MOLECULAR BIOLOGY OF *SHAKER*, A DROSOPHILA GENE THAT ENCODES MULTIPLE POTASSIUM CHANNEL COMPONENTS

Thesis by

Alexander Kamb

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1988

(Submitted March, 1988)

ACKNOWLEDGEMENTS

I am aware how hopeless it is to try to acknowledge all those who have contributed to my development, both scientifically and personally. Perhaps when we finally understand how the brain works, we can rectify the problem of selective recall of names and events that prevents any acknowledgement from being truly comprehensive. With this disclaimer, I take the opportunity to single out those who have assisted me in various ways during the completion of my doctoral thesis.

First and foremost, I mention my parents, Linda and Barclay, whom I love and admire greatly. By a fortunate accident of paternity, I was reared by two exceptional people who were, first, superior role models, and now, terrific friends.

I mention other members of my family: my three brothers Barclay, Anthony, and Linus, who have been for the most part a source of pleasure and pride for me; my maternal grandparents, Linus and the late Ava Helen, who showed me that people can be not only good, but great; my uncle and aunt, Linus and Stephanie, who have shared their varied interests with me and who have been exceptionally generous in other ways.

I mention the friends I have had outside Caltech: Alex Levinson, a very loyal friend and an intelligent tennis player; Gifford Combs, a man of impeccable taste (if not manners) and a great sense of fun; Alan Korman, a rare example of charisma and soul in science; Brian Seed, a man whose cleverness is matched only by his spirit; Janet Manning, a person whose directness and world view I respect; Kathy Price, a charming woman whose radiance and support I have appreciated for many years; Arletta Sturtivant, a person who shows that age doesn't matter in a friend; the Shapiros and Zachariasens, friends of my parents who have become good friends of mine as well; and of course, Sara Smith.

I mention special friends I have had at Caltech: Tadmiri Venkatesh (Venky), the little bainchoot, the C.K.G., a mahatma and citizen of the world; Mitch Kronenberg, a

man of style and integrity; Ulf Landegren, a true intellectual with sensibility and values; Rick Barth, a creative thinker whose sense of humor lifted my spirits; Dennis Ballinger, a rock-solid person who deserves the best in science and life; Archie Pollock, a free-spirited scientist; Vijay and Usha VijayRaghavan, people whose generosity I partook of, and didn't completely repay; Chi-Bin Chien, a remarkably well-rounded prodigy; Jane Robinson (J.R.), a gentle soul who gave us that tender sadness of the Irish, as well as wholesome meals on occasion; Tobi Delbruck, an optimist who tried to keep me from becoming too serious; Jude Rosenthal, one of a kind, whose warmth, fun, and total lack of pride provided countless hours of entertainment; and, Jean Grinols, a classy dance partner and a great friend.

I mention some members of the Caltech faculty: Barbara Wold, a thoughtful and helpful scientist whose friendliness and lively conversation I enjoyed; Lee Hood, a man whom power has not corrupted, who lacks insecurity and pettiness, and who forgave me for leaving immunology; Seymour Benzer, my scientific grandfather, whose scientific style I admire, who gave me space in his lab, and who initiated the field I worked in; Bernardo Rudy, a warm and interesting scientist with whom I wish I could have been associated longer; the members of my thesis commitee, John Abelson, Henry Lester, Ed Lewis, and Barbara Wold, who have been helpful along the way.

I mention members of the Tanouye laboratory and hangers-on: Julie Tseng-Crank and Linda Iverson, who were extremely able collaborators; Medha Gautam, who helped preserve a sense of sanity; Jonathan Pollock, whose generosity, style, and dedication to science we all appreciated; Ken McCormack, who made me feel as if I wasn't the only square peg in the round hole of science; Mathew K. Mathew, whose Buddhalike smile calmed my nerves; Jim Campanelli, who will doubtless carry on the dubious tradition of Tanouye grad students; Ross McMahon, whose easygoing nature and responsible assistance we all benefited from greatly; Doug Fambrough, whose talents and no-nonsense attitude made him such good company; Martin

Heisenberg, whose genuine modesty belied great intelligence, a true gentleman; finally, Mani Ramaswami, who is like a brother to me, without whom I would certainly not have come through.

I mention my advisor, Mark Tanouye. Someone told me early on in my graduate days that every graduate student ends up hating his advisor. It never happened. I have always felt that I would work only for someone whom I respect, and I have not faltered. His philosophy, to motivate by creating a pleasant lab atmosphere, appears to work.

Last, I mention Caltech itself, its people and what it stands for. I can hardly imagine a better place to spend one's graduate years. The environment is peaceful and pleasant; the staff is fabulous; and the faculty and students in general foster a non-competitive atmosphere that is conducive to the enjoyment of science. I hope that Caltech will continue in the future to strive for the ideals that it represents to me.

ABSTRACT

Shaker (*Sh*) mutants of Drosophila suffer from a characteristic leg-shaking behavioral defect. Previous genetic and physiological experiments suggested that *Sh* encodes at least one component of a fast, transient, or A-type K^+ channel. To address questions pertaining to the structure, function and heterogeneity of K^+ channels, we have undertaken a molecular analysis of *Sh*. We have isolated molecular clones for the genomic region encompassing *Sh* as part of a 350 kb chromosomal walk. Using a combination of classical and molecular genetics, we have mapped several *Sh* mutations within this region, and localized the *Sh* gene. *Sh* mutations scatter over at least 65 kb of genomic DNA, and the *Sh* gene itself is large, spanning at least 95 kb. Comparative studies on a collection of *Sh* cDNA clones show that *Sh* encodes a diverse array of gene products. The basis for this diversity is a mechanism that generates a limited number of different 5' and 3' end segments, and splices these segments onto a central constant region. This differential splicing mechanism produces at least 10, and possibly 28 or more, predicted *Sh* proteins that differ at the carboxyl and/or amino terminus. The primary structures of *Sh* proteins deduced from the cDNAs reveal two general types of polypeptide: a protein that contains seven potential membrane-spanning domains, including a positively charged segment that is similar to a sequence called S4 in Na^+ channels, and a smaller protein that lacks S4 and contains only three potential membrane-spanning regions. Variants of the first protein type range in size from 493 a.a. to 656 a.a., whereas variants of the second type range from 303 a.a. to 337 a.a. These polypeptides may assemble as homomultimers and/or as heteromultimers to produce K^+ channels with different features.

TABLE OF CONTENTS

Introduction

Electrical Signalling in the Nervous System

Information is transmitted through the nervous system as electrical impulses that depend on the transient and selective movement of ions across cell membranes. An electrochemical gradient, generated by pumping K^+ into the cell and Na^+ and Ca^{+2} out of the cell, furnishes the driving force for these ion fluxes. Because the membrane is slightly permeable to K^+ in the resting state, the membrane voltage normally remains near the Nernst equilibrium potential for K^+, about -75 mV. Temporary fluctuations away from the resting potential are the basis for neuronal signalling.

A primary form of electrical activity in the nervous system is the action potential, first described in detail by A. Hodgkin and A. Huxley. They showed that in response to an excitatory stimulus, the membrane permeability to Na^+ sharply increases. This results in an influx of Na^+ that drives the membrane potential toward the Nernst equilibrium potential for Na^+, normally about +55 mV (Hodgkin and Huxley, 1952). The inward current carried by Na^+ is responsible for the rising phase of the action potential. The falling phase of the action potential has two components. First, within a fraction of a millisecond following the activation of the Na^+ current, the membrane permeability to Na^+ begins to decrease, or inactivate. Second, as the action potential nears its peak value, the membrane permeability to K^+ ions increases. This allows an efflux of K^+ ions from the cell that tends to counteract the depolarization caused by the Na^+ current. The combination of these two processes, inactivation of the Na^+ current and activation of K^+ currents, drives the membrane potential back toward the resting value.

Electrical impulses are transmitted between cells in either of two ways: by direct electrical coupling through gap junctions, or via a chemical intermediate called a neurotransmitter. Gap junctions are regions of membrane apposition where the cytoplasms of two cells are in physical and electrical contact (Spray and Bennett, 1985). Cells joined by gap junctions are electrically coupled such that action potentials can travel

directly from one cell to the other. This type of connection is called an electrical synapse. The second and more common mechanism for intercellular signalling in the nervous system involves the conversion of an electrical impulse into a chemical signal, and back again into an electrical form. This process takes place at a specialized composite organelle called a chemical synapse. The basic features of the chemical synapse were uncovered by B. Katz and his collaborators in their work on neuromuscular transmission at the nerve/muscle synapse, and by subsequent electron microscopic examination of the synapse structure (Katz, 1969; Heuser et al., 1974). The chemical synapse consists of three parts: 1/ a presynaptic membrane on one cell from which neurotransmitter molecules are released; 2/ a synaptic cleft across which the neurotransmitter molecules diffuse; and 3/ a postsynaptic membrane on a second cell to which the molecules bind. Depolarization caused by an incoming action potential triggers a rapid increase in Ca^{+2} permeability across the presynaptic membrane. Ca^{+2} ions flow into the nerve terminal along a Ca^{+2} electrochemical gradient and cause vesicles containing neurotransmitter molecules to fuse with the presynaptic membrane, dumping their contents into the synaptic cleft. These molecules diffuse across the cleft and bind to the postsynaptic membrane, where they induce changes in postsynaptic membrane ion permeability. Depending on the type of synapse, the permeability changes lead either to a depolarization episode called an excitatory postsynaptic potential (epsp), or to a hyperpolarization event called an inhibitory postsynaptic potential (ipsp). If the epsp reaches a threshold value, normally about -45 mV, an action potential ensues.

Ion Channels Mediate Membrane Permeability Changes

The changes in membrane permeability that underlie action potentials and synaptic transmission are regulated by integral membrane proteins called ion channels. Three-dimensional structures of ion channel molecules are so far mostly a subject for speculation; however, it is a matter of agreement that ion channels undergo

conformational changes that enable them to act as pores in the membrane through which ions can pass. These conformational changes occur in some cases as a response to membrane potential, and in other cases as a reponse to neurotransmitter binding.

Ion channels that open in response to neurotransmitter binding are called ligand-gated channels. These channels are generally clustered at chemical synapses, but may also exist in other regions of the cell. Ligand-gated channels are named for the substances that they interact with, for example, gamma-amino-butyric acid (GABA) receptors, glutamate receptors, and nicotinic acetylcholine receptors (AChR) (Hartzell, 1981; Hille, 1984). Some ligand-gated channels, such as the AChR, are selective for small cations. Other ligand-gated channels, such as the GABA receptor, are selective for small anions. In addition, there is a large class of receptors that bind neurotransmitter molecules, but do not act themselves as ion permeable pores, for example, serotonin receptors, muscarinic acetylcholine receptors, and substance K receptors (Hartzell, 1981). Instead, these receptors are coupled via second-messenger systems to other ion channels, and they mediate the electrical response to neurotransmitters indirectly.

By far the best characterized ligand-gated channel is the AChR. This receptor is found throughout vertebrate and invertebrate nervous systems, but is most prominent as the ligand-gated channel present at the vertebrate neuromuscular junction. The electric organ of the fish Torpedo provides a particularly rich source of receptor, and most structural information about the AChR has its origin in material collected from this tissue. The AChR is a pentamer composed of four different subunits in the stoichiometry alpha:beta:gamma:delta, 2:1:1:1, which range in size from 40 kD to 65 kD (Raftery et al., 1980). The subunits have several regions of homology at the amino acid level and are grossly similar in their hydrophobicity profiles: each one includes four hydrophobic domains and one amphipathic segment that are thought to span the membrane (Noda et al., 1983a). In addition, all subunits are believed to be oriented with the amino terminus on the extracellular side of the membrane and the carboxyl terminus

on the cytoplasmic side. Electron microscopy and x-ray scattering studies have revealed the approximate dimensions and shape of the AChR structure (Stroud, 1983; Kubalek et al., 1987). It is roughly 85 Angstroms across with fivefold rotational symmetry. In the center of this structure is a deep well that may form the entrance to the channel pore. There are a multitude of models and a paucity of solid experimental facts relating to the tertiary and quaternary structures of the AChR. Most models assume that the various transmembrane domains are packed as alpha helices around a central pore much as the staves in a barrel (Guy, 1984; Young et al., 1985). The interior of the barrel could be lined by the hydrophilic faces of the five amphipathic helices. It is not known how the binding of acetylcholine to the AChR triggers the opening of the channel pore.

Ion channels of the second type, voltage-gated channels, open and close as a function of membrane potential (Adams and Galvan, 1986). These channels display remarkable ion selectivity and are named for the ion species that they normally permit to pass through the cell membrane. Thus, there are Ca^{+2} channels, Cl^- channels, Na^+ channels, and K^+ channels. Each channel class, and in some cases each subclass, has a characteristic conductance, rate of opening and closing, voltage sensitivity, and pharmacological behavior.

Na^+ channels have been studied intensively and serve as a good departure point for a discussion of voltage-gated channel structure and function. Na^+ channels are present in essentially all vertebrate and invertebrate neurons. They regulate the change in membrane permeability to Na^+ that Hodgkin and Huxley first described in the action potential of the squid giant axon. In general, Na^+ channels are a homogeneous group: Na^+ channels from different cell types and different species all have similar physiological properties (Hille, 1984). In some tissues such as the vertebrate brain, Na^+ channels appear to consist of three subunits, a large glycosylated alpha subunit of 260kd, and two smaller subunits, beta and gamma (Hartshorne and Catterall, 1984). In

other tissues, only the large alpha subunit is present. From expression studies using the cloned alpha subunit gene, it seems that the alpha subunit is sufficient for reconstitution of most if not all of the physiological properties of Na^+ channels observed *in vivo* (Noda et al., 1986). The alpha subunit is composed of four tandemly repeated segments, each of which includes five hydrophobic domains and one positively charged sequence called S4 (Noda et al., 1984). Thus, the entire protein consists of twenty hydrophobic segments and four amphipathic S4 domains. The voltage dependence of the channel is thought to derive at least in part from the S4 sequence, which may traverse the membrane as an alpha helix or as a 3_{10} helix, and change conformation as the membrane potential changes (Catterall, 1986; Guy, 1986). Most models place both the amino and the carboxyl terminus of the protein within the cytoplasm.

Some primary structural information is also available for a dihydropyridine-binding protein purified from muscle T-tubules that may be a type of voltage-gated Ca^{+2} channel (Noda et al., 1987). This protein bears a remarkable resemblance in gross structure to Na^+ channels: it has four tandemly repeated domains, each consisting of five hydrophobic segments and one S4like sequence.

Potassium Channels: a Diverse Population of Ion Channels

K^+ channels are an exceptionally diverse family of channels that encompass both channels that are voltage-gated and channels that are more properly classified as ligand-gated (Rudy, 1988). Nearly thirty different channel variants have been described. These various channel types have identical ion selectivity features: in order of most permeable to least permeable, $K^+ > Na^+ > Cs^+$. However, they differ in details of voltage dependence, channel kinetics, sensitivity to pharmacological agents, and other properties. Traditionally, K^+ channels have been grouped into three general classes based on their activation and inactivation properties: fast opening, transient (A-type);

delayed opening (K-type); and, Ca^{+2}-activated (C-type). Both A-type and K-type channels are voltage-gated, whereas the C-type channel behaves as a ligand-gated channel, since it appears to open in response to Ca^{+2}, not strictly as a function of membrane voltage. In addition, there are K^+ channel classes that have been described, which are modulated by second-messemger systems, for example, serotonin (S-type) and muscarinic acetylcholine (M-type). Finally, there is a family of K^+ channels called inward rectifiers which does not seem to be either voltage- or ligand-gated, but instead, is blocked by intracellular Mg^{+2}.

The cellular distribution of different K^+ channels is also highly variable. Distinct cell types appear to possess certain combinations of K^+ channel classes in carefully controlled proportions. For example, neurons in the guinea pig nucleus tractus solitarius can be divided into three types based on electrophysiological properties (Dekin and Getting, 1984; 1987). These neurons have distinguishable K^+ currents that are differentially expressed among the three cell types: type I neurons have high levels of a C-type K^+ current and a relatively slow A-type current; type II neurons have reduced levels of the C-type current and a relatively fast A-type current; type III neurons lack these currents entirely; they possess only a K-type conductance common to all three neuronal classes.

The extensive heterogeneity of K^+ channels follows from the variety of functions they serve both in the nervous system and in other cells. K^+ channels not only play a primary role in the repolarization phase of the action potential, but also regulate the level of electrical excitability in neurons, as well as the action potential wave form. Even simple nervous systems possess a multitude of cell types that exhibit different firing patterns. There are fast-bursting neurons, slow-bursting neurons, single-spiking neurons, rapidly repolarizing neurons, slowly repolarizing neurons, and many others that differ in their excitability levels and in the precise features of their action potential wave forms. The bulk of this variation derives from differences in K^+ channel

distribution. In the case of the nucleus tractus solitarius described above, each of the three neuronal types fires in a unique repetitive manner. These firing patterns appear to be regulated mainly by the various K^+ currents each cell type possesses (Dekin and Getting, 1987).

Another basis for K^+ channel heterogeneity stems from the role of K^+ currents in short-term modulation of neuronal activity. The most exhaustively studied case involves the serotonin-sensitive, or S-type channel, in the giant sea snail Aplysia californica. E. Kandel and coworkers have shown that in certain Aplysia neurons, serotonin activates a process that ultimately results in the temporary closure of the S channel (Siegelbaum et al., 1982; Camardo et al., 1983). This process is thought to involve the stimulation of adenylate cyclase, an increase in intracellular cAMP, the activation of a cAMP-dependent protein kinase, and the phosphorylation of the S-type channel. The phosphorylated form of the channel is unable to open and therefore cannot contribute to the repolarizing K^+ current. The more this neuron is stimulated by a serotoninlike input, the more S-type channels are closed, and the less outward repolarizing current is generated following a depolarizing stimulus. Thus, repeated stimulation of this neuron results in a gradual broadening of the action potential. The ultimate physiological and behavioral effects of this mechanism are striking: a broadened action potential causes an enhancement of synaptic output that persists, in essence, a memory. These consequences follow from the fact that the amount of neurotransmitter released at the presynaptic terminal is roughly proportional to the duration of the action potential. Thus, a prolonged action potential increases the quantity of neurotransmitter released presynaptically and hence increases the magnitude and duration of the epsp. The two neurons become more strongly coupled, and the animal is sensitized to the input stimulus. In the case of Aplysia, this phenomenon appears to underlie short term memory processes such as sensitization in the gill and tail withdrawal reflexes (Klein et al., 1982).

Temporary modulation of channel activity has been described in other systems as well. For example, the M-type current is a K^+ conductance in vertebrate neurons which is inhibited by acetylcholine acting through muscarinic receptors (Adams et al., 1982). This modulation may occur via a direct coupling between the channel and G protein factors, or via inositides (Higashida and Brown, 1986; Yatani et al., 1987). Several A-type and C-type channels have been shown to be inhibited by neurotransmitter-induced modifications (Rudy, 1988). The majority of these modifications are regulated either by cAMP or Ca^{+2} second-messenger systems. Thus, K^+ channels are the targets for modifications that alter channel properties, a feature that contributes to the diversity observed within this family of channels.

Origin of Potassium Channel Diversity

It is not obvious why special roles for modulation and regulation of excitability levels are reserved primarily for K^+ channels. For instance, it is easy to imagine that Na^+ channels could satisfy some of the same physiological needs. In fact, there are at least two examples of regulation of neuronal firing patterns by Na^+ channels. In the giant axons of the Annelid Myxicola, repeated stimulation causes the nerve to become refractory to further stimulation (Rudy, 1981). This adaptation appears to be the result of an unusual Na^+ channel that is capable of entering an inactive state from which recovery is extremely slow. Purkinje cells also possess an atypical Na^+ channel, a channel that fails to inactivate, resulting in vastly broadened depolarization episodes (Llinas and Sugimori, 1980). This mode of regulation involving Na^+ channels, however, appears to be much less prevalent than K^+ channel-mediated control.

One plausible explanation for the predominance of K^+ channel types involves the energetics of ion pumps. Following an action potential, the interior of a cell is slightly depleted of K^+ and slightly richer in Na^+. To counteract the ion fluxes that contribute to the action potential, ion pumps must use energy to remove the excess Na^+

and replenish the K^+. It costs much more energy to pump Na^+ while a cell is in the resting state, than to pump K^+. This follows from the fact that the resting potential of cells is near the Nernst equilibrium potential for K^+. Thus, an ion pump must work against an electrochemical potential of roughly 130 mV (55-(-75)) to pump Na^+ out, whereas K^+ removal requires much less energy. It is sensible, therefore, to minimize the quantity of Na^+ that enters the cell during each depolarization event. Apparently, neurons are energetically efficient, because in most cases they turn off Na^+ currents as rapidly as possible and vary the excitatory wave forms through a family of K^+ currents. These arguments for energy utilization also apply to more primitive cells such as Paramecium that have inward currents carried by Ca^{+2} instead of Na^+.

Another possible explanation for the unique regulatory role of K^+ channels involves the evolutionary origin of K^+ channels. Ancestral cells probably possessed negative resting potentials because of the anionic nature of nucleic acids and most proteins. The net negative charge of these macromolecules results in an asymmetric distribution of ions across cell membranes and a negative membrane potential described by the Donnan equilibrium equation. To balance the difference in osmotic pressure caused by the high internal protein concentration, it was necessary to evolve pumps, and possibly, ion selective channels. Given the major ionic constituents of sea water, Ca^{+2}, Cl^-, Na^+, and K^+, the ionic gradients that could be generated and used by cells were restricted to this set of ions. Since the concentration of K^+ in sea water is roughly one-twentieth that of Na^+ and Cl^-, and since a high concentration of Ca^{+2} inside the cell would cause $Ca_3(PO_4)_2$ precipitate, it was essential to evolve a pump that exchanged K^+ for Na^+ and/or Ca^{+2} so that the interior of the cell became relatively K^+ rich. Na^+ ions were excluded from the cell for two reasons: first, to balance the osmotic pressure, and second, to maintain a reasonable ionic strength inside the cell. Because the resting potential of a cell is naturally negative, it makes sense to evolve pumps and channels that maintain this state, rather than to use energy in creating a positive resting potential.

These primordial ion channels, descended either from ion pump molecules or some other integral membrane protein, most likely were K^+ channels or Cl^- channels, since leakage of K^+ ions out of the cell or Cl^- ions into the cell results in a membrane potential that is negative.

This scenario is purely speculative; however, based on their broad distribution, a case can be made for the proposition that K^+ channels evolved before other ion channels. They are probably the most widely disseminated channel type and have been observed in all eukaryotic cells that have been studied using electrophysiology, including yeast (Gustin et al., 1986). By the time Paramecium evolved many millions of years ago, several classes of K^+ channel, similar to vertebrate K^+ channel types, already existed (Hille, 1984; Rudy, 1988).

If K^+ channels evolved before other ion channels, they were in a position to begin diversification first. This head start may have led to the rapid recruitment of K^+ channels for service in several different capacities. Hence, K^+ channels are both ubiquitous and diverse. They appear to subserve a variety of functions, depending on the cells in which they reside. These functions include maintaining a negative resting potential, repolarizing the membrane following an action potential, controlling spike latencies, frequencies and burst terminations, modulating adaptive responses of neurons, and regulating cell volume and turgor pressure (Rudy, 1988).

Hints about the evolutionary basis for the diversity of K^+ channels may lie in the structure of the K^+ channel itself. For example, vertebrate and invertebrate Na^+ channels are encoded by a single large gene, and it is likely that such a structure does not lend itself to rapid evolutionary change and diversification. A more sensible strategy would be to draw the appropriate polypeptides for a spectrum of channel properties from a pool of smaller subunits. To investigate this possibility, it is necessary to examine the structure of the K^+ channel.

The Biochemistry of Potassium Channels

The biochemistry and molecular biology of K^+ channels has lagged behind several other ion channels, for example, the Na^+ channel, the Ca^{+2} channel, and the AChR. This is in large part due to the fact that until recently there were no well-characterized high affinity ligands specific for K^+ channels. Much of the structural information about the Na^+ channel, Ca^{+2} channel and AChR comes either directly or indirectly from the availability of ligands that bind with high affinity to these molecules. The AChR was purified first using a neurotoxin from the banded krait snake, called alpha-bungaro toxin. Partial protein sequences were obtained from the purified subunits, and these sequences were used to construct oligonucleotide probes to identify cDNA clones for all the subunits (Noda et al., 1982; 1983b; Claudio et al., 1983). The Na^+ channel was purified in a similar fashion using the puffer fish venom component, tetrodotoxin. Antibodies were generated against the purified material and then used to screen cDNA expression libraries to obtain clones (Noda et al., 1984). A putative Ca^{+2} channel protein was purified using a high affinity dihydropyridine as a biochemical label; as with the AChR, oligonucleotides were synthesized based on the protein sequence and used to identify cDNA clones (Noda et al., 1987).

The biochemical purification of K^+ channel molecules has not been reported. Recently, several groups have described reagents that specifically block certain classes of K^+ channel. For example, the mamba snake toxins Dendrotoxin and Toxin I block transient K^+ currents at nanomolar concentrations or less (Harvey and Anderson, 1985; Benoit and Dubois, 1986); Charybdotoxin blocks brain C-type channels at concentrations in the nanomolar range (Miller et al., 1985); and Apamin, a component of bee venom, appears to bind one variety of C-type channel in brain with high affinity (Romey et al., 1984). Of these K^+ channel blocking agents, only Apamin appears to block a single type of channel specifically. Thus, most schemes that involve biochemical purification of K^+ channels using a ligand-binding assay might yield a heterogeneous

collection of proteins. Although Apamin appears to specifically label one membrane protein (30 kD), the Apamin-binding channel is present in such small amounts in the brain that biochemical purification would be extremely difficult. There is no tissue especially rich in K^+ channels that might serve the same purpose for K^+ channel biochemistry as the electric organ of Torpedo served for AChR purification.

The Genetics of Potassium Channels in Drosophila

An alternative approach to the molecular biology of K^+ channels is to utilize genetics. A major advantage of this strategy is that prior biochemical purification of the channel can be circumvented. Instead, genetic mutations serve as a starting point for the physiological, genetic, and ultimately, molecular study of gene(s) that are likely to encode K^+ channel components (Tanouye et al., 1986). This approach involves the following steps: first, mutations that affect K^+ channels are identified on the basis of behavior, electrophysiology, and pharmacology; second, genetic methods are used to localize the gene and to facilitate molecular cloning; third, the gene is cloned and studied using molecular techniques; finally, the nucleotide sequence of the gene is used to generate specific antibodies that allow the biochemical study of the protein. The primary uncertainty in this strategy is the reliability of the mutant phenotype as an indication of the gene in question. Even with the strictest possible criteria of phenotype, it may be difficult to demonstrate that a particular mutation affects the structural gene for a K^+ channel.

This type of approach has been used to study several ion channels in Drosophila melanogaster. Drosophila is uniquely suited to the molecular and genetic study of the nervous system because it is amenable to classical genetics, molecular genetics, and even electrophysiology. Several types of K^+ channel have been observed in Drosophila: an A-type channel, a K-type channel, and a C-type channel. Mutants have been isolated that appear to have defects in each of these channel types (Ganetzky and Wu, 1986).

slowpoke (*slo*) is a mutation that specifically alters a Ca^{+2}-activated K^+ conductance present in Drosophila muscle. Action potentials in adult muscle of *slo* mutants are prolonged greatly, and voltage clamp experiments show that muscle A-type and K-type currents are normal, but a C-type current is eliminated completely by *slo*. At 22^o, *slo* adults fly only with difficulty and display mild leg-shaking behavior under ether anesthesia. At higher temperatures, *slo* flies become even more uncoordinated.

ether a go-go (*eag*) flies shake their legs vigorously when etherized. *eag* appears to affect both an A-type current and a K-type current in muscle. However, none of the *eag* alleles abolishes either I_A or I_K completely. Furthermore, different *eag* alleles preferentially reduce I_A or I_K, depending on the allele. These data suggest that *eag* may encode either a component shared between A-type and K-type channels, or a modulator of channel activity.

Hyperkinetic (*Hk*) is another leg-shaking mutation that appears to affect K^+ channels. Patch-clamp studies of cultured Drosophila neurons reveal that *Hk* mutants possess an inward rectifying K^+ channel with an unusually large unit conductance. No defects have been detected in motor neurons or neuromuscular junctions. However, the kinetics of activation of the muscle A-type channel is slower than normal and results in a smaller peak amplitude. It is not clear if *Hk* encodes a component or a modulator of K^+ channels.

Shaker (*Sh*) mutations affect only A-type currents in Drosophila muscle. Like *eag* and *Hk* mutants, *Sh* flies move their legs violently under ether anesthesia. A variety of evidence described in detail below suggests that *Sh* encodes a component of a Drosophila A-type channel that is present in muscle and in nerve.

The Behavior and Physiology of *Shaker*

The first *Sh* mutant was isolated in 1944 by Catsch in Germany. Since then, at least 25

Sh alleles have been isolated, virtually all on the basis of their abnormal leg-shaking behavior. *Sh* flies suffer from other behavioral defects as well, such as wing-scissoring, abdominal spasms, and antennal twitching under ether anesthesia. Fully awake mutant flies also twitch and shudder on occasion. These behavioral defects seem not to be caused exclusively by abnormalities within the central nervous system, because legs severed from *Sh* flies continue to shake. This suggests that defects in the peripheral nerve and muscle are sufficient to account for the behavioral phenotype.

Several groups have examined the physiological effects of *Sh* mutations in muscle, in nerve, and at the neuromuscular junction. Experimental evidence from all these preparations suggests that *Sh* mutations specifically affect a single class of K^+ channel. The behavioral phenotype of *Sh* is believed to be a direct consequence of improper membrane repolarization caused by inadequate K^+ current in motor neurons.

Jan et al (1977) first described the physiological defects of *Sh* mutants at the larval neuromuscular junction. They detected a vastly prolonged Ca^{+2} conductance in the presynaptic terminal of the *Sh* mutant Sh^{KS133} compared to wildtype larvae. The time course of the conductance increase was measured in low Ca^{+2} medium by iontophoretic application of Ca^{+2} to the nerve terminal at various times after nerve stimulation. In normal larvae, the increased Ca^{+2} conductance lasts 1-2 msec. In Sh^{KS133} larvae, it lasts 60 msec. The authors considered three possible explanations for the *Sh* defect: 1) defective Na^+ channels that fail to inactivate properly and cause a prolonged presynaptic depolarization epoch; 2) abnormal Ca^{+2} channels that remain open after repolarization of the nerve terminal; and 3) defective K^+ channels that fail to repolarize the nerve terminal effectively. The possibility of a Na^+ channel defect was excluded by using the Na^+ channel antagonist tetrodotoxin to block Na^+ currents and eliminate action potentials. Under these conditions, Ca^{+2} conductance was still prolonged in the mutant when the nerve terminal was depolarized directly, suggesting that defects in Na^+ channel inactivation are not responsible for the abnormal *Sh*

physiology. The possibility of altered Ca^{+2} channels was refuted by electrotonically repolarizing the mutant nerve terminals and showing that Ca^{+2} conductance can be turned off. Thus, the *Sh* defect does not stem from the failure of Ca^{+2} channels to inactivate. Since application of the K^+ channel-blocking agent 4-aminopyridine caused a normal neuromuscular junction to mimic a mutant terminal, Jan et al. concluded that the probable underlying basis for the *Sh* phenotype was a defect in repolarizing K^+ current.

More direct evidence for a repolarization defect in *Sh* mutants because of abnormal K^+ channels comes from interacellular recordings of action potentials in wildtype and *Sh* flies (Tanouye et al, 1981; Tanouye and Ferrus, 1985). Measurements of action potential duration in the cervical giant axon of adult Drosophila reveal that repolarization is delayed in several *Sh* mutants including Sh^{KS133}. In normal flies, the action potential lasts about 0.5 msec., whereas in Sh^{KS133} flies the action potential duration is increased to 5 msec. Ca^{+2} channel blockers such as Co^{+2} and Mn^{+2} fail to remove the repolarization delay in mutants, so the defect is not due to abnormally large inward Ca^{+2} currents that act to maintain a state of depolarization. Moreover, Sh^{KS133} defects can be mimicked in normal fibers by application of 4-aminopyridine. The drug has little effect on Sh^{KS133} giant fibers.

The most illuminating analysis of *Sh* physiology involves experiments on larval and pupal muscle using the voltage-clamp technique (Salkoff and Wyman, 1981a; 1981b; 1983; Salkoff, 1983; Haugland and Wu, 1988). At 72 hours of pupal development, the developing flight muscle contains a single A-type K^+ current. In Sh^{KS133} muscle, however, this current is completely absent. Normal fullydeveloped flight muscles contain several other ion currents in addition to I_A: a voltage-activated inward Ca^{+2} current, a glutamate-activated synaptic current, a K-type current, and a C-type current. Of these currents, only the A-type current is affected by *Sh* mutations. Similar results have been obtained from voltage-clamp experiments on larval muscle. Thus, the *Sh* defect is specific and it also can be mimicked under voltage clamp by

application of 4-aminopyridine to the preparation.

Aldrich and coworkers have carried the analysis of *Sh* one step further, using the patch-clamp technique (Solc et al., 1987). They find two classes of A-type K^+ currents, one present in cultured myotubes (A_1) and one present in cultured larval brain neurons (A_2). These two channels are distinct, differing in conductance, voltage dependence, and gating kinetics. Only the A_1 channel present in myotubes appears to be affected by *Sh* mutations; a neuronal K^+ channel that is altered in *Sh* mutants has not been observed so far. Since *Sh* is known to affect nerve as well as muscle, these data suggest that either: 1) the A_1 channel may be expressed in only a small subpopulation of larval brain neurons; 2) the A_1 channel may be present only in the axons, not in the cell bodies of neurons; or 3) the A_1 channel may be only one among several types of K^+ channel affected by *Sh* mutations; that is, *Sh* may alter other K^+ channels present in neurons. In fact, Solc et al. (1987) have observed other classes of A-type currents in both neurons and myotubes but have not analyzed these currents in detail.

The Nature of *Sh* Mutations

Many different *Sh* alleles have been examined in muscle under voltage-clamp and all appear to suffer from the same general defect in an A-type K^+ current (Salkoff, 1983; Timpe and Jan, 1987; Haugland and Wu, 1988). However, there are quantitative differences in the amount of I_A produced by various mutants: four ethylmethanesulfonate- (EMS-)induced mutations (Sh^{KS133}, Sh^{102}, Sh^{M}, Sh^{K82a}) and two x-ray-induced mutations (Sh^{LC}, W32) completely abolish I_A; two EMS-induced mutations (Sh^{E62}, Sh^{rKO120}) and one x-ray-induced mutation (B55) reduce I_A in amplitude but leave its kinetic and voltage-dependent properties unaltered.

Voltage-clamp experiments on transheterozygote combinations of the various EMS-induced *Sh* alleles provide further insight into the nature of *Sh* mutations. In an exhaustive study of *Sh* transheterozygote I_A levels in larval muscle, Haugland and Wu (1988) showed that most allelic combinations produce a quantity of I_A that is the simple average of the I_A produced by either allele alone. Two alleles, however (Sh^{KS133}, Sh^{102}), appear to interfere with the wildtype gene product: transheterozygote combinations that involve either of these alleles produce a level of I_A that is significantly less than the simple average expected. Furthermore, this discrepancy between the I_A value predicted from the average and the real I_A value does not involve a constant factor, but instead, varies from one allelic combination to another. This type of behavior that involves antagonistic and specific interallelic effects is suggestive of subunit interactions in an oligomeric protein assembly.

The properties of one allele, Sh^5, is particularly intriguing under voltage-clamp. In Sh^5 pupal muscle, the peak conductance for A-type K^+ currents is near normal. However, the kinetics of the currents is altered in the mutant so that channels both inactivate more rapidly and recover from inactivation faster than channels in normal flies. In Sh^5 larval muscle, A-type currents have altered activation and inactivation voltage-dependence. Moreover, transheterozygote combinations that involve Sh^5 display inactivation voltage relations that are intermediate between Sh^5 and wildtype. These results suggest that Sh^5 encodes an altered gene product. Analysis of Sh^5 transheterozygote combinations implies that the gene product affected by this mutation is present in the multimeric structure as a single copy. Otherwise, a more complicated inactivation voltage relationship would be expected from heterozygous animals if the multimer involved several combinations of wildtype and Sh^5 subunits.

Sh mutations thus span a range of phenotypes that has implications for the

nature of the *Sh* gene product. There are amorphic (loss of function) alleles, hypomorphic (reduced function) alleles, antimorphic (poison function) alleles, and one neomorphic (gain of function) allele. All these alleles have very specific effects on the Drosophila nervous system; they alter one class of K^+ current. Morover, the physiology of the various *Sh* transheterozygotes suggests that the *Sh* gene product participates in a multimeric structure, and possibly, a heteromultimer.

Further insight into the nature of *Sh* has been achieved through detailed genetic studies of the *Sh* locus carried out primarily by M. Tanouye and A. Ferrus (Tanouye et al., 1981; Tanouye and Ferrus, unpublished). This genetic analysis has uncovered other interesting aspects of *Sh* apart from its role in K^+ channel function. In addition, attempts to define the molecular nature of *Sh* have relied heavily on the genetic and cytological characterization of the region encompassing *Sh* described below.

Genetics and Cytology of *Sh*

Sh mutations map to position 57.6 on the Drosophila X chromosome. Most, if not all, *Sh* mutations are dominant with respect to leg-shaking and cannot be placed into genetic complementation groups using this behavioral phenotype. However, several EMS-induced *Sh* alleles have been grouped together on the basis of interallelic recombination frequency. Sh^{E62} lies 0.01cM distal to a group of three *Sh* alleles: Sh^{KS133}, Sh^{102}, and Sh^{5}. These three alleles cannot be separated from each other by recombination and must lie closer than 0.01cM.

The region surrounding *Sh* has been saturated for recessive lethal mutations, and several lethal complementation groups have been mapped with respect to each other and with respect to *Sh* (Ferrus and Tanouye, unpublished). Three recessive lethal complementation groups, called l_1, l_2, and l_3, fall immediately distal to *Sh*. These three loci are located 0.07 cM, 0.05 cM, and 0.03 cM from *Sh*, respectively. Several recessive lethal complementation groups lie just proximal to *Sh*. The nearest one, called

l_4, maps 0.07 cM in the proximal direction. These recessive lethal mutations are not dominant leg-shakers and therefore differ from the *Sh* mutations themselves. The lethal period varies for the different loci, but in general, lethality occurs relatively late: in the first or second larval instar. The function of these genes is unknown. However, two types of evidence suggest that at least some of the recessive lethals may be involved in nervous system function. First, mosaic analyses show that lethality foci for these mutations are located in the ventral area of the gynandromorph fate map, a region from which the nervous system and musculature originate. Second, intracellular recordings from lethal/+ heterozygotes show abnormal giant fiber action potentials.

Several chromosome rearrangements have been used to map *Sh* cytologically (Tanouye et al., 1981). Flanking *Sh* on the left is the breakpoint of the insertional translocation T(X:3)JC153 located at 16E2-4. This breakpoint causes neither behavioral nor physiological abnormalities and thus probably lies just distal to *Sh*. Flanking *Sh* on the right is the T(X:Y)V7 translocation breakpoint at 16F6-8. The V7 translocation does not appear to have any behavioral or physiological defects and therefore serves as a proximal cytological boundary for *Sh*. Located between JC153 and V7 are three translocation breakpoints, all of which show leg-shaking and electrophysiological abnormalities: T(X:Y)B55, T(X:3)Sh^{LC}, and T(X:Y)W32. Sh^{LC} and W32 flies have no detectable A-type currents in their muscle fibers. B55 flies have reduced levels of I_A.

Small deficiencies and duplications of the *Sh* region have been constructed using these X-Y translocations (Tanouye et al., 1981; Salkoff, 1983). A hemizygous deficiency that completely deletes the material between the B55 and W32 breakpoints is viable and eliminates A-type K^+ currents from muscle. These animals shake their legs violently and have abnormal action potentials in their giant fibers (Tanouye et al., 1981). Animals carrying duplications of the same region possess levels of I_A similar to B55 animals. Deletions of the chromosomal material between the W32 and V7 breakpoints

results in haplolethality; that is, a heterozygous deficiency is lethal and animals die in the embryonic stage. Haplolethal regions are relatively rare in Drosophila: only two other haplolethal regions are known (Lindsley et al., 1972; Spencer et al., 1982).

These X-Y translocations have been used in addition to map the recessive lethals with respect to chromosome breakpoints and to map *Sh*. Recessive lethal loci l_1, l_2, and l_3 lie distal to both the B55 and W32 breakpoints, whereas the l_4 locus falls proximal to these breakpoints (Ferrus and Tanouye, unpublished). Timpe and Jan (1987) showed that the ability to produce measurable I_A in B55 flies segregates with the proximal arm of the B55 X chromosome. Flies that carry *Sh* null mutations produce no I_A. However, flies that carry a *Sh* null mutation and, in addition, carry the proximal arm of B55, possess roughly the same amount of I_A as B55 flies themselves. This suggests that at least part of *Sh* maps proximal to the B55 translocation breakpoint.

Hypothesis: *Sh* Encodes Potassium Channel Components That Are Differentially Expressed

Several genetic observations suggest that *Sh* may be a rather complex locus: *Sh* mutant phenotypes are associated with several mutations and chromosome breakpoints located at distinct sites within the *Sh* region. These include sites occupied by the Sh^{E62} and Sh^{KS133} mutations, as well as the B55 and W32 breakpoints. It is difficult to determine from the genetic analysis whether mutations at these various sites correspond to abnormalities in different gene products, or represent aberrant expression of a single gene product. In addition, loci that neighbor *Sh*, l_1, l_2, and l_3 appear to play a role in nervous system function. This role may also involve membrane excitability, but the relationship of the recessive lethals to *Sh* is not clear.

Comparison of the different *Sh* mutant phenotyes suggests that the K^+ channel

affected by *Sh* mutations is present in at least some muscle fibers and some neurons. Moreover, there is evidence that the expression of *Sh* in these tissues is differentially regulated. For example, among the *Sh* translocation mutations, B55 appears to have reduced function, producing roughly 20% of the wildtype I_A level in pupal muscle (Timpe and Jan, 1987). Yet, B55 mutants shake their legs much more vigorously than the I_A null mutant W32. Sh^{LC} mutants, on the other hand, lack I_A completely, and also are strong leg-shakers. Because the vigor of leg-shaking appears to reflect a defect in nerve, one interpretation of these results is that B55 preferentially affects a channel in nerve, W32 preferentially affects a channel in muscle, and Sh^{LC} eliminates channel expression in both tissues. Among the EMS-induced mutations, Sh^{rKO120} is the best example of this type of phenotypic discordance. Sh^{rKO120} flies are extreme leg-shakers, more extreme than several I_A null mutants such as Sh^{102}. However, a sizeable fraction of I_A is present in the muscle fibers of Sh^{rKO120} animals, suggesting that Sh^{rKO120} also differentially affects expression of the *Sh* gene product in muscle and nerve.

The simplest way to account for the results presented in the preceding sections is to propose that the *Sh* locus encodes one, and possibly several, K^+ channel subunits whose expression may be regulated in complex ways. The specificity of the defect, the gene dosage behavior of the locus, and the physiology of Sh^5 argue that *Sh* mutations affect the structural gene for a K^+ channel. Furthermore, the transheterozygote analysis suggests that the A-type channel affected by *Sh* is a multimeric structure and is possibly a heteromultimer. Finally, the genetic complexity of the locus and the discordance between the behavioral and physiological phenotypes suggest that *Sh* may encode several products. *Sh* is, therefore, a prime candidate for molecular analysis because: 1) it is likely to encode a subunit of a K^+ channel; 2) it appears to be genetically complex; and 3) it may be subject to interesting patterns of tissue specific gene expression.

The Molecular Biology of *Sh*

Two approaches are used typically in Drosophila to isolate molecular clones of genetically characterized loci. The first method, transposon tagging, involves identification of P element transposon-induced mutations, and retrieval of the flanking genomic DNA using P element probes. P elements are a family of transposons in Drosophila that can be induced to transpose at a relatively high frequency by certain genetic crosses called hybrid dysgenic crosses (Rubin et al., 1982). In a dysgenic cross, males of a P strain that carry a large number of P elements on their chromosomes are crossed to females of an M strain that lack P elements. Because M strains do not produce a transposition repressor substance, P elements on the parental P strain complement of chromosomes are induced to transpose. These transposition events cause insertion mutations that can be selected for genetically. Mutants obtained in this way are used as a source of DNA for genomic libraries that are screened in order to recover clones containing P element DNA. These clones are sorted through to determine which ones contain flanking DNA from the gene of interest. This method has been used to isolate several genes in Drosophila (Searles et al., 1982; Chang et al., 1986; Banerjee et al., 1987). Attempts to isolate *Sh* by transposon tagging, however, have not been successful (Jan et al.,1983).

The second method for cloning genes in Drosophila, chromosome walking, relies on the availability of a probe that lies close to the cytological location of the gene of interest (Spoerel and Kafatos, 1987). This probe is used first to screen genomic libraries to obtain a family of overlapping genomic clones. The extreme end fragments from this set of clones are used next to isolate additional genomic clones that extend further along the chromosome. This process is repeated until the gene of interest is identified within the cloned sequence. Several Drosophila genes, including *Sh*, have been isolated by this method (Bender et al., 1983; Garber et al., 1983).

The goals of the research described in this thesis were: 1) to isolate the *Sh* gene as part of a chromosomal walk originating from adjacent sequences; 2) to provide some form of positive identification for the *Sh* gene; and 3) to provide evidence in support of or in opposition to the hypothesis that *Sh* encodes a K^+ channel subunit. The first chapter of this thesis discusses efforts to reach these goals. The second chapter presents material that further supports the notion that *Sh* encodes a K^+ channel component. In addition, it describes results that pertain to the structure of this type of K^+ channel in Drosophila, as well as to the phenomenon of K^+ channel diversity.

References

Adams, P.R., Brown, D.A. and Constanti, A. (1982). M-currents and other currents in bullfrog sympathetic neurons. J. Physiol. *330*, 537-572.

Adams, P.R. and Galvan, M. (1986) Voltage-dependent currents of vertebrate neurons and their role in membrane excitability. In *Basic Mechanisms of the Epilepsies*. ed. Delgado-Escueta, A., Ward, A.A., Woodbury, D.M. and Porter, R. Raven Press: New York.

Banerjee, U., Renfranz, P.J., Pollock, J.A. and Benzer, S. (1987). Molecular characterization and expression of *sevenless*, a gene involved in neuronal pattern formation in the Drosophila eye. Cell *49* , 281-291.

Bender, W., Spierer, P. and Hogness, D.S. (1983). Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the *bithorax* complex in Drosophila melanogastor. J. Mol. Biol. *168*, 17-33.

Benoit, E. and Dubois, J.M. (1986). Toxin I from the snake Dendroaspis polylepis polylepis: a highly specific blocker of one type of K^+ channel in myelinated nerve fiber. Brain Res. *377*, 374-377.

Camardo, J.S., Schuster, M.J., Siegelbaum, S.A. and Kandel, E.R. (1983). Modulation of a specific potassium channel in sensory neurons of Aplysia by serotonin and cAMP-dependent protein phosphorylation. Cold Spring Harbor Symp. Quant. Biol. *48*, 213-220.

Catterall, W.A. (1986). Molecular properties of voltage-sensitive sodium channels. Ann. Rev. Biochem. *55*, 953-985.

Chang, D.Y., Wisely, B., Huang, S.M. and Voelker, R.A. (1986). Molecular cloning of *suppressor of sable*, a Drosophila melanogastor transposon-mediated suppressor. Mol. Cell. Biol. *6*, 1520-1528.

Claudio, T., Ballievet, M., Patrick, J. and Heinemann, S. (1983). Nucleotide and deduced amino acid sequences of Torpedo californica acetylcholine receptor gamma subunit. Proc. Natl. Acad. Sci. *80*, 1111-1115.

Dekin, M.S. and Getting, P.A. (1984). Firing pattern of neurons in the nucleus tractus solitarius: modulation by membrane hyperpolarization. Brain Res. *324* , 180-184.

Dekin, M.S. and Getting, P.A. (1987). *In vitro* characterization of neurons in the central part of the nucleus tractus solitarious. II. Ionic basis for repetitive firing patterns. J. Neurophysiol. *58*, 215-229.

Ganetzky, B. and Wu, C.-F. (1986). Neurogenetics of membrane excitability in Drosophila. Ann. Rev. Genet. *20*, 13-44.

Garber, R.L., Kuroiwa, A. and Gehring, W.J. (1983). Genomic and cDNA clones of the homeotic locus *Antennapedia* in Drosophila. EMBO J. *2*, 2027-2036.

Gustin, M.C., Martinac, B., Saimi, Y., Culberston, M.R. and Kung, C. (1986). Ion channels in yeast. Science *233*, 1195-1197.

Guy, H.R. (1984). A structural model of the acetylcholine receptor channel based on partition energy and helix packing calculations. Biophys. J. *45*, 249-261.

Guy, H.R. and Seetharamulu, P. (1986). Molecular model of the action potential sodium channel. Proc. Natl. Acad. Sci. USA *83*, 508-512.

Hartshorne, R.P. and Catterall, W.A. (1984). The sodium channel from rat brain. Purification and subunit composition. J. Biol. Chem. *259*, 1667-1675.

Hartzell, H.C. (1981). Mechanisms of slow postsynaptic potentials. Nature *291* , 539-544.

Haugland, F.N. and Wu, C.-F. (1988). A voltage clamp analysis of gene-dosage effects of the *Shaker* locus on larval muscle potassium currents in Drosophila. J. Neurosci. (in press).

Harvey, A.L. and Anderson, A.J. (1985). Dendrotoxins: snake toxins that block potassium channels and facilitate neurotransmitter release. Pharmacol. Ther. *31*, 33-55.

Heuser, J.E., Reese, T.S. and Landis, D.M.D. (1974). Functional changes in frog neuromuscular junctions studied with freeze-fracture. J. Neurocytol. *3*, 109-131.

Higashida, H. and Brown, D.A. (1986). Two polyphosphatidylinositide metabolites control two K^+ currents in a neuronal cell. Nature. *323*, 333-335.

Hille, B. (1984). *Ionic Channels of Excitable Membranes*. Sinauer: Sunderland, Ma.

Hodgkin, A.L. and Huxley, A.F. (1952). The components of membrane conductance in the giant axon of Loligo. J. Physiol. *116*, 473-496.

Jan, Y.N., Jan, L.Y. and Dennis, M. (1977). Two mutations of synaptic transmission in Drosophila. Proc. R. Soc. Lond. B *198*, 87-108

Jan, L.Y., Barbel, S., Timpe, L., Laffer, C., Salkoff, L., O'Farrel, P. and Jan, Y.N. (1983). Mutating a gene for a potassium channel by hybrid dysgenesis: an approach to cloning of the *Shaker* locus in Drosophila. Cold Spring Harbor Symp. Quant. Biol. *48*, 233-245.

Katz, B. (1969). *The Release of Neural Transmitter Substances*. Ryerson, Liverpool University: Toronto.

Klein, M., Camardo, J. and Kandel, E.R. (1982). Serotonin modulates a specific potassium current in the sensory neurons that show presynaptic facilitation in Aplysia. Proc. Natl. Acad. Sci. USA *79*, 5713-5717.

Kubalek, E., Ralston, S., Lindstrom, J. and Unwin, N. (1987). Location of subunits within the acetylcholine receptor by electron image analysis of tubular crystals from Torpedo marmorata. J. Cell. Biol. *105*, 9-18.

Lindsley, D.L., Sandler, L., Baker, B.S., Carpenter, A.T.C., Dennell, R.E., Hall, J.C., Jacobs, P.A., Miklos, G.L.G., Davis B.K., Gethmann, R.C., Hardy, R.W., Hessler, A., Miller, S.M., Nozawa, H., Parry, D.M. and Gould-Somero, M. (1972). Segmental aneuploidy and the genetic gross structure of the Drosophila genome. Genetics *71*, 157-184.

Llinas, R. and Sugimori, M. (1980). Electrophysiological properties of *in vitro* Purkinje cell dendrites in mammalian cerebellar slices. J. Physiol. *305*, 197-213.

Miller, C., Moczydlowski, E., Latorre, R. and Phillips, M. (1985). Charybdotoxin, a protein inhibitor of single Ca^{+2}-activated K^{+} channels from mammalian skeletal muscle. Nature *313*, 316-318.

Noda, M., Ikeda, T., Suzuki, H., Takeshima, H., Takahashi, T., Kuno, M. and Numa, S. (1986). Expression of functional sodium channels from cloned cDNA. Nature *322*, 826-828.

Noda, M., Shimizu, S., Tanabe, T., Takai, T., Kayano, T., Ikeda, T., Takahashi, H., Nakayama, Y., Kanaoka, Y., Minamino, N., Kangawa, K., Matsuo, H., Raftery, M.A., Hirose, T., Inayama, S., Hayashida, H., Miyata, T. and Numa, S. (1984). Primary structure of the Electrophorus electricus sodium channel deduced from cDNA sequence. Nature *312*, 121-127.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Furatani, Y., Hirose, T., Asai, M., Inayama, S., Miyata, T. and Numa, S. (1982). Primary structure of alpha-subunit precursor of Torpedo californica acetylcholine receptor deduced from cDNA sequence. Nature *299*, 793-797.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, T., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H. Inayama, S., Miyata, T. and Numa, S. (1983a). Structural homology of Torpedo californica acetylcholine receptor subunits. Nature *302*, 528-532.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Hirose, T., Asai,

M., Takashima, H., Inayama, S., Miyata, T. and Numa, S. (1983b). Primary structures of the beta- and gamma-subunit precursors of Torpedo californica acetylcholine receptor deduced from cDNA sequences. Nature *301*, 251-255.

Noda, M., Tanabe, T., Takeshima, H., Mikami, A., Flockerzi, V., Takahashi, H., Kangawa, K., Konjima, M., Matsuo, M., Hirose, T. and Numa, S. (1987). Primary structure of the receptor for calcium channel blockers from skeletal muscle. Nature *328*, 313-318.

Raftery, M.A., Hunkapillar, M.W., Strader, C.D. and Hood, L.E. (1980). Acetylcholine receptor: complex of homologous subunits. Science *208*, 1454-1457.

Romey, G., Hugues, M., Schmid-Antomarchi, H. and Lazdunski, M. (1984) Apamin: a specific toxin to study a class of Ca^{+2}-dependent K^+ channels. J. Physiol., Paris *79*, 259-264.

Rubin, G.M., Kidwell, M.G. and Bingham, P.M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell *29*, 987-994.

Rudy, B. (1981). Inactivation in Myxicola giant axons responsible for slow and accumulative adaptation phenomena. J. Physiol. *312*, 531-549.

Rudy, B. (1988). Diversity and ubiquity of K^+ channels. Neuroscience (in press).

Salkoff, L. (1983). Genetic and voltage clamp analysis of a Drosophila potassium channel. Cold Spring Harbor Symp. Biol. *48*, 221-231.

Salkoff, L. and Wyman, R.J. (1981a). Genetic modification of potassium channels in

Drosophila *Shaker* mutants. Nature *293*, 228-230.

Salkoff, L. and Wyman, R.J. (1981b). Outward currents in developing Drosophila flight muscle. Science *212*, 461-463.

Salkoff, L. and Wyman, R.J. (1983). Ion currents in Drosophila flight muscles. J. Physiol. *337*, 687-709.

Searles, L.L., Jokerst, R.S., Bingham, P.M., Voelker, R.A. and Greenleaf, A.L. (1982). Molecular cloning of sequences from a Drosophila RNA polymerase II locus by P element transposon tagging. Cell *31*, 585-592.

Siegelbaum, S.A., Camardo, J.S. and Kandel, E.R. (1982). Serotonin and cyclic AMP close single K^+ channels in Aplysia sensory neurons. Nature *299*, 413-417.

Solc, C.K., Zagotta, W.N. and Aldrich, R.W. (1987). Single-channel and genetic analyses reveal two distinct A-type potassium channels in Drosophila. Science *236*, 1094-1098.

Spencer, F.A., Hoffmann and F.M., Gelbart (1982). *Decapentaplegic*: a gene complex affecting morphogenesis in Drosophila melanogaster. Cell *28*, 451-461.

Spoerel, N.A. and Kafatos, F.C. (1987). Isolation of full-length genes: walking the chromosome. Methods Enzymol. *152*, 598-603.

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

Stroud, R.M. (1983). Acetylcholine receptor structure. Neurosci. Commentar. *1*, 124-138.

Tanouye, M.A. and Ferrus, A. (1985). Action potentials in normal and *Shaker* mutant Drosophila. J. Neurogenet. *2*, 253-271.

Tanouye, M.A., Ferrus, A. and Fujita, S.C. (1981). Abnormal action potentials associated with the *Shaker* complex locus of Drosophila. Proc. Natl. Acad. Sci. USA.*78*, 6548-6552.

Tanouye, M.A., Kamb, C.A. and Iverson, L.E. (1986). Genetics and molecular biology of ionic channels in Drosophila. Ann. Rev. Neurosci. *9*, 255-276.

Timpe, L.C. and Jan, L.Y. (1987). Gene dosage and complementation analysis of the *Shaker* locus in Drosophila. J. Neurosci. *7*, 1307-1317.

Yatani, A. Codina, J., Brown, A.M. and Birnbaumer, L. (1987). Direct activation of mammalian atrial muscarinic potassium channels by GTP regulatory protein G_k. Science *235*, 207-211.

Young, E.F., Ralston, E., Blake, J., Ramachandran, J., Hall, Z.W. and Stroud R.M. (1985). Topological mapping of acetylcholine receptor: evidence for a model with five transmembrane segments and a cytoplasmic COOH-terminal peptide. Proc. Natl. Acad. Sci. USA *82*, 626-630.

Chapter 1

Molecular Characterization of *Shaker*, a Drosophila Gene That Encodes a Potassium Channel

Alexander Kamb, Linda E. Iverson, Mark A. Tanouye

Division of Biology 216-76

California Institute of Technology

Pasadena, California 91125 USA

(Cell, Vol. 50, July 31, 1987, 405-413)

Summary

The Drosophila *Shaker* (*Sh*) gene appears to encode a type of voltage-sensitive potassium (K^+) channel called the A channel. We have isolated *Sh* as part of a 350 kb chromosomal walk. The region around *Sh* contains four identified transcription units. We find that *Sh* corresponds to a very large transcription unit encompassing a total of about 95 kb of genomic DNA and split by a major 85 kb intron. *Sh* has multiple hydrophobic domains that have a high probability of being membrane-spanning, consistent with the proposal that it encodes an ion channel.

Introduction

Potassium (K^+) channels are a heterogeneous group of ion channels that are similar in their ability to select for K^+ over other ions, but are different in details of activation, inactivation, and kinetics (Latorre and Miller, 1983; Hille, 1984). They contribute significantly to several physiological functions, for example, action potential repolarization, cardiac pacemaking, neuron bursting, and possibly learning and memory (Hodgkin and Huxley, 1952; Noble, 1975; Giles and Noble, 1976; Connor and Stevens, 1971; Carew and Sahley, 1986). Despite the physiological importance of K^+ channels, virtually nothing is known about their biochemical nature. Several obstacles stand in the way of K^+ channel purification: they are rare membrane proteins; they have no antibodies and few well-characterized ligands; and they have not, as yet, been amenable to molecular cloning via cross-hybridization with other known channel genes.

Given these difficulties, we undertook the molecular cloning of the first K^+ channel gene using primarily genetic criteria. As we shall explain, the major problem is to develop stringent enough standards to: 1) be certain that we are dealing with a K^+

channel structural gene; and 2) identify the gene once it has been cloned.

Several investigators have proposed that the Drosophila *Shaker* (*Sh*) gene encodes a structural component of a voltage-sensitive K^+ channel called the A channel (reviewed in Salkoff and Tanouye, 1986; Tanouye et al., 1986). Voltage-clamp of muscle membrane shows that all *Sh* mutations alter A current (I_A), but not other ionic currents (Salkoff and Wyman, 1981; Salkoff, 1983; Wu and Haugland, 1985). Phenotypes vary from mutants that completely lack I_A (e.g. Sh^{102}, Sh^{KS133}, Sh^{M}, W32) to others that possess reduced levels (e.g. Sh^{E62}, Sh^{rKO120}, B55). In addition, one mutant has alterations in kinetic and voltage sensitivity properties that may reflect a subtle change in channel structure (Sh^{5}). Gene dosage analysis shows that some mutations encode abnormal gene products, whereas others cause channel underproduction (Tanouye et al., 1981; Salkoff, 1983; Timpe and Jan, 1987). The proposal that *Sh* encodes a K^+ channel arises from the specificity of the mutant defects for a single channel class, the variety of phenotypes seen among the different mutants, and the gene dosage behavior.

The *Sh* K^+ channel defect has a variety of physiological and behavioral consequences: nerve fibers show broadened action potentials and multiple firings; neuromuscular junctions show prolonged transmitter release; all appendages show violent shaking and twitching under ether anesthesia (Tanouye and Ferrus, 1985; Jan et al., 1977; Ganetzky and Wu, 1982a, b; Trout and Kaplan, 1973). Many of these *Sh* phenotypes are altered in double mutant combinations with other excitability mutations. The mutation *ether-a-go-go*, which may also affect K^+ currents, greatly enhances many *Sh* phenotypes. Interestingly, many *Sh* phenotypes are completely suppressed by the mutation *no action potential*, which affects Na^+ channels (Ganetzky and Wu, 1982, 1983).

A genetic description of the X chromosome 16F region that contains *Sh*

provides a basis for localizing the cloned gene (Figure 1). Central to this analysis is that we have obtained and genetically mapped mutations for virtually every gene in the vicinity of *Sh* (about 0.1% of the X chromosome) (Tanouye et al., 1981; Ferrus and Tanouye, 1981; Ferrus and Tanouye, unpublished results). The availability of such a detailed map allows a categorical allotment of cloned DNA segments to sites either flanking or within the *Sh* locus. *Sh* is defined by several apparent point mutations (e.g. Sh^{E62}, Sh^{KS133}, Sh^{102}, Sh^{5}) and translocation breakpoints (e.g. B55, LC, W32) all of which have IA defects. *Sh* point mutations map to two sites that have been separated by recombination. *Sh* is flanked by several closely linked genes identified by recessive lethal mutations. Two of these genes that lie distal to *Sh* (l_2, l_3) have been useful in the isolation and identification of the *Sh* gene described in this paper.

Results

Cloning *Shaker* by Chromosomal Walking

Sh was cloned in a chromosomal walk through the 16F region of the X chromosome that covered 350 kb of overlapping phage and cosmid clones. The walk initiated from the cDNA clone adm 135 H4 (Wolfner, 1980) that mapped to a 3-8 band interval defined distally by the breakpoint of JC153 (16E2-4) and proximally by the breakpoint of B55 (16F1-3) (Figure 2A-C). We walked bidirectionally until the identification of the most distal *Sh* translocation breakpoint (B55) allowed us to determine proximal-distal orientation. The walk continued proximally until all three *Sh* translocation breakpoints had been identified (see below). This report is concerned with the proximal 160 kb of the walk, which is depicted in Figure 3.

Mapping Three *Shaker* Translocation Breakpoint Mutations

We anticipated that, as for other Drosophila genes, *Sh* translocation breakpoint mutations

would cluster near or within a single transcription unit, thereby pinpointing the gene on the molecular map. We mapped all three *Sh* translocation breakpoint mutations, B55, LC, and W32, by using genomic DNA blots to examine restriction fragment-length polymorphisms (Figure 4A). The locations of the B55 and LC breakpoints were verified by *in situ* hybridization to mutant polytene chromosomes (Figure 2C-E), whereas the locations of the B55 and W32 breakpoints were confirmed by genomic DNA blots, using Df(1)B55d/W32p flies (see Experimental Procedures).

All three *Sh* breakpoints map to the chromosomal walk, suggesting that the gene must be within or very close to the cloned region (Figure 5). However, the breakpoints do not cluster at a single location on the molecular map; rather, they are scattered over 65 kb. The B55 breakpoint is located at about molecular map position 34, LC at about 56, and W32 at about 97. These results are somewhat surprising since each of the three breakpoints causes similar phenotypes. Several possibilities are consistent with these observations: 1) *Sh* may be a large gene; 2) *Sh* may be a complex locus containing more than one gene; or 3) *Sh* may be subject to position effects; that is, breakpoints located some distance away from the gene may affect its expression.

Identification of Transcription Units in the *Shaker* Region

Two transcription units were identified by examining RNA blots with subcloned DNA fragments from throughout the proximal part of the chromosomal walk. RNA for these blots was isolated from second instar larvae, third instar larvae, mid-late stage pupae, and adults. I_A is known to be present at several of these stages. One of the identified transcription units (F-a) gives rise to a family of transcripts related to the adm 135 H4 cDNA. F-a is located at about molecular map position 3 (Figure 5). The second identified transcription unit (F-b) gives rise to a moderately abundant, 4.5 kb transcript. F-b is located at about molecular map position 18.

Interestingly, we were unable to detect transcripts in the region most likely to

contain *Sh*, the 65 kb region bounded by the *Sh* translocation breakpoints. Using nick-translated probes we estimate that our RNA blots were sensitive enough to detect transcripts in this region at 1/1000 the level of cytoplasmic actin 5C mRNA, or about one copy of mRNA per 100,000 poly (A+) RNA molecules. As an alternative method for detecting transcription units, we screened several cDNA libraries. Hybridization probes were pools of subcloned fragments representing an 80 kb stretch of DNA encompassing the *Sh* translocation breakpoints. Two transcription units (F-c and F-d) were identified by this method. Each is defined by only a single cDNA clone isolated from one late pupal library. The F-c cDNA is located at about molecular map position 49 (however, see below); F-d is located about position 73. Transcripts corresponding to F-c and F-d have not yet been identified on RNA blots from several developmental stages using the cDNAs as probes, indicating that these transcription units are expressed only at a low level (see Experimental Procedures).

Any of these four identified transcription units might be affected by the *Sh* translocation breakpoints. F-c and F-d may be slightly better *Sh* candidates because they fall within the region flanked by the translocation breakpoints.

Demonstration that the F-a Transcription Unit Cannot Encode *Shaker*

We eliminated one of the four *Sh* gene candidates by using P element-mediated germline transformation to show that F-a corresponds to the 16F l_2 gene (Figure 1). An 8.8 kb BamHI-SalI genomic DNA fragment centered at molecular map position 3 contains the entire F-a transcription unit. This fragment was subcloned into the P element plasmid vector pPSXIA and the construct was injected into Drosophila embryos. Genetic crosses using a stably transformed line showed that the transferred gene complements the l_2 allele, $l_2\ ^{1359}$; but not l_1 or l_3 alleles (see Experimental Procedures). In addition to identifying F-a as the l_2 gene, these results provide a firm distal boundary for the region

containing *Sh*, since l_2 ^{1359} lies 0.05 cM distal to Sh^{KS133} (Ferrus and Tanouye, unpublished results).

Localization of Two *Shaker* Mutations: One Point Mutation and One Spontaneous Mutation

The mapping of one apparent point mutation (Sh^{E62}) and one spontaneous mutation (Sh^{M}) provided compelling evidence that at least part of *Sh* must be encoded by the F-c transcription unit. Our strategy again exploited genetic information available for *Sh* as we generated recombinant chromosomes that, from genetic criteria, had crossover sites located between the Sh^{E62} mutation and the Sh^{102} mutation (Figure 6). The fact that Sh^{E62} was originally induced in a Canton-Special (CS) chromosome and Sh^{102} in an Oregon-R (OR) chromosome provides a means for mapping the crossover site since these two Drosophila strains show frequent restriction fragment length polymorphisms. A wildtype chromosome generated by recombination between Sh^{E62} and Sh^{102} will show an OR polymorphism pattern that shifts to a CS pattern at the point of crossover. By examining polymorphisms in the recombinant, we can find where the pattern shifts and thereby determine where the crossover occurred.

In this way we mapped two recombinants each of which had crossover points between Sh^{E62} and Sh^{102} (Figure 6). We performed a similar analysis on one recombinant with a crossover point between l_3^{583} (OR chromosome) and Sh^{KS133} (CS chromosome). All three recombinant chromosomes tested have crossover points that map to the same 60 kb interval (between map positions 11 and 70.9) (Figure 5). The results indicate that Sh^{E62} and l_3 must be distal to map position 70.9; therefore, F-d (located between positions 70.8 and 75.3) cannot harbor these mutations. The results indicate further that the Sh^{KS133} and Sh^{102} mutations must be proximal to a polymorphic middle repetitive sequence between map positions 9.8 and 12.3. It is

formally possible that the F-b transcription unit (located between map positions 14.6 and 21.3) could harbor Sh^{KS133} and Sh^{102}; however, we consider this unlikely since it would require that the crossover points of all three recombinants fall within the same 2.3 kb interval, separating the gene and the restriction site polymorphism.

The spontaneous mutation Sh^{M} was mapped by examining restriction-site polymorphisms on genomic DNA blots (Figure 4B). The mutation is associated with a 2.2 kb insertion at about molecular map position 46, apparently within the F-c transcription unit. This finding also provides good evidence that F-c is *Sh*, since many spontaneous mutations in Drosophila arise from the insertion of transposable elements into a gene (Zachar and Bingham, 1982; Modolell et al., 1983; Karch et al., 1985).

Genomic Organization of the F-c Transcription Unit

Taken together, the mapping of Sh^{M} and Sh^{E62} pointed to F-c as a good candidate for encoding at least part of the *Sh* gene. However, breakpoint and recombinant mapping suggested that *Sh* could be, among other things, a large gene or two closely linked genes. Further examination of the F-c cDNA clone revealed interesting features about its genomic organization that help resolve this issue.

The F-c transcription unit encompasses a total of about 96 kb and is split by a major 85 kb intron (Figure 7). A distal exon unit is centered at about map position 49; a proximal exon unit is centered at about position 139. The F-d cDNA maps within the major F-c intron. At present we have no evidence that these two transcription units are related, either structurally or functionally, since the two cDNAs have no sequences in common. Given the structure of the F-c transcription unit, it is difficult to explain how the cDNAs might be partial clones of the same transcript. F-c and F-d may follow the precedent of other Drosophila genes that contain a functionally unrelated gene within their introns (O'Hare, 1986).

Amino Acid Sequence of F-c

The nucleotide sequence for the 938 base pair F-c cDNA was determined by the dideoxynucleotide chain termination method (Sanger et al., 1977). The cDNA sequence is apparently incomplete, since one long open reading frame extends through its entire length. The cDNA nucleotide sequence and an additional 139 nucleotides of 3' sequence, obtained by genomic sequencing, are depicted in Figure 8.

A hydrophobicity profile of the deduced amino acid sequence reveals three putative transmembrane segments (Figure 9). They are: 1) a stretch of 19 uncharged amino acid residues (hydrophobicity index: 3.3); 2) a stretch of 19 amino acid residues with 18 of them uncharged (hydrophobicity index: 1.4); 3) a stretch of 18 amino acid residues with 17 of them uncharged (hydrophobicity index: 1.8). Hydrophobicity indices greater than 1.6 have a high probability of being membrane-spanning (Kyte and Doolittle, 1982).

A computer search of the National Biological Research Foundation protein sequence database detected no significant homology with any of more than 4000 published amino acid sequences, including known kinases, phosphatases, and G proteins. The F-c sequence shows only weak homology to other membrane proteins, including the nicotinic acetylcholine receptor (AChR) and the sodium (Na^+) channel (Figure 8). However, the F-c sequence is rich in arginine residues, and it has been suggested that arginine rich sequences in the Na^+ channel may constitute a voltage-sensing domain (Noda et al., 1984; Catterall, 1986).

Discussion

We propose that the *Sh* gene is contained in the region covered by our chromosomal walk. We suggest further from genetic criteria that one of the transcription units identified in this region, F-c, corresponds to *Sh*. Finally, we provide arguments

supporting the proposition that *Sh* encodes a voltage-sensitive K^+ channel. These propositions are consistent with all the facts and are attractive because they provide reasonable explanations for several distinct features of *Sh* genetics and electrophysiology. A rigorous demonstration of some points remains difficult, however, because of the large size and the low level of expression of *Sh*.

The F-c Transcription Unit Corresponds to the *Shaker* Gene

A variety of evidence indicates that three 16F genes that had been previously identified genetically (l_2, l_3, and *Sh*) are all contained in the cloned region. Mapping l_2 and l_3 mutations provides a firm distal limit for the *Sh* gene. Rigorous evidence for having cloned all of *Sh* requires the identification of a non-*Sh* proximal marker, and we have not yet located l_4, which is the closest known gene proximal to *Sh*. Evidence that at least part of the *Sh* gene has been cloned comes from the inclusion of all three known *Sh* translocation breakpoint mutations (B55, LC, and W32), one spontaneous allele (Sh^M), and one apparent point mutation (Sh^{E62}) within the chromosomal walk.

Four transcription units, F-a, F-b, F-c, and F-d, were identified in the cloned region by isolating cDNA clones and probing RNA blots. A major question remains: Do these four account for all of the transcription units in the region? If so, then arguments presented below identifying F-c as the *Sh* gene are convincing. Although it is possible that we have missed transcription units with very low levels of expression, we were successful in detecting two with expression at less than about one copy of mRNA per 100,000 poly (A+) RNA molecules (F-c and F-d).

The physical mapping of the Sh^{E62} and Sh^M mutations provide the strongest evidence that the F-c transcription unit corresponds to the *Sh* gene. Sh^{E62} is an apparent point mutation: It is ethylmethanesulfonate-induced; it causes no cytological aberration on polytene chromosomes; it does not alter frequencies of recombination; and

it shows no restriction fragment-length polymorphisms on genomic DNA blots. The Sh^{E62} defect is likely to be within the distal exon unit of F-c, as shown by genomic DNA blot analysis of recombinant chromosomes. The possibility that F-b harbors Sh^{E62} is not rigorously excluded by our analysis. However, a recent report showing that the proximal element of B55 complements Sh^{E62} reinforces this view, since we show here that this element does not contain the F-b transcription unit (Timpe and Jan, 1987). F-c remains the only candidate for harboring Sh^{E62} of the transcription units we have identified. A similar conclusion comes from mapping the spontaneous mutation Sh^{M}, which is associated with an insertion into the distal exon unit of F-c. This result also suggests a correspondence between the F-c transcription unit and Sh^{M}, since many spontaneous mutations in Drosophila arise from the insertion of transposable elements into a gene (Zachar and Bingham, 1982; Modolell et al., 1983; Karch et al., 1985). Taken together, the results on mapping these two mutations, Sh^{E62} and Sh^{M}, strongly implicate F-c as encoding at least part of *Sh*.

Further evidence that the F-c transcription unit encodes the *Sh* gene comes from an analysis of the translocation breakpoints that are scattered over a 65 kb region of DNA. The F-c transcription unit is also large, encompassing a total of about 95 kb of genomic DNA. The large target size of F-c provides an explanation for why *Sh* is such a highly mutable locus (Figure 1). The LC and W32 breakpoints and Sh^{M} interrupt F-c providing an explanation for why I_A is completely eliminated in these mutants (Salkoff, 1983; Wu and Haugland, 1985). However, the B55 breakpoint appears not to interrupt F-c, providing a plausible reason that I_A is present in this mutant, albeit at reduced levels. Finally, the widely spaced exon units of F-c can explain why apparent point mutations of *Sh* map to sites that can be separated by recombination (Figure 1); the large size of F-c makes intragenic recombination likely. These explanations satisfy the major arguments contained in a previous proposal that *Sh* is a complex of closely linked and

functionally related genes (Tanouye et al., 1981) and offer the simpler explanation that *Sh* is a single large gene with an unusual structure.

Arguments that *Shaker* Encodes a Potassium Channel

The proposal that *Sh* encodes a structural component of the A-type K^+ channel arises from the specificity of the defect for this channel class, the variety of channel phenotypes affected, and the gene dosage behavior (reviewed in Salkoff and Tanouye, 1986; Tanouye et al., 1986). Most alternative explanations for *Sh*, for example, that it is a channel regulator, are excluded. However, the possibility that *Sh* encodes a protein that specifically modifies A channels, such as a kinase, a phosphatase, or a glycosyl transferase, has never been rigorously dealt with. The main difficulty with the modifier model is the notion that it must have several independently mutable functions that control channel number, kinetics, activation voltage, and probably conductance. From gene dosage analyses we must also postulate that mutant and normal modifiers are able to compete for functionally independent modification sites. These limitations leave the possibility of a modifier open, but unlikely.

The amino acid sequence of F-c presented here argues further against a channel modifier model for *Sh*, since the gene shows no apparent homology to known kinases, phosphatases, or G proteins. F-c has multiple hydrophobic domains that have a high probability of being membrane-spanning. In this aspect it generally resembles the only two sequenced ion channels, the Na^+ channel and the nicotinic AChR, which also have multiple membrane-spanning domains (Noda et al., 1983, 1984, 1986). Compelling sequence arguments are difficult, however, since there is little direct amino acid homology between the sequenced regions of F-c and these other channels (Figure 8); and the F-c sequence is still incomplete.

In summary, we propose that the F-c transcription unit encodes a structural component of the type A K^+ channel. Its large size, genomic structure, and location

provide good evidence that F-c corresponds to the *Sh* gene, thereby linking it with previous genetic and electrophysiological studies supporting *Sh* as a K^+ channel structural gene. Evidence that it is a membrane protein with no obvious homology to known kinases or phosphatases argues against other explanations for *Sh*. Rigorous confirmation of these suggestions awaits a thorough *Sh* mutant transcript analysis, gene transfer studies, and full-length nucleotide sequence determination, although these may be difficult because of the large gene size and low expression. The complete sequence of *Sh* may reveal whether it encodes a subunit of a K^+ channel or the entire channel protein. Detailed structural comparisons with the Na^+ channel will be particularly interesting because of similarities in activation, inactivation, and kinetics. Both channels are highly ion-selective, albeit to different ions. Finally, there is hope that the A channel nucleotide sequence will help provide access to other major K^+ channel classes such as the delayed rectifier, the inward rectifier, and the Ca^{+2}-dependent K^+ channels.

Experimental Procedures

Fly Strains

The x-ray-induced chromosomal translocations, T(1;Y)B55, T(1;Y)W32, and T(1;3)Sh^{LC} are referred to as B55, W32, and LC, respectively. The duplication Dp(1;3)JC153 is obtained from the insertional translocation, T(1;3)JC153. Df(1)B55d/W32p, constructed using the distal element of B55 and the proximal element of W32, is deficient for DNA between the two breakpoints. These chromosomes and their associated *Sh* phenotypes have been described previously (Tanouye et al., 1981; Salkoff, 1983).

Standard Techniques

The following standard methods are described by protocols and combined references in

Davis et al. (1980) and Maniatis et al. (1982): screening of recombinant DNA libraries, preparation of plasmid and phage DNA, restriction enzyme analysis, agarose gel electrophoresis, Southern blot transfers to nitrocellulose filters, DNA fragment isolation, nick translations, hybridizations to DNA on filters, and subcloning of DNA fragments into plasmid vectors. Drosophila genomic DNA was isolated as described in Bender et al. (1983).

Recombinant DNA Libraries

Lambda phage libraries were constructed in the EMBL 3 vector with wildtype Canton-Special (CS), Oregon-R (OR), and Urbana-Special (US) genomic DNA. Larval DNA was partially digested with Sau3A to an average size of 25 kb as judged by electrophoresis through 0.3% agarose gels. The partially digested DNA was centrifuged (26K rpm, SW27 rotor, 18 h) through a linear sucrose gradient (10-40%). Fractions containing 15-22 kb fragments were pooled, ethanol-precipitated, and used for ligation with EcoRI- and BamHI-digested EMBL 3 vector DNA. *In vitro* packaging extracts were from Amersham. Partial CS and OR genomic libraries were constructed similarly, except that insert DNA was digested to completion with either BamHI or EcoRI and used without size selection. A CS library in the Charon 4 vector is described in Maniatis et al. (1978). An OR cosmid library in the pTL5 vector was constructed in collaboration with Dr. H. Sun using methods similar to Pirrotta (1986).

Mapping *Sh* breakpoints and Sh^M

Mapping of the translocation breakpoints by genomic DNA blot analysis was carried out using DNA extracted from Sh^M, B55, LC, W32, and Df(1)B55d/W32p flies. CS and OR DNA were used as controls. DNA was digested with restriction endonucleases, size-fractionated by agarose gel electrophoresis, and blotted to nitrocellulose. Blots were probed with ^{32}P-labeled genomic DNA clones or subcloned fragments from the

appropriate chromosomal region. The breakage and rejoining of chromosomes that occur in a simple translocation cause a restriction fragment-length polymorphism between mutant and normal DNA. In a blot comparing mutant and normal genomic DNA, the restriction fragment carrying the breakpoint is absent and novel fragments are created. DNA between the B55 and W32 breakpoints is deleted in Df(1)B55d/W32p flies and the corresponding restriction fragments are missing in DNA blots.

RNA Analysis and Screening cDNA Libraries

RNA was prepared from animals frozen and ground to a powder in liquid nitrogen. The powder was added to a solution of guanidinium isothiocyanate (4 M), 2-mercaptoethanol (1 M), sodium acetate (0.05 M, pH 4.5), and EDTA (0.001 M), and RNA was pelleted through a cushion of 5.7 M CsCl. Poly (A)+ RNA was purified by oligo (dT)-cellulose column chromatography, size-fractionated by agarose-formaldehyde gel electrophoresis (usually 10 ug/lane), and blotted to nitrocellulose. For identifying transcription units, the entire chromosomal region from map positions 10 to 130 was subcloned as fragments averaging about 5 kb. These subcloned fragments were used to probe RNA blots representing second instar larva, third instar larva, mid-late stage pupa, and adult stages. RNA blots covering similar developmental stages plus embryonic stages (0-5 h, 5-10 h, 10-15 h, and 15-20 h) were examined for F-c and F-d expression, using their respective cDNAs as a probe.

Several cDNA libraries that we screened were gifts from Dr. L. Kauvar. They were constructed from RNA of second instar larvae, third instar larvae, early pupae, late pupae, or adults in lambda gt10 (Poole et al., 1985). A single cDNA clone representing F-c and one representing F-d were identified from among about 500,000 clones in a late pupal library.

Germ-Line Transformation and Establishment of Transformed Lines

For P element-mediated germ-line transformation, a recombinant EMBL 3 phage containing an 18 kb BamHI fragment located between molecular map positions -1.7 and +16.3 was cleaved with SalI to generate 8.8 and 9.2 kb Drosophila fragments. This DNA was mixed with XhoI-cleaved pPSXIA vector DNA (a gift from Dr. J. Posakony), ligated, and used to transform HB101 E. coli. A recombinant plasmid containing the 8.8 kb fragment was identified by restriction enzyme analysis and designated pPF-a. The plasmid uses *Adh+* as a selectable marker in flies. Using methods similar to those described by Goldberg et al. (1983), purified pPF-a plasmid DNA was mixed 5:1 with helper P element DNA (pPI25.1) and injected into 30 min.-old Adh^{fn23} *pr cn* embryos.

Four ethanol-resistant lines were generated. A second chromosome transformant, (pPF-a) Adh^{fn23} *pr cn*, was homozygosed and used in genetic studies. Males (genotype: (pPF-a) Adh^{fn23} *pr cn*) were crossed to females (genotype: f^5 lethal *os*/FM7a). Progeny males (genotype: f^5 lethal *os* /Y; (pPF-a) Adh^{fn23} *pr cn*/+) were produced when the lethal allele was l_2 ^{1359} (the only allele of the l_2 complementation group). No such males were observed when the lethal allele was l_1 ^{1579} (an l_1 allele) or l_3 ^{583} (an l_3 allele).

Restriction Fragment-Length Polymorphism Mapping of Recombinants

Two recombinants (genotype: f^+ Sh^+ os^+), each with a crossover point between the Sh^{E62} and Sh^{102} mutations, were obtained from among 20,581 male progeny of heterozygous females (genotype: f^5 Sh^{E62}/Sh^{102} *os*). Control flies were f^5 Sh^{E62} and Sh^{102} *os* males.

One recombinant (genotype: f^+ $l_3{}^+$ Sh^+ *os*), with a crossover point between l_3 ^{583} (an l_3 allele) and Sh^{KS133}, was obtained from among about 8,000 male progeny of

transheterozygous females (genotype: $f^5\ l_3\ ^{583}\ os/Sh^{KS133}$; note that hemizygous l_3 males do not survive). Control flies were Sh^{KS133} males, $f^5\ l_3\ ^{583}\ os$/FM6 females, and FM6 females.

The following restriction site polymorphisms were scored by genomic DNA blots:

1) a 5.9 kb middle repetitive sequence between molecular map positions -11.0 and -16.9 present in CS but not OR chromosomes and revealed in a HindIII digest probed with a subcloned 0.7 kb BamHI fragment from positions -9.3 to -10.0 (note: this BamHI fragment is duplicated at about position -2);

2) A 2.5 kb middle repetitive sequence between positions 9.8 and 12.3 is present in CS but not OR or Urbana-S (US) chromosomes and revealed in a BamHI digest probed with adm 135 H4 or in a HindIII digest probed with the US genomic DNA clone lambda 5.5 (located between positions 3.5 to 20.8).

3) An EcoRI site at position 70.9 is present in OR but not CS chromosomes and revealed in an EcoRI digest probed with a subcloned 5.8 kb EcoRI fragment from position 66.6 to 72.4.

4) An EcoRI site at position 106.5 is present in CS but not in OR chromosomes and is revealed in an EcoRI digest probed with a 5.6 kb SalI fragment corresponding to position 103.1 and 108.7 isolated from lambda 22.5.

DNA Sequence Analysis

Both strands of the F-c cDNA clone were sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977), using templates prepared from recombinant M13 virions (Sanger et al., 1980) and synthetic primers prepared in the Caltech microchemical facility. Genomic DNA from the phage clone lambda 14.6 was sequenced after subcloning restriction fragments into pUC8.

Acknowledgements

We thank D. Ballinger, U. Banerjee, M. Gautam, J. Pollock, M. Ramaswami, J. Tseng-Crank, and K. VijayRaghavan for comments. This research was supported by the Pfeiffer Research Foundation, and USPHS grant NS21327-01 to M. T. L. I. was supported by a Muscular Dystrophy Association Fellowship and the E. S. Gosney Fund. A. K. was supported by NRSA training grant GM07616. M. T. is a McKnight Foundation Scholar and a Sloan Foundation Fellow.

References

Bender, W., Spierer, P. and Hogness, D.S. (1983). Chromosomal walking and jumping to isolate DNA from the *Ace, rosy*, and *bithorax* complex in Drosophila melanogaster. J. Mol. Biol. *168*, 17-33.

Carew, T.J. and Sahley, C.L. (1986). Invertebrate learning and memory: from behavior to molecules. Ann. Rev. Neurosci. *9*, 435-487.

Catterall, W.A. (1986). Molecular properties of voltage-sensitive sodium channels. Ann. Rev. Biochem. *55*, 953-985.

Connor, J.A. and Stevens, C.F. (1971). Prediction of repetitive firing behaviour from voltage clamp data on an isolated neurone soma. J. Physiol. (Lond.) *213*, 31-53.

Davis, R.L. and Davidson, N. (1984). Isolation of the Drosophila melanogaster *dunce* chromosomal region and recombinational mapping of dunce sequences with restriction

site polymorphisms as genetic markers. Molec. Cell. Biol. *4*, 358-367.

Davis, R.W., Botstein, D. and Roth, J.R. (1980). *Advanced Bacterial Genetics. A Manual for Genetic Engineering*. Cold Spring Harbor Laboratory: Cold Spring Harbor, New York.

Ferrus, A. and Tanouye, M.A. (1981). Fine genetic analysis of the *Shaker* locus. Calif. Inst. Techn. Ann. Rep, 247.

Ganetzky, B. and Wu, C.-F. (1982a). Drosophila mutants with opposing effects on nerve excitability: genetic and spatial interactions in repetitive firing. J. Neurophysiol. *47*, 501-514.

Ganetzky, B. and Wu, C.-F. (1982b). Indirect suppression involving behavioral mutants with altered nerve excitability in Drosophila melanogaster. Genetics *100*, 597-614.

Ganetzky, B. and Wu, C.-F. (1983). Neurogenetic analysis of potassium currents in Drosophila: synergistic effects on neuromuscular transmission in double mutants. J. Neurogenet. *1*, 17-28.

Giles, W. and Noble, J. (1976). Changes in membrane currents in bullfrog atrium produced by acetylcholine. J. Physiol. (Lond.) *261*, 103-123.

Goldberg, D.A., Posakony, J.W. and Maniatis, T. (1983). Correct developmental expression of a cloned alcohol dehydrogenase gene transduced into the Drosophila germ line. Cell *34*, 59-73.

Hafen, E., Basler, K., Edstrom, J.-E. and Rubin, G. (1987). *Sevenless*, a cell-specific homeotic gene of Drosophila, encodes a putative transmembrane receptor with a tyrosine kinase domain. Science *236* , 55-63.

Hille, B. (1984). *Ionic Channels of Excitable Membranes*. Sinauer: Sunderland, Ma.

Hodgkin, A.L. and Huxley, A.F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. (Lond.) *117*, 500-544.

Jan, Y.N., Jan, L Y, and Dennis, M. (1977). Two mutations of synaptic transmission in Drosophila. Proc. R. Soc. Lond. B *198*, 87-108.

Karch, F., Weiffenbach, B., Peifer, M., Bender, W., Duncan, I., Celniker, S., Crosby, M. and Lewis, E.B. (1985). The abdominal region of the *bithorax* complex. Cell *43*, 81-96.

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

Langer-Safer, P.R., Levine, M. and Ward, D. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. Proc. Natl. Acad. Sci. USA *79*, 4381-4385.

Latorre, R. and Miller, C. (1983). Conduction and selectivity in potassium channels. J. Memb. Biol. *7*, 11-30.

Lipman, D.J. and Pearson, W.R. (1985). Rapid and sensitive protein similarity searches. Science *227*, 1435-1441.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Estratiadis, A. (1978). The isolation of structural genes from libraries of eukaryotic DNA. Cell *15*, 687-701.

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

Modolell, J., Bender, W, and Meselson, M. (1983). Drosophila melanogaster mutations suppressible by the *suppressor of hairy-wing* are insertions of a 7.3-kilobase mobile element. Proc. Natl. Acad. Sci. USA *80*, 1678-1682.

Noble, D. (1975). *The Initiation of the Heartbeat.* Clarendon Press: Oxford.

Noda, M., Ikeda, T., Kayano, T., Suzuki, H., Takashima, H., Kurasaki, M., Takahashi, H, and Numa, S. (1986). Existence of distinct sodium channel messenger RNAs in rat brain. Nature (Lond.) *320*, 188-192.

Noda, M., Shimizu, S., Tanabe, T., Takai, T., Kayano, T., Ikeda, T., Takahashi, H., Nakayama, H., Kanaoka, Y., Minamino, N., Kangawa, K., Matsuo, H., Raftery, M. A., Hirose, T., Inayama, S., Hayashida, H., Miyata, T. and Numa, S. (1984). Primary structure of Electrophorus electricus sodium channel deduced from cDNA sequence. Nature (Lond.) *312*, 121-127.

Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y.,

Hirose, T., Takashima, H., Inayama, S., Miyata, T. and Numa, S. (1983). Structural homology of Torpedo californica acetylcholine receptor subunits. Nature (Lond.) *302*, 528-532.

O'Hare, K. (1986). Genes within genes. Trends in Genetics *2*, 33.

Pardue, M. and Gall, J. (1975). Nucleic acid hybridization to the DNA of cytological preparations. In *Methods in Cell Biology 10*. ed. Prescott, D. Academic Press: New York.

Pirrotta, V. (1986). Cloning Drosophila genes. In *Drosophila, A Practical Approach*. ed. Roberts, D.B. IRL Press: Oxford.

Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985). The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell *40*, 37-43.

Rubin, G.M. and Spradling, A.C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science *218*, 348-353.

Salkoff, L. (1983). Genetic and voltage clamp analysis of a Drosophila potassium channel. Cold Spring Harbor Symp. Quant. Biol. *48*, 221-231.

Salkoff, L. and Tanouye, M.A. (1986). Genetics of ion channels. Physiol. Rev. *66*, 301-329.

Salkoff, L. and Wyman, R.J. (1981). Genetic modification of potassium channels in Drosophila *Shaker* mutants. Nature *293*, 228-230.

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

Sanger, F., Coulson, A.R., Barrell, G. G., Smith, A.J.F.,and Roe, B. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. *143*, 161-164.

Spradling, A.C. and Rubin, G.M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. Science *218*, 341-347.

Tanouye, M.A. and Ferrus, A. (1985). Action potentials in normal and *Shaker* mutant Drosophila. J. Neurogenet. *2*, 253-271.

Tanouye, M.A., Ferrus, A. and Fujita, S.C. (1981). Abnormal action potentials associated with the *Shaker* complex locus of Drosophila. Proc. Natl. Acad. Sci. USA *78*, 6548-6552.

Tanouye, M.A., Kamb, C.A., Iverson, L.E. and Salkoff, L. (1986). Genetics and molecular biology of ionic channels in Drosophila. Ann. Rev. Neurosci. *9*, 255-276.

Timpe, L.C. and Jan, L.Y. (1987). Gene dosage and complementation analysis of the *Shaker* locus in Drosophila. J. Neurosci. *7*, 1307-1317.

Trout III, W.E. and Kaplan, W.D. (1973). Genetic manipulation of motor output in *Shaker* mutants of Drosophila. J. Neurobiol. *4*, 495-512.

Wolfner, M. (1980). Ecdysone-responsive genes of the salivary gland of Drosophila

melanogaster. Ph.D. thesis, Stanford Univ., Stanford, Calif.

Wu, C.-F. and Haugland, F. (1985). Voltage clamp analysis of membrane currents in larval muscle fibers of Drosophila: alteration of potassium currents in *Shaker* mutants. J. Neurosci. *5*, 2626-2640.

Zachar, Z. and Bingham, P.M. (1982). Regulation of *white* locus expression: the structure of mutant alleles at the *white* locus of Drosophila melanogaster. Cell *30*, 529-541.

Figure Legends

Figure 1. Genetic Representation of the X Chromosome 16F Region that Contains *Sh*

Depicted are x-ray-induced mutations that are the breakpoints of chromosomal translocations. Three of the breakpoints disturb I_A and provide a cytological location for *Sh*. These are: B55 (16F1-4), LC (16F1-2), and W32 (16F3-6). Distal to these breakpoints is one, JC153 (16E2-4), that does not affect I_A and hence is not considered a *Sh* mutation (note that JC153 also contains a proximal X chromosomal breakpoint not dealt with in this report).

Ethylmethanesulfonate-induced *Sh* and recessive lethal mutations that do not alter frequencies of recombination are referred to as point mutations, although this has not been confirmed molecularly. Because *Sh* mutations are leg-shaking dominants, complementation tests are difficult; however, by recombination, three point mutations map to a single site (Sh^{KS133}, Sh^{102}, Sh^{5}). Recombination defines a second site, placing Sh^{E62} 0.01 cM distal to Sh^{KS133}. *Sh* appears to be a highly mutable locus with a total of 32 identified alleles. All have been mapped generally to the 16F region, but are excluded from this figure since they have not been localized well enough to assign them to either the Sh^{E62} site or the Sh^{KS133} site. The relative positions between *Sh* breakpoints and point mutations have not been determined rigorously, since the breaks interfere with recombination.

Three recessive lethal loci map between JC153 and B55 (l_1, 3 alleles; l_2, one allele; l_3, 5 alleles). These recessive lethal mutations are the nearest genetic markers that border *Sh* distally. Several lethal complementation groups that map proximal to *Sh* have also been identified but are omitted from the figure for clarity. In several cases the linear

order of lethals is estimated from recombination frequency and has yet to be confirmed by three factor crosses. This very precise genetic and cytogenetic mapping of 16F make the cloning strategy feasible. The figure depicts data from several sources (Tanouye et al., 1981; Ferrus and Tanouye, 1981; Ferrus and Tanouye, unpublished).

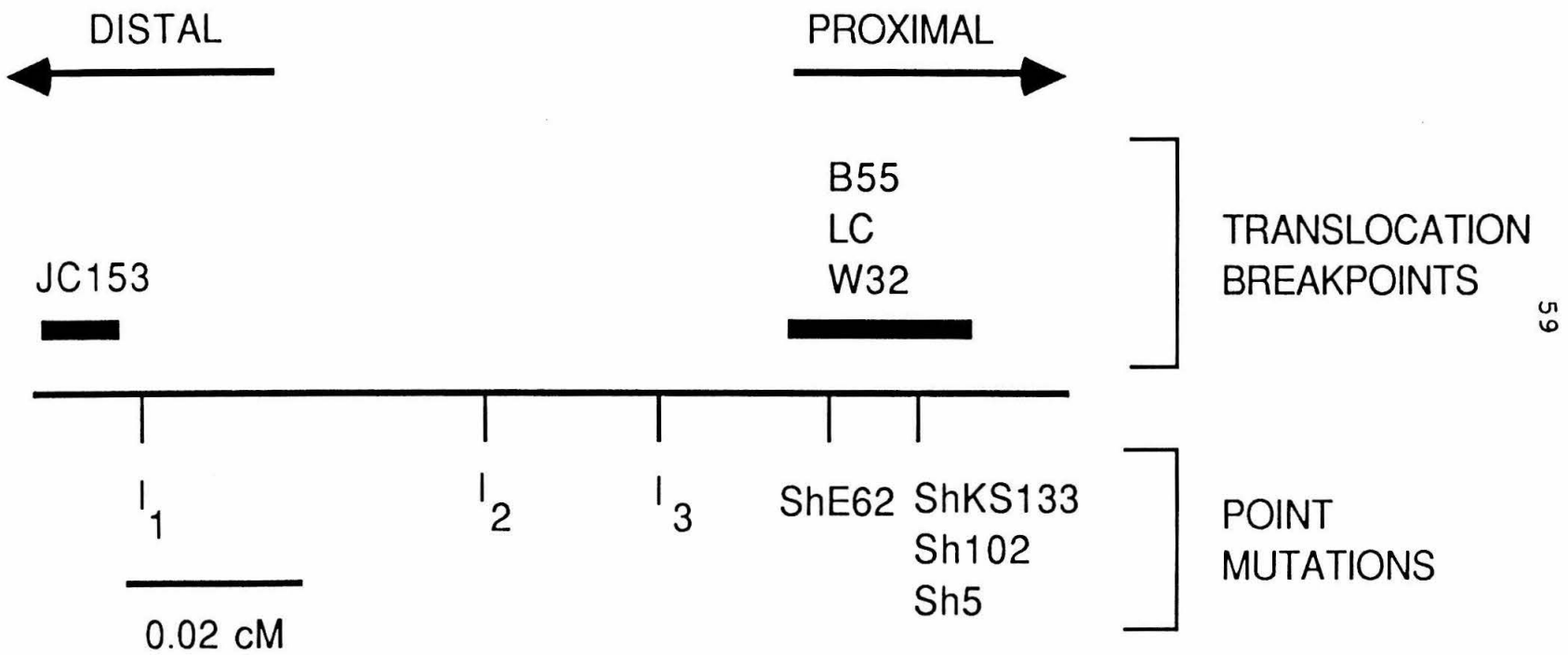

DISTAL
PROXIMAL
B55
LC
W32
JC153
TRANSLOCATION
BREAKPOINTS
I1
I2
I3
ShE62
ShKS133
Sh102
Sh5
POINT
MUTATIONS
0.02 cM

Figure 2. Cytological Representation of the Chromosomal Walk and Mapping of the B55 Translocation Breakpoint by *In Situ* Hybridization to Polytene Chromosomes

These data show that the start of our chromosomal walk was between the breakpoints of JC153 and B55. They also show that the B55 breakpoint is located between molecular map positions 27 and 47.

(a) The entire recombinant plasmid adm 135 H4, the startpoint of the walk, hybridizes to the 16F region of a normal Canton-S polytene chromosome (arrow).

(b) Adm 135 H4 sequences are proximal to 16E2-4 (and distal to 17A-B) since there is hybridization to a Dp(1;3)JC153 chromosome (arrow).

(c) Adm 135 H4 sequences are distal to 16F1-4 since there is hybridization to the distal (arrow) but not the proximal element of a B55 chromosome.

(d) The genomic DNA clone lambda 14.6 (map position 26.2 to 46.1) spans the B55 breakpoint since the recombinant phage hybridizes to both the proximal and distal elements (arrows) of the B55 chromosome.

(e) The proximal 9.0 kb of lambda 14.6 (an EcoRI-SalI fragment) hybridizes to the proximal (arrow) but not the distal element of a B55 chromosome.

Hybridization probes were labeled by nick translation using biotin-11-dUTP and detected by a streptavidin-peroxidase complex (D) (Langer-Safer et al., 1982; Hafen et al., 1987) or labeled using ^{35}S-dATP (A,B) or ^{3}H-dCTP (C,E) and detected by autoradiography (Pardue and Gall, 1975). Similar hybridization experiments were used to examine the LC translocation breakpoint (data not shown).

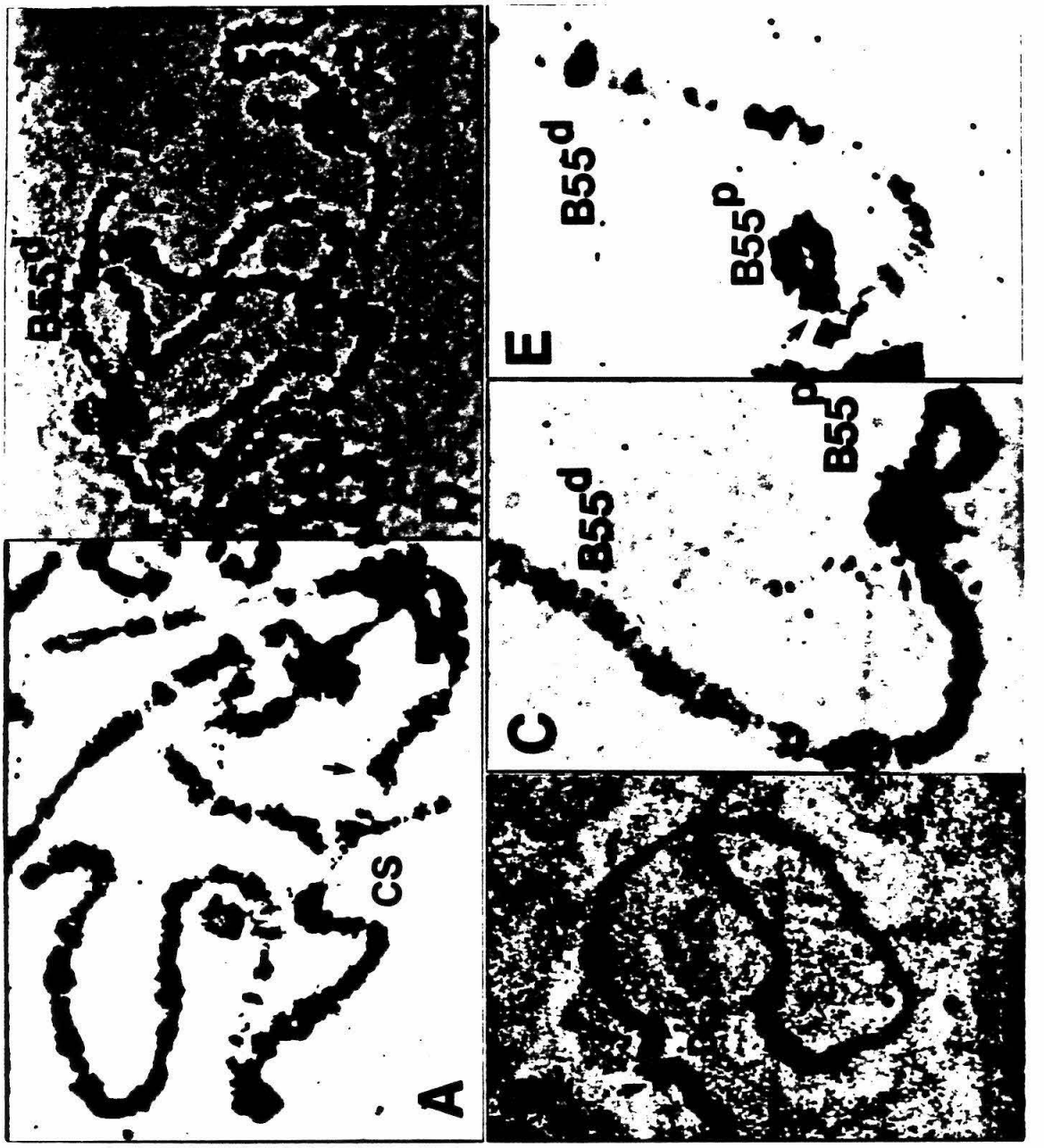

A
CS
B55d
C
B55d
B55p
E
B55d
B55p

Figure 3. Restriction Map of the *Shaker* Region and Genomic DNA Clones Obtained from Chromosomal Walking

Starting from the cDNA clone adm 135 H4, overlapping DNA clones were isolated from genomic DNA libraries constructed in bacteriophage lambda and cosmid vectors. Unique DNA sequences from the ends of identified clones were used to isolate additional genomic DNA clones, thereby extending the walk bidirectionally. For simplicity, only representative clones from the proximal extent of the walk are depicted.

The consensus map resulting from digestion with several restriction enzymes has the EcoRI sites shown. Distances from the EcoRI site in adm 135 H4 are shown in kilobases.

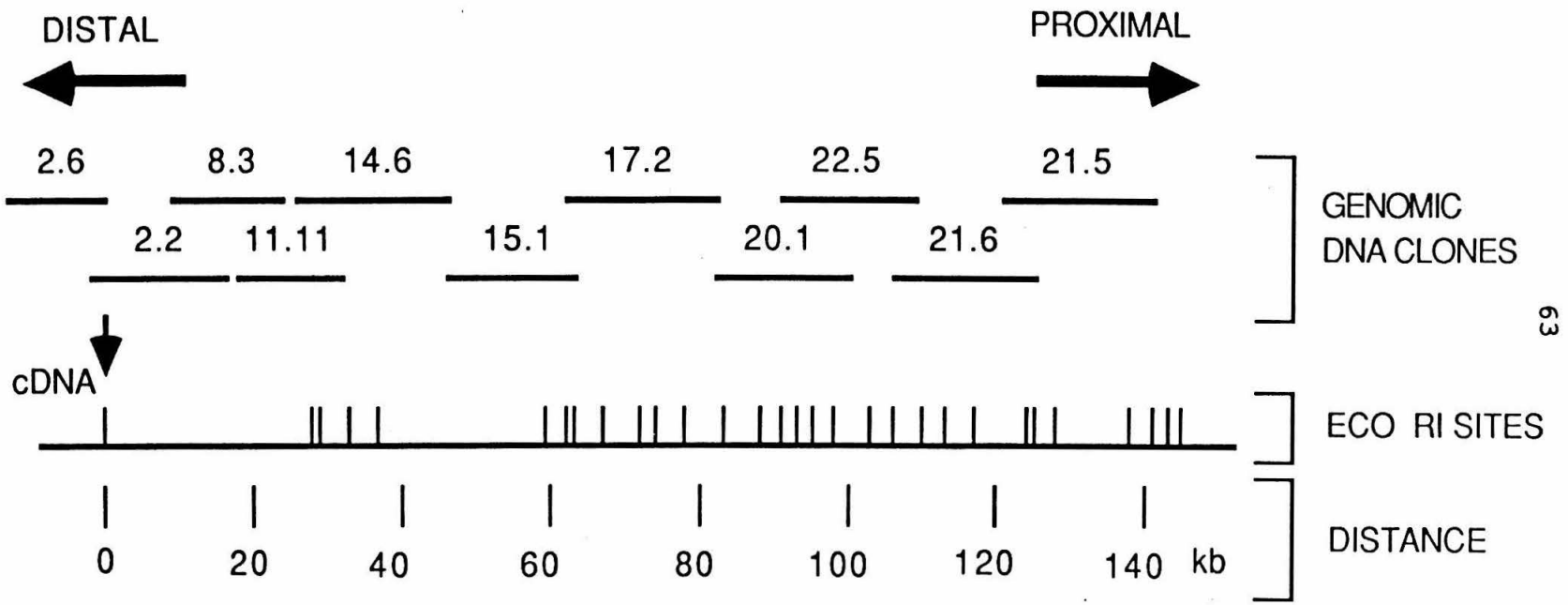

DISTAL
PROXIMAL
2.6
8.3
14.6
17.2
22.5
21.5
2.2
11.11
15.1
20.1
21.6
GENOMIC
DNA CLONES
cDNA
ECO RI SITES
0
20
40
60
80
100
120
140
kb
DISTANCE

Figure 4. Localizing the B55 Translocation Breakpoint and the Sh^M Mutation by Restriction Fragment-Length Polymorphisms on Genomic DNA Blots

(a) The B55 translocation breakpoint maps to a HindIII-BamHI fragment between molecular map positions 33.4 and 34.9. OR (lane 1) and B55 (lane 2) genomic DNA was digested with EcoRI, separated by electrophoresis in an agarose gel, transferred to nitrocellulose, and probed with the phage clone lambda 14.6 that extends from map position 26.2 to 46.1. A 4.4 kb OR fragment is absent in the mutant and two novel fragments (marked by stars) are created. Similar blots with other enzymes were used to complete the localization (data not shown). Similar blots were used to identify polymorphisms associated with the LC and W32 translocation breakpoints (data not shown).

(b) The spontaneous allele, Sh^M, is associated with a 2.2 kb insertion in an XbaI fragment located between molecular map positions 45.5 and 46.1. A blot of XbaI-digested CS (lane 3) and Sh^M (lane 4) genomic DNA was probed with a subcloned EcoRI-SalI fragment of lambda 14.6 that extends from molecular map position 37.1 to 46.1. A 0.6 kb CS fragment is absent in the mutant and a novel 2.8 kb fragment (marked by a star) is created.

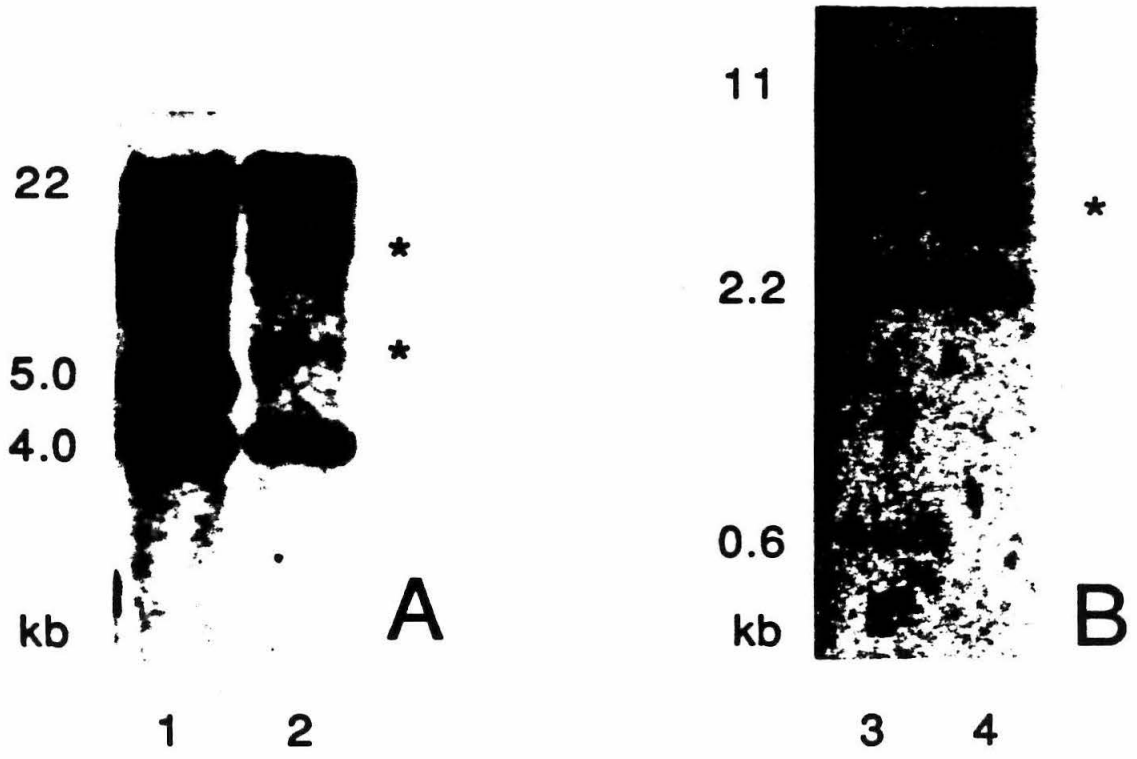

22
5.0
4.0
kb
*
*
A
1
2
11
2.2
0.6
kb
*
B
3
4

Figure 5. Molecular Organization of the *Shaker* Region: Locations of Transcription Units, Several *Shaker* Mutations, and Recombination Crossover Points

The *Sh* mutations were localized by examining whole genome DNA blots for restriction fragment-length polymorphisms (see Figure 4). The breakpoints of B55 and LC mapped to single restriction fragments: B55 to a HindIII-BamHI fragment between molecular map positions 33.4 and 34.9 and LC to a HindIII-EcoRI fragment between 54.1 and 59.1. W32 eliminates an EcoRI site at map position 98.0. Restriction fragments distal to an EcoRI site at 95.2 and proximal to a HindIII site at 98.7 are normal. Thus, we assign the W32 breakpoint to the interval between 95.2 and 98.7. Sh^{M} mapped to an XbaI fragment between positions 45.5 and 46.1.

The F-a and F-b transcription units were localized by probing RNA blots with subcloned genomic fragments. F-a, which corresponds to the original cDNA clone adm 135 H4, is located between molecular map positions -1.7 and +7.1. F-b is located between positions 14.6 and 21.3. F-c and F-d locations were determined by mapping their respective cDNA clones. Based on hybridization to blots of restriction enzyme-digested genomic DNA clones lambda 14.6 and lambda 15.1, the F-c cDNA is located between positions 45.9 and 51.9 (see, however, Figure 7). The F-d cDNA localized between positions 70.8 and 75.3.

Crossover points of Sh^{E62}-Sh^{102} and l_3-Sh^{KS133} recombinants fell between polymorphic restriction sites at about map positions 11 and 70.9. Sh^{E62} and l_3 must be located distal to 70.9. Sh^{KS133} and Sh^{102} must be located proximal to l_1 (see Figure 6).

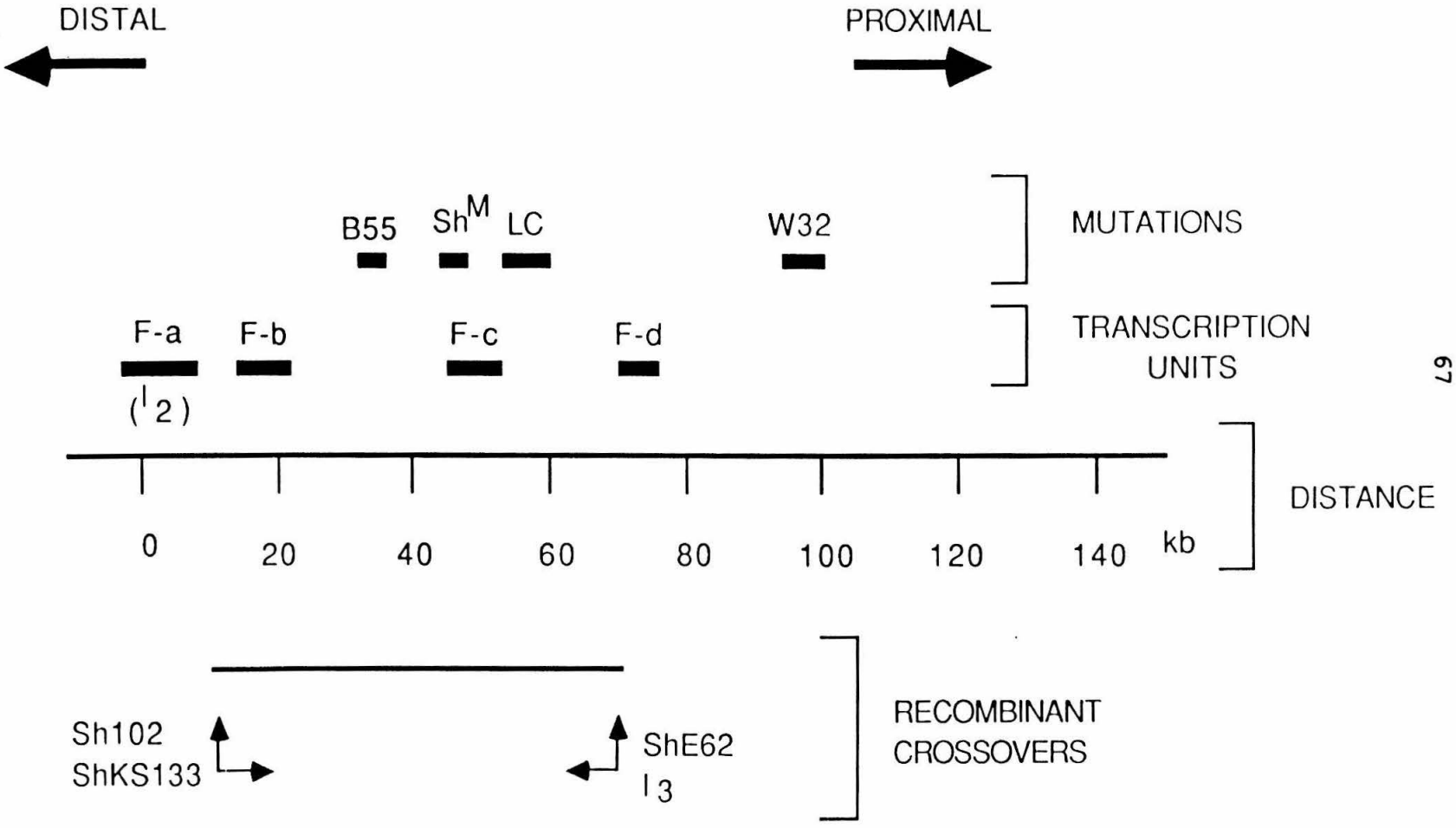

DISTAL
PROXIMAL
B55
Sh^M
LC
W32
MUTATIONS
F-a
F-b
F-c
F-d
TRANSCRIPTION UNITS
(I 2)
0
20
40
60
80
100
120
140
kb
DISTANCE
Sh102
ShKS133
ShE62
I 3
RECOMBINANT CROSSOVERS

Figure 6. Generation of a Wildtype Recombinant Chromosome with a Crossover Point Between Sh^{E62} and Sh^{102}; and Mapping of the Crossover Point Using Restriction Fragment-Length Polymorphisms on Genomic DNA Blots

(a) A recombinant with a crossover point between Sh^{E62} and Sh^{102} (genotype: f^{+} Sh^{+} os^{+}/Y) was isolated from among the male progeny of heterozygous females (genotype: f^{5} Sh^{E62}/Sh^{102} os) (see Experimental Procedures).

(b) The recombinant chromosome has an Oregon-R (OR) restriction pattern distal to the crossover point and a Canton-S (CS) pattern proximal to it. The portion derived from OR is depicted by the solid bar; the portion from CS by the open bar.

(c) A genomic DNA blot was used to score a polymorphic restriction site at about molecular map position 11. The recombinant (lane 1) shows an OR pattern like the Sh^{102} os parental type (lane 2), so this position must be distal to the crossover point.

(d) A DNA blot scoring a polymorphic restriction site at map position 70.9 shows that the recombinant (lane 4) has a CS pattern like the f^{5} Sh^{E62} parental type (lane 6); therefore, this site must be proximal to the crossover point.

These DNA blots show that the crossover point in this recombinant occurs between map positions 11 and 70.9. For further verification additional polymorphic sites at at about map positions -14 and +106 were also scored (data not shown). Genomic DNA was from f^{+} Sh^{+} os^{+}/Y recombinant males (lane 1, 4); Sh^{102} os/Y males (lane 2, 5); and f^{5} Sh^{E62}/Y males (lane 3, 6). Results similar to these were obtained for a second, independently derived Sh^{E62}-Sh^{102} wildtype recombinant and one $l_3{}^{583}$-Sh^{KS133} wildtype recombinant (data not shown).

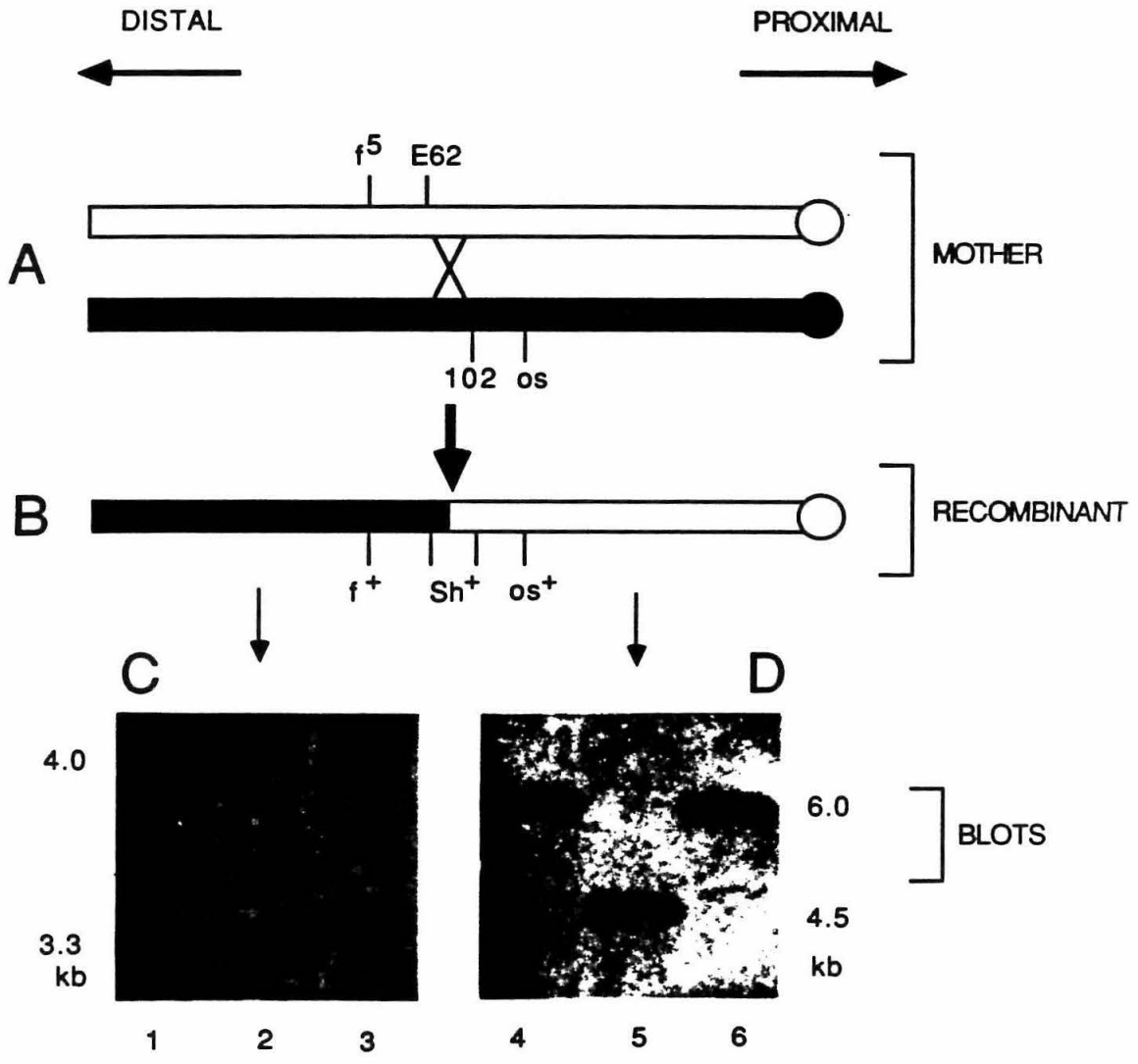

DISTAL
PROXIMAL
f5
E62
A
MOTHER
102
os
B
RECOMBINANT
f+
Sh+
os+
C
D
4.0
3.3
kb
6.0
4.5
kb
BLOTS
1
2
3
4
5
6

Figure 7. F-c is a Large Transcription Unit

The genomic organization of the F-c and F-d transcription units was determined by using the cDNAs to probe DNA blots of recombinant phage clones covering the entire cloned region. F-c is large, encompassing a total of at least 95.6 kb. It is split by a major, 85.1 kb intron. F-d is located within the major F-c intron. A schematic representation of this organization is shown relative to Sh^M and the *Sh* translocation breakpoints. Since the precise boundaries of all the exons have not been determined, the areas indicated by the blocks are not necessarily distinct exons; rather, they represent genomic regions within which exons reside.

(a) The distal exon unit of the F-c cDNA is located between molecular map positions 45.9 and 51.9 based on cDNA hybridization to a set of contiguous restriction fragments of lambda 14.6 and lambda 15.1. The blot is of HindIII-digested lambda 14.6 DNA probed with the F-c cDNA clone.

(b) The proximal exon unit of the F-c transcription unit is located between map positions 137.0 and 141.5 based on cDNA hybridization to a set of contiguous restriction fragments of lambda 21.5. The blot is of HindIII-digested lambda 21.5 DNA probed with the F-c cDNA clone.

To exclude the possibility that the lambda clones shared sequences, lambda 21.5 insert DNA was used to probe cloned DNA blots containing lambda 14.6 and lambda 15.1 DNA; no hybridization was seen (data not shown). To test the possibility that we were missing other major exon units, F-c was used to probe a genomic DNA blot. All genomic restriction fragments could be accounted for by corresponding fragments identified in lambda 14.6, lambda 15.1, and lambda 21.5 (data not shown). Based on location, Sh^M and the breakpoints of LC and W32 should interrupt F-c; B55 may break outside of the transcription unit.

(c) F-d is located between map positions 70.8 and 75.3 based on cDNA hybridization to

a set of contiguous restriction fragments of lambda 17.2. The blot is of HindIII-digested lambda 17.2 DNA probed with the F-d cDNA clone. F-d hybridization to a genomic DNA blot showed that all restriction fragments are accounted for by the lambda 17.2 hybridization (data not shown).

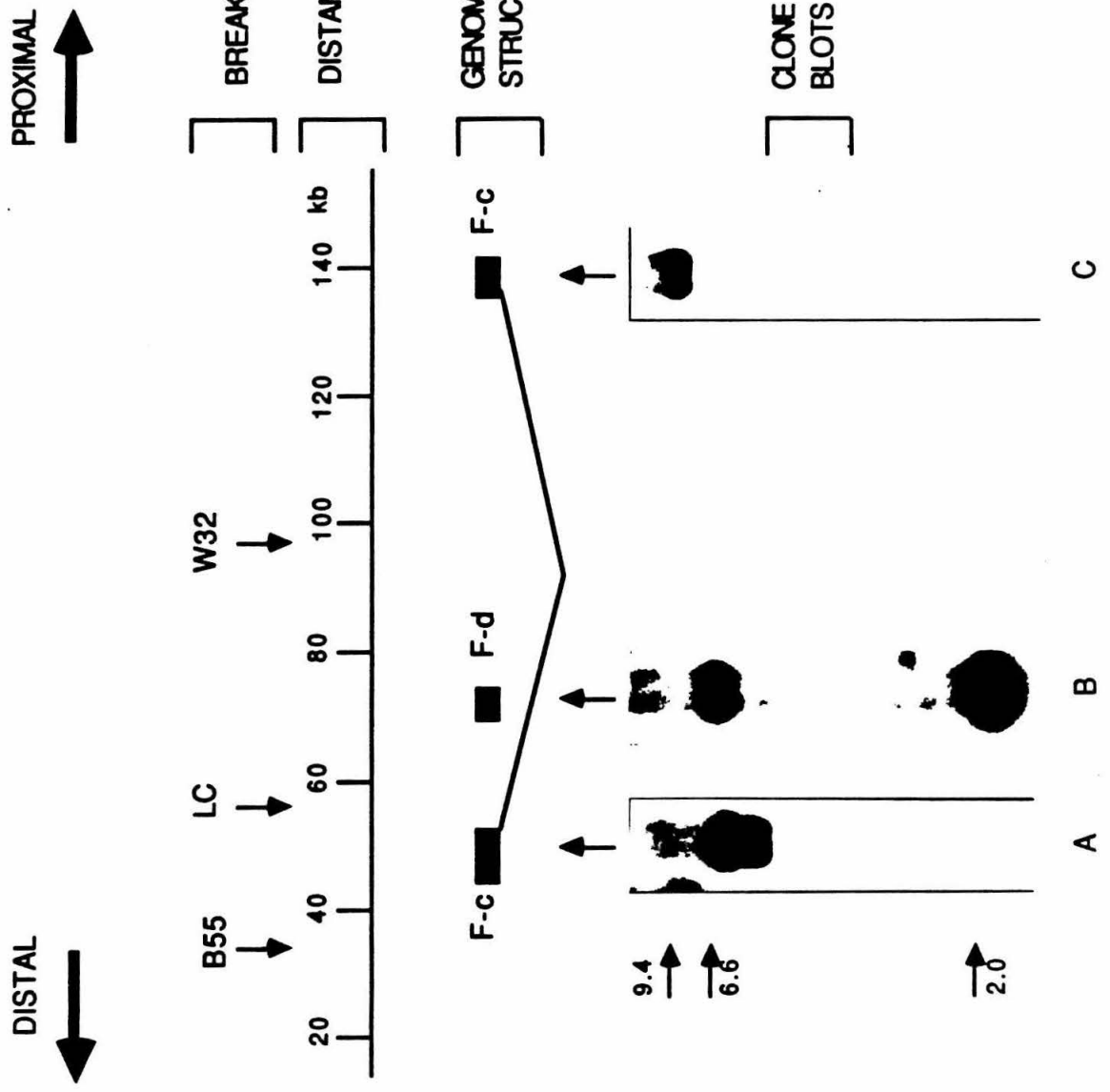

PROXIMAL
DISTAL
BREAKPOINTS
DISTANCE
GENOMIC STRUCTURE
CLONE BLOTS
B55
LC
W32
20
40
60
80
100
120
140
kb
F-c
F-d
F-c
9.4
6.6
2.0
A
B
C

Figure 8. Nucleotide and Deduced Amino Acid Sequence of F-c

Depicted are a 938 nucleotide sequence derived from the F-c cDNA and an additional 139 nucleotides of 3' sequence derived from genomic sequencing. Potential transmembrane sequences are underlined. A potential asparagine-linked N-glycosylation site is marked with a triangle. Arginine residues are marked with circles. The sequence derived from genomic DNA should be interpreted cautiously since it probably runs into an intron and is not double stranded.

A comparison of the deduced F-c sequence with other ion channel sequences (Noda et al., 1983, 1984) by the algorithm of Lipman and Pearson (1985) showed surprisingly limited homology. An 117 amino acid stretch corresponding to the 154-271 region of the nicotinic AChR (Noda et al., 1983) showed the best homology (25 identical residues and 47 conservative substitutions). The homology score, however, was only 55, since 4 gaps are needed to make the alignment. A score of about 100 is generally considered to be significant.

F-c showed some homology (8 identical residues and 6 conservative substitutions) with a 22 amino acid stretch that corresponds to the 1407-1429 region of the eel Na^+ channel (Noda, 1984). The homology score was 42.

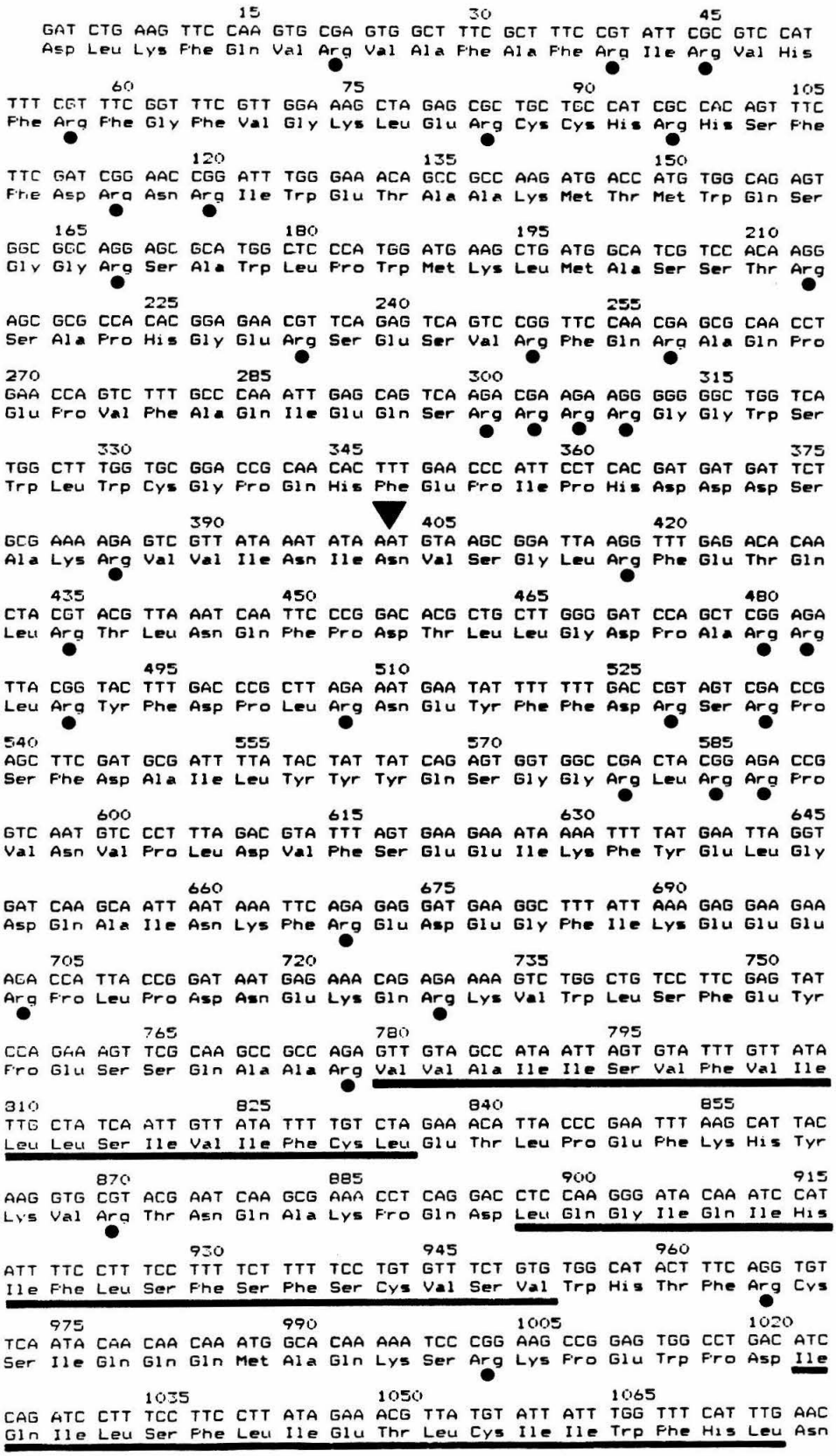
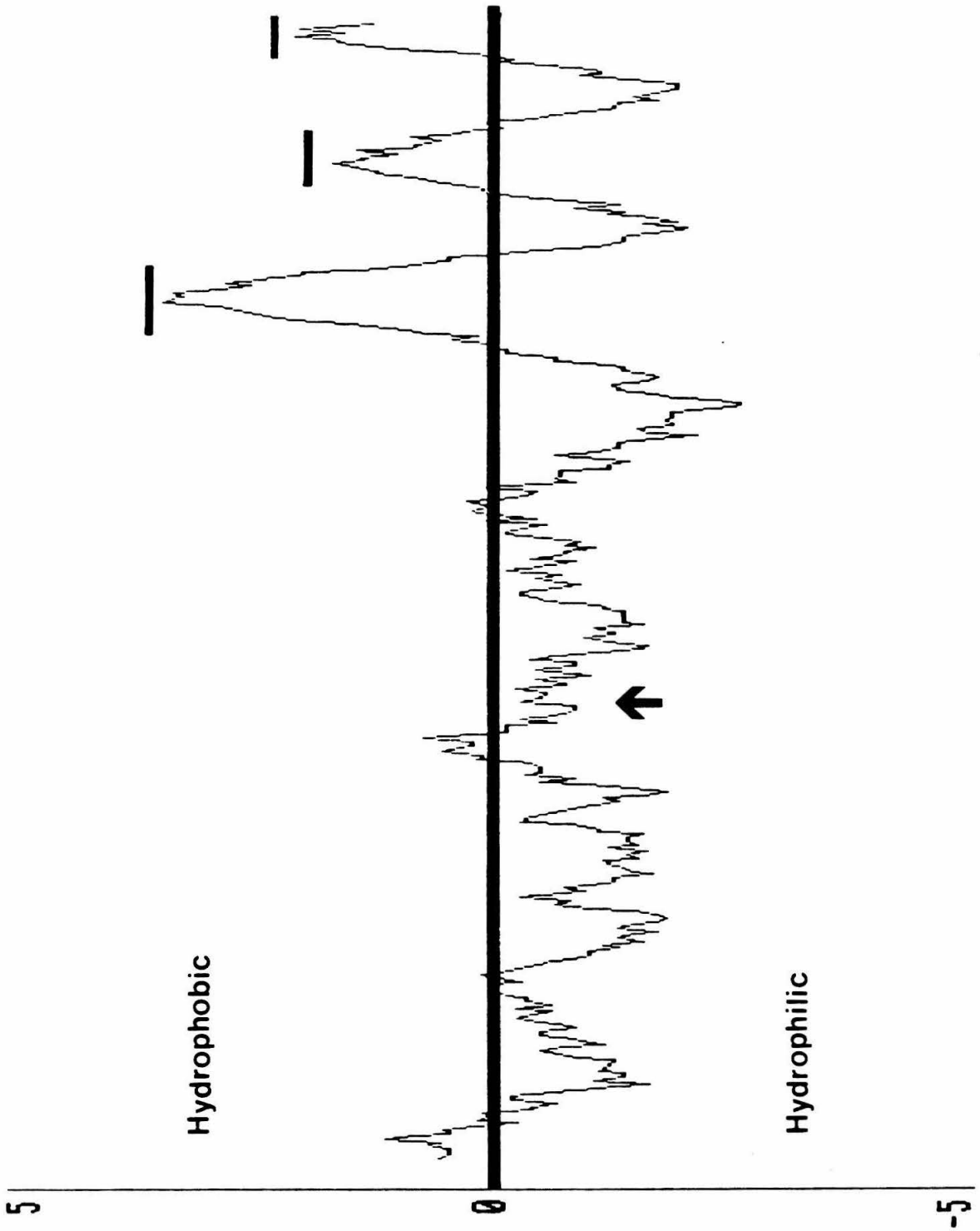

15 30 45
GAT CTG AAG TTC CAA GTG CGA GTG GCT TTC GCT TTC CGT ATT CGC GTC CAT
Asp Leu Lys Phe Gln Val Arg Val Ala Phe Ala Phe Arg Ile Arg Val His
60 75 90 105
TTT CGT TTC GGT TTC GTT GGA AAG CTA GAG CGC TGC TGC CAT CGC CAC AGT TTC
Phe Arg Phe Gly Phe Val Gly Lys Leu Glu Arg Cys Cys His Arg His Ser Phe
120 135 150
TTC GAT CGG AAC CGG ATT TGG GAA ACA GCC GCC AAG ATG ACC ATG TGG CAG AGT
Phe Asp Arg Asn Arg Ile Trp Glu Thr Ala Ala Lys Met Thr Met Trp Gln Ser
165 180 195 210
GGC GGC AGG AGC GCA TGG CTC CCA TGG ATG AAG CTG ATG GCA TCG TCC ACA AGG
Gly Gly Arg Ser Ala Trp Leu Pro Trp Met Lys Leu Met Ala Ser Ser Thr Arg
225 240 255
AGC GCG CCA CAC GGA GAA CGT TCA GAG TCA GTC CGG TTC CAA CGA GCG CAA CCT
Ser Ala Pro His Gly Glu Arg Ser Glu Ser Val Arg Phe Gln Arg Ala Gln Pro
270 285 300 315
GAA CCA GTC TTT GCC CAA ATT GAG CAG TCA AGA CGA AGA AGG GGG GGC TGG TCA
Glu Pro Val Phe Ala Gln Ile Glu Gln Ser Arg Arg Arg Arg Gly Gly Trp Ser
330 345 360 375
TGG CTT TGG TGC GGA CCG CAA CAC TTT GAA CCC ATT CCT CAC GAT GAT GAT TCT
Trp Leu Trp Cys Gly Pro Gln His Phe Glu Pro Ile Pro His Asp Asp Asp Ser
390 405 420
GCG AAA AGA GTC GTT ATA AAT ATA AAT GTA AGC GGA TTA AGG TTT GAG ACA CAA
Ala Lys Arg Val Val Ile Asn Ile Asn Val Ser Gly Leu Arg Phe Glu Thr Gln
435 450 465 480
CTA CGT ACG TTA AAT CAA TTC CCG GAC ACG CTG CTT GGG GAT CCA GCT CGG AGA
Leu Arg Thr Leu Asn Gln Phe Pro Asp Thr Leu Leu Gly Asp Pro Ala Arg Arg
495 510 525
TTA CGG TAC TTT GAC CCG CTT AGA AAT GAA TAT TTT TTT GAC CGT AGT CGA CCG
Leu Arg Tyr Phe Asp Pro Leu Arg Asn Glu Tyr Phe Phe Asp Arg Ser Arg Pro
540 555 570 585
AGC TTC GAT GCG ATT TTA TAC TAT TAT CAG AGT GGT GGC CGA CTA CGG AGA CCG
Ser Phe Asp Ala Ile Leu Tyr Tyr Tyr Gln Ser Gly Gly Arg Leu Arg Arg Pro
600 615 630 645
GTC AAT GTC CCT TTA GAC GTA TTT AGT GAA GAA ATA AAA TTT TAT GAA TTA GGT
Val Asn Val Pro Leu Asp Val Phe Ser Glu Glu Ile Lys Phe Tyr Glu Leu Gly
660 675 690
GAT CAA GCA ATT AAT AAA TTC AGA GAG GAT GAA GGC TTT ATT AAA GAG GAA GAA
Asp Gln Ala Ile Asn Lys Phe Arg Glu Asp Glu Gly Phe Ile Lys Glu Glu Glu
705 720 735 750
AGA CCA TTA CCG GAT AAT GAG AAA CAG AGA AAA GTC TGG CTG TCC TTC GAG TAT
Arg Pro Leu Pro Asp Asn Glu Lys Gln Arg Lys Val Trp Leu Ser Phe Glu Tyr
765 780 795
CCA GAA AGT TCG CAA GCC GCC AGA GTT GTA GCC ATA ATT AGT GTA TTT GTT ATA
Pro Glu Ser Ser Gln Ala Ala Arg Val Val Ala Ile Ile Ser Val Phe Val Ile
810 825 840 855
TTG CTA TCA ATT GTT ATA TTT TGT CTA GAA ACA TTA CCC GAA TTT AAG CAT TAC
Leu Leu Ser Ile Val Ile Phe Cys Leu Glu Thr Leu Pro Glu Phe Lys His Tyr
870 885 900 915
AAG GTG CGT ACG AAT CAA GCG AAA CCT CAG GAC CTC CAA GGG ATA CAA ATC CAT
Lys Val Arg Thr Asn Gln Ala Lys Pro Gln Asp Leu Gln Gly Ile Gln Ile His
930 945 960
ATT TTC CTT TCC TTT TCT TTT TCC TGT GTT TCT GTG TGG CAT ACT TTC AGG TGT
Ile Phe Leu Ser Phe Ser Phe Ser Cys Val Ser Val Trp His Thr Phe Arg Cys
975 990 1005 1020
TCA ATA CAA CAA CAA ATG GCA CAA AAA TCC CGG AAG CCG GAG TGG CCT GAC ATC
Ser Ile Gln Gln Gln Met Ala Gln Lys Ser Arg Lys Pro Glu Trp Pro Asp Ile
1035 1050 1065
CAG ATC CTT TCC TTC CTT ATA GAA ACG TTA TGT ATT ATT TGG TTT CAT TTG AAC
Gln Ile Leu Ser Phe Leu Ile Glu Thr Leu Cys Ile Ile Trp Phe His Leu Asn

Figure 9. Hydrophobicity Plot of the Deduced Amino Acid Sequence of F-c

The plot was generated using the method of Kyte and Doolittle (1982). Horizontal lines indicate the positions of putative transmembrane sequences. The arrow indicates the position of the asparagine residue as a possible site of glycosylation.


Hydrophobic
Hydrophilic
5
0
-5

Chapter 2

Multiple Products of the Drosophila *Shaker* Gene Contribute to Potassium Channel Diversity

Alexander Kamb, Julie Tseng-Crank and Mark A. Tanouye

Division of Biology 216-76

California Institute of Technology

Pasadena, California 91125

(submitted to Neuron)

Summary

K^+ channels are known through electrophysiology and pharmacology to be an exceptionally diverse group of channels. Molecular studies of the *Shaker (Sh)* locus in Drosophila have provided the first glimpse of K^+ channel structure. The sequences of 3 *Sh* cDNA clones have been reported to date; none of the 3 is identical. We have isolated and examined 18 additional *Sh* cDNAs in an attempt to understand the origin, the extent, and the significance of the variability. The diversity is extensive: We have already identified cDNAs representing at least 10 distinct types, and *Sh* could potentially encode 28 or more products. This diversity, however, fits a simple pattern where variable 3' and 5' ends are spliced onto a central constant region to yield different cDNA types. These different *Sh* cDNAs encode proteins with distinct structural features.

Introduction

The molecular biology of potassium (K^+) channels is of interest for several reasons: First, K^+ channels may be present in all eukaryotic cells. Second, they may be the first channel type that arose in evolution. Third, they appear to be targets for modulation by second messenger systems, a trait that may be fundamental to learning and memory. Finally, they are an exceptionally diverse group of channels, with individual classes that vary greatly in kinetics, voltage-dependence, pharmacology, and other properties (Latorre and Miller, 1983; Hille, 1984; Rudy, 1988).

Because of their prevalence and diversity in the nervous system, K^+ channels play a leading role in the control of neuronal activity. Each neuronal type has its own combination of K^+ channel classes, which seems to govern cell-specific firing patterns (Dekin and Getting, 1987; Jahnsen and Llinas, 1984; Llinas, 1984). Understanding the molecular basis for K^+ channel diversity may reveal important details as to how cells

express different channel types, and hence, set excitability properties. In addition, knowledge of K^+ channel structure is central to understanding K^+ channel function, and possibly, channel function in general, since K^+ channels comprise a multitude of channel types with a variety of characteristics.

Recently, we and others have identified and cloned a Drosophila gene called *Shaker (Sh)* that encodes a component of a fast, transient, voltage-dependent, or A-type, K^+ channel (Kamb et al., 1987; Tempel et al., 1987; Baumann et al., 1987). The 3 *Sh* cDNA clones so far described are surprisingly different from each other: The genomic organization and the nucleotide sequences of theses clones show gross differences. This variation results in conceptual protein products that differ both at the amino terminus and at the carboxyl terminus.

The discrepancies observed among these *Sh* cDNA clones lead to several questions: What general features are conserved among different *Sh* cDNA clones? What is the extent of the diversity among clones? What possible explanations may account for this variation? And could the diversity manifested by these *Sh* clones contribute to the diversity observed in K^+ channels?

To address these questions, we have isolated and analyzed 18 new *Sh* cDNA clones. Comparisons reveal that *Sh* cDNA diversity is indeed extensive; however, the clones can be grouped into general classes that possess distinct structural features. In addition, the structure of these cDNAs and the proteins they encode prompt us to speculate about the nature of the A channel in Drosophila and about a possible mechanism for generating K^+ channel protein diversity.

Results

Structure of *Sh* cDNA Clones:

A Constant Central Region with Variable 5' and 3' Ends

The general features of 21 different *Sh* cDNA clones discussed in this report are summarized in Table 1. Eighteen cDNAs are new isolates; 17 clones derive from a single adult head library (H clones), and one clone comes from an embryo library (E1). For comparison we also discuss 3 cDNAs described previously: pupal clone P1 (Kamb et al., 1987), adult head clone ShA1 (Tempel et al., 1987), and embryo clone gt141 (Baumann et al., 1987).

These clones encompass a diverse group of structures, but the pattern of diversity is simple: variable 5' and 3' ends are joined to a central constant region yielding the different cDNAs. This pattern is evident from a comparison of cDNA restriction maps (Figure 1, Table 1). In the 863 b.p. constant region, the maps of nearly all *Sh* cDNAs are identical. Flanking the constant region, the maps diverge; however, it is possible to group variable ends into at least three 5' end classes and four 3' end classes based on restriction map similarities and other criteria (see Experimental Procedures).

The diversity of *Sh* cDNA 5' ends is extensive; we have observed a total of 7 different types. Three major classes account for 14 of the 21 cDNAs: class I (P1, E1, H9, H15, H29); class II (H1, H4, H19, H21, H34, ShA1); and class III (H2, H6, H37). Four other cDNAs have unique 5' ends (H11, H13, H20, gt141) and may constitute additional classes. We are unable to classify 3 clones, because they do not extend far enough in the 5' direction (H31, H33, H38).

Four major classes account for 19 of the twenty-one 3' ends: class i (H6, H19, H20, H33, H37, H38, ShA1); class ii (H9, H15, H29); class iii (E1, H2, gt141); and class iv (H1, H4, H21, H31, H34). Two clones (P1, H11) have unique ends, one

clone (H13) does not extend far enough in the 3' direction to catalog.

There are enough combinations of ends represented among the clones to suggest that 5' ends may assort independently with 3' ends. For example, 3 different 5' ends are associated with class iii 3' ends: E1 (class I), H2 (class II), and gt141 (unique). Two different 3' ends are found with class II 5' ends: H19 (class i) and H34 (class iv). Thus, in order to describe fully the cDNA type, it is necessary to designate both the 5' and 3' class. We describe the classification of individual cDNAs in the form: (5' class.3' class).

Genomic Organization of *Sh* cDNAs: Complex Splicing Pathways

The genomic organization of *Sh* cDNAs (Figure 2) was determined by mapping the clones onto the *Sh* chromosomal walk (Kamb et al., 1987). The cDNA sequences span at least 108 kb of genomic DNA from map position +33 to +141, consistent with our earlier description of *Sh* as a large gene. Sequences from the cDNA 5' ends account for most of this distance: they scatter over 95 kb on the chromosome, from +46 to +141. The genomic locations of cDNA 3' ends are much less variable: they appear to map over a distance of only 4 kb, from +33 to +37.

All *Sh* cDNA clones that fall into the same 5' end class have similar genomic DNA hybridization patterns; for example, 5' end class I clones hybridize to two sets of restriction fragments separated by 85 kb, as described previously for P1 (Kamb et al., 1987). Class II cDNAs have a genomic organization similar to ShA1 (Tempel et al., 1987), and characteristically hybridize to contiguous restriction fragments between +64 and +80. The hybridization pattern of class III cDNA clones, distinguished by hybridization to restriction fragments between +54 and +62, has not been described previously. Clones that possess unique 5' ends show a variety of hybridization patterns.

Complex splicing pathways are apparently used to generate relatively short 5' end cDNA sequences from a large chromosomal region. For example, most cDNAs

extend only 200-600 b.p. in the 5' direction from a conserved SalI site (Figure 1), but this cDNA sequence can be distributed over 95 kb of genomic DNA (Figure 2). The splicing pathways are diverse: Class I clones are derived from the removal of an 85 kb intron sequence, whereas class II and class III clone 5' ends are assembled from many scattered microexons. The microexonic structure of class II ends is illustrated by H34: the 600 b.p. of cDNA sequence 5' to the conserved SalI site is spread over 70 kb of genomic DNA contained in 5 sets of discontinuous restriction fragments. The genomic structure of class III ends is illustrated by H37: The 300 b.p. of cDNA sequence 5' to the SalI site is scattered over roughly 45 kb contained in 3 sets of fragments.

Variation in the 3' ends of *Sh* cDNA clones is not as readily apparent at this level of analysis as 5' end diversity. Several clones (H1, H4, H9, H15, H20, H21, H29, H31, H33, H34, H37) extend further distally than the others; they hybridize to a 4 kb restriction fragment centered at map position +35. However, it is not clear in many cases if this differential hybridization reflects bona fide diversity or simply incomplete extension of the cDNA strands during cloning.

Analysis of *Sh* sequences: The Constant Region

In order to analyze *Sh* cDNA diversity at the nucleotide level, we compared the DNA sequences of 14 clones (Figure 3, Table 2): complete nucleotide sequences for the 3 previously reported clones (Kamb et al., 1987; Tempel et al., 1987; Baumann et al., 1987); complete nucleotide sequences for 4 new clones (E1, H2, H4, H29); and partial nucleotide sequences for 7 new clones (H1, H9, H11, H15, H21, H34, H37). Multiple members of each 5' and 3' class are represented among these sequences in order to provide a good basis for comparison (see Experimental Procedures).

The constant region for each of the clones is essentially identical and extends 863 b.p. The 5' divergence point, where clones from the three 5' end classes (I, II, III) and the unique 5' end of gt 141 all diverge, occurs 251 b.p. upstream of the conserved

SalI site. The 3' divergence point, where class iii 3' ends diverge, occurs 612 b.p. downstream of the SalI site. A secondary 3' divergence point, where the other 3' end classes (i, ii, iv) all diverge from one another, lies 913 b.p. downstream of the SalI site.

Conceptual translation shows that the central constant region contains a single long open reading frame stretching from end to end. It encodes a 287 a.a. segment which is conserved in each *Sh* polypeptide (Figure 4). Three potential membrane-spanning domains, described previously (Tempel et al., 1987; Baumann et al., 1987; Kamb et al., 1987), are observed in hydropathy plots (Figure 5). The constant region contains 2 potential N-linked glycosylation sites and a particularly interesting sequence (YFIT), which is present in the putative G protein binding domain of the beta-adrenergic receptor (Dohlman et al., 1987). A related sequence (YFIS) is present at amino acid position 1360 in a dihydropyridine binding protein, a putative Ca^{+2} channel (Noda et al., 1987).

Variable Amino Termini

Clones from the different 5' end classes all have translational start codons in-frame with the constant region, which lie 146 b.p. (class I), 182 b.p. (class II), and 92 b.p. (class III) upstream of the 5' divergence point. The amino termini of proteins encoded by class I, II, and III cDNA variable 5' ends are 49 a.a., 61 a.a., and 31 a.a. long, respectively. They have no significant homology to one another, nor to any of the 4,000 sequences in the PIR protein sequence database. There are differences among the classes in the number of potential N-linked glycosylation sites: class I-type amino termini have one site; class II-types have no sites; and class III-types have two sites (Figure 4). No potential membrane-spanning domains are apparent from an examination of hydropathy plots of the amino terminal ends (Figure 5).

Variable Carboxyl Termini:

Potential Membrane-Spanning Domains

Clones from the different 3' end classes all have stop codons in-frame with the constant region, which lie 804 b.p. (class i), 471 b.p. (class ii), 3 b.p. (class iii), and 924 b.p. (class iv) downstream of the primary 3' divergence point. The carboxyl termini of proteins encoded by class i, ii, iii, and iv cDNA variable 3' ends are 268 a.a., 157 a.a., 1 a.a., and 308 a.a. long, respectively. Because their carboxyl ends are so short, class iii polypeptides are considerably smaller than the other *Sh* polypeptides, varying in complete size from 304 a.a. (gt 141, unique.iii) to 337 a.a. (E1, I.iii). Proteins encoded by the other cDNA classes range in size from 493 a.a. (H29, I.ii) to 656 a.a. (H4, II.iv). Hydropathy plots of all class iii proteins reveal only the 3 hydrophobic domains contributed by their respective constant segments (Figure 5). Three additional hydrophobic domains are observed in hydropathy plots of class i, ii, and iv carboxyl ends. Thus, each of the complete polypeptides of these classes contains 6 hydrophobic potential membrane-spanning domains: 3 contributed by the constant segment and 3 from the carboxyl end.

In addition to hydrophobic domains, the class i, ii, and iv carboxyl ends each includes an amphipathic potential membrane-spanning sequence that is similar to the S4 sequence present in Na^+ channels and in a dihydropyridine-binding protein (Noda et al., 1984; Noda et al., 1987). The S4 domain is believed to constitute an ion channel voltage sensor (Catterall, 1986; Guy and Seetharamulu, 1986). The *Sh* S4 consists of 7 repeated copies of the three amino acid sequence: Arg-X-X (where X is any hydrophobic or uncharged residue and Lys is occasionally substituted for Arg). S4 is absent from all class iii polypeptides; in the other classes it is interposed between the third and fourth hydrophobic domains.

The amino acid sequences of proteins encoded by classes i, ii, and iv cDNA variable 3' ends do not diverge until a point within the fifth hydrophobic segment.

Thus, these three classes are essentially identical, not only through the constant region (that includes the first 3 hydrophobic domains), but also through S4, the fourth hydrophobic domain, and part of the fifth hydrophobic domain.

Variable Carboxyl Termini: Other Homologies

Sh protein sequences have other noteworthy features: class i carboxyl ends contain a consensus phosphorylation site, as reported by Tempel et al. (1987); a similar site is present in class iv, but not in class ii or iii ends (Figure 4). Class ii and iv ends each possess a single potential asparagine-linked N-glycosylation site. The carboxyl termini of class i and class iv encoded proteins have a disproportionate number of glutamine residues near their C-terminal ends: Class i carboxyl ends are 30% glutamine (19 of the last 60 residues) including a stretches of 3 and 4 glutamines; class iv carboxyl ends are 35% glutamine (35 of the last 100 residues) with stretches of 5 and 11 glutamines.

A particularly interesting form of variation occurs in a 258 b.p. sequence near the 3' end of class i and iv clones (Figure 6, Figure 7). In this segment, the nucleotide sequences are similar but not identical (83% identity). Apparently, these classes of mRNA are differentially spliced such that one of these two 258 b.p. sequences is chosen. The majority of nucleotide changes (74%) do not alter the amino acid sequence, suggesting that the 2 sequences are subject to selection pressure through their respective polypeptides.

Discussion

In this report, we describe 2 basic forms of conceptual proteins encoded by *Sh* cDNAs: the more common form contains 7 potential membrane-spanning domains (6 hydrophobic domains and one amphipathic S4 domain); the other form contains 3 potential membrane-spanning hydrophobic domains. The surprising structural diversity

seen among different *Sh* cDNAs encoding these 2 basic protein forms is sufficient to account for differences among 3 previously published *Sh* cDNA sequences (Kamb et al., 1987; Tempel et al., 1987; Baumann et al., 1987). The general pattern of *Sh* cDNA diversity is simple: The different clones contain variable 5' and 3' ends joined to a constant central region.

How is *Sh* cDNA Diversity Generated?

Sh cDNA diversity apparently reflects genuine *in vivo Sh* gene product diversity; that is, the different clones probably correspond to mRNAs that collectively encode a variety of polypeptides. This diversity appears to be due mainly to a differential splicing mechanism that generates variable sequences and splices them onto the 5' and 3' ends of a constant segment. Several observations support this explanation: 1) in several instances the same 5' or 3' end has been independently isolated from different libraries; 2) clones often diverge at precisely the same nucleotide; 3) sequences at the 5' ends of clones are non-random, with splicing patterns that link only particular exons together; and 4) similar sequences, such as the two 258 b.p. sequences in class i and iv 3' ends, are subject to selection pressure through the polypeptides they encode since the majority of nucleotide changes are silent with respect to the protein sequence. Differential processing is known in several other instances to be a general mechanism for producing multiple protein products from a single gene (Breitbart et al., 1987)

Superimposed on the diversity generated by differential splicing may be an additional level of complexity caused by the use of alternative transcriptional start and stop sites. For example, differential polyadenylation sites may be used to generate variable 3' end lengths in class iv clones such as H4 and H34 (Figure 3). It is less clear whether or not some of the variability observed among 5' ends is due to different transcriptional promotors, but class III mRNAs may initiate from a site distal to class I initiation sites.

Other explanations do not account for the bulk of *Sh* cDNA diversity in a reasonable way. For example, artifacts that are due to tandem ligations are unlikely, since none of the *Sh* inserts has the internal EcoRI sites expected for tandem ligations in a library prepared using EcoRI linkers. Also, different *Sh* cDNAs are isolated with greater frequency than expected for tandem ligations, typically rare events in library construction. P1, H11, and H21 cDNAs might result from incomplete RNA processing; each contains a small insertion that could be spliced out in the mature mRNA (Figure 1, Figure 3, Table 1). However, incomplete processing is not likely to account for the main features of *Sh* cDNA variation. The best example is the five 5' end class I clones; none contains any sequences from the 85 kb genomic region within which the 5' sequences of the other clones map. If class II and III 5' sequences derive from introns or intron fragments, it is surprising that none of the sequences is found in the class I clones. Also, class ii clones (H9, H15, H29) have an identical structure at both the 5' and 3' end; therefore, if they represent immature mRNAs, the splicing process yields intermediates that consistently possess matched 5' and 3' ends. Finally, the possibility that these clones correspond to splicing intermediates carries the odd implication that splicing proceeds via a pathway that uses nested splice sites to pare down a large intron before the final splice is made.

How Many Products Does *Sh* Encode?

The degree of diversity among *Sh* cDNA clones is astonishing: Among the 21 clones discussed in this report are at least 10 distinct types and no more than 3 examples of any one type. Thus, it is unlikely that this is a complete collection of *Sh* cDNAs. Two variables apparently dictate the number of potential products: a) the total number of 5' and 3' end classes; and b) the combinatorial assortment of these ends.

At least 90 kb of the *Sh* gene is used to encode cDNA 5' ends. Three 5' end classes (I, II, III) have been characterized from this region; however, this is likely to be

an incomplete list because 4 unique ends that may represent additional classes have already been found. Thus, we estimate that the number of 5' end classes is at least 7. We find that *Sh* cDNA 3' ends are encoded over a chromosomal region of only about 4 kb. However, taking data presented by Papazian et al. (1987) into account, 10 kb may be a more accurate estimate (see Experimental Procedures). Four 3' end classes (i, ii, iii, iv) and 2 unique ends are encoded in this region. The 2 unique ends (P1, H11) may not represent genuine classes; both may result from aberrant splicing (Figure 3, Table 1). Thus, the total number of 3' end classes may be limited to 4.

Sh can produce 28 different protein products if seven 5' and four 3' end classes assort independently. We have not examined a large enough sample of cDNAs to determine if there are preferred combinations of 3' and 5' ends. However, this may be the case; for example, all 3 of the class ii cDNAs analyzed thus far contain class I 5' ends (although class I ends have been found to associate with other 3' ends).

Features of Potassium Channel Diversity

Functional descriptions of K^+ channels provide a framework within which to consider *Sh* gene product diversity. Voltage-dependent K^+ channels are diverse, with a dozen or more classes distinguished primarily by different activation, inactivation, and kinetic parameters (Rudy, 1988). Furthermore, many K^+ channels are modulated by specific cyclic nucleotide-dependent second-messenger processes. Sophisticated mechanisms may be required to regulate expression of different K^+ channel types, since neighboring neurons can express distinctly different populations (Dekin and Getting, 1987); moreover, intracellular targeting might be complex, since any given neuron can have markedly different channel populations distributed among its dendritic, somatic, axonal, and synaptic membranes (Hille, 1984; Llinas, 1984). Despite the heterogeneity of K^+ channels in distribution and intrinsic channel features, the K^+ selectivity filter appears identical among all classes. Thus, one might expect that homologous structures are

responsible for the similar ion selectivity and conductivity seen across K^+ channel classes; relatively more diverse structures might be needed to generate the variety of activation, inactivation, kinetic, modulatory, differential expression, and intracellular targeting features.

The major limitation in our analysis of *Sh* gene product diversity is an incomplete understanding of K^+ channel structure and biochemistry. For example, we have no biochemical indication of the K^+ channel subunit composition (however, see below). Also, the membrane orientation of the K^+ channel is not known, although Tempel et al. (1987) have proposed that the *Sh* gene product amino terminus is intracellular, based on the apparent lack of a signal sequence. This placement dictates that the carboxyl terminus is extracellular for a *Sh* product with 3 or 7 membrane-spanning segments. Finally, we would like to know which segments constitute the channel pore, since we anticipate that many of the most interesting sequences, those affecting conductance, selectivity, and gating, should be close to this location. Although the precise location is not clear, pore sequences must span the membrane. Consequently, the constant region and carboxyl termini, with their hydrophobic domains and amphipathic S4 sequences, are good candidates for pore-forming sequences. An interesting possibility is that sequences in and around the hydrophobic segments of the constant region are involved in conserved K^+ channel functions, such as ion selectivity, while carboxyl termini sequences are involved in more variable channel functions, such as kinetics. Amino termini sequences are remote from potential membrane-spanning domains, but may also affect channel properties depending on protein folding patterns.

How Many Potassium Channel Classes are Encoded by *Sh*?

Previous voltage-clamp analyses suggested that *Sh* mutations affect only a single K^+ channel class, an A-type channel present in Drosophila muscle. Several other currents

are not affected: a Ca^{+2} current, a delayed rectifier K^+ current, a Ca^{+2}-dependent K^+ current, and a transient neuronal K^+ current (Salkoff and Wyman, 1981; Salkoff, 1983; Wu and Haugland, 1985; Elkins et al., 1986; Solc et al., 1987). Recently, 3 types of voltage-dependent, inactivating K^+ current have been seen in voltage clamp analysis of Xenopus oocytes injected with *Sh* mRNA (Iverson et al., 1988; Timpe et al., 1988). One type is similar to the A current described in Drosophila muscle and myocytes and shown to be altered by *Sh* mutations. The other 2 distinct types have not been described previously in Drosophila; they inactivate slowly and have long recovery times following inactivation. Iverson et al. (1988) have suggested that differences among these 3 currents may be defined by particular classes of carboxyl and amino termini. Thus, it appears that *Sh* is capable of generating at least 3 types of K^+ current. Depending on the subunit composition of the channel (see below), the number could be much larger.

Is the A Channel Composed of Multiple Subunits?

Several lines of argument imply that *Sh* encodes channels composed of multiple subunits. For example, A channels are as complex as Na^+ channels in terms of inactivation, voltage sensitivity, and ion selectivity. However, the largest potential *Sh* protein (656 a.a., 6 hydrophobic domains) is considerably smaller than the eel Na^+ channel (1820 a.a., 20 hydrophobic domains; Noda et al., 1984). A dihydropyridine-binding protein, a putative Ca^{+2} channel, is also large (1873 a.a., 20 hydrophobic domains; Noda et al., 1987). The A channel may more closely resemble multimeric channels such as the ligand-gated nicotinic acetylcholine receptor from Torpedo (five subunits 437 to 501 a.a. with 5 hydrophobic domains each; Numa, 1987). Channels with subunits smaller than the AChR have been reported; for example, liver gap junction proteins are hexamers composed of relatively small subunits (226 a.a. and 283 a.a) that have at most four transmembrane domains apiece (Paul, 1986; B. Nicholson, personal communication). Class iii *Sh* polypeptides are nearly as small as

these gap junction subunits and have even fewer hydrophobic domains.

Analysis of *Sh* mutations also suggests a multimeric A channel (Haugland and Wu, 1988; Timpe and Jan, 1987). For example, subunit interactions are indicated from observations that Sh^{KS133} and Sh^{102} act antagonistically when placed in combination with a normal Sh^{+} gene or with other *Sh* alleles. The interactions are allele-specific, providing additional evidence in favor of protein-protein contacts. Experiments on the Sh^{5} mutation suggest further that the A channel may exist normally as a heteromultimer, since one component, the one affected by Sh^{5}, appears to be present as a single copy. *Sh* gene product diversity presented in this report is consistent with a heteromultimeric model of the A channel, since the various cDNAs have the potential of generating radically different subunits. However, a heteromultimeric channel is apparently not strictly required, since functional K^{+} channels are generated in Xenopus oocytes following injection with a single *Sh* mRNA species (Iverson et al., 1988; Timpe et al., 1988).

A number of issues remain to be resolved in the analysis of *Sh* diversity. For example, we would like to know if all *Sh* polypeptides are capable of forming homomultimeric (or monomeric) channels. In addition, it is possible that channels composed of different combinations of *Sh* subunits may display a spectrum of physiological properties. We would particularly like to know the role of the small, class iii-type proteins: are they capable of forming channels; do they function as subunits; or do they perform regulatory functions? Since *Sh* mRNA can be injected into Xenopus oocytes singly and in mixtures, this expression system provides a powerful means for approaching these questions of K^{+} channel structure, function, and variation.

Experimental Procedures

Standard Techniques

Standard methods of screening recombinant DNA libraries, preparation of plasmid DNA, agarose and polyacrylamide electrophoresis, restriction enzyme analysis, Southern blot transfers, DNA fragment isolation, hybridization to DNA on filters, and subcloning of DNA fragments into plasmid vectors are described by protocols and combined references in Davis et al. (1980) and Maniatis et al. (1982). Radioisotope labelling by the random hexamer primer method is described in Feinberg and Vogelstein (1983).

Isolation of cDNA clones

Standard methods (Maniatis et al., 1982) were used to screen 500,000 phage from a 9-12 hour embryo cDNA library (gift from Dr. K. Zinn), and 500,000 phage from an adult head cDNA library (gift from Dr. P. Salvaterra). The probe was prepared from a *Sh* cDNA clone previously named F-c (Kamb et al., 1987) and called P1 in this report. We obtained one cDNA clone from the embryo library and 27 cDNAs from the head library. Clone inserts were sized on 1% agarose gels. We chose 17 of the adult head *Sh* clones for further analysis because, based on insert size, these clones apparently were generated independently during library construction. In addition, we analyzed the single embryo clone.

For naming clones in our collection, P is for pupal, H is for head, and E is for embryo, denoting the cDNA library of origin. Data on the *Sh* cDNAs of Tempel et al. (1987), ShA1, and Baumann, et al. (1987), gt141, are included for comparison. For these clones, the published nucleotide sequences were used to: 1) assign the variable 3' and 5' ends to the classes of Table 1; 2) generate the restriction maps of Figure 1; and 3) align nucleotide sequences for the schematic diagram of Figure 3. The pattern of

genomic hybridization for ShA1 and gt141 (Figure 2) was reconstructed from published data and matched to the resolution of the present analysis. Papazian et al. (1987) report hybridization of the ShA1 clone to sequences centered at about map position +24. Our analysis does not extend so far distally and this fragment is not represented in Figure 2.

A small pupal cDNA clone (0.6 kb) named F-d (Kamb et al., 1987) is likely to be a *Sh* cDNA clone since: 1) it hybridizes to genomic sequences between positions +68.5 and +75, and 2) it hybridizes to cDNAs H1, H4, H11, H13, H19, H20, H21, and H34 on clone blots. Restriction mapping and limited nucleotide sequence analysis did not allow us to assign the clone to classes or to align it with any of the other *Sh* cDNAs reported here.

Restriction Enzyme Mapping of cDNA Clones

Sh cDNAs were subcloned into the EcoRI site of the Bluescript vector (Stratagene) for restriction mapping. Plasmids containing cDNA inserts were linearized and labelled at the unique BamHI site by filling in with Klenow. The plasmid was then digested with EcoRI, and the two insert fragments purified by agarose gel electrophoresis. These fragments were subjected separately to partial restriction enzyme digestion, 5% polyacrylamide gel electrophoresis, and autoradiography.

DNA Sequence Analysis

DNA sequences of *Sh* cDNA clones were obtained by the dideoxynucleotide chain termination method using recombinant m13 virions (mp18 and mp19)(Sanger et al., 1977, 1980). Sequencing reactions were performed either with Klenow (Boehringer and Mannheim) or with Sequenase (U.S. Biochemicals). Overlapping sequences were obtained either by subcloning various restriction fragments or by using synthetic oligonucleotides prepared in the Caltech microchemical facility. Computer analysis of DNA sequences was carried out using either a Macintosh 512 and the DNA Inspector

software package or an IBM XT and software developed at Caltech by Dr. A Goldin.

Four cDNA clones were sequenced completely: 1) E1 (class I.iii); 2) H2 (class III.iii); 3) H4 (class II.iv); 4) H29 (class I.ii). In addition, the 3 previously published complete *Sh* cDNA sequences were compared: 1) P1 (class I. unique); 2) ShA1 (class II.i); 3) gt141 (class unique.iii) (Kamb et al., 1987; Tempel et al., 1987; Baumann et al., 1987). Partial nucleotide sequences were obtained for 7 cDNA clones: 1) H1 (class II.iv, 450 b.p. at 3' end); 2) H9 (class I.ii, 350 b.p. at 5' end); 3) H11 (class unique.unique, 1000 b.p. at 5' end); 4) H15 (class I.ii, 350 b.p. at 3' end); 5) H21 (class II.iv, 430 b.p. at 3' end); 6) H34 (class II.iv, 600 b.p. at 5' end, 2000 b.p. at 3' end); 7) H37 (class III.i, 400 b.p. at 5' end, 900 b.p. at 3' end). Among these sequences are 4 examples of class I clones; 3 examples of class II; 2 examples of class III; 2 examples of class i; 2 examples of class ii; 3 examples of class iii; and, 4 examples of class iv.

Assignment of *Sh* cDNAs to 5' and 3' Classes

Each *Sh* cDNA 5' and 3' end was defined as unique or assigned to a class: unique ends were represented only once in our collection of cDNAs; a class was defined when an end was observed in two or more clones. *Sh* cDNA classification was based on comparisons of genomic hybridization patterns, restriction maps, and DNA sequences. In some cases, class i and iv 3' ends were further checked by using specific oligonucleotides to probe cDNA clone blots. One pair of specific oligonucleotides hybridized to sequences 220 b.p. downstream of the 3' divergence site; a second pair hybridized 290 b.p. downstream of the divergence site. Differences in the sequence and/or length of untranslated regions are not considered in grouping cDNA clones into classes.

Acknowledgments

We thank M. Gautam, L. Iverson, M. Mathew, M. Ramaswami and B. Rudy for helpful discussions, and R. McMahan for technical assistance. This research was supported by the Pfeiffer Research Foundation, and by USPHS grant NS21327-01 to M.T. J.T.-C. was supported by NIH training grant NS07251. M.T. is a McKnight Foundation Scholar and a Sloan Foundation Fellow.

The costs of publication of this article were defrayed in part by payment of page charges. This article must be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734, solely to indicate this fact.

References

Baumann, A., Krah-Jentgens, I., Mueller, R., Mueller-Holtkamp, F., Seidel, R., Kecskemethy, N. Casal, J., Ferrus, A. and Pongs, O. (1987). Molecular organization of the maternal effect region of the *Shaker* complex of Drosophila: characterization of an I_A channel transcript with homology to vertebrate Na^+ channel. EMBO J. *6*, 3419-3429.

Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Ann. Rev. Biochem. *56*, 467-495.

Catterall, W.A. (1986). Molecular properties of voltage-sensitive sodium channels. Ann. Rev. Biochem. *55*, 953-985.

Davis, R.W., Botstein, D. and Roth, J.R. (1980). *Advanced Bacterial Genetics. A Manual for Genetic Engineering*. Cold Spring Harbor Laboratory: Cold Spring Harbor,

New York.

Dekin. M.S. and Getting, P.A. (1987). *In vitro* characterization of neurons in the ventral part of the nucleus tractus solitarius. II. Ionic basis for repetitive firing patterns. J. Neurophysiol. *58*, 215-229.

Dohlman, H.G., Caron, M.G. and Lefkowitz, R.J. (1987). A family of receptors coupled to guanine nucleotide regulatory proteins. Biochemistry *26*, 2657-2664.

Elkins, T., Ganetzky, B. and Wu, C.-F. (1986). A Drosophila mutation that eliminates a calcium-dependent potassium current. Proc. Natl. Acad. Sci. USA *83*, 8415-8419.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *132*, 6-13.

Guy, H.R. and Seetharamulu, P. (1986). Molecular model of the action potential sodium channel. Proc. Natl. Acad. Sci. USA *83*, 508-512.

Haugland, F.N. and Wu, C.-F. (1988) A voltage clamp analysis of gene-dosage effects of the *Shaker* locus on larval muscle potassium currents in Drosophila. J. Neurosci. (in press).

Hille, B. (1984). *Ionic Channels of Excitable Membranes*. Sinauer: Sunderland, Ma.

Iverson, L.E., Tanouye, M.A., Lester, H.A., Davidson, N. and Rudy, B. (1988). Potassium channels from *Shaker* RNA expressed in Xenopus oocytes (in press).

Jahnsen, H. and Llinas, R. (1984). Ionic basis for the electroresponsiveness and oscillatory properties of guinea-pig thalamic neurones i*n vitro*. J. Physiol. *349*, 227-247.

Kamb, A. (1988). Molecular biology of *Shaker*, a Drosophila gene that encodes multiple potassium channel components. Ph.D. thesis (California Institute of Technology).

Kamb, A., Iverson, L.E. and Tanouye, M.A. (1987). Molecular characterization of *Shaker*, a Drosophila gene that encodes a potassium channel. Cell *50*, 405-413.

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

Latorre, R. and Miller, C. (1983). Conduction and selectivity in potassium channels. J. Memb. Biol. *7*, 11-30.

Llinas, R. (1984). Comparative electrobiology of mammalian central neurons. In *Brain Slices*. ed. Dingledine, R. Plenum: New York.

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

Noda, M., Shimizu, S., Tanabe, T., Takai, T., Kayano, T., Ikeda, T., Takahashi, H., Nakayama, Y., Kanaoka, Y., Minamino, N., Kangawa, K., Matsuo, H., Raftery, M.A., Hirose, T., Inayama, S., Hayashida, H., Miyata, T. and Numa, S. (1984). Primary structure of the Electrophorus electricus sodium channel deduced from cDNA sequence. Nature *312*, 121-127.

Noda, M., Tanabe, T., Takeshima, H., Mikami, A., Flockerzi, V., Takahashi, H., Kangawa, K., Konjima, M., Matsuo, M., Hirose, T. and Numa, S. (1987). Primary structure of the receptor for calcium channel blockers from skeletal muscle. Nature *328*, 313-318.

Numa, S. (1987). Molecular basis for the function of ionic channels. Biochem. Soc. Symp. *52*, 119-143.

Papazian, D.M., Schwarz, T.L., Tempel, B.L., Jan, Y.N. and Jan L.Y. (1987). Cloning of genomic and complementary DNA from *Shaker*, a putative potassium channel gene. Science *237*, 749-753.

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

Rudy, B. (1988). Diversity and ubiquity of K^+ channels. Neuroscience (in press).

Salkoff, L. (1983). Genetic and voltage clamp analysis of a Drosophila potassium channel. Cold Spring Harbor Symp. Biol.*48*, 221-231.

Salkoff, L. and Wyman, R.J. (1981). Genetic modification of potassium channels in Drosophila *Shaker* mutants. Nature *293*, 228-230.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463-5467.

Sanger, F., Coulson, A.R., Barrell, G.G., Smith, A.J.F. and Roe, B. (1980). Cloning

in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. *143*, 161-164.

Schwarz, T.L., Tempel, B.L., Papazian, D.M., Jan, Y.N. and Jan, L.Y. (1988). Multiple potassium-channel components are produced by alternative splicing at the *Shaker* locus in Drosophila. Nature (in press).

Solc, C.K., Zagotta, W.N. and Aldrich, R.W. (1987). Single-channel and genetic analyses reveal two distinct A-type potassium channels in Drosophila. Science *236*, 1094-1098.

Tempel, B.L., Papazian, D.M., Schwarz, T.L., Jan Y.N. and Jan, L.Y. (1987). Sequence of a probable potassium channel component encoded at the *Shaker* locus of Drosophila. Science *237*, 770-775.

Timpe, L.C. and Jan, L.Y. (1987). Gene dosage and complementation analysis of the *Shaker* locus in Drosophila. J. Neurosci. *7*, 1307-1317.

Timpe, L.C., Schwarz, T.L., Tempel, B.L., Papazian, D,M., Jan, Y.N. and Jan, L.Y. (1988). Expression of functional potassium channels from *Shaker* cDNA in Xenopus oocytes. Nature (in press).

Wu, C.-F. and Haugland, F. (1985). Voltage clamp analysis of membrane currents in larval muscle fibers of Drosophila: alteration of potassium currents in *Shaker* mutants. J. Neurosci. *5*, 2626-2640.

Figure Legends

Figure 1. Restriction Maps of *Sh* cDNA Clones, and a Schematic Diagram that Illustrates Their Pattern of Variation

Restriction maps for 18 cDNAs are aligned with respect to shared sites; in particular, a unique SalI site labelled with a star. Maps of all the clones are essentially identical in a central constant region denoted by a thick horizontal line. The constant region extends 863 b.p. and includes a conserved 5' AvaII site and a conserved 3' CfoI site. Variable 5' and 3' ends are depicted by thin horizontal lines; groupings of clones are according to 5' class.

Ava, Bam, Bgl, Cfo, Hae, Hin, Pst, Pvu, Rsa, Sal, and Xba are AvaII, BamHI, BglI, CfoI, HaeIII, HinfI, PstI, PvuII, RsaI, SalI, and XbaI sites, respectively. To simplify the figure, map locations of Sau3a sites and about 150 b.p. from the 5' end of H13 were omitted. Where possible, restriction sites are adjusted to conform to nucleotide sequences. The H6 map is tentative. Clones H31, H33, and H38 were not mapped. Several restriction sites 3' of the 3'-most PstI site are not mapped. H21 appears to have a small insertion (about 100 b.p.) just 5' to the conserved central region, shown here as a shift in several restriction sites to the left.

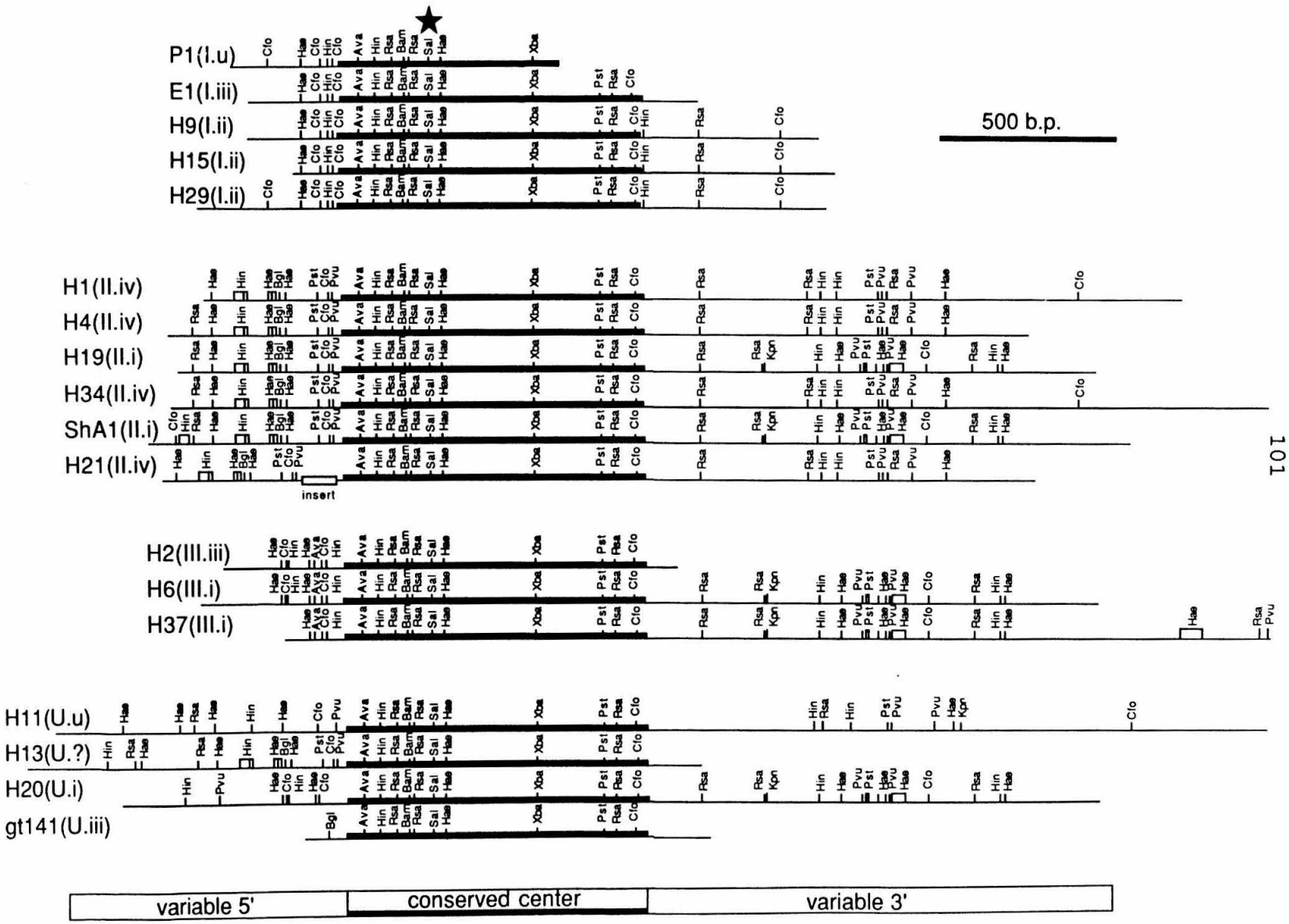

P1(I.u)
E1(I.iii)
H9(I.ii)
H15(I.ii)
H29(I.ii)
500 b.p.
H1(II.iv)
H4(II.iv)
H19(II.i)
H34(II.iv)
ShA1(II.i)
H21(II.iv)
insert
H2(III.iii)
H6(III.i)
H37(III.i)
H11(U.u)
H13(U.?)
H20(U.i)
gt141(U.iii)
variable 5'
conserved center
variable 3'

Figure 2. Genomic Organization of *Sh* cDNA Clones

These data show that *Sh* cDNAs map over a large genomic region that covers at least 108 kb. They also show that the pattern of genomic organization varies markedly among the various clones. The cDNAs were mapped using gel-purified inserts to probe DNA blots of recombinant phage, plasmid, and cosmid genomic DNA clones covering 143 kb of the original *Sh* chromosomal walk, from +27 to +170 (Kamb et al., 1987). Map coordinates and locations of *Sh* translocation breakpoints (B55, Sh^{LC}, and W32) are taken from this earlier study. The positions of restriction fragments that hybridize to each cDNA are shown as solid bars below the coordinate map. Since the precise boundaries of exons have not been determined, the bars represent genomic regions within which exons reside.

Clones are grouped according to their 5' end class. The most proximal genomic hybridization is characteristic of the 5' ends of cDNAs P1, E1, H9, H15, and H29 (class I clones). These clones hybridize to a 3.2 kb EcoRI restriction fragment centered at map position +139. The most distal genomic hybridization is characteristic of the 3' ends of several clones (H1,H4, H9, H15, H20, H21, 29, H31, H33, H34, H37). These clones hybridize to a 3.5 kb EcoR1/HindIII restriction fragment centered at +35.

Clones from all 5' end classes hybridize to three contiguous restriction fragments between +37 and +52. A conserved SalI restriction site at position +47.5 is marked. Proximal to the conserved fragments, the patterns of genomic hybridization vary among the classes. E1, representative of other class I clones, hybridizes to a restriction fragment that lies between +137 and +141. H34, representative of class II cDNAs, hybridizes to four sets of restriction fragments between: 1) +64 and +80; 2) +89 and +94; 3) +99 and +101; and, 4) +110 and +113. H37, representative of class III clones, hybridizes to two sets of restriction fragments between: 1) +54 and +59; and, 2) +85 and +89. The unique 5' ends of H13 and H20 resemble the genomic organization

of class II and III clones, respectively.

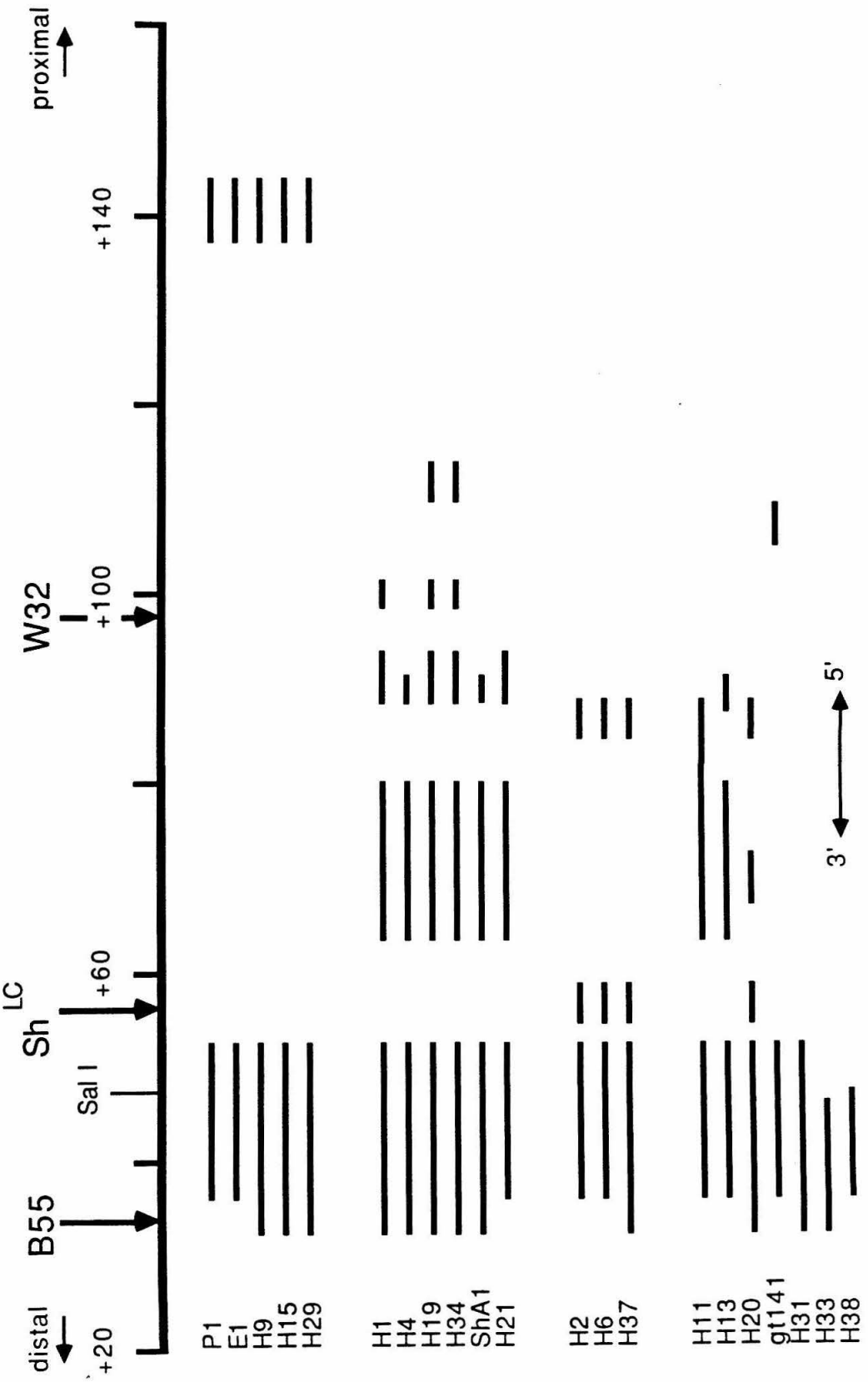

distal
proximal
+20
B55
Sal I
Sh LC
+60
+100
W32
+140
P1
E1
H9
H15
H29
H1
H4
H19
H34
ShA1
H21
H2
H6
H37
H11
H13
H20
gt141
H31
H33
H38
5'
3'

Figure 3. DNA Sequence Comparisons of *Sh* cDNA Clones

A schematic diagram of the nucleotide sequences of 7 *Sh* cDNA clones P1, E1, H2, H4, H29, H37, ShA1, and gt141 is shown with the sequences aligned with respect to a conserved SalI site (star). Dotted lines delineate the approximate borders of the constant segment. Similar shadings in the 5' and 3' end segments reflect similar DNA sequences. Translational start and stop signals are shown as arrows. The approximate locations of the S4 amphipathic domain and the six hydrophobic domains in the conceptual translation products are depicted with horizontal bars.

Several cDNA clones may be polyadenylated: class iii clones E1 and H2 have 18 and 7 adenosine residues at their 3' ends, respectively. A potential polyadenylation signal (GATAAA), similar to the consensus signal (AATAAA), lies 27 bases upstream of the A-tracts. These cDNAs are not likely to be derived from incompletely spliced mRNA, because there is not enough room in their 3' end segment to encode 3 additional hydrophobic domains and an S4 domain. All 3 clones have a relatively short sequence (181 b.p.) separating the central core region from the presumptive site of polyadenylation. Moreover, class iii clones have been isolated 3 times from 3 separate cDNA libraries.

Class iv clones H4 and H21 have 7 and 10 adenosine residues at their 3' ends, respectively. The same putative polyadenylation signal (GATAAA) lies 13 bases upstream. Interestingly, the 3' end of the class iv clone H34 is longer than H4 and H21 by 750 b.p.; it may represent differential utilization of polyadenylation signals. H34 has only 3 adenosine residues at the site corresponding to the H4 and H21 A-tracts. The alternative, that H4 and H21 derive from internal oligo(dT) priming during cDNA synthesis, has not been ruled out.

For reference to the previously published ShA1 cDNA sequence: the common SalI restriction site is at nucleotide +433; the site of 5' divergence is after +181; the site

of 3' divergence is after +1044 (Tempel et al., 1987). For reference to the previously published gt141 cDNA sequence: the SalI site is at position 492; the site of 5' divergence is after 240; the site of 3' divergence is after 1103 (Baumann et al., 1987). For reference to the previously published P1 cDNA sequence: the SalI site is at nucleotide 530; the site of 5' divergence is after 273; the site of 3' divergence occurs within the constant region after nucleotide 867 and is due to the presence of a 71 b.p. sequence absent from all other cDNAs (see Table 1)(Kamb et al., 1987). The complete nucleotide sequences for the clones reported here are available from: Kamb, 1988; the Genbank database; and the authors upon request.

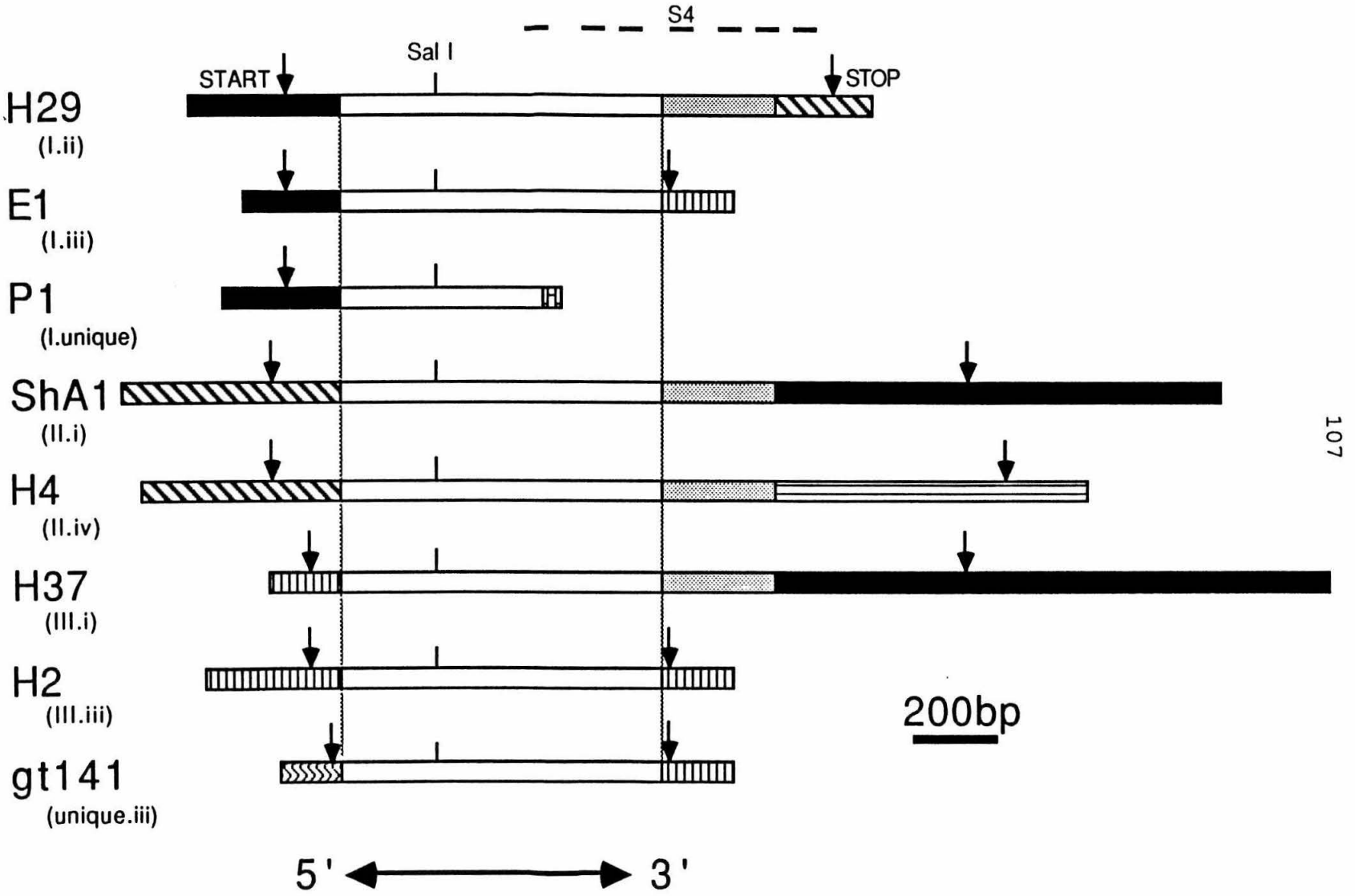

S4
Sal I
START
STOP
H29
(I.ii)
E1
(I.iii)
P1
(I.unique)
ShA1
(II.i)
H4
(II.iv)
H37
(III.i)
H2
(III.iii)
gt141
(unique.iii)
200bp
5'
3'

Figure 4. Deduced Amino Acid Sequences of Potential *Sh* Proteins

The polypeptides potentially encoded by cDNA clones representing each of the major 5' and 3' classes are depicted; H4 (class II.iv); H37 (class III.i); H29 (class I.ii); H2 (class III.iii); and, E1 (class I.iii). Each of the sequences is apparently full-length as judged by the presence of translational start and stop codons in their nucleotide sequences. Amino acid identities relative to the H4 encoded protein are shown by dashes. Potential transmembrane sequences are boxed and labeled HD1-6 or S4. HD6 for the H29 protein is not particularly hydrophobic, and there is considerable uncertainty about whether or not it traverses the membrane (see Figure 5). Potential sites for asparagine-linked glycosylation are marked with diamonds. Potential sites for cAMP-dependent phosphorylation are marked with circles. The sequence YFIT, present in the *B*-adrenergic receptor, is underlined. Utilization of these sites depends partly on the membrane orientation of *Sh* proteins, which is unknown.

The H2 polypeptide has an amino acid difference at position 131 derived from the nucleotide sequence that is either the result of intrastrain variation or the result of cDNA cloning. The P1 sequence reported previously contained sequencing errors in the region upstream of the SalI site that resulted in reading frame shifts (Kamb et al., 1987). We report here the corrected amino acid sequence in this region (see E1).

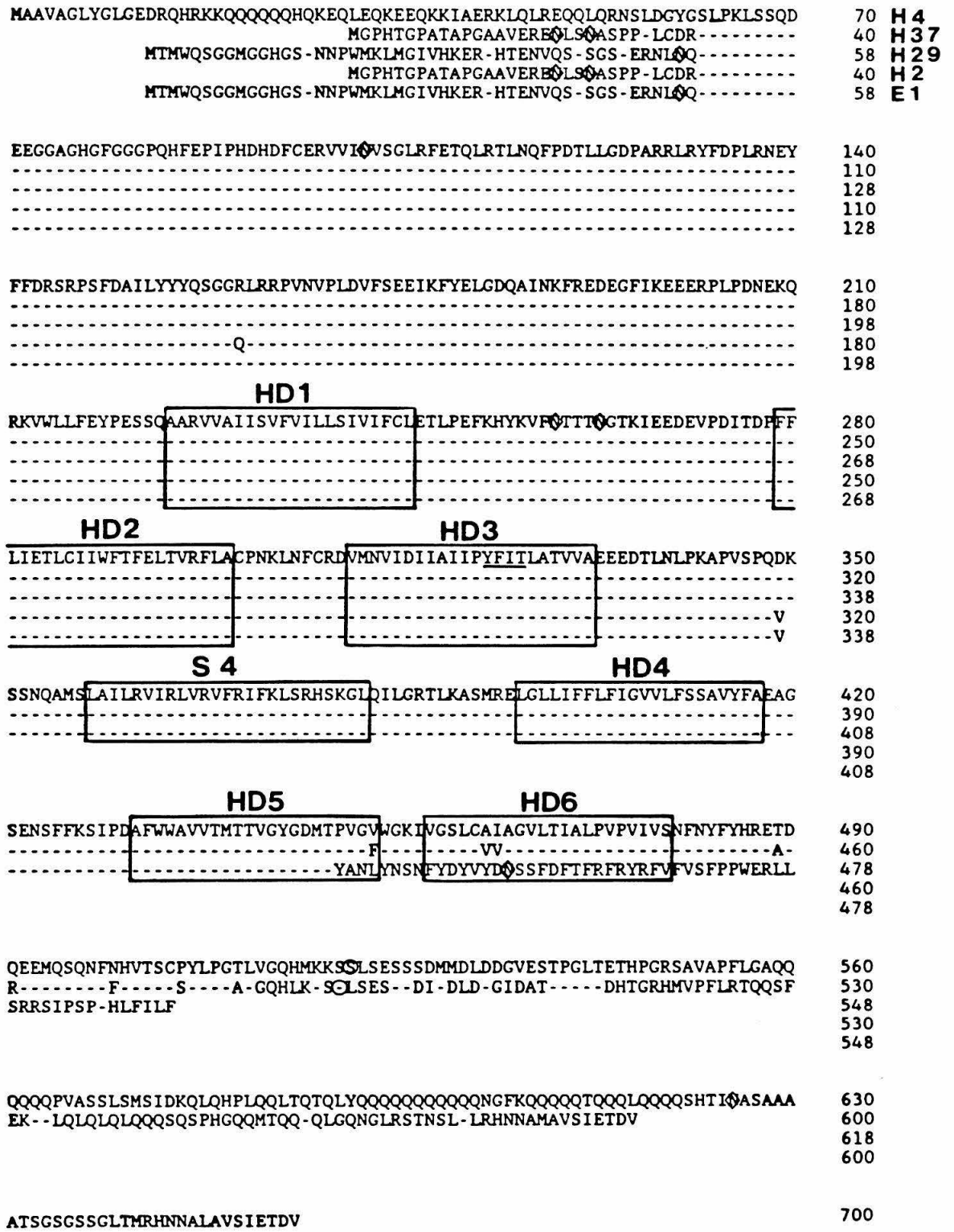

HD1
HD2
HD3
S 4
HD4
HD5
HD6

Figure 5. Hydrophobicity Plots of Potential *Sh* Proteins

The hydrophobicity plots of amino acid sequences derived from H29, E1, H4, H37, and H2 were calculated according to the algorithm of Kyte and Doolittle (1982). The size of the averaging window was 11 amino acids. Dotted vertical lines show the approximate extent of the constant segment. The location of the S4 amphipathic segment is shown in H4 with an arrow. The sixth hydrophobic domain of H29 is tentative and contains several charged residues.

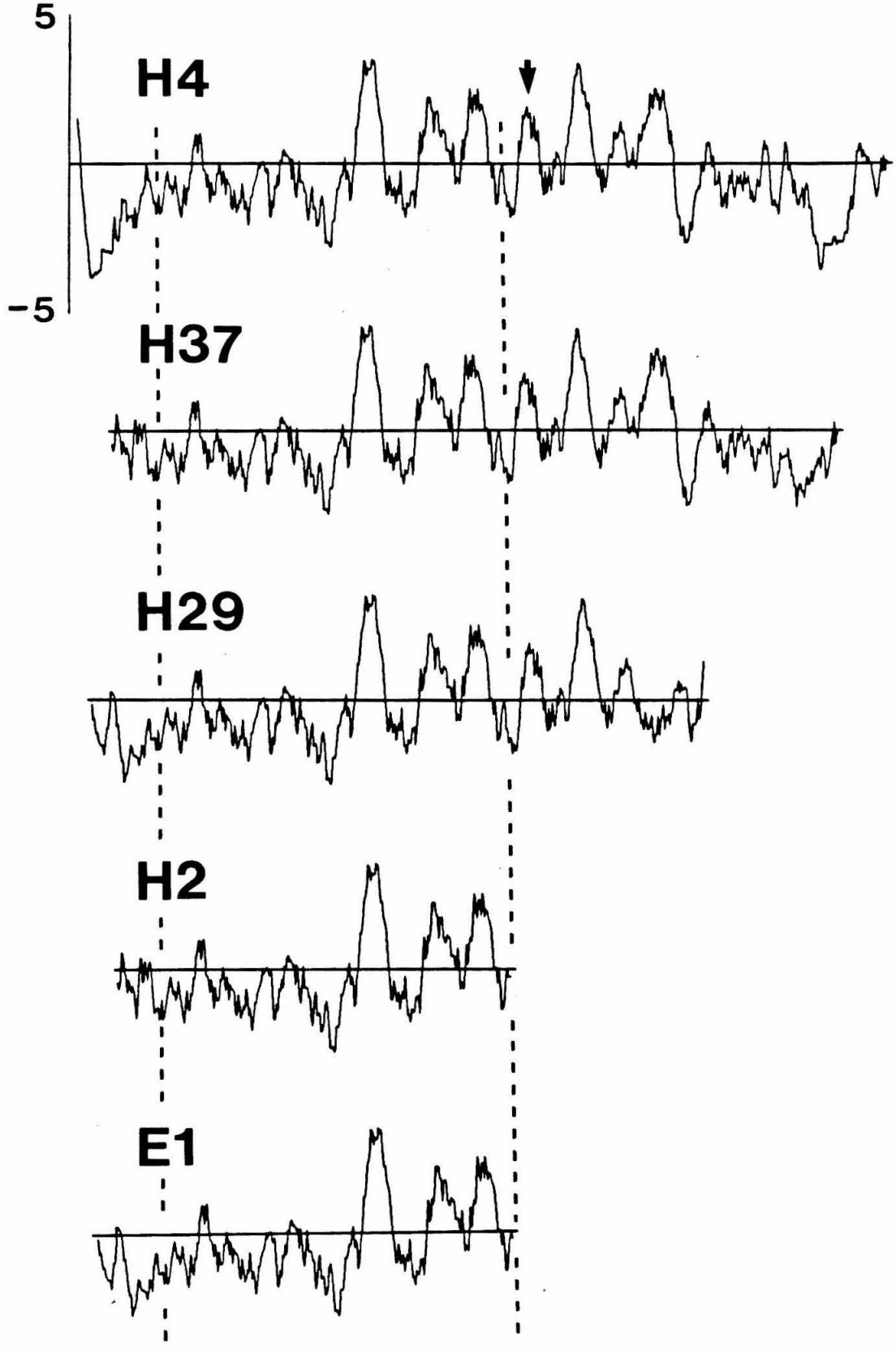
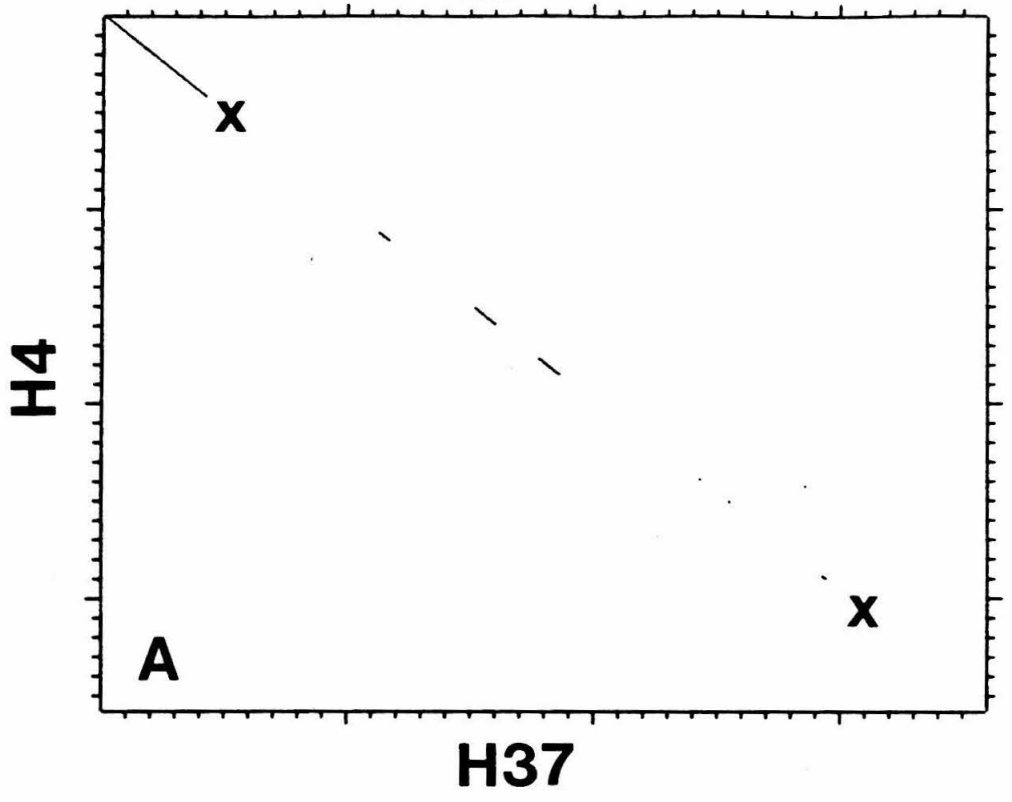

5
-5
H4
H37
H29
H2
E1

Figure 6. Dot Matrix Comparisons for a Portion of H4 and H37 Nucleotide Sequences

A 358 b.p. sequence near the 3' ends of H4 (3' class iv, Y-axis) and H37 (3' class i, X-axis) is compared using a dot matrix algorithm. A. Under stringent conditions (complete identity), the two sequences diverge at a position corresponding to amino acid 448 in the H4 protein (in the fifth hydrophobic domain). B. Under more relaxed conditions (80% identity), the sequences do not diverge until they reach a position corresponding to amino acid 534 in the H4 protein (beyond the sixth hydrophobic domain). Thus, nucleotide sequences in a 258 b.p. region (delimited with a small x) are similar, but not identical. The H4 sequence is 354 b.p. and extends from nucleotide 1578 to 1936; the H37 sequence is 358 b.p. and extends from nucleotide 1217 to 1571.


H4
x
x
A
H37

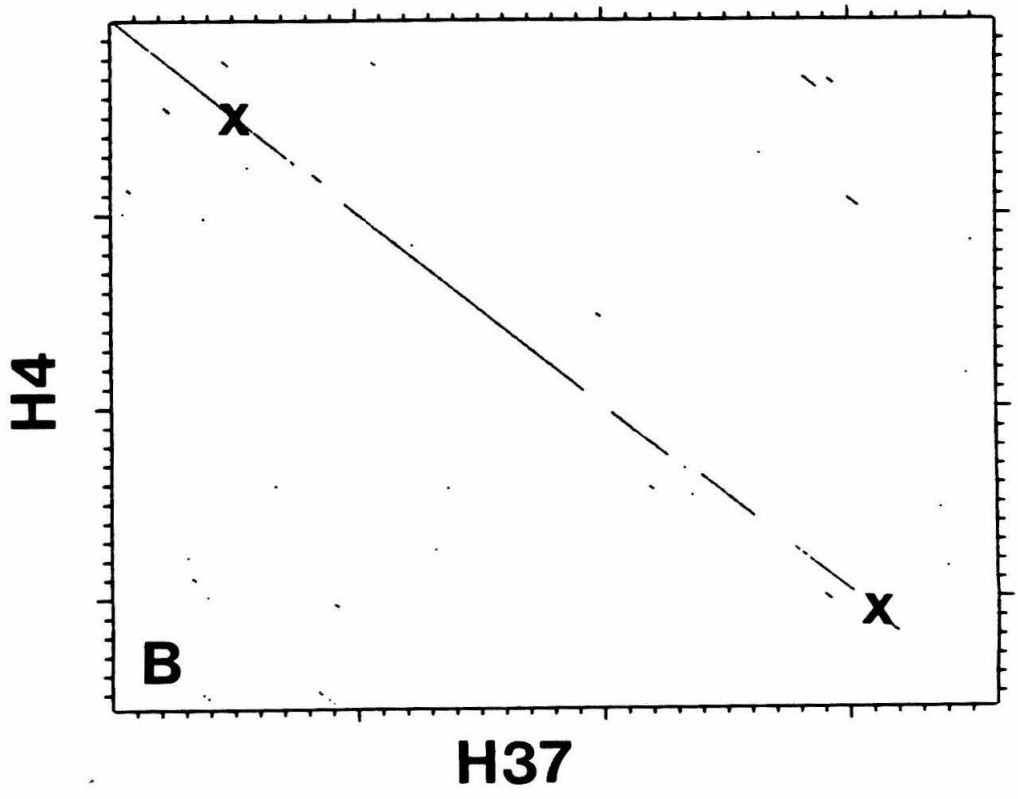

H4
x
x
B
H37

Figure 7. Sequence alignment for a Portion of H4 and H37

Depicted are the nucleotide and predicted amino acid sequences for a portion of H4 (258 b.p.) and H37 (255 b.p.) located near their respective 3' ends. Identities are indicated by dashes in the H37 sequence. The nucleotide sequences are 83% identical (213/258) in this region (A). However, 28 of 38 codons (74%) containing nucleotide changes are unaltered in coding sense, and the protein sequence is 88% identical (75/85); there are 10 a.a. substitutions (3 are conservative) and one a.a. deletion (B). The H4 sequence extends from a position 913 b.p. to 1171 b.p. downstream of the conserved SalI site; the H37 sequence extends from position 913 to 1168 b.p. downstream of SalI.

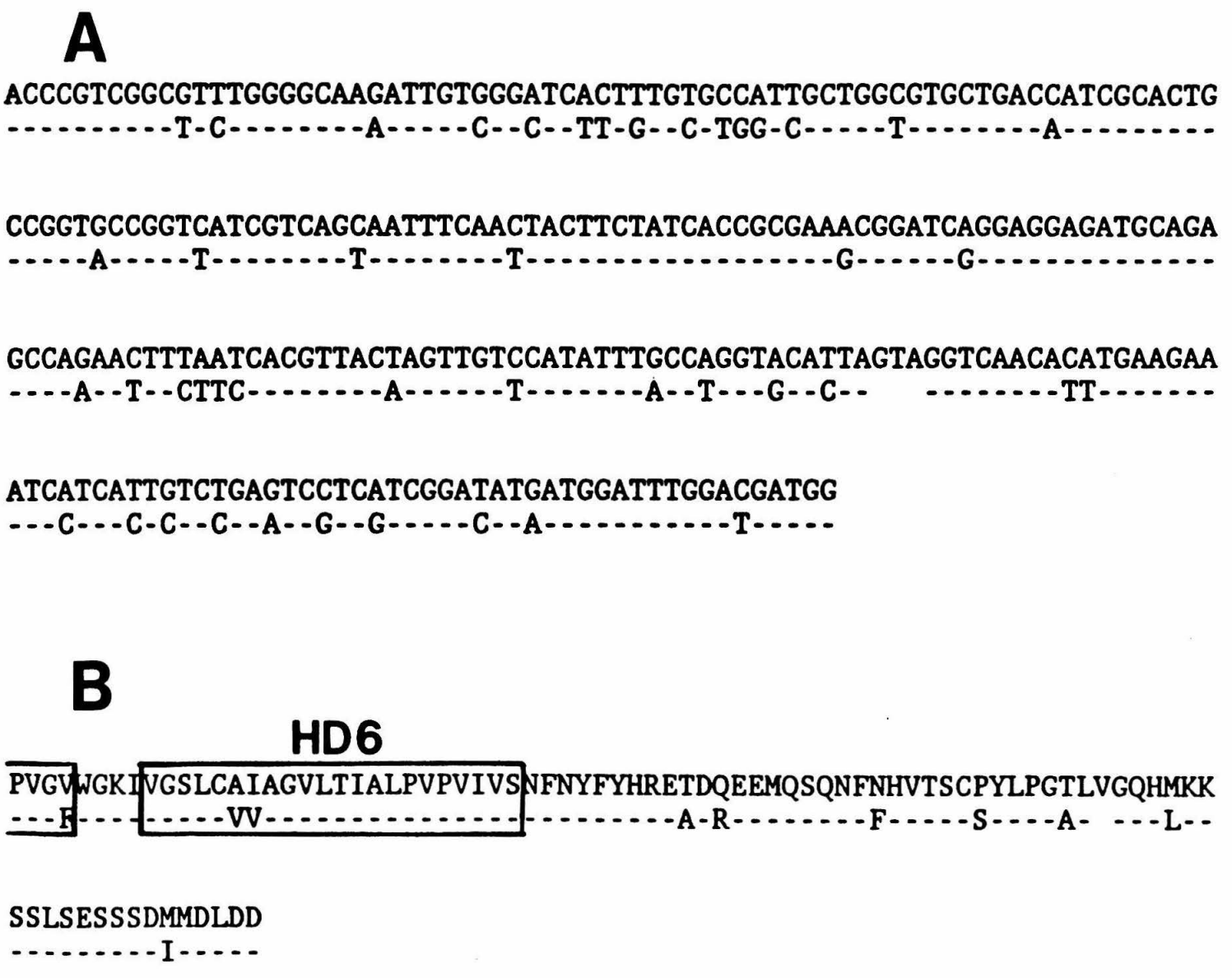

A
ACCCGTCGGCGTTTGGGGCAAGATTGTGGGATCACTTTGTGCCATTGCTGGCGTGCTGACCATCGCACTG
----------T-C--------A-----C--C--TT-G--C-TGG-C-----T--------A---------
CCGGTGCCGGTCATCGTCAGCAATTTCAACTACTTCTATCACCGCGAAACGGATCAGGAGGAGATGCAGA
-----A-----T--------T--------T------------------G------G--------------
GCCAGAACTTTAATCACGTTACTAGTTGTCCATATTTGCCAGGTACATTAGTAGGTCAACACATGAAGAA
----A--T--CTTC--------A------T-------A--T---G--C--   --------TT-------
ATCATCATTGTCTGAGTCCTCATCGGATATGATGGATTTGGACGATGG
---C---C-C--C--A--G--G-----C--A-----------T-----
B
HD6
PVGVWGKIVGSLCAIAGVLTIALPVPVIVSNFNYFYHRETDQEEMQSQNFNHVTSCPYLPGTLVGQHMKK
---F----------VV---------------------A-R--------F-----S----A- ---L--
SSLSESSSDMMDLDD
---------I-----

Table 1. Sizes and Class Designations of 21 *Sh* cDNAs

Assignment of classes is described in Experimental Procedures. Clones that do not extend far enough in the 5' or 3' direction to catalog are denoted with a question mark. Although cataloged as unique, the 5' ends of H4 and H13 are similar to class II; that of H20 to class III. P1, H11, and H21 may represent incompletely processed mRNAs: P1 contain a 71 b.p. sequence at its 3' end that may be normally removed by splicing; H11 appears to possess sequences specific to both class i and class iv ends based on hybridization to oligonucleotides; and H21 appears to have an insertion of roughly 100 b.p. near its 5' end.

Table 1.

Sh cDNA clone	length	5' end class	3' end class
P1	0.9 kb	I	U
E1	1.3 kb	I	iii
H1	2.8 kb	II	iv
H2	1.4 kb	III	iii
H4	2.5 kb	II	iv
H6	2.6 kb	III	i
H9	1.7 kb	I	ii
H11	3.5 kb	U	U
H13	2.6 kb	U	?
H15	1.6 kb	I	ii
H19	2.7 kb	II	i
H20	2.9 kb	U	i
H21	2.5 kb	II	iv
H29	1.8 kb	I	ii
H31	1.9 kb	?	iv
H33	1.9 kb	?	i
H34	3.2 kb	II	iv
H37	2.8 kb	III	i
H38	1.9 kb	?	i
ShA1	2.9 kb	II	i
gt141	1.2 kb	U	iii

Table 2. Features of the Constant Region and Different 5' and 3' End Classes

Depicted are consensus features of *Sh* potential proteins determined by comparing complete and partial sequences from 14 cDNA clones.

Table 2.

Constant	cDNA	Peptide	Hydrophobic	S4
	863 b.p.	287 a.a.	3	0
5' End Class				
I	146 b.p.	49 a.a.	0	0
II	182 b.p.	61 a.a.	0	0
III	92 b.p.	31 a.a.	0	0
3' End Class				
i	804 b.p.	268 a.a.	3	1
ii	471 b.p.	157 a.a.	3	1
iii	3 b.p.	1 a.a.	0	0
iv	924 b.p.	308 a.a.	3	1

Appendix

Nucleotide Sequences of Five Different *Shaker* cDNA Clones

E1

```
        10         20         30         40         50         60
ACAGTTTCTT CGATCGGAAC CGGATTTGGG AAACAGCCGC CAAGATGACC ATGTGGCAGA

        70         80         90        100        110        120
GTGGCGGCAT GGGAGGCCAT GGCTCCCAGA ACAATCCATG GATGAAGCTG ATGGGCATCG

       130        140        150        160        170        180
TCCACAAGGA GCGGCGCCAC ACGGAGAACG TTCAGAGTCA GTCCGGTTCC AACGAGCGCA

       190        200        210        220        230        240
ACCTGAACCA GTCTTTGCCC AAATTGAGCA GTCAAGACGA AGAAGGGGGG GCTGGTCATG

       250        260        270        280        290        300
GCTTTGGTGG CGGACCGCAA CACTTTGAAC CCATTCCTCA CGATCATGAT TTCTGCGAAA

       310        320        330        340        350        360
GAGTCGTTAT AAATGTAAGC GGATTAAGGT TTGAGACACA ACTACGTACG TTAAATCAAT

       370        380        390        400        410        420
TCCCGGACAC GCTGCTTGGG GATCCAGCTC GGAGATTACG GTACTTTGAC CCGCTTAGAA

       430        440        450        460        470        480
ATGAATATTT TTTTGACCGT AGTCGACCGA GCTTCGATGC GATTTTATAC TATTATCAGA

       490        500        510        520        530        540
GTGGTGGCCG ACTACGGAGA CCGGTCAATG TCCCTTTAGA CGTATTTAGT GAAGAAATAA

       550        560        570        580        590        600
AATTTTATGA ATTAGGTGAT CAAGCAATTA ATAAATTCAG AGAGGATGAA GGCTTTATTA

       610        620        630        640        650        660
AAGAGGAAGA AAGACCATTA CCGGATAATG AGAAACAGAG AAAAGTCTGG CTGCTCTTCG

       670        680        690        700        710        720
AGTATCCAGA AAGTTCGCAA GCCGCCAGAG TTGTAGCCAT AATTAGTGTA TTTGTTATAT

       730        740        750        760        770        780
TGCTATCAAT TGTTATATTT TGTCTAGAAA CATTACCCGA ATTTAAGCAT TACAAGGTGT

       790        800        810        820        830        840
TCAATACAAC AACAAATGGC ACAAAAATCG AGGAAGACGA GGTGCCTGAC ATCACAGATC
```

```
       850         860         870         880         890         900
CTTTCTTCCT TATAGAAACG TTATGTATTA TTTGGTTTAC ATTTGAACTA ACTGTCAGGT

       910         920         930         940         950         960
TCCTCGCATG TCCGAACAAA TTAAATTTCT GCAGGGATGT CATGAATGTT ATCGACATAA

       970         980         990       1,000       1,010       1,020
TCGCCATCAT TCCGTACTTT ATAACACTAG CGACTGTCGT TGCCGAAGAG GAGGATACGT

     1,030       1,040       1,050       1,060       1,070       1,080
TAAATCTTCC AAAAGCGCCA GTCAGTCCAC AGGTATGAGA TTTCTGTTTG CCCAATGCCA

     1,090       1,100       1,110       1,120       1,130       1,140
ACTACACCAA AAATATCAAT ATTAATCACC CACACACGAC CACACACACA CACATGCATA

     1,150       1,160       1,170       1,180       1,190       1,200
TACAGTTGAA CCGATTAAAG TAATAAATGT TGCCATCATC ATTTATGAGT TGCCTGACAT

     1,210       1,220       1,230       1,240       1,250       1,260
GATAAAGAAG GGGTTGCCTG CCCGCCCCGA AACCAAAAAA AAAAAAAAAA AA
```

H2

```
        10         20         30         40         50         60
GAAAGCCAAG GAGCTCGCCA AATAACAGAA GGTGCTCCGG ACTAGAAGCA TCAGAAGCAG

        70         80         90        100        110        120
CAGCACCAGC AGCAGCACCC CAACCACCGC CAGCATCGCA AGTGGATTAC CGTAAAGTGG

       130        140        150        160        170        180
AGCAGCCGCA GGACGAGGTC GAGGGCGCGG CAGGATACGC CATCCGAGCA GGCTGAGCAT

       190        200        210        220        230        240
TGGCGCAATT AGCGATTCGC ACCACTTGCT CGTCCTCAAC CATTTCCCGC ATCGATTTCA

       250        260        270        280        290        300
TCGAGCGATG GGGCCTCACA CCGGTCCTGC GACCGCGCCA GGAGCAGCCG TTGAGCGTGA

       310        320        330        340        350        360
AAACCTATCC AACGCCTCGC CCCCTCTTCT CTGCGATAGG TCTTTGCCCA AATTGAGCAG

       370        380        390        400        410        420
TCAAGACGAA GAAGGGGGGG CTGGTCATGG CTTTGGTGGC GGACCGCAAC ACTTTGAACC

       430        440        450        460        470        480
CATTCCTCAC GATCATGATT TCTGCGAAAG AGTCGTTATA AATGTAAGCG GATTAAGGTT

       490        500        510        520        530        540
TGAGACACAA CTACGTACGT TAAATCAATT CCCGGACACG CTGCTTGGGG ATCCAGCTCG

       550        560        570        580        590        600
GAGATTACGG TACTTTGACC CGCTTAGAAA TGAATATTTT TTTGACCGTA GTCGACCGAG

       610        620        630        640        650        660
CTTCGATGCG ATTTTATACT ATTATCAGAG TGGTGGCCAA CTACGGAGAC CGGTCAATGT

       670        680        690        700        710        720
CCCTTTAGAC GTATTTAGTG AAGAAATAAA ATTTTATGAA TTAGGTGATC AAGCAATTAA

       730        740        750        760        770        780
TAAATTCAGA GAGGATGAAG GCTTTATTAA AGAGGAAGAA AGACCATTAC CGGATAATGA

       790        800        810        820        830        840
GAAACAGAGA AAAGTCTGGC TGCTCTTCGA GTATCCAGAA AGTTCGCAAG CCGCCAGAGT
```

```
       850        860        870        880        890        900
TGTAGCCATA ATTAGTGTAT TTGTTATATT GCTATCAATT GTTATATTTT GTCTAGAAAC

       910        920        930        940        950        960
ATTACCCGAA TTTAAGCATT ACAAGGTGTT CAATACAACA ACAAATGGCA CAAAAATCGA

       970        980        990      1,000      1,010      1,020
GGAAGACGAG GTGCCTGACA TCACAGATCC TTTCTTCCTT ATAGAAACGT TATGTATTAT

     1,030      1,040      1,050      1,060      1,070      1,080
TTGGTTTACA TTTGAACTAA CTGTCAGGTT CCTCGCATGT CCGAACAAAT TAAATTTCTG

     1,090      1,100      1,110      1,120      1,130      1,140
CAGGGATGTC ATGAATGTTA TCGACATAAT CGCCATCATT CCGTACTTTA TAACACTAGC

     1,150      1,160      1,170      1,180      1,190      1,200
GACTGTCGTT GCCGAAGAGG AGGATACGTT AAATCTTCCA AAAGCGCCAG TCAGTCCACA

     1,210      1,220      1,230      1,240      1,250      1,260
GGTATGAGAT TTCTGTTTGC CCAATGCCAA CTACACCAAA AATATCAATA TTAATCACCC

     1,270      1,280      1,290      1,300      1,310      1,320
ACACACGACC ACACACACAC ACATGCATAT ACAGTTGAAC CGATTAAAGT AATAAATGTT

     1,330      1,340      1,350      1,360      1,370      1,380
GCCATCATCA TTTATGAGTT GCCTGACATG ATAAAGAAGG GGTTGCCTGC CCGCCCCGAA

     1,390
ACCAAAAAAA
```

H4

```
        10         20         30         40         50         60
GTCGTCAGTC AGTCAATCAA TCACTCTTGT AACCATGTAC CAAAGTTCTT TGCCGCGAAA

        70         80         90        100        110        120
ACTAAAATGA AAACGAAAGT GAAAATGAGC GAATGGCAGC CGCGGCCACA GCAATCGATC

       130        140        150        160        170        180
CATGACACAA CCAGTGACAA GCAGTCCCCC AGTGAAACCG CATCCGCATC CGAGTCCGAT

       190        200        210        220        230        240
ACCGATAAAG ATTCTGAATC GGAGTGAGTG CCGCGTCCGA GAGCGTTCCC TGTCCACGTC

       250        260        270        280        290        300
CACCATCGGC GGAGCAGGTG TGCCTGAGGC CCACCTGGTG GCATGGCCGC CGTTGCCGGC

       310        320        330        340        350        360
CTCTATGGCC TTGGGGAGGA TCGCCAGCAC CGCAAGAAGC AGCAGCAACA GCAGCAGCAC

       370        380        390        400        410        420
CAGAAGGAGC AGCTCGAGCA GAAGGAGGAG CAAAAGAAGA TCGCCGAGCG GAAGCTGCAG

       430        440        450        460        470        480
CTGCGGGAGC AGCAGCTCCA GCGCAACTCC CTCGATGGTT ACGGGTCTTT GCCCAAATTG

       490        500        510        520        530        540
AGCAGTCAAG ACGAAGAAGG GGGGGCTGGT CATGGCTTTG GTGGCGGACC GCAACACTTT

       550        560        570        580        590        600
GAACCCATTC CTCACGATCA TGATTTCTGC GAAAGAGTCG TTATAAATGT AAGCGGATTA

       610        620        630        640        650        660
AGGTTTGAGA CACAACTACG TACGTTAAAT CAATTCCCGG ACACGCTGCT TGGGGATCCA

       670        680        690        700        710        720
GCTCGGAGAT TACGGTACTT TGACCCGCTT AGAAATGAAT ATTTTTTTGA CCGTAGTCGA

       730        740        750        760        770        780
CCGAGCTTCG ATGCGATTTT ATACTATTAT CAGAGTGGTG GCCGACTACG GAGACCGGTC

       790        800        810        820        830        840
AATGTCCCTT TAGACGTATT TAGTGAAGAA ATAAAATTTT ATGAATTAGG TGATCAAGCA
```

```
       850        860        870        880        890        900
ATTAATAAAT TCAGAGAGGA TGAAGGCTTT ATTAAAGAGG AAGAAAGACC ATTACCGGAT

       910        920        930        940        950        960
AATGAGAAAC AGAGAAAAGT CTGGCTGCTC TTCGAGTATC CAGAAAGTTC GCAAGCCGCC

       970        980        990      1,000      1,010      1,020
AGAGTTGTAG CCATAATTAG TGTATTTGTT ATATTGCTAT CAATTGTTAT ATTTTGTCTA

     1,030      1,040      1,050      1,060      1,070      1,080
GAAACATTAC CCGAATTTAA GCATTACAAG GTGTTCAATA CAACAACAAA TGGCACAAAA

     1,090      1,100      1,110      1,120      1,130      1,140
ATCGAGGAAG ACGAGGTGCC TGACATCACA GATCCTTTCT TCCTTATAGA AACGTTATGC

     1,150      1,160      1,170      1,180      1,190      1,200
ATTATTTGGT TTACATTTGA ACTAACTGTC AGGTTCCTCG CATGTCCGAA CAAATTAAAT

     1,210      1,220      1,230      1,240      1,250      1,260
TTCTGCAGGG ATGTCATGAA TGTTATCGAC ATAATCGCCA TCATTCCGTA CTTTATAACA

     1,270      1,280      1,290      1,300      1,310      1,320
CTAGCGACTG TCGTTGCCGA AGAGGAGGAT ACGTTAAATC TTCCAAAAGC GCCAGTCAGT

     1,330      1,340      1,350      1,360      1,370      1,380
CCACAGGACA AGTCATCGAA TCAGGCTATG TCCTTGGCAA TATTACGAGT GATACGATTA

     1,390      1,400      1,410      1,420      1,430      1,440
GTTCGAGTAT TTCGAATATT TAAGTTATCT AGGCATTCGA AGGGTTTACA AATATTAGGA

     1,450      1,460      1,470      1,480      1,490      1,500
CGAACTCTGA AAGCCTCAAT GCGGGAATTA GGTTTACTTA TATTTTTCTT ATTTATAGGC

     1,510      1,520      1,530      1,540      1,550      1,560
GTCGTACTCT TCTCATCGGC GGTTTATTTT GCGGAAGCTG GAAGCGAAAA TTCCTTCTTC

     1,570      1,580      1,590      1,600      1,610      1,620
AAGTCCATAC CCGATGCATT TTGGTGGGCG GTCGTTACCA TGACCACCGT TGGATATGGT

     1,630      1,640      1,650      1,660      1,670      1,680
GACATGACAC CCGTCGGCGT TTGGGGCAAG ATTGTGGGAT CACTTTGTGC CATTGCTGGC
```

1,690 1,700 1,710 1,720 1,730 1,740
GTGCTGACCA TCGCACTGCC GGTGCCGGTC ATCGTCAGCA ATTTCAACTA CTTCTATCAC

1,750 1,760 1,770 1,780 1,790 1,800
CGCGAAACGG ATCAGGAGGA GATGCAGAGC CAGAACTTTA ATCACGTTAC TAGTTGTCCA

1,810 1,820 1,830 1,840 1,850 1,860
TATTTGCCAG GTACATTAGT AGGTCAACAC ATGAAGAAAT CATCATTGTC TGAGTCCTCA

1,870 1,880 1,890 1,900 1,910 1,920
TCGGATATGA TGGATTTGGA CGATGGTGTC GAGTCCACGC CGGGATTGAC AGAAACACAT

1,930 1,940 1,950 1,960 1,970 1,980
CCTGGACGCA GTGCGGTGGC TCCATTTTTG GGAGCCCAGC AGCAGCAGCA ACAACCGGTA

1,990 2,000 2,010 2,020 2,030 2,040
GCATCCTCAC TGTCGATGTC GATCGACAAA CAACTGCAGC ACCCACTGCA GCAGCTGACG

2,050 2,060 2,070 2,080 2,090 2,100
CAGACGCAAC TGTACCAACA GCAGCAACAG CAGCAGCAGC AGCAGCAAAA CGGCTTCAAG

2,110 2,120 2,130 2,140 2,150 2,160
CAGCAGCAGC AACAGACGCA GCAGCAGCTG CAACAGCAAC AGTCCCACAC AATAAACGCA

2,170 2,180 2,190 2,200 2,210 2,220
AGTGCAGCAG CGGCGACGAG CGGCAGCGGC AGTAGCGGTC TCACCATGAG GCACAATAAT

2,230 2,240 2,250 2,260 2,270 2,280
GCCCTGGCCG TTAGTATCGA GACCGACGTT TGACTACTGG TGCAAAAGAC GTTGCGTGGT

2,290 2,300 2,310 2,320 2,330 2,340
ATAAATTTGG CTTGACAGGA GTTACGTTGG ATGCCAGAAA CGACTACAAA AGCTGTTTAT

2,350 2,360 2,370 2,380 2,390 2,400
ATTTAATTTA AGTAGAACAA ATAACAAAAA CAAATTTAAT CTATTGCTAA ATTAAATTAA

2,410 2,420 2,430 2,440 2,450 2,460
AATCTAAATT AAAATCTAAA TTAATTTAAT TAAATTATAC ATTTAATGAT AAACAACACT

H29

```
        10         20         30         40         50         60
CAAACTGAAA AAAAAAAAAG AAAGATAAAA AACCTATTGA TGCGGCTGCC GAACCCAGAA

        70         80         90        100        110        120
ACAACTCAGT TTCAACTCTT GCCCTGTGGC ATCTGTCCTC TATATATATT TGTATAGTAT

       130        140        150        160        170        180
ATACATATAT ATCTGATCTG AAGTTCCAAG TCGCGAGTGC GTTTCCGTTT CCGTATTCGC

       190        200        210        220        230        240
GTCCATTTTC GTTTCGGTTT CGTTGGAAAG CTAGAGCGCT GCTGCCACTG CCACAGTTTC

       250        260        270        280        290        300
TTCGATCGGA ACCGGATTTG GGAAACAGCC GCCAAGATGA CCATGTGGCA GAGTGGCGGC

       310        320        330        340        350        360
ATGGGAGGCC ATGGCTCCCA GAACAATCCA TGGATGAAGC TGATGGGCAT CGTCCACAAG

       370        380        390        400        410        420
GAGCGGCGCC ACACGGAGAA CGTTCAGAGT CAGTCCGGTT CCAACGAGCG CAACCTGAAC

       430        440        450        460        470        480
CAGTCTTTGC CCAAATTGAG CAGTCAAGAC GAAGAAGGGG GGGCTGGTCA TGGCTTTGGT

       490        500        510        520        530        540
GGCGGACCGC AACACTTTGA ACCCATTCCT CACGATCATG ATTTCTGCGA AAGAGTCGTT

       550        560        570        580        590        600
ATAAATGTAA GCGGATTAAG GTTTGAGACA CAACTACGTA CGTTAAATCA ATTCCCGGAC

       610        620        630        640        650        660
ACGCTGCTTG GGGATCCAGC TCGGAGATTA CGGTACTTTG ACCCGCTTAG AAATGAATAT

       670        680        690        700        710        720
TTTTTTGACC GTAGTCGACC GAGCTTCGAT GCGATTTTAT ACTATTATCA GAGTGGTGGC

       730        740        750        760        770        780
CGACTACGGA GACCGGTCAA TGTCCCTTTA GACGTATTTA GTGAAGAAAT AAAATTTTAT

       790        800        810        820        830        840
GAATTAGGTG ATCAAGCAAT TAATAAATTC AGAGAGGATG AAGGCTTTAT TAAAGAGGAA
```

```
       850        860        870        880        890        900
GAAAGACCAT TACCGGATAA TGAGAAACAG AGAAAAGTCT GGCTGCTCTT CGAGTATCCA

       910        920        930        940        950        960
GAAAGTTCGC AAGCCGCCAG AGTTGTAGCC ATAATTAGTG TATTTGTTAT ATTGCTATCA

       970        980        990      1,000      1,010      1,020
ATTGTTATAT TTTGTCTAGA AACATTACCC GAATTTAAGC ATTACAAGGT GTTCAATACA

     1,030      1,040      1,050      1,060      1,070      1,080
ACAACAAATG GCACAAAAAT CGAGGAAGAC GAGGTGCCTG ACATCACAGA TCCTTTCTTC

     1,090      1,100      1,110      1,120      1,130      1,140
CTTATAGAAA CGTTATGTAT TATTTGGTTT ACATTTGAAC TAACTGTCAG GTTCCTCGCA

     1,150      1,160      1,170      1,180      1,190      1,200
TGTCCGAACA AATTAAATTT CTGCAGGGAT GTCATGAATG TTATCGACAT AATCGCCATC

     1,210      1,220      1,230      1,240      1,250      1,260
ATTCCGTACT TTATAACACT AGCGACTGTC GTTGCCGAAG AGGAGGATAC GTTAAATCTT

     1,270      1,280      1,290      1,300      1,310      1,320
CCAAAAGCGC CAGTCAGTCC ACAGGACAAG TCATCGAATC AGGCTATGTC CTTGGCAATA

     1,330      1,340      1,350      1,360      1,370      1,380
TTACGAGTGA TACGATTAGT TCGAGTATTT CGAATATTTA AGTTATCTAG GCATTCGAAG

     1,390      1,400      1,410      1,420      1,430      1,440
GGTTTACAAA TATTAGGACG AACTCTGAAA GCCTCAATGC GGGAATTAGG TTTACTTATA

     1,450      1,460      1,470      1,480      1,490      1,500
TTTTTCTTAT TTATAGGCGT CGTACTCTTC TCATCGGCGG TTTATTTTGC GGAAGCTGGA

     1,510      1,520      1,530      1,540      1,550      1,560
AGCGAAAATT CCTTCTTCAA GTCCATACCC GATGCATTTT GGTGGGCGGT CGTTACCATG

     1,570      1,580      1,590      1,600      1,610      1,620
ACCACCGTTG GATATGGTGA CATGACGTAT GCAAATCTAT ACAATTCTAA CTTTTATGAT

     1,630      1,640      1,650      1,660      1,670      1,680
TACGTTTACG ATAACTCCAG TTTTGATTTT ACGTTTCGTT TCCGTTATCG TTTCGTTTTC
```

```
     1,690       1,700       1,710       1,720       1,730       1,740
GTTTCGTTCC CGCCTTGGGA GCGCCTTCTT TCTAGGCGTT CCATTCCTTC TCCGTTCCAC

     1,750       1,760       1,770       1,780       1,790       1,800
CTCTTTATTC TATTCTAATT TTAATTTTCA ACTGTTTCGT TACTTGCGTT TTTATGAGTG

ATTA
```

H37

```
        10         20         30         40         50         60
TCATCGAGCG ATGGGGCCTC ACACCGGTCC TGCGACCGCG CCAGGAGCAG CCGTTGAGCG

        70         80         90        100        110        120
TGAAAACCTA TCCAACGCCT CGCCCCCTCT TCTCTGCGAT AGGTCTTTGC CCAAATTGAG

       130        140        150        160        170        180
CAGTCAAGAC GAAGAAGGGG GGGCTGGTCA TGGCTTTGGT GGCGGACCGC AACACTTTGA

       190        200        210        220        230        240
ACCCATTCCT CACGATCATG ATTTCTGCGA AAGAGTCGTT ATAAATGTAA GCGGATTAAG

       250        260        270        280        290        300
GTTTGAGACA CAACTACGTA CGTTAAATCA ATTCCCGGAC ACGCTGCTTG GGGATCCAGC

       310        320        330        340        350        360
TCGGAGATTA CGGTACTTTG ACCCGCTTAG AAATGAATAT TTTTTTGACC GTAGTCGACC

       370        380        390        400        410        420
GAGCTTCGAT GCGATTTTAT ACTATTATCA GAGTGGTGGC CGACTACGGA GACCGGTCAA

       430        440        450        460        470        480
TGTCCCTTTA GACGTATTTA GTGAAGAAAT AAAATTTTAT GAATTAGGTG ATCAAGCAAT

       490        500        510        520        530        540
TAATAAATTC AGAGAGGATG AAGGCTTTAT TAAAGAGGAA GAAAGACCAT TACCGGATAA

       550        560        570        580        590        600
TGAGAAACAG AGAAAAGTCT GGCTGCTCTT CGAGTATCCA GAAAGTTCGC AAGCCGCCAG

       610        620        630        640        650        660
AGTTGTAGCC ATAATTAGTG TATTTGTTAT ATTGCTATCA ATTGTTATAT TTTGTCTAGA

       670        680        690        700        710        720
AACATTACCC GAATTTAAGC ATTACAAGGT GTTCAATACA ACAACAAATG GCACAAAAAT

       730        740        750        760        770        780
CGAGGAAGAC GAGGTGCCTG ACATCACAGA TCCTTTCTTC CTTATAGAAA CGTTATGTAT

       790        800        810        820        830        840
TATTTGGTTT ACATTTGAAC TAACTGTCAG GTTCCTCGCA TGTCCGAACA AATTAAATTT
```

```
       850        860        870        880        890        900
CTGCAGGGAT GTCATGAATG TTATCGACAT AATCGCCATC ATTCCGTACT TTATAACACT

       910        920        930        940        950        960
AGCGACTGTC GTTGCCGAAG AGGAGGATAC GTTAAATCTT CCAAAAGCGC CAGTCAGTCC

       970        980        990      1,000      1,010      1,020
ACAGGACAAG TCATCGAATC AGGCTATGTC CTTGGCAATA TTACGAGTGA TACGATTAGT

     1,030      1,040      1,050      1,060      1,070      1,080
TCGAGTATTT CGAATATTTA AGTTATCTAG GCATTCGAAG GGTTTACAAA TATTAGGACG

     1,090      1,100      1,110      1,120      1,130      1,140
AACTCTGAAA GCCTCAATGC GGGAATTAGG TTTACTTATA TTTTTCTTAT TTATAGGCGT

     1,150      1,160      1,170      1,180      1,190      1,200
CGTACTCTTC TCATCGGCGG TTTATTTTGC GGAAGCTGGA AGCGAAAATT CCTTCTTCAA

     1,210      1,220      1,230      1,240      1,250      1,260
GTCCATACCC GATGCATTTT GGTGGGCGGT CGTTACCATG ACCACCGTTG GATATGGTGA

     1,270      1,280      1,290      1,300      1,310      1,320
CATGACGCCC GTCGGCTTCT GGGGCAAAAT TGTCGGCTCT TTGTGCGTGG TCGCTGGTGT

     1,330      1,340      1,350      1,360      1,370      1,380
GCTGACAATC GCACTGCCGG TACCGGTTAT CGTCAGTAAT TTCAATTACT TCTATCACCG

     1,390      1,400      1,410      1,420      1,430      1,440
CGAAGCGGAT CGGGAGGAGA TGCAGAGCCA AAATTTCTTC CACGTTACAA GTTGTTCATA

     1,450      1,460      1,470      1,480      1,490      1,500
TTTACCTGGT GCACTAGGTC AACATTTGAA GAAATCCTCA CTCTCCGAAT CGTCGTCGGA

     1,510      1,520      1,530      1,540      1,550      1,560
CATAATGGAT TTGGATGATG GCATTGATGC AACCACGCCA GGTCTGACTG ATCACACGGG

     1,570      1,580      1,590      1,600      1,610      1,620
CCGCCACATG GTGCCTTTTC TCAGGACACA GCAGTCATTC GAGAAGCAGC AGCTCCAGCT

     1,630      1,640      1,650      1,660      1,670      1,680
TCAGCTGCAG CTGCAGCAGC AGTCGCAGTC GCCGCACGGC CAACAGATGA CGCAGCAGCA
```

1,690 1,700 1,710 1,720 1,730 1,740
GCAGCTGGGC CAGAACGGCC TAAGGAGCAC AAATAGTTTA CAGTTAAGGC ATAATAACGC

1,750 1,760 1,770 1,780 1,790 1,800
GATGGCCGTC AGTATTGAGA CCGACGTCTG ACTACTAGTC AAACAAATGG AAAATGGACG

1,810 1,820 1,830 1,840 1,850 1,860
AAATTTGCGC AGTGAAATGC TACGTTGGAT GCCAGAAACG TCATCAAAAG CAGTCTAATT

1,870 1,880 1,890 1,900 1,910 1,920
TAGAATTTTA TTAATAAATA CAATTAAAAT ATAATTATAA TAATTAGTAA GCAACGTAGT

1,930 1,940 1,950 1,960 1,970 1,980
TGTAAATTAA ACAGCAAATG TACACAGACA CAACACACAC ACAGACACAG TGCCAGTTCA

1,990 2,000 2,010 2,020 2,030 2,040
CTCAGCTTGA ATTACACTAT TTGTAGACAC CAAAAAGAGT CAAATATGGA CTGGCCTTCT

2,050 2,060 2,070 2,080 2,090 2,100
ATAGGGATTT CCTTGTTTCT CCTTTCATTT TCCTTCTGGT AATCTACACA CCGAAAACAC

2,110 2,120 2,130 2,140 2,150 2,160
TTACACACAC ACGTCCACAC ACACTCAAAG TAAAAAACTC TACTTGATAC CTATGTTCAA

2,170 2,180 2,190 2,200 2,210 2,220
ATTTAGCAAT TAACAACTAA CAATCGTTAA CAACAACAAA ACAAAAAACA TATAAAACCA

2,230 2,240 2,250 2,260 2,270 2,280
AAAAACGAGA GAAAAAAAAA AACAAACAAA ACCAAAATCT AATTATCTTA GTAGACTAAT

2,290 2,300 2,310 2,320 2,330 2,340
CTAATTGGAG TTTCTTCCTT TCTTTAGAAG CTAGCAAAAC AAAAACAAAG AACAACAACA

2,350 2,360 2,370 2,380 2,390 2,400
ACCAGACAAA AACAAAACAT ACAATATCTG CTAATTTTAT TTTCATCTTT AAATTATGCT

2,410 2,420 2,430 2,440 2,450 2,460
CTATTATTAA ATATTAGTCA GAATATTAGT AAAACAAACG GAATTCAAAT TTGTTTTTTG

2,470 2,480 2,490 2,500 2,510 2,520
TTATTTGTTC TACTTAAAAT TAAAATATAA AACAGCTTTT GTAGTCGTTT CTGGCATCCA

```
     2,530       2,540       2,550       2,560       2,570       2,580
ACGTAACTCC TGTCAAGGCC AAATTTATAC CACGCAACGT CTTTTGCACC AGTAGTCAAA

     2,590       2,600       2,610       2,620       2,630       2,640
CGTCGGTCTC GATACTAACG GCCAGGGCAT TATTGTGCCT CATGGTGAGA CCGCTACTGC

     2,650       2,660       2,670       2,680       2,690       2,700
GTGCCCGCTC GTCGCCGCTG CTGCACTTGC GTTTATTGTG TGGGACTGTT GCTGTCTAGG

     2,710       2,720       2,730       2,740       2,750       2,760
CTGCTGCTGC GTCTGTTGCT GCTGCTGCTT GAAGCCGTTT TGCTGCTGCT GCTGCTGCTG

     2,770       2,780       2,790       2,800
TTGCTGCTGT TGGTACAGTT GCGTCTGCGT CAGCTGCTG
```