Commercial trypsin also usually contains chymotryptic activity that cleaves preferentially at bulky hydrophobic residues such as phenylalanine and tryptophan.

Cleavage of the cytoplasmic loop is usually ascertained by the diagnostic appearance of protein fragments in the 10-16 kD range coupled with the loss of the 24-28 kD bands in the gel. As shown in Figure 8, the cytoplasmic loop is remarkably resistant to proteolysis by all the enzymes used here, except trypsin. This is despite the fact that there are several potential cleavage sites in the predicted sequence. The fragments generated by trypsin, EndoLys-C, EndoArg-C, and V8-Glu have estimated masses of 15.9/13 kD, 24.0 kD, 24.9 kD, and 24.0 kD respectively, while the control sample migrated at 27.5 kD. The trypsin fragment sizes are probably overestimated because of non-linear migration of the lower standards, and actually represent the two 10 kD fragments previously described (Nicholson, 1981). In some gel systems these two fragments have been resolved, as is seen here (Hertzberg, 1988).

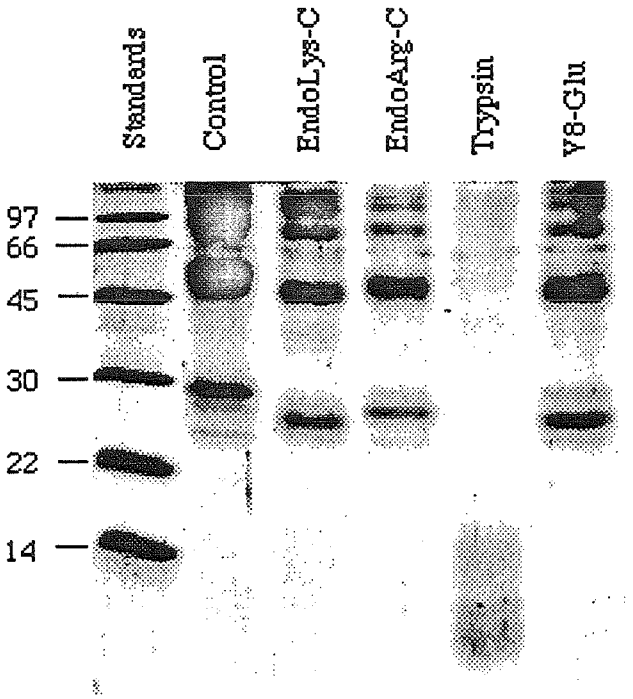


Figure 8. SDS-PAGE of page of isolated liver gap junctions proteolysed in membrane protection experiments. Molecular mass standards on the left are in kD. The control sample was incubated in PBS at 37° C for 3 hours, and migrates at 27.5 kD. The estimated sizes of the gap junction protein treated with EndoLys-C, EndoArg-C, trypsin or V8-Glu is 24.0, 24.9, 15.9/13 or 24.0 kD respectively. The bands above 45 kD are all multimeric aggregates. The 21 kD protein is poorly visible in this digitized image, but can be seen in the control sample and the EndoArg-C sample, and is present in all except the trypsin sample.

Previously used enzymes all cleaved the cytoplasmic loop, and because the resulting fragments were diffuse did not provide any information on cleavage sites in the C-terminus. The use of EndoLys-C, EndoArg-C, and V8-Glu, that do not cleave the loop, allows for a closer examination of the C-terminal tail of Cx32. The three fragments at 24.0 and 24.9 kD, generated by these enzymes are all detected by the αCT238 antibody on western blots, but none is detected by αCT274 (see Figure 12 for position of peptides in model of protein). This shows that these three enzymes all cleave the cytoplasmic tail of Cx32. It is unlikely that there is any cleavage at the N-terminus of the protein since the only

potential site is Arg^{15}, which is only 3-4 amino acids away from the predicted start of the transmembrane domain.

Based on the size of the fragments and the antibody binding, EndoArg-C must cleave the protein at Arg^{264} or Arg^{265}, EndoLys-C must cleave at Lys^{244} or Lys^{250}, and V8-Glu must cleave at Glu^{242}, Glu^{247}, or Glu^{254}. According to models from hydropathy analysis of the predicted protein sequence, the fourth transmembrane domain of Cx32 exits the membrane at approximately Glu^{216}. This means that the cleavage sites are all unexpectedly far from the membrane and that several closer sites such as Arg^{238} and Lys^{231} are not attacked. The fact that EndoArg-C does not cleave Arg^{238}, or that EndoLys-C does not cleave Lys^{231}, can be explained in three ways. First, it is possible these sites are too close to the membrane, even though a 15-22 amino acid peptide segment would protrude 5.4-7.9 nm from the membrane if fully extended. This seems unlikely since chymotrypsin cleaves the N-terminus of Cx32 within 10 amino acids of the membrane (Hertzberg et al. 1988), and for rhodopsin it has been shown that a V8 site 8 amino acids from the presumed transmembrane segment can be cleaved (Laird et al., 1987). Alternatively the model could be wrong and the sites in question could be substantially closer to the membrane or even directly protected. This also seems unlikely, and leaves the more interesting prospect that the C-terminal tail of the gap junction protein is a highly organized domain. This region is the most diverged part of the different connexins (Chapter 3), and is thought to contain specific regulatory domains. Because it is so diverged, it is unlikely that there is a common organization for the C-terminus of the connexins, and no structure for Cx32 has been proposed that explains the results obtained here. However, it has been shown by X-ray diffraction that most of the material removed by trypsinization is at least 2.8 nm from the 6 fold axis, and the C-terminus is therefore probably at the periphery of the connexon (Makowski et al., 1984).

It is also observed that the 21 kD protein (Cx26) is resistant to digestion by EndoArg-C, V8-Glu, or EndoLys-C, but not to trypsin. This is consistent with the

predicted organization of the protein based on analysis of the protein predicted from cDNA cloning (Zhang et al., 1989). Cx26 has a very small cytoplasmic tail of 10-20 amino acids that probably is not available to proteases, and has a cytoplasmic loop similar to Cx32. The fact that Cx26 is not affected, also provides further evidence that these enzymes do not have access to the N-terminus.

As described above, V8-Glu at pH 8 only cleaved the protein at a site near the C-terminus. However, when V8-Glu digestions were carried out at lower pH's, a cleavage site in the cytoplasmic loop was exposed and two fragments of 14 and 10 kD were produced. This indicates that at least one of the four glutamic acids in the loop has become available, and is cleaved by the protease. A titration of the availability of the cleavage site versus pH shows that the threshold is between pH 6 and pH 7 (Figure 9). This pH effect does not appear to arise from an effect on the activity of the enzyme. According to the Drapeau et al. (1972), V8 protease is 25-75% active over the entire pH range used and control digestion of BSA over the same pH range shows no reduction in the enzyme activity. A possible explanation for the pH sensitivity of the cytoplasmic loop is a change in the conformation of the loop, resulting in the exposure or protection of susceptible residues. Such a change could be unrelated to any natural activity of the channel, or it may be an indicator of a functionally significant pH sensitive conformational change in the gap junction protein. Physiological activity of the gap junction is known to be regulated by intracellular pH (Turin and Warner, 1977). For hepatocytes, the pKa for channel closure is 6.3 (Spray et al., 1984; Spray and Bennett, 1985), consistent with the pH sensitive effect on proteolysis described here. Unwin and Ennis (1984) have proposed a model for gap junction gating in which the entire connexon closes by twisting and tilting the protein subunits. Such change in conformation could explain how the glutamic acid residue(s) becomes exposed. Similar effects of proteases have also been observed on other proteins such as the Na^+/K^+ ATPase, for which trypsin is a probe of conformational changes related to functional activity (Joergensen and Farley, 1988).

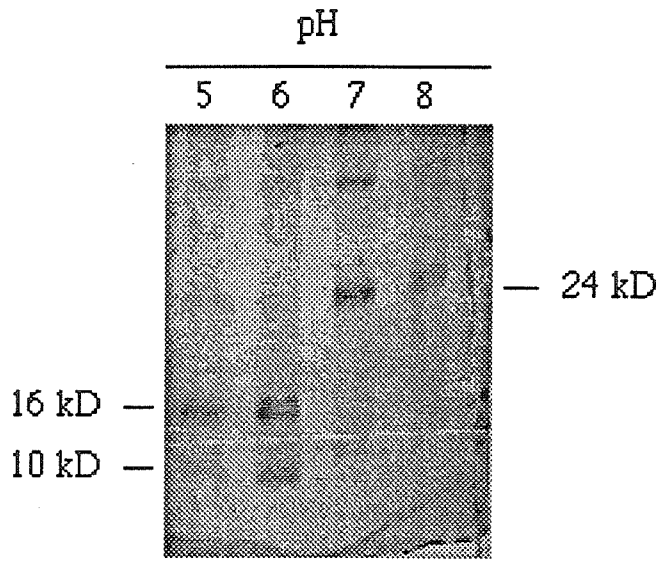


Figure 9. Apparent conformational change of the cytoplasmic loop in response to pH detected by V8-Glu protease. At higher pH's only a single band at 24 kD is seen, while at pH's less than 7 the cytoplasmic loop is cleaved to give two fragments of 16 and 10 kD. The V8-Glu is well known to be active across this pH range, and controls with BSA show no change in digestion pattern. The pH sensitivity of V8-Glu cleavage of the cytoplasmic loop correlates with pH at which the channel is gated (pKa 6.3), and may represent a conformational change in the protein related to gating.

Antibodies Against Connexin-32

Four peptides corresponding to predicted protein sequence from the Cx32 cDNA, CT222, CT238, CT274, and IL100, were synthesized and characterized for the production of antibodies (Table 1). The IL100 peptide was purified by reverse phase high pressure liquid chromatography (HPLC) and the other peptides were judged by their HPLC traces to be sufficiently pure immediately after synthesis. Antisera was raised as described and initially screened against western blots of isolated liver gap junction protein and subsequently by ELISA against a variety of substrates.

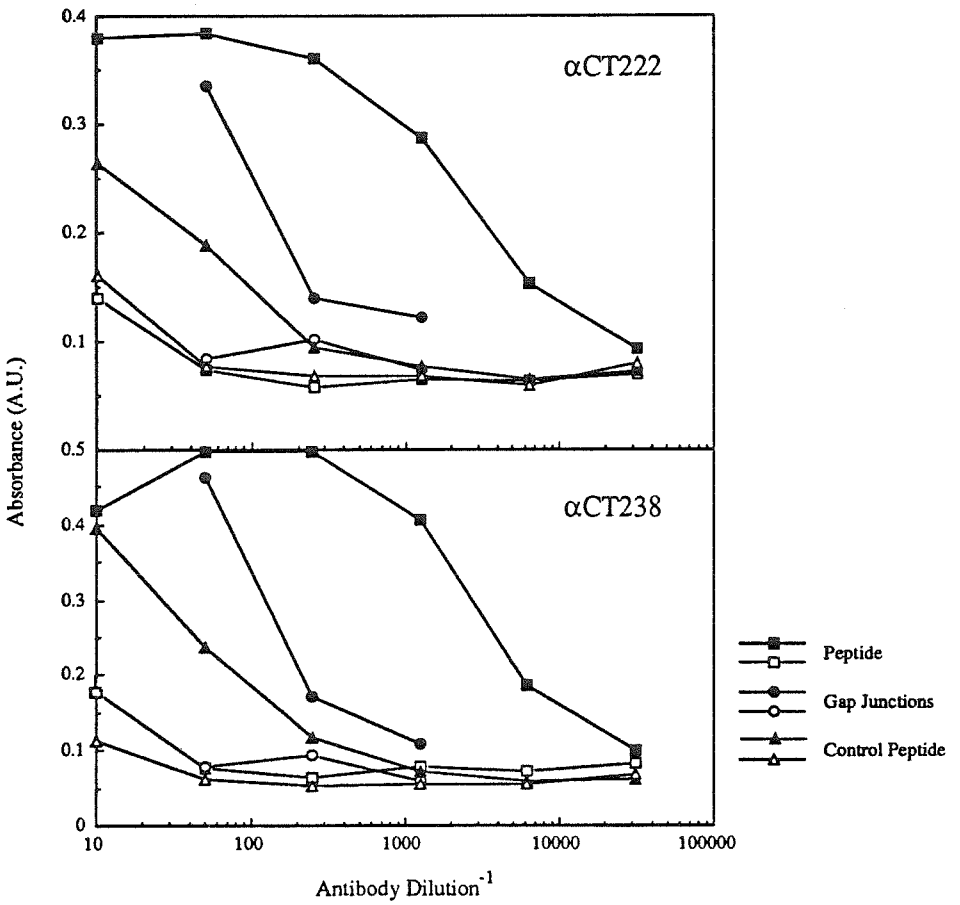
The CT222 peptide coupled to KLH did not produce antiserum that recognized the target molecule (Cx32) on western blots after several months of injections. In a further attempt to elicit an antigenic response, the CT222 peptide was coupled to OVA and used to

immunize the same rabbits. Again the antiserum showed no activity against Cx32 on western blots or by ELISA, however, it did react with the CT222 peptide by ELISA (Figure 10). It is well known that antibodies raised against specific peptides often do not recognize either the native or denatured form of the target molecule. There are several explanations for why the CT222 antiserum recognized only the synthetic peptide. The peptide may have been improperly synthesized and had the wrong sequence, though amino acid analysis provided by the manufacturer showed the proper ratios. It is also possible that the antiserum does contain activity against the native form of Cx32, but that the $Cx32^{222-231}$ sequence segment begins just 5-6 amino acids after the protein is thought to exit the membrane, and may be sterically unavailable to the antibodies. This explanation is made less likely by the observation of Milks et al. (1988) who produce antibodies against the $Cx32^{217-234}$ peptide that bind to isolated gap junction plaques. Finally the antibodies against CT222 may require the amino-teminal or C-terminal ends of the peptide, which are not present in the native or denatured protein. This peptide was not useful for any of the experiments it was intended for.

Similar to the CT222 peptide, CT238 coupled to KLH did not elicit a response from the rabbit. It was also re-coupled to OVA and used to immunize the same rabbit. This reimmunization was successful and produced antiserum that recognized both the peptide and the Cx32 protein on western blots. However, the αCT238 did not react with isolated gap junction plaques by ELISA (Figure 10). The possible reasons of this are as described above for αCT222, though again it is unlikely that the membrane would prevent binding of the antibody. Since αCT238 did not react with whole gap junctions, it could not be used directly to determine the location of the $Cx32^{238-247}$ peptide. But it was used indirectly together with proteases, as described above, to demonstrate that a large part of the C-terminus of Cx32 is protected from cleavage.

The CT274 peptide coupled to KLH produced a high titer antiserum against the peptide by ELISA and the Cx32 protein on western blots. The αCT274 antibodies reacted

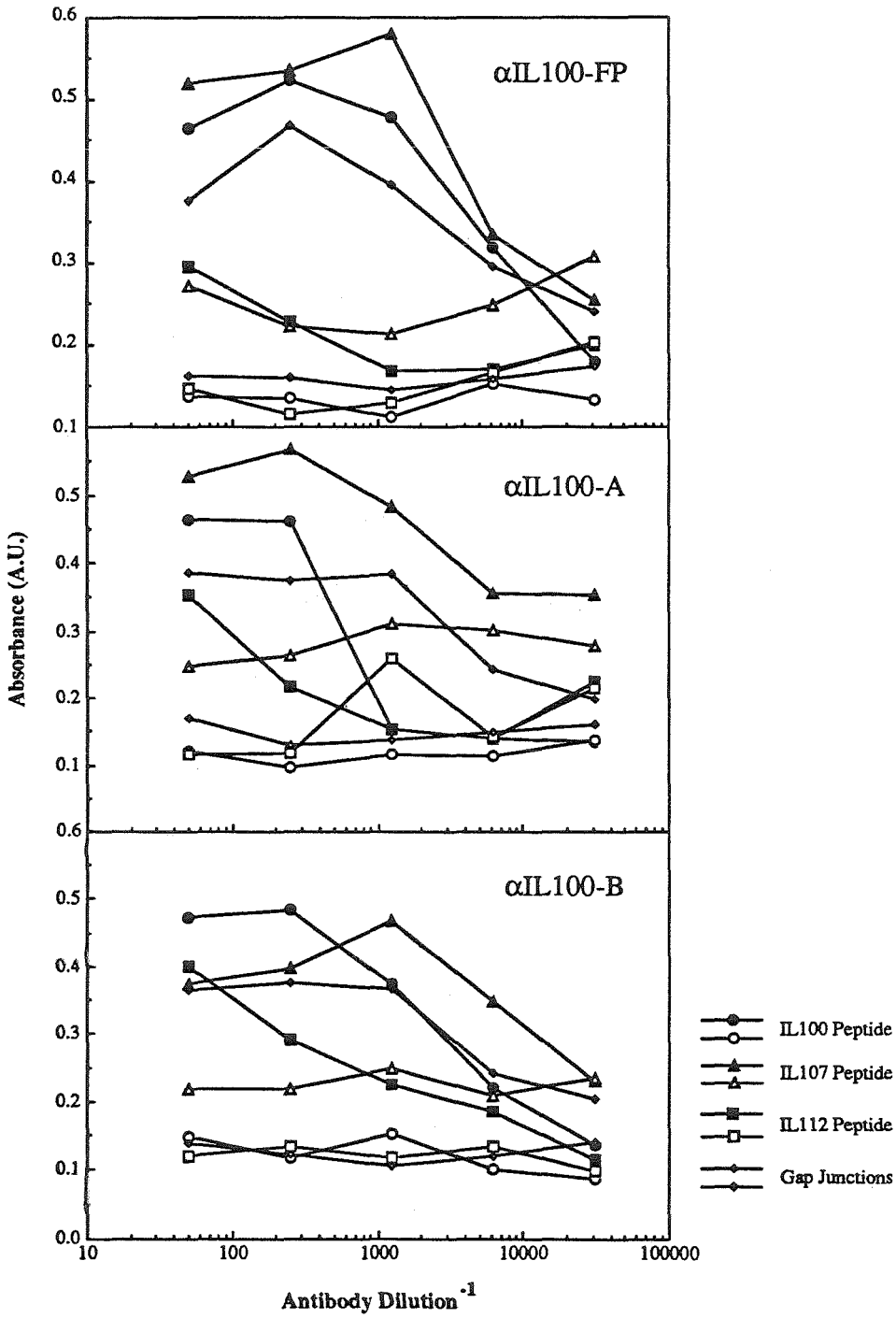
Figure 10. Reactivity of αCT222, αCT238 antibodies against different peptides and isolated gap junctions determined by ELISA. Solid symbols are for immune serum, and open symbols are for pre-immune serum.


αCT222
αCT238
Absorbance (A.U.)
Antibody Dilution^{-1}
0.4
0.3
0.2
0.1
0.5
0.4
0.3
0.2
0.1
0
10
100
1000
10000
100000
Peptide
Gap Junctions
Control Peptide

only with the intact Cx32 protein and not with the 24/26 kD breakdown products on Western blots, consistent with the C-terminal position of the peptide. Curiously αCT274 reacted well with freshly prepared rat liver gap junctions (less than 1-2 weeks old) but lost activity against gap junctions that had been stored at -20° C for longer periods of time. This was probably not due to slow proteolysis of the protein, though several amino acids could have been lost without affecting the migration of the protein on SDS-PAGE. A more likely explanation is the gradual oxidation of one or both of the cysteines (Cys^{280}, Cys^{283}) in the $Cx32^{274-283}$ peptide. Cysteines have several oxidation states of which the lowest is $-CH_2SH$, followed by $-CH_2S-SH_2C-$ (or other disulfides). The latter is easily reduced with the reducing agents such as mercaptoethanol or dithiothreitol. The next oxidation form of cysteine is $-CH_2SO_2$, which is only reduced by extensive exposure to reducing agents and finally cysteine can be irreversibly oxidized to $-CH_2SO_3$. These oxidized forms of cysteine could easily reduce or eliminate the antibody binding to the protein. The synthetic peptide was supplied and stored in a reducing reagent. As the case with both αCT222 and αCT238, αCT274 did not react with native isolated gap junctions.

The IL100 peptide alone or coupled to OVA produced antiserum that reacted with Cx32 on Western blots, and with the peptide and isolated gap junctions by ELISA (Figure 11). The αIL100 antibodies reacted well with the IL100 and IL107 peptides by ELISA, but reacted poorly or not at all with the IL112 peptide. This suggests that there is at least one epitope for αIL100 is in the $Cx32^{107-111}$ peptide GHGDP, and possibly more epitopes are in the preceding $Cx32^{100-106}$ sequence KKMLRLE. Antibodies, like proteases are too large to enter the gap or the pore of native gap junctions, and therefore serve as probes for cytoplasmic domains. The binding of the αIL100 to isolated native gap junctions therefore demonstrates that these epitopes are accessible to the antibody and are located on the cytoplasmic side. This is consistent with the proteolytic data, and in agreement with the models that have been proposed for Cx32 topology.

Figure 11. Reactivity of three different preparations of αIL100 antibodies against the nested IL peptides and isolated gap junctions determined by ELISA. The -FP antibody was produced against free peptide, and the -A and -B antibodies were produced against peptide coupled to ovalbumin. Solid symbols are for immune serum, and open symbols are for pre-immune serum.


αIL100-FP
αIL100-A
αIL100-B
Absorbance (A.U.)
Antibody Dilution -1
IL100 Peptide
IL107 Peptide
IL112 Peptide
Gap Junctions

The activity of the αCT238, αIL100 and αCT274 against the isolated gap junction protein confirm the protein sequence predicted by cDNA cloning (Paul, 1986). The CT274 antibody also shows that the entire C-terminal sequence predicted by the cDNA, less possibly a few amino acids, is present in the mature protein. The migration of the protein at 27-28 kD, despite the predicted molecular mass of 32,000 Da, is therefore likely due to the highly hydrophobic nature of the protein. Anomalous migration of hydrophobic proteins has been well documented in, for example, cytochrome c oxidase and lac permease (Isaac, 1985; Wright et al., 1986). Both those proteins appear substantially smaller by SDS-PAGE than predicted, probably because of an increased negative charge resulting from increase SDS binding to the hydrophobic residues.

Topology of Connexin-32

Hydropathy analysis of Cx32 initially revealed 4 highly hydrophobic segments (Paul, 1986). When combined with early proteolytic data, this led to a model of the protein with cytoplasmic amino and C-termini, and 4 transmembrane segments. Subsequently, a number of similar models have been for Cx32 (Milks et al., 1988; Hertzberg et al., 1988)and other connexins (Yancey et al., 1989; Zhang and Nicholson, 1989). Based on these models and the data presented in this thesis, a modified model for Cx32 topology in the membrane is presented in Figure 12.

Many aspects of this model have now been tested. The cytoplasmic localization of the N-terminus has been established with both antibodies and proteases (Zimmer et al., 1987; Hertzberg et al., 1988), and the two predicted extracellular loops have been confirmed with antibodies against specific peptides (Goodenough et al., 1988). These two loops have also been shown to be linked by at least one, and at the most three disulfide bonds (B. Nicholson, personal communication; John and Revel, in preparation).

The predicted cytoplasmic loop is highly sensitive to proteases as discussed above, and antibodies such as αIL100 bind to the cytoplasmic surface of isolated gap junction plaques. The potential V8-Glu sites in the cytoplasmic loop, Glu^{102}, Glu^{109}, Glu^{118} and

Glu^{119}, at least one of which is used at pH <7, are shown on the model. These sites are consistent with specific sites known to be available to other proteases that can cleave the protein between Leu^{108} and Lys^{124} (Hertzberg et al., 1988). Together the antibody binding and protease accessibility show that this region of the molecule is indeed cytoplasmic. However, the exact boundaries at which the cytoplasmic loop enters and exits the membrane have not been determined. This is particularly important for the boundary to TM3, which is thought to be involved in the formation of the pore.

The C-terminus is also cytoplasmic as determined by proteases, as shown here, and antibodies (Milks et al., 1988). The data presented here further suggests that parts of this region are protected from some proteases, and may be highly organized. Specific sites for EndoArg-C, EndoLys-C and V8-Glu cleavage are shown in the model.

There is little direct data available on the structure of the transmembrane segments. TM1, TM2, and TM4 are approximately 20 hydrophobic amino acids long and therefore could form transmembrane α-helices similar to other membrane proteins. TM3 has a slightly longer hydrophobic stretch. Based purely on these sequence characteristics most models predict that the gap junction protein has 4 transmembrane α-helices. The third predicted helix has a conserved amphipathic nature for most connexins represented by the sequence TYX_2SX_3K/RX_3E, which is thought to line the pore. Unwin (1989) has suggested that several ion channels including the acetylcholine receptor, MP26, GABA receptor and the gap junction have similar general organization and that they all have an amphipathic α-helix that lines their pores. However, there is no direct evidence for what residues actually form the channel of the gap junction. In fact, the current models do not account for X-ray diffraction data that has an intense β-sheet signature near the pore (Caspar et al., 1988), though this data has recently been reinterpreted as resulting from a tilted α-helix, which would give a similar X-ray pattern (Tibbitts et al., 1990).

While the 4 α-helix model is consistent with the current data, the discovery of a sequence similarity between the $Cx32^{118-147}$ and the H5 region of voltage gated K^+

channels described in Chapter 4 suggests an alternate organization. The H5 region immediately precedes the 6th transmembrane segment (S6, see Catterall, 1988 for review and nomenclature) of K^+ channels and is now thought to line the pore, possibly as a β-sheet (Hartman et al., 1991; Yellen et al., 1991; Yool and Schwarz, 1991). Based on the

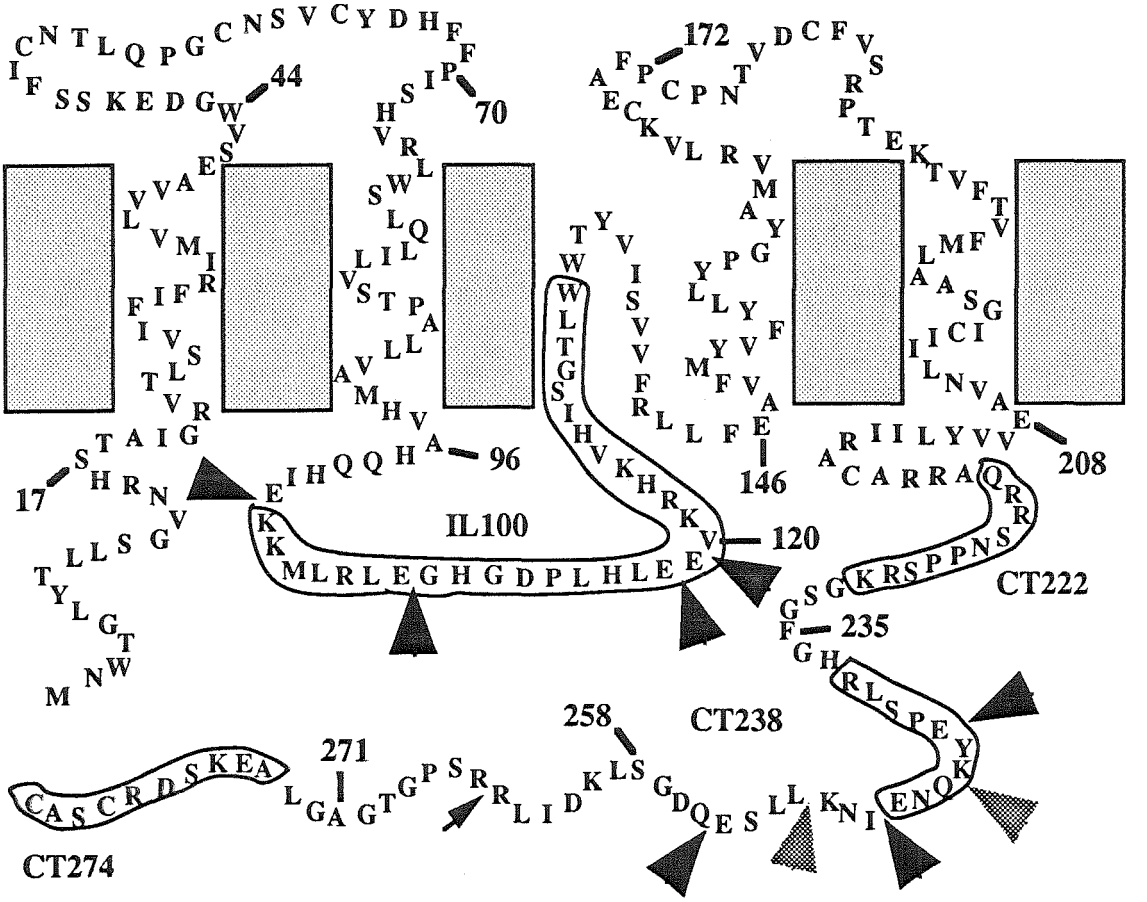


Figure 12. Model of the topological organization of Cx32 in the membrane. IL100, CT222, CT238 and CT274 peptides are boxed, and possible cleavage sites for V8-Glu (solid arrows), EndoLys-C (shaded arrows) and EndoArg-C (small arrow). The αIL100 antibody must bind within the peptide GHGPP. The cytoplasmic side of the membrane is down, and the extracellular side is on the top.

sequence similarity, the Cx32 model presented here has been modified to form a structure similar to that being proposed for K^+ channels. This results in the segment $Cx32^{147-164}$ being moved into the membrane, putting $Cx32^{118-147}$ in a position similar to H5 with

respect to the membrane (except on the intracellular side). The $Cx32^{118-147}$ does not have any striking features that suggest how it may be organized, however if it does form a β-sheet that lines the pore it would account for the β-sheet signature seen in the X-ray diffraction data. The remaining part of TM3, $Cx32^{147-164}$, would likely have to cross the membrane once, possibly as an α-helix. It is important to note that since the putative amphipathic helix and the pore lining sequence proposed here overlap, many experiments such as mutagenesis would not easily distinguish the two possibilities. The final determination of the structure and organization of the gap junction will likely have to be by high resolution physical methods.

References

Baker, T. S., et al. (1985). "Gap junction structures. VII. Analysis of connexon images obtained with cationic and anionic negative stains." J. Mol. Biol. **184**: 81-98.

Benedetti, E. L. and P. Emmelot (1965). "Electron microscopic observations on negatively stained plasma membranes isolated from rat liver." J. Cell Biol. **26**: 299-305.

Benedetti, E. L. and P. Emmelot (1968). "Hexagonal array of subunits in tight junctions separated from isolated rat liver plasma membranes." J. Cell Biol. **38**: 15-24.

Beyer, E. C., D. L. Paul and D. A. Goodenough (1987). "Connexin43: A protein from rat heart homologous to a gap junction protein from liver." J. Cell Biol. **105**(6): 2621-2629.

Brink, P. R. and M. M. Dewey (1980). "Evidence for fixed charge in the nexus." Nature: 101-102.

Caspar, D. L. D., et al. (1988). Gap junction structure. Gap Junctions. New York, Alan R. Liss. 117-133.

Catterall, W. A. (1988). "Structure and function of voltage-sensitive ion channels." Science **242**: 50-61.

Drapeau, G. R., Y. Boily and J. Houmard (1972). "Purification and properties of an extracellular protease of *Staphylococcus aureus*." J. Biol. Chem. **247**(20): 6720-6726.

Finbow, M. E., S. B. Yancey, R. G. Johnson and J. Revel (1980). "Independent lines of evidence suggesting a major gap junctional protein with a molecular weight of 26,000." Proc. Nat. Acad. Sci. U.S.A. **77**: 970-974.

Goodenough, D. A., D. L. Paul and L. Jesaitis (1988). "Topological distribution of two connexin32 antigenic sites in intact and split rodent hepatocyte gap junctions." J. Cell Biol. **107**: 1817-1824.

Goodenough, D. A. and J.-P. Revel (1971). "The permeability of isolated and *in situ* mouse hepatic gap junctions studied with enzymatic tracers." J. Cell Biol. **50**: 81-91.

Henderson, D., H. Eibl and K. Weber (1979). "Structure and biochemistry of mouse hepatic gap junctions." J. Mol. Biol. **132**: 193-218.

Hertzberg, E. L. (1984). "A detergent-independent procedure for the isolation of gap junctions from rat liver." J. Biol. Chem. **259**: 9936-9943.

Hertzberg, E. L., et al. (1988). "Topology of the Mr 27,000 liver gap junction protein." J. Biol. Chem. **263**(35): 19105-19111.

Hertzberg, E. L. and N. B. Gilula (1979). "Isolation and characterization of gap junctions from rat liver." J. Biol. Chem. **254**(6): 2138-2147.

Hertzberg, E. L., D. C. Spray and R. Cook (1987). Biochemical, immunochemical, and topological studies of gap junctions. Gap junctions. New York, Alan R. Liss. 9-28.

Hirokawa, N. and J. Heuser (1982). "The inside and outside of gap junction membranes visualized by deep-etching." Cell **30**: 395-406.

Hartman, H. A., et al. (1991). "Exchange of conduction pathways between two related K^+ channels." Science **251**: 942-944.

Hoh, J. H., et al. (1991). "Atomic force microscopy and dissection of gap junctions." Science Submitted.

Isaac, P. G., V. P. Jones and C. P. Leaver (1985). "The maize cytochrome c oxidase subunit I gene." EMBO **4**: 1617-1623.

Joergensen, P. L. and R. A. Farley (1988). "Proteolytic cleavage as a tool for studying structure and conformation of pure membrane-bound Na^+, K^+ -ATPase." Meth. Enzymol. **156**: 291-301.

Kyhse-Andersen, J. (1984). "Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose." J. Biochem. Biophys. Methods **10**: 203-209.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydrophobic character of a protein." J. Mol. Biol. **157**: 105-132.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**: 680-686.

Laird, D. W. and J.-P. Revel (1990). "Biochemical and immunochemical analysis of the arrangement of connexin43 in rat heart gap junction membranes." J. Cell Sci. **97**: 109-117.

Laird, D. W., S. Y. C. Wong and R. S. Molday (1987). Antigenic determinants on rhodopsin as defined by monoclonal antibodies. Membrane Proteins Symposium, Richmond, California, BioRad Laboratories.

Makowski, L. (1985). Structural domains in gap junctions: Implications for the control of intercellular communication. Gap Junctions. Cold Spring Harbor, Cold Spring Harbor Laboratory. 5-12.

Makowski, L., D. L. D. Caspar, D. A. Goodenough and W. C. Phillips (1982). "Gap junction structures. III. The effect of variations in the isolation procedure." Biophys. J. **37**: 189-191.

Makowski, L., D. L. D. Caspar, W. C. Philips and D. A. Goodenough (1977). "Gap junction structures. II. Analysis of the X-Ray diffraction data." J. Cell Biol. **74**: 629-645.

Makowski, L., D. L. D. Caspar, W. C. Phillips and D. A. Goodenough (1984). "Gap junction structures. V. Structural chemistry inferred from X-ray diffraction measurements on sucrose accessibility and trypsin susceptibility." J. Mol. Biol. **174**: 449-481.

Merril, C. R., D. Goldman and M. L. Van Keuren (1984). "Gel protein stains: Silver stain." Meth. Enzymol. **104**: 441-447.

Milks, L., et al. (1988). "Topology of the 32-kd liver gap junction protein determined by site-directed antibody localizations." EMBO J. 7(10): 2967-2975.

Nicholson, B. J. (1983). Biochemistry and diversity of the gap junction protein: A study of liver, heart and lens. California Institute of Technology.

Nicholson, B. J., et al. (1981). "Rat liver gap junction protein: properties and partial sequence." Proc. Nat. Acad. Sci. USA 78(12): 7594-7598.

Nicholson, B. J., et al. (1985). "The Mr 28,000 gap junction proteins from rat heart and liver are different but related." J. Biol. Chem. **260**(11): 6514-6517.

Nicholson, B., et al. (1987). "Two homologous protein components of hepatic gap junctions." Nature **329**(6141): 732-734.

Nicholson, B. J. and J.-P. Revel (1983). "Gap junctions in liver: Isolation, morphological analysis, and quantitation." Meth. Enzymol. **98**: 519-537.

Paul, D. L. (1986). "Molecular cloning of cDNA for rat liver gap junction protein." J. Cell Biol. **103**: 123-134.

Revel, J.-P. and M. Karnovsky (1967). "Hexagonal array of subunits in intercellular junctions of the mouse heart and liver." J. Cell Biol. **33**: C7-C12.

Spray, D. C., et al. (1984). "Physiological and pharmacological properties of gap junctions between dissociated pairs of rat hepatocytes." J. Cell Biol. **99**: 344a.

Spray, D. C. and M. V. L. Bennett (1985). "Physiology and pharmacology of gap junctions." Ann. Rev. Physiol. **47**: 281-303.

Tager, H. S. (1976). "Coupling of peptide to albumin with difluorodinitrobenzene." Anal. Biochem. **71**: 367-375.

Tibbitts, T. T., D. L. D. Caspar, W. C. Philips and D. A. Goodenough (1990). "Diffraction diagnosis of protein folding in gap junction connexons." Biophys. J. **57**: 1025-1036.

Turin, L. and A. Warner (1977). "Carbon dioxide reversibly abolishes ionic communication between cells of early amphibian embryo." Nature **270**: 56-57.

Unwin, N. (1986). "Is there a common design for membrane channels?" Nature **323**: 12-13.

Unwin, P. N. T. and P. D. Ennis (1984). "Two configurations of a channel-forming membrane protein." Nature **307**: 609-613.

Unwin, P. N. T. and G. Zampighi (1980). "Structure of the junction between communicating cells." Nature **283**: 545-549.

Wright, J. K., R. Seckler and P. Overath (1986). "Molecular aspects of sugar:ion transport." Ann. Rev. Biochem. **55**: 225-248.

Yancey, S. B., et al. (1989). "The 43-kD polypeptide of heart gap junctions: immunolocalization (I), Topology (II), and functional domains (III)." J. Cell Biol. **108**: 2241-2254.

Yellen, G., M. E. Jurman, T. Abramson and R. MacKinnon (1991). "Mutations affecting internal TEA blockade identify the probable pore-forming region of a K^+ channel." Science **251**: 939-942.

Yool, A. J. and T. L. Schwarz (1991). "Alteration of ionic selectivity of a K^+ channel by mutation of the H5 region." Nature **349**: 700-704.

Zhang, J.-T. and B. J. Nicholson (1989). "Sequence and tissue distribution of a second protein of hepatic gap junctions, CX26, As Deduced from its cDNA." J. Cell Biol. **109**(6, pt2): 3391-3401.

Zimmer, D., C. R. Green, W. H. Evans and N. B. Gilula (1987). "Topological analysis of the major protein in isolated intact rat liver gap junctions and gap junction-derived single membrane structures." J. Biol. Chem. **262**: 7751-7763.

Chapter 2

Atomic Force Microscopy and Dissection of Gap Junctions

Gap junctions consist of two apposed plasma membranes that contain an array of cell to cell channels (1). These channels form aqueous pores that allow the free passage of molecules less than 1000 Da in vertebrates and provide a low resistance electrical pathway between cells (2). Proposed biological functions for gap junctions include regulation of growth, transmission of developmental signals, coordination of smooth muscle contraction, synchronization of myocardial contractions, and maintainence of metabolic homeostasis (3).

The structure of the gap junction has been studied extensively by physical and biochemical methods. By electron microscopy it was first described as a close membrane apposition (4) with a quasi crystalline array of particles (5) and subsequently a gap between the membranes was defined (1). Models of the gap junction have been constructed using data from X-ray diffraction, electron microscopy, and Fourier three dimensional reconstruction techniques (6, 7). The current view, based on these approaches, is that the gap junction is composed of two apposed membranes with a 2-3 nm gap between them, and a closely packed array of cell-cell channels. The most regular samples show hexagonal packing with a lattice constant of 8-10 nm, but the degree of order varies depending on the preparation. Each channel is composed of two connexons, one from each membrane, aligned head to head across the gap. The connexon is shaped roughly as a cylinder 7.5 nm tall and 7 nm in diameter, with a 1.5-2.0 nm pore through the center. Each connexon exhibits six fold symmetry and is thought to consist of six identical or homologous protein subunits. The major protein components, connexins, from several tissues have been isolated (8, 9), and shown to be members of a gene family (10). The organization of the connexins in the membrane has been examined in some detail using proteases (11, 12) and antibodies against specific peptides (13, 14). However, the organization of the polypeptides within each connexon is not known and the nature of the interactions between connexons across the gap are poorly understood.

A new class of instruments, refered to as scanning probe microscopes, offers the possibility of high resolution imaging in a physiologic environment. The Atomic Force Microscope (AFM) is one of the most promising of these new instruments for the imaging of biological specimens (15). Details of its operation have been presented elsewhere (16, 17). Briefly, the sample is mounted on a piezoelectric xyz translator and scanned in the x and y directions below a tip attached to a microcantilever. Deflections are measured by reflecting a laser beam off the cantilever onto a two segment photodiode. A feedback circuit maintains the cantilever deflection, keeping the force exerted on the sample constant, by raising or lowering the sample as it is scanned. An image of the surface is then reconstructed by plotting the height (z) versus the lateral position (x and y) of the sample (Fig. 1). The AFM has been used to image a number of biological specimens (16, 18, 19). The purple membrane from *Halobacterium halobium,* in particular, has been studied in some detail by Hansma and his coworkers and they have reported imaging of bacteriorhodopsin at molecular resolution (19). In the work described here, we have used an AFM equipped with a fluid cell to probe the structure and organization of the gap junction and its channels in phosphate buffered saline (PBS) (20), and to manipulate the bilayers.

Imaging of Gap Junction Plaques. Gap junctions from rat liver are isolated as membrane pairs, often refered to as plaques, with densely packed cell to cell channels. We have used a routinely applied method, the key feature of which is the solubilization of most non-junctional membranes by sarkosyl followed by centrifugation on a sucrose step gradient (8, 21). These gap junctions are highly purified based on electron microscopy and polyacrylamide gel electrophoresis. A commercially available NanoScope II equipped with an 18 μm x 18 μm AFM stage and a fluid cell was used for imaging (22). Images were acquired using V shaped 100 μm cantilevers with a spring constant of 0.38 N/m and pyramidal tips (22, 23). The cantilevers were coated on their top surface with a thin layer of gold, and on the bottom surface, including the tip, with a thin layer of chromium.

Isolated gap junction membranes in PBS were adsorbed to glass cover slips (24). To control and minimize the amount of force the sample was subjected to, the tip was routinely engaged onto a 49 x 49 nm scan area in the center of the field. The force was then adjusted to approximately 1 nanoNewton (nN), and the scan area was gradually enlarged until it included a sample to be imaged. The force during imaging was always monitored closely and assessed within 15-30 seconds, of acquiring an image.

Gap junctions imaged under PBS appear similar in general shape and distribution to ones seen by electron microscopy (Fig. 2A-B). They are flat structures 0.5 - 1 μm large with irregular edges and are 14.4 nm thick (Fig. 2C). Occasionally a step of 6-7 nm is seen at the edge of a plaque. This represents either a single membrane from a gap junctional plaque or a piece of non-junctional membrane attached to the gap junction. The surface of the gap junction has height variations of about a nanometer, but sometimes bumps of 50-100 nm in width and several nanometers in height are seen. At high magnification, the surface has no discernible regular features and is smoothly undulating. High magnification images of the glass itself are smooth with z variation of only 1-2 nm and no detectable regular pattern (Fig. 2D).

Glutaraldehyde fixation of hepatic gap junctions does not result in any discernible changes in morphology. Interestingly all samples that were glutaraldehyde fixed were more easily scrapped off the glass substrate. This may be explained by the reaction of glutaraldehyde with the amino groups of molecules such as phospholipids and proteins. Reaction of these groups with glutaraldehyde would reduce the total positive charge of the gap junction membrane and thereby reduce the strength of the electrostatic interactions between the gap junction and the negatively charged glass. It has been reported previously that the adsorption of purple membranes to mica requires positively charged membranes (19).

Force Dissection of Gap Junction Membranes. We examined the effect of force on the structure and appearance of the gap junction. Samples of untreated plaques

were imaged in PBS in an increasing force series from less that 1 nN to approximately 15 nN. After locating a gap junction at a low force, the plaque appeared stable for several passes of the tip. Remarkably, upon increasing the force, the top membrane of the gap junction began to become distorted, and after several passes of the tip it would be completely removed exposing a new surface that we believe is the extracellular surface of the bottom membrane of the gap junction (Fig. 3). During this "force dissection," the thickness of the native plaques changed from 14.4 nm for the double bilayer, to 6.4 nm for the single bilayer (compare Fig. 3A with 3D, or 3E with 3H). The remaining half of the junction plaque could only be removed at extremely high forces suggesting that the interactions between the glass and the gap junction are significantly stronger than between the two membranes of the gap junction. Force dissection was also carried out on both trypsinized and glutaraldehyde fixed samples.

Splitting of gap junctions has previously been accomplished in a variety of ways (13, 25), and such junctions have been used to demonstrate the extracellular localization of specific segments of the connexins by labeling with antibodies against specific peptides (13, 14). However, the conditions for splitting are somewhat harsh and it is often difficult morphologically to know which side is which in the resulting single membranes. Force dissection has the advantage that it gives access to the extracellular domains in a controlled fashion and these domains are then available for immediate experimental manipulation.

The observation that fixed gap junctions can be force dissected at forces similar to unfixed material, suggests that glutaraldehyde does not cause the formation of cross-links across the gap. The phospholipid head groups in the plane of the membrane are separated by a 2-3 nm gap and are therefore unlikely to be cross-linked. However, the extracellular domains in connexin-32, the major hepatic gap junction protein, span the gap to interact with the corresponding domains in the opposing connexon. These extracellular domains have several residues that could be cross-linked with glutaraldehyde. Since there is apparently little or no cross-linking across the gap, we conclude that these residues are not

near each other in the connexon-connexon interactions or that they are not accessible to the glutaraldehyde. We are currently attempting to use other cross-linking reagents and "force dissection" as an assay to learn more about which groups are involved in the contacts between connexons.

The "force dissection" always removes membrane in the forward scan direction suggesting that there is a bias in the lateral force applied to the sample. While the image normally generated by the NanoScope II only includes forward scans of the AFM tip, the tip remains in contact with the sample as it scans in both directions. The nature of the interactions between the tip and the sample which result in "force dissection" are not known. However, the cantilever is mounted at a 12° angle with respect to the sample, which provides an obvious asymmetry in the forces exerted on the sample in the two scan directions, and may be the basis for the direction of "force dissection" (Fig. 1).

Hemi-Channel Imaging. We have looked for the cell-cell channels on the surface of gap junctions under a variety of conditions. In intact gap junctions there is no substructure and the surface appears smooth as it does by electron microscopy of deep etched samples (26). However, a distinct hexagonal pattern is revealed on the surface exposed by "force dissection" of gap junction plaques that have been both trypsinized and glutaraldehyde fixed (Fig. 4A). This hexagonal array is remarkably similar to the pattern of cell-cell channels seen in negative stained isolated gap junctions in the electron microscope (7). We believe that it represents an image of connexons protruding into the extracellular space. The connexons appear 4-6 nm in diameter, which is somewhat less than suggested by current models (6, 7). Some connexons have a small depression in the center which could represent part of the channel pore.

Sometimes the surface of gap junction hemi-channels appeared to be in rows (Fig. 4B). This is also evident in the Fourier transform where the intensity of the six symmetrical spots sometimes varied suggesting a more defined order in one direction. These rows were

not an artifact of the scanning, since the rows moved relative to the scan direction when it was altered up to 35 degrees.

In images of "force dissected" gap junctions the connexons protrude 0.4-0.5 nm from the surface of the plaque (Fig. 4C), not enough to account for the 2-3 nm gap seen by electron microscopy. This may result from the protein obstructing the scanning tip, preventing it from reaching the membrane and thus making the height of the exposed part of the connexons appear too small, or it may be due to the deformation of the protein by the tip at the forces used. Cross-links introduced by glutaraldehyde should stabilize the protein in the connexon, perhaps making it rigid enough to deflect the cantilever and explaining why the connexons can only be visualized after fixation. The role of trypsinization of the gap junction in revealing the hemi-channels is not known. It may alter the conformation of the connexons, but more likely influences the interaction between the plaque and the glass substrate. Trypsin is known to remove much of the protein on the cytoplasmic surface of the plaques without disrupting their structure.

A two dimensional Fourier transform of the AFM image produces a distinct six fold symmetry in the frequency domain (Fig. 4D). The lattice constant for the hexagonal array determined from the Fourier transform is 9.1 nm at a lateral resolution of 1-2 nm (27). The six fold pattern is clearly visible to one order and second order spots are often seen. This suggests that there is short range disorder in the hexagonal array, which is also seen in the filtered image (Fig. 4E). Previous data suggests that the degree of order in the hexagonal arrays of cell-cell channels varies depending on the preparation, and there have been suggestions that the variation in packing has physiological significance (28).

The images we show here represent the first visualization of the extracellular domains of the gap junction, and are to our knowledge the first images of an ion channel acquired by atomic force microscopy. There are currently efforts under way directed toward improving the cantilevers, reducing their spring constants, and changing the conditions under which imaging is carried out. While imaging of untreated samples is

obviously a goal, our results show that fixation can be an important tool in achieving high resolution AFM images of biological specimens.

Thickness Measurements of Gap junctions and the Effect of Force. While the "force dissection" somewhat complicated the measurement of the thickness of gap junctions and the effect of force, we performed measurements on gap junction plaques in a series of increasing and decreasing forces. Because of the "force dissection", the decreasing series would often include only the 7 nm profiles. These measurements were carried out for isolated native hepatic gap junctions, trypsinized gap junctions, and glutaraldehyde fixed gap junctions (Fig. 5). We find no statistically significant effect of force on the thickness of any of the gap junction samples, and there is no significant effect of the treatments. The thickness distribution versus force shows the full and half thicknesses of the plaques, and since there is no observable effect of force, all measurements were combined. The means of thicknesses (+/- S. D.) are of 14.4 (+/- 0.9) nm and 6.4 (+/- 0.8) nm for native plaques, 15.5 (+/-1.3) nm and 7.1 (+/-0.7) nm for trypsinized plaques, and 14.8 (+/-1.0) nm and 7.1 (+/-0.3) nm for glutaraldehyde fixed plaques. The thicknesses of native gap junctions we have measured here are in general agreement with the thickness of 15-18 nm determined by X-ray diffraction and electron microscopy (29). However, it has been reported that the apparent thickness of purple membranes depends on the substrate used (19). Until the basis for this substrate dependence is understood, the significance of the absolute measurements reported here will not be known.

Trypsinization is known to remove 60-70 amino acids from the cytoplasmic surface of connexin-32, and cleave the protein into two 10 kDa fragments that contain the four transmembrane domains and the two extracellular loops (11). AFM images of trypsinized and untrypsinized gap junctions appear identical and there is no apparent change in shape or thickness. A priori one might have expected that the thickness would have been affected by the removal of protein from the surface, this is not so. In contrast, native cardiac gap

junctions that have a larger cytoplasmic protein domain are significantly thicker than hepatic gap junctions when imaged with the AFM, but are reduced to 14-15 nm upon trypsinization (30). This would suggest that there is not enough protein mass removed from the hepatic gap junction by trypsinization to affect the thickness, or that the protein removed was closely associated with the membrane and thus did not contribute significantly to the overall thickness of the gap junction.

The AFM has the ability to image at high resolution under near physiological conditions and to provide real time experimental access to the sample. We have here been able to physically manipulate the gap junction membrane, and thereby provide a unique view of the extracellular regions of the channels. We hope this will provide encouragement to use the AFM to address other biological problems, particularly structural features of membranes and membrane proteins such as ion channels and receptors.

Figure 1. Schematic of the atomic force microscope fluid cell, cantilever and a gap junction plaque. The glass fluid cell filled with PBS is sealed against the cover slip with an o-ring and encloses the sample and cantilever. The cover slip is mounted on a piezoelectric translator which scans the sample in a raster pattern (x and y) below the tip. The translator maintains the force constant by raising or lowering (z) the sample in response to deflections in the cantilever monitored by a laser reflected off the cantilever onto a segmented photodiode. This drawing is not to scale.

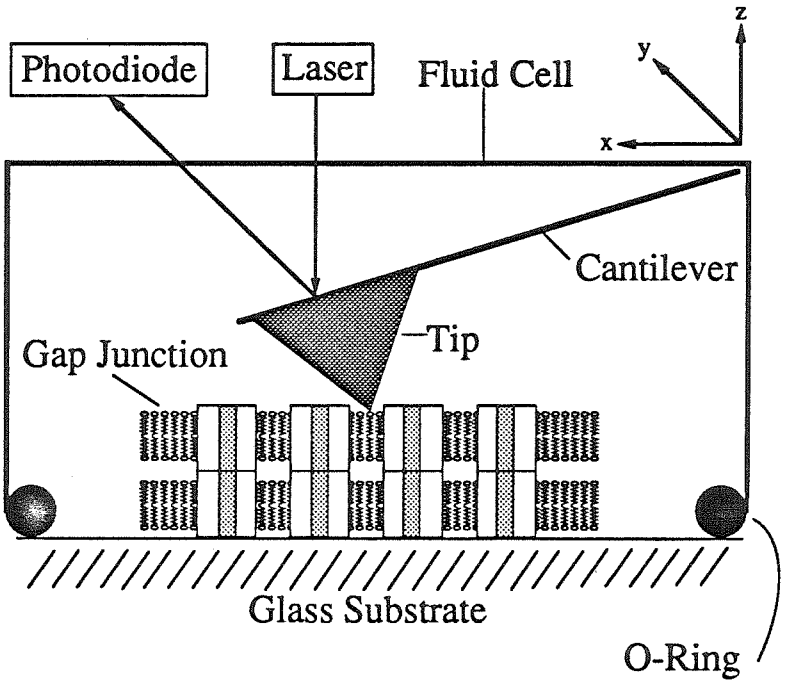

Photodiode
Laser
Fluid Cell
z
y
x
Cantilever
Tip
Gap Junction
Glass Substrate
O-Ring

Figure 2. Low magnification images of isolated rat liver gap junctions negatively stained with phosphotungstic acid taken with the electron microscope (A), and of isolated gap junctions in PBS taken with the AFM (B). The high magnification insert (A) shows the connexons in the gap junction plaques as seen by electron microscopy. The gap junctions in the AFM appear similar in shape and distribution to ones in the electron microscope. Also shown are a single gap junction plaque (C) at an intermediate magnification (1.5 μm x 1.5 μm), and a surface view of a high magnification image (147 nm x 147 nm) of the glass cover slip (D). The glass is extremely smooth, but does have a ripple that is 0.5-1.0 nm high in the scan direction. The gap junction plaque has the typical shape, is about 15 nm thick, and shows the bumps that are sometimes present on the gap junction surface. Gap junctions containing 200-300 ng connexin-32 protein were diluted into 50-75 ul PBS on a glass cover slip (24) and adsorbed for 10-20 minutes and subsequently washed twice in 20 ml PBS for 5 minutes. Samples were stored for up to 8 hours in PBS before being imaged. Temperature near AFM was routinely 21-25 °C.

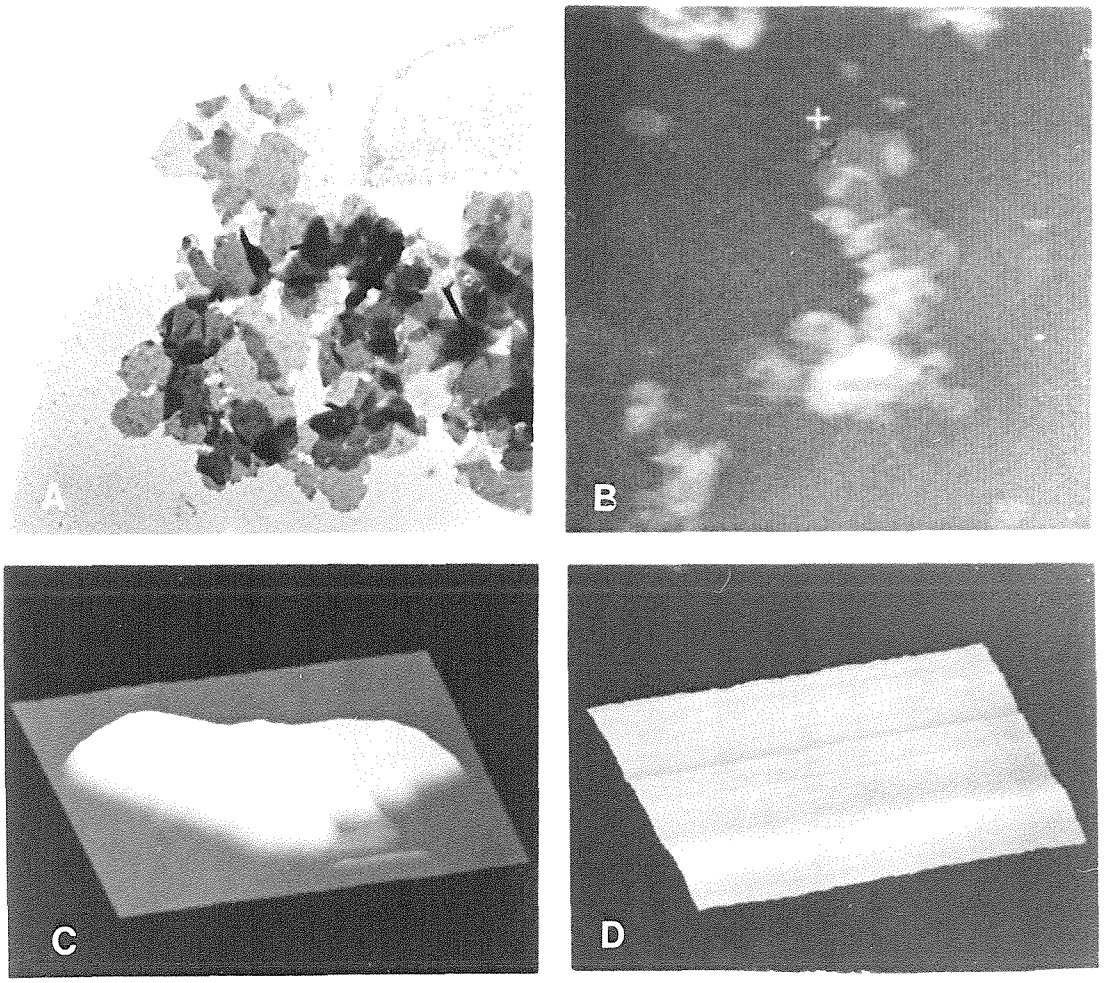

A
B
C
D

Figure 3. Force dissection of native gap junction membranes in PBS adsorbed to glass using the AFM. The top row of images shows a surface view of a gap junction plaque subjected to a sequential series of scans at increasing forces of 0.8 nN (A), 3.6 nN (B), 6.1 nN (C), and 9.6 nN (D). The field size is 1.5 μm x1.5 μm and the plaque image is about 15 nm thick. A piece of single membrane is seen attached to one edge of the plaque. At low force the shape of the plaque is stable for several scans, but as the force is increased the top membrane begins to smear from right to left and is eventually removed completely leaving the extracellular side of the 7 nm thick bottom membrane exposed. The second series of images show a top view of another gap junction plaque as it is subjected to sequential scans at increasing forces of 0.8 nN (E), 3.1 nN (F), 10.1 nN (G), and a repeat scan at 10.1 nN (H). The field size is 1.5 μm x 1.5 μm and the line cut marked on each image is shown on the right. These line cuts show clearly how the top membrane is removed at higher forces, reducing the thickness of the structure to half the original. The thickness of the structure in the first image (E) is about 15 nm and about 7 nm in the last (H). The remaining membrane could only be removed at extremely high forces, suggesting that the interactions between the gap junction plaque and the glass are stronger than the interactions between the two membranes.

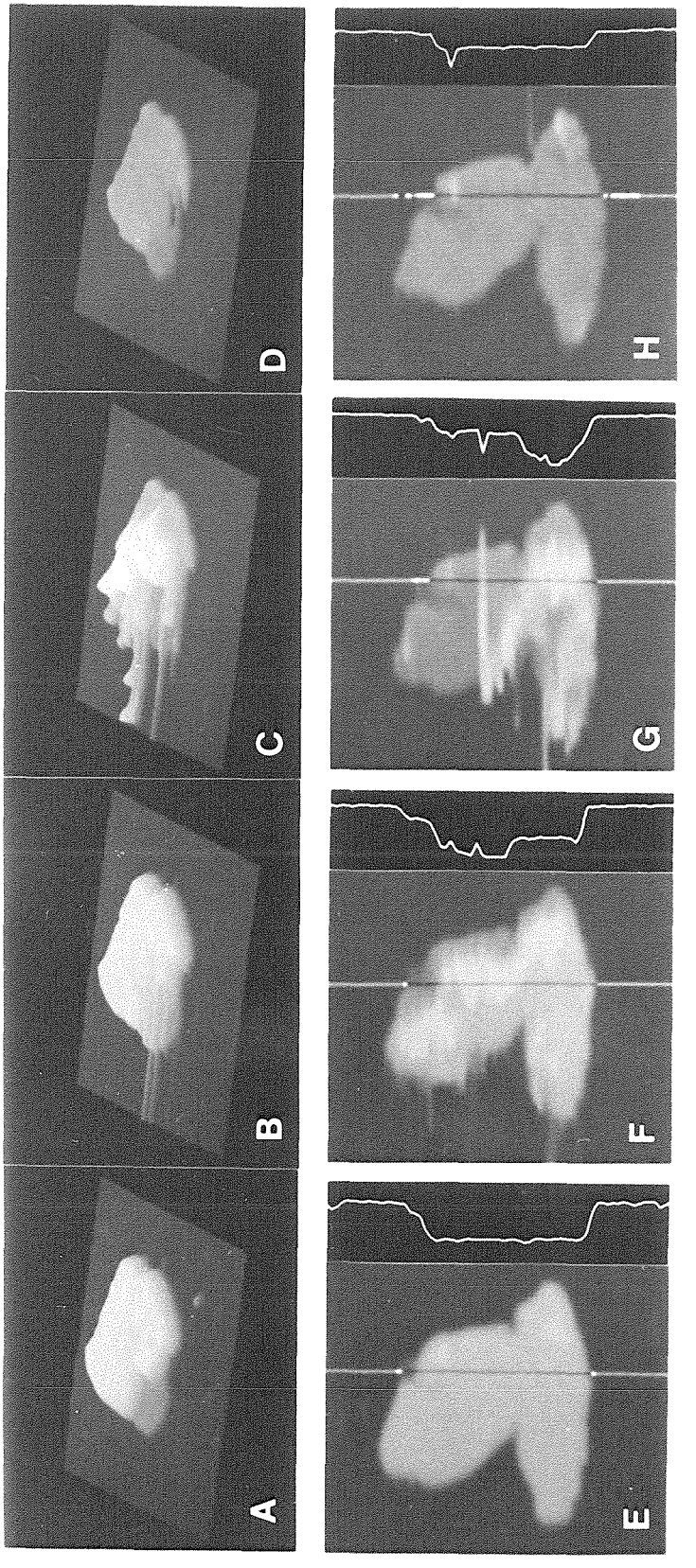

A
B
C
D
E
F
G
H

Figure 4. Imaging of extracellular surface and connexons of a gap junction plaque that has been both trypsinized and glutaraldehyde fixed (31). High magnification (120 nm wide) images show the hexagonal array of connexons (**A**). The connexons appear to be approximately 4-6 nm in diameter and sometimes have a depression in the center, indicated by arrows, which may represent part of the pore of the channel. In some samples a row like appearance was apparent (**B**). The surface view shows the connexons protruding approximately 0.4-0.5 nm (**C**). Images **A-C** were all plane fitted, but not processed further. Fourier transforms (**D**) reveal a distinct hexagonal array with a 9.1 nm lattice constant (27). The hexagonal pattern is only visible to one order in most samples, though occasionally second order spots are seen. A top view of the filtered inverse Fourier transform (91 nm x 91 nm) clearly shows the hexagonal packing (**E**). The surface view of a filtered inverse Fourier transform (**F**) also shows the hexagonal pattern but reveals some irregularity in the packing.

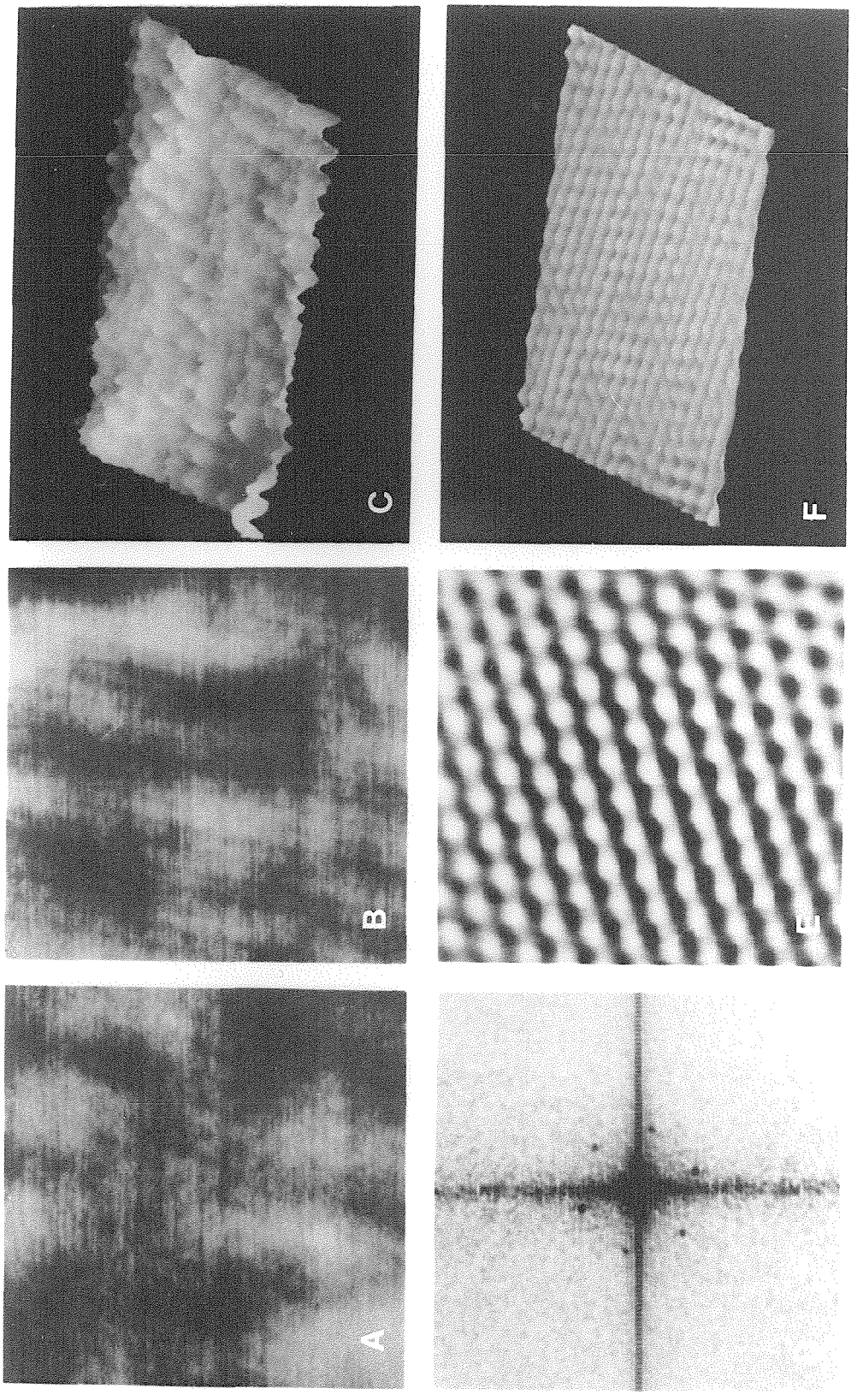
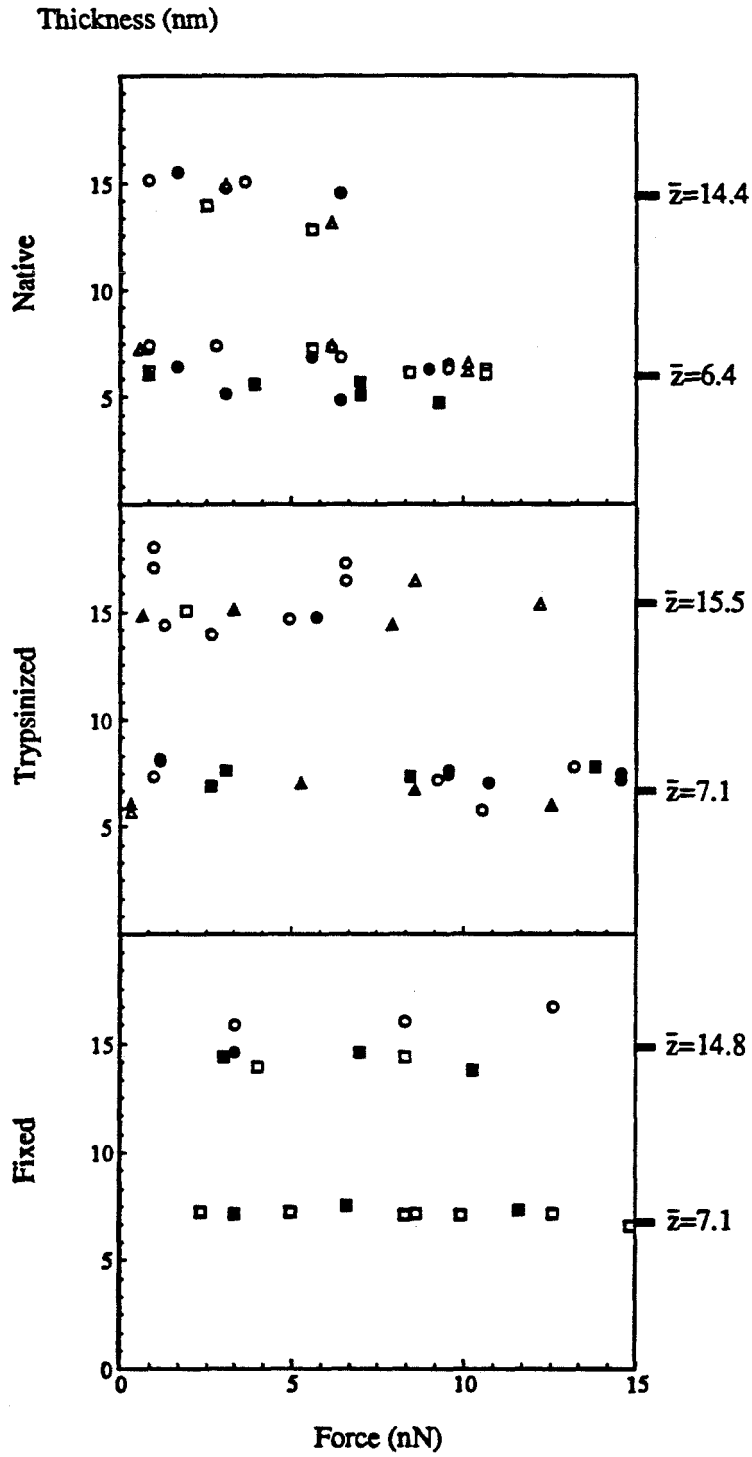

C
F
B
E
A

Figure 5. Force versus thickness distributions for native, glutaraldehyde fixed, and trypsinized gap junctions (31). Thicknesses were measured from line cuts of each plaque taken at 0, 90, and 45 degrees from the scan direction. Three measurements from each cut were taken. There was no significant difference in the thickness measured at the different angles, so all measurements from a single plaque at a given force were averaged (z). Each symbol in a given graph represents a different plaque.


Thickness (nm)
Native
Trypsinized
Fixed
z̄=14.4
z̄=6.4
z̄=15.5
z̄=7.1
z̄=14.8
z̄=7.1
0
5
10
15
Force (nN)

References and Notes

1. J.-P. Revel, M. Karnovsky, *J. Cell Biol.* **33**, C7 (1967).

2. J. Flagg-Newton, I. Simpson, W. R. Loewenstein, *Science* **205**, 404 (1979).

3. E. L. Hertzberg, R. G. Johnson, *Modern Cell Biology: Gap Junctions* (Alan R. Liss, Inc., New York, 1988).

4. M. M. Dewey, L. Barr, *Science* **137**, 670 (1962).

5. J. D. Robertson, *J. Cell Biol.* **19**, 201 (1963).

6. L. Makowski, D. L. D. Caspar, W. C. Philips, D. A. Goodenough, *J. Cell Biol.* **74**, 629 (1977); L. Makowski, in *Gap Junctions* M. V. L. Bennett, D. C. Spray, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1985), pp. 5-12.

7. P. N. T. Unwin, G. Zampighi, *Nature* **283**, 545 (1980).

8. M. E. Finbow, S. B. Yancey, R. G. Johnson, J. Revel, *Proc. Nat. Acad. Sci. U.S.A.* **77**, 970-974 (1980); D. Henderson, H. Eibl, K. Weber, *J. Mol. Biol.* **132**, 193 (1979).

9. R. W. Kensler, D. A. Goodenough, *J. Cell Biol.* **86**, 755 (1980); C. K. Manjunath, G. E. Goings, E. Page, *Am. J. Physiol.* **246**, H865 (1984).

10. B. J. Nicholson, D. B. Gros, S. B. Kent, L. E. Hood, J.-P. Revel, *J. Biol. Chem.* **260**, 6514 (1985); D. L. Paul, *J. Cell Biol.* **103**, 123 (1986); E. C. Beyer, D. L. Paul, D. A. Goodenough, *ibid.*, **105**, 2621 (1987); J.-T. Zhang, B. J. Nicholson, *ibid.*, **109**, 3391 (1989); J. H. Hoh, S. A. John, J.-P. Revel, *J. Biol. Chem.* **In Press**, (1991)

11. B. J. Nicholson, M. W. Hunkapiller, L. B. Grim, L. E. Hood, J.-P. Revel, *Proc. Nat. Acad. Sci. USA* **78**, 7594 (1981).

12. S. B. Yancey, S. A. John, R. Lal, B. J. Austin, J.-P. Revel, *J. Cell Biol.* **108**, 2241 (1989); L. Milks, N. M. Kumar, R. Houghton, N. Unwin, N. B. Gilula, *EMBO J.* **7**, 2967 (1988).

13. D. A. Goodenough, D. L. Paul, L. Jesaitis, *J. Cell Biol.* **107**, 1817 (1988).

14. D. W. Laird, J.-P. Revel, *J. Cell Sci.* **97**, 109 (1990).

15. G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **56**, 930 (1986).

16. B. Drake, *et al.*, *Science* **243**, 1586 (1989).

17. S. A. C. Gould, *et al.*, *J. Vac. Sci. Technol. A* **8**, 369 (1990); D. Rugar, P. Hansma, *Physics Today* **October**, 23 (1990).

18. R. D. Edstrom, *et al.*, *Biophys. J.* **58**, 1437 (1990).

19. H.-J. Butt, K. H. Downing, P. K. Hansma, *ibid.*, **58**, 1473 (1990).

20. Phosphate buffered saline was 137 mM NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4, and 4.3 mM Na_2HPO_4 at pH 7.2.

21. Animals used for the preparation of gap junctions were treated in accordance with institutional guidelines.

22. Imaging was carried out at Imaging Services, Box 9981, Truckee, California 95737. The NanoScope II and cantilevers were from Digital Instruments, 6780 Cortona Drive, Santa Barbara, California 93117. The stage that we used was calibrated at Digital Instruments on a diffraction grating (xy) and by interferometry (z).

23. T. R. Albrecht, S. Akamine, T. E. Carver, C. F. Quate, *J. Vac Sci. Technol. A* **8**, 3386 (1990).

24. No. 1 thickness glass cover slips were from Fisher Scientific, Cat. # 12-545-80. They were mounted on metal stubs obtained from Digital Instruments (Santa Barbara, CA). Cover slips were cleaned with a mild detergent, rinsed with distilled water. All solutions used for imaging were filtered through a 0.2 μm filter to remove large debris.

25. C. K. Manjunath, G. E. Goings, E. Page, *Biochem. J.* **205**, 189 (1982).

26. N. Hirokawa, J. Heuser, *Cell* **30**, 395 (1982).

27. The lattice constant (L) is determined from the Fourier transform by L=(1/sin120)x(1/a*).

28. G. Bernardini, C. Peracchia, *Invest. Opthalmol. Visual Sci.* **21**, 291 (1981).

29. G. E. Sosinsky, J. C. Jesior, D. L. D. Caspar, D. A. Goodenough, *Biophys. J.* **53**, 709 (1988).

30. Hoh et al., unpublished observation.

31. Trypsinization was carried out in PBS at an enzyme to substrate of 1:10 for 2 hours at 25°C after which an excess of soybean trypsin inhibitor was added. The gap junctions were pelleted at 14000xg for 30 min. and the supernatent was aspirated. Complete digestion was assessed by polyacrylamide gel electrophoresis, which showed the 10 kD band known to result from trypsinization. The trypsinized gap junctions were resuspended in PBS and stored at less than -20°C until used. Fixation was carried out on trypsinized gap junctions adsorbed to glass by submerging the entire stub with the cover slip and sample into 1% glutaraldehyde in PBS at room temperature for 20-30 minutes. The gap junctions were subsequently rinsed several times in PBS and kept in PBS at room temperature for up to several hours.

32. This work was supported by NIH grants R37 HL21788 to MFA, and HL37109 and BRSG-RR07003 to JPR. The authors express special thanks to Barney Drake of Imaging Services for his help with the operation of the AFM. We also thank Dr. R. Johnson, Dr. P. Lampe, Dr. K. Puranum, and Dr. B. Yancey for comments on the manuscript, Dr. P. Bjorkman for help in interpreting the Fourier transform, and Marit Jentoft-Nilsen for helpful discussions.

Chapter 3

Molecular Analysis of The Gap Junction Gene Family

Introduction

The identification of a major protein component of the hepatic gap junction opened the door to a variety of biochemical and molecular approaches to understanding its structure, activity and function (Henderson et al., 1979; Finbow et al., 1980; Nicholson et al., 1981). This prompted several efforts, including myself and others in the Revel laboratory, to isolate cDNA clones for the liver 28 kD protein. In 1986 two groups published papers describing cDNAs for the molecule now called connexin-32 (Cx32) or β_1 (Paul, 1986; Kumar and Gilula, 1986).

The cDNA clones for Cx32 provided a starting point to isolate connexin homologues from different tissues and different species. Given the poor understanding of the biological function of the gap junction, identifying connexin homologues in an organism amenable to genetic manipulation seemed the best approach to gaining further insight to the function of this cell to cell channel. A priori this seemed like a feasible task based on the wide distribution of the gap junctions and their extraordinary structural conservation. Therefore several investigators, including myself, screened several *Drosophila* libraries, both cDNA and genomic (D. Paul and T. Bargiello, personal communications). However, these attempts were all unsuccessful for reasons that are not clear, though some light is shed on the problem by the divergence rate analyses described in this chapter.

It was already known based on N-terminal protein sequencing that there was at least one homologue to Cx32, the Mr 45,000 protein found in heart (Nicholson et al., 1985; Manjunath et al., 1984). Therefore this was one of the first organs screened for cDNA clones (Beyer et al., 1987). Since then a variety of cDNA libraries have been screened for connexin homologues (for example: Beyer et al., 1988; Zhang and Nicholson, 1989; Ebihara et al., 1989). This chapter describes efforts toward isolating new members of the connexin gene family by screening a genomic library. Connexin gene structure is examined

and the expression of several connexins in different organs is determined by Northern blot analysis.

The availability of sequences of different connexin isoforms, and homologues from different species, also allows a variety of analyses. Multiple alignment of the sequences reveals conserved regions that are probably essential for the proper structure and function of the gap junctions, and diverged regions that may contain elements that confer specific functions to the various connexin isoforms. Evolutionary analysis, such as tree building, provides further insights to relationships between connexins. And on a grander scale, comparison with other ion channels reveals a possible common ancestry.

Materials and Methods

Isolation of Genomic Clones

A rat genomic library in Charon 4A (Sargent, 1979) was screened at low stringency with a Cx32 probe as described (Hoh et al., 1991; Chapter 4). Briefly, 5-6 genome equivalents of phage were plated and replicated onto Hybond-N filters (Amersham, Arlington Heights, Illinois). These filters were prehybridized for 6 hours at 45° C in 5 x SSPE (1 x SSPE = 150 mM NaCl, 10 mM $Na_2H(PO_4)$, 0.1 mM EDTA, pH 7.2), 5 x Denhardt's (50 x Denhardt's = 1% w/v bovine serum albumin, 1% w/v polyvinylpyrrolidone, and 1% w/v Ficoll), 30% deionized formamide, 20 µg/ml poly(A) RNA and 30 mg/ml yeast tRNA, and then probed at 45° C for 36 hours with a P^{32} labeled RNA probe prepared from a Cx32 cDNA (Paul, 1986). The probe had a specific activity of 2.5 µCi/ng, and was used at 0.2-0.5 ng/ml. The filters were washed in SSPE at increasing stringency to a final wash of 1 x SSPE at 60° C for 4 hours. Positive clones were rescreened at least twice at successively lower plaque density, to clonal purity.

Restriction mapping was performed by a series of single and double digests with a variety of enzymes. The digests were separated by agarose gel electrophoresis and fragments were sized with by linear regression analysis against *Hin*dIII/*Eco*RI digested

lambda DNA and *Hin*dIII digested lambda DNA. Restriction maps were assembled by hand.

Southern Blots

Clone blots were prepared from genomic DNA of positive clones. This genomic DNA was digested with several restriction enzymes, separated by agarose gel electrophoresis, blotted onto Hybond-N, and probed with random primer labeled probes prepared from the Cx26, Cx32, or Cx43 cDNA (Zhang and Nicholson, 1989; Paul, 1986; Beyer et al., 1987). Probes with a specific activity of 0.5 μCi/ng were prepared by the method of Feinberg and Vogelstein (1984). High stringency prehybridization and hybridization were carried out in 5 x SSPE, 5 x Denhardt's solution, 50% deionized formamide, 1 mM sodium pyrophosphate, 1 mM ATP, 0.1% SDS, 20 μg/ml salmon sperm DNA, 30 μg/ml yeast tRNA and 2-5 ng/ml probe DNA at 45° C for 24 and 36 hours respectively. Blots were washed in increasing stringency in 1 x SSPE at 40° C (low), 55° C (medium), and 70° C (high).

Genomic DNAs from several different organisms were isolated by the method of Strauss (1988) or purchased from Clontech Laboratories (Palo Alto, California). They were digested with several different restriction enzymes, separated by agarose gel electrophoresis, and blotted onto Hybond-N as described (Ausubel et al., 1988).

DNA Sequencing

Segments of several clones were sequenced in by the standard dideoxy chain-termination method (Sanger et al., 1977; Tabor, 1987). Sequencing was performed on a single template with a series of unique oligonucleotide primers, deleted templates generated by exonuclease III digestion (Ausubel et al., 1988), or a new method that used the random insertion of a transposon into the template to generate known start sites for sequencing (Strathmann et al., 1991). In some cases deoxyinosine was substituted for deoxyguanosine to resolve compressions.

Northerns Blots

Total RNA was isolated from liver, heart, tail skin including the dermis, tail connective tissue (everything left after the skin was removed), Harderian gland, eye, epidydimal fat pads, brain without the cerebellum, blood, stomach, femur including marrow, pancreas, spleen, ovary, uterus, thigh skeletal muscle, lung, duodenum, testis, and kidney, from both male and female Sprague-Dawley rats by the guanidinium thiocyanate (GTC) method or the modified GTC method that includes a centrifugation through CsCl (MacDonald, et al., 1987). RNA was separated by electrophoresing 10-20 μg of each RNA in a formaldehyde agarose gel. The fractionated RNA was capillary transferred onto Hybond-N, and probed under conditions identical with the high stringency southern blots except that the temperature was increased to 48-50° C.

Computers and Programs

A variety of computers and programs were used for manipulating and analyzing data. Early work, including database searching and sequence assembly was performed on the now defunct BIONET system and the available software (Intelligenetics, Mountain View, California). Subsequently two packages were used extensively, the PC-GENE programs (PCG) running on an Epson Equity II or an IBM PC, or the Wisconsin Computer Genetics Group (GCG) package V. 6.2 or earlier (Devereux et al., 1984) running on a VAX Station 3100 (Model m38). Other programs that were used include ALP3 (Murata et al. 1985) for triple alignments and GEL, both public domain programs (from Intelligenetics, Mountain View, California) that run under MS-DOS. MacPattern V. 1.2.2 (Freeware by Rainer Fuchs, Heidelberg, FRG) with V. 6.0 of the PROSITE data base was run on a Macintosh IIcx, and used to detect consensus sequences for 430 known structural and functional sequence motifs.

Multiple Alignment

The multiple alignment of connexins in Chapter 4 (Hoh et al., 1991) was extended with the RCx46 (sequence kindly supplied by Dr. E. C. Beyer), CCx42 and CCx45 (see

Table 1 for definitions and naming of connexins). These sequences were aligned three at a time with RCx43 using ALP3, and added to the previous alignment using LINEUP (GCG). Final adjustments to the alignment were made by eye, taking into account protein modification sites in the PROSITE database.

Consensus and variation type of representations of the multiple alignment were produced. The multiple alignment was initially transfered to an EXCEL (Microsoft, V. 2.2) spreadsheet, one sequence per column. All blank spaces in the alignment were replaced with the letter B. Functions that calculated total number of amino acids, number of different amino acids and the number of the most common amino acid (consensus) at each position (row) were prepared. Missing amino acids (B's) in the alignment were not scored in any of these. A threshold of 14 identities was used for the consensus plot, resulting in only the perfectly conserved residues being represented. Variation plots were produced in two ways, on a position by position basis with a threshold of 5 different amino acids and average over a 5 amino acid window with no threshold.

Divergence Rates and Phylogenetic Tree

Divergence analysis on all connexins above was carried out using the method of Perler et al. (1980) as implemented in the DIVERGE program (GCG). For divergence analysis the connexins were divided into two homology domains; the first domain (DI) was from amino acid 2 to 98 in the RCx31, the second domain (DII) corresponded to position 125-211 in RCx31 and the aligned domains in the other connexins (see Figure 8). To allow proper alignment of the sequences by DIVERGE, the gaps in the second domain were removed manually resulting in 3 segments corresponding to 125-153, 155-165, and 168-211 in RCx31. The divergence is only described as a function of replacement sites since all the divergence times in our analysis were greater than 100 MYr, except the human/rat branch. Analysis of silent sites, which do not result in amino acid replacements, is not useful for times greater than 100 MYr (Perler et al., 1980).

Divergence rates for Cx32 and Cx43 were estimated by linear regression analysis of plots of the percent corrected divergence versus the divergence times. The times used are human/rodent 75 MYr, mammal/bird 275 MYr, and mammal/amphibian 350 MYr (Dayhoff, 1972, Doolittle, et al., 1989, Perler, et al., 1980). Domains I and II were plotted separately.

Phylogenetic trees were generated using the program CLUSTAL (PCG) (Higgins and Sharp, 1988) for both domains I and II using rat Cx26, Cx31, Cx32, Cx43 and Cx46, *Xenopus* Cx38, and chicken Cx42 and Cx45, or for DI using all 14 available connexin sequences. This program generates all possible pair wise alignments using the algorithm of Needleman and Wunsch (1970). It then generates dendrograms using the unweighted pair group maximum averages method of Sneath and Sokal (1973).

Connexin Alignment With Other Channel Proteins

Alignments of connexins with other channel proteins was performed manually. The initial alignment of a 30 amino acid segment between RCx32 and several K^+ channels was identified by visual inspection. The remaining connexin sequences were added based on their alignment with RCx32, and the Na^+ and Ca^{++} sequences were added based on their previously described alignment with K^+ channels (Jan and Jan, 1990). Amino acids conserved between connexins and the other channels are bolded. The criterion for bolding is that any two amino acids in the connexins must be identical with any two in the other channels at a given position. This bolding scheme was selected empirically to emphasize the similarites between the sequences. The names and abbreviations of the K^+, Na^+, and Ca^{++} channels used are shown in Table 4.

Statistical significance of the alignments was tested by pair wise scrambled Needleman and Wunsch type alignments (PCOMPARE, PCG). These comparisons used the MDM-78 matrix (Dayhoff, 1978), the gap penalties were default values and sequences were scrambled 100 times.

Results and Discussion

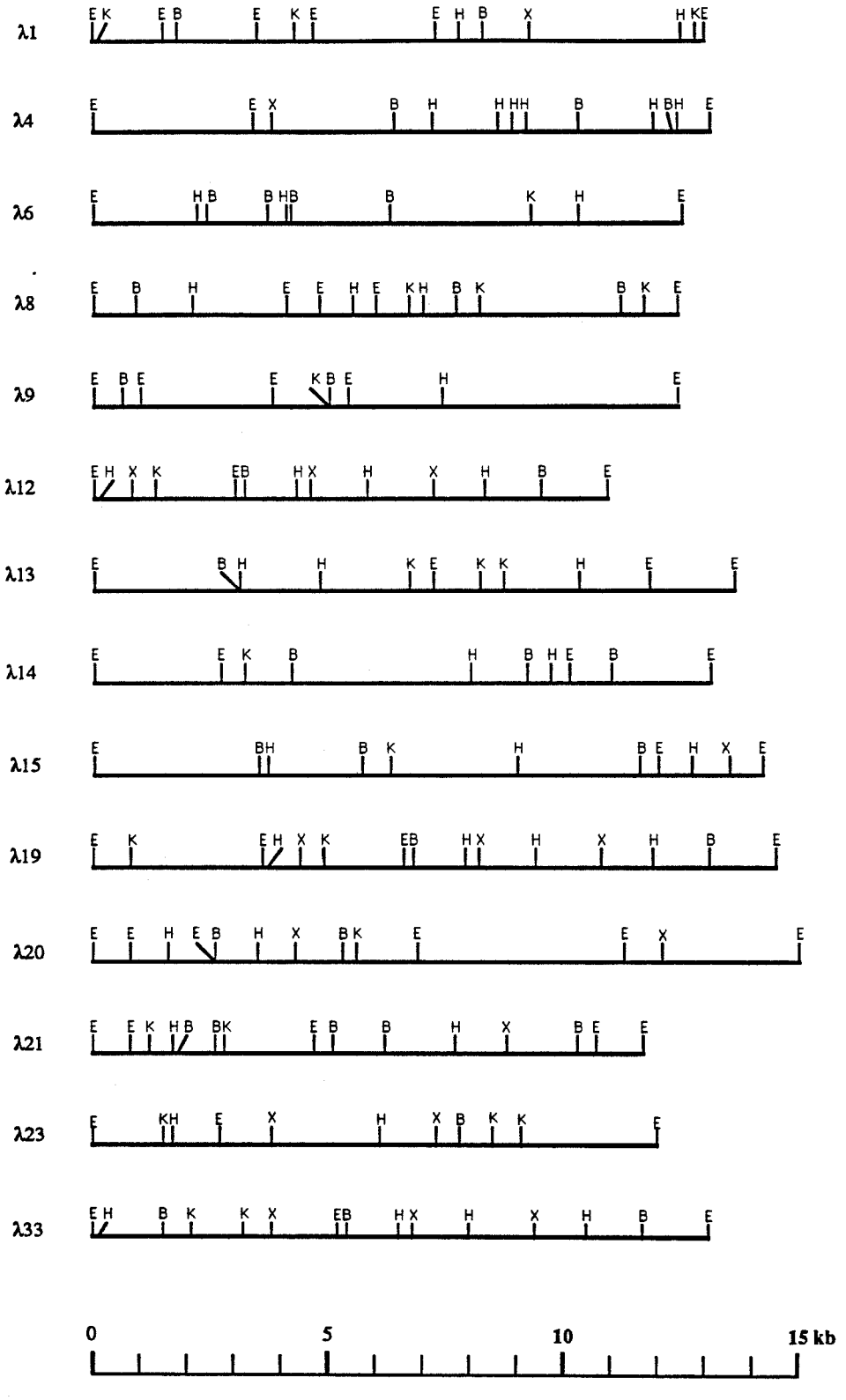
Isolation of Connexin Genes

Prior to the isolation of the Cx32 cDNA, several investigators had suggested that gap junctions in different tissues were composed of a single major protein (Hertzberg and Skibbens, 1984; Finbow et al., 1984; Zervos et al., 1985). However, Nicholson et al. (1985, 1987) showed by N-terminal sequence analysis that the major protein components of gap junctions from liver and heart formed at least a small gene family, and that in some cases the proteins appeared restricted to certain tissues. This observation in conjunction with the cloning of a cDNA for Cx32 prompted several groups to search for connexin homologues to further explore the diversity and distribution of gap junctions. Most of these groups screened organ or tissue specific libraries at low stringency with a Cx32 probe. Given that different gap junction proteins appeared to be expressed in different cell types, I decided to pursue an alternate approach and screen a genomic library at low stringency in the hope of identifying connexin homologues, particularly ones that were of low abundance, temporally or spatially restricted in expression, or otherwise not available in most cDNA libraries.

The low stringency screen of an *Eco*RI partial digest rat genomic library produced 30 positive clones. Six of these clones hybridized particularly strongly, and were therefore thought to represent the *Cx32* gene. Many clones were digested with several different restriction enzymes and probed under various conditions to identify potential new connexin homologues. These clones were then grouped into low, medium and high sequence similarity based on the temperature at which the probe could be washed off. The two medium similarity clones λ4 and λ21 had more than one band that hybridized to the probe. These clones with "split homology units" were considered particularly promising candidates for connexin homologues.

Many of the clones were also restriction mapped (Figure 1). These maps revealed that clones λ12, λ19, λ33, all high similarity clones, contained an identical 8 kb *Eco*RI

Figure 1. Restriction maps of 15 genomic clones isolated in a low stringency screen with the Cx32 cDNA as a probe. Enzymes used were *Eco*RI (E), *Bam*HI (B), *Kpn*I (K), *Hin*dIII (H), and *Xba*I (X).


λ1
λ4
λ6
λ8
λ9
λ12
λ13
λ14
λ15
λ19
λ20
λ21
λ23
λ33
0
5
10
15 kb

fragment indicating they were related, however, these clones also contained non-identical adjacent *Eco*RI fragments of 3.2 kb or 5.0 kb suggesting that they were unique. Furthermore, these two fragments both contained the hybridizing sequence, based on the blots described above. There were no other obviously related clones.

The Cx32 Gene

The high similarity clones were examined in greater detail by DNA sequence analysis. Sequencing of several kb of the λ12 clone revealed major portions of the sequence contained in the Cx32 cDNA, confirming that this clone indeed represented the *Cx32* gene. That sequencing also demonstrated that there was at least one exon in the *Cx32* gene at -19 with respect to the translation start site, and that the entire coding region for the protein appeared to be in a single exon. Further sequencing was abandoned, leaving approximately 200 bp within the coding region not sequenced, after the publication of the structure and sequence of the *Cx32* gene (Miller et al., 1988), which was identical with the data obtained for the λ12 clone.

To understand the two classes of high similarity clones, they were restriction mapped in more detail. These restriction maps show that both classes of clones contain the entire second exon of the *Cx32* gene, and are related by a 1.8 kb insertion/deletion approximately 300 bp to the 3' side of the second exon. These two classes of clones therefore appear to represent two restriction length polymorphisms (RFLP) of the Cx32 gene. Examination of genomic DNA from 45 Sprague-Dawley, 11 Fisher 344 and 10 Wistar rats shows that the RFLP is naturally occurring, and is strain specific. This *Eco*RI RFLP will serve as a convenient genetic marker for the Cx32 gene. These results are explained in greater detail in Appendix II (Hoh and Revel, 1991).

Other Connexin Genes

Several candidate clones for connexin homologues, identified by probing clone blots at low stringency, were sequenced. The first segment sequenced was a 5 kb *Eco*RI fragment from clone λ21. This revealed an open reading frame encoding a protein of 270

amino acids that was 50, 40, and 58% identical with RCx32, RCx43, and RCx26 respectively. This molecule has several characteristic features of the connexins including four predicted transmembrane domains, six perfectly conserved cysteines in the predicted extracellular loops (with a single amino acid insertion), and several conserved potential phosphorylation sites in the C-terminus. Based on the sequence similarity to other connexins, the conserved structural domains and a predicted molecular mass of 30,960 Da, this molecule is designated connexin-31, or β3 according to the nomenclature of Gimlich et al. (1990). A detailed analysis of Cx31 is presented in Chapter 4 (Hoh et al., 1991). Large portions of the λ1 and λ4 clones were also sequenced, however, no sequences with similarity to the connexins were discovered. Since no regions of substantial similarity between clones and the probe were detected, the reason that these clones hybridized with the probe is unknown.

There are many clones that have not been restriction mapped or sequenced and little is known about them, except that based on southern blots, they do not encode the *Cx43* or *Cx26* genes. It is curious and worrisome that neither of these two known connexins were isolated in this screen, particularly Cx26, which is more similar to Cx32 than Cx31 is over the conserved domains. There are several possible explanations for this. The hybridization conditions may have been too stringent for Cx43 and most other connexins, except Cx26, which are less similar to the Cx32 probe than Cx31. The library was prepared from genomic DNA partially digested with *Eco*RI, and therefore does not include parts of the genome in which *Eco*RI sites are more than 20 kb, the size limit of the vector, apart. Also, the library was amplified, which of course may cause biases in the regions of the genome represented. Finally the clones containing the Cx26 and Cx43 genes may have been under represented in the original library and the 6 genome equivalents screened were not sufficient. The fact that only two connexins, one unique, have resulted from this screen so far would suggest that further screenings of different libraries may be fruitful.

The identification of Cx31 brings the total number of cloned connexins to 14 (Table 1). These represent 8 different isoforms, 5 α connexins and 3 β connexins. Cx43, or α_1 has been isolated from 5 different species and Cx32 or β^1 has been isolated from 3. The rest have only been isolated from a single species. The types of cDNA libraries screened include heart for Cx43, liver for Cx32 and Cx26, lens for Cx46, and embryonic libraries from chicken and *Xenopus* for Cx42, Cx45, and Cx38. The probes used for all of these screens were either the Cx32 or Cx43.

Abbrev.	Protein	Species	Names		Reference
BCx43	N.I.	Bovine	Connexin-43	α_1	Lash et al., 1990
CCx42	N.I.	Chicken	Connexin-42	α_4	Beyer, 1990
CCx43	N.I.	Chicken	Connexin-43	α_1	Musil et al., 1990
CCx45	N.I.	Chicken	Connexin-45	α_5	Beyer, 1990
HCx32	N.I.	Human	Connexin-32	β_1	Kumar and Gilula, 1986
HCx43	N.I.	Human	Connexin-43	α_1	Fishman et al., 1990
RCx26	21 kD	Rat	Connexin-26	β_2	Zhang and Nicholson, 1989
RCx31	N.I.	Rat	Connexin-31	β_3	Hoh et al., 1991
RCx32	28 kD	Rat	Connexin-32	β_1	Paul, 1986
RCx43	45 kD	Rat	Connexin-43	α_1	Beyer et al., 1987
RCx46	N.I.	Rat	Connexin-46	α_3	Beyer et al., 1988
XCx30	N.I.	*Xenopus*	Connexin-30	β_1	Gimlich et al., 1988
XCx38	N.I.	*Xenopus*	Connexin-38	α_2	Ebihara et al., 1989
XCx43	N.I.	*Xenopus*	Connexin-43	α_1	Gimlich et al., 1990

Table 1. Abbreviations and names of connexins for which cDNA or genomic clones have been described to date. In the connexin-XX naming system proposed by Paul (1986), the XX is the calculated molecular mass of the predicted protein sequence. Most proteins have not been isolated (N.I.). The greek letter naming system proposed by Gimlich et al.(1990) is based on poorly defined sequence characteristics. It has been extended to connexins not previously named based on the phylogenetic analysis described in this chapter.

Distribution of Connexins

The current central issue in the study of gap junctions is biological function. There are many ways in which this problem can be addressed, including examining the distribution of different connexins. Correlating various connexins with different physiological or biological functions known for specific organs, tissues or cell types may provide clues to the biological role of the gap junction.

As previously noted, the gap junction has been described morphologically in almost every metazoan that has been examined, and in present in most cells types within an organism. Its structure is also highly conserved in the various organisms and cell types in which it is found, though some minor variations between vertebrate and invertebrate, and hepatic and cardiac, for example, gap junctions have been described. Based on the wide distribution and high degree of structural conservation, it was initially suggested that the gap junction in different tissues was composed of the same protein(s). Peptide mapping, and subsequent protein sequencing of the cardiac gap junction protein demonstrated that this was not the case and led to the proposal that different gap junction proteins may be germ line restricted, i.e., that there is an endodermal (hepatic type), a mesodermal (cardiac type), and an ectodermal (lens type) gap junction protein. The further isolation of the gap junction protein, and the cloning of cDNAs and genes for connexins, has now made available a variety of probes to further examine the distribution of connexins.

Northern blots of several tissues with probes for RCx26, RCx31, RCx32, or RCx43 reveals a wide and varied distribution of the different connexins (Figure 2, see Chapter 4 Figure 5 for RCx31 Northern blot). The 2.5 kb RCx26 mRNA is detected in the kidney and skin, while the 1.7 kb RCx32 mRNA is found in the brain, kidney, liver, duodenum, and stomach. RCx43 is the most widely distributed, its 3.0 kb mRNA being found in brain, connective tissue, heart, lung, skin, bone, fat, eye, ovary, and stomach. Finally, the 1.7 kb RCx31 mRNA is found in eye, placenta, Harderian gland and skin. Northern blots of course have lower limits of detection that usually are not determined. In these blots it is therefore possible that many tissues that appear negative, actually do express a particular connexin. It is for example known that Cx26 is expressed in liver (Nicholson et al., 1987; Zhang and Nicholson, 1989), which is not detected in these blots.

The distribution of connexins has been examined by several investigators with both antibodies and nucleic acid probes. A list of some reported localizations, which is now growing weekly, with results from the Northern blots described here is shown in Table 2.

Figure 2. Northern blot of total RNAs from a variety of tissues probed sequentially with the (A) RCx32, (B) RCx26 and (C) RCx43 cDNA at high stringency. The probes detected mRNAs of 1.7 kb, 2.5 kb, and 3.0 kb in the respective tissues, and were applied in the above order to avoid problems with residual hybridization. Such residual hybridization can be seen in kidney, for example.

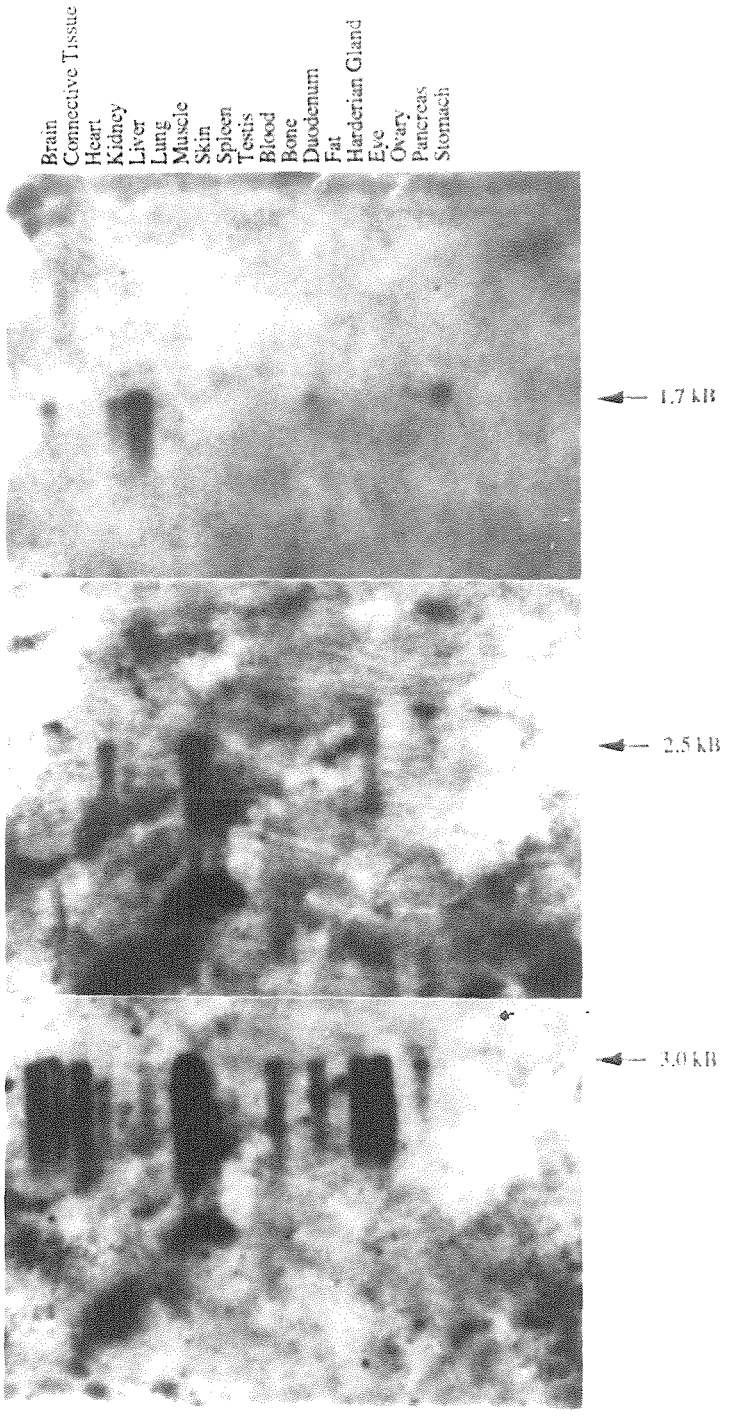

Brain
Connective Tissue
Heart
Kidney
Liver
Lung
Muscle
Skin
Spleen
Testis
Blood
Bone
Duodenum
Fat
Harderian Gland
Eye
Ovary
Pancreas
Stomach
1.7 kB
2.5 kB
3.0 kB

Table 2. Summary of connexin distribution. Symbols refer to reference for localization and are defined as: H (this chapter, Hoh et al., 1991), A (Larson et al., 1990), B (Beyer et al., 1987), C (Beyer, 1990), D (Dermietzel et al., 1984), E (Dermietzel et al., 1989), G (Beyer et al., 1988), L (Lash et al., 1990), M (Musil et al., 1990), P (Paul, 1986), R (Beyer et al., 1989), T (Traub et al., 1989), Y (Yamamoto et al., 1990), and Z (Zhang and Nicholson, 1989). The superscript c refers to data from permanent cell lines.

	Cx26	Cx31	Cx32	Cx38	Cx42	Cx43	Cx45	Cx46
Liver	ZT		PZD		C			
Heart					C	BLMH	C	G
Kidney	Z		Z R			BH		
Eye		H						
Lens						BM[c]		
Epithilium						R		
Fiber cells								G
Cornea								
Brain			P				C	
Astrocytes						E		
Glia						Y		
Oligodendro.			E					
Leptomenin.	E					E		
Ependymal	E					EY		
Pineal	E					E		
Arachnoid						Y		
Lung			Z					
Skin	H	H				H		
Stomach	Z		Z		C	R	C	
Small Intestine	Z		Z					
Large Intestine						R		
Pancreas			R					
Exocrine			D					
Endocrine			D					
Fat						H		
Bone						H		
Blood						H		
Harderian Gland		H						
Plancenta		H						
Ovaries						B		
Granulosa						R		
Theca Externa						R		
Testis			Z					
Spleen			Z					
Smooth Muscle						L[c]A		
Aorta						A		
Skeletal Muscle					C		C	
Uterus						B		
Endometrium			DR					
Myometrium			D					
Fallopian Tube			D					
Embryonic				E	C		C	
Pulm. Endothel.						L[c]		

This table reveals that Cx43 is the most widely distributed isoform of this gene family. The distributions also show that there is no germline restriction of these connexins as was previously proposed. Most of the data in the table is derived from Western blots or Northern blots, of whole organs or tissues, which do not have very high spatial resolution. In addition there are a number of reports of antibody staining of sections or in situ hybridization that provide more spatial information (for example Micevych and Abelson, 1991; Nicholson et al., 1987; Beyer et al., 1989; Yamamoto et al., 1990). However, even when these are considered there is no discernable pattern for the expression.

It seems extraordinarily unlikely that a given cell purely at random expresses one or a few different connexins. Therefore, there must be some pattern, set of rules or requirements that determine where and when a certain connexin is expressed. Based on the distribution data available to date one cannot make any such rationalizations for the connexins yet. Having said this, it is also well known that biological systems tend to be redundant and it is possible that the diversity in connexins, and expression of more than one connexin in a single cell, is in part just that, redundancy. This would severely complicate any attempts to correlate distribution with physiology or function.

Multiple Alignment of Connexins

Comparison of sequences from different homologues, or of sequences from different species of a protein, can provide a great deal of information about structure and function. In principle such a comparison reveals two things about the compared sequences, similarities or differences. Highly conserved regions are of course candidates for essential and common, structural or functional elements of the protein, while diverged regions are more likely to be less important or have specific functions for the particular homologue or isoform. A comparison can be carried out pairwise between related sequences, or many at once in a multiple alignment. While I have examined most pairwise alignments between connexins, only the multiple alignment has yielded any useful information.

Constructing a multiple alignment has some steps that are largely arbitrary, the most significant of which is the order in which the alignment is assembled. The most common way of ordering a multiple alignment is based on sequence similarity, the most similar sequences are aligned first and the other sequences are added in decreasing similarity. Because of the obvious conserved domains containing the transmembrane segments, and the diverged C-terminal tail and intracellular loop, it seemed appropriate here to order the alignment by decreasing (or increasing) size of the sequences. Therefore, a series of three fold alignments were performed to detect similar regions between the various connexins, and a multiple alignment was constructed by hand to take into account the regions detected in the three fold alignments. The alignment was then adjusted to minimize the number of single or small groups of amino acids, and to take into account the alignment of predicted protein kinase C phosphorylation sites (Figure 3).

The multiple alignment shows the general features of connexins that have now been well described (Paul, 1986; Beyer et al., 1987; Zimmer et al., 1987; Zhang and Nicholson, 1989; Beyer et al., 1990; Hoh et al., 1991). Two highly conserved domains of c. 100 amino acids containing the transmembrane spanning regions separated by a cytoplasmic loop of 20-60 amino acids, followed by a highly diverged C-terminal tail c. 15-170 amino acids long. Within these large domains a number of specific features have also been described. The conservation of three cysteines in each extracellular loop with the pattern CX_6CX_3C and CX_4CX_5C, and the putative amphipathic helix, TM3, that may line the pore.

The information in a multiple alignment is extremely dense, and is often cumbersome to extract information from. There are many schemes such as bolding similarities, boxing similarities, or replacing conserved amino acids with dashes, that are designed to make information in a multiple alignment more useful. However, these too are often inadequate. Therefore I have chosen to show graphical representations of the multiple alignment designed to be informative (Figures 4 and 5). These representations are of two

Figure 3. Multiple alignment of 14 connexin protein sequences (see table 1 for nomenclature). Alignment was generated by hand from a series of three fold alignments using ALP3, and was adjusted to minimize the number of breaks and take into account the predicted protein kinase C phosphorylation sites (bolded).

```
RCx26 M.DWGTLQSI LGGVNKHSTS IGKIWLTVLF IFRIMILVVA AKEVWGDEQA DFVCNTLQPG CKNVCYDHYF PISHIRLWAL QLIMVSTPAS LVAMHVAYRR
XCx30 M.NWAGLYAI LSGVNRHSTS IGRIWLSVVF IFRIMVLVAA AESVWGDEKS AFTCNTQQPG CNSVCYDHFF PISHIRLWAL QLIIVSTPAL LVAMHVAHLQ
RCx31 M.DWKKLQDL LSGVNQYSTA FGRIWLSVVF VFRVLVYVVA AERVWGDEQK DFDCNTRQPG CTNVCYDNFF PISNIRLWAL QLIFVTCPSM LVILHVAYRE
HCx32 M.NWTGLYTL LSGVNRHSTA IGRVWLSVIF IFRIMVLVVA AESVWGDEKS SFICNTLQPG CNSVCYDQFF PISHVRLWSL QLILVSTPAL LVAMHVAHQQ
RCx32 M.NWTGLYTL LSGVNRHSTA IGRVWLSVIF IFRIMVLVVA AESVWGDEKS SFICNTLQPG CNSVCYDHFF PISHVRLWSL QLILVSTPAL LVAMHVAHQQ
XCx38 MAGWELLKLL LDDVQEHSTL IGKVWLTVLF IFRIFILSVA GESVWTDEQS DFICNTQQPG CTNVCYDQAF PIYHVRYWVL QFLFVSTPTL TYLGHMVYLS
CCx42 MGDWSFLGEF LEEVHKHSTV VGKVWLTVLF IFRMLVLGTA AGPLWGDEQS DFMCDTQQPG CENVCYDKAF PISHVRFWVL QIIFVSTPSL VYMGHAMHTV
BCx43 MGDWSALGKL LDKVQAYSTA GGKVWLSVLF IFRILLLGTA VESAWGDEQS AFRCNTQQPG CENVCYDKSF PISHVRFWVL QIIFVSVPTL LYLAHVFYVM
CCx43 MGDWSALGKL LDKVQAYSTA GGKVWLSVLF IFRILLLGTA VESAWGDEHV AFRCNTQQPG CENVCYDKSF PISHVRFWVL QIIFVSVPTL LYLAHVFYVM
HCx43 MGDWSALGKL LDKVQAYSTA GGKVWLSVLF IFRILLLGTA VESAWGDEQS AFRCNTQQPG CENVCYDKSF PISHVRFWVL QIIFVSVPTL LYLAHVFYVM
RCx43 MGDWSALGKL LDKVQAYSTA GGKVWLSVLF IFRILLLGTA VESAWGDEQS AFRCNTQQPG CENVCYDKSF PISHVRFWVL QIIFVSVPTL LYLAHVFYVM
XCx43 MGDWSALGRL LDKVQAYSTA GGKVWLSVLF IFRILLLGTA VESAWGDEQS AFVCNTQQPG CENVCYDKSF PISHVRFWVL QIIFVSTPTL LYLAHVFYLM
CCx45 M.SWSFLTRL LEEIHNHSTF VGKIWLSVLI VFRIVLTAVG GESIYYDEQS KFVCNTEQPG CENVCYDAFA PLSHVRFWVF QIILVATPSV MYLGYAIHKI
RCx46 MGDWSFLGRL LENAQEHSTV IGKVWLTVLF IFRILVLGAA AEEVWGDEQS DFTCNTQQPG CENVCYDRAF PISHIRFWAL QIIFVSTPTL IYLGHVLHIV

RCx26 .......... ........HE KKRKFMKGEI KNEFKDIEEI KTQ..... .......... .......... ...KVRIE.G SLWWTYTTSI FFRVIFEAVF
XCx30 .......... .........H QEKKELRLSR HVKDQELAEV KKH..... .......... .......... ...KVKIS.G TLWWTYISSV FFRIIFEATF
RCx31 .......... ........ER ERKHRQKHGE HCAKLYSHPG KKH..... .......... .......... .........G GLWWTYLFSL IFKLIIELVF
HCx32 .......... .........H IEKKMLRLEG HGDPLHLEEV KRH..... .......... .......... ...KVHIS.G TLWWTYVISV VFRLLFEAVF
RCx32 .......... .........H IEKKMLRLEG HGDPLHLEEV KRH..... .......... .......... ...KVHIS.G TLWWTYVISV VFRLLFEAVF
XCx38 .......... .......... ......KKEE KERQKENESR IL.VANEAQT EVY..... .........SS ATKKIRIQ.G PLMCTYTTSV VFKSIFEAGF
CCx42 .......... .......... ......RMEE KRKMKEAERE AQ.EMKNSGD TYYQQKCPVA EKTELSCWDE SGGKIILR.G SLLNTYVYSI LIRTAMEIAF
BCx43 .......... .......... ......RKEE KLNKKEEELK VVAQTDGANV DMHLKQ...I EIKKFKYGIE EHGKVKMR.G GLLRTYIISI LFKSVFEVAF
CCx43 .......... .......... ......RKEE KLNKREEELK VV.QNDGVNV DMHLKQ...I EIKKFKYGIE EHGKVKMR.G GLLRTYIISI LFKSVFEVAF
HCx43 .......... .......... ......RKEE KLNKKEEELK VA.QTDGVNV DMHLKQ...I EIKKFKYGIE EHGKVKMR.G GLLRTYIISI LFKSIFEVAF
RCx43 .......... .......... ......RKEE KLNKKEEELK VA.QTDGVNV EMHLKQ...I EIKKFKYGIE EHGKVKMR.G GLLRTYIISI LFKSVFEVAF
XCx43 .......... .......... ......RKEE KLNRKEEELK MV.QNEGGNV DMHLKQ...I EIKKFKYGLE EHGKVKMR.G GLLRTYIISI LFKSVFEVGF
CCx45 ARMVEHSDVD RRFRSKSFST RWKQHRGLEE AEDDHEEDPM MY.PEIELES ERENKEQQPP AK......AK HDGRRRIRED GLMRIYVLQL LVRATFEVGF
RCx46 .......... .......... ......RMEE KKKEREEELL RR.DNPQHGR GREPMRTGSP RDPPLR...D DRGKVXIA.G ALLRTYVFNI IFKTLFEVGF

RCx26 MYVFYIMYNG FFMQRLVKCN .AWPCPNTVD CFISRPTEKT VFTVFMISVS GICILLNITE LCYLFIRYCS GKSKRPV... .......... ..........
XCx30 MYIFYLIYPG YSMIRLLKCD .AYPCPNTVD CFVSRPTEKT IFTVFMLVAS GVCIVLNVAE VFFLIAQACT RRARR..... HRDSGSISKE HQ........
RCx31 LYVLHTLWHG FTMPRLVQCA SVVPCPNTVD CYIARPTEKK VFTYFMVGAS AVCIILTICE ICYLIF.... .......... HRIMRGLSKD KST.......
HCx32 MYVFYLLYPG YAMVRLVKCD .VYPCPNTVD CFVSRPTEKT VFTVFMLAAS GICIILNVAE VVYLIIRACA RRAQRRSNPP SRKGSGFGHR LSPEYK....
RCx32 MYVFYLLYPG YAMVRLVKCE .AFPCPNTVD CFVSRPTEKT VFTVFMLAAS GICIILNVAE VVYLIIRACA RRAQRRSNPP SRKGSGFGHR LSPEYK....
XCx38 LLGQWYIY.G FVMSPIFVCE .RIPCKHKVE CFVSRPMEKT IFIIFMLVVS LISLLLNLME LIHLSFKCFQ HGIKE..... .......... GATCPPTGIP
CCx42 IVGQYILY.G IFLETLYICQ .RAPCPHPVN CYVSRPTEKN VFIIFMLAVA VLSLFLSLAE LYHLGWKKAK ERCSRAYKPS PST..... .....APRRL
BCx43 LLIQWYIY.G FSLSAVYTCK .RDPCPHQVD CFLSRPTEKT IFIIFMLVVS LVSLALNIIE LFYVFFKGVK DRVKGKSDPY HTTTGPLSPS KDCGSPKYAY
CCx43 LLIQWYIY.G FSLSAIYTCE .RDPCPHRVD CFLSRPTEKT IFIVFMLVVS LVSLALNIIE LFYVFFKGVK DRVKGKTDPY .SHSGTMSPS KDCGSPKYAY
HCx43 LLIQWYIY.G FSLSAVYTCK .RDPCPHQVD CFLSRPTEKT IFIIFMLVVS LVSLALNIIE LFYVFFKGVK DRVKGKSDPY HATSGALSPA KDCGSQKYAY
RCx43 LLIQWYIY.G FSLSAVYTCK .RDPCPHQVD CFLSRPTEKT IFIIFMLVVS LVSLALNIIE LFYVFFKGVK DRVKGRSDPY HATTGPLSPS KDCGSPKYAY
XCx43 IIIQWYMY.G FSLSAIYTCK .RDPCPHQVD CFLSRPTEKT IFIWFMLIVS IVSLALNIIE LFYVTYKSIK DGIKGKKDPF SATNDAVISG KECGSPKYAY
CCx45 LIGQYLLY.G FEVSPVFVCS .RKPCPHKID CFISRPTEKT IFLLIMYGVS CMCLLLNVWE MLHLGFGTIR DTLNNKRKEL EDSGTYNYPF TWNTPSAPPG
RCx46 IAGQYFLY.G FQLQPLYRCD .RWPCPNTVD CFISRPTEKT IFVIFMLAVA CASLVLNMLE IYHLGWKKLK QGVTNHFNPD ASECRHKPLD PLSEAANSGP

RCx26 .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
XCx30 .......... .......... .......... .......... .......... .......... .......... .....QNEMN LLITGGS... ..........
RCx31 .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
HCx32 .......... .......... .......... .......... .......... .......... .......... .....QNEIN KLLSEQDGS. ..........
RCx32 .......... .......... .......... .......... .......... .......... .......... .....QNEIN KLLSEQDGS. ..........
XCx38 FNG..... .......... .......... .AGNRMPPQE YTNPPSSNQD IDLPAYNKMS GGHNWSSIQM E....QQVNG LVKPKCQCD. ..........
CCx42 ESAPQVERA. .......... .......... .QMYTPPPDF NQCLASPNGK FISPFSNKMA SQQNTANFAT ERVHSQEDAA GEGPFMKSS. ..........
BCx43 FNGCSSPTA. .......... .......... .PLSPMSPPG YKLVTGDRNN SSCRNYNKQA SEQNWANYSA E....QNRMG QAGSTISNS. ..........
CCx43 YNGCSSPTA. .......... .......... .PLSPMSPPG YKLVTGDRNN SSCRNYNKQA SEQNWANYSA E....QNRMG QAGSTISNS. ..........
HCx43 FNGCSSPTA. .......... .......... .PLSPMSPPG YKLVTGDRNN SSCRNYNKQA SEQNWANYSA E....QNRMG QAGSTISNS. ..........
RCx43 FNGCSSPTA. .......... .......... .PLSPMSPPG YKLVTGDRNN SSCRNYNKQA SEQNWANYSA E....QNRMG QAGSTISNS. ..........
XCx43 FNGCSSPTA. .......... .......... ....PMSPPG YKLVTGERNP SSCRNYNKQA SEQNWANYSA E....QNRMG QAGSTISNT. ..........
CCx45 YNIAVKPDQ. .......... .......... .MQYTELSNA KMAYKQNKAN IAQEQQYGSN EENIPADLEN L....QREIK VAQERLDMAI QAYNNQNNP.
RCx46 PSVSIGLPPY YTHPACPTVQ GKATGFPGAP LLPADFTVVT LNDAQGRGHP VKHCNGHHLT TEQNWASLGA E...PQTPAS K.PSSAASSP HGRKGLTDSS

RCx26 .......... .......... .......... .......... .......... .......... .......
XCx30 .......... .......... .........II KRSAG..... QEKGDHCSTS .......... .......... .......
RCx31 .......... .......... .........K SISSPKSSSR ASTCRCHHKL LESGDLEAVP ADDKLQASAP SLTPI..
HCx32 .......... .......... .....LKDIL RRSPGTGAGL AEKSDRCSAC .......... .......... .......
RCx32 .......... .......... .....LKDIL RRSPGTGAGL AEKSDRCSAC .......... .......... .......
XCx38 .......... .CWSQSNISV VVSGAPGIIS NMDAVKRNHQ TSSKQQYV.. .......... .......... .......
CCx42 .......YME SPEVASECAA PALPESYFNE KRRFSKA.SR ABS..KARSD DLSV..... .......... .......
BCx43 HAQPFDFPDD HQNSKKLDAG HELQPLAIVD QRPSSRASSR ABS..RPRPD DLEI..... .......... .......
CCx43 HAQPFDFADE HQNTKKLASG HELQPLTIVD QRPPSRASSR ABS..RPRPD DLEI..... .......... .......
HCx43 HAQPFDFPDD NQNSKKLAAG HELQPLAIVD QRPSSRASSR ABS..RPRPD DLEI..... .......... .......
RCx43 HAQPFDFPDD NQNAKKVAAG HELQPLAIVD QRPSSRASSR ABS..RPRPD DLEI..... .......... .......
XCx43 HAQPFDFSDE HQNTKKMAPG HEMQPLTILD QRPSSRASSH ABS..RPRPD DLEI..... .......... .......
CCx45 .......... .......... ..GSSSREKK SKAGSNKSS. ABSKSGDGKN SVWI..... .......... .......
RCx46 GSSLEESALV VTPEGEQALA TTVEMHSPPL VLLDPERSSK SSS.GRARPG DLAI..... .......... .......
```

Figure 4. Plot of residues identical in all 14 connexins in the multiple alignment (Figure 3) Center figure shows the entire length of the alignment. Above and below are the highly conserved domains enlarged. Bars show the position of the predicted transmembrane spans.

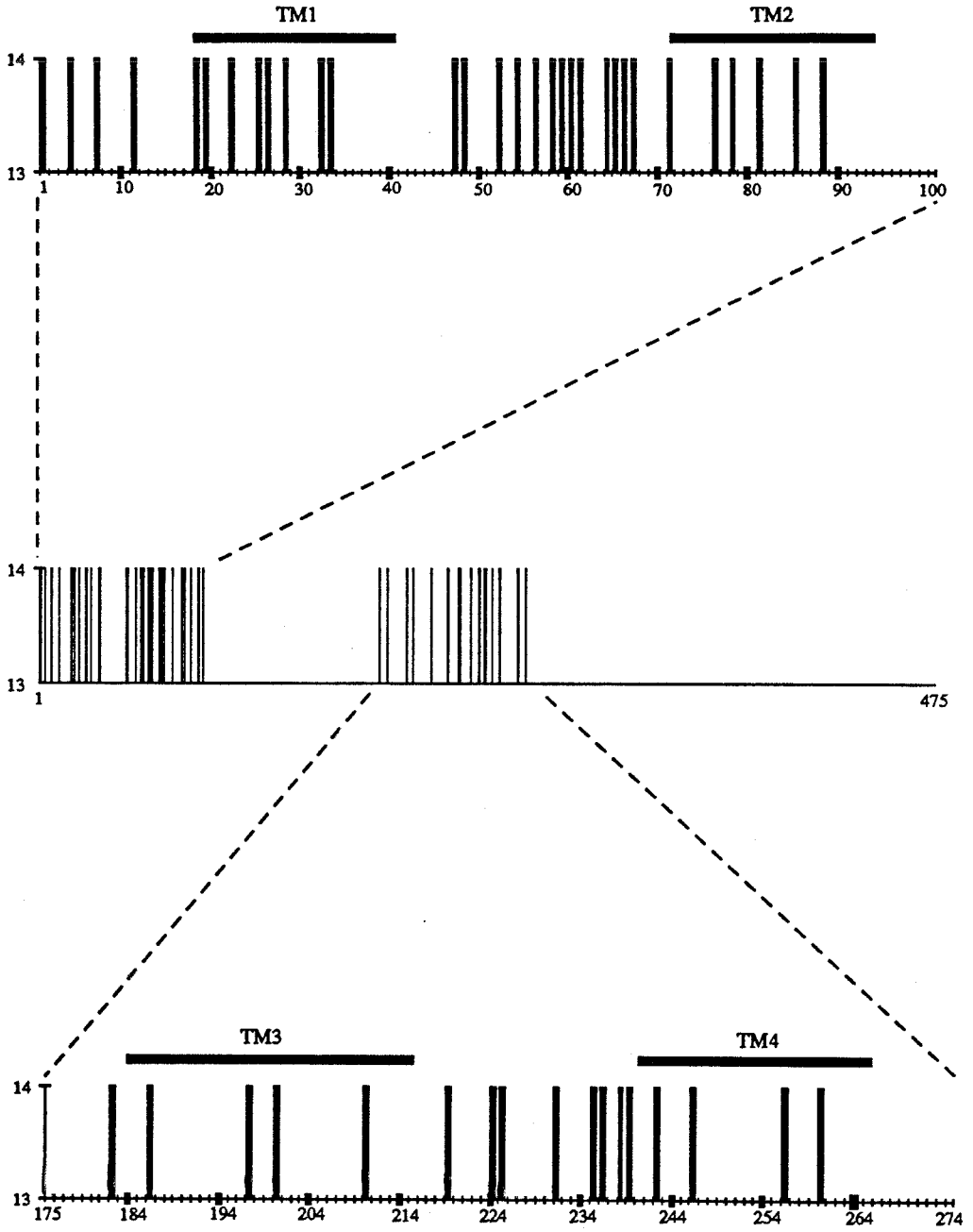
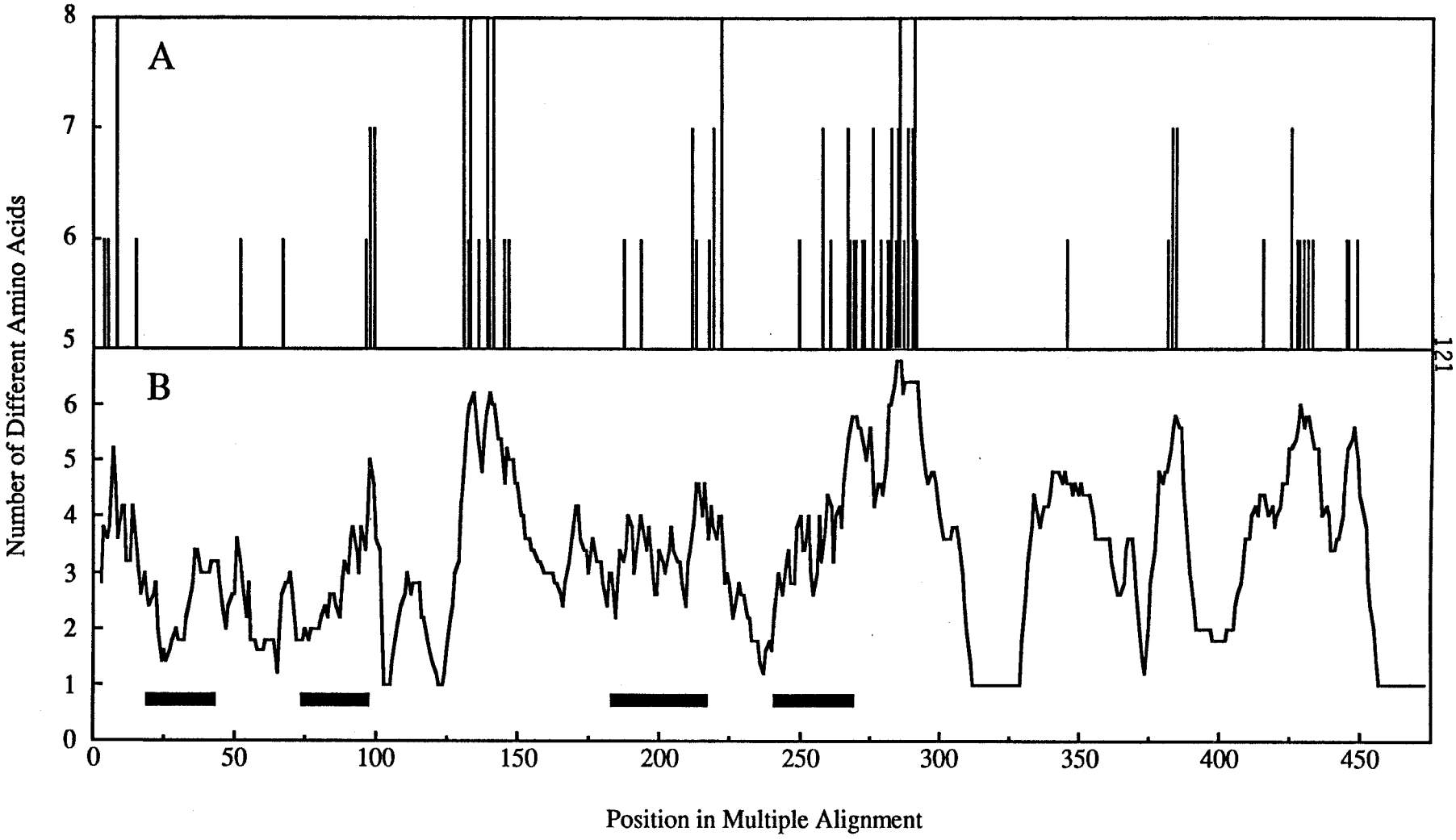

TM1
TM2
14
13
1
10
20
30
40
50
60
70
80
90
100
14
13
1
475
TM3
TM4
14
13
175
184
194
204
214
224
234
244
254
264
274

Figure 5. Plots of positions in the multiple alignment (Figure 3) with at least 5 different residues. A. Shows the exact number of different amino acids at each position. B. Plot of the number of different amino acids averaged over a 5 residue window. Bars show the position of predicted transmembrane spans.


A
B
Number of Different Amino Acids
Position in Multiple Alignment
0
1
2
3
4
5
6
7
8
0
50
100
150
200
250
300
350
400
450

types, ones that emphasize similarities and ones that emphasize differences. Because of the natural division between the β connexins and α connexins, and the structural similarities within the groups, analyses of each group alone as well as together were carried out (not shown).

Representing conserved regions between several sequences has a number of difficulties. The most notable is the alignment of the highly diverged sequences of different length. There are for any given set of such sequences several ways in which they can be aligned, and of course a sequence from one homologue will not appear conserved if it is not aligned with the corresponding sequence in another homologue. In addition what is conserved and not conserved depends on the criteria used. Conserved residues can be scored quantitatively: based on known evolutionary replacement statistics using the MDM-78 matrix, by estimated genetic relatedness using the genetic matrix, by a modified genetic matrix that takes into account the structure of amino acids, or by the standard unitary matrix (which per se is not quantitative, see Feng et al., 1985, for discussion of different matrices). There are also chemical and structural criteria for considering alignments that group amino acids into sets of, for example, negatively charged (D and E), positively charged (K and R), or hydrophobic (I, L, V, etc.) residues. Further, the representation of similarities can be position by position, or an average value over a window that emphasizes similar regions. The representation of conserved of all connexins shown here (Figure 4) is the simplest kind. It uses the multiple alignment in Figure 3, and shows only the exact positions in which an amino acid is perfectly conserved.

Representing diverged regions also depends significantly on the initial alignment of the sequences, however, the problem here is directly opposite to representation of conserved regions. That is, alignment of regions that are not truly homologous sometimes gives rise to the appearance they are conserved. This is particularly true for small regions of a few amino acids. Scoring of the alignment can in principle be the exact inverse of that for conserved regions. In addition, the number of different amino acids at a given position is a

useful number that tends to compensate for cases when the consensus number is low but there is only a few different amino acids used (Figure 5).

Examination of the multiple alignment and its various representations shows that the amino terminal 22 amino acids are a highly conserved domain with 4 perfectly conserved residues. This part of the protein is thought to be cytoplasmic and immediately precedes TM1 (Yancey et al., 1990; Zimmer et al., 1987). No function has been proposed for this domain, though for Cx43, antibodies against the N-terminus can affect the activity of the channel (Yancey et al., 1990).

The putative 1:st, 2:nd and 4:th transmembrane segments are also highly conserved. They each consist of a hydrophobic stretch of c. 20 amino acids that is predicted by a variety of algorithms to form a transmembrane α-helix. Just to form such a structure should not require the high degree of conservation seen, since there are no known absolute requirements for specific residues in transmembrane α-helices (Heijne and Gavel, 1988). This therefore suggests that the conserved residues are essential for a specific function or structure common to connexins. Gap junctions have two highly conserved features, the general morphology and structure, including connexon diameter, center to center spacing, and size of the gap, and the ability to pass small molecules from one cell to the next. The known variable features of gap junctions are the quantity of molecules that can be passed (conductance) and control of gating of the channels, by pH for example. Given that the variable conductances suggest differences in the pore region between the connexins, the highly conserved regions of the molecule such as TM1, TM2 and TM4 are more likely to be essential to the overall structure of the gap junction.

The third putative transmembrane segment is a relatively long hydrophobic stretch of 30-40 amino acids across which there are 5 perfectly conserved residues. Several groups have proposed that part of this segment forms an amphipathic helix that lines the pore (Milks et al., 1988; Beyer et al., 1990). This is largely based on the then perfectly conserved motif TYX_2SX_3K/RX_3E. However, the recent availability of the CCx45 and

RCx46 sequences has disrupted this motif with an isoleucine instead of threonine, and asparagine or glutamine instead of the serine. In addition the amphipathic helix model does not account for the differences in conductances of Cx26, Cx32 or Cx43, though it is possible that conductance is regulated outside the pore proper. Based on sequence similarity with other channel proteins, an alternate hypothesis is put forth below, in which the pore is lined by a segment of amino acids that begins just before the long hydrophobic stretch, and may span the membrane as many as three times instead of one (see model, Chapter 1, Figure 12).

The two extracellular loops have the characteristic three cysteines, and several other perfectly conserved residues. These loops are presumably the regions that are responsible for the connexon-connexon interactions, and must form a seal between the pore and the extracellular fluids. It is known that the cysteines in this region form at least one disulfide bond between the loops (John and Revel, in preparation). Such disulfide bonding between extracellular loops has also been seen in other proteins such as the acetylcholine receptor, β adrenergic receptor and rhodopsin (Dohlman et al., 1990; Karnik and Khorana, 1990). These disulfides presumably place severe structural constraints on the loops and how they interact across the gap. Consistent with structural conservation in this region of the molecule is the observation that functional heterologous channels, with for example one connexon made of Cx32 and one of Cx43, can be formed (Swenson et al., 1989).

The intracellular loop that separates TM2 from TM3 varies in length for c. 20-60 amino acids and is highly diverged. Its most notable feature is a large number of charged residues. Clusters of lysine and arginine are common in membrane proteins and appear to act as stop transfer signals (Heijne and Gavel, 1988). Beside this possible effect on assembly, there are no known or proposed functions for this domain.

The C-terminal tail of the connexins is the most diverged region of the molecule. It varies in length between 15 amino acids for Cx26 and 170 for Cx46. It is presumed that this region contains some of the regulatory domains for gating and other functions. A

comparison of sequences might detect elements responsible for common functions, however it will miss functions specific to various connexins. Therefore, the highly diverged nature of the C-terminus may be taken as evidence that it does contain elements that confer specific capabilities to the different connexins. From the variation plot of the connexins (Figure 5), there is a short stretch of sequence after TM4 that allows for an unusual number of different amino acids. Even between just the Cx43s, which are highly conserved along their entire length, this region is diverged. Because of this it seems unlikely that the region is directly responsible for any essential function, but may be a spacer to provide appropriate conformation of the cytoplasmic tail.

There are two conserved regions in the C-terminal tail that have been designated the QNE and SSRA domains at positions 376 and 438 respectively in the multiple alignment (Hoh et al., 1991; Chapter 4). There is no known or proposed function for the QNE domain. On the other hand, the SSRA domain has several predicted phosphorylation sites that may play a role in the regulation of channel activity. There is direct biochemical evidence that at least some of these sites are actually phosphorylated in Cx43 (Laird and Revel, 1990; Musil et al., 1990; Saez et al., 1986).

Evolution of the Gap Junction Gene Family

The unique alignments at the C-termini shown in the multiple alignment, have implications for how the connexins evolved. Without any homologies in the C-terminal region it would appear that a simple addition/removal of sequence at the C-terminus of a progenitor connexin could have given rise to the various isoforms. The fact that there are at least two conserved domains, QNE and SSRA, suggests instead that the C-termini of these connexins are related through several internal deletions/insertions (Figure 6). Of course, this does not preclude that additions or removals also occured.

To describe the evolution of the gap junction gene family, a phylogentetic tree was constructed based on the protein sequences. There are essentially two parts to building a tree, first one must quantitatively compare the traits, in this case the sequences, and second

the branch order is determined. There are many algorithms and approaches to generating trees, the one used here is one of the conceptually simplest. I have used the

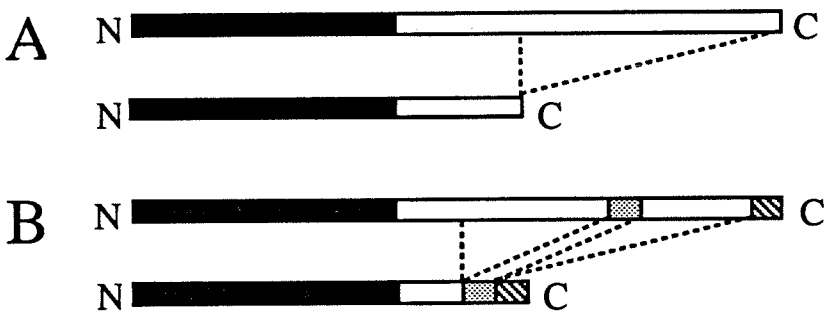


Figure 6. Schematic of connexin evolution. Earlier analyses of the connexin sequences detected no substantial similarities along the C-terminal tails of the molecules, which suggested that one or more major addition/removal events from the end of the protein, as in (A), could explain the divergent nature of this region. The data presented in the multiple alignment here, however, reveals several stretches of sequence similarity in the C-terminal tails which means that this region of the molecule probably evolved through a series of insertion/deletion events as in (B).

computer program CLUSTAL (PCG), which first generates all possible pairwise alignments, using the Needleman and Wunsch algorithm, between the sequences and scores the alignments using the MDM-78 matrix. These scores are then used to assemble a tree by the unweighted pair group maximum averages method of Sneath and Sokal (1973), which starts with the tree with the most similar pair. The average of the scores of these two sequences compared with all other sequences, is then compared to all the scores from pairs not including either of the first two sequences. If any of the averages is higher than all the other pairs, then that sequence is made a branch below the first two sequences. If none of the averages is higher, then the highest scoring second pair is made a branch completely separate from the first. This process is repeated until the tree is complete (Figure 7).

The tree of 14 connexins shows two major branches, one corresponding to the longer α connexins and one corresponding to the shorter β connexins. The current criteria for the greek letter based naming system for gap junction proteins are poorly defined

sequence characteristics. Given the clustering of α and β connexins on two major branches, phylogenetic relatedness may be a suitable criterion for this naming system in the

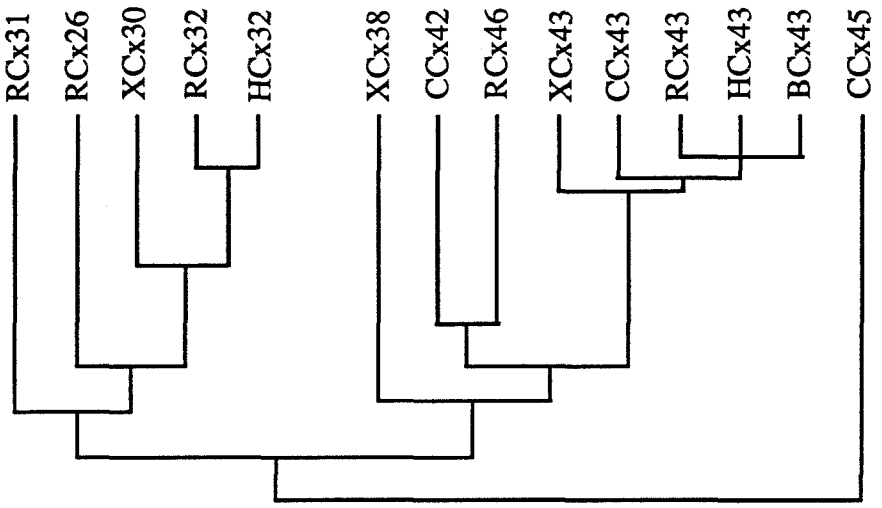


Figure 7. Phylogenetic tree of connexins described in Table 1. It shows two major branches corresponding to the β and α connexins. CCx45 is tentatively classified as an α though it may actually form a new branch, γ connexins. The tree was constructed by first carrying out all possible pair wise comparisons using DI, roughly the first 100 amino acids, and scoring with MDM-78. The branch order was the determined by the unweighted pair group maximum averages method. The known species divergence times are 75 MYr for mammals, 275 MYr for birds and 350 MYr for amphibians. The time of divergence between the two main branches is estimated below to be 1.3-1.9 BYr.

future. Cx45 is tentatively designated an α connexin, however, it is very different than any other connexin and may form a new group, the γ connexins. Otherwise, it bears most structural resemblance to the α connexins, and is treated as such here. It should be noted that all the connexins for which sequences from different species are available, i.e., Cx32 and Cx43, the branch order is consistent with the known divergence times between the animal species. This gives additional credence to the validity of the tree.

While the phylogenetic tree shows the relative relatedness of various connexins, it does not, except at species branches, give the time of the divergence of different connexins. These times can be estimated for connexins for which there are sequences from different species. Such sequences are available for Cx32 (HCx32, RCx32, XCx30) and Cx43

(RCx43, XCx43, CCx43, BCx43, HCx43). Using these, the divergence rate of a given connexin can be estimated by the method of Perler et al. (1980). In principle this method measures the rate from a plot of percent sequence divergence versus the known divergence time of the animal species used. The way in which percent divergence is calculated is based on changes at the nucleotide level that cause changes in the amino acid sequence. These changes are then corrected for multiple events, that is mutations that have changed back to the original residue, to give the final "corrected percent divergence." This number is similar to the percent amino acid identity between sequences when the differences are small, but is much larger when the differences are great. Percent corrected divergence of domains I and II were calculated for the connexins above except BCx43 and HCx43, which were not available at the time (Table 3). These divergence numbers are then plotted versus estimated divergence times for human/rodent of 75 MYr, mammal/bird of 275 MYr, and mammal/amphibian of 350 MYr (Dayhoff, 1972, Doolittle, et al., 1989, Perler, et al., 1980). Linear regression analysis, including the origin once for each curve, reveals

	XCx30	RCx32	HCx32	RCx43	CCx43	XCx43
XCx30		16	14	37	36	43
RCx32	10		1.2	49	50	54
HCx32	9.3	0.6		49	49	51
RCx43	38	35	35		2.7	12
CCx43	39	38	37	1.5		11
XCx43	37	34	33	3.1	4.8	

Table 3. Percent corrected divergence numbers for connexins used to estimate the rate of divergence of Cx32 and Cx43. The numbers above the diagonal correspond to DI and below to DII. These numbers were calculated from the amino acid replacements and are corrected for multiple mutations.

domain I divergence rates of 2.7%/100 MYr (R^2=0.99) for Cx32 and 1.1%/100 MYr (R^2=0.74) for Cx43. Domain II appears to change significantly faster, 4.3%/100 MYr (R^2=0.99) for Cx32 and 3.0%/100 MYr (R^2=0.72) for Cx43. These rates of 1-3% per 100

MYr for DI are similar to some of the most slowly evolving proteins known such as cytochrome C (c. 3%) and phosphoglyceraldehyde dehydrogenase (c. 2%) (Dayhoff,

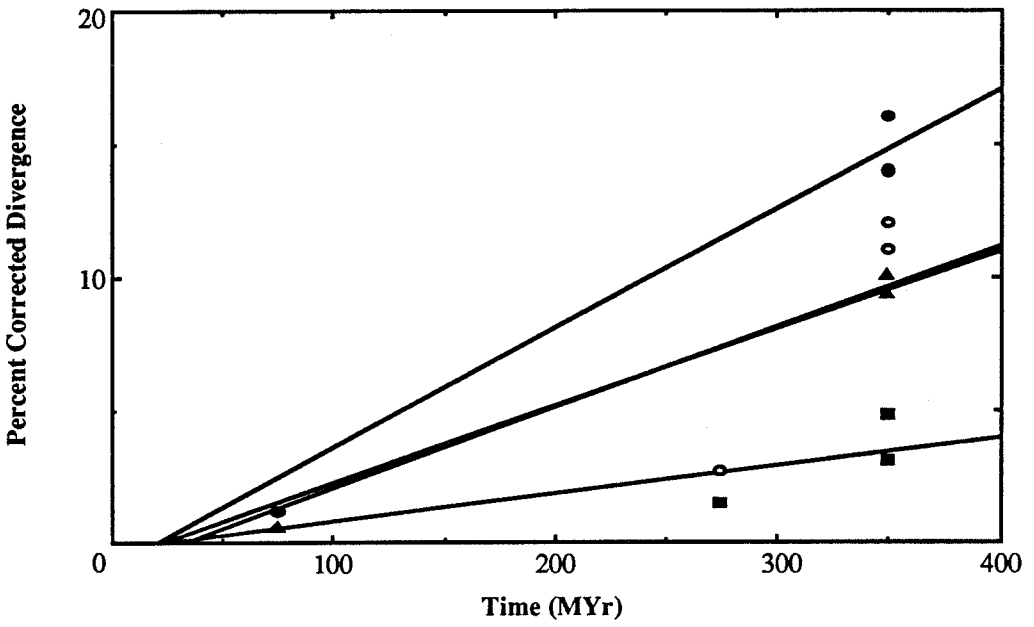


Figure 8. Divergence rates of DI and DII for Cx32 and Cx43. Cx43 DI and DII are represented by squares and open circles respectively, and Cx32 DI and DII by triangles and solid circles.

1972). In theory this would allow a "lookback time," i.e., time before the sequences mutate beyond recognition, of about 10 BYr, well before the estimated origin of life on our planet (Doolittle et al., 1987). The connexins are a curious mix of extremely slowly changing regions such as DI and DII, and rapidly diverging regions such as the IL and C-terminus.

Using these rates, 1.3-1.9 BYr would be required to account for the average 36% domain I and 46% domain II divergence between the Cx32 and Cx43 groups observed. This time can be compared estimated the prokaryote/eukaryote divergence at 1.8 BYr, the plant/animal at 1.0 BYr, and the vertebrate/invertebrate at 0.6 BYr (Doolittle, et al., 1989). The rate analysis therefore predicts that connexin homologues, for both Cx32 and Cx43, should be present in both plants and invertebrates. There is immunological evidence for connexin homologues in the plasmodesmata of plants, however, no connexin homologues

from invertebrates have been found despite extensive efforts by myself and others (D. Paul, personal communication). Besides providing evidence for the existence of an invertebrate connexin homologue, the analysis of divergence rates suggests an improved probe for low stringency screening. The most slowly evolving region of the known connexins is DI of Cx43, which would therefore be a better probe than Cx32 that was used in previous screens. Consistent with this suggestion is the fact that CCx45, the most diverged connexin so far isolated was cloned with a Cx43 probe (Beyer, 1990). Probes for polymerase chain reaction based screens should also probably be directed against this region.

Homology With Other Channels

To date the primary sequence of the connexins has appeared unique, and no homologies to other proteins have been identified. Unwin (1989) has suggested the putative pore lining segment (TM3) of the gap junction proteins is similar to several other ion channels including acetylcholine receptor, MIP and the GABA receptor. However, these alignments are not convincing and are not supported by statistical analysis of the sequences. Recently a series of papers has provided strong evidence for the H5 region of voltage gated K^+ channels as the sequence that lines the pore (Hartman et al., 1991; Yellen et al., 1991; Yool and Schwarz, 1991). These groups used similar approaches based on site directed mutagenesis and domain swapping. K^+ channels have previously been shown to be members of a superfamily of ion channels that includes, K^+, Na^+, Ca^{++}, and cGMP gated channels (Jan and Jan, 1990). This observation prompted me to examine the gap junction proteins for a region similar to H5.

Visual inspection of Cx32 immediately revealed a region with some sequence similarities to H5. These sequences were aligned with many of the channels from the Jan and Jan superfamily (Table 4, Figure 9). The alignment of connexins, which extends about 10 amino acids to the N-terminal side of H5, versus these other channels shows a strong sequence similarity between all the channels. There is a bias for negatively charged amino

Abrev.	Name		Accession #
Na/1	Rat Brain Sodium Channel	Repeat 1	a25019
Na/2		Repeat 2	
Na/3		Repeat 3	
Na/4		Repeat 4	
Ca/1	Rabbit Heart Calcium Channel	Repeat 1	s05054
Ca/2		Repeat 2	
Ca/3		Repeat 3	
Ca/4		Repeat 4	
ShA	*Drosophila* Potasium Channel		s00479
mbk1	Mouse Brain Potasium Channel		s06378
rck4	Rat Brain Potasium Channel		s06708
shab	*Drosophila* Potasium Channel		m32659
drk1	Rat Brain Potasium Channel		s05448
shaw	*Drosophila* Potasium Channel		m32661

Table 4. Names and sources of channel sequences compared with connexins. Accession numbers prefixed with a and s are for protein sequences from the PIR database (National Biomedical Research Foundation, Georgetown, Washington D.C.), and the m prefix indicates nucleotide sequence from Genbank (Intelligenetics, Mountain View, CA). The different domains from Na^+ and Ca^{++} channels are homologous internal repeats. The sequence used in the comparison is the H5 region from K^+ channels and approximately 10 residues preceding it, and the corresponding sequences from Na^+ and Ca^{++} channels as described by Jan and Jan (1990).

acids in the first 3-4 positions, followed by several positively charged residues. A pair tryptophans in the middle of the segment are conserved in many sequences. This is followed by a S/T that is conserved in 17 of the 22 positions, and 8 amino acids later a negatively charged D/E that is present in 19 of the 22 sequences. While any given pair of different channel sequences is not obviously similar, the several residues conserved in most sequences and general similarity suggests that they are homologous. The similarities between the β connexins (Cx26, Cx30, Cx31, Cx32) and the K^+ channels appear particularly strong. However, other pairs such as CCx45 and Ca/3 also show a great deal of similarity (Figure 9 B).

To provide further evidence that a pair of sequences are homologous, i.e., that they have a common ancestor, based on the observed sequence similarities is often a difficult proposition, particularly if the sequence similarity is weak and short as is the case here. There are well established methods for comparing two sequences, however, methods for

simultaneously aligning many sequences are poorly developed. Therefore, the sequences in Figure 9 were compared pairwise using the scrambled Needleman and Wunsch type

A

```
RCx31   AKLYSHPGKKHGGLWWTYLFSLIFKLIIEL
RCx26   EEIKTQKVRIEGSLWWTYTTSIFFRVIFEA
XCx30   AEVKKHKVKISGTLWWTYISSVFFRIIFEA
RCx32   EEVKRHKVHISGTLWWTYVISVVFRLLFEA
XCx38   YSSATKKIRIQGPLMCTYTTSVVFKSIFEA
CCX42   WDESGGKIILRGSLLNTYVYSILIRTAMEI
RCx43   GIEEHGKVKMRGGLLRTYIISILFKSVFEV
CCX45   AKHDGRRRIREDGLMRIYVLQLLVRATFEV
Na/1    GRNPNYGYTSFDTFSW.AFLSLFRLMTQDF
Na/2    TDCKLPRWHMNDFFHS.FLIVFRVLCGEWI
Na/3    TARWKNVKVNFDNVGF.GYLSLLQVATFKG
Na/4    REVGIDDMFNFETFGN.SMICLFQITTSAG
Ca/1    WDGPKHGITNFDNFAF.AMLTVFQCITMEG
Ca/2    FDEMQTRRSTFDNFPQ.SLLTVFQILTGED
Ca/3    PRSWENSKFDFDNVLA.AMMALFTVSTFEG
Ca/4    DTTEINRNNNFQTFPQ.AVLLLFRCATGEA
ShA     GSENSFFKSIPDAFWW.AVVTMTTVGYGDM
mbk1    EEAESHFSSIPDAFWW.AVVSMTTVGYGDM
rck4    DEPTTHFQSIPDAFWW.AVVTMTTVGYGDM
shab    DEKDTKFVSIPEAFWW.AGITMTTVGIGDM
drk1    DEDDTKFKSIPASFWW.ATITMTTVGYGDM
shaw    PNPHNDFNSIPLGLWW.ALVTMTTVGYGDM
```

B

```
CCx45   AKHDGRRRIREDGLMRIYVLQLLVRATFEV
Ca/3    PRSWENSKFDFDNVLA.AMMALFTVSTFEG
```

Figure 9. Multiple alignment of 8 connexins, 6 K^+ channels, 4 domains of a Na^+ channel, and 4 domains of a Ca^{++} channel as defined in Tables 1 and 4. Amino acids conserved between the connexins and other channel proteins are bolded. The structurally similar residues D/E, K/R and S/T are considered identical, and the bolding criterion is any residues at a given position in the connexins must be identical with two in the other proteins.

alignments. This method aligns the two sequences and scores the alignment using one of several possible scoring matrices. The most common matrix is the MDM-78, or log odds matrix, that was used here. One sequence in the pair is then randomly scrambled, aligned, and scored in the same way. This process is repeated 100 times. The score of the original alignment is then compared with all the scrambled scores, and the results are recorded as

the number of standard deviations (S. D.) from the scores of the alignments of scrambled sequences. This method controls for compositional bias in the sequences.

The results of the pairwise comparison of all the channel sequences are shown in Table 5. As expected, each group shows similarities within themselves, 3.3-11.5 S.D. for connexins, 8.3-11.5 S.D. for K^+ channels, 1.5-4.4 S.D. for the Na^+ channel, and 3.7-7.0 S.D. for the Ca++ channel. Besides these, there are two clusters high similarity scores, the β connexins (Cx26-Cx32) with the K^+ channels, 1.5-4.1 S.D., and the Na^+ channel with the Ca^{++} channel 1.7-6.9 S.D. There are no obvious similarities between connexins and Na^+ or Ca^{++} channels. Many of the numbers for connexins compared with K+ are above the 3 S. D. cutoff often used to indicate homology, and the fact that some connexins appear more similar to K^+ channels than Na^+ or Ca^{++} compared to K^+ channels, which have been shown to be related (Jan and Jan, 1990), further suggests that the connexin sequence is related to K^+ channels.

The statistical support for the homology between connexins and K^+ channels, and by extension the other members of the ion channel superfamily, is further strengthened by the context of the sequences. The H5 region of K^+ channels precedes by several amino acids S6, which is thought to span the membrane (for review see Catterall, 1988). The homologous region in the connexins (CxH3) is very hydrophobic, and in most models spans the membrane. However, it too precedes a hydrophobic segment that could span the membrane, placing CxH3 in the same position as H5 relative to the membrane, except on the opposite side. It is interesting in this context, that Yool and Schwarz (1991) have suggested that H5 may form a Beta sheet that lines the pore, similar to the Beta barrel proposed for porin and the voltage dependent anion selective channel (Blachly-Dyson et al., 1990; Keffel et al., 1985). Based on X-ray diffraction, the gap junction has a significant b-sheet component that appears to line the pore (Caspar et al., 1988). This β-sheet component is not accounted for in most current models of connexin organization. It is also possible that the a reevaluation of the K^+ channel topology is necessary and that H5

Table 5. Scores from pairwise scrambled comparisons of connexins and other channel proteins. The sequences used are those in Figure 9. Alignments were by the method of Needleman and Wunsch (1970), scored using the MDM-78 matrix and scrambled 100 times.

	RCx26	RCx31	XCx30	RCx32	XCx38	CCx42	RCx43	CCx45	ShA	mbk1	rck4	shab	drk1	shaw	Na-1	Na-2	Na-3	Na-4	Ca-1	Ca-2	Ca-3	Ca-4
RCx26		6.7	11.2	11.0	7.9	3.4	9.0	4.7	1.7	3.4	2.9	2.8	3.0	2.4	0.8	0.7	0.1	-0.2	0.1	0.3	0.2	-0.2
RCx31			7.7	9.1	4.4	3.3	5.7	4.4	1.5	2.5	2.4	2.0	1.7	2.6	1.5	0.1	-0.5	-1.2	0	-0.4	-1.0	-0.9
XCx30				11.5	7.5	3.6	7.9	4.7	1.9	3.7	3.4	3.0	2.9	3.0	2.2	0.5	0.4	-0.4	0.6	-0.2	0.4	-0.5
RCx32					6.3	3.4	8.6	4.6	2.0	4.1	4.0	2.9	3.3	2.6	2.2	1.7	0.3	0	1.0	0.6	-0.4	0.1
XCx38						5.7	7.9	5.7	-0.7	0.6	-0.3	-0.1	0.6	-0.2	-1.5	0.3	1.1	-0.7	1.8	0.2	1.3	0.4
CCx42							8.3	6.7	-0.1	0	0	1.0	-0.1	-1.0	0.7	0	0.5	2.5	0.3	1.0	0.5	-1.0
RCx43								7.4	-0.1	0.8	-0.4	0.8	1.1	0.9	0.8	0.4	0.5	-0.8	1.1	0.6	1.3	1.4
CCx45									0.4	0.8	0.1	0.2	0.5	0.5	0.9	1.1	1.8	0.4	0.7	1.6	2.8	0.8
ShA										11.5	10.1	8.3	9.9	8.3	2.7	0.7	-1.3	-0.6	-0.7	0.1	-0.7	0
mbk1											10.5	10.8	11.9	9.6	2.1	0.2	-0.7	-0.2	-0.2	-1.0	-0.6	1.6
rck4												9.8	10.5	9.0	2.5	0.2	-0.1	-0.6	0.1	0.5	-0.3	0.9
shab													9.5	8.4	3.0	-0.5	-0.6	-0.9	-0.3	0.6	-1.2	0.3
drk1														9.2	1.7	0.7	-0.6	-1.2	-1.0	0.5	-0.9	0.5
shaw															2.7	0	0.1	-0.4	0.1	0.7	0	0.3
Na-1																2.9	3.3	3.9	4.9	4.2	2.4	4.8
Na-2																	1.5	2.0	1.9	5.3	1.7	3.3
Na-3																		4.4	4.8	4.2	6.9	2.7
Na-4																			5.8	6.1	5.1	7.6
Ca-1																				7.0	4.7	6.0
Ca-2																					3.7	7.6
Ca-3																						3.7
Ca-4																						

actually forms a transmembrane α-helix that lines the pore. The analysis here suggests that CxH3 and H5 are homologous, and that the gap junction and other channels described here share a common ancestor.

References

Ausubel, F. M., et al., Ed. (1988). Current protocols in molecular biology. New York, Greene and Associates and Wiley Interscience.

Beyer, E. C. (1990). "Molecular cloning and developmental expression of two chick embryo gap junction protein." J. Biol. Chem. **265**(24): 14439-14443.

Beyer, E. C., D. A. Goodenough and D. L. Paul (1988). The connexins, a family of related gap junction proteins. Gap Junctions. New York, Alan R. Liss, Inc. 167-176.

Beyer, E. C., J. Kistler, D. L. Paul and D. A. Goodenough (1989). "Antisera directed against Connexin43 peptides react with a 43-kD protein localized to gap junctions in myocardium and other tissues." J. Cell Biol. **108**: 595-605.

Beyer, E. C., D. L. Paul and D. A. Goodenough (1987). "Connexin43: A protein from rat heart homologous to a gap junction protein from liver." J. Cell Biol. **105**(6): 2621-2629.

Beyer, E. C., D. L. Paul and D. A. Goodenough (1990). "Connexin family of gap junction proteins." J. Membr. Biol. **116**: 187-194.

Blachly-Dyson, E., S. Peng, M. Colombini and M. Forte (1990). "Selectivity changes in site-directed mutants of the VDAC ion channel - structural implications." Science **247**: 1233-1236.

Caspar, D. L. D., et al. (1988). Gap junction structure. Gap Junctions. New York, Alan R. Liss. 117-133.

Catterall, W. A. (1988). "Structure and function of voltage-sensitive ion channels." Science **242**: 50-61.

Dayhoff, M. O., Ed. (1972). Atlas of protein sequence and structure 1972. Silver Spring, Maryland, The National Biomedical Research Foundation.

Dayhoff, M. O., Ed. (1978). Atlas of protein sequence and structure: Supplement 3. Silver Spring, Maryland, The National Biomedical Research Foundation.

Dermietzel, R., et al. (1984). "Gap junctions in several tissues share antigenic determinants with liver gap junctions." EMBO J. **3**(10): 2261-2270.

Dermietzel, R., et al. (1989). "Differential expression of the three gap junction proteins in developing and mature brain tissues." Proc. Nat. Acad. Sci. USA **86**: 10148-10152.

Devereux, J., P. Haeberli and O. Smithies (1984). "A complete set of sequence analysis programs for the VAX." Nucl. Acid Res. **12**(1): 387-394.

Dohlman, H. G., et al. (1990). "Role of extracellular disulfide-bonded cysteines in the ligand binding function of the β_2-adrenergic receptor." Biochem. **29**: 2335-2342.

Doolittle, R. F. (1987). Of urfs and orfs. Mill Valley, University Science Books.

Doolittle, R. F., K. L. Anderson and D.-F. Feng (1989). Estimating the prokaryote-eukaryote divergence time from protein sequences. The hierarchy of life. New York, Exerpta Medica. 73-85.

Ebihara, L., et al. (1989). "Cloning and expression of a *Xenopus* embryonic gap junction protein." Science **243**: 1194-1195.

Feng, D. F., M. S. Johnson and R. F. Doolittle (1985). "Aligning amino acid sequences: comparison of commonly used methods." J. Mol. Evol. **21**: 112-125.

Finbow, M. E., J. Shuttleworth, A. E. Hamilton and J. D. Pitts (1984). "Analysis of the vertebrate gap junction protein." EMBO J. **2**: 1479-1486.

Finbow, M. E., S. B. Yancey, R. G. Johnson and J. Revel (1980). "Independent lines of evidence suggesting a major gap junctional protein with a molecular weight of 26,000." Proc. Nat. Acad. Sci. U.S.A. **77**: 970-974.

Fishman, G. I., D. C. Spray and L. A. Leinwand (1990). "Molecular characterization and functional expression of the human cardiac gap junction channel." J. Cell Biol. **111**: 589-598.

Gimlich, R. L., N. Kumar and N. B. Gilula (1988). "Sequence and developmental expression of mRNA coding for a gap junction protein in *Xenopus*." J. Cell Biol. **107**: 1065-1073.

Gimlich, R. L., N. Kumar and N. B. Gilula (1990). "Differential regulation of the levels of three gap junction mRNAs in Xenopus embryos." J. Cell Biol. **110**: 597-605.

Hartman, H. A., et al. (1991). "Exchange of conduction pathways between two related K^+ channels." Science **251**: 942-944.

Heijne, G. v. and Y. Gavel (1988). "Topogenic signals in integral membrane proteins." Eur. J. Biochem. **174**: 671-678.

Henderson, D., H. Eibl and K. Weber (1979). "Structure and biochemistry of mouse hepatic gap junctions." J. Mol. Biol. **132**: 193-218.

Hertzberg, E. L. and R. V. Skibbens (1984). "A protein homologous to the 27,000 dalton liver gap junction protein is present in a wide variety of species and tissues." Cell **39**: 61-69.

Higgins, D. G. and P. M. Sharp (1988). "CLUSTAL: a package for performing multiple sequence alignment on a microcomputer." Gene **73**: 237-244.

Hoh, J. H., S. A. John and J.-P. Revel (1991). "Molecular cloning and characterization of a new member of the gap junction gene family, connexin-31." J. Biol. Chem. **266**: 6524-6531.

Hoh, J. H. and J.-P. Revel (1991). "A strain specific RFLP near the rat connexin-32 (Cx32) gap junction gene." Mammalian Genome In Press.

Jan, L. Y. and Y. N. Jan (1990). "A superfamily of ion channels." Nature **345**: 672.

Karnik, S. S. and H. G. Khorana (1990). "Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187." J. Biol. Chem. **265**(29): 17520-17524.

Keffel, B., R. M. Garavito, W. Banmeister and J. P. Rosenbusch (1985). "Secondary structure of a channel-forming protein: porin from *E. coli*." EMBO J. **4**: 1589-1592.

Kumar, N. M. and N. B. Gilula (1986). "Cloning and characterization of the human and rat liver cDNAs coding for a gap junction protein." J. Cell Biol. **103**: 767-776.

Laird, D. W., K. L. Puranam and J.-P. Revel (1991). "Turnover and phosphorylation dynamics of connexin43 gap junction protein in cultured cardiac myocytes." Biochem. J. **273**: 67-72.

Laird, D. W. and J.-P. Revel (1990). "Biochemical and immunochemical analysis of the arrangement of connexin43 in rat heart gap junction membranes." J. Cell Sci. **97**: 109-117.

Larson, D. M., C. C. Haudenschild and E. C. Beyer (1990). "Gap junction messenger RNA expression by vascular wall cells." Circ. Res. **66**: 1074-1080.

Lash, J. A., E. S. Critser and M. L. Pressler (1990). "Cloning of a gap junctional protein from vascular smooth muscle and expression in two-cell mouse embryos." J. Biol. Chem. **265**(22): 13113-13117.

MacDonald, R. J., G. H. Swift, A. E. Przybyla and J. M. Chirgwin (1987). "Isolation of RNA using guanidinium salts." Meth. Enzymol. **152**: 219-227.

Manjunath, C. K., G. E. Goings and E. Page (1984). "Cytoplasmic surface and intermembrane components of rat heart gap junctional proteins." Am. J. Physiol. **246**: H865-H875.

Micevych, P. E. and L. Abelson (1991). "Distribution of mRNAs coding for liver and heart gap junction proteins in the rat central nervous system." J. Comp. Neurol. **305**: 96-118.

Miller, T., G. Dahl and R. Werner (1988). "Structure of a gap junction gene: Rat connexin-32." Biosci. Rep. **8**(5): 455-464.

Murata, M., J. S. Richardson and J. L. Sussman (1985). "Simultaneous comparison of three sequences." Proc. Nat. Acad. Sci. USA **82**: 3073-3077.

Musil, L. S., E. C. Beyer and D. A. Goodenough (1990). "Expression of the gap junction protein connexin43 in embryonic chick lens: molecular cloning, ultrastructural localization, and post-translational phosphorylation." J. Membrane Biol. **116**: 163-175.

Needleman, S. B. and C. D. Wunsch (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins." J. Mol. Biol. **48**: 443-453.

Nicholson, B., et al. (1987). "Two homologous protein components of hepatic gap junctions." Nature **329**(6141): 732-734.

Nicholson, B. J., et al. (1985). "The Mr 28,000 gap junction proteins from rat heart and liver are different but related." J. Biol. Chem. **260**(11): 6514-6517.

Nicholson, B. J., et al. (1981). "Rat liver gap junction protein: properties and partial sequence." Proc. Nat. Acad. Sci. USA **78**(12): 7594-7598.

Paul, D. L. (1986). "Molecular cloning of cDNA for rat liver gap junction protein." J. Cell Biol. **103**: 123-134.

Perler, F., et al. (1980). "The evolution of genes: the chicken preproinsulin gene." Cell (20): 555-566.

Saez, J., et al. (1986). "cAMP increases junctional conductance and stimulates phosphorylation of the principal gap junction polypeptide." Proc. Nat. Acad. Sci. U.S.A **83**: 2473.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc. Nat. Acad. Sci. U.S.A. **74**: 5463-5467.

Sargent, T. D., et al. (1979). "The rat serum albumin gene: Analysis of cloned sequences." Proc. Nat. Acad. Sci. U.S.A. **76**(7): 3256-3260.

Sneath, P. H. A. and R. R. Sokal (1973). Numerical Taxonomy. New York, Freeman.

Strathmann, M., et al. (1991). "Transposon-facilitated DNA sequencing." Proc. Nat Acad. Sci. USA **88**: 1247-1250.

Strauss, W. M. (1988). Preparation of genomic DNA from mammalian tissue. In: Current protocols in molecular biology. New York, Greene and Associates and Wiley Interscience. 2.2.1-2.2.3.

Swenson, K. I., J. R. Jordan, E. C. Beyer and D. L. Paul (1989). "Formation of gap junctions by expression of connexins in *Xenopus* oocyte pairs." Cell **57**: 145-155.

Tabor, S. and C. C. Richardson (1987). "DNA sequence analysis with modified bacteriophage T7 DNA polymerase." Proc. Nat. Acad. Sci. U.S.A. **84**: 4767-4771.

Traub, O., et al. (1989). "Comparative characterization of the 21-kD and 26-kD gap junction proteins in murine liver and cultured hepatocytes." J. Cell Biol. **108**: 1039-1051.

Unwin, N. (1989). "The structure of ion channels in membranes of excitable cells." Neuron **3**: 665-676.

Wilbur, W. J. and D. J. Lipman (1983). "Rapid similarity searches of nucleic acid and protein data banks." Proc. Nat. Acad. Sci. U.S.A. **80**: 726-730.

Yamamoto, T., A. Ochalski, E. L. Hertzberg and J. I. Nagy (1990). "LM and EM immunolocalization of the gap junctional protein connexin 43 in rat brain." Brain Res. **508**: 313-319.

Yellen, G., M. E. Jurman, T. Abramson and R. MacKinnon (1991). "Mutations affecting internal TEA blockade identify the probable pore-forming region of a K^+ channel." Science **251**: 939-942.

Yool, A. J. and T. L. Schwarz (1991). "Alteration of ionic selectivity of a K^+ channel by mutation of the H5 region." Nature **349**: 700-704.

Zhang, J.-T. and B. J. Nicholson (1989). "Sequence and tissue distribution of a second protein of hepatic gap junctions, Cx26, As Deduced from its cDNA." J. Cell Biol. **109**(6, pt2): 3391-3401.

Zervos, A. S., J. Hope and H. W. Evans (1985). "Preparation of a gap junction fraction from uteri of pregnant rats: the 28-kD polypeptides of uterus, liver and heart gap junctions are homologous." J. Cell Biol. **101**: 1363-1370.

Zimmer, D., C. R. Green, W. H. Evans and N. B. Gilula (1987). "Topological analysis of the major protein in isolated intact rat liver gap junctions and gap junction-derived single membrane structures." J. Biol. Chem. **262**: 7751-7763.

Chapter 4

Molecular Cloning and Charaterization of a New Member of The Gap Junction Gene Family, Connexin-31

THE JOURNAL OF BIOLOGICAL CHEMISTRY 147 Vol. 266, No. 10, Issue of April 5, pp. 6524-6531, 1991

Printed in U.S.A.

Molecular Cloning and Characterization of a New Member of the Gap Junction Gene Family, Connexin-31*

(Received for publication, September 24, 1991)

Jan H. Hoh, Scott A. John, and Jean-Paul Revel‡

From the Division of Biology 156-29, California Institute of Technology, Pasadena, California 91125

A new member of the connexin gene family has been identified and designated rat connexin-31 (Cx31) based on its predicted molecular mass of 30,960 daltons. Cx31 is 270 amino acids long and is coded for by a single copy gene. It is expressed as a 1.7-kilobase mRNA that is detected in placenta, Harderian gland, skin, and eye. Cx31 is highly conserved and can be detected in species as distantly related to rat as *Xenopus laevis*. It exhibits extensive sequence similarity to the previously identified connexins, 58, 50, and 40% amino acid identity to Cx26, Cx32, and Cx43, respectively. When conservation of predicted phosphorylation sites is used to adjust the alignment of Cx31 to other connexins, a unique alignment of three predicted protein kinase C phosphorylation sites near the carboxyl terminus of Cx31 with three sites at the carboxyl terminus of Cx43 is revealed.

The gap junction is a structure composed of two closely apposed plasma membranes with a tightly packed array of cell to cell channels (Revel and Karnovsky, 1967). The physiology of the channels has been characterized in some detail in several experimental systems (Loewenstein, 1981; Spray and Bennett, 1985). In vertebrates, they have been shown to provide a low resistance electrical pathway between cells and to allow the passage of molecules <1000 Da with little or no selectivity (Flagg-Newton *et al.*, 1979). These channels are thought to have many important biological functions including the regulation of growth control (Mehta *et al.*, 1986), synchronization of cellular activity including synchronized contraction of myocardial cells (Barr *et al.*, 1965), regulation of embryonic development and differentiation (Pitts, 1978), and metabolic homeostasis (Sheridan *et al.*, 1979).

Gap junctions were identified first in the gold fish Mauthner cell by Robertson (Robertson, 1963). They have now been identified in almost every metazoan that has been examined, and they have also been described morphologically in a wide variety of tissues. This wide distribution and conservation of structure suggests that it is involved in a fundamental biological function shared by all multicellular animals.

Despite their wide distribution, gap junctions have been isolated only from a few organs in a few organisms. Hepatic gap junctions from mouse and rat were isolated, and two proteins of M_r 21,000 and M_r 28,000 were identified as major structural components (Finbow *et al.*, 1980; Henderson *et al.*, 1979). Cardiac gap junctions isolated from rat, rabbit, and mouse have been shown to have a principal protein of M_r 47,000 (Kensler and Goodenough, 1980; Manjunath *et al.*, 1982 and 1984). Gap junctions have also been isolated from rat and bovine lens, and a protein of M_r 70,000 has been identified as its major component in the lens fiber cell (Kistler *et al.*, 1988). This protein has the same amino-terminal sequence as a protein with a predicted molecular mass of 46,000 Da (Beyer *et al.*, 1988). The relationship between the two proteins is not yet understood. Finally, gap junctions from the arthropods *Nephrops* and *Drosophila* have been isolated, and several putative protein components have been identified (Berdan and Gilula, 1988; Buultjens *et al.*, 1988; Ryerse, 1989).

The major structural proteins of gap junctions identified by isolation are now called connexins. They are members of a gene family that was first identified on the basis of protein sequence (Nicholson *et al.*, 1985; Nicholson *et al.*, 1981). Several cDNAs coding for connexins now have been isolated. In rat, the three proteins for which cDNAs have been described are designated Cx26,[1] Cx32, and Cx43 based on their predicted molecular mass (Beyer *et al.*, 1987; Paul, 1986; Zhang and Nicholson, 1989). These cDNAs correspond to the hepatic M_r 21,000 and M_r 28,000 proteins and the cardiac M_r 47,000 protein, respectively. Homologues to these connexins from several other species have been isolated also (Gimlich *et al.*, 1990; Lash *et al.*, 1990; Musil *et al.*, 1990). Another member of the connexin gene family expressed in early *Xenopus* development, Cx38, has also been identified (Ebihara *et al.*, 1989). Through the use of the cDNAs as probes, the distribution of the various connexins has been described in several tissues and cell lines (Beyer *et al.*, 1987; Crow, *et al.*, 1990; Larson *et al.*, 1990; Musil *et al.*, 1990; Zhang and Nicholson, 1989). All connexins identified to date have unique distributions. Some tissues and organs have a single connexin while others have more than one, but no two are always found together. In rodent hepatocytes it appears that a single cell coexpresses both Cx26 and Cx32 (Nicholson *et al.*, 1987; Traub *et al.*, 1989).

Currently there is little known about the genes that code for the connexins. The gene coding for one connexin, rat Cx32, has been isolated and described (Miller *et al.*, 1988). It is a single copy gene with two exons, of which the second contains the entire coding region for the protein. The genes coding for Cx26 and Cx43 are also single copy (Musil *et al.*,

* This work was supported in part by Grants HL37109 and BRSG-RR07003 from the National Institutes of Health and the Ruddock Fund. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M59936.

‡ To whom correspondence should be addressed.

[1] The abbreviations used are: all connexins are abbreviated Cx followed by their predicted molecular mass. When appropriate, the species from which the connexin was identified is designated by the prefixes C, H, R, or X for chicken, human, rat, and *Xenopus*, respectively. Divergence times are given in million years (MYr) or billion years (BYr). Every 1000 bases is a kilobase (kb). DNA melting temperature is designated T_m.

1990; Zhang and Nicholson, 1989). The gene coding for Cx26 appears to have at least two exons, one of which contains the entire coding region as determined by Southern blotting (Zhang and Nicholson, 1989).

All connexins cloned to date were isolated first from cDNA libraries. The current study uses instead low stringency screening of a rat genomic library to isolate genes that code for connexin homologues to understand further the diversity, distribution, and phylogeny of this family of proteins. Using this alternate strategy we have identified a new member of the connexin gene family designated Cx31. Characterization of the gene reveals a 270-amino acid open reading frame with a high degree of sequence similarity to other connexins. This connexin shares many features with previously identified members of the family, including an alignment and conservation of three potential phosphorylation sites at the extreme carboxyl terminus of Cx43. We have also produced the first phylogenetic tree for the known connexins, which shows that Cx31 is closely related to Cx26 and Cx32 and distantly related to Cx43.

EXPERIMENTAL PROCEDURES

Materials—Restriction endonucleases and T7 RNA polymerase were obtained from Boehringer Mannheim. Modified T7 DNA polymerase and sequencing reagents were purchased from U. S. Biochemical Corp. Chemicals were from Sigma or Boehringer Mannheim. Sprague-Dawley rats were supplied by Simonsen Laboratories (Gilroy, CA).

Isolation of Connexin Homologue Gene—To identify connexin homologues we screened a Sprague-Dawley rat genomic *Eco*RI partial digest Charon 4A library using the Cx32 cDNA as a probe under low stringency conditions (Paul, 1986; Sargent *et al.*, 1979). Briefly, the Cx32 cDNA in pGEM-3 (Promega Biotec, Madison, WI) was linearized in the polylinker on the 5′ side of the insert with *Bam*HI, and an antisense RNA probe with specific activity 2.5 μCi/ng was transcribed using T7 RNA polymerase and [α-^{32}P]UTP (Melton *et al.*, 1984; Tabor and Richardson, 1985). The RNA probe was used to screen approximately five genome equivalents (1×10^6 clones) of the genomic library plated at 5×10^4 plaques per 150-mm plate and replicated in duplicate onto Hybond-N (Amersham Corp.) membranes. Prehybridization and hybridization were carried out in 5 × SSPE (1 × SSPE = 150 mM NaCl, 10 mM $Na_2H(PO_4)$, and 0.1 mM EDTA, pH 7.2), 5 × Denhardt's solution (50 × Denhardt's = 1% w/v bovine serum albumin, 1% w/v polyvinylpyrrolidone, and 1% w/v Ficoll), 30% deionized formamide, 20 μg/ml poly(A) RNA, 30 μg/ml yeast tRNA, and 0.5–2 ng/ml RNA probe at 45 °C for 6 and 36 h, respectively. Membranes were treated at increasing stringency to a final wash of 1 × SSPE at 60 °C for 4 h and subjected to autoradiography. Positive clones were identified, and the procedure was repeated until clonal purity was achieved. The genomic clones isolated were characterized further by restriction site analysis and Southern blotting using the rat Cx26, Cx32, and Cx43 cDNAs as probes (Beyer *et al.*, 1987; Paul, 1986; Zhang and Nicholson, 1989). A 4.4-kb *Eco*RI fragment from one λ clone, RGJ21, which cross-hybridized with all three probes, was subcloned into pBluescript II KS(+) (Stratagene Cloning Systems, La Jolla, CA), characterized in more detail, and both strands of a 1.1-kb *Eco*RI-*Sac*II fragment were sequenced by standard dideoxy sequencing (Sanger *et al.*, 1977; Tabor and Richardson, 1987).

Genomic Southern Blots—Southern blots of genomic DNA from rat (Sprague-Dawley), mouse (Balb/c), pig, and frog (*Xenopus laevis*) were performed under high and moderate stringency conditions. The rat DNA was isolated as described by Strauss (Strauss, 1988), the mouse and pig DNA were obtained commercially (Clontech, Palo Alto, CA), and the *Xenopus* DNA was provided by R. Wagner (California Institute of Technology). The rat DNA was digested separately with *Bgl*I, *Eco*RI, *Hin*dIII, *Kpn*I, *Nhe*I, *Sac*II or *Xba*I, separated on a 0.8% agarose gel, and capillary transferred onto Hybond-N as described by Maniatis *et al.* (1982). The blot was probed with the Cx31 *Eco*RI-*Sac*II fragment random primer labeled to a specific activity of 1×10^9 cpm/μg with [α-^{32}P]dATP (Feinberg and Vogelstein, 1984). High stringency prehybridization and hybridization were carried out in 5 × SSPE, 5 × Denhardt's solution, 50% deionized formamide, 1 mM sodium pyrophosphate, 1 mM ATP, 0.1% sodium dodecyl sulfate, 20 μg/ml salmon sperm DNA, 30 μg/ml yeast tRNA, and 2–5 ng/ml probe DNA at 45 °C for 24 and 36 h, respectively. Blots were washed at increasing stringency to a final 0.1 × SSPE at 68 °C for 4 h. For the moderate stringency "zoo" blot all DNAs were digested with *Eco*RI or *Hin*dIII and treated as above, except that the hybridization conditions were 5 × SSPE, 5 × Denhardt's, 1 mM sodium pyrophosphate, 1 mM ATP, 0.1% sodium dodecyl sulfate, 20 μg/ml salmon sperm DNA, 30 μg/ml yeast tRNA, and 2–5 ng/ml probe DNA at 60 °C. The zoo blots were washed at increasing stringency to a final 1 × SSPE at 65 °C for 2–10 h. T_m values were estimated by the method of Meinkoth (Meinkoth and Wahl, 1984). All fragments were sized using the GEL[2] regression program.

Northern Blots of Organ RNA—To determine the expression pattern of RCx31 we isolated total RNA from several tissues and organs by the guanidinium thiocyanate method or the modified guanidinium thiocyanate method that includes a centrifugation through CsCl (MacDonald *et al.*, 1987). The sources of RNA were liver, heart, tail skin including the dermis, tail connective tissue (everything left after the skin was removed), Harderian gland, eye, placenta from a 19-day pregnant animal, epididymal fat pads, brain without the cerebellum, blood, stomach, femur including marrow, pancreas, spleen, ovary, uterus, thigh skeletal muscle, lung, duodenum, testis, and kidney, all from both male and female Sprague-Dawley rats. Northern blots were performed by electrophoresing 10–20 mg of each RNA in a formaldehyde gel. The separated RNA was capillary transferred onto Hybond-N and probed under conditions identical with the high stringency Southern blots except that the temperature was increased to 48–50 °C (Ausubel *et al.*, 1988).

Analysis of RCx31 Sequence—All computer analyses were carried out with either PC-Gene version 6.01 or 6.25 (PCG) running on an Epson Equity II or the University of Wisconsin Genetics Computer Group package version 6.2 (GCG) (Devereux *et al.*, 1984) running on a VAX Station 3100 (Model m38) unless otherwise specified. All parameters are default unless otherwise indicated. Kyte and Doolittle hydropathy analysis was carried out on RCx31 using a 15-amino acid window in SOAP (PCG) (Kyte and Doolittle, 1982). The RCx31 sequence was scanned for known consensus sites for post-translational protein modification with PROSITE (PCG). Version 3.0 of PROSITE detects possible glycosylation, phosphorylation, sulfonation, amidation, fatty acylation, hydroxylation, carboxylation, phosphopantetheine attachment, and farnesyl group binding sites.

Multiple Sequence Alignment—Multiple alignments of the new rat Cx31 with protein sequences for rat Cx43, Cx32, and Cx26; chicken Cx43; *Xenopus* Cx43, Cx30, and Cx38; and human Cx32 were generated three sequences at a time using the program ALP3.[3] The ALP3 algorithm has been described by Murata *et al.* (1985). The triple alignments were compiled and adjusted by eye, taking into account the predicted phosphorylation sites, using the program LINEUP (GCG).

Construction of Phylogenetic Tree—To produce a phylogenetic tree and and estimate divergence times, the connexins were divided into two homology domains; the first domain was from amino acid 2 to 99 in the Cx31, the second domain corresponded to positions 125–153, 155–165, and 168–211 in Cx31 and the aligned domains in the other connexins (see Fig. 6). Phylogenetic trees where generated using the program CLUSTAL (PCG) (Higgins and Sharp, 1988) for both domains I and II using rat Cx26, Cx31, Cx32, Cx43, and *Xenopus* Cx38. This program generates all possible pair-wise alignments using the algorithm of Wilbur and Lipman (1983). It then generates dendrograms using the unweighted pair group maximum averages method of Sneath and Sokal (1973). Divergence rates for Cx32 and Cx43 were estimated by linear regression analysis of plots of the percent corrected divergence for RCx32, HCx32, and XCx30; and RCx43, CCx43, and XCx43 *versus* divergence times using the method of Perler (Perler *et al.*, 1980). The times used are human/rodent 75 MYr, mammal/bird 275 MYr, and mammal/amphibian 350 MYr (Dayhoff, 1972; Doolittle *et al.*, 1989; Perler *et al.*, 1980).

RESULTS

Isolation of a New Connexin Gene—Thirty genomic clones that cross-hybridized to the RCx32 cDNA probe were isolated. Southern blotting and restriction site analysis of these clones

[2] GEL is a public domain program available from Intelligenetics, Mountain View, CA.

[3] ALP3 is a public domain program available from Intelligenetics, Mountain View, CA.

revealed that six corresponded to two different polymorphisms of the RCx32 gene.[4] A single clone designated RGJ21 that cross-hybridized with RCx26, RCx32, and RCx43 was identified and restriction-mapped (Fig. 1). The remaining clones are currently under further investigation, but three others that have been examined closely reveal no connexin homologues. The 4.4-kb fragment containing the homologous sequence was subcloned, and 982 bases between the *Eco*RI and *Sac*II sites were sequenced. Fig. 2 shows the nucleotide sequence and the translation of the nucleotide sequence that revealed a 270-amino acid open reading frame coding for a protein with a predicted molecular mass of 30,960 Da. The predicted protein exhibits a high degree of similarity to previously characterized connexins, 58, 50, and 40% amino acid identity, and 65, 58, and 51% nucleotide identity to RCx26, RCx32, and RCx43, respectively. Based on its homology to other connexins and predicted characteristics we designate this protein rat connexin-31 (RCx31).

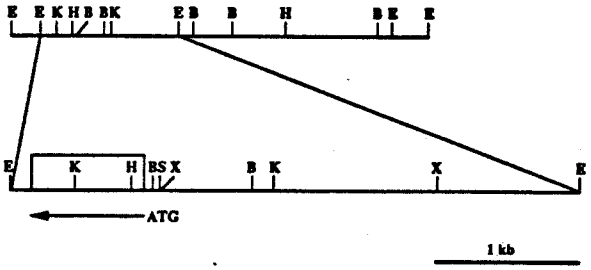


FIG. 1. **Restriction map of the genomic clone RGJ21.** The 4.4-kb *Eco*RI fragment that contains the gene coding for Cx31 is enlarged. The open reading frame is *boxed*, and the ATG start codon and coding direction of the protein are shown. The map was generated by single and double digests with the restriction endonucleases *Eco*RI (*E*), *Bam*HI (*B*), *Hin*dIII (*H*), *Kpn*I (*K*), *Sac*II (*S*), and *Xba*I (*X*).

[4] Hoh, J. H., and Revel, J.-P. (1991) *Mamm. Genome*, in press.

Genomic Southern Blots—High stringency Southern blots of rat genomic DNA digested with seven different enzymes probed with Cx31 all revealed only a single band, including the 1.6-kb *Kpn*I and 4.4-kb *Eco*RI fragments predicted from the genomic clone (Fig. 3). To control for cross-hybridization the same blot was hybridized with the RCx32 cDNA. It produced no bands identical with the Cx31 blot. This would suggest that RCx31 is a single copy gene. Southern blots of rat, mouse, pig, and frog genomic DNA were probed with Cx31 under conditions that allow sequences of greater than 75% nucleotide identity to be detected based on the estimated T_m (Fig. 4). Single *Eco*RI bands of 4.4, 4.1, 1.0, and 3.2 kb are seen in the rat, mouse, pig, and frog DNAs, respectively. Single *Hin*dIII bands of 3.0, 18, and 4.1 kb are also seen in the rat, mouse, and frog DNAs, respectively. In addition, a strong 3.3-kb *Hin*dIII band and three larger but weaker bands are seen in the pig DNA. This suggests that Cx31 is highly conserved and present in these species.

Distribution of Cx31 mRNA—A Northern blot of RNA

```
GGTGAATCACTAGAATAATGGGTTTTTTTCCTCCATTGTTCCCTCAGGTGCACAGCCCCCCGAACCCTGAGCAGGCACC

ATG GAT TGG AAG AAG CTT CAG GAC CTG TTG AGC GGT GTG AAC CAG TAC TCC ACC GCA TTC 60
MET Asp Trp Lys Lys Leu Gly Asn Leu Leu Ser Gly Val Asn Gln Tyr Ser Thr Ala Phe

GGG CGC ATC TGG CTG TCG GTA GTG TTC GTC TTC CGG GTG CTG GTG TAT GTG GTG GCT GCA 120
Gly Arg Ile Trp Leu Ser Val Val Phe Val Phe Arg Val Leu Val Tyr Val Val Ala Ala

GAG CGT GTG TGG GGC GAC GAG CAA AAA GAC TTT GAC TGT AAC ACC AGG CAG CCC GGT TGT 180
Glu Arg Val Trp Gly Asp Glu Gln Lys Asp Phe Asp Cys Asn Thr Arg Gln Pro Gly Cys

ACC AAC GTG TGC TAT GAC AAC TTC TTC CCC ATC TCC AAC ATC CGC CTC TGG GCC CTG CAG 240
Thr Asn Val Cys Tyr Asp Asn Phe Phe Pro Ile Ser Asn Ile Arg Leu Trp Ala Leu Gln

CTC ATC TTC GTC ACG TGT CCC TCT ATG CTG GTC ATC CTG CAC GTA GCC TAC CGC GAG GAG 300
Leu Ile Phe Val Thr Cys Pro Ser MET Leu Val Ile Leu His Val Ala Tyr Arg Glu Glu

CGG GAA CGG AAG CAT CGC CAG AAG CAC GGG GAG CAT TGC GCC AAA CTG TAC AGC CAC CCG 360
Arg Glu Arg Lys His Arg Gln Lys His Gly Glu His Cys Ala Lys Leu Tyr Ser His Pro

GGC AAG AAG CAC GGC GGC CTG TGG TGG ACC TAC CTG TTC AGT CTC ATC TTC AAG CTC ATC 420
Gly Lys Lys His Gly Gly Leu Trp Trp Thr Tyr Leu Phe Ser Leu Ile Phe Lys Leu Ile

ATT GAA TTG GTC TTC CTG TAT GTT CTA CAC ACG CTC TGG CAT GGC TTC ACC ATG CCG CGT 480
Ile Glu Leu Val Phe Leu Tyr Val Leu His Thr Leu Trp His Gly Phe Thr Met Pro Arg

CTG GTA CAG TGC GCC AGC GTG GTA CCT TGC CCC AAC ACC GTG GAT TGC TAC ATC GCT CGG 540
Leu Val Gln Cys Ala Ser Val Val Pro Cys Pro Asn Thr Val Asp Cys Tyr Ile Ala Arg

CCC ACG GAG AAG AAA GTC TTT ACC TAC TTC ATG GTA GGC GCG TCT GCC GTC TGC ATT ATT 600
Pro Thr Glu Lys Lys Val Phe Thr Tyr Phe Met Val Gly Ala Ser Ala Val Cys Ile Ile
    (P)

CTC ACC ATC TGT GAG ATC TGC TAC CTC ATC TTC CAC AGG ATT ATG CGA GGC CTG AGC AAG 660
Leu Thr Ile Cys Glu Ile Cys Tyr Leu Ile Phe His Arg Ile Met Arg Gly Leu Ser Lys
    (C)

GAC AAA TCG ACG AAG AGC ATC AGC TCC CCG AAG TCC TCC AGC CGG GCC TCC ACC TGT CGC 720
Asp Lys Ser Thr Lys Ser Ile Ser Ser Pro Lys Ser Ser Ser Arg Ala Ser Thr Cys Arg
        (P)                 (P)             (P)                 (P)

TGT CAC CAC AAG CTG CTG GAG AGT GGT GAT CTG GAA GCA GTA CCA GCC GAT GAC AAG CTG 780
Cys His His Lys Leu Leu Glu Ser Gly Asp Leu Glu Ala Val Pro Ala Asp Asp Lys Leu

CAG GCT TCA GCG CCT AGC CTG ACC CCC ATT TAA CCACGCGCTGGAGAAGGGGTGAGGCTGGGAGGTG
Gln Ala Ser Ala Pro Ser Leu Thr Pro Ile ---

TGGAGGGGTCCTGGGGTGCTGAGTGCCCCCCACTTTGAATTCCTGCAG
```

FIG. 2. **Nucleotide sequence and predicted protein sequence for Cx31.** Predicted protein kinase C phosphorylation and casein kinase II sites as predicted by PROSITE are marked *P* and *C*, respectively.

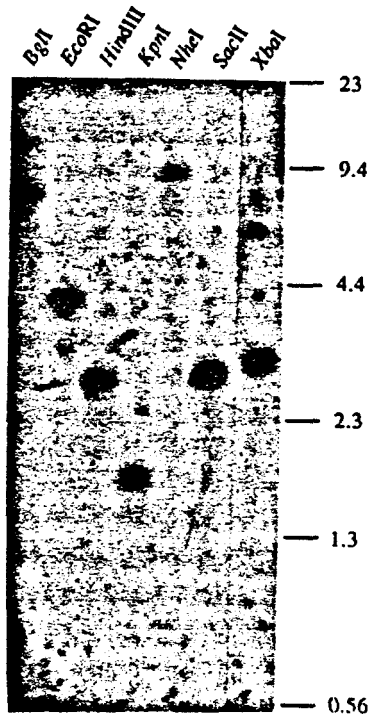


FIG. 3. **Southern blot of rat genomic DNA probed with Cx31.** The DNA was digested with various enzymes and probed at high stringency with a random primed Cx31 *Eco*RI-*Sac*II fragment. Only a single band is seen in all lanes suggesting that Cx31 is a single copy gene. Hybridization of the same blot with a rat Cx32 probe revealed no bands identical with Cx31.

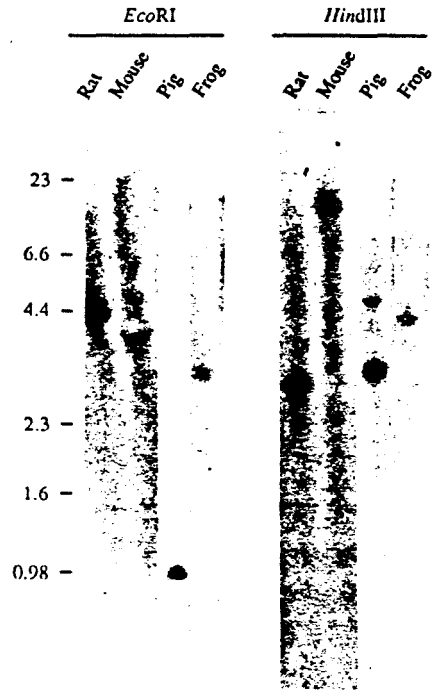


FIG. 4. **Southern "zoo" blot of genomic DNA probed with Cx31.** Genomic DNA from rat, mouse, pig, and frog was digested with *Eco*RI and *Hin*dIII and probed at moderate stringency with a random primed Cx31 *Eco*RI-*Sac*II fragment. The figure is a composite from two different washes of the same blot. The rat and mouse lanes had excessive background after the first wash, and the pig and frog lanes were too weak to permit adequate reproduction after the second wash. All washes were done in 1 × SSPE at 65 °C.

isolated from 20 different rat organs was probed with Cx31. It shows the presence of a 1.7-kb mRNA in placenta, Harderian gland, skin, and eye (Fig. 5). Some other organs show weak hybridization, but these cannot be described with confidence as positive signals. Control blots probed with RCx26, RCx32, and RCx43 show hybridization to 2.5-, 1.7-, and 3.0-kb mRNAs, respectively, in organs known to produce these transcripts, but no cross-hybridization with the RCx31 mRNA was detected. In addition to the organs known to express Cx26 and Cx43, we find a strong signal for both Cx26 and Cx43 in skin, and we detect Cx43 in bone, whole blood, and epididymal fat pads (data not shown). The control blots for Cx26 and Cx43 did not have RNA from placenta on them.

Analysis of RCx31 Protein Sequence—Hydropathy analysis of the RCx31 protein sequence reveals four highly hydrophobic amino acid segments similar to other connexins. In Cx32 and Cx43 these segments have been shown to be transmembrane spanning (Yancey *et al.*, 1989; Milks *et al.*, 1988; Zimmer *et al.*, 1987). Analysis of the protein for post-translational consensus modification sequences reveals potential phosphorylation and amidation sites. The protein kinase C consensus (S/T)-*X*-(RK) (Kishimoto *et al.*, 1985; Woodget *et al.*, 1986) is present at amino acid positions 182, 223, 229, 233, and 238 (Fig. 2). Position 182 is in the predicted second extracellular loop and would likely not be available to a cytoplasmic kinase. The other sites are all predicted to be exposed in the cytoplasm, in the carboxyl-terminal region. A single potential casein protein kinase II site is predicted at position 202, which is in the middle of the putative fourth transmembrane helix (Kuenzel and Mulligan, 1987). An amidation consensus site *X*-G-(RK)-(RK) is present at position 120 (Kreil, 1984).

Multiple Alignment—Multiple alignment of RCx31 with eight other connexins was carried out (Fig. 6; positions in the multiple alignment are referred to with the prefix MA). It includes all connexin homologues identified for which the full protein sequence is available, except for the bovine Cx43 recently published (Lash *et al.*, 1990). The alignment shows a perfect conservation of the three cysteines in the first extracellular loop, but RCx31 has a single amino acid inserted between the first two cysteines in the second extracellular loop. The cysteines have been shown to form at least one disulfide bond between the extracellular loops in Cx32[5] and Cx43.[6] RCx31 has a 22-amino acid deletion relative to the Cx43's cytoplasmic loop, leaving it with the smallest cytoplasmic loop of all the connexins. As is the case for Cx43s and XCx38, Cx31 has an arginine at MA162 in the third highly amphipathic putative transmembrane segment. The carboxyl-terminal regions of the different connexins are diverged highly. To improve the alignment of RCx31 over this region, we have taken into consideration the predicted post-translational modification sites. We find three predicted protein kinase C phosphorylation sites conserved when RCx31 and RCx43 are compared. These sites are at positions MA367, MA371, and MA376 in the multiple alignment and characterized by the sequence SSRAS. The third predicted phosphorylation site has a threonine substituted for a serine and is offset one amino acid from the corresponding Cx43 site that is at MA375. In addition the alignment shows two other sequence similarities among the other proteins that have not been

[5] B. Nicholson, personal communication.
[6] S. A. John, unpublished observations.

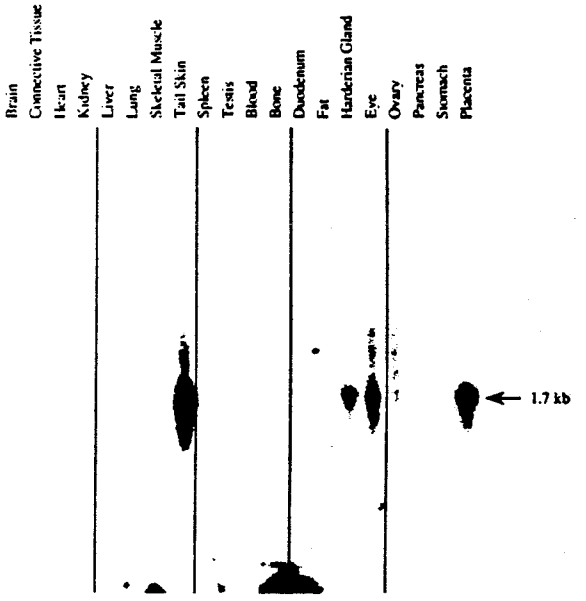


FIG. 5. **Northern blot of total RNA isolated from 20 different rat organs probed with Cx31.** RNAs were separated on a formaldehyde agarose gel and probed at high stringency with a random primed Cx31 *Eco*RI-*Sac*II fragment. The 1.7-kb mRNA band is detected in skin, Harderian gland, eye, and placenta. Control blots probed with Cx26, Cx32, and Cx43 did not cross-react with any Cx31 bands. The Cx32 mRNA comigrated with the Cx31 mRNA when the above blot was reprobed. However, Cx31 and Cx32 were not detected in any of the same tissues. Control blots did reveal the presence of Cx26 in skin, and of Cx43 in skin, bone, whole blood, and epididymal fat pads.

described previously. All connexins except RCx31 and RCx26 share a 11–13-amino acid stretch with four to six identities beginning at position MA319 in the multiple alignment characterized by a QNX_6S sequence. Also, the carboxyl termini of Cx32 and Cx43 show a weak alignment across the SSRAS segment.

Phylogenetic Tree of Connexins—Using the program CLUSTAL, a tree of the five unique connexins identified to date, for which nucleotide sequences are available, was constructed (Fig. 7). The tree has two major branches, one with Cx38 and Cx43, and a second with Cx26, Cx31, and Cx32. Based on estimated divergence rates for Cx32 and Cx43, the two branches these molecules represent diverged 1.3–1.9 BYr ago.

DISCUSSION

All previously identified connexins were isolated as cDNAs from organ- or tissue-specific libraries. Because it has become clear that the connexins form a moderately large gene family and that the various connexins exhibit tissue specificity, we decided to use an alternate approach to isolate new connexin homologues by screening a rat genomic library at low stringency. This more general approach will identify new connexins irrespective of the tissue in which they are expressed. In this way we have identified so far one new connexin, rat Cx31, that we describe here. The gene for Cx32 was isolated six times from the five genome equivalents we screened, but only a single Cx31 gene was found. In addition, the low stringency screening did not identify genes for either Cx43 or Cx26. There are several possible explanations for these anomalies. The library we used was amplified, which is known to cause biases in representation. In addition, the library was generated from an *Eco*RI partial digest, which obviously will result in some parts of the genome not being represented because the *Eco*RI fragments they generate are too large for the vector chosen.

The isolation of Cx31 now extends the number of connexins identified in rat to five, including Cx26, Cx32, Cx43, and Cx46. The nucleotide sequences for all of these, except Cx46, have been described. The number of genes isolated that code for connexins is now two, rat Cx31 and Cx32.

A significant issue with respect to Cx31 is whether the entire coding region for the protein for the gene is represented in the sequence from the single exon we describe. The answer to this cannot be known with certainty until a cDNA for Cx31 has been isolated and characterized. However, we believe that the coding region sequence is complete for the following reasons. The connexin genes coding for Cx26 and Cx32, for which there is sufficient structural detail, and that appear to be related closely to Cx31, both contain the entire coding region in a single exon. The mRNA size of 1.7 kb, although large enough to accommodate a larger coding region, is the same size as the Cx32 protein that is similar in size. Finally, there are no predicted exon/intron borders in Cx31 between the sequence coding for the fourth transmembrane helix and the carboxyl terminus (data not shown).

As in the case of the genes coding for rat Cx26, Cx32, and Cx43 (Miller *et al.*, 1988; Musil *et al.*, 1990; Zhang and Nicholson, 1989), the gene coding for Cx31 is a single copy gene. The fact that it shares no bands with Cx32 on a genomic Southern blot suggests that the Cx31 gene is not located near the Cx32 gene. Despite this, Cx31 and Cx32 may be part of a gene cluster with other connexins.[7] Further understanding of the Cx31 gene structure must await the isolation of a cDNA. By Southern blotting the Cx31 gene appears to be highly conserved, and a hybridization signal is seen in an organism as diverged from rat as the amphibian *X. laevis*. Of course it is possible that the signal seen in *Xenopus* with the Cx31 probe is a connexin isoform. However, all the connexins, Cx32 and Cx43, which have been isolated from multiple species are much more conserved between species than between isoforms. Further comparison of Cx32 and Cx43 sequences demonstrates that the amino-terminal region, corresponding to the first 99 amino acids, of these molecules is particularly well conserved and changes at a rate similar to the slowly evolving cytochrome *c* and glyceraldehyde-3-phosphate dehydrogenase (Dayhoff, 1972).

Cx31 has a unique distribution being found in skin, eye, Harderian gland, and placenta. It is found in the skin with Cx43 and Cx26. The skin is a complex tissue with many cell types including fibrobasts that have been shown to express Cx43 (Beyer *et al.*, 1989; Crow *et al.*, 1990). Intercellular communication in the epidermis has been studied in some detail (Kam *et al.*, 1986), but the exact localization of the different connexin molecules, their relationship to each other, and physiological roles have yet to be examined. Cx31 is found with Cx43 and MP70 (Cx46?) in the eye, where Cx43 is localized to the fibroblasts in the cornea and the cells of the lens epithelium (Beyer *et al.*, 1989) and MP70 is localized to the lens fiber cell (Kistler *et al.*, 1988).

To date Cx31 is the only connexin found in the placenta. Gap junctions have been described in the mature rat placenta in all three layers of the labyrinth. The barrier between maternal and fetal blood supplies is at the boundary between layers II and III. Tracers applied through the maternal blood supply easily pass through layer I and accumulate at the border between II and III. These layers have been shown to contain gap junctions, and it has been proposed that these

[7] J. H. Hoh, unpublished observations.

```
                                                                                                                    90
RCx31 M.DWkkLqdL LSGVNqYSTA fGRiWLSVvF vFRvLVyVVA AErVWGDEQk dFdCNTrQPG CtNVCYDaFF PISniRLWaL QLIFVtcPsm
RCx26 M.DWgtLqsi LgGVNkHSTe IGKiWLtVLF IFRIMiLVVA AkeVWGDEQa dFvCNTlQPG CkNVCYDhYF PISHiRLWaL QLImVSTPAs
XCx30 M.nWagLyai LSGVNrHSTe IGRiWLSVvF.IFRIMVLVaA AESVWGDEkS AFtCNTQQPG CnaVCYDhFF PISHiRLWaL QLIiVSTPAL
RCx32 M.nWtgLytL LSGVNrHSTA IGRVWLSViF IFRIMVLVVA AESVWGDEkS sFiCNTlQPG CnaVCYDhFF PISHVRLWsL QLIlVSTPAL
HCx32 M.nWtgLytL LSGVNrHSTA IGRVWLSViF IFRIMVLVVA AESVWGDEkS sFiCNTlQPG CnaVCYDqFF PISHVRLWsL QLIlVSTPAL
XCx38 MagWalLklL LDdVQeHSTl IGKVWLtVLF IFRIfiLsVA gESVWtDEQS dFiCNTQQPG CtNVCYDqaF PIyHVRyWVL QflFVSTPTL
XCx43 MGDWsaLgrL LDkVQaYSTA gGKVWLSVLF IFRILlLgtA vESaWGDEQS AfvCNTQQPG CeNVCYDksF PISHVRfWVL QiIFVSTPTL
CCx43 MGDWsaLgkL LDkVQaYSTA gGKVWLSVLF IFRILlLgtA vESaWGDEhv AFzCNTQQPG CeNVCYDksF PISHVRfWVL QiIFVSvPTL
RCx43 MGDWsaLgkL LDkVQaYSTA gGKVWLSVLF IFRILlLgtA vESaWGDEQS AFzCNTQQPG CeNVCYDksF PISHVRfWVL QiIFVSvPTL
                                                                                                                   180
RCx31 LVilHVAYre erErKhrqKh gEhcak.lys hpgKkH.... .......... ........GG LWWTYlfSli FRlIiElVFL YVlhtlwhGF
RCx26 LVAMHVAYrr H.EKKrkfmk gEiKnefkdi eEiktq.... .......... ...KVrIeGs LWWTYttSIf FrvIFEAVFM YVFYimYnGF
XCx30 LVAMHVAhlq HqEKKe..lr lsrhvkdqel aEVKkH.... .......... ...KVKIsGt LWWTYIsSVf FrlIFEAtFM YIFYlIYPGy
RCx32 LVAMHVAhqq HiEKKm..lr lEghgdplhl eEVKrH.... .......... ...KVhIsGt LWWTYvISVv FrlIFEAVFM YVFYllYPGy
HCx32 LVAMHVAhqq HiEKKm..lr lEghgdplhl eEVKrH.... .......... ...KVhIsGt LWWTYvISVv FrlIFEAVFM YVFYllYPGy
XCx38 tyLgHmvYlm kKEEKerqKE nEsrilvane aqtevy.... ........se stkKirIqGp LmcTYttSVv FKSIFEAgFL lgQWYIY.GF
XCx43 LyLaHVfYlm rKEEKlNrKE eElKmvqneg gnVdmHLKQI EIKKFKYGIE EHGKVKmrGG LlrTYIISIl FKSvFEvgFi iIQWYmY.GF
CCx43 LyLaHVfYvm rKEEKlNkrE eElKvvqndg vnVdmHLKQI EIKKFKYGIE EHGKVKmrGG LlrTYIISIl FKSvFEvaFL iIQWYIY.GF
RCx43 LyLaHVfYvm rKEEKlNkKE eElKvaqtdg vnVemHLKQI EIKKFKYGIE EHGKVKmrGG LlrTYIISIl FKSvFEvaFL iIQWYIY.GF
                                                                                                                   270
RCx31 tKpRLVqCas vvPCPNTVDC yiaRPTEKkV FTyFMvgaSa VCIiLtIcEi cYLIF..... .......... HrimrglSkd Kst.......
RCx26 fMqRLVKCn. awPCPNTVDC FiSRPTEKTV FTVFMisVSG ICIiLNItEL cYLfiryCag ksKRpv.... .......... ..........
XCx30 SMiRLlKCd. ayPCPNTVDC FVSRPTEKTI FTVFMLVaSG VCIvLNvaEv FfLIaqaCtr RarR..... HrdSGeiSke hq........
RCx32 aMvRLVKCe. afPCPNTVDC FVSRPTEKTV FTVFMLaaSG ICIiLNvaEv vYLIiraCar RaqRRSnPpS rkgSGfghrl speyk.....
HCx32 aMvRLVKCd. vyPCPNTVDC FVSRPTEKTV FTVFMLaaSG ICIiLNvaEv vYLIiraCar RaqRRSnPpS rkgSGfghrl speyk.....
XCx38 vMSpifvCe. RiPCkHkVeC FVSRPmEKTI FIiFMLVVSl ISLlLNlmEL ihLeFKcfqh giKe..... .......... gatcpPtgip
XCx43 SlSaiytCk. RdPCPHqVDC FlSRPTEKTI FIwFMLiVSi VSLaLNIiEL FYvtyKsikd giKgkkDPF. sstndaviag KscGSPKYAY
CCx43 SlSaiytCe. RdPCPHrVDC FlSRPTEKTI FIVFMLVVSl VSLaLNIiEL FYvfFKgvkd RvKgktDPY. .shSGtmSps KdcGSPKYAY
RCx43 SlSavytCk. RdPCPHqVDC FlSRPTEKTI FIiFMLVVSl VSLaLNIiEL FYvfFKgvkd RvKgRSDPY. HattGplSps KdcGSPKYAY
                                                                                                                   360
RCx31 ..... ..... ..... ..... ..... ..... ..... ..... ..... .....
RCx26 ..... ..... ..... ..... ..... ..... ..... ..... ..... .....
XCx30 ..... ..... ..... ..... .....Q NeMnllitgg S..... ..... .....
RCx32 ..... ..... ..... ..... .....Q NeinkllSeq dgS..... ..... .....Lk
HCx32 ..... ..... ..... ..... .....Q NeinkllSeq dgS..... ..... .....Lk
XCx38 FNG.....a gnrtpPqeYt nppsenqdid lpaYNKmagg hNWasiqmEQ qvnGlvkpkc qcd..... ....cwsqsa ISvvvsgapg
XCx43 FNGCSSPTA. ..PMSPPGYK LVTGERNpSS CRNYNKQASE QNWANYSAEQ NzMGqagSti SNtHAQPFDF aDEHQNTKKm IldqRpssra
CCx43 yNGCSSPTAP LSPMSPPGYK LVTGDRNNSS CRNYNKQASE QNWANYSAEQ NzMGqagSti SNSHAQPFDF aDEHQNTKKl ASGHELQPLt
RCx43 FNGCSSPTAP LSPMSPPGYK LVTGDRNNSS CRNYNKQASE QNWANYSAEQ NzMGqagSti SNSHAQPFDF pDDnQNaKKv AaGHELQPLa
                 *   *   *                                  408
RCx31 ..ksiSspks SSrAStcrch hkLleSGDLE AVPADDKLQA SAPSLTPI
RCx26 ..... ..... ..... ..... ..... .....
XCx30 .IikRSag.. ...qekgdhc sts..... ..... .....
RCx32 dIlrRSpgtg aglAekadRc eac..... ..... .....
HCx32 dIlrRSpgtg aglAekadRc eac..... ..... .....
XCx38 IIsnmdavkr nhqtSSkqqy v..... ..... .....
XCx43 Apghemqplt SShASSrpRp ddLEI..... ..... .....
CCx43 IvdqRppsra SSrASSrpRp ddLEI..... ..... .....
RCx43 IvdqRpssra SSrASSrpRp ddLEI..... ..... .....
                 *   *   *
```

FIG. 6. **Multiple alignment of rat connexins.** Rat connexin-31 (RCx31), -26 (RCx26), -32 (RCx32), and -43 (RCx43), *Xenopus* connexin-30 (XCx30), -38 (XCx38), and -43 (XCx43), chicken connexin-43 (CCx43), and human connexin-32 (HCx32). Consensus amino acids are shown in *upper case letters* and differences are shown in *bold lower case letters*. The criterion for a consensus is the identity of half, or four or more of the sequences at a given position. Alignments were carried out three at a time using the program ALP3, then compiled using LINEUP, and adjusted by eye. It shows the characteristic highly similar sequence for the first 230 amino acids except for a deletion of varying lengths with respect to all Cx43s at positions MA126–MA149. Further it reveals a conserved segment in several but not all connexins at MA320–MA333. When the alignment was adjusted further to take into account predicted phosphorylation sites, three conserved putative protein kinase C sites, marked with *, were discovered in the RCx31 and RCx43 sequences. RCx32 also has a weak similarity across this region.

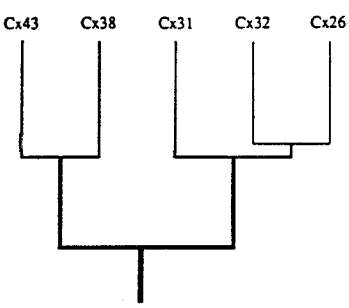


FIG. 7. **Phylogenetic tree of connexins.** The tree is based on the CLUSTAL analysis of domains I and II and of the protein sequences for rat Cx26, Cx31, Cx32, and Cx43, and *Xenopus* Cx38. The estimated time for the divergence of the two major branches is 1.3–1.9 BYr ago.

gap junctions are responsible for the exchange of fluids, gases, and small metabolites between the mother and fetus in rat (Metz *et al.*, 1978).

The Harderian gland is a secretory organ located on the posterior aspect of the eye in animals with a third eye lid. The biological function of the gland is not well understood. In mammals it has been shown to consist of two major cell types, secretory cells and myoepithelial cells (Woodhouse and Rhodin, 1963). In the hamster, both these cell types are innervated, and it has been suggested that the secretory functioning of the gland is regulated neuronally (Bucana and Nadakavukaren, 1972). A neuronal signal that activates secretion might be transmitted electrically or biochemically between cells through gap junctions. Further study of Cx31, which is the only connexin yet identified in the Harderian gland, may shed light on this process.

Analysis of the Cx31 protein sequence reveals several potential protein modification sites. The single potential amidation site is probably not utilized, because amidation usually is associated with processed proteins such as neuropeptides (Kreil, 1984). There are five predicted protein kinase C sites. The first, at position 182, is in the predicted second extracellular domain and probably not available to a kinase. The same is true for the single casein protein kinase II site at position 202, which is in the middle of the predicted fourth transmembrane helix. The remaining four protein kinase C sites are in the carboxyl-terminal region. Whether these sites are utilized must await biochemical analysis of the protein. It is interesting to note that protein kinase C and cAMP-dependent protein kinase have been implicated in regulation of junctional communication in a number of systems (Murray and Gainer, 1989). Cx32 has been shown to be phosphorylated by cAMP-dependent protein kinase *in vivo* and *in vitro* (Saez *et al.*, 1986; Takeda *et al.*, 1987). In addition, Cx43 in fibroblasts is phosphorylated (Crow *et al.*, 1990).

The multiple alignments (Fig. 6) show the general characteristics of the gene family that have been described in previous comparisons (Beyer *et al.*, 1990; Gimlich *et al.*, 1990). The molecules have a 200–220 amino acid core which is highly conserved, except for the intracellular loop. The loop is of variable length but highly conserved between homologous connexins from different species. The high degree of sequence similarity through the first 200–220 amino acids is particularly interesting considering the high degree of morphological conservation of the gap junction observed. Two notable features

in this region are the conserved cysteines and the putative amphipathic helix. The two sets of cysteines have the pattern CX_5CX_3C in the first extracellular loop and CX_3CX_5C in the second extracellular loop. The only exception is in Cx31 where the second set of cysteines has a single amino acid inserted which results in the pattern CX_4CX_5C. The putative amphipathic helix, which has been proposed to line the pore of the channel, begins at position MA154 and has the conserved motif TX_3SX_3K/RX_3E. The carboxyl termini are highly diverged and vary in length from approximately 10 to 150 amino acids and may provide regulatory specificity to the various connexin isoforms. Several aspects of the alignment we show here are different from ones published previously. We find a unique alignment in the carboxyl-terminal region characterized by the sequence QNX_6S or QNX_9S, starting at position MA320, conserved in all connexins except RCx31 and RCx26. The significance of this sequence is not known, and it does not match any known structural or functional motif as determined by PROSITE. Its conservation in many connexins would suggest that it may be involved in a shared structure or function. Further study of RCx31 and its physiology may reveal more about the sequence, since it is not present in that molecule. The approach of using predicted phosphorylation sites to adjust the alignment of the sequences revealed a segment characterized by the sequence SSRAS, starting position at MA371, that is shared between RCx31 and RCx43. There is also a weak alignment with RCx32, which contains a predicted protein kinase C site at position MA377. The SSRAS segment has several putative protein kinase C phosphorylation sites in RCx31 or RCx43, but no actual positions of protein kinase C phosphorylation in any of these molecules has been reported. RCx32 has been shown to be phosphorylated *in vitro* by protein kinase C (Takeda *et al.*, 1989).

The phylogenetic tree of connexins generated from analysis of the divergence rates, and the CLUSTAL program, predicts a branch between Cx43/Cx38 and Cx26/Cx31/Cx32 at 1.3–1.9 BYr. That estimated divergence time assumes that the amino acid replacement method of Perler is an accurate clock (Perler *et al.*, 1980). It is relevant to note that the prokaryote/eukaryote divergence is estimated at 1.8 BYr, the plant/animal at 1.0 BYr, and the vertebrate/invertebrate at 0.6 BYr (Doolittle *et al.*, 1989). It is also interesting to note that the phylogenetic tree of connexins is consistent with the nomenclature proposed for the gap junction protein family by Gimlich *et al.* (1990). That nomenclature is based on unspecified sequence similarities between connexins and uses greek letters to identify homologous proteins. Eventually it will be useful to have a naming system independent of calculated molecular weight, because a name based on the size of the molecule will become useless once the repertoire of connexin isoforms and connexins from different species becomes too large. We suggest that phylogenetic relatedness would be a suitable criterion for naming connexins, but such a system must await further definition and analysis of the gene family. Until that time the current system of using the prefix Cx followed by the molecular mass is a practical approach.

Acknowledgments—We thank Dr. D. Paul for the Cx32 cDNA clone, Dr. E. Beyer for the Cx43 cDNA, Dr. B. Nicholson for the Cx26 cDNA, Dr. J. Bonner for the rat genomic library, and R. Wagner for the *Xenopus* genomic DNA. We are also grateful to Dr. B. Yancey and Dr. K. Puranam for comments and suggestions on the manuscript.

REFERENCES

Ausubel, F. M., Brent, R., Kingston, R., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1988) *Current Protocols in Molecular Biology*, Greene and Associates and Wiley Interscience, New York

Barr, L., Dewey, F. F., and Berger, W. (1965) *J. Gen. Physiol.* **48,** 797–823

Berdan, R. C., and Gilula, N. B. (1988) *Cell Tissue Res.* **251,** 257–274

Beyer, E. C., Paul, D. L., and Goodenough, D. A. (1987) *J. Cell Biol.* **105,** 2621–2629

Beyer, E. C., Goodenough, D. A., and Paul, D. L. (1988) in *Gap Junctions* (Hertzberg, E. L., and Johnson, R. G., eds) Vol. 7, pp. 167–176, Alan R. Liss, Inc., New York

Beyer, E. C., Kistler, J., Paul, D. L., and Goodenough, D. A. (1989) *J. Cell Biol.* **108,** 595–605

Beyer, E. C., Paul, D. L., and Goodenough, D. A. (1990) *J. Membr. Biol.* **116,** 187–194

Bucana, C. D., and Nadakavukaren, M. J. (1972) *Science* **175,** 205–206

Buultjens, T. E. J., Finbow, M. E., Lane, N. J., and Pitts, J. D. (1988) *Cell Tissue Res.* **251,** 571–580

Crow, D. S., Beyer, E. C., Paul, D. L., Kobe, S. S., and Lau, A. F. (1990) *Mol. Cell. Biol.* **10,** 1754–1763

Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure 1972,* The National Biomedical Research Foundation, Silver Spring, MD

Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–394

Doolittle, R. F., Anderson, K. L., and Feng, D.-F. (1989) in *The Hierarchy of Life* (Fernholm, B., Bremer, K., and Joernvall, H., eds) pp. 73–85, Exerpta Medica, NewYork

Ebihara, L., Beyer, E. C., Swenson, K. I., Paul, D. L., and Goodenough, D. A. (1989) *Science* **243,** 1194–1195

Feinberg, A. P., and Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267

Finbow, M. E., Yancey, S. B., Johnson, R. G., and Revel, J. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 970–974

Flagg-Newton, J., Simpson, I., and Loewenstein, W. R. (1979) *Science* **205,** 404–407

Gimlich, R. L., Kumar, N., and Gilula, N. B. (1990) *J. Cell Biol.* **110,** 597–605

Henderson, D., Eibl, H., and Weber, K. (1979) *J. Mol. Biol.* **132,** 193–218

Higgins, D. G., and Sharp, P. M. (1988) *Gene (Amst.)* **73,** 237–244

Kam, E., Melville, L., and Pitts, J. D. (1986) *J. Invest. Dermatol.* **87,** 748–756

Kensler, R. W., and Goodenough, D. A. (1980) *J. Cell Biol.* **86,** 755–764

Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsuji, Y., Nomura, H., Takeyama, Y., and Nishizuka, Y. (1985) *J. Biol. Chem.* **260,** 12492–12499

Kistler, J., Christie, D., and Bullivant, S. (1988) *Nature* **331,** 721–723

Kreil, G. (1984) *Methods Enzymol.* **106,** 218–223

Kuenzel, E. A., Mulligan, J. A., Sommercorn, J., and Krebs, E. G. (1987) *J. Biol. Chem.* **262,** 9136–9140

Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132

Larson, D. M., Haudenschild, C. C., and Beyer, E. C. (1990) *Circ. Res.* **66,** 1074–1080

Lash, J. A., Critser, E. S., and Pressler, M. L. (1990) *J. Biol. Chem.* **265,** 13113–13117

Loewenstein, W. R. (1981) *Physiol. Rev.* **61,** 829–913

MacDonald, R. J., Swift, G. H., Przybyla, A. E., and Chirgwin, J. M. (1987) *Methods Enzymol.* **152,** 219–227

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Manjunath, C. K., Goings, G. E., and Page, E. (1982) *Biochem. J.* **205,** 189–194

Manjunath, C. K., Goings, G. E., and Page, E. (1984) *Am. J. Physiol.* **246,** H865–H875

Mehta, P., Bertram, J., and Loewenstein, W. R. (1986) *Cell* **44,** 187–196

Meinkoth, J., and Wahl, G. (1984) *Anal. Biochem.* **138,** 267–284

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984) *Nucleic Acids Res.* **12,** 7035–7056

Metz, J., Aoki, A., and Forsmann, W. G. (1978) *Cell Tissue Res.* **192,** 391–407

Milks, L., Kumar, N. M., Houghton, R., Unwin, N., and Gilula, N. B. (1988) *EMBO J.* **7,** 2967–2975

Miller, T., Dahl, G., and Werner, R. (1988) *Biosci. Rep.* **8,** 455–464

Murata, M., Richardson, J. S., and Sussman, J. L. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 3073–3077

Murray, A. and Gainer, H. S. C. (1989) in *Cell Interactions and Gap Junctions* (Sperelakis, N., and Cole, W. C., eds) Vol. 1, pp. 97–107, CRC Press, Boca Raton, FL

Musil, L. S., Beyer, E. C., and Goodenough, D. A. (1990) *J. Membr. Biol.* **116,** 163–175

Nicholson, B. J., Hunkapiller, M. W., Grim, L. B., Hood, L. E., and Revel, J.-P. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 7594–7598

Nicholson, B. J., Gros, D. B., Kent, S. B. H., Hood, L. E., and Revel, J.-P. (1985) *J. Biol. Chem.* **260,** 6514–6517

Nicholson, B., Dermietzel, R., Teplow, D., Traub, O., Willecke, K., and Revel, J.-P. (1987) *Nature* **329,** 732–734

Paul, D. L. (1986) *J. Cell Biol.* **103,** 123–134

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) *Cell* **20,** 555–566

Pitts, J. D. (1978) in *Intercellular Junctions and Synapses, Receptors and Recognition, Series B* (Feldman, J., Gilula, N. B., and Pitts, J. D., eds) Vol. 2, pp. 63–79, Chapman and Hall, London

Revel, J.-P., and Karnovsky, M. J. (1967) *J. Cell Biol.* **33,** C7–C12

Robertson, J. D. (1963) *J. Cell Biol.* **19,** 201–221

Ryerse, J. S. (1989) *Cell Tissue Res.* **256,** 7–16

Saez, J., Spray, D. C., Nairin, A. C., Hertzberg, E. L., Greengard, P., and Bennett, M. V. L. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 2473

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467

Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A., and Bonner, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 3256–3260

Sheridan, J. D., Finbow, M. E., and Pitts, J. D. (1979) *Exp. CellRes.* **123,** 111–117

Sneath, P. H. A., and Sokal, R. R. (1973) *Numerical Taxonomy,* W. H. Freeman & Co., New York

Spray, D. C., and Bennett, M. V. L. (1985) *Annu. Rev. Physiol.* **47,** 281–303

Strauss, W. M. (1988) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R., Moore, D. D., Seidman, J. G., Smith, J. A., and Stuhl, K., eds.) Vol. 1, pp. 2.2.1–2.2.3, Greene and Associates and Wiley Interscience, New York

Tabor, S., and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 1074–1078

Tabor, S., and Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U. S. A.,* **84,** 4767–4771

Takeda, A., Hashimoto, E., Yamamura, H., and Shimazu, T. (1987) *FEBS Lett.* **210,** 169–172

Takeda, A., Saheki, S., Shimazu, T., and Takeuchi, N. (1989) *J. Biochem.* **106,** 723–727

Traub, O., Look, J., Dermietzel, R., Brummer, F., Hulser, D., and Willecke, K. (1989) *J. Cell Biol.* **108,** 1039–1051

Wilbur, W. J., and Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 726–730

Woodget, J. R., Gould, K. L., and Hunter, T. (1986) *Eur. J. Biochem.* **161,** 177–184

Woodhouse, M. A., and Rhodin, J. A. G. (1963) *J. Ultrastruct. Res.* **9,** 76

Yancey, S. B., John, S. A., Lal, R., Austin, B. J., and Revel, J.-P. (1989) *J. Cell Biol.* **108,** 2241–2254

Zhang, J.-T., and Nicholson, B. J. (1989) *J. Cell Biol.* **109,** 3391–3401

Zimmer, D., Green, C. R., Evans, W. H., and Gilula, N. B. (1987) *J. Biol. Chem.* **262,** 7751–7763

Appendix I

Sequence Diversity of Gap Junction Proteins

Sequence diversity of gap junction proteins

J. P. REVEL, S. B. YANCEY, B. NICHOLSON and J. HOH

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

Abstract. This paper summarizes our understanding of the molecular organization of gap junction proteins. There appear to be overall similarities in the organization of heart and liver junctions in terms of general domains, even though the molecular sizes of the two proteins are quite different. Sequence data on the amino-terminal regions of these two proteins show 43% of the residues to be identical and 25% more to be homologous. The major intrinsic protein of lens (MIP), believed by many to be the lens-fibre junction protein, does not show such sequence homology with the known portions of junction proteins from either heart or liver. Yet the sequence of MIP, which is completely known, suggests a conformation for this molecule quite compatible with a junctional role. It thus appears that molecules potentially involved in junction formation will prove to form a rather diverse family, with special characteristics of organ-specific molecules that may well be related to their function.

1986 Junctional complexes of epithelial cells. Wiley, Chichester (Ciba Foundation Symposium 125) p 108–127

With the finding of many structural and physiological similarities between gap junctions in an enormous range of multicellular organisms and tissues (see review by Peracchia 1980), it has become quite common to think of gap junctions as highly conserved structures. With more refined biochemical analysis, however, it has become increasingly evident that several junctional proteins exist, which raises the exciting possibility that the differences seen reflect functional specializations. This paper reviews the evidence for diversity based on sequence analysis of junction proteins, which holds the key to controlled progress in understanding the function of these cell-to-cell channels.

The gap junction protein(s)

Size of the gap junction proteins

The problem of the size of the junction protein has been a thorny one for several years. We have recently discussed this topic in some detail (Revel et al 1985) and will only summarize, except where new information is available.

Since it is becoming clear that junction proteins in different organs can be rather dissimilar, we have chosen to discuss individual organs separately.

The major junction molecule of mammalian liver. In spite of many years of work (see Goodenough 1976, Duguid & Revel 1976), arguments about the size of the liver junction protein were squelched only in autumn 1985, when Paul (1985b) deduced the size of the protein from the analysis of its cloned cDNA, confirming the consensus molecular mass of 27–28 kDa. This was the size of the protein obtained in different laboratories by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) of preparations from different origins, when proteolysis and aggregation were adequately controlled (Henderson et al 1979, Hertzberg & Gilula 1979, Finbow et al 1980, Nicholson et al 1981, Hertzberg 1984). The reported molecular mass of the junction protein has varied from 26 kDa to 28 kDa, but the apparent differences presumably reflect differences in the calibration of sizing gels, rather than real differences in molecular mass. Affinity-purified antibodies that recognize the 28 kDa protein stain recognizable junctions (Dermietzel et al 1984, Hertzberg & Skibbens 1984, Paul 1985a) both *in situ* (Fig. 1) and in isolated fractions. In Western blots molecules of about 28 kDa bind antibody preparations generated against junctions isolated by different approaches (Paul 1985b, Traub et al 1983, Dermietzel et al 1984, Hertzberg & Skibbens 1984).

Unfortunately, there are other, contradictory results. The antisera that Warner et al (1984) used in their studies of gap junctions in *Xenopus* were raised against a 27 kDa junctional component of liver, yet bind to a 54 kDa molecule in *Xenopus*, and in rat liver homogenates. Paul (1985a) has also reported cross-reactivity with a 54 kDa component found in homogenates. The

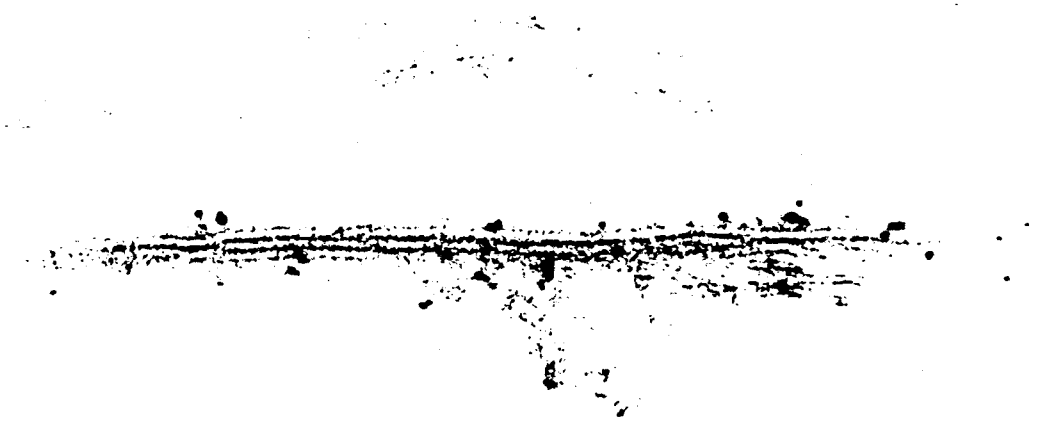


FIG. 1. Liver gap junction stained *in situ* with an affinity-purified antibody to liver junctions followed by a gold-labelled second antibody. Little or no label is found on adjacent liver cell membranes. (Frozen thin section of material embedded in sucrose by the technique of Tokuyasu; prepared by Dr R. Dermietzel.)

size of the cloned cDNA coding for the liver gap junction protein now seems to exclude the possibility that this larger molecule is a precursor. Carefully controlled experiments will be needed to determine whether it is (1) related to the 28 kDa protein and involved in gap junction structure; (2) a contaminant in the material used as antigen; or (3) a non-junctional but abundant protein showing homology with the gap junction protein (e.g. β-fibrinogen).

Extensive arguments in support of another molecule, which migrates in SDS–PAGE as a 16 kDa protein, have been presented by Finbow et al (1983 and this volume). Although one could imagine this molecule to represent a proteolytic breakdown product, this suggestion is not supported by peptide mapping data. The sequence of the liver protein will show whether the 16 kDa polypeptide represents a fragment of the 28 kDa molecule. If not, it will then be necessary to determine whether the 16 kDa entity is a particularly troublesome contaminant or whether it has to be considered yet another junction protein.

How many different junction molecules exist in liver? Besides the major protein(s) detected on gels of solubilized purified junctions, there are always minor bands to be seen if the lanes are heavily loaded. Some of these could represent contaminants (Finbow et al 1980), but others might be minor junctional components.

Our understanding of the chemical composition of gap junctions is largely derived from the analysis of isolated junction fractions. In the absence of enzymic activity, a major criterion of purity has been the morphological examination of junction fractions by electron microscopy. Because junction components could be lost or modified by proteolysis during isolation, one cannot be certain of the composition of junctions *in situ*. The rearrangement of membrane constituents might lead to the inclusion in isolated junctions of molecules normally present in the extrajunctional membrane. The alkaline extraction procedure developed by Hertzberg (1984) and the treatment of liver junctions with 0.6 M-KI (Gros et al 1983), two significant departures from the usual isolation steps, do not introduce detectable differences into the protein patterns seen by SDS–PAGE, so that these theoretical concerns may not be justified in practice.

A semi-quantitative estimate of minor junctional components present in junction fractions can be made by densitometric analysis of gels stained with Coomassie Blue (Nicholson et al 1981). In rat liver only about 10% of the stain is not in the 28 kDa band, and most of this is found in a band of about 21 kDa. In mouse liver junctions, the 21 kDa component (Henderson et al 1979) and its dimer represent as much as 30–40% of the junction protein. Partial hepatectomy, which in the rat reduces the number of junctions (Meyer et al 1981) and the amount of 28 kDa protein (Finbow et al 1980), does not affect the intensity of the 21 kDa band (Traub et al 1983). The 21 kDa entity might thus be considered a contaminant, were it not for the fact that its amino-terminal

sequence (see below) shows 40–50% homology with that of the 28 kDa molecule (B. Nicholson et al, unpublished work). One must therefore consider the existence of junctions that contain several different but related proteins, as the acetylcholine receptor does. Alternatively, there may be two separate types of junctional structures which co-isolate but are each composed of a single major protein. We presently favour the second hypothesis, given the independent behaviour of the 21 and 28 kDa bands after hepatectomy and the fact that, in the mouse, the two proteins reaggregate separately after solubilization.

Thus, when we speak of the junction protein to denote the major component presumably involved in the construction of the channel, we must be aware that there may well be several other proteins. The precise relationship between these remains to be worked out.

Heart junction molecules. Until recently it was believed that, as in the liver, the junction protein in the heart had a molecular mass of about 28–30 kDa (Kensler & Goodenough 1980, Manjunath et al 1982, Gros et al 1983). Several authors had also observed polypeptides of 40–50 kDa, but since they were absent from the cleanest preparations, they were considered contaminants (Manjunath et al 1982, Colaco & Evans 1982). The recent work of Manjunath et al (1985) now suggests that the native protein is, in fact, a molecule of about 45–47 kDa, which we will refer to here as the 45 kDa entity. The 28–30 kDa fragment is believed to be produced by the action of an endogenous serine protease, which can be inhibited only if a sufficient concentration of phenylmethylsulphonyl fluoride is present continuously. A cytoplasmic fuzzy coat present in intact heart junctions is lost as the 45 kDa protein disappears and is replaced with polypeptides of lower molecular mass (Manjunath & Page 1985). The 45 kDa band and the 28 kDa polypeptide have identical N-terminal sequences (B. Nicholson et al, unpublished work).

Thus, the changes in molecular size we have described are due to the removal of a cytoplasmic tail fragment totalling about 17 kDa from the carboxyl end of the molecule. The N-terminal sequence of the heart protein is related to but substantially different from that of the liver protein (details are presented below). Thus liver and heart junction proteins are probably distinct even though they show sufficient homology to permit some (Hertzberg & Skibbens 1984) but not all (Paul 1985a, Dermietzel et al 1984) antibodies to cross-react.

Eye lens. A discussion of the size of the junction protein in eye lens is made particularly difficult by lingering doubts about the true junctional nature of major intrinsic protein (MIP), the molecule originally assumed to represent the gap junction between lens fibres (see Revel et al 1985, Revel & Yancey 1985, for discussion). On the basis of models derived from the amino acid sequence of MIP, we have weighed the possibility that MIP is a channel-forming

molecule, and thus has at least one attribute essential to a junction protein (Gorin et al 1984, Revel & Yancey 1985).

Among other suggestive features is the fact that MIP can reconstitute channels when introduced into artificial membranes (Peracchia & Girsch 1985, Girsch & Peracchia 1985, Zampighi et al 1985, Gooden et al 1985, Nikaido & Rosenberg 1985). These experiments are less than conclusive because the starting material is usually not a highly purified protein, but consists of purified junction fractions with their associated contaminants. Unfortunately other approaches, such as immunocytochemistry, also give results that are difficult to interpret (see below).

Recently, Kistler et al (1985) have identified a molecule of about 70 kDa in lens junctions. By immunofluorescence this molecule has the punctate distribution expected of a junction protein. However the 70 kDa entity is present in unexpectedly small amounts in purified junction fractions. It is also easily removed from membranes by mild detergent treatment, a property more reasonably expected of adventitious material than of a transmembrane channel-forming protein, and especially one that requires interactions between proteins for channel formation and assembly into the characteristic junctional plaques. Such interactions would be expected to stabilize and anchor the protein in the lens membranes.

Immunocytochemistry of potential junction proteins

The evidence that any of the molecules described above are junctional is, as we have seen, somewhat circumstantial, since it is not absolutely certain that they exist in junctions, as distinguished from junction fractions, which may have gained or lost components. One way to try to overcome this problem is to show that the molecules of interest are appropriately distributed *in situ*. Although rearrangements are also likely to occur during preparation of the samples for immunocytochemistry, the artifacts are likely to be different from those due to the isolation procedures.

Lens MIP. Much of the work with MIP has unfortunately been done with isolated fractions. This entails potential problems (discussed above), which could be responsible for the widely conflicting results obtained after different isolation procedures. The immunocytochemical study by Paul & Goodenough (1983) suggests that MIP may be present in all lens membranes except where 'thick' contacts, of morphology similar to that expected for gap junctions, are found. 'Thin' junctions, where no gap is apparent between adjacent membranes, display MIP only on one side. These experiments are contradicted by the work of Keeling et al (1983) and of Sas et al (1985), which suggests that MIP antibodies (including monoclonals) bind, although not exclusively, to junc-

tional areas of lens-fibre fractions. One could argue that different determinants are labelled in the two sets of experiments, perhaps because MIP molecules located in the junction and elsewhere in the membrane are not in the same configuration.

A different approach was used by Bok et al (1982) and Fitzgerald et al (1983), who detected MIP by immunostaining of sections of eye lenses. They found MIP in all lens-fibre membranes, including regions identified as gap junctions. These data are not completely compelling because the lens morphology appears severely damaged by the freeze fixation used to preserve the tissue, but the idea that MIP is a junction protein is supported by recent data from Benedetti's laboratory (Vallon et al 1985). Overall, immunocytochemistry has not yet completely resolved the ambiguities in the localization of MIP.

Immunocytochemistry of the liver 28 kDa protein. In the liver, immunostaining reveals the close association of antibody against junction protein and gap junctions. There is little or no extrajunctional staining of the plasma membrane. In regenerating liver, labelled vesicles found in the cytoplasm may represent junction precursor material being brought to the cell surface for insertion (R. Dermietzel et al, unpublished work). At present there are no data on other tissues which can be interpreted at the ultrastructural level. However, since the association of the 28 kDa component with junctions is good in liver, one is encouraged to believe that the protein detected by the antibody to liver junctions in other tissues is also associated with junctions. The lens results, however, and the differences in sequence to be discussed below make it clear that each tissue will have to be examined independently before definite conclusions can be drawn.

Sequence data

A major step in understanding the functioning of junction proteins will be to obtain sequence data which will permit modelling of the protein itself. To achieve this goal two complementary approaches, amino acid sequencing and cDNA sequencing, have been used. Amino acid sequencing from small amounts of protein (Hunkapiller et al 1984) can give a partial sequence for the N-terminal moiety of the protein or derived peptides. Such sequences can be used to synthesize oligonucleotide probes or peptides to be used as immunogens in the production of antibody probes; with these, libraries can be searched for partial or complete clones, which can then be efficiently and completely sequenced. A complete sequence is available only for MIP, although the successful sequencing of the liver protein, under way in several laboratories (including ours), has been announced (Paul 1985b, N.B. Gilula, this volume).

Although gap junction protein is available only in trace amounts, the sensitivity of modern techniques has permitted us to obtain information about the proteins from liver, lens and heart. Because of technical limitations it has only been possible to obtain sequence data on the first 25 or so amino acids for the heart-derived molecule and about 50 for the liver protein.

Paul (1985b) has now confirmed by cDNA sequencing the liver N-terminal amino acid sequence we had previously published (Nicholson et al 1981). For the eye lens protein, both the amino-terminal and the cyanogen-bromide peptide sequences we had obtained by protein sequencing have been confirmed by DNA sequencing (Gorin et al 1984). We are thus in a position to compare at least fragmentary sequences for proteins extracted from three organs.

MIP sequences. Protein-based data are available for bovine and rat MIP and cDNA-derived data for bovine lens MIP. There are very few species differences between the proteins, and those that exist are mostly conservative substitutions. This is as expected from the work of Zigler & Horwitz (1981), who showed widespread immunological cross-reactivity, and from the comparative peptide mapping data of Takemoto et al (1981). Bovine and rat lens MIP differ in only three of 39 known residues, as shown below for residues 11–30:

Bovine MIP	. . .R A I C A E F F A S L F Y V F F G L G A . . .
Rat MIP	. . .R A I F A E F F A T L F Y V F F G L G S . . .

Residues 14, 20 and 30 are cysteine, serine and alanine respectively in bovine MIP, but phenylalanine, threonine and serine in rat MIP. Only a very small degree of correspondence has been found between MIP sequences and the sequences of gap junction proteins in liver or heart (the only ones known).

Sequences for liver junction proteins. All of the liver sequence information available has been obtained by direct protein microsequencing. The rat and mouse 28 kDa proteins show complete homology throughout the region available for comparison, with one exception: the mouse sequence is missing the first amino acid of the rat sequence, a methionine residue. Otherwise, as in lens MIP, there seems to be excellent conservation of residues between species.

Rat 28 kDa protein	M N W T G L Y T L L S G V N R H S T A I . . .
Mouse 28 kDa protein	N W T G L Y T L L S G ? N R

We not know of course whether the same degree of conservation persists through the rest of the molecule, or whether conservation is limited to the N-terminal region that can be analysed. The 21 kDa proteins found in relative abundance in junction fractions from mouse liver and in lesser amounts in rat liver also show a close relationship. The only differences detected between these two molecules in the first 17 residues are at position 1 (methionine in the rat, tryptophan in the mouse) and at position 16 (histidine in the rat and alanine in the mouse, although the latter assignment is not completely certain). Thus there is a nearly 90% identity between rat and mouse 21 kDa molecules.

Rat 21 kDa protein	WDWGTLQAIL GGVNKHS
Mouse 21 kDa protein	MDWGTLQ?IL ?GVNKA

Comparison of the 21 kDa with the 28 kDa sequence shows identities (residues 3, 6, 10, 12–14, 16 and 17), homologies (residues 9, 15) and inversions W T G becomes W G T). Clearly the 28 kDa and 21 kDa molecules are related to each other. As shown below, both sets of molecules also show homology with the heart junction protein. All three thus appear to be part of a family.

Comparison of sequences for heart and liver junction proteins. In the region available for comparison between heart and rat liver 28 kDa molecules, 43% of the residues are identical and another 25% represent conservative replacements (Nicholson et al 1985). None of the identities extend over more than four amino acids. Most workers believe that such regions of sequence identity are too small to constitute sets of serologically related determinants. Since at least some of the antibodies raised against the liver protein cross-react with the heart protein, one would expect regions of greater homology elsewhere in the molecule. We have already indicated that the heart protein is a molecule of 45 kDa with the same amino-terminal sequence as the polypeptide of 30 kDa derived from the larger parent by proteolysis. If this is correct, then the liver and heart proteins differ by a large 17 kDa portion that is easily accessible to proteolytic enzymes, i.e. is probably exposed at the cytoplasmic face of the junctions. One might argue that this portion of the molecule controls the physiological uniqueness of heart junctions, e.g. their unique gating characteristics after pH changes (Spray 1985). The portions of the remaining 27–30 kDa sequences that are common to the liver and heart molecules must specify the common properties of the channels (as well as others).

Heart protein	ADWSALGKLL KKVSQAY ST...
Liver 28 kDa protein	MNWTGLYTLL SGVNRHS TA...
Liver 21 kDa protein	WDWGTLQAIL GGVNHKS

Of special interest is the first amino acid of the sequence, a methionine in the rat liver 28 kDa protein, a tryptophan in the 21 kDa molecule and an alanine in the heart protein. There is no question that the assignment of a methionine as residue 1 for the rat liver junction protein is correct. It has, in fact, been confirmed by cDNA sequencing (D.L. Paul, personal communication).

We cannot be as sanguine about alanine as the N-terminus of the heart protein, for technical reasons. It has, however, been detected several times, and even should the alanine assignment prove erroneous, any revision of the sequence will clearly not specify a methionine in its place. Because of the good correspondence of the heart and liver sequences elsewhere one must therefore question whether methionine in the liver protein does represent the initiation codon, or whether there is post-translational processing of the polypeptide, as suggested by the sequence of the mouse liver 28 kDa protein.

Structure of potential junction proteins

Organization of MIP

Analysis of the deduced amino acid sequence of MIP reveals that it is likely to consist of six transmembrane segments, with a short cytoplasmic N-terminal region and a larger moiety, 43 amino acids long, at the carboxyl terminus also exposed in the cytoplasm. Antibodies raised against synthetic peptides modelled on one cytoplasmic and one ab-cytoplasmic region suggest that they are accurately placed in the model. The most interesting feature of the model is the

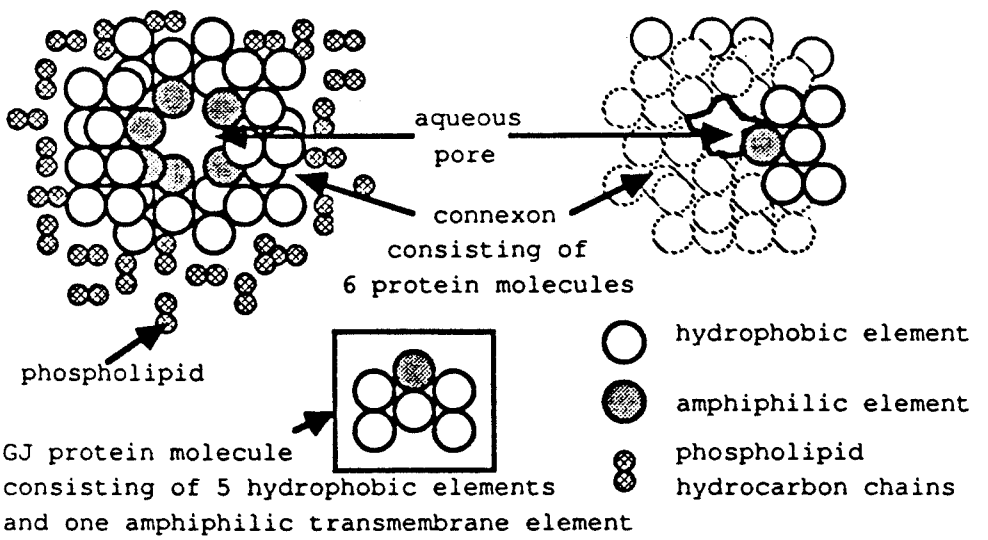


FIG. 2. Schematic diagram based on the postulated structure of MIP (Revel & Yancey 1985), showing how an amphiphilic segment of a transmembrane protein could be accommodated in the hydrophobic core of the membrane by allowing several molecules to associate and form a gap junction (GJ) channel.

existence of an amphiphilic transmembrane helix, which is most easily accommodated by allowing MIP molecules to cluster in tetramers or hexamers to form a pore-like structure (Fig. 2). This defines both a potential channel-lining region and portions of the molecule where interactions between connexons in different membranes can be supposed to take place (Revel & Yancy 1985). Experiments in progress are aimed at defining the consequence of interfering with this region of the molecule.

Stick models of junctions

For heart and liver junction proteins, the sequence information publicly available still consists entirely of peptide sequence data and therefore represents only a small portion (10%) of the whole molecules. There are some indications of similarities in overall structure, but they are not very strong at this point, because there is too little information to make meaningful comparisons. For what it is worth, there are signs of the presence of a transmembrane helix close to the amino terminus. This may be a characteristic feature of junctional proteins or may denote only that all these molecules are integral membrane proteins. In view of the paucity of data, the best that can be done at present is to make 'stick' models which help to summarize the information available (Fig. 3).

The sites of proteolytic cleavage depicted in Fig. 3 are those observed when

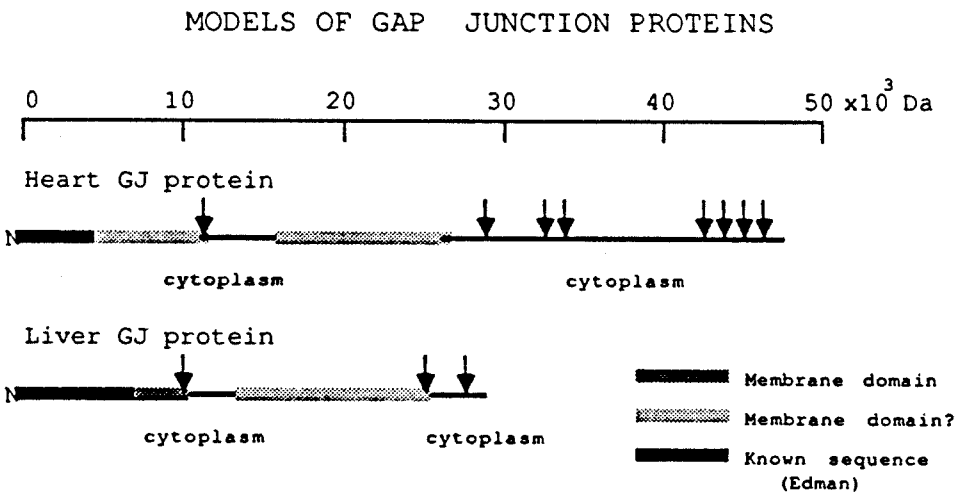


FIG. 3. Diagrammatic representation of liver and heart gap junction (GJ) proteins. The bars represent junction molecules, with arrows indicating possible cleavage points deduced from the sizes of peptides recovered after proteolysis. The cleavage sites indicated are not necessarily the only ones which explain the data. Stippled sections: regions that, because of their resistance to proteolytic attack, are believed to be mostly transmembrane; black sections: portions of the molecule known from microsequencing. Since enzyme attack will occur only on exposed, largely cytoplasmic portions of the molecules, possible locations of cytoplasmic loops can be assigned.

whole junctions are treated with enzymes. In the liver, proteolysis produces peptides of 26 kDa and 24 kDa and eventually two 10 kDa fragments. In the heart, an exposed carboxyl-terminal region is easily removed to produce a 28–30 kDa moiety, which in turn yields two 11 kDa pieces. Presumably some of the other regions are also exposed to enzymic attack. The 17 kDa peptide shown at the carboxyl terminus is not recovered as such, presumably because it becomes degraded to fragments too small to be recovered by PAGE. However, because bands are recovered at 34 kDa and 31 kDa, there must be cleavage sites there. The broad band from 45–47 kDa suggests multiple cleavages at the carboxyl end. The cytoplasmic location of parts of the molecules is surmised from the ease of proteolytic attack; in whole junctions one expects groups at the external face to be split only slowly, if at all, because the enzymes used are too large to penetrate between apposed membranes with only a 2 nm gap between them (Goodenough & Revel 1971). In heart junctions, this assumption is supported by the demonstration (Manjunath et al 1985) that a cytoplasmic fuzz layer is lost as the 45 kDa cardiac protein is cleaved to a 28–30 kDa entity. In both the liver and the heart, several molecules — probably six, as implied by the results of Unwin & Zampighi (1980), Caspar et al (1977), Zampighi & Simon (1985) and of Makowski (1985) — come together to form a channel in a manner similar to that postulated for the lens-fibre protein. The interaction of proteins with each other to form connexons is being actively investigated in several laboratories.

Conclusions

Although gap junctions in general can be experimentally shown to have many properties and characteristics in common, they can be constructed of proteins quite different from each other. The differences between proteins of different organs contrast with the similarities between proteins from the same organs of different species and permit speculation about physiological diversity. It is also becoming apparent that there may well be several proteins in each connexon and several different junctional types linking cells to each other. Working out the meaning of these differences from the point of view of the organism will be a challenging but rewarding task.

Acknowledgements

The authors would like to express their appreciation to their collaborators, whose contributions have made this synthesis possible. In particular we thank Dr C.K. Manjunath and Professor E. Page and his laboratory at the University of Chicago; Dr R. Dermietzel and Professor K. Willecke and his laboratory in Essen; Dr M. Finbow and Dr J. Pitts and his laboratory in Glasgow. We have enjoyed the expert help of Drs M. Hunkapiller, S. Kent, D. Teplow and S. Horvath and the microchemical facility at Caltech. The original work was carried out under grants from the US National Institutes of Health (GM06965 and BRSG RR07003) and from the Ruddock Fund.

REFERENCES

Bok D, Dockstader J, Horwitz J 1982 Immunocytochemical localization of the lens main intrinsic polypeptide (MIP 26) in communicating junctions. J Cell Biol 92:213–220

Caspar D, Goodenough D, Makowski L, Phillips W 1977 Gap junction structures. I. Correlated electron microscopy and x-ray diffraction. J Cell Biol 74:605–628

Colaco CA, Evans WH 1982 Partial purification of an intercalated disc containing cardiac plasma membrane fraction. Biochim Biophys Acta 684:40–46

Dermietzel R, Leibstein A, Frixen U, Janssen-Timmen U, Traub O, Willecke K 1984 A protein homologous to the 27 000 dalton liver gap junction protein is present in a wide variety of species and tissues. EMBO (Eur Mol Biol Organ) J 3:2261–2270

Duguid J, Revel JP 1976 The protein components of the gap junction. Cold Spring Harbor Symp Quant Biol 40:45–47

Finbow M, Yancey SB, Johnson R, Revel JP 1980 Independent lines of evidence suggesting a major gap junctional protein with a molecular weight of 26 000. Proc Natl Acad Sci USA 77:970–974

Finbow M, Shuttleworth J, Hamilton AE, Pitts JD 1983 Analysis of vertebrate gap junction protein. EMBO (Eur Mol Biol Organ) J 2:1479–1486

Finbow ME, Buultjens TEJ, John S, Kam E, Meagher L, Pitts JD 1986 Molecular structure of the gap junctional channel. In: Junctional complexes of epithelial cells. Wiley, Chichester (Ciba Found Symp 125) p 92–107

Fitzgerald PG, Bok D, Horwitz J 1983 Immunocytochemical localization of the main intrinsic polypeptide (MIP) in ultrathin frozen sections of rat lens. J Cell Biol 97:1491–1499

Gilula NB 1986 Topology of gap junction protein and channel function. In: Junctional complexes of epithelial cells. Wiley, Chichester (Ciba Found Symp 125) p 128–139

Girsch SJ, Peracchia C 1985 Lens cell-to-cell channel protein: I. Self-assembly into liposomes and permeability regulation by calmodulin. J Membr Biol 83:217–226

Gooden M, Rintoul D, Takehana M, Takemoto L 1985 Major intrinsic polypeptide (MIP26K) from lens membrane: reconstitution into vesicles and inhibition of channel-forming activity by peptide antiserum. Biochem Biophys Res Commun 128: 993–999

Goodenough DA 1976 In vitro formation of gap junction vesicles. J Cell Biol 68:220–231

Goodenough DA, Revel JP 1971 The permeability of isolated and in situ mouse hepatic gap junctions studied with enzymatic tracers. J Cell Biol 50:81–91

Gorin MB, Yancey SB, Cline J, Revel JP, Horwitz J 1984 The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning. Cell 39:49–58

Gros DB, Nicholson BJ, Revel JP 1983 Comparative analysis of the gap junction protein from rat heart and liver: is there a tissue specificity of gap junctions? Cell 35:539–549

Henderson D, Eibl H, Weber K 1979 Structure and biochemistry of mouse hepatic gap junctions. J Mol Biol 132:193–218

Hertzberg EL 1984 A detergent independent procedure for the isolation of gap junctions from rat liver. J Biol Chem 259:9936–9943

Hertzberg EL, Gilula NB 1979 Isolation and characterization of gap junctions from rat liver. J Biol Chem 254:2138–2147

Hertzberg EL, Skibbens RV 1984 Gap junctions in several tissues share antigenic determinants with liver gap junctions. Cell 39:61–69

Hunkapiller MW, Strickler JE, Wilson KJ 1984 Contemporary methodology for protein structure determination. Science (Wash DC) 226:304–311

Keeling P, Johnson K, Sas D, Klukas K, Donahue P, Johnson R 1983 Arrangement of MP26 in lens membranes: analysis with proteases and antibodies. J Membr Biol 74:217–228

Kensler RW, Goodenough DA 1980 Isolation of mouse myocardial gap junctions. J Cell Biol 86:755–764

Kistler J, Kirkland B, Bullivant S 1985 Identification of a 70,000-D protein in lens membrane junctional domains. J Cell Biol 101:28–35

Makowski L 1985 Structural domains in gap junctions: implications for the control of intercellular communication. In: Bennett MV, Spray DC (eds) Gap junctions. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, p 5–12

Manjunath CK, Page E 1985 Cell biology and protein composition of cardiac gap junctions. Am J Physiol 248:H783–H791

Manjunath CK, Goings GE, Page E 1982 Isolation and protein composition of gap junctions from rabbit hearts. Biochem J 205:189–194

Manjunath CK, Goings GE, Page E 1985 Proteolysis of cardiac gap junctions during their isolation from rat hearts. J Membr Biol 85:159–168

Meyer DJ, Yancey SB, Revel JP 1981 Intercellular communication in normal and regenerating rat liver: a quantitative analysis. J Cell Biol 91:505–523

Nicholson B, Hunkapiller M, Grim L, Hood L, Revel JP 1981 The rat liver gap junction protein: properties and partial sequence. Proc Natl Acad Sci USA 78:7594–7598

Nicholson B, Gros D, Kent S, Hood L, Revel JP 1985 The Mr 28000 gap junction proteins from rat heart and liver are different but related. J Biol Chem 260:6514–6517

Nikaido H, Rosenberg EY 1985 Functional reconstitution of lens gap junction proteins into proteoliposomes. J Membr Biol 85:87

Paul D 1985a Structural domains in gap junctions: implications for the control of intercellular communication. In: Bennett MV, Spray DC (eds) Gap junctions. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, p 107–122

Paul D 1985b Molecular cloning of cDNA for rat liver gap junction protein. J Cell Biol 101:393a

Paul DL, Goodenough DA 1983 Preparation characterization and localization of antisera against bovine MP26, an integral membrane protein of the lens fiber plasma membrane. J Cell Biol 96:625–632

Peracchia C 1980 Structural correlates of gap junction permeation. Int Rev Cytol 66:81–146

Peracchia C, Girsch SJ 1985 Permeability and gating of lens gap junction channels incorporated into liposomes. Curr Eye Res 4:431–440

Revel JP, Yancey SB 1985 Molecular conformation of MIP: is it a junction protein? In: Bennett MV, Spray DC (eds) Gap junctions. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, p 33–48

Revel JP, Nicholson B, Yancey SB 1985 Chemistry of gap junctions. Annu Rev Physiol 47:263–279

Sas D, Sas J, Johnson K, Menko A, Johnson R 1985 Junctions between lens fiber cells are labelled with a monoclonal antibody shown to be specific for MP26. J Cell Biol 100:216–225

Spray DC, White RL, Mazel F, Bennett MVL 1985 Regulation of gap junctional conductance. Am J Physiol 248:H753–H764

Takemoto LA, Hansen JS, Nicholson BJ, Hunkapiller M, Revel JP, Horwitz J 1981 Major intrinsic polypeptide (MIP) of the lens membrane: biochemical and immunological characterization of a major cyanogen bromide fragment. Biochim Biophys Acta 731:267–274

Traub O, Drüge PM, Willecke K 1983 Degradation and resynthesis of gap junction protein in plasma membranes of regenerating liver after partial hepatectomy or cholestasis. Proc Natl Acad Sci USA 80:755–759

Unwin PNT, Zampighi G 1980 Structure of the junctions between communicating cells. Nature (Lond) 283:489–504

Vallon O, Dunia I, Favard-Sereno C, Hoebeke J, Benedetti EL 1985 MP 26 in the bovine lens: a post-embedding immunocytochemical study. Biol Cell 53:85–88

Warner AE, Guthrie SC, Gilula NB 1984 Antibodies to gap junctional protein selectively disrupt junctional communication in the early amphibian embryo. Nature (Lond) 311:127–131

Zampighi GA, Simon SA 1985 The structure of gap junctions revealed by electron microscopy. In Bennett MV, Spray DC (eds) Gap junctions. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, p13–22.

Zampighi GA, Hall JE, Kreman M 1985 Purified lens junctional protein forms channels in planar lipid films. Proc Natl Acad Sci USA 82:8468–8472

Zigler JS, Horwitz J 1981 Immunochemical studies on the major intrinsic polypeptides from the human lens membrane. Invest Ophthalmol Visual Sci 21:46–51

DISCUSSION

Evans: Has it been shown clearly that the heart 47kDa protein is not the dimer of the 27kDa protein, as has been found in the liver?

Revel: Yes. It is definitely not a dimer.

Garrod: Aren't there several different types of junctions in the lens? Which one does MIP come from?

Revel: All of the evidence indicates that MIP is associated with membranes and most likely with junction(s), but it is not established beyond doubt that it is involved in gap junctions. Girsch & Peracchia (1985), Nikaido and Rosenberg (1985), Zampighi et al (1985) and Gooden et al (1985) have all done reconstitution experiments using solubilized lens membranes; but these have only shown us that channels (as distinct from junctions) can form. From the experiments of Zampighi et al (1985) it can be argued that one is dealing with a junction since two membranes are needed, but it is not absolutely certain that MIP is the protein causing the observed physiological effects. Conclusive experiments with highly purified MIP that show the formation of gap junction channels both morphologically and physiologically remain to be done. Dr Benedetti and his coworkers have begun filling this gap.

Evans: Do antibodies against synthetic peptides from the N-terminus of MIP block intercellular communication in lens cells?

Revel: I haven't tried that yet.

Stoker: Lens is curious in its coupling properties and is different from other cell systems, so the junctions won't necessarily be the same.

Stevenson: Have you used antibodies against MIP to stain whole tissue or isolated lens junctions, Professor Revel?

Revel: No.

Benedetti: Amino acid analysis and the sequence of MP26 (MIP), based on cDNA cloning, indicate that the polypeptide has a prevailing hydrophobicity, which is common for intrinsic membrane proteins (Do et al 1985, Gorin et al 1984, Eisenberg et al 1984). Moreover, *in vitro* experiments on the translation of RNA encoding MP26 show that the polypeptide is cotranslationally inserted into microsomal membranes (Anderson et al 1983) and into isolated plasma membranes (Dunia et al 1985). Another interesting feature of MP26 is that its chemical properties vary during the differentiation and ageing of lens fibres. In particular, MP26 is affected by endogenous protease, which during the ageing of the lens generates a polypeptide of 22 kDa. This protein fragment accumulates in the plasma membrane of the lenticular 'nucleus' (Benedetti et al 1981).

The bidimensional long-range and short-range distribution of MP26 within the lipid matrix is striking. The protein oligomers, which are probably associated with cell-to-cell communication, may form either randomly distributed particle arrays in the plane of the membrane or an eventually orthogonal geometrical lattice. In the lens, the formation of these various membrane domains can be depicted as a self-assembly of identical or quasi-equivalent protein subunits, which accommodate in the lipid phase in different spatial arrangements in parallel with the differentiation and ageing of the fibres (Dunia et al 1985).

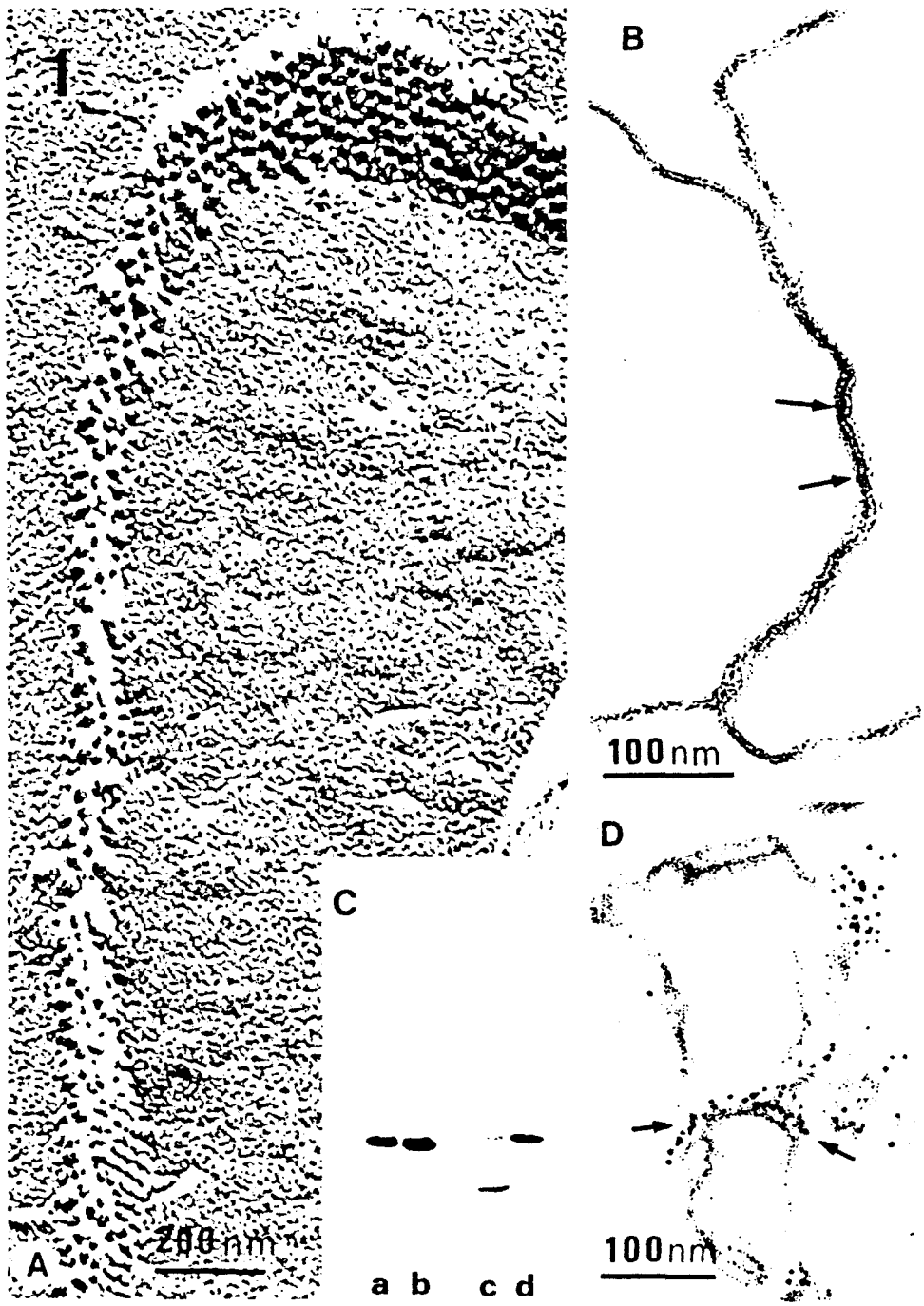
We have recently studied the reassociation of purified MP26 in reconstituted liposomes. Isolated cortical plasma membranes extracted with urea and alkali are solubilized in octyl glucoside. After high speed centrifugation the supernatant, which contains the solubilized MP26 and the membrane lipids, is complemented with azolectin and cholesterol and then exhaustively dialysed against Tris buffer. After dialysis, bilayer vesicles and multilamellar liposomes are reconstituted with the association of MP26 (Figs. 1 & 2). Since the SDS–PAGE pattern of these proteoliposomes (Fig. 1C) shows that MP26 is the only protein component of the reconstituted material, we assume that the intramembrane particles in the fracture faces of reconstituted material are composed of MP26 randomly oriented within the thickness of the bilayer. Most of these 8 nm particles have a polygonal shape (Fig. 2B) and may form either small clusters or arrays (Fig. 2A, C) or even a bidimensional geometrical lattice (Fig. 1A). The unilamellar vesicles appear to aggregate into pairs or clusters and the

---

FIG. 1. (*Benedetti*) Reassociation of purified MP26 in reconstituted liposomes. (A) Bidimensional lattice of repeating subunits in freeze-fractured reconstituted proteoliposomes. (B) Pentalaminar structure visualized in thin sections of reconstituted proteoliposomes (arrows). Note the structural identity of this reconstituted membrane domain with gap junctions. (C) SDS–PAGE pattern of reconstituted proteoliposomes (lanes a, b) and of the octyl glucoside-insoluble fraction (lanes c, d). Note that in lanes a and b MP26 is the major protein component of the reconstituted proteoliposomes. (D) Immuno-gold labelling of reconstituted proteoliposomes showing that MP26 is concentrated in regions of membrane-to-membrane contact (arrows).

intramembrane particles form arrays in sites where two unilamellar vesicles are in close contact (Fig. 2A, B).

The fracture aspect of large multilamellar liposomes is also characterized by the presence of intramembrane particles with two different types of distribution. One corresponds to randomly dispersed particulate entities and com-



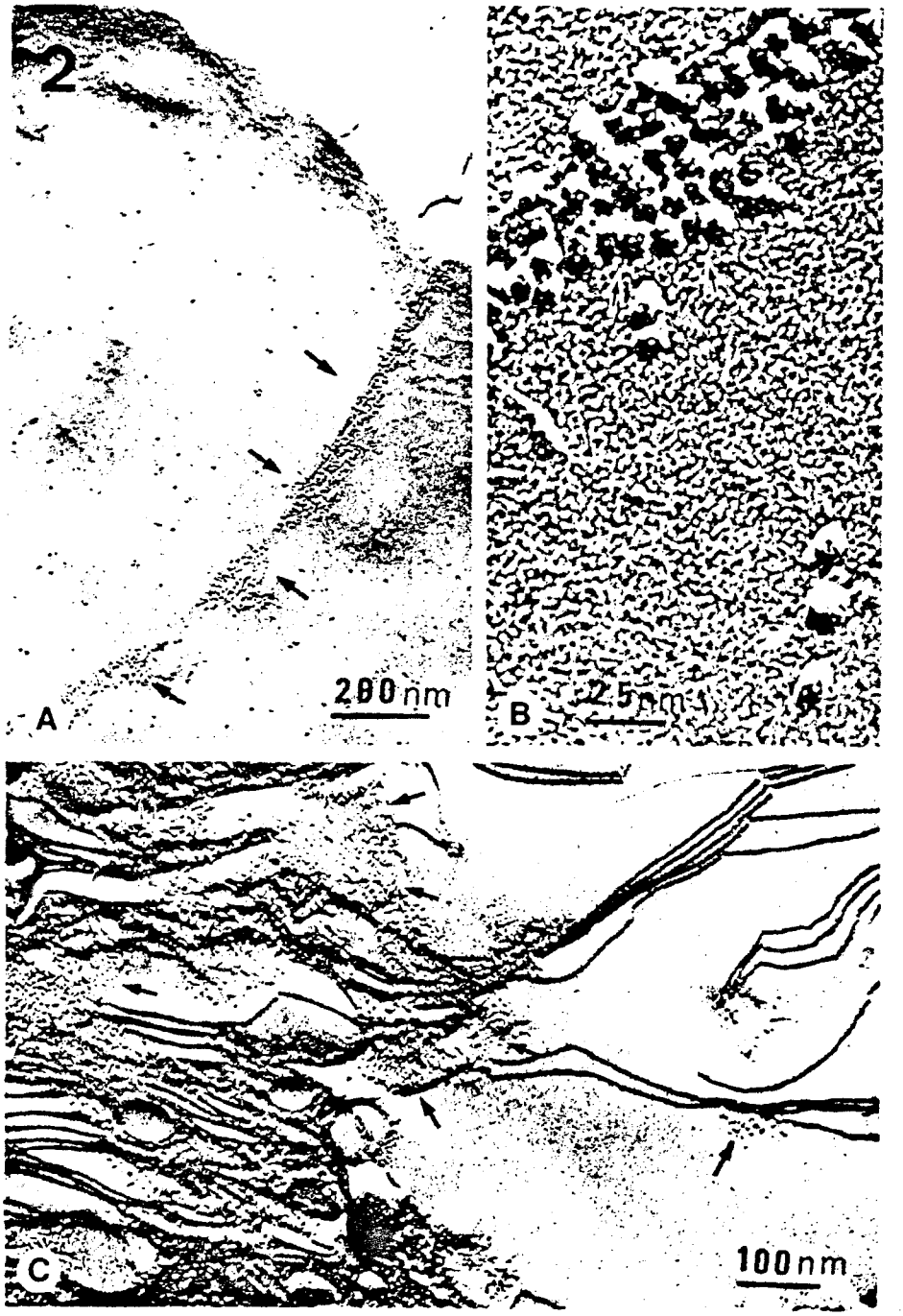


FIG. 2. (*Benedetti*) Reassociation of purified MP26 in reconstituted liposomes. (A) Particle arrays (arrows) concentrated in regions of contact between two freeze-fractured reconstituted vesicles. (B) High magnification of A, showing the polygonal aspect (arrow) of the intramembrane particles. (C) Freeze-fractured multilamellar proteoliposome showing the 'junctional domains' (arrows) between the bilayers.

plementary pits; the other is represented by assemblies of repeating subunits which connect the bilayers where the interlamellar space is reduced (Fig. 2C). It is noteworthy that even in a single fracture face the arrays of particles are intermixed with rows of pits. Thin sections of fixed and embedded proteoliposomes reveal a pentalaminar aspect at sites where vesicular membranes appear in close apposition (Fig. 1B).

Immuno-gold labelling with a polyclonal antibody against MP26 shows that the polypeptide is not evenly distributed, but forms clusters of various dimensions, concentrated at sites where two vesicular membranes are in contact (Fig. 1D) (I. Dunia, S. Manenti & E.L. Benedetti, unpublished work). These data show that MP26, like other channel-forming proteins, can be incorporated into reconstituted proteoliposomes.

The conformation of MP26, which probably forms oligomers of four subunits, favours not only lateral interaction in the plane of the bilayer (where a geometrical lattice of repeating subunits may form) but also the vertical association of MP26 oligomers in two apposed reconstituted bilayers.

Revel: These are the nicest reconstitution experiments I have seen. Most of the starting materials used by others have not been purified MIP, but simply redissolved lens junctions, so you don't know what is responsible for the physiological activity you find. Experiments with purified MIP make a great step forward by showing that MIP can form junction-like structures.

Gilula: It has been very difficult to determine whether or not MIP functions exclusively as a junctional protein *in vivo*. Some antibody experiments indicate that the antigen, which is supposed to be MIP, exists in non-junctional membrane regions, and others have localized the antigen to junctional sites. So it is important to find out whether the protein functions differently in these two locations on the same cell surface. Secondly, there have not been any extensive electrophysiological studies of intercellular junction channels in the lens. What is needed is a demonstration of a regulatory mechanism operating in a reconstituted system in the same relevant manner as it does *in vivo*. Finally, we must contend with the fact that the 'other' junction, which exists in the liver and in the heart, is also present in the lens. This gap junction is found in the undifferentiated lens epithelium, but we are not certain whether the gene product is carried through the differentiation process into the highly differentiated lens region to provide some membrane channel activity. As long as these uncertainties exist, I think we should be cautious, as Professor Revel is, about projecting how this protein is involved in channel activity in the lens. Certainly the work on reconstituted systems by Girsch & Peracchia (1985), Zampighi et al (1985) and Lucio Benedetti and his colleagues (this discussion) suggests that MIP can produce a channel with some functional and structural meaning. But we are still dealing with a potential rather than demonstrable function, and it remains to be seen what other kinds of molecules in the same tissue might provide this function *in vivo*.

Revel: The point I was trying to make in my talk is that one cannot think of a *single* junction protein, even though they all show some sort of similarity and perhaps cross-reactivity with various antibody preparations. There is now evidence for a multiplicity of junction proteins.

Edelman: Has anyone detected genetic polymorphisms in these proteins, or determined the chromosomal locations of the genes with probes?

Gilula: We have done some phenotypic mapping of the expression of nucleotide transfer in somatic cell hybrids with Frank Ruddle and his colleagues,, and have tentatively assigned the expression of this property to a single human chromosome. We used communication-defective mouse A9 cells for the analysis, and we have subsequently learned that these cells are making a junction gene product.

Franke: Studies with cDNA should help us to decide on the diversity of these proteins and sequences. It should be possible to compare Paul's cDNA sequence for the liver protein with other fragment data that exist. And if there is any reading frame information further downstream from the 3′ end than where David Paul thinks the stop codon is, that sequence should show up.

Revel: As I have not seen the actual sequence, I can't really comment. At the moment I only know what David Paul told us on the telephone, i.e. that our N-terminal sequence matches up with what he finds by sequencing his cDNA.

REFERENCES

Anderson DJ, Mostov KE, Blobel G 1983 Mechanisms of integration of *de-novo* synthetized polypeptides into membranes: signal-recognition particle is required for integration into microsomal membranes of calcium ATPase and lens MP26 but not cytochrome b_5. Proc Natl Acad Sci USA 80:7249-7253

Benedetti EL, Dunia I, Ramaekers FCS, Kibbelaar MA 1981 Lenticular plasma membranes and cytoskeleton. In: Bloemendal H (ed) Molecular and cellular biology of the eye lens. John Wiley, New York, p 137-188

Do NL, Paroutaud P, Dunia I, Benedetti EL, Hoebeke J 1985 Sequence analysis of peptide fragments from the intrinsic membrane protein of calf lens fibers MP26 and its natural maturation product MP22. FEBS (Fed Eur Biochem Soc) Lett 181:74-77

Dunia I, Do NL, Manenti S, Benedetti EL 1985 Dilemmas of the structural and biochemical organization of lens membranes during differentiation and aging. Curr Eye Res 4:1219-1234

Eisenberg D, Schwartz E, Komarony M, Wall R 1984 Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 179:125-142

Girsch SJ, Peracchia C 1985 Lens cell-to-cell channel protein: I. Self assembly into liposomes and permeability regulation by calmodulin. J Membr Biol 83:217-226

Gooden M, Rintoul D, Takehana M, Takemoto L 1985 Reconstitution into vesicles and inhibition of channel forming activity by antipeptide antibody. Biochem Biophys Res Commun 128:993-999

Gorin MB, Yancey SB, Cline J, Revel JP, Horwitz J 1984 The major intrinsic protein

(MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning. Cell 39:49-59
Nikaido H, Rosenberg EY 1985 Functional reconsititution of lens gap junction proteins into proteoliposomes. J Membr Biol 85:87-92
Zampighi GA, Hall JE, Kreman M 1985 Purified lens junctional protein forms channels in planar lipid films. Proc Natl Acad Sci USA 82:8468-8472

Appendix II

A Strain Specific Restriction Fragment Length Polymorphism Near the Rat Connexin-32 (*Cx32*) Gap Junction Gene

Mammalian Genome

A strain specific restriction fragment length polymorphism near the rat *connexin-32* (*Cx32*) gap junction gene

Jan H. Hoh and Jean-Paul Revel

California Institute of Technology, Division of Biology, 156-29, Pasadena, California 91125, USA

Received January 8, 1991; accepted February 18, 1991

Gap junctions are cellular structures that consist of cell to cell channels. They have been identified morphologically in almost every metazoan organism that has been examined and have been isolated from several different sources. The protein components identified from isolated gap junctions have been described extensively (see, for example, Stevenson and Paul 1989). These proteins, called connexins, form a gene family that was first described on the basis of protein sequence (Nicholson et al. 1985; Nicholson et al. 1987). Several cDNAs coding for gap junction proteins have now been isolated. In the rat, the three proteins for which cDNAs have been described are designated connexin-26 (Cx26), connexin-32 (Cx32), and connexin-43 (Cx43) based on their predicted molecular mass (Paul 1986; Beyer et al. 1987; Zhang et al. 1990). The genes for two connexins from the rat, connexin-31 (Cx31) and Cx32, have been cloned and described in some detail (Hoh et al. 1991; Miller et al. 1988). The *Cx32* gene is composed of two exons separated by a 6 kb intron. The first exon is approximately 0.2 kb and contains most of the 5′ untranslated region, while the second exon is 1.4 kb and contains 19 base pairs (bp) of the 5′ untranslated region, the entire coding sequence, and the 3′ untranslated ~~part of the cDNA~~. We here identify a restriction fragment length polymorphism (RFLP) near the *Cx32* gene. The polymorphism results from a 1.8 kb insertion/deletion on the 3′ side of the gene. At last count, the rat genetic map had more than 150 loci in 11 linkage groups onchromosomes (O'Brien 1990). However, few polymorphisms have been identified in or near any of these loci. The polymorphism described here will provide a convenient genetic marker for the *Cx32* gene in the rat.

Genomic clones for Cx32 were isolated by screening a Charon 4A library prepared from an *Eco*R I partial digest of rat genomic DNA with the cDNA for *Cx32* (Sargent et al. 1978; Paul 1986). The restriction maps for three unique clones isolated are shown in Fig. 1A. All three clones share an 8 kb *Eco*R I restriction fragment that has been shown to contain the first exon of the gap junction gene (Miller et al. 1988). In the Cx32E8/3 and Cx32E8/3/2 clones, the second exon is located on a 3.2 kb *Eco*R I fragment, while in Cx32E8/5 the second exon is located on a 5 kb *Eco*R I fragment. A more detailed restriction site analysis of the polymorphic *Eco*R I fragments containing the second exon, including the entire coding region, is shown in Fig. 1B. It reveals a 1.8 kb insertion/deletion of DNA beginning 300–400 bp to the 3′ side of the second exon. One end of the actual insertion/deletion site is between the *Bgl* I and *Dra* II sites in the Cx32E8/5. The current data provide no basis for distinguishing an insertion of the 1.8 kb to yield the Cx32E8/5 polymorphism from a deletion that would result in Cx32E8/3. The restriction map also predicts that the insertion/deletion should produce polymorphic restriction fragments by using the restriction enzymes *Kpn* I, *Bam*H I, *Dra* II, *Pst* I, and *Hind* III. Several of these enzymes have been used on Sprague-Dawley genomic DNA and the fragments produced are consistent with the Cx32E8/5 map (data not shown). It should be noted that the Cx32E8/5 clone appears to be identical with the 1a clone for the *Cx32* gene previously described (Miller et al. 1988). The clones were isolated from the same library using the same probe and have almost identical restriction maps. We have detected the presence of two *Bg*l I sites in the 5 kb *Eco*R I fragment in our isolate of Cx32E8/5 that are not defined by Miller et al (1988). One of the *Bgl* I sites is in the coding region and is consistent with the nucleotide sequence of the cDNA. This difference is probably the result of a restriction mapping error by Miller et al. Currently there is no evidence that the transcription of the two different *Cx32* genes is actually affected by the insertion/deletion. It has been observed that the *Cx43* gene uses at least two different 3′ untranslated ends in different tissues (Risek 1990). Also, in some cell types the *Cx43* gene produces a mRNA significantly shorter

Offprint requests to: J-P. Revel

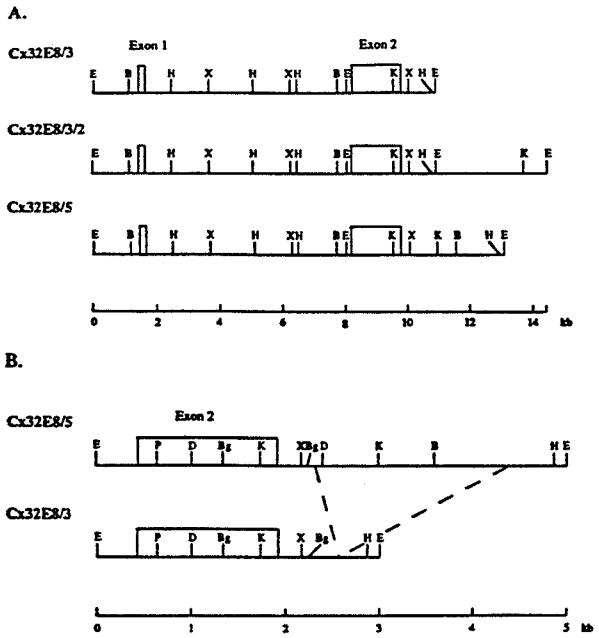


Fig. 1. Restriction maps for *Cx32* genomic clones. (A). Restriction map of three unique genomic clones for the *Cx32* gene. Restriction enzymes used are E = *Eco*R I, B = *Bam*H I, H = Hind III, K = *Kpn* I, X = Xba I. (B) Detailed restriction map of the second exon and the polymorphic region of the Cx32E8/3 and Cx32E8/5 clones. Restriction enzymes used in addition to above, Bg = *Bg*l I, P = *Pst* I, and D = *Dra* II.

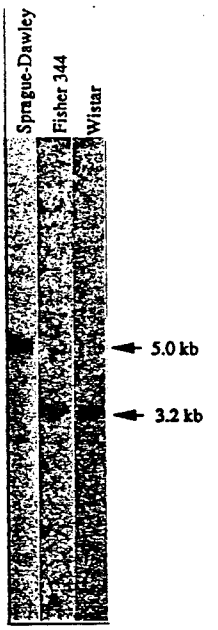


Fig. 2. *Eco*R I polymorphism of *Cx32* Gene. Southern blot of genomic DNA from Sprague-Dawley, Wistar, and Fisher 344 rats digested with *Eco*R I and probed at high stringency with the *Cx32* cDNA. The Sprague-Dawley sample has the 5 kb polymorphism while then Wistar and Fisher 344 samples have the 3.2 kb polymorphism.

than the 3.0 kb commonly seen, providing further evidence that it may be multiply spliced (D. Polacek and P.F. Davies, personal communication). While there is currently no direct evidence that the *Cx32* gene utilizes multiple 3′ ends, if it does, the proximity of the insertion/deletion may have an effect on transcription.

To ensure that the two clones did not arise from a cloning artifact, and to investigate the distribution of this RFLP, we isolated genomic DNA from 45 Sprague-Dawley rats. These rats were ordered from two different vendors (Simonsen Laboratories, 1180 C Day Road, Gilroy, Calif. 95020, USA and Harlan-Sprague-Dawley, P.O. Box 29176, Indianapolis, Ind. 46229, USA). Animals of both sexes were used in all experiments. Southern blots of *Eco*R I digests of these DNAs were probed with the *Cx32* cDNA. All 45 genomic DNAs from Sprague-Dawley rats revealed a 5 kb *Eco*R I fragment. Subsequently, we isolated genomic DNA from 11 Wistar and 10 Fisher 344 rats (Simonsen Laboratories). Southern blots of *Eco*R I digests of these DNAs produced only the 3.2 kb *Eco*R I fragment (Fig. 2). These results, summarized in Table 1, demonstrate that the RFLP near the *Cx32* gene is naturally occurring and is not a cloning artifact. In addition, the RFLP is highly strain-specific; it may therefore provide one suitable marker for some rat strains. It should be noted that both Sprague-Dawley and Wistar are outbred, while Fisher 344 is an inbred strain of rats. The *Cx32* RFLP will also be useful in determining genetic linkage of the *Cx32* gene to other genetic markers. This may be particularly important since at least some gap junction genes appear to form a gene cluster (J. Hoh and J-P. Revel, unpublished data). One notable inconsistency in the specificity of the RFLP is the origin of the clones. The genomic library from which the clones were isolated was prepared from genomic DNA from a single Sprague-Dawley rat (Sargent et al. 1979), yet both forms of the polymorphism are represented in it. There are several possible explanations for this observation, the most obvious being a sampling error. Breeders maintain populations that are isolated from each other often started from a small number of animals. We have tried to overcome this problem by testing rats from two different breeders, Simonsen Laboratories, the supplier of the rat from which the library was prepared, and Harlan-Sprague-Dawley, the company that originally derived the strain. Alternately, the 3.2 kb poly-

Table 1. RFLP distribution.

Distribution of the 3.2 and 5 kb *Eco*R I RFLPs in Sprague-Dawley, Wistar, and Fisher 344 rats.

Rat strain (Source)	3.2 kb	5.0 kb
Sprague-Dawley (Simonsen)	0	22
Sprague-Dawley (Harlan-Sprague-Dawley)	0	23
Wistar (Simonsen)	11	0
Fisher 344 (Simonsen)	10	0

morphism may be present in Sprague-Dawley at extremely low frequency and the 45 animals (90 chromosomes) we have screened are not enough to detect it. It is also possible that the rat used to prepare the library was not a true Sprague-Dawley, or that the 3.2 kb polymorphism has become fixed in the population since the library was prepared. At present we cannot distinguish among any of these explanations.

Acknowledgments. We thank Drs. S.A. John K. Puranam, and S.B. Yancey for their comments and suggestions. We also thank Dr. D. Paul for the *Cx32* cDNA. This work was supported by grants HL37109 and BRSG RR07003 from the National Institutes of Health, and the Ruddock Fund.

References

Beyer, E.C., Paul, D.L., and Goodenough, D.A.: Connexin-43: A protein from rat heart homologous to a gap junction protein from liver. *J Cell Biol 105:*2621–2629, 1987.

Hoh, J.H., John, S.A., and Revel, J.-P.: Molecular cloning, characterization, and phylogenetic analysis of a new member of the gap junction gene family, *connexin-31. J Biol Chem,* in press, 1991.

Miller, T., Dahl, G., and Werner, R.: Structure of a gap junction gene: Rat *connexin-32. Biosci Rep 8:* 455–464, 1988.

Nicholson, B.J., Gros, D.B., Kent, S.B., Hood, L.E., and Revel, J-P.: The Mr 28,000 gap junction proteins from rat heart and liver are different but related. *J Biol Chem 260:* 6514–6517, 1985.

Nicholson, B.J., Dermietzel, R., Teplow, D., Traub, O., Willecke, K., and Revel, J-P.: Two homologous protein components of hepatic gap junctions. *Nature 329:* 732–734, 1987.

Paul, D.L.: Molecular cloning of cDNA for rat liver gap junction protein. *J Cell Biol 103:* 123–134, 1986.

O'Brien, S.J.: *Genetic Maps: Locus Maps of Complex Genomes,* Cold Spring Harbor Laboratory, Cold Spring Harbor, 1990.

Risek, B., Guthrie, S., Kumar, N., and Gilula, N.B.: Modulation of gap junction transcript and protein expression during pregnancy in the rat. *J Cell Biol 110:* 269–282, 1990.

Sargent, T.D., Wu, J-R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A., and Bonner, J.: The rat serum albumin gene: Analysis of cloned sequences. *Proc Natl Acad Sci USA 76:* 3256–3260, 1979.

Stevenson, B.R., and Paul D.L.: Molecular constituents of intercellular junctions. *Curr Opinion in Cell Biol 1:* 884–891, 1989.

Zhang, J-T., and Nicholson, B.J.: Sequence and tissue distribution of a second protein of hepatic gap junctions, Cx26, as deduced from its cDNA. *J Cell Biol 109:* 3391–3401, 1989.