EXPRESSION, CHARACTERIZATION, AND LIGAND STUDIES INVOLVING DOMAINS OF THE CHICK CELL SURFACE PROTEIN NEOGENIN

Thesis by

Frank Miskevich

In Partial Fullfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1997

(Submitted October 15, 1996)

Acknowledgements

I would like to thank a number of people who have contributed in various ways to the completion of this thesis: Bill Dreyer, my advisor who was involved in virtually all phases of the research (except writing it up); Luis Sanchez in Pamela Bjorkman's lab who helped designing the bacterial chromatography scheme; Dirk Krapf and Gary Hathaway in the Protein Microanalytical Facility; Susan Ou in the Monoclonal Antibody Facility; Chantal Morgan in Steve Mayo's lab who ran the CD experiments; my thesis committee of Pamela Bjorkman, Scott Fraser, Steve Mayo, and Kai Zinn; fellow Dreyer lab members Anita Gould, Bob Lane, and Jost Vielmetter for regular consultations; and everyone else at Caltech and around LA who has helped to keep me somewhat sane over the last 6+ years.

Abstract

The vertebrate retinotectal projection provides an excellent developmental system to study mechanisms and molecules involved in precisely patterning the nervous system. The retina sends out a single topographic projection which maps in a one to one correspondence in the tectum. This correspondence is brought about by the interplay of numerous factors, including electrical activity, extracellular signals, and the interaction of various signal cascades within the retinal ganglion cell.

Neogenin is an alternatively spliced transmembrane protein homologous to a number of genes involved in neurite outgrowth and pathfinding. Its developmental expression in the retina suggested that it could be involved in differentiation or signaling events during the period when optic fibers are making initial connections in the tectum. The aim of this study was to identify extracellular and cytoplasmic signals carried by neogenin.

The immunoglobulin domains of neogenin were heterologously expressed in the yeast *Pichia pastoris* and biochemically characterized. This protein was then used to generate monoclonal antibodies against various epitopes in these domains hopefully to identify function blocking or cross-species reactive antibodies.

The intracellular isoforms of neogenin were expressed and characterized in *E. coli* to identify proteins which interact with neogenin. Proteins of 200, 140, 110, and 55 kD were specifically labeled in brain lysates. Both neogenin isoforms react with the proteins in a calcium dependent fashion. Affinity chromatography, antibody co-precipitation, and expression library screening were then attempted to identify these labeled proteins.

Table of Contents

Chapter 1: Introduction and Backround	I-1
Retinotectal Development	I-2
Ganglion Cell Signals During Development	I-6
Redundant Signalling Pathways	I-8
Signalling Pathways During Retinotectal Development	I-10
Cell Surface Molecules in Retinotectal System Formation	I-19
Secreted Factors in Retinotectal Development	I-29
Neogenin	I-31
Tumor Supressor Genes	I-35
References	I-37
Chapter 2: Domain Expression of Neogenin	II-1
Protein Expression Overview	II-2
Research Approach	II-6
Materials and Methods	II-7
Results	II-18
Discussion	II-27
References	II-32
Chapter 3: Monoclonal Antibodies Against Neogenin Ig Domains	III-1
Introduction to Monoclonal Antibodies	III-1
Materials and Methods	III-3
Monoclonal Antibody Results	III-5

Discussion III-8

References III-10

Chapter 4: A Search for Protein Ligands to Neogenin IV-1

Introduction to Ligand Binding Studies IV-1

Experimental Ligand Identification Plan IV-4

Materials and Methods IV-5

Ligand Results IV-9

Discussion IV-18

References IV-21

Appendix I: Expression and Purification of Protein Domains of Neogenin and Bravo/NrCAM Expressed in *P. pastoris* and *E. coli* for Production of Monoclonal Antibodies A I-1

Appendix II: Human Neogenin a Member of the Ig Superfamily Shows Strong Evolutionary Conservation and is Related to the Tumor Suppressor Molecule DCC A II-1

Chapter 1

Introduction and Backround to the Retinotectal System

Introduction

Perhaps the most complex of all biological structures, the nervous system co-ordinates an extraordinary array of processes. In order for the nervous system to develop properly, the appropriate signals must be sent and received both locally and globally. Axons must grow out from their cell body and navigate using extrinsic and intrinsic cues to find their correct target location. During neurite outgrowth, the environment of the cell and growth cone are constantly changing as neurons migrate from their place of birth to their eventual location, cells differentiate into new cell types and express different proteins than they did a few hours before, and axons cross the paths or fasciculate with other axons. Cells must also compete for survival with their neighbors, as perhaps half of all cells born in the embryonic nervous system do not survive through birth. Clearly amazing control and signalling methods are at work to generate stereotypic connections amidst such a confusing array of potential signals.

In order to gain insights into how such a bewildering array of connections can take place, it is usually useful to study a simpler system within the greater collection in order to dissect out signals that are relevant there, and then relate that back to the more complex case. Within the central nervous system, several different regions could be used as a simple model to study aspects of neuronal behavior. Of these, the retino-tectal system is perhaps the most studied and best understood in terms of the anatomy of the projection

and molecules that are present and may play a role in setting up the stereotypic pattern found in the system.

Secondly, it is necessary to choose an organism which is experimentally tractable in the desired system. Naturally the choice of organism at least partially determines what aspect of the system will be studied. Chickens make an excellent experimental organism for studying the developing retino-tectal system for a variety of reasons, and therefore has been studied quite thoroughly. First, the retina of a developing chick is extremely large compared to other organisms or even other areas of the chick central nervous system. Additionally, the pattern of connections are essentally made entirely pre-hatching, allowing all aspects of the development to be studied in a consistent environment. Although genetic analysis is not generally possible in chickens, excellent culture systems have been developed which allow relatively easy manipulation of other aspects of the system. For *in vivo* work, facile access to the developing embryo is possible allowing for ideas to be tested directly in the organism. Finally, large numbers of eggs can be obtained and incubated at low cost, which makes biochemistry in this system much easier.

Development of the Retino-Tectal System

The retina is the only primary sensory system to develop from the neurectoderm, the early cells which divide and form the central nervous system. It is generated as an outgrowth from the diencephalon at the 5-7 somite stage in the chick and develops into the optic cup. The optic cup is composed of two different layers, one differentiating into the pigmented epithelial cells and the second layer differentiating into ventricular cells, the

precursor cells of the other cell types found in the retina (reviewed in Graw, 1996). Ventricular cells differentiate into the majority of cell types found in the retina, including various glial cells and the five types of neurons: photoreceptor cells, amacrine cells, horizontal cells, bipolar cells, and retinal ganglion cells. Photoreceptor cells are found in the outermost layer of neuronal cell bodies, the outer nuclear layer. They connect to bipolar cells of the inner nuclear layer which passively connects the photoreceptors to the ganglion cells of the innermost cell layer, the ganglion cell layer. Amacrine and horizontal cells are found in the inner nuclear layer with the bipolar cells and mediate lateral communication between bipolar and ganglion cells and photoreceptor and bipolar cells, respectively.

The sole output cells of the retina are the ganglion cells. They project out through the optic fissure and form a bundle of axons called the optic nerve. The ganglion cells of the optic nerve, however, do not all project to the optic fissure at the same time. There is a distinct temporal/nasal gradient of projection and maturation, such that temporal axons tend to reach the fissure before nasal axons. Even at this stage, it is clear that the axons are responding to some sort of signal determining their direction of growth. The ganglion cells all project to the optic fissure, fasiculating with each other but suggesting that there is a molecular signal telling the axons to grow in a particular direction.

Once the axons have left the retina as the optic nerve, the axons continue to grow along a defined route until they reach the next choice point at the optic chiasm. In the chick visual system all axons project to the contra-lateral tectum, but in some other vertebrate systems axons project contra-laterally or ipsi-laterally depending upon the

portion of the visual field from which they receive input. These axon tracts cross at the midline and growth cone morphology becomes more complex and axonal growth slows in this region (Sretavan and Reichardt, 1993). The axons must then make a choice to follow one or the other path. Here again, axons must respond to an environmental signal based upon their location in the retina and choose the correct path to the tectum.

The optic nerve then continues growing along the outside of the diencephalon and enters the most superficial layer of the optic tectum, the stratum opticum. Just prior to entering the tectum, axons defasciculate and are able to grow independently in the tectum. The retinal axons make precise topographic connections in the tectum such that in the chick temporal axons connect to anterior tectum and nasal axons connect to posterior tectum. At least two different types of signals seem to be involved in generating the retinotopic map. First, axons grow along the surface of the tectum and their growth cones extend to the approximate location in the tectum in the absence of activity in the retina. This argues that molecular cues signal the growth cone that it is in the correct location in the tectum, and that retinal cells respond based upon their location. However, this initial map is imprecise. To refine the connections, electrical activity in the retina is required. This activity allows competion between the various axons sharing overlapping terminal arbors and eliminates redundant connections.

During or after the refinement process, numerous cells of the retina undergo programmed cell death between E12 and E17 (Clarke, 1992). Perhaps half of the retinal ganglion cells die during this period in the chick, although the percentage of cells dying appears to vary significantly between organisms. The regulation of cell numbers by

programmed cell death is thought to be mediated by competion for a neuronal survival factor such as BDNF (Herzog et al., 1994; Rodriguez-Tebar et al., 1989; Sawai et al., 1996). Thus a cell is thought to require a signal from its target in the tectum for the ganglion cell to survive.

Just before or during the time period when cell death is occurring, the ganglion cell axons leave the stratum opticum and project to particular retinorecipient lamina within the SGFS (stratum granulosum et fibrosum superficiale) of the tectum. Retinal axons are born with this information encoded in their array of receptors expressed on the growth cone and are able to recognize these signals even on fixed tissue in culture (Yamagata and Sanes, 1995a). Again, particular cell surface molecules are present either generally in the retinorecipient layers of the tectum (such as N-cadherin or the N-acetyl galactosamine (gal-NAc) carbohydrate) (Yamagata et al., 1995), or restricted to individual layers in the tectum (i.e., β-2 nicotinic acetylcholine receptor or substance P receptor) (Yamagata and Sanes, 1995b). These molecular address cues suggest that there may be instructive molecules telling the retinal axons to arborize in a particular layer due to their restricted expression on subsets of axons and complimentary patterns in individual layers of the tectum.

Each stage of axonal outgrowth during development of the retinotectal projection requires that axons make choices based upon environmental cues that they encounter. First, the appropriate signal must be generated in the appropriate place at the correct time during development to tell the axons what to do at that point. Secondly, the retinal axons must be able to receive this signal, which typically means that they must have proteins on

the surface of the growth cone to recognize the signal. Next, the retina must be able to transduce the signal from the outside to the inside of the cell, generally via a second messenger system. Finally, the retinal cell must be able to respond to the signal in an appropriate fashion based upon its location in the retina and its particular molecular specification telling it where to make synapses in the tectum.

Types of Signals Detected by Ganglion Cells During Retinotectal Development

The first signal presumptive ganglion cells must receive is to differentiate into a ganglion cell from the primitive ventricular cell. Future ganglion cells appear to have altered their gene transcription before outward phenotypic signs of commitment are evident (Matter et al., 1995). Ganglion cells have a number of properties unique among retinal neurons, including extending long projecting axons, expression of particular classes of neurotransmitters and receptors, and migrating to particular layers in the retina. Each of these activities require a particular combination of expressed proteins that grant a given cell a particular phenotype. However, since it is believed that all retinal ventricular cells are born equipotent (Maslim et al., 1986), individual cells must respond to an environmental signal to determine whether to become a ganglion cell or a different cell type.

Once formed, ganglion cells must extend their axons to their correct targets. In general, there are four different types of neuronal guidance signals that can be distingushed in experimental systems. Two of these, chemoattraction and chemorepulsion, act at a distance from the secreting cell and are typically thought of as

gradients of attractive or repulsive signal which orient an extending neurite to grow toward or avoid a particular region. The other two signals are mediated by contact and involve either growth over a permissive substrate or inhibition of growth by substrate molecules. In all of these cases, there can be blurred areas between these divisions, such as secreted molecules that are immobilized on the extracellular matrix. These molecules have diffused somewhat from the secreting cells, but they are likely to form extremely sharp gradients in the region of the secreting cells.

Chemoattractants are typically characterized by the ability to attract axons toward a diffusable signal imbedded in a collagen gel to limit and slow diffusion (Goodman, 1996). This was one of the first postulated mechanisms of directed axonal growth (Cajal, 1893) and there are several culture systems which exhibit this type of extension. Perhaps the most influential studies of this type are the chemotrophic effects of neuronal growth factor, NGF, (Gunderson and Barrett, 1979) on sensory axons, and more recently the discovery of the netrin family of diffusable attractants which are secreted from floor plate cells and attract commissural axons from spinal cord explants (Tessier-Lavigne et al., 1988; Serafini et al., 1994; Kennedy et al., 1994).

Currently there are no known chemoattractants of this sort involved in the developing retino-tectal system. However, diffusable factors are still very likely to play important roles in setting up connections between the optic fibers and tectal neurons. For example, brain derived neurotrophic factor, BDNF, has been shown to increase branching of retinal axons in rats *in vivo* (Sawai et al., 1996). This increase in branching may be required to form the synaptic arbor at the eventual site of synapse formation.

In addition to cues directing axon guidance, axons also appear to receive signals determining if a cell will survive or die. During the period of retinal ganglion cell death after the axons have reached the tectum, there appears to be a competition for a limiting factor produced by the tectum that is required for the ganglion cells to survive. Axons that are able to obtain enough of this factor, presumably those that make a sufficient number of tight contacts with the appropriate tectal cells, are preserved, whereas the others degenerate. Thus the number of retinal ganglion cells is at least partially controlled by the number and type of synaptic targets. BDNF is able to act as a cell survival factor for chick retinal ganglion cells in culture (Rodriguez-Tebar, 1989) and possibly *in vivo* as well (Herzog et al., 1994). Thus ganglion cells may need to integrate a number of different diffusable signals simultaneously in order to survive and form correct connections.

Multiple, Partially Redundant Signalling Pathways Appear to Direct Axon Guidance

Because of the bewildering number of potential signals that ganglion cells encounter, there must be mechanisms to distinguish between different types of signals received at the same time. Typically these signalling molecules fall into groups of related molecules which often share similar functions during development. Depending upon the particular subset of receptor expressed and ligands enocountered, the growing axon decides how it should react. However, only rarely is any protein required to do a single complex job. Rather, there are probably multiple systems that have partial overlap of

function such that disruption of one type of interaction may not be sufficient to disrupt the overall structure and accuracy of the system.

The *Drosophila* nervous system provides an excellent example where two different signal transduction pathways are involved in setting up the neuronal architecture. Mutations in Fasciclin I and the tyrosine kinase Abl have only slight perturbations individually, such that the overall scaffold of axons in either single mutant is relatively unchanged. However, the commissures in double mutant animals are completely disrupted, preventing axons that normally cross the midline from projecting to their proper locations (Henkemeyer et al., 1987; Zinn et al., 1988; Elkins et al., 1990). These results suggest that there are two different signalling pathways involved in directing axons to cross the midline. Either signal alone is sufficient to generate the gross anatomical features, but if both genes are knocked out, the animals have no other sufficient methods to direct the axons along their proper courses.

This *Drosophila* result, while not being directly related to development of the retinotectal system, may explain a few puzzling results of single mutants generated in mice. Several widely expressed proteins such as N-CAM and tenascin have been implicated in a number of neuronal functions for a variety of different systems and were believed to play major roles in neuronal development. However, when knockout mice were generated in these genes, the mice developed mostly normally, at odds with their presumed functions and widespread expression patterns (Cremer et al., 1994; Saga et al., 1992). In fact, for N-CAM one of the primary phenotypes (a slightly smaller olfactory bulb) could be duplicated by removal of polysialic acid (Ono et al., 1994). It seems likely

that the reason gross errors were not generally detected was because there are a number of pathways that partially overlap with the functions of these molecules and were able to compensate for their loss. While no single protein may be able to assume all of the roles of N-CAM, for example, other signalling molecules may be able to adjust for its absence and direct the axon accordingly. Recently, however, a transgenic mouse synthesized secreted NCAM which acts as a dominant lethal gene, killing the embryonic mouse (Rabinowitz et al., 1996). Presumably the secreted NCAM disrupted a vital redundant function by overstimulating a signalling pathway (otherwise the deletion should have a similar phenotype). Thus eliminating one critical protein may be unlikely to disrupt crucial axonal connections, making a very complicated system also a very robust one. It also suggests that deletion experiments may be incomplete when examining a redundant system.

Signalling Pathways During Retinotectal Development

Since cells must be able to receive and respond to a variety of signals, it therefore follows that there must be multiple signalling pathways in the cell to differentiate between the signals recieved. These signalling pathways need not be independent; in fact, interaction between two different pathways allow different signals to be integrated in a boolean manner. For example, if this receptor AND that receptor are activated, then differentiate--otherwise continue dividing. If this event has happened or these two other events have occurred, then commit suicide. How a cell signals and what other receptors

use a particular signalling pathway is essential to understand the biochemical events leading to differentiation and development.

Kinases

Virtually every pathway at some level or another relies upon phosphorylation to regulate signal transduction. There are three distinctive families of kinases which are characterized by the types of residues they are able to phosphorylate. Tyrosine kinases are often associated with cell membranes, either as transmembrane molecules (as in the Trk family of neurotrophin receptors), or as soluble kinases that often associate with transemembrane proteins. Serine/threonine kinases may also be associated with the cell surface (such as the insulin receptor or protein kinase C) and their activity often modulates the activity of another protein based on their phosphorylation state. The third type of kinase family is able to phosphorylate tyrosine, serine, and threonine residues (Gomez et al., 1990; Matsuda et al., 1993). It is often further downstream of the receptor, and in at least some documented cases mediates the transfer of the cytoplasmic signal to the nucleus. Examples of the first two kinase classes from the retinotectal system or other neuronal developmental models will be used to demonstrate how these kinases mediate developmental signals.

Tyrosine kinases are perhaps the most studied type of kinase. They are often associated with the cell membrane, either as a transmembrane protein or as a soluble tyrosine kinase. The trk family of tyrosine kinases are extremely important in the development of the retinotectal system. TrkB is the primary receptor for BDNF and is

widely expressed in the retina (Garner et al., 1996). It likely mediates axon sprouting (Sawai et al., 1996) at the tectum, as well as neuronal survival (Rodriguez-Tabar, 1989) through retrograde transport from its target tissue (Primi and Clarke, 1996; von Bartheld et al., 1996). TrkC, the NT-3 receptor, is expressed in virtually all cells of the developing retina and antibodies against NT-3 interfere with the differentiation of retinal ganglion cells and appears to disrupt the synapses between ganglion cells and amacrine as well as bipolar cells (Bovolenta et al., 1996).

Soluble tyrosine kinases also appear to be involved in outgrowth or target identification in ganglion cell growth cones. Inhibitors of the soluble protein kinases inhibit outgrowth of ganglion cell fibers in the optic tract (Worley and Holt, 1996). Furthermore, some biochemical evidence links the src tyrosine kinase with N-CAM dependent neurite outgrowth (Beggs et al., 1994) and the related fyn kinase seems to interact with the Ig superfamily member F11/contactin (Zisch et al., 1995) which is unexpected as contactin is anchored to the cell membrane by a glycosyl phosphatidylinositiol linkage. Thus tyrosine kinases play a vital role in the devleopment of the retinotectal system and neurite outgrowth in general. The soluble tyrosine kinase family also includes the abl kinase which is part of the redundant pathway generating the *Drosophila* neuronal archetecture.

Protein kinase C is a common serine/threonine kinase that is involved in retinal neurite development. Numerous different forms of protein kinase C (at least 6, with more being discovered regularly) are expressed in developing retina (McCord et al., 1996). These proteins have characteristic distributions that suggest they have various roles in

development. It appears to be involved in several functions, including the sharpening neurite connections during the development of the retinotopic map (Schmidt, 1994). Unlike the soluble tyrosine kinases or the serine/threonine calmodulin dependent protein kinase, however, PKC mechanisms do not appear to be involved in neurite outgrowth from retinal explants (Jian et al., 1994). PKC also seems to be involved in regulating hyperglycemia in a mouse model of diabetes (Kunisaki et al., 1995) as well as GABA responses in bipolar cells (Feigenspan and Bormann, 1994). Because PKC recognizes a relatively common motif (serines or threonines that are two or three amino acids downstream of two basic residues (Kennelly and Krebs, 1991)), the specific proteins phosphorylated during sharpening of the retinotopic map or regulating hyperglycemia may be restricted by their subcellular localization. The restricted expression of various PKC or PKC-like proteins would also somewhat limit its potential phosphorylation targets and hopefully restrict its activity.

Phosphatases

Naturally if there are enzymes to add phosphates, there must logically be a way of removing the phosphates once their function has been completed. Although not as fervently studied as the kinases, it is becoming increasingly obvious that they play an essential role in neuronal development. As with kinases, phosphatases are divided into tyrosine phosphatases and serine/threonine phosphatases, although there are dual specificity phosphatases as well. Many of the known phosphatases are transmembrane molecules, often with extracellular Ig domains that theoretically should interact with

ligands and thus potentially generate a signal dependent phosphatase activity. CRYP-alpha is an abudant CAM-like phosphatase found in the tectum, and multiple alternative splicing isoforms have been isolated that could alter the ligand specificity of the extracellular domain (Stoker, 1994). There is also a soluble tyrosine phosphatase, HVH5, that is expressed in the retina and has been demonstrated to have phosphatase activity against MAP-kinase (Martel et al., 1995), a dual specificity kinase that seems to be a central regulator of cytoplasmic signalling into the nucleus (reviewed in Ferrell, 1996). While functional studies in the retinotectal system are not yet published, motor neuron axonal targetting in several *Drosophila* phosphatase mutants appears to be disrupted (Desai et al., 1996) and various phosphatases are expressed in the vertebrate retina (Shock et al., 1995). Presumably these are functioning to regulate the activity of the various kinases which are required during retinotectal development.

G Proteins

G proteins fall into two different classes, the small monomeric type typified by ras, and the heterotrimeric G proteins. These two different types of G proteins have a number of features in common. First, they are each composed of multiple members that have different tissue and temporal distributions. Second, their activity is controlled by a common mechanism-- the exchange of GDP for GTP. They are inactivated by the conversion of GTP to GDP intrinsically, but the dephosphorylation rate is modulated by other protein exchange factors in the cell. Finally, their biological function (whether

activating, inhibiting, or both) depends upon the proteins with which they interact as well as what particular member of the family was activated.

Heterotrimeric G proteins are cytoplasmic proteins that are linked to the 7 transmembrane family of receptors. They are composed of three subunits, an alpha subunit which contains the GTP binding motif, and the beta-gamma dimer that appears to function as a unit. The alpha subunit is able to cause several different types of second messenger pathways to be activated including phospholipase C (Ashkenazi et al., 1989) and calcium flux (Ashkenazi et al., 1987), or inhibited such as adenylate cyclase (Dell'Acqua et al., 1993), depending upon the subfamily of alpha subunit coupled to the receptor. More recently, a single alpha subunit has been shown to mediate both inhibitory and stimulatory responses (Hunt et al., 1994), suggesting that the type of response generated by a particular receptor is controlled by both the subunit composition of the bound heterotrimeric G protein and the expression of particular G protein activated second messenger systems near where the signal is generated. The beta-gamma dimer is also important for generating a signal. Beta-gamma is able to activate the β-adrenergic receptor kinase (BARK) which can regulate the activity of the β adrenergic receptor (Hausdorff et al., 1990). Phosducin, a cytoplasmic protein expressed abundantly in the retina, is able to bind some (but not all) beta-gamma dimers and inhibit their functions (Muller et al., 1996).

Heterotrimeric G proteins mediate a number of processes in the retinotectal system. Transducin, the original G protein, and rhodopsin, a 7 transmembrane photoreceptor, form the prototype of a G protein signal transduction system. When 11-cis

retinal absorbs a photon of light, rhodopsin undergoes a conformational change that causes transducin to release GDP and bind a molecule of GTP, thus activating the G alpha subunit. The beta-gamma dimer is released (where it may interact with a kinase to regulate rhodopsin signalling by homology to activating βARK (Hausdorff et al., 1990)) and transducin activates an adenylate cyclase, generating the second messenger cAMP which closes a sodium channel. The intrinsic phosphatase activity of G-alpha eventually converts the GTP to GDP and thus ends the signal. A heterotrimeric G protein based signal transduction mechanism has also been shown to be involved in retinal axon outgrowth (Bates and Meyer, 1996).

Small G proteins of the ras or rho families are monomeric proteins that also exchange GTP for GDP when activated. Perhaps the best understood ras signalling pathway is in *Drosophila* ommatidia development, where ras is activated by the sevenless receptor (reviewed in Zipursky and Rubin, 1994). The activity of ras has been shown to be modified by at least two other proteins in the developing fly eye, drk and son of sevenless (sos) which modulate the nucleotide exchange of ras (Rogge et al., 1991; Simon et al., 1993). In mammalian cells, ras signalling has been closely associated with the MAP kinase in signal transduction to the nucleus (Russell, 1995). Ras has also been claimed to be able to transdifferentiate retinal pigment epithelial cells into neurons (Dutt et al., 1993). The small G protein rho (a protein related to ras but with different properties and expression patterns) has been implicated in the extension and retraction of lamellapodia and fillapodia during growth cone growth and extension (Mackay et al., 1995). Thus it

appears that various small G proteins are involved in controlling cell fate and differentiation and axonal extension.

SH2 and SH3 Domains

SH2 and SH3 domains refer to particular protein motifs first identified in the oncogene src and later found in a number of different proteins. Both of these domains are responsible for coupling cell surface proteins to particular cytoplasmic signalling molecules, typically kinases (Cicchetti et al., 1992) but in at least one case to the G protein exchange factor sos important for *Drosophila* ommatidia development (Olivier et al., 1993). SH2 domains recognize phosphotyrosine as a key element of their binding domain (Songyang et al., 1993). Each particular SH2 domain is then able to distinguish the environment around the phosophotyrosine to provide specificity to the interaction. A second phosphotyrosine recognition domain PTB (Kavanaugh et al., 1995) also recognizes phosphotyrosine but does not show homology to the SH2 domains. Thus SH2 recognition events are able to provide a protein interaction that requires a specific post-translational modification before occurring.

SH3 domains share no sequence homology to SH2 domains, and instead of binding a phosphotyrosine recognize a core motif of Pro - X - X - Pro. The PXXP motif adopts a polyproline type II helix (Lim et al., 1994) that is essential to generate the binding motif. There are two general classes of SH2 binding sites, and they differ in their composition outside of the core motif. Class 1 binding sites recognize P- X- X- P- X- X- R/K, and require the basic residue 3 positions downstream of the second proline (Feng et al., 1994;

Wu et al., 1995). Crystal structures show that a postively charged binding pocket recognizes the basic residue, and the individual residues of the binding pocket provide the specificity for either argenine or lysine. Class 2 binding sites recognize R/K- X- P- X- X- P using similar mechanisms (Feng et al., 1994). In both cases, the basic residues must be approximately one turn of the polyproline helix away from the core motif to be in the proper position for binding. Recent experiments have identified other residues outside of the core motif with basic residue that are involved in additional binding interactions to provide specificity of the binding event (Feng et al., 1995). Another possible specificity determining event in SH3 domain binding is the presence of multiple P-X-X-P motifs in the binding protein that require both motifs to bind SH3 domains in order to generate a biologically relevant interaction.

Cytoskeletal Mechanisms

The cytoskeleton is an intracellular framework consisting of a number of different polymers of individual units that provide structural integrity for the cell. Different cell types can contain different components. For example, while β tubulin is found in virtually all cell types, there is a specific isoform which is found primarily in projection neurons and not interneurons. Thus antibodies against the neural isoform of β tubulin can be used in the retina to fairly specifically label retinal ganglion cells and not other neurons (Snow and Robson, 1994).

Numerous transmembrane proteins are known to be able to bind to proteins of the cytoskeleton. For example, neurofascin, an Ig superfamily protein, contains a motif that

interacts with the actin binding protein ankyrin (Davis et al., 1993). The ankyrin binding motif is conserved in related molecules from *Drosophila* to chickens to humans (Davis and Bennett, 1994), and is therefore probably critical in the functioning of these molecules, some of which have been directly implicated in neurite outgrowth or axonal pathfinding (de la Rosa et al., 1990; Doherty et al., 1995). The integrin and cadherin superfamilies discussed below also seem to interact with the cytoskeleton through vinculin and the catenins, respectively (Otey et al., 1990; Stappert and Kemler, 1993).

The prevalence of transmembrane proteins being linked to the cytoskeleton suggests that the association is important for the functions of the molecules, possibly in axonal growth. As a growth cone extends filapodia and lamellapodia and senses its environment, the cytoskeleton is undergoing rapid extension and retraction (Gordon-Weeks, 1991; Challacombe et al., 1996). These movements must be based upon signals received from the extracellular environment and somehow transmitted to the cytoskeleton, possibly by tyrosine phosphorylation (Atashi et al., 1992). The second messenger Ca^{2+} is known to regulate actin filament assembly (Kater et al., 1988), and provides yet another mechanism for molecules at the growth cone to control cytoskeleton assembly and axonal growth. As other fillapodia or lamellapodia retract, those with filaments anchored to membrane bound proteins in a macromolecular complex could be maintained. In this way it is possible for the cytoskeleton to mediate signals for growing in a particular direction, the essential feature of axonal pathfinding in any neuronal system.

Cell Surface Molecules That May Be Involved in Retinotectal System Formation

Cell surface molecules are likely to play a key role in receiving signals before transducing them into cellular events. Typically the relevant receptor molecules fall into families of related proteins. Different family members share structural features in common, but subserve discreet functions during development and can involve widely varying methods of signal transduction as well.

Immunoglobulin Superfamily

The immunoglobulin (Ig) superfamily is perhaps the most diverse cell surface family of proteins known, and includes a wide variety of proteins involved in many different functions. They all share in common a protein domain called an immunoglobulin fold that consists of about 100 amino acids (the exact number varies considerably between different domains) arranged in two anti-parallel beta sheets that face each other (reviewed in Williams and Barclay, 1988). This superfamily is not well conserved between different family members, as the basic structure of the domain appears capable of tolerating a variety of different residues at many positions. The most conserved features of the domain are a pair of cysteine residues that form a disulfide bond between the two beta sheets to stabilize the structure (Williams and Barclay, 1988), although they are not absolutely required (Wang et al., 1990), several residues adjacent to the cysteines to allow the protein to fold properly, and a core hydrophobic region composed of the residues facing inward on the juxtaposed strands. These domains are divided into a number of subtypes, each having a characteristic number of strands in each sheet as well as slightly different conserved residues (Vaughn and Bjorkman, 1996).

Many of the Ig superfamily proteins contain multiple Ig domains, typically at the amino-terminus of the molecule. These domains are believed to be oriented as beads on a string or as semi-rigid rods with flexible hinges between adjacent interacting domains as known between the Fab and Fc regions of antibodies (Williams and Barclay, 1988). This arrangement is believed to help the protein extend away from the cell surface so that the molecule can interact with its binding partner or partners more easily. Ig domains can also be extended away from the cell surface by other domains, particularly fibronectin type III (FNIII) repeats that adopt a similar folding topology to Ig domains (Leahy et al., 1992).

Ig superfamily members also have adopted a number of different strategies for signalling (or not signalling). Some molecules such as one isoform of N-CAM and TAG-1 or its homolog axonin do not have a transmembrane region, but instead are linked to the cell surface by the use of a GPI anchor (Owens et al., 1987; Furley et al., 1990). Other family members contain domains which interact with the cytoskeleton such as bravo/NrCAM or neurofascin and may transduce a signal that way (Davis and Bennett, 1994). Some family members contain intracellular domains with enzymatic activities such as kinases (Jaye et al., 1992) or phosphatases (Streuli et al., 1989). Still others such as neogenin are not known to contain any enzymatic function but yet may still transduce a signal (Vielmetter et al., 1994), perhaps by other proteins which interact with the intracellular domains. The following list, while not necessarily complete, is intended to summarize specific functions of some Ig superfamily molecules that are found in the retinotectal system and what their roles might be during development. Ig superfamily

members which have been identified as kinases or phosphatases will be discussed under the appropriate sections.

N-CAM is a widely expressed protein of the Ig superfamily and is believed to play a role in a number of different types of interactions. N-CAM contains 4 Ig domains followed by 5 FNIII repeats that reach to approximately the cell surface. The molecule contains three different alternative isoforms (plus other microexons elsewhere) that give the molecule either a GPI- linked (120 KD), a transmembrane region plus short intracellular domain (140 kD), or a long intracellular domain (180 kD) (Owens et al., 1987). The protein is expressed uniformly along the optic nerve from the earliest ages examined and is thought to have a general adhesive role in the brain.

Numerous different functions have been ascribed to N-CAM and its various isoforms. N-CAM is known to be a homophilic adhesion molecule which can be blocked by the peptide K- Y- S- F- N- Y- D- G- S- E, a sequence found in the third Ig domain (Rao et al., 1993) or modified by sialylation (Edelman et al., 1983). Functionally, N-CAM supports neurite outgrowth (Bixby et al., 1987), and the amount of polysialic acid added to N-CAM has been shown to affect axonal growth characteristics (Edelman et al., 1983). A microexon in the extracellular domain (the VASE exon) has also been shown to differentially allow neurite outgrowth in developing cerebellum and hippocampus, and may be playing a similar role elsewhere as alternative splicing of NCAM is very common (Walsh et al., 1992). N-CAM, particularly sialylated N-CAM, is believed to be involved in learning and memory formation (Ronn et al., 1995). N-CAM transgenic mice, however, have been generated and appear mostly normal except for small defects in olfactory bulb

development and impaired learning (Cremer et al., 1994). It does, however, exert a dominant lethal effect on embryos when transgenically expressed as a secreted molecule (Rabinowitz et al., 1996). While its role in the formation of the retinotectal system is far from clear, it is perhaps the best studied Ig superfamily molecule in this family to date.

Another similar set of molecules is the NgCAM, Bravo/NrCAM, neurofascin subgroup of the Ig superfamily. This group of molecules contains 6 Ig domains, 4 or 5 FNIII motifs (determined by alternative splicing), a transmembrane domain, and a cytoplasmic domain that shares a sequence motif (NEDGSFIGQY) that is not only shared among all of this subgroup but the *Drosophila* protein neuroglian as well (Grumet et al., 1991; Volkmer et al., 1992). This peptide sequence is likely to be involved in the binding to ankyrin because the other related molecules have also been shown to bind to ankyrin (Davis and Bennett, 1994). Their tethering to the cytoplasm and its conservation between vertebrates and flies suggests that their cytoplasmic localization is functionally significant, perhaps involved in signal transduction.

Different members of this family seem to have somewhat different roles in the developing organism. NgCAM is known to bind to itself as well as axonin-1 (Kuhn et al., 1991), restrictin, and F11/contactin (Brummendorf et al., 1993), although the precise molecular nature of this interaction is unknown. It is also able to support neurite outgrowth of cultured retinal ganglion cells (Lemmon et al., 1989). Bravo/NrCAM does not, however, support neurite outgrowth of the same cells but does interact homophilically (de la Rosa et al., 1990; Mauro et al., 1992). Instead, it is thought that Bravo mediates fasciculation of the optic nerve, as axons grown in collagen matrix appear to fasiculate less

when treated with anti-Bravo antibody (de la Rosa et al., 1990), and Bravo expression is significantly decreased just prior to the optic tectum where axons must spread out over the surface in order to find their correct targets (Kayyem et al., 1991). Outside of the retinotectal system, Bravo/NrCAM is believed to influence the turning of commissural axons in the spinal cord (Stoeckli and Landmesser, 1995) and this effect is thought to be mediated by Bravo/Nr-CAM interacting with contactin (Morales et al., 1993).

Extracellular Matrix

Extracellular matrix molecules constitute a large class of proteins that are found outside of cells and provide a general framework through which the cells grow and migrate. These molecules are typically large, multi-domain proteins that interact with specific receptors for particular purposes. Two of the major proteins of this type in the retina are laminin and tenascin, both of which potentially affect development.

Laminin is an extremely common protein that is composed of three subunits. There are different isoforms of these subunits that are expressed in restricted patterns throughout the body and nervous system. S-laminin, for example, is expressed in the retina and may regulate photoreceptor cell morphogenesis (Hunter et al., 1992). Laminin is a widely recognized adhesive substrate for many different types of neurons, including retinal ganglion cells. It could provide a generally permissive environment through which axons must grow. Several different proteins have been identified as laminin receptors, including α-6 β-1 integrin (de Curtis et al., 1991). In the tectum laminin expression is limited to the basal lamina (Yamagata et al., 1995) and may be a general molecule

supporting the growth of axons over the surface of the tectum while they are finding their appropriate topographic targets.

Tenascin is an extracellular matrix molecule containing FNIII type repeats that can be alternatively spliced and assembles *in vivo* as a hexamer (reviewed in Reichardt and Tomaselli, 1991). It contains a variety of fibronectin repeats in complex patterns of alternative splicing, so that the specific protein expressed varies depending upon its location in the body. Tenascin is known to bind contactin through the use of a tenascin affinity column (Zisch et al., 1992). The interaction probably takes place through contactin's three most amino-terminal Ig domains, as proteolytic fragments of contactin containing these domains were the smallest proteins able to bind tenascin (Zisch et al., 1992). Tenascin's role *in vivo*, however, is somewhat uncertain as tenascin knockout mice develop normally (Saga et al., 1992). This may occur because there are multiple, partially redundant signalling pathways that are active in development. Although tenascin may be a component of several of these pathways, other proteins may be able to substitute for different roles of the molecule, thus giving no detectable phenotype in the knockout mutant. Overexpression of portions of tenascin, as already shown with N-CAM, might display significant neuronal phenotypes.

Integrins

Integrins are a dimer composed of an alpha chain and a beta chain in close apposition. Many different alpha and beta chains are known in multiple organisms, and it is not yet clear whether all alpha chains are able to associate with all beta chains. The

molecules are known to be involved in several different types of cell adhesion, and a number of different integrins are expressed in the retinotectal system in various patterns (Reichardt and Tomaselli, 1991).

Alpha-3 beta-1 integrin is known to bind a particular tripeptide (RGD) found in a fibronectin repeat (Pierschbacher and Ruoslahti, 1984). The tripeptide is found in a loop between the F and G strands in the tenth FNIII domain of fibronectin (Leahy et al., 1992), and free peptide is able to inhibit cells expressing the integrin from binding to fibronectin. This type of structural motif (a free loop extended out from the compact body of a protein) may be a general type of structure recognized by the integrin family. Multiple other short peptides are known to bind various other integrin dimers, including REDV and IDAPS binding to alpha-4 beta-1 (Massia and Hubbell, 1992; Mould and Humphries, 1991), and Mac-1 interacts with Gly^{190}- Val^{202} of fibrinogen (Altieri et al., 1993). Various other binding sites for integrins are likely to be found using combinatorial phage expression libraries (Koivunen et al., 1994).

Integrins transmit signals by at least two signal transduction pathways including ras- like small G proteins and tyrosine kinases (Parsons, 1996). In addition, integrins have been found to bind to the cytoskeleton using the protein vinculin (reviewed in Hynes, 1992). The cytoskeletal binding has been shown to be particularly important in point contacts with extracellular matrix, particularly with laminin (Reichardt and Tomasselli, 1991).

Cadherins

Cadherins are a family of calcium dependent homophilic adhesion molecules first identified due to their calcium dependent adhesion (Takeichi, 1988). Structurally, cadherins are a beta sheet protein that resembles Ig domains, FNIII repeats, and growth factor receptor domains (Shapiro et al., 1995), but contain a calcium binding loop that determines if homophilic binding will take place. Cadherins have been implicated in a number of aspects of cell adhesion and/or guidance functions, particularly during early development in neural crest cell migrations (Ranscht and Bronner-Fraser, 1991).

In the retio-tectal system, N-cadherin has a particularly striking distribution. As axons are initially growing into the tectum, N-cadherin is broadly expressed in many of the tectal laminae. As retinal axons dive into the tectum and begin to form synapses in specific laminae, N-cadherin becomes restricted to the same retinorecipient laminae (Yamagata et al., 1995). Furthermore, antibodies to N-cadherin are able to somewhat interfere with targetting to particular retinorecipient laminae suggesting that the molecule may be involved in stabilizing synapse formation (A. Inoue and J. Sanes, personal communication). This role in syanpse stabilization or strengthening is also being investigated in relation to LTP formation in the hippocampus (Erin Schuman, personal communication). N-cadherin also seems to act in a general adhesive nature in the retina as early retinal cells dissociate when treated with an N-Cadherin blocking antibody (Matsunaga et al., 1988). Retinas from later stages failed to dissociate, although the stereotypical patterning of the retina was disrupted.

Cadherins have several potentially independent methods of transducing a signal in the cell. First, they are associated with a group of three proteins known as alpha, beta,

and gamma catenins (Ozawa et al., 1989). The first of these proteins was discovered independently in *Drosophila* associated with segment boundaries (Riggleman et al., 1990), and they are vital for the biological role of cadherin function. Catenins are known to also associate with the cytoskeleton (Stappert and Kemler, 1993) and may be required to mediate cadherin function by directing cytoskeletal assembly or disassembly. Secondly, the cadherin/catenin complex associates with a 120 kD tyrosine kinase substrate related to beta catenin (Shibamoto et al., 1995). This protein may regulate cadherin function based upon its phosphorylation state. The catenins have also been assembled *in vitro* with E cadherin, with β catenin directly interacting with the cadherin (Aberle et al., 1994). Finally, there is some evidence that cadherins, like L1, may use the FGF as a co-receptor in cell signalling (Williams et al., 1994).

EPH Tyrosine Kinases and Ligands

EPH tyrosine kinases and their ligands are rapidly growing families of proteins that are involved in pattern formation in many different nervous systems during development (Lai and Lemke, 1991). These molecules seem to be extremely important in the retinotectal system, and appear to play key roles in the formation of the topographic map in the tectum. Mek4 is an EPH kinase that is expressed in a temporal-nasal gradient in the retina, with higher expression on the temporal side (Cheng et al., 1995). Its ligand, ELF-1, is expressed in a complimentary posterior to anterior gradient in the tectum (Cheng et al., 1995). A second EPH ligand homologue, RAGS, has been cloned independently and shown to have a growth cone collapsing ability and is expressed in a

posterior to anterior gradient (Drescher et al., 1995). These molecules therefore represent potential cell surface signals for the chemoaffinity hypothesis (Sperry, 1963) for axonal guidance. However, at least one other EPH kinase is expressed in a gradient (Kenny et al., 1995) of this sort and other proteins such as engrailed in the tectum (Itaskaki and Nakamura, 1992) or TOP in the retina (Savitt et al., 1995) may also be involved in the process.

Although there is not much information available on the downstream targets of the EPH kinases, expressed kinase domains are able to autophosphorylate probably as they would in the cell after activation of the receptor. Autophosphorylation of tyrosine residue 602 in the EPH family member Sek enabled p59fyn, a cytoplasmic tyrosine kinase, to associate with the EPH kinase domain via an SH2 domain interaction (Ellis et al., 1996). This offers another potential link between a cell surface molecule involved in axon outgrowth and pathfinding with a soluble tyrosine kinase, similar to the genetic link of the abl tyrosine kinase to axonal scaffold formation in *Drosophila* fasciclin I double mutants (Elkins et al., 1990).

Secreted Factors (Netrins, Neurotrophins and Semaphorins)

Secreted factors of various families play a number of vital roles in the development of the nervous system, including the retinotectal system. The neurotrophins and their receptors (the trk family of proto-oncogenes and the co-receptor p75) are perhaps the most recognized and most intensely studied secreted factors and receptors in the nervous system. The retinotectal system contains both brain derived neurotrophic factor (BDNF)

and NT-3 as well as their receptors, TrkA and TrkC (Garner et al., 1996; Primi and Clarke, 1996; Bovolenta et al., 1996; von Barheld et al., 1996). NGF and TrkB do not appear expressed in this system (von Bartheld et al., 1996). BDNF is known to enhance cell survival of cultured retinal ganglion cells (Rodriguez-Tebar et al., 1989), and the *in vivo* expression pattern (early expression in retina, followed by a decline of expression there and heavier expression in the tectum and ischimo-optic nucleus during the period of ganglion cell death) (von Bartheld et al., 1996) suggest that it plays a similar role in the developing organism. Indeed, BDNF has been shown to be retrogradely transported in retinal ganglion cells (Primi and Clarke, 1996; von Barheld et al., 1996). BDNF has also been implicated in branching of retinal ganglion cells (Sawai et al., 1996) and thus may be involved in causing axons to slow their growth and sense their chemical environment at key decision points during axon outgrowth.

The role of NT-3 in the retinotectal system is perhaps more general but certainly not less important. NT-3 seems to have similar branching enhancement functions as BDNF (Sawai et al., 1996). However, functional blocking antibodies to NT-3 disrupt multiple aspects of retinal formation and overall structure (Bovolenta et al., 1996). Possibilities as to the function of NT-3 in the retina include it being a survival factor for a type of non-ganglion neuron in the retina that may be responsible for helping to form the overall architecture of the retina or it may function more directly as a differentiation factor for cells of the retina. In any case, it is clear that BDNF and NT-3 have important roles in generating the adult retinotectal system.

Semaphorins, also known as collapsins, are a family of secreted proteins that act as inhibitors of axon outgrowth. Semaphorins have been isolated from many different organisms including *Drosophila* (originally as fasciclin IV), grasshopper, mouse, chicken, and human (reviewed in Kolodkin, 1996). Structurally they contain a large "sema" domain at the N-terminus, followed by a single immunoglobulin domain, a transmembrane region, and a unique intracellular domain. Their particular function in the retinotectal system is not completely understood. There appear to be at least two semaphorins that are expressed in the retinotectal system, but in the case of collapsin-1 retinal ganglion cells are insensitive to its inhibitory effects whereas sensory axons are sensitive to it (Puschel et al., 1995; Luo et al., 1993). One possibility is that the semaphorins of the retinotectal system are present to prevent axons other than optic fibers from innervating the tectum.

Netrins are a family of secreted, membrane associated proteins identified from floor plate cells that attract commissural axons in mouse spinal cord (Serafini et al., 1994), and are somewhat related to the amino terminal domains of the laminin β2 family of extracellular matrix proteins. The netrins are important in that they are a chemoattractant protein, a biochemical function that has been implicated in trigeminal nerve guidance (Lumsden and Davies, 1983) and cortical projections to subcortical targets (Heffner et al., 1990) as well as floor plate attracting commissural axons. While netrin expression in the retinotectal system has yet to be reported, it is possible that they may be involved in outgrowth of retinal ganglion cell axons (discussed later).

Neogenin

Neogenin is a chicken cell surface molecule isolated in a monoclonal antibody screen looking for proteins distributed in provocative patterns during retinotectal development (Vielmetter et al., 1994). It is expressed early in retinal development and can be seen in the optic fiber layer on retinal ganglion cell axons essentially as soon as they can be visualized. Expression also appears in the inner and outer plexiform layers as synapses begin to be made in those areas as well. At E7.5 when the slightly more mature temporal axons have reached the tectum while nasal axons have not, neogenin is distinctly down-regulated on the temporal axons, while its expression appears unchanged on nasal axons. By E10 when all axons have reached the tectum, neogenin is essentially not visible on optic fibers, although it is still visible in both the inner and outer plexiform layers. This expression is probably due to neurons in the retina other than retinal ganglion cells who are still making connections within these layers. At later stages, neogenin also is down regulated in the plexiform layers as well. Because neogenin expression is rapidly downregulated while axons are forming their topographic connections in the tectum, neogenin does not appear to be involved in this process per se. Rather, neogenin may be required for neurite outgrowth and synapse formation, or it could also be involved in the differentiation of the neuronal cells as these are the events taking place just prior to formation of the map.

Neogenin's expression in the cerebellum supports the view that it is expressed on newly forming neurites and is not found on mature, connected neurons. Early in the formation of the cerebellum, neogenin is expressed in the posterior of the half-spheroid structure. As the cerebellum grows and foliae develop, neogenin's expression is restricted

to the most posterior folds in the exterior granule cell layer. This pattern of posterior expression is maintained throughout embronic development. This region of the cerebellum is particularly informative because this is where new neurons are being born (LeDouarin, 1993).

The antibody used to identify neogenin was used to screen a bacterial expression library, and identified a clone that encoded an open reading frame. The gene was fully sequenced and found to contain four amino terminal Ig domains followed by 6 FNIII repeats, a single transmembrane domain, and a proline rich alternatively spliced intracellular domain. Although no specific enzymatic motifs were found in the cytoplasmic domain, there were 7 P-X-X-P core motifs that are core residues of a motif able to bind to SH3 domain containing proteins (Wu et al., 1995). Numerous potential phosphorylation sites were also identified within this domain, including a potential syk tyrosine kinase site.

A human clone of neogenin was identified by BLASTing the GenBank EST database of expressed human sequences, and that clone was used to probe a human library (Vielmetter et al., in preparation). The probe isolated a full length copy of human neogenin that was greater than 90% IDENTICAL to the chicken protein at the amino acid level, with even higher identity in the intracellular domain (94%). The intracellular domain contains the same potential alternative exon as the chicken gene, and the splice junctions are also identical between the two species. This striking evolutionary conservation suggests they almost certainly are playing the same roles in development.

Neogenin Homologues and their Roles

Neogenin appears to be a member of a growing subfamily within the Ig superfamily that contains 4 Ig domains, 6 FNIII repeats, and a cytoplasmic domain. Three other proteins have this same structure-- *unc-40*, a *C. elegans* mutant, *frazzled*, a *Drosophila* gene, and DCC, a human gene believed to be a tumor supressor protein.

Frazzled is the least studied member of the neogenin homologues. It is a *drosophila* mutation that shows a prominent midgut phenotype as well as axon pathfinding errors (Kolkdziej et al., 1995). The protein contains the same number of Ig domains and fibronectin repeats, but it has an intracellular domain that does not appear to be closely related to that of neogenin.

Deleted in colorectal carcinoma, DCC, is a human gene identified on chromosome 18q21.1 in loss of heterozygosity studies that is absent in some, but not all, colon tumors (Hedrick et al., 1994) and has subsequently been found absent in a number of other tumor types as well (reviewed in Cho and Fearon, 1995). DCC is the closest homologue to neogenin (Vielmetter et al., 1994), but closer homologues has been found in several different organisms including chicken (Chuong et al., 1994), *Xenopus* (Pierceall et al., 1994), and mice (Cooper et al., 1995) that argue neogenin and DCC are related proteins but serving different functions in the organism. Like neogenin, DCC is most abundantly expressed in the brain. DCC has also been shown to be required for neurite outgrowth, as outgrowing axons from cells where DCC expression has been interrupted, which causes the growing axons to stop extending their process and freeze (Lawlor and Narayanan, 1992). The axons do not extend further, nor do they retract. This suggests that DCC's

role in axon outgrowth is a permissive one, while absence of expression prevents further extension, an action which would be very appropriate for axons that have reached their correct locations in the brain.

The final protein related to neogenin, *Unc-40,* is a *C. elegans* mutation that affects the migration of circumferential axons and leads to an uncoordinated phenotype (Hedgecock et al., 1990). It appears to be required for ventral growth of numerous (but not all) axons including two of the three commissures that normally join the ventral nerve cord. Not all axons display pathfinding errors; some enter the ventral nerve cord at approximately their normal position, others wander around somewhat and enter it at an atypical position, while still others grow laterally and never join the ventral nerve cord. Once the axons make contact with the nerve cord, however, they fasciculate with it normally. This suggests that *unc-40* is specifically required for ventral migration rather than it causing a general growth defect. In addition, because some axons do grow correctly, and others grow incorrectly but still in a generally ventral direction, there are likely to be additional cues in the environment that the axons sense to orient their growth.

Unc-40 has been shown to interact genetically with *Unc-6,* a mutation that affects both ventral and dorsal projecting axons (Hedgecock et al., 1990). *Unc-6* is related to the netrin family of vertebrate proteins that attract spinal cord commissural axons (Serafini et al., 1994). In addition, *frazzled* mutants in flies appear similar to *Drosophila* netrin mutants (P. Kolodziej, personal communication). DCC, the other protein related to neogenin, has also been suggested to interact with netrin-1 (Marc Tessier-Lavigne, personal communication). This suggests the intriguing idea that perhaps neogenin is also

able to interact with netrin or a netrin related protein. Because of neogenin's pattern of expression, it could be involved in directing axon outgrowth from the ganglion cells toward the optic fissure or tectum. Once they have arrived at the tectum, neogenin expression would no longer be required and could therefore be down regulated.

Tumor Supressor Genes

Tumor supressor genes are a class of genes which, when mutated, increase the likelihood of tumor formation (reviewed in Knudson, 1993). Tumor supressor genes are usually discovered by collecting different independent examples of particular tumor types and examining them for common genetic defects. Typically this is done by examining loss of heterozygosity, which looks for large scale chromosomal deletions and narrow the candidate region to a few chromosomal bands (althogh in practice hundreds of open reading frames may be found in that region). Once a candidate molecule is identified, the gene can be used as a probe to examine both copies of a gene as well as families which have a history of a particular tumor type.

Currently there are about a dozen tumor supressor gene candidates. Several of these such as RB1 or WT1 (Friend et al., 1986; Call et al., 1990) are expressed in the nucleus and could be involved in cell cycle control or transcription. Other gene products including APC and NF1 are found in the cytoplasm (Su et al., 1992; Cawthon et al., 1990) and may regulate growth signals and intercept them, preventing them from affecting the nucleus. DCC is the only currently known cell surface candidate tumor supressor gene (although two other putative genes also may lie on the cell surface (Knudson, 1993)) and

presumably also acts as a control mechanism limiting the uncontrolled growth of cells in tumors. However, DCC also has at least one other function during development; it is required for axonal growth as shown in the antisense experiments. This places DCC (and by homology neogenin) as a key player in a pathway that controls growth and proliferation based upon signals received from outside the cell. The nature of these signals, both the extracellular signalling molecule and the intracellular pathways triggered is therefore of profound interest in understanding how cells integrate information and execute developmental programs leading to normal tissue formation.

References

Aberle, H., S. Butz, J. Stappert, H. Weissig, R. Kemler, and H. Hoschuetzky. (1994) *J. Cell Sci.* 107, 3655-3663.

Altieri, D. C., J. Plescia, and E. F. Plow. (1993) *J. Biol. Chem.* 268, 1847-1853.

Ashkenazi, A., E. G. Peralta, J. W. Winslow, J. Ramachandran, and D. J. Capon. (1989) *Cell* 56, 487-493.

Ashkenazi, A., J. W. Winslow, E. G. Peralta, G. L. Peterson, M. I. Schimerlik, D. J. Capon, and J. Ramachandran. (1987) *Science* 238, 612-615.

Atashi, J. R., S. G. Klinz, C. A. Ingraham, W. T. Matten, M. Schachner, and P. Maness. (1992) *Neuron* 8, 831-842.

Bates, C. A., and R. L. Meyer. (1996) *Brain Res.* 714, 65-75.

Beggs, H. E., P. Soriano, and P. F. Maness. (1994) *J. Cell Biol.* 127, 825-833.

Bixby, J. L., R. S. Pratt, J. Lilien, and L. F. Reichardt. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2555-2559.

Bovolenta, P., J.-M. Frade, E. Marti, M.-A. Rodriquez-Pena, Y.-A. Barde, and A. Rodriquez-Tebar. (1996) *J. Neurosci.* 16, 4402-4410.

Brummendorf, T., M. Hubert, U. Treubert, R. Leuschner, A. Tarnok, and F. G. Rathgen. (1993) *Neuron* 10, 711-727.

Call, K. M., T. Glaser, C. Y. Ito, A. J. Buckler, J. Pelletier, D. A. Haber, E. A. Rose, A. Kral, H. Yeger, W. H. Lewis, C. Jones, and D. E. Housman. (1990) *Cell* 60, 509-520.

Cawthon, R. M., R. Weiss, G. Xu, D. Viskochil, M. Culver, J. Stevens, M. Robertson, D. Dunn, R. Gesteland, P. O'Connell, and R. White. (1990) *Cell* 62, 193-201.

Challacombe, J. F., D. M. Snow, and P. C. Letourneau. (1996) *Semin. Neurosci.* 8, 67-80.

Cheng, H.- J., M. Nakamoto, A. D. Bergemann, and J. G. Flanagan. (1995) *Cell* 82, 371-381.

Cho, K. R., and E. R. Fearon. (1995) *Euro. J. Cancer* 31A, 1055-1060.

Chuong, C.-M., T.- X. Jiang, E. Yin, and R. B. Widelitz. (1994) *Dev. Biol.* 164, 383-397.

Cicchetti, P., B. J. Mayer, G. Thiel, and D. Baltimore. (1992) *Science* 257, 803-806.

Clarke, P. G. H. (1992) *J. Neurobiol.* 23, 1140-1158.

Cooper, H. M., P. Armes, J. Britto, J. Gad, and A. F. Wilks. (1995) *Oncogene* 11, 2243-2254.

Cremer, H., R. Lange, A. Christoph, M. Plowman, and G. Hopper. (1994) *Nature* 367, 455-459.

Davis, J. Q., and V. Bennett. (1994) *J. Biol. Chem.* 269, 27163-27166.

Davis, J. Q., T. McLaughlin, and V. Bennett. (1993) *J. Cell Bio.* 121, 121-133.

de Curtis, I., V. Quaranta, R. N. Tamura, and L. F. Reichardt. (1991) *J. Cell Biol.* 113, 405-416.

de la Rosa, E. J., J. F. Kayyem, J. M. Roman, Y.- D. Stierhof, W. J. Dreyer, and U. Schwarz. (1990) *J. Cell Biol.* 111, 3087-3096.

Dell'Acqua, M. L., R. C. Carroll, and E. G. Peralta. (1993) *J. Biol. Chem.* 268, 5676-5685.

Desai, C. J., J. G. Gindhart, L. S. B. Goldstein, and K. Zinn. (1996) *Cell* 84, 599-609.

Doherty, P., E. Williams, and F. S. Walsh. (1995) *Neuron* 14, 57-66.

Drescher, U., C. Kremoser, C. Handwerker, J. Loschinger, M. Noda, and F. Bonhoeffer. (1995) *Cell* 82, 359-370.

Dutt, K., M. Scott, P. P. Sternberg, P. J. Linser, and A. Srinivasan. (1993) *DNA Cell Biol.* 12, 667-673.

Edelman, G. M., S. Hoffman, C.- M. Chuong, J.-P. Thiery, R. Brackenbury, W. J. Gallin, M. Grumet, M. E. Greenberg, J. J. Hermperly, C. Cohen, and B. A. Cunningham. (1983) *Cold Spring Harbor Symp. Quant. Biol.* 48, 515-526.

Elkins, T., K. Zinn, L. McAllister, F. M. Hoffmann, and C. S. Goodman. (1990) *Cell* 60, 565-575.

Ellis, C., F. Kasmi, P. Ganju, E. Walls, G. Panayotou, and A. D. Reith. (1996) *Oncogene* 12, 1727-1736.

Feigenspan, A., and J. Bormann. (1994) *J. Physiol.* 481, 325-330.

Feng, S., J. K. Chen, H. Yu, J. A. Simon, and S. L. Schreiber. (1994) *Science* 266, 1241-1247.

Feng, S., C. Kasahara, R. J. Rickles, and S. L. Schreiber. (1995) *Proc. Natl. Acad. Sci. USA* 92, 12408-12415.

Ferrell, J. E. (1996) *Cur. Topics Dev. Bio.* 33, 1-60.

Friend, S. H., R. Bernards, S. Rogelj, R. A. Weinberg, J. M. Rapaport, D. M. Albert, and T. P. Dryja. (1986) *Nature* 323, 643-646.

Furley, A. J., S. B. Morton, D. Manalo, D. Karagogeos, J. Dodd, and T. M. Jessell. (1990) *Cell* 61, 157-170.

Garner, A. S., H. J. Menegay, K. L. Boeshore, X.- Y. Xie, J. M. Voci, J. E. Johnson, and T. H. Large. (1996) *J. Neurosci.* 16, 1740-1752.

Gomez, N., N. K. Tonks, C. Morrison, T. Harmar, and P. Cohen. (1990) *FEBS Lett.* 271, 119-122.

Goodman, C. S. (1996) *Annu. Rev. Neurosci.* 19, 341-377.

Gordon-Weeks, P. R. (1991) *BioEssays* 13, 235-238.

Graw, J. (1996) *Dev. Gen.* 18, 181-197.

Grumet, M., V. Mauro, M. P. Burgoon, G. E. Edelman, and B. A. Cunningham. (1991) *J. Cell Biol.* 113, 1399-1412.

Gundersen, R. W., and J. N. Barrett. (1979) *Science* 206, 1079-1080.

Hausdorff, W. P., M. G. Caron, and R. J. Lefkowitz. (1990) *FASEB J.* 4, 2881-2889.

Hedgecock, E. M., J. G. Colotti, and D. H. Hall. (1990) *Neuron* 2, 61-85.

Hedrick, L., K. R. Cho, E. R. Fearon, T. C. Wu, K. W. Kinzler, and B. Vogelstein. (1994) *Genes Devel.* 8, 1174-1183.

Heffner, C. D., A. G. S. Davies, and D. D. M. O'Leary. (1990) *Science* 247, 217-220.

Henkemeyer, M. J., Gertler, F. B., W. Goodman, and F. M. Hoffman. (1987) *Cell* 51, 821-828.

Herzog, K.-H., K. Bailey, and Y.-A. Barde. (1994) *Development* 120, 1643-1649.

Hunt, T. W., R. C. Carrol, and E. G. Peralta. (1994) *J. Biol. Chem.* 269, 29565-29570.

Hunter, D. D., M. D. Murphy, C. V. Olsson, and W. J. Brunken. (1992) *Neuron* 8, 399-413.

Hynes, R. O. (1992) *Cell* 69, 11-25.

Itasaki, N., and H. Nakmura. (1992) *Neuron* 8, 787-798.

Jaye, M., J. Schlessinger, and C. A. Dionne. (1992) *Biochem. Biophys. Acta* 1135, 185-199.

Jian, X., H. Hidaka, and J. T. Schmidt. (1994) *J. Neurobiol.* 25, 1310-1328.

Kater, S. B., M. P. Mattson, C. Cohan, and J. Connor. (1988) *Trends Neurosci.* 11, 315-320.

Kavanaugh, W. M., C. W. Turck, and L. T. Williams. (1995) *Science* 268, 1177-1179.

Kayyem, J. F., J. M. Roman, E. de la Rosa, U. Schwarz, and W. J. Dreyer. (1992) *J. Cell Biol.* 118, 1259-1270.

Kennedy, T. E., T. Serafini, J. R. de la Torre, and M. Tessier-Lavigne. (1994) *Cell* 78, 425-435.

Kennelly, P. J., and E. G. Krebs. (1991) *J. Biol. Chem.* 266, 15555-15558.

Kenny, D., M. Bronner-Fraser, and C. Marcelle. (1995) *Dev. Biol.* 172, 708-716.

Knudson, A. G. (1993) *Proc. Natl. Acad. Sci. USA* 90, 10914-10921.

Koivunen, E., B. Wang, and E. Rouslahti. (1994) *J. Cell Biol.* 124, 373-380.

Kolodkin, A. L. (1996) *Trends Cell Biol.* 6, 15-22.

Kolodziej, P. A., L. Y. Jan, and Y. N. Jan. (1995) *Neuron* 15, 273-286.

Kuhn, T. B., E. T. Stoeckli, M. A. Condrau, F. G. Rathgen, and P. Sonderegger. (1991) *J. Cell Biol.* 115, 1113-1126.

Kunisaki, M., S.-E. Bursell, A. C. Clermont, H. Ishii, L. M. Ballas, M. R. Jirousek, F. Umeda, H. Nawata, and G. L. King. (1995) *Am. J. Physiol.* 269, E239-E246.

Lai, C., and G. Lemke. (1991) *Neuron* 6, 691-704.

Lawlor, K. G., and R. Narayanan. (1992) *Cell Growth Different.* 3, 609-616.

Leahy, D. J., W. A. Hendrickson, I. Aukhil, and H. P. Erickson. (1992) *Science* 258, 987-991.

LeDouarin, N. M. (1993) *Trends Neurosci.* 16, 64-72.

Lemmon, V., K. L. Farr, and C. Lagenaur. (1989) *Neuron* 2, 1597-1603.

Lim, W. A., F. M. Richards, and R. O. Fox. (1994) *Nature* 372, 375-379.

Lumsden, A., and A. M. Davies. (1983) *Nature* 306, 786-788.

Luo, Y., D. Raible, and J. A. Raper. (1993) *Cell* 75, 217-227.

Mackay, D. J. G., C. D. Nobes, and A. Hall. (1995) *Trends Neurosci.* 18, 496-501.

Martell, K. J., A. F. Seasholtz, S. P. Kwak, K. K. Clemens, and J. E. Dixon. (1995) *J. Neurochem.* 65, 1823-1833.

Maslim, J., M. Webster, and J. Stone. (1986) *J. Comp. Neurol.* 254, 382-402.

Massia, S. P., and J. A. Hubbell. (1992) *J. Biol. Chem.* 267, 14019-14026.

Matsuda, S., Y. Gotoh, and E. Nishida. (1993) *J. Biol. Chem.* 268, 3277-3281.

Matsunaga, M., K. Hatta, and M. Takeichi. (1988) *Neuron* 1, 289-295.

Matter, J.- M., L. Matter-Sadzinski, and M. Ballivet. (1995) *J. Neurosci.* 15, 5919-5928.

Mauro, V. P., L. A. Krushel, B. A. Cunningham, and G. E. Edelman. (1992) *J. Cell Biol.* 119, 191-202.

McCord, R., A. Klein, and N. N. Osborne. (1996) *Neurochem. R.* 21, 259-266.

Morales, G., M. Hubert, T. Brummendorf, U. Treubert, A. Tarnok, U. Schwarz, and F. G. Rathjen. (1993) *Neuron* 11, 1113-1122.

Mould, A. P., and M. J. Humphries. (1991) *EMBO J.* 10, 4089-4095.

Muller, S., A. Straub, S. Schroder, P. H. Bauer, and M. J. Lohse. (1996) *J. Biol. Chem.* 271, 11781-11786.

Olivier, J. P., T. Raabe, M. Henkenmeyer, B. Dicson, G. Mbamalu, B. Margolis, J. Schlessinger, E. Hafen, and T. Pawson. (1993) *Cell* 73, 179-191.

Ono, K., H. Tomasiewicz, and T. Magnuson. (1994) *Neuron* 13, 595-609.

Otey, C. A., F. M. Pavalko, and K. Burridge. (1990) *J. Cell Biol.* 111, 721-729.

Owens, G. C., G. E. Edelman, and B. A. Cunningham. (1987) *Proc. Natl. Acad. Sci. USA* 84, 294-298.

Ozawa, M., H. Baribault, and R. Kemler. (1989) *EMBO J.* 8, 1711-1717.

Parsons, J. T. (1996) *Curr. Op. Cell Biol.* 8, 146-152.

Pierceall, W. E., M. A. Reale, A. F. Candida, C. V. E. Wright, K. R. Cho, and E. R. Fearon. (1994) *Dev. Biol.* 166, 654-665.

Pierschbacher, M. D., and E. Ruoslahti. (1984) *Nature* 309, 30-33.

Primi, M.- P., and P. G. H. Clarke. (1996) *Dev. Neurosci.* 7, 473-476.

Puschel, A. W., R. H. Adams, and H. Betz. (1995) *Neuron* 14, 941-948.

Rabinowitz, J. E., U. Rutishauser, and T. Magnuson. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6421-6424.

Ranscht, B. and M. Bronner-Fraser. (1991) *Development* 111, 15-22.

Reichardt, L. F., and K. J. Tomaselli. (1991) *Annu. Rev. Neurosci.* 14, 531-570.

Riggleman, B., P. Schedl, and E. Wieschaus. (1990) *Cell* 63, 549-560.

Rodriguez-Tebar, A., P. L. Jeffrey, H. Thoenen, and Y.- A. Barde. (1989) *Dev. Bio.* 136, 296-303.

Rogge, R. D., C. A. Karlovich, and U. Banerjee. (1991) *Cell* 64, 39-48.

Ronn, L. C. B., E. Bock, D. Linneman, and H. Jahnsen. (1995) *Brain Res.* 677, 145-151.

Russell, M., C. A. Lange-Carter, and G. L. Johnson. (1995) *J. Biol. Chem.* 270, 11757-11760.

Saga, Y., T. Yagi, Y. Ikawa, T. Sakakura, and S. Aizawa. (1992) *Genes & Dev.* 6, 1821-1831.

Savitt, J. M., D. Trisler, and D. C. Hilt. (1995) *Neuron* 14, 253-261.

Sawai, H., D. B. Clarke, P. Kittlerova, G. M. Bray, and A. J. Aguayao. (1996) *J. Neurosci.* 16, 3887-3894.

Schmidt, J. T. (1994) *J. Neurobiol.* 25, 555-570.

Serafini, T., T. E. Kennedy, M. J. Galko, C. Mirzayan, T. M. Jessel, and M. Tessier-Lavigne. (1994) *Cell* 78, 409-424.

Shapiro, L., A. M. Fannon, P. D. Kwong, A. Tompson, M. S. Lehmann, G. Grubel, J. F. Legrand, J. Alsnielsen, D. R. Colman. (1995) *Nature* 374, 327-337.

Shibamoto, S., M. Hayakawa, K. Takeuchi, T. Hori, K. Miyazawa, N. Kitamura, K. R. Johnson, M. J. Wheelock, N. Matsuyoshi, M. Takeichi, and F. Ito. (1995) *J. Cell Biol.* 128, 949-957.

Shock, L. P., D. J. Bare, S. G. Klinz, and P. F. Maness. (1995) *Mol. Brain Res.* 28, 110-116.

Simon, M. A., G. S. Dodson, and G. M. Rubin. (1993) *Cell* 73, 169-177.

Snow, R. L., and J. A. Robson. (1994) *Neuroscience* 58, 399-409.

Songyang, Z., S. E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W. G. Haser, F. King, T. Roberts, S. Ratnofsky, R. J. Lechleider, B. G. Neel, R. B. Birge, J. E. Fajardo, M. M. Chou, H. Hanafusa, B. Schaffhausen, and L. C. Cantley. (1993) *Cell* 72, 767-778.

Sperry, R. W. (1963) *Proc. Natl. Acad. Sci. USA* 50, 703-710.

Sretavan, D. W., and L. F. Reichardt. (1993) *Neuron* 10, 761-777.

Stappert, J. and R. Kemler. (1993) *Curr. Opin. Neurobiol.* 3, 60-66.

Stoeckli, E. T., and L. T. Landmesser. (1995) *Neuron* 14, 1165-1179.

Stoker, A. W. (1994) *Mech. Dev.* 46, 201-217.

Streuli, M., N. Krueger, L. Hall, S. Schlossman, and H. Saito. (1989) *Proc. Natl. Acad. Sci. USA* 86, 8698-8702.

Su, L. K., K. W. Kinzler, B. Vogelstein, A. C. Preisinger, A. R. Moser, C. Luongo, K. A. Gould, W. F. Dove. (1992) *Science* 256, 668-670.

Tessier-Lavigne, M., M. Placzek, A. G. S. Lumsden, J. Dodd, and T. M. Jessel. (1988) *Nature* 336, 775-778.

Vaughn, D. E., and P. J. Bjorkman. (1996) *Neuron* 16, 261-273.

Vielmetter, J., J. F. Kayyem, J. M. Roman, and W. J. Dreyer. (1994) *J. Cell Biol.* 127, 2009-2020.

Volkmer, H., B. Hassel, J. M. Wolff, R. Frank, and F. G. Rathgen. (1992) *J. Cell Biol.* 118, 149-161.

von Bartheld, C. S., R. Williams, F. Lefcort, D. O. Clary, L. F. Reichardt, and M. Bothwell. (1996) *J. Neurosci.* 16, 2995-3008.

Walsh, F. S., J. Furness, S. E. Moore, S. Ashton, and P. Doherty. (1992) *J. Neurochem.* 59, 1959-1962.

Wang, J., Y. Yan, T. P. J. Garrett, J. Liu, D. W. Rodgers, R. L. Garlick, G. E. Tarr, Y. Husain, E. L. Reinherz, and S. C. Harrison. (1990) *Nature* 348, 411-418.

Williams, A. F., and A. N. Barclay. (1988) *Ann. Rev. Immunol.* 6, 381-405.

Worley, T., and C. Holt. (1996) *J. Neurosci.* 16, 2294-2306.

Wu, X., B. Knudsen, S. M. Feller, J. Zheng, A. Sali, D. Cowburn, H. Hanafusa, and J. Kuriyan. (1995) *Structure* 3, 215-226.

Yamagata, M., J.- P. Herman, and J. R. Sanes. (1995) *J. Neurosci.* 15, 4556-4571.

Yamagata, M., and J. R. Sanes. (1995a) *Development* 121, 189-200.

Yamagata, M., and J. R. Sanes. (1995b) *Development* 121, 3763-3776.

Zinn, K., L. McAllister, and C. S. Goodman. (1988) *Cell* 53, 577-587.

Zipursky, S. L., and G. M. Rubin. (1994) *Annu. Rev. Neurosci.* 17, 373-397.

Zisch, A. H., L. D'Alessandri, K. Amrein, B. Ranscht, K. H. Winterhalter, and L. Vaughan. (1995) *Mol. Cell Neurosci.* 6, 263-279.

Zisch, A. H., L. D'Alessandri, B. Ranscht, R. Falchetto, K. H. Winterhalter, and L. Vaughan. (1992) *J. Cell Biol.* 119, 203-213.

Chapter 2

Domain Expression from the Chick Cell Surface Protein Neogenin

Introduction

We are interested in finding proteins which interact with the chick cell surface protein neogenin, both as ligands in the extracellular domain and as signal transducing molecules in the intracellular domain. We chose to take a biochemical approach, one that would look directly for protein-protein interaction. However, neogenin is expressed at relatively low levels in the cell, and it seemed somewhat impractical to isolate the amount of native neogenin required from chick tissue. In addition, because neogenin is a large and complex protein, we felt that it might be difficult to identify regions where interactions were occurring, making the experiments generally more difficult, and perhaps increasing the opportunities for non-biological interactions to take place.

Therefore, we turned toward the idea of domain expression as a way of studying the biochemistry of neogenin. This approach has many advantages over biochemical isolation of the intact protein. First, heterologous protein expression systems could produce massive amounts of protein that greatly facilitates any type of biochemical studies. Secondly, since we were engineering the protein, we could incorporate a purification tag in the recombinant protein to aid in purification and in analyzing proteins that bound to the expressed construct. Finally, we could focus our attention upon the regions of neogenin that appeared most interesting by only expressing those regions of the protein. We concentrated upon two regions of the protein--the immunoglobulin domains

for extracellular ligands based upon a large body of work suggesting Ig or Ig-like domains mediate numerous protein interactions (see Chapter 1) as well as the two alternative isoforms of the intracellular domain as they would be the region most likely to be involved in signal transduction.

Protein Expression Overview

In general, there is no single system that is best for protein expression. Numerous different systems are available, depending upon the requirements of the protein, including expression level, folding requirements, and post-translational modifications. We desired milligram quantities of protein, preferably from a stable producer instead of transient expression for all of the proteins. For the Ig domains in particular, we definitely wanted to secrete the protein because several studies had shown that bacterially produced Ig domains were only functional when secreted (Skerra et al., 1991). Thus there were several expression systems that would theoretically meet our requirements, including bacterial, yeast, insect (bacculovirus), and mammalian expression systems.

Bacterial expression systems are the oldest and most common heterologous expression organism as well as the simplest. There are numerous different types of expression vectors, including the T7 based promoter family of vectors (Moffat and Studier, 1987), and even a vector specifically designed to secrete proteins into the periplasmic space (Skerra et al., 1991). The T7 promoter is perhaps the strongest of these, and able to drive production of the mRNA to extremely high levels. The promoter itself is a short DNA sequence from the T7 bacterial phage, followed by a ribosome

binding sequence and a multi-cloning site where the insert is placed, which is then followed by a tϕ mRNA terminator region. Expression is carried out by addition of T7 RNA polymerase which specifically recognizes the T7 promoter. The polymerase is added either by IPTG induction of a transposon carrying the polymerase in particular bacterial strains or by viral infection with a phage that expresses the polymerase. Several bacterial strains carry the IPTG inducible T7 RNA polymerase gene, including BL-21 and BL-21 pLYS-S or pLYS-E (Studier and Moffat, 1986; Rosenberg et al., 1987). The pLYS-S strain expresses low amounts of intracellular T7 lysozyme that inhibits the T7 RNA polymerase and reduces transcription of genes downstream of the T7 promoter in the absence of IPTG induction. The pLYS-E strain expresses a higher amount of T7 lysozyme. The pT7SC plasmid is a special T7 promoter construct which includes two transcriptional repressor regions of the rrn gene upstream and downstream of the T7 expression cassette that reduces the transcription of the heterologous gene from cryptic or weak promoter sites (Brown and Campbell, 1993). Many proteins expressed using this promoter are packaged into insoluble inclusion bodies which facilitates their purification tremendously.

Yeast expression systems have several advantages over bacterial systems. Yeast is a eukaryotic organism, and as such it contains many of the same post-translational modifications found in other eukaryotes. It also contains a secretory system that is very similar to mammalian cells, whereas bacteria do not contain an endoplasmic reticulum. *Sarcomyces cerevisiae* is a popular choice for an expression system. Due to excellent genetics and biochemistry, the yeast is very well understood at the molecular level and

Gal4 (Brent and Ptashne, 1985) has proven to be an extremely robust promoter in this yeast. However, expression levels for many constructs are somewhat low for large scale protein production. In addition, secreted protein in *cerevisiae* often are hyperglycosylated (Ziegler et al., 1988) which can inhibit functionality of the protein.

To circumvent some of these problems, the yeast expression system *Pichia pastoris* was developed. *Pichia* typically adds only core mannose units to glycosylation sites, which, while not necessarily the correct modification for mammalian proteins, is nevertheless much smaller than that typically added by *cerevisiae* and less likely to interfere with the functions of the protein (Grinna and Tschopp, 1989). *Pichia* expression typically uses the alcohol oxidase promoter AOX1 to drive heterologous protein expression (Tschopp et al., 1987). This is an extremely strong promoter, able to generate up to 50% of the total cellular protein. It is also a tightly regulated promoter, being almost silent when carbon sources other than methanol are present. This allows an extensive cell mass to be generated even when a toxic gene product is being expressed. Finally *Pichia* is able to grow to extremely high density in fermentor cultures, with over 100 grams dry cell weight per liter (Cregg et al., 1993). Because of the high density and strength of the AOX1 promoter, *Pichia* can produce grams of protein per liter (Cregg et al., 1993), which is greater than levels usually obtained from the T7 promoter in bacteria yet with all of the advantages of a eukaryotic organism.

Insect cells have also become a popular host for heterologous protein expression due to the bacculovirus system. Bacculovirus relies upon the strong polyhedron promoter (which normally drives the expression of a viral coat protein) to give high levels of protein

expression, up to 30% of the total cell protein (Kidd and Emery, 1993). Recombinant bacculovirus clones are generated by co-transfecting bacculovirus DNA with a plasmid containing the desired gene under the control of the polyhedron promoter and a gene required for bacculovirus replication. Only bacculovirus that recombine with the plasmid are able to generate a plaque in the insect cell lawn, and recombinant clones typically also express a marker so they are easy to recognize. The bacculovirus clone is then plaque purified and expanded to generate a high titer virus stock that is used to produce quantities of recombinant protein. The expression system has been shown to reproduce many of the normal post-translational modifications found in other eukaryotes (Kloc et al., 1991) which make it ideal for some applications. However, it requires continual regeneration of the viral stocks and fresh cells must be reinfected to produce more protein. Furthermore, insect cell culture media is significantly more expensive than yeast or bacteria, making it less desirable for large scale protein production.

Mammalian cells are perhaps the most complex of the potential heterologous protein expression systems. However, they are often used for secreted protein expression because they are the most likely system to properly express, fold, and secrete a given protein. Furthermore, their post-translational modifications should be virtually identical to the native protein (with the single exception of tissue specific glycosylation events). Chinese hamster ovary (CHO) cells are a common choice for heterologous expression. They have been used for a variety of proteins (Davis et al., 1990; Harfst et al., 1992) and culture conditions are well developed. However, culturing the cells is expensive, and

purification of secreted proteins can be a problem. Purification can be improved by using serum free media, but media changes may alter the expression characteristics of the clones.

Another mammalian expression system that is exciting is a myeloma based cell line for secreting proteins, particularly antibodies. The expression plasmid includes a selectable marker that is dose dependent, so that it is possible to find cells that contain the most number of inserted genes and therefore the potentially highest levels of recombinant protein expression (Bebbington et al., 1992). This expression system has yielded up to 500 mg/L of secreted protein. However, to obtain the high yield cultures it is necessary to select for increasing levels of glutamine synthetase which may take up to 6 months.

Our Research Approach

We chose to explore two different expression systems, *Pichia pastoris* for extracellular Ig domain expression and *E. coli* for intracellular domain protein expression. Both systems are relatively inexpensive and easy to culture, give potentially high levels of protein production, and cells may be stably maintained. For *Pichia*, we decided to use the pPIC-9K vector which includes the expression cassette, the his4 gene for yeast selection, ampicillin resistance and an f1 origin of replication for bacterial propagation, and neomycin resistance to select for multi-copy integrants in *Pichia* which has been shown to increase expression levels for some proteins. For bacterial expression, we decided to use the T7 expression construct because of its high yield and relatively low uninduced promoter activity.

In order to purify the protein, we wanted to incorporate a purification tag on the recombinant protein. We incorporated a C-terminal six histidine tail in our PCR primers to purify the protein. This was chosen for several reasons, including the small size and expected low antigenicity of the tag, the same tag could be used for both expression systems, and the high degree of purification obtained by researchers using this purification method (Hochuli and Piesecki, 1992; Lindner et al., 1992).

Materials and Methods

Construction of Pichia pastoris *Expression Clones*

A bacterial clone (A6-BAA1-1) isolated by screening a lambda-ZAP II expression library with the monoclonal antibody 10-22A8 was used as a template to PCR amplify the Ig domain constructs for eventual expression. All cloning primers were purified using Nen-Sorb columns and diluted to 10 μM before use in PCR reactions. The primers used were as follows:

Neo1.for- 5' TCT AGA GAA TTC GTA GTG AGA ACC TTC ACT C 3'

Neo4.rev- 5' GC GGC CGC CCC GGG TTA ATG GTG ATG GTG GTG ATG TCG TGA AGT GGG AGG TAA TGT TG 3'

Neo2.for- 5' GGA TCC TTG TAA GCA GAA CAG 3'

Neo3.rev- 5' CTG CAG TTA ATG GTG ATG GTG GTG ATG GCG GCC GCG CTT CAG AAA CTC AGG 3'

PCR was carried out using the following PCR mix: 8 λ distilled water, 4 λ 1.25 mM dNTP's, 2 λ 10X Vent buffer, 2 λ 1 mg/mL acetylated BSA, 1 λ each forward and reverse primer (10 μM concentration), 1 λ diluted template DNA, and 1 λ Vent polymerase (0.5 units/ λ, diluted in Vent buffer) added during initial 72 °C incubation step. PCR was carried out under the following conditions: 92 °C for 5 minutes, 72 °C 5 min., followed by 25 cycles of 92 °C 0.5 min., 56 °C 1 min., 72 °C 2 min., and then a 72 °C 5 min. extension. The PCR product was analyzed for specificity on a 1% agarose-TAE gel, and the product was cloned using the TA Cloning Kit (Stratagene) according to the instructions.

Positive bacterial clones were selected by white-blue selection, double checked by PCR using the conditions above, and plasmid prepared using Wizard minipreps according to standard protocols. Plasmids were double digested with Not I and EcoRI to guarantee the presence of the restriction sites, and the plasmids were sequenced by either manual or automated dye terminator sequencing according to instructions. Positive clones with no PCR artifacts were then selected for expression.

Yeast expression plasmid pPIC-9K was prepared from XL-1 bacteria using Wizard minipreps according to standard procedures. The plasmid was digested with Not I, followed by EcoRI. The double digested plasmid was purified on a 1% agarose-TAE gel, spun through a 0.2 μm filter, and the agarose left was washed with 100 λ TAE buffer and spun through the filter. Plasmid was then dephosphorylated with calf intestine alkaline phosphatase and test ligated to guarantee that dephosphorylation was efficient.

Insert DNA was prepared by digesting TA vector+cloned insert with EcoRI followed by Not I and purifying the insert on a 1% agarose-TAE gel. Insert was cut from the agarose gel and spin filter purified as above. The ligation reactions were carried out at two different insert: vector ratios in order to optimize the reaction. T4 ligase (2 units) was incubated in ligase buffer (supplied with enzyme) and supplemented to 1 mM ATP using 1:1 and 3:1 molar ratios of insert to vector. The reaction was carried out at 12 °C for 6 hours to overnight. The ligation reaction was then heated to 65 °C for 10 minutes and precipitated with glycogen and 3 volumes isopropanol at -70 °C for at least 1 hour. The DNA was then spun down, washed with ethanol, dried, and resuspended in distilled water for electroporation in XL-1 bacteria. Positive colonies were selected using ampicillin resistance, and PCR was used to identify clones containing insert.

Electroporation of Pichia pastoris

A single colony of *Pichia pastoris* (strain GS115) was picked from a MD-his plate and grown in 25 mL MD-his broth at 30 °C to an $O.D._{600}$= 0.6 to 1.0. The cells were then spun down and washed 3x with 25 mL ice cold distilled water and resuspended in 25 mL 1.0 M sorbitol. The cells were then spun down again and resuspended in 250λ 1.0 M sorbitol and stored for up to 1 week at 4 °C.

Recombinant plasmid was miniprepped using Wizard miniprep kits according to the manual, eluting with TE buffer, pH 7.5. 25λ (approximately 4 μg DNA) of the prep was digested using 10 units of Bgl II enzyme at 37 °C for 2 hours, extracted once with an equal volume of equilibrated phenol:chloroform (1:1), extracted with chloroform, and 1λ

of the digest was analyzed on a 1% agarose gel to diagnose cutting. The remaining plasmid was precipitated with 40 mg glycogen and 2 volumes isopropanol at -20 °C for at least 1 hour. Plasmid was spun down at 4 °C in a microfuge, washed with 100λ 95% ethanol, and air dried. Linearized plasmid was resuspended in 5 λ distilled water for electroporation.

40λ of yeast was placed in a 0.1 cm. electroporation cuvette along with 1λ of linearized plasmid on ice. Electroporation was carried out using 400 ohms resistance, 1.25 keV, 25 μF; these conditions generally gave a time constant between 8.0 and 9.0 when the transformation was successful. 900λ of ice cold 1.0 M sorbitol was immediately added to the suspension, removed from the electroporation cuvette, and placed on ice. The yeast suspension was then plated on 2 MD plates: 100 λ was placed on one plate, and the rest was briefly centrifuged and resuspended in 100 λ distilled water and plated on a second plate. The plates were incubated 2-4 days at 30 °C until colonies appeared.

Screening for Expressing Clones

His independent colonies were picked into 150 λ distilled water in a microtiter plate. 2 λ of the suspension was spotted onto an MD plate, and 2 λ was spotted onto an MM plate in an identical position and incubated at 30 °C for 24 to 48 hours until clear differences between slow growing colonies and normal colonies were apparent.

Slow growing colonies were then examined by PCR to test for presence of the recombinant insert. Slow growing colonies were grown up in 1 mL MD broth and 100 λ

of yeast suspension was spun down in a microfuge. The yeast was then resuspended in 50 λ TE buffer, pH 7.5 and the cell wall was digested with 50 U lyticase (Sigma) for 2 hours at room temperature. The presence of insert was tested using specific primers for the insert (the same ones used for cloning). PCR conditions were as follows: 92 °C 5 min. (lysis step); 72 °C 5 min. (add enzyme here; i.e., hot start); and then 30 cycles of 92 °C 0.5 min.; 60 °C 1 min.; 72 °C 1.5 min. This was followed by 72 °C 5 min. (extension). PCR bands were then analyzed on 1% TAE- agarose gels.

PCR positive, methanol slow growing colonies were then screened for protein expression according to the method of (Barr et al., 1992). Single colonies were grown up in 10 mL BMMY broth at 30 °C with rapid shaking for two days, cells centrifuged down and resuspended in 2 mL BMMY for 2 days of induction at 30 °C with rapid shaking. Methanol (10 λ) was added after 1 day to replace that lost by evaporation. Yeast cells were spun down 10 minutes in a microfuge and supernatant was removed. 15 λ of supernatant was then run on a 10% SDS-PAGE gel in Tris-glycine-EDTA buffer, stained with coomassie blue for 4-8 hours, and destained overnight. Positive expressing colonies were clearly visible above background bands.

Preparation of Recombinant Protein from Pichia pastoris

Positive expressing colonies were grown 2 days in 10 mL MD broth at 30 °C to generate a starter culture for expression. The starter culture was used to inoculate a 2L culture of BMGY in a 6 L Erlenmeyer flask and shaken at 30 °C for 2 days to generate a large cell mass. The yeast were then spun down and resuspended in 400 mL BMMY,

transferred to a 2 L Erlenmeyer flask and shaken at 30 °C for 2 days covered only in 3 layers of sterile cheesecloth to guarantee adequate aeration. 2 mL of methanol were added after 1 day to replace that lost by evaporation.

The suspension was spun down at 4,000 g, 4 °C for 15 minutes, and the supernatant spun a second time for 1 hour at 10,000 g, 4 °C. The supernatant was then filtered through a 0.2 μm filter with glass fiber prefilter, and 1 λ/mL aprotinin and PMSF to 1 mM were added as protease inhibitors. The supernatant was then concentrated using a 76 mm YM10 10 kD or YM3 3 kD cutoff membrane (amicon) to approximately 40 mL. The concentrated supernatant was then washed with 400 mL of 0.1 M ammonium bicarbonate and concentrated to less than 40 mL.

Recombinant Protein Purification Using IMAC Chromatography

The IMAC column was prepared by taking 1 mL IDA-agarose (Pharmacia) and washing it with 10 volumes distilled water at 1mL/min. The agarose was then charged with 2.5 mL 10% nickel chloride in distilled water and washed with 20 volumes of distilled water. The column was then washed with 10 volumes running buffer.

Concentrated protein supernatant was diluted 1:1 with 2x running buffer and loaded onto the column at 1 mL/min. The column was then washed with 20 volumes running buffer. Weakly bound proteins were removed from the column by washing with running buffer supplemented with 10 mM imidazole. Ten 0.5 mL fractions were collected, followed by a 5 mL long wash which was collected in batch. Protein was then eluted into 10 0.5 mL fractions using running buffer plus 50 mM imidazole, and all remaining protein

was removed by using running buffer plus 200 mM imidazole, again collecting 10 0.5 mL fractions. Fractions were then analyzed on polyacrylamide gels, stained with coomassie blue for 4 or more hours, and destained overnight. Fractions containing protein were then concentrated in a centricon-10 or -30 spin concentrator and stored in 50% glycerol at -20 °C until used.

Construction of Bacterial Expression Clones

Chick neogenin clones containing both alternative splicing forms of the intracellular domain (NE3 and NE4) were used as starting material to generate PCR fragments for use in T7 promoter based bacterial expression systems. PCR primers were derived from the chicken neogenin sequence, and appropriate restriction sites and a carboxy- terminal purification tag were added to the oligos.

neo-cyt.S- 5' GAA TTC CAT ATG TGC ACT CGT CGT ACC ACT- 3'

neo-cyt-his.A- 5' GCGG CCGC GGA TCC TTA GTG ATG ATG GTG GTG ATG TCG TGC TGT AGT GAT GGC ATT -3'

PCR was carried out using the following PCR mix: 8 λ distilled water, 4 λ 1.25 mM dNTP's, 2 λ 10X Vent buffer, 2 λ 1 mg/mL acetylated BSA, 1 λ each forward and reverse primer (10 μM concentration), 1 λ diluted template DNA, and 1 λ Vent polymerase (0.5 units/ λ, diluted in Vent buffer) added during initial 72 °C incubation step. PCR was carried out under the following conditions: 92 °C for 5 min., 72 °C 5 min., followed by 25 cycles of 92 °C 0.5 min., 56 °C 1 min., 72 °C 2 min., and then a 72 °C 5 min. extension. The PCR product was analyzed for specificity on a 1% agarose-TAE gel,

and the product was cloned using the TA Cloning Kit (Stratagene) according to the instructions.

Positive bacterial clones were selected by white-blue selection, double checked by PCR using the conditions above, and plasmid prepared using Wizard minipreps according to standard protocols. Plasmids were double digested with Not I and EcoRI to guarantee the presence of the restriction sites, and the plasmids were sequenced by either manual or automated dye terminator sequencing according to instructions. Positive clones with no PCR artifacts were then selected for expression.

Bacterial expression plasmid pET-3a or pT7SC was prepared from XL-1 bacteria using Wizard minipreps according to standard procedures. The plasmid was digested with Nde I, followed by BamHI. The double digested plasmid was purified on a 1% agarose-TAE gel, spun through a 0.2 μm filter, and the agarose left was washed with 100 λ TAE buffer and spun through the filter. Plasmid was then dephosphorylated with calf intestine alkaline phosphatase and test-ligated to guarantee that dephosphorylation was efficient.

Insert DNA was prepared by digesting TA vector+cloned insert with NdeI followed by BamH I and purifying the insert on a 1% agarose-TAE gel. Insert was cut from the agarose gel and spin filter purified as above. The ligation reactions were carried out at two different insert: vector ratios in order to optimize the reaction. T4 ligase (2 units) was incubated in ligase buffer (supplied with enzyme) and supplemented to 1 mM ATP using 1:1 and 3:1 molar ratios of insert to vector. The reaction was carried out at 12 °C for 6 hours to overnight. The ligation reaction was then heated to 65 °C for 10 minutes

and precipitated with glycogen and 3 volumes isopropanol at -70 °C for at least 1 hour. The DNA was then spun down, washed with ethanol, dried, and resuspended in distilled water for electroporation in XL-1 bacteria. Positive colonies were selected using ampicillin resistance, and PCR was used to identify clones containing insert.

Bacterial Expression Conditions

The positive expression constructs were then transformed into BL21-pLYS-S using calcium chloride/heat shock transformation. Colonies were grown overnight on 2xYT/amp plates, and a single colony was picked into a 10 mL culture of 2xYT/chlor/amp. The culture was then grown to an $O.D_{600}$= 0.6 and induced with 2 mM IPTG to begin transcription of the T7 RNA polymerase. Cultures were then grown at room temperature for 24 hours.

Bacteria were spun down and resuspended at 1 mL inclusion body lysis buffer + DNAse I salt supplements per 100 mL of initial culture. The resuspended bacteria were then frozen at -70 °C for at least 1 hour to lyse the cells. Lysate was then quickly thawed and 40 λ of 1 mg/mL DNAse I was added. The lysate was incubated at room temperature until the viscosity of the solution was significantly diminished. The lysate was then centrifuged at 12,000 g for 15 minutes at 4 °C and the clear top solution was removed. The pellet was then washed with the same volume of inclusion body lysis buffer, recentrifuged, and the supernatant added to the previous supernatant.

IMAC Chromatography of Recombinant Protein

The IMAC column was prepared by taking 1 mL IDA-agarose (Pharmacia) and washing it with 10 volumes distilled water at 1mL/min. The agarose was then charged with 2.5 mL 10% nickel chloride in distilled water and washed with 20 volumes of distilled water. The column was then washed with 10 volumes running buffer.

Combined protein supernatant was diluted 1:1 with IMAC supplement buffer and loaded onto the column at 1 mL/min. The column was then washed with 20 volumes running buffer. Weakly bound proteins were removed from the column by washing with running buffer supplemented with 10 mM imidazole. Ten 0.5 mL fractions were collected, followed by a 5 mL long wash which was collected in batch. Protein was then eluted into 10 0.5 mL fractions using running buffer plus 50 mM imidazole, and all remaining protein was removed by using running buffer plus 200 mM imidazole, again collecting 10 0.5 mL fractions. Fractions were then analyzed on polyacrylamide gels, stained with coomassie blue for 4 or more hours, and destained overnight.

Ion-Exchange Chromatography

Positive IMAC fractions were then adjusted to low salt (< 50 mM) and 50 mM sodium phosphate buffer, pH 6.0. Protein was then loaded onto a 1 mL Mono-S FPLC column (Pharmacia) at 1 mL/min and washed with 3 mL 50 mM sodium phosphate, pH 6.0. Sodium chloride concentration was adjusted up to 1.0 M over 20 minutes using a linear gradient at a flow rate of 1 mL/min. Fractions containing protein were then analyzed on 12% polyacrylamide gels and positive fractions collected, concentrated, and used for further experiments.

Hydrophobic Affinity Chromatography

As a final step in purification, NE3 and NE4 protein was adjusted to 0.8 M ammonium sulfate, 50 mM sodium phosphate, pH 7.0 and loaded onto a 1 mL Phenyl Superose column (Pharmacia) at the recommended flow rate of 0.5 mL/min. Protein was eluted in a linear gradient of ammonium sulfate to 0 M, 50 mM sodium phosphate. Fractions were analyzed on 12% polyacrylamide gels and fractions containing the recombinant protein were pooled and concentrated.

N-Terminal Sequence Analysis

Recombinant protein was run on a 10% polyacrylamide gel under standard conditions. Pro-Blott (Bio-Rad) PVDF membrane was prepared according to manufacturer's recommendations, and proteins were blotted overnight at 150 mA in Towbin's buffer (39 mM glycine, 48 mM tris base, 0.074% SDS, 20% methanol). The blot was then rinsed in distilled water and stained briefly in 0.05% coomassie blue in 45% methanol, 45% acetic acid until desired protein bands were visible. The gel was then destained in 45% methanol, 45% acetic acid and thoroughly rinsed with distilled water. The desired protein bands were then excised from the PVDF filter and sequenced in the Caltech Protein/Peptide Microanalytical Core Facility using and ABI 373A protein sequencer.

Gel Filtration Chromatography

Gel filtration chromatography was carried out on a 25 mL Superdex 200 FPLC column (Pharmacia). Protein in PBS buffer was loaded onto the column at 0.5 mL/ minute and 1 mL fractions collected. Proteins were eluted isocratically at 0.5 mL/min. in PBS buffer and retention times calculated from the chromatograph. Protein molecular weight standards used were blue dextran (void volume), γ-globulin (150 kD), BSA (66 kD), ovalbumin (42 kD), and cytochrome C (14 kD).

Circular Dichroism

Circular dichroism experiments were performed using a 0.1 cm path length cell on an AVIV 62A spectrophotometer running the supplied IGOR software. Wavelength scans were performed going from 250 nm to 205 nm, with 2 nm increments using a time constant of two seconds averaged over 3 scans. Wavelength scans were obtained with 3 minute equilibration time, 1 nm band width, and +/- 5° deadband. Thermal denaturation scans were calculated from 25 °C to 99 °C sampling for 10 seconds with a time constant of 0.1 seconds (equilibration time 3 minutes). Thermal scanning was done at 230 nm using a 1 nm band width, +/- 0.5° deadband, and a heating slope of 10 °C per minute. Resulting molar ellipticities were graphed against wavelength using Kaleidograph. The extracellular Ig domains were used at 50 μM, and the intracellular portion was scanned at 20 μM concentration. All scans were conducted in 10 mM sodium phosphate buffer, and 1 mM DTT was added to the phosphate buffer for scans involving the intracellular portion of neogenin.

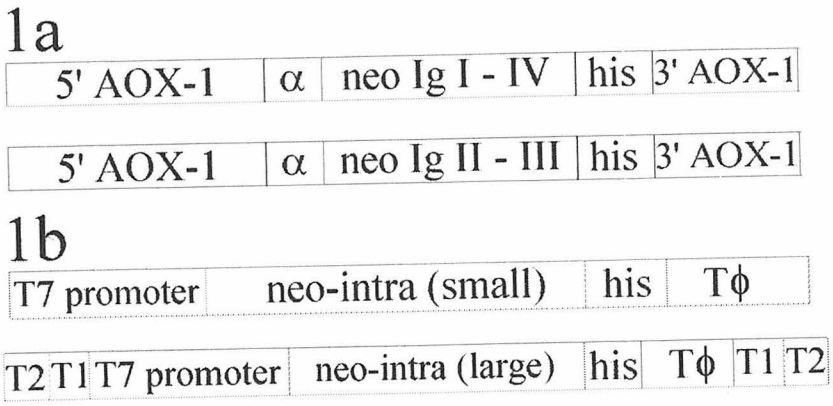


Figure 1: *Expression Casettes* Figure 1a shows the *Pichia* expression casette for both secreted constructs which use the AOX-1 promoter, α factor leader sequence, and 3' AOX-1 terminator. Figure 1b shows both expressed bacterial constructs with T7 promoter and tϕ terminator. The large isoform casette also includes T2 and T1 elements from the rrn operator of *E. coli.*

Results

Pichia *Expression Vector Construction and Screening*

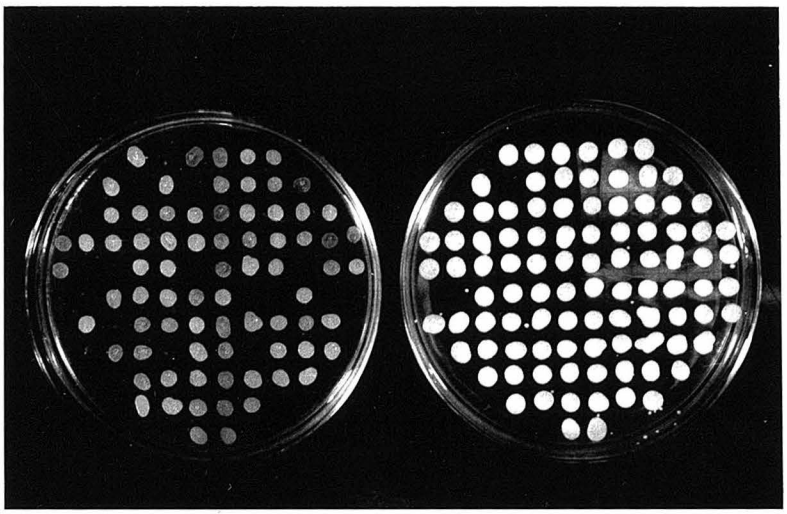


Figure 2: *Replica Plates* Figure 2 shows replica plates with a MM plate on the left and MD plate on the right. 19 of the 95 colonies (20%) grow poorly on MM plates.

The expression cassette controlling *Pichia* expression is diagrammed in Figure 1a. The expression vector was transformed into GS115 his4$^-$ yeast strain and screened for plasmid transformation on minimal media. Colonies were then replicated on MM and MD

plates looking for slow growing colonies (Mut^S phenotype). The percentage of slow growing colonies varied depending upon the insert present. The 4 Ig domain expression

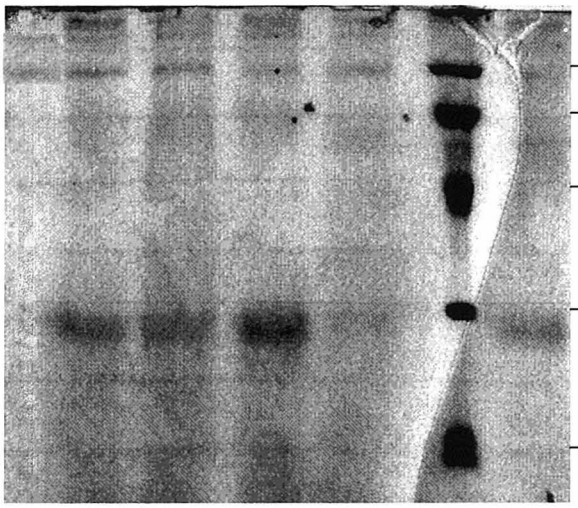


Figure 3: *Expression Screening* Figure 4 shows unconcentrated expression supernatants from screening cultures for the smaller 2 Ig domain construct. Protein levels vary from nearly undetectable to relatively high.

construct was integrated at approximately 30%, while the smaller construct only integrated in 20% of the colonies. An example of a replica plate is shown in Figure 2.

Colonies that grew slowly on MM plates were then picked from the MD plates and screened in shake flask cultures for protein expression. An example of this screen is

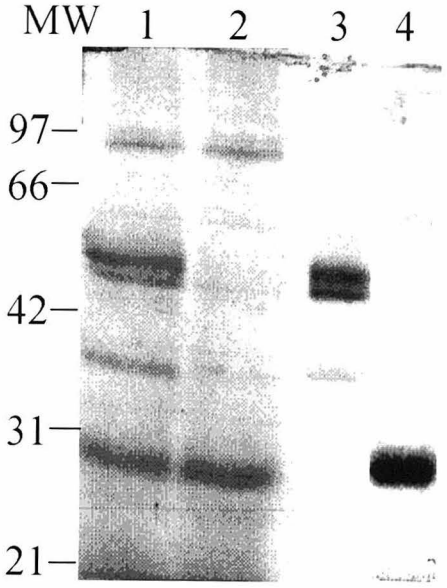


Figure 4: *IMAC Purification* lane 1--4 Ig domain construct concentrated supernatant. lane 2-- 4 Ig domain flow through. lane 3-- 4 Ig domain elution. lane 4-- 2 Ig domain elution.

shown in Figure 3 for the 2 Ig domain construct. The second and third Ig domain construct was expressed in detectable amounts in 11 of 18 colonies, while only 1 of 12 slow growing colonies expressed the larger 4 Ig domain construct.

Pichia Expression and Purification

Protein was prepared and purified by IMAC chromatography as shown in Figure 4. The IMAC flow through lane (lane 2) shows two bands specifically removed as compared to the original supernatant (lane 1). Protein was eluted using imidazole as a competitive chelating agent (lane 3) as an essentially pure protein. Note that there is a protein at approximately 36 kD that copurifies with the larger and more abundant 47 kD protein. Lane 4 shows only the purified small 2 Ig domain protein eluted from the IMAC column. This band is purified without bands of unexpected sizes.

Bacterial Expression Vector Construction and Screening

When the DNA encoding either isoform of neogenin's intracellular domain was placed in frame with the T7 promoter, it was difficult to obtain full length inserts. Figure 5 shows 40 colonies of the large intracellular isoform cloned into the expression vector. Several clones show the expected 1+ kb insert, but most clones show no insert (although plasmid is present in at least some colonies) and other inserts appear to be mixed or of smaller size. When plasmids giving PCR bands were digested with the enzymes used to clone the insert, only two colonies gave the expected band insert size.

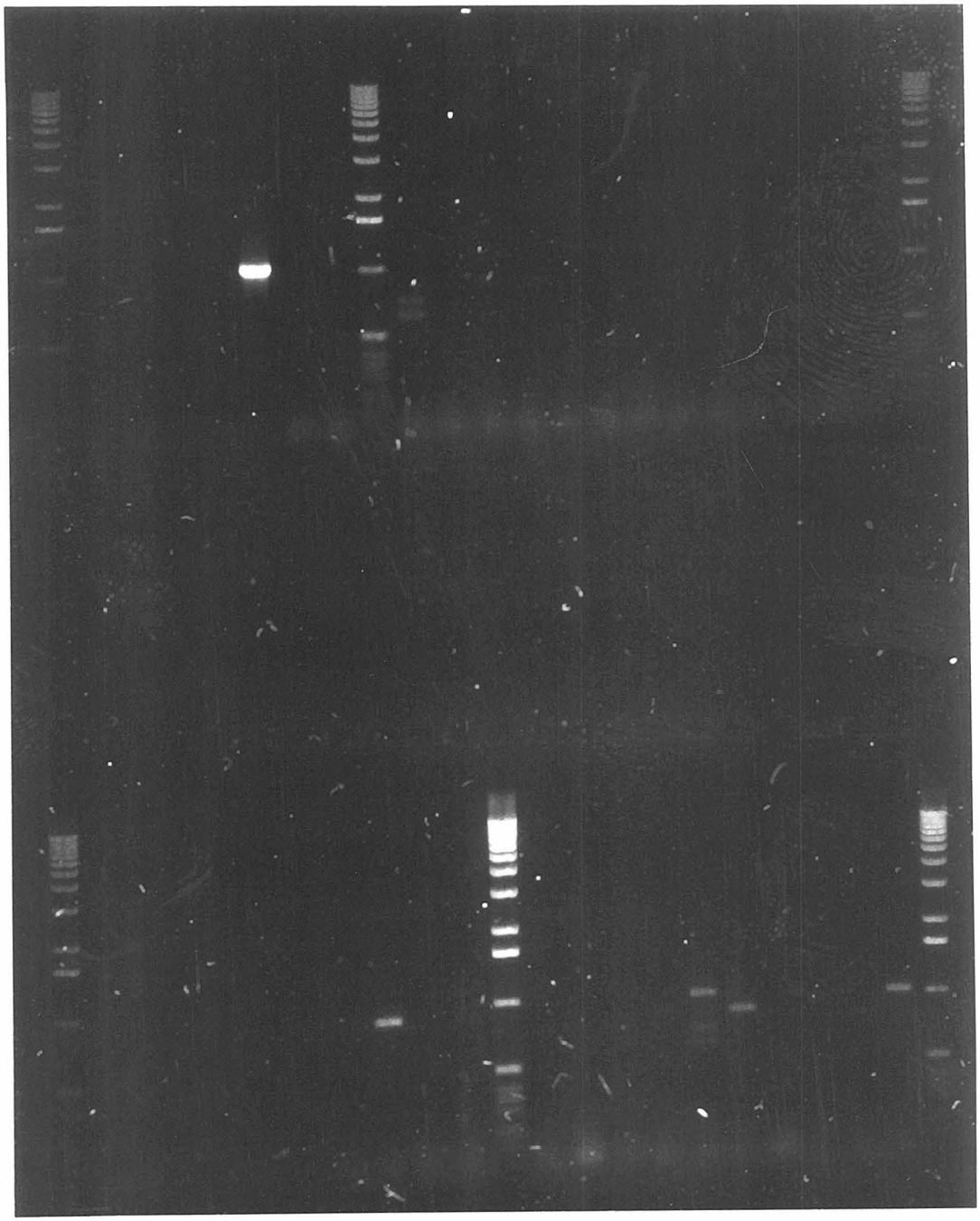


Figure 5: *Bacterial Expression Casette Construction* Figure 5 shows a 1% agarose gel examining 40 colonies cloning the large isoform of neogenin's intracellular domain under the control of the T7 promoter. Few colonies show the expected 1+ kb band, although several colonies show fragments of various sizes.

Protein expression levels are low compared to that expected from a T7 promoter based system as seen in Figure 6. The protein is found in the bacterial supernatant, and is

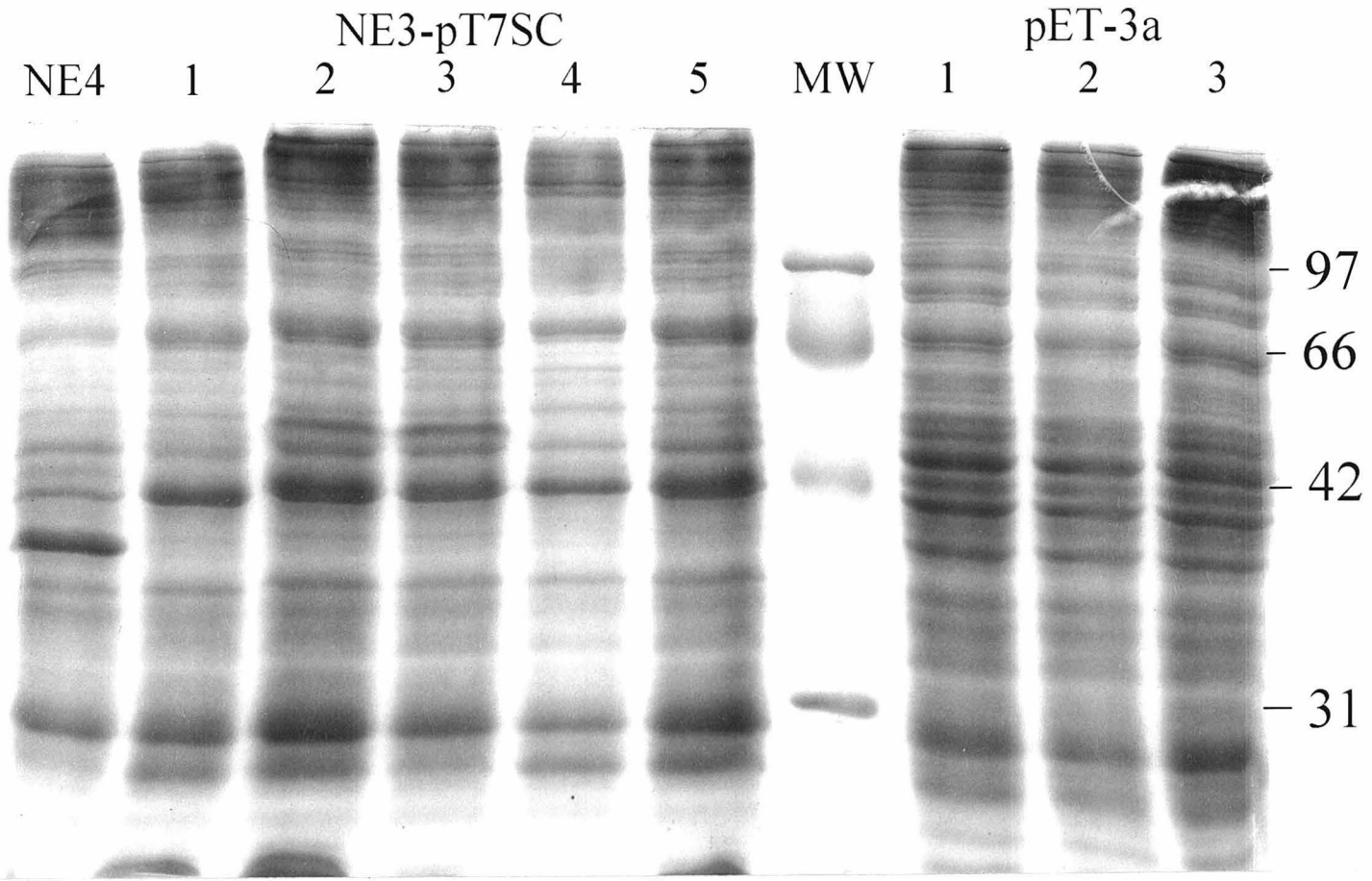


Figure 6: *Bacterial Expression* Figure 6 shows the expression of the various intracellular domain isoforms in bacteria. lane NE4-- small isoform in pET-3a vector. NE3-pT7SC lanes 1-5 -- bacterial lysate from pT7SC plasmid containing the large intracellular isoform. pET-3a lanes 1-3 -- bacterial lysates from the large isoform in pET-3a vector. The expected sizes are 36 and 42 kD, respectively.

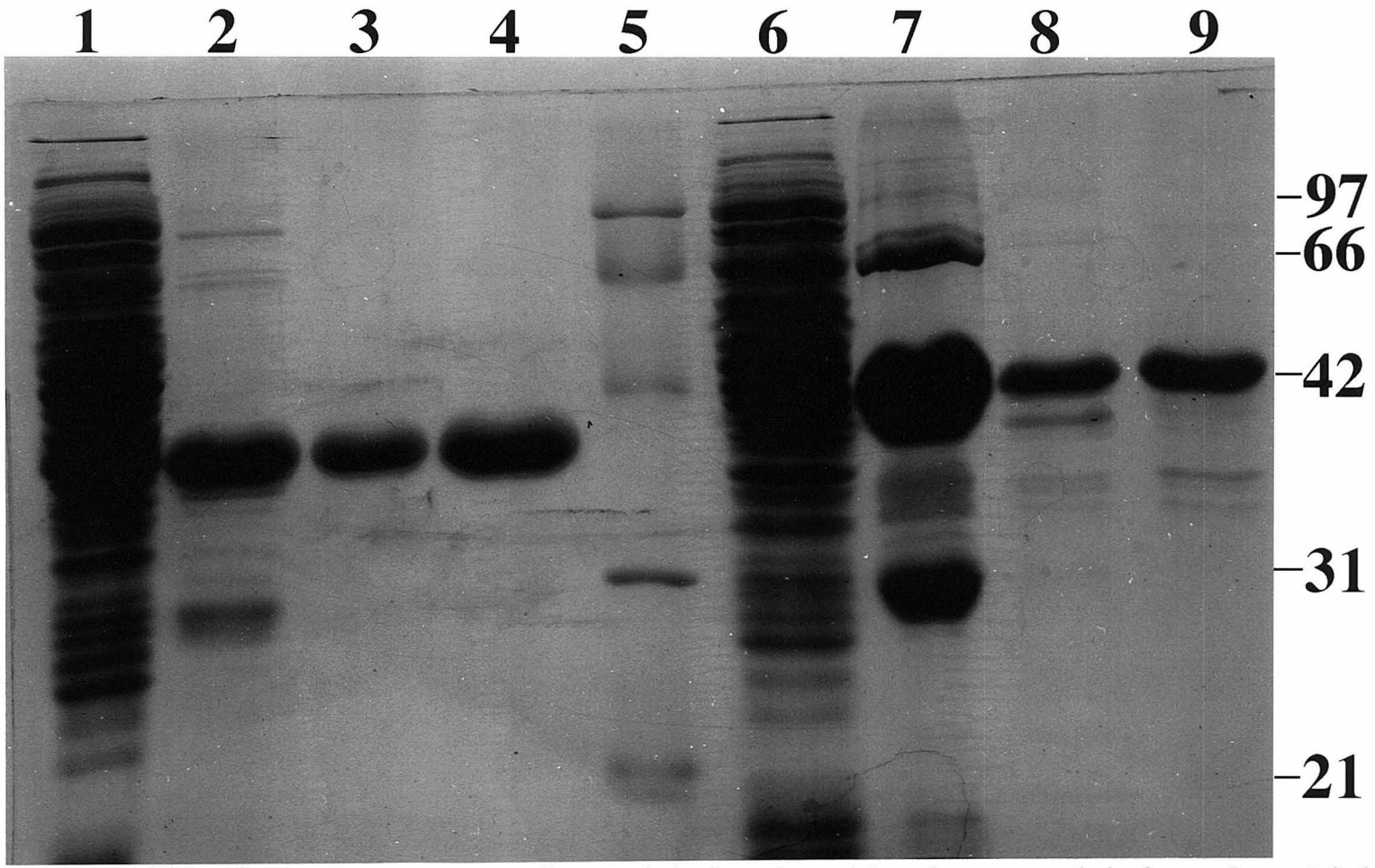


Figure 7: *Bacterial Purification* Lanes 1-4 are small neogenin isoform, lanes 6-9 are large neogenin isoforms. Lanes 1 & 6 -- total bacterial lysate. Lanes 2 & 7-- IMAC purified protein. Lanes 3 & 8-- Mono-S ion exchange chromatography. Lanes 4 & 9-- phenyl superose hydrophobic affinity chromatography.

visible as a 36 kD band for the small isoform (lane NE4), and as a 42 kD band in expression lanes from the pT7SC plasmid for the large (NE3) isoform. The NE3 band does not appear to be prominent in extracts from the pET expression vector. There does not appear to be an inclusion body in any bacteria expressing either isoform of neogenin's intracellular domain.

Bacterial Protein Purification

The purification of bacterially produced intracellular domain is shown in Figure 7. IMAC chromatography significantly purified both isoforms of the expressed proteins (compare lane 1 with 2 and lane 6 with 7), although there are clearly contaminating proteins that copurify with the expressed protein. Mono-S cation exchange chromatography (lanes 3 and 8) followed by hydrophobic affinity chromatography (lanes 4 and 9) purified the proteins nearly to homogeneity. The overall yield for either recombinant protein is approximately 1.5 mg pure protein per liter of expression culture.

Amino Acid Sequencing

Table 1 shows the amino terminal sequence obtained from the purified recombinant proteins. Several features are apparent from the table. First, there is variable processing of amino termini in the *Pichia* expression system. The large 4 Ig domain construct includes all residues encoded in the DNA sequence, including those added due to restriction sites incorporated into the PCR primers. There is also a protein of 36 kD that co-purifies on the IMAC column with the 4 Ig domain construct. This band was also

sequenced, and found to contain a frayed N-terminus with three predominant ends that all begin approximately midway between first and second Ig domains based on their distance from the conserved cysteines. These residues correspond to residues 165, 166, and 169 of the full length neogenin protein. The smaller 2 Ig domain construct, however, had 20 amino acids removed from the expected N-terminus. Bacterial processing is consistent, in that each isoform contains identical residues as expected, beginning two amino acids downstream from the predicted beginning based upon DNA sequence. The first residue was indeterminate, most likely because the predicted amino acid, a cysteine, is only visible in the protein sequencer when alkylated.

Table 1

Amino terminal Sequencing Results

large intracellular isoform	X-T-R-R-T-T
small intracellular isoform	X-T-R-R-T-T
two Ig domain construct	S-N-P-E-L-S
four Ig domain construct	Y-V-E-F-V-V-R-T-F-T-P-F
36 kD Ig domain fragment	G-L-P-R-F-T-S-Q, T-V-A-G-L-P-R-F, L-T-V-A-G-L-P-R

Gel Filtration Chromatography

Gel filtration chromatography was performed upon all of the expressed proteins to characterize their biochemical properties. Table 2 shows the retention times of all the recombinant proteins as well as those of the included standards. Molecular weight was calculated from a semi-log plot based upon the retention time and known molecular

weights of the standards. All proteins ran approximately as expected based upon their apparent molecular weights determined by polyacrylamide gels.

Table 2

Gel Filtration Table

Protein	retention	mol. weight
blue dextran	8.03	2000 kD
gamma globulin	11.73	150 kD
bovine albumin	13.18	66 kD
ovalbumin	13.51	42 kD
cytochrome C	16.7	14 kD
small intracellular isoform	13.62	42 kD
large intracellular isoform	13.4	52 kD
4 Ig domains	13.39	52 kD
2 Ig domains	14.22	34 kD

Circular Dichroism

Circular dichroism spectroscopy was performed in order to assess the folding of the expressed proteins, and the data are shown in Figure 8. The small intracellular isoform of recombinant neogenin does not appear to have any destinguishing characteristics for CD in 10 mM sodium phosphate buffer. The large intracellular isoform was not analyzed by CD. Both the 4 Ig domain and 2 Ig domain proteins, however, exhibit a negative ellipticity around 216 nm and a plateau between 222 nm and 230 nm under non-denaturing conditions. When heated to denaturation, both proteins lose this peak, with the 2 Ig domain construct ressembling the small intracelluar domain isoform, while the 4 Ig domain protein has a single broad, deep negative ellipticity around 220 nm.

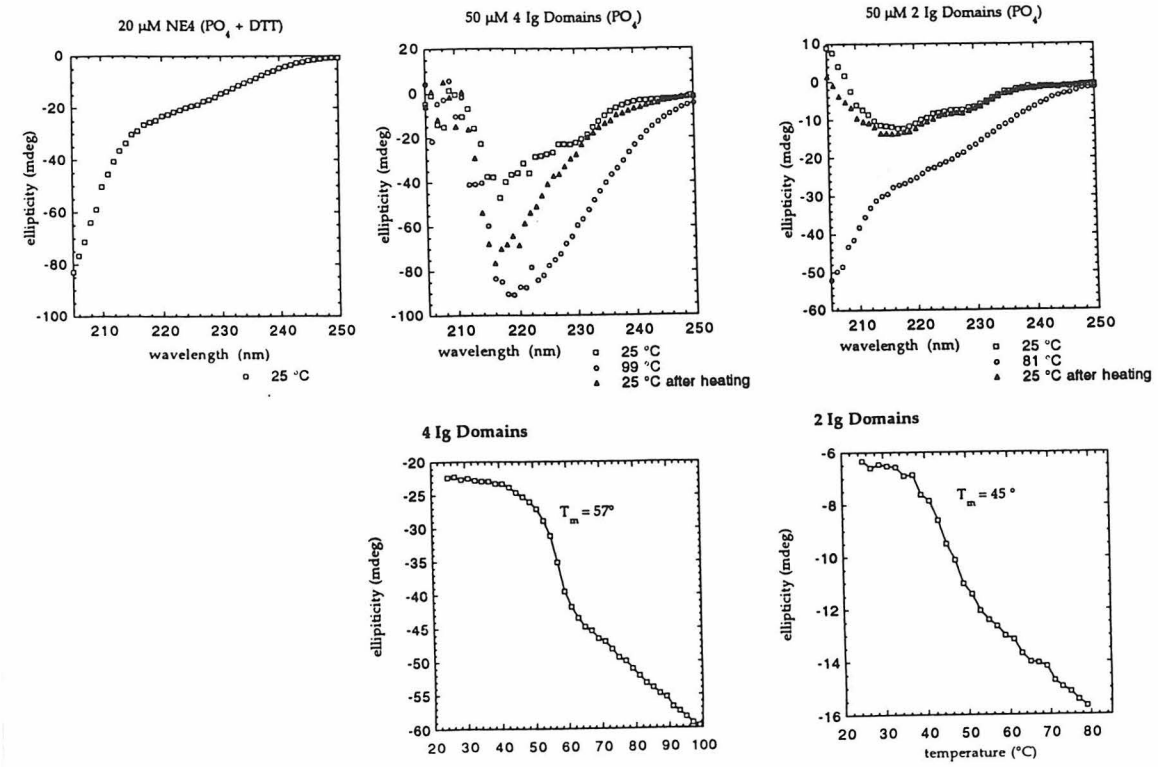


Figure 8: *CD Spectra* Figure 8 shows the CD spectra for each of the expressed proteins at 25 and 90 °C in the range from 250 to 205 nm (plotted as molar ellipticity vs. wavelength), and the thermal denaturation curves for both recombinant Ig domain constructs.

Discussion

Protein expression systems differ in their ability to produce recombinant protein effectively. Currently, there is no known way to predict which host will be able to make any given protein efficiently. *Pichia pastoris* appears to be an excellent vector for

producing the extracellular domains of neogenin. A high percentage of histidine independent colonies (20 to 30%) showed the Mut^S phenotype, diagnostic for integration at the AOX1 locus. This is within the expected range of integration based on published results (Cregg et al., 1993). Recombinant protein is clearly visible as a distinct band in the culture supernatant relatively free of contaminating yeast proteins. Purification from the media using IMAC chromatography results in essentially pure protein after a single step. The 6 histidine tag appears to be a very efficient purification method for secreted proteins expressed in this organism.

The intracellular domains of neogenin were relatively poorly expressed in *E. coli*. The T7 promoter is an extremely powerful one, often making proteins that get packed into inclusion bodies. Neogenin's intracellular domain, however, is soluble and is produced at levels only 10% of that obtained with a fibronectin domain of Bravo (R. Lane, in preparation). One possible reason this protein is poorly produced is that it may be toxic to the bacteria. Several different lines of evidence support this theory. Identification of bacterial clones before protein production but after putting the insert in frame under the control of the T7 promoter was very difficult, with only 2 of 40 colonies containing a full length insert DNA with the appropriate restriction sites at the 3' and 5' ends. The large isoform of neogenin required terminator elements of the *rrn* operon found in the expression plasmid pT7SC (Brown and Campbell, 1993) for the protein to be produced at all. Both isoforms were produced at low levels even though high levels of IPTG and long induction times were used. Finally, BL21-pLYS-S hosts needed to be freshly transformed with the expression plasmid in order to maintain protein expression. All of these pieces of

evidence suggests that the protein is toxic to bacteria. There is a special BL21-pLYS-E host strain for expression of particularly toxic gene products that contains a T7 lysozyme gene under the control of the ampicillin promoter which acts as a natural inhibitor of the T7 RNA polymerase (Rosenberg et al., 1987), but this expression strain was never tested.

IMAC purification of the recombinant product from bacterial lysate was not as efficient as from *Pichia* supernatant. Several copurifying proteins needed to be removed by cation exchange and hydrophobic affinity chromatography in order to obtain essentially pure protein as assayed by SDS gels. Interestingly, both proteins elute at approximately the same salt concentration from the Mono-S ion exchange column (0.67 M vs. 0.70 M), but the small isoform elutes significantly earlier from the hydrophobic affinity column than the larger isoform. This suggests that the alternative exon is rather hydrophilic (as can be seen by examining the amino acid sequence) and that it is exposed to solvent in the recombinant protein. Since it appears to be on the surface, it is therefore able to interact with other proteins in the cytoplasm and could alter the functions of the intact neogenin protein.

The amino acid sequences determined from the purified recombinant proteins unambiguously identified the purified material as being the correct proteins. Cleavage of the Ig domains produced in *Pichia* appeared to be variable, as the larger 4 Ig domain construct was cleaved just after the leader peptide and includes all residues expected from the DNA insert, while the smaller 2 Ig domain construct begins a full 20 residues downstream from its expected starting location. Proteolytic cleavage of the 4 Ig domain construct also takes place between the first and second Ig domains, but none of these sites

correspond to the beginning of the 2 Ig domain construct. This cleavage results in a frayed amino terminus, which is why three different sequences are found. Culture conditions have been found to alter the proteolytic processing of Ghilanten expressed in *Pichia*, and it is possible that an alternative expression strategy could yield different processing of the proteins (Brankamp et al., 1995). It is also possible that proteases are more able to digest loose structures than compactly folded ones. The beginning of the smaller construct was arbitrarily selected to begin between the first and second Ig domains of the neogenin molecule, and the tightly folded beta sandwich characteristic of Ig domains (Williams and Barclay, 1988) only includes residues starting 20 amino acids from where the coding frame begins. The larger construct starts with native sequence, and therefore probably adopts a tightly folded structure. The proteolysis of the 4 Ig domain construct between the first and second Ig domains also argues that there is a relatively flexible protein sequence within this region of neogenin. This flexibility could also have functional consequences for the *in vivo* ligand binding properties of neogenin.

Gel filtration chromatography provides a second estimation of a proteins molecular weight based upon its migration through a porous gel matrix. The molecular weights predicted by DNA sequences are in fairly good agreement with those determined by SDS-PAGE and gel filtration, with gel filtration giving somewhat higher estimates across all of the proteins.

The expressed proteins were further characterized using circular dichroism. The small isoform of neogenin's intracellular domain showed a generally increasing ellipticity that contained no characteristic peaks. This suggests that the protein contains no defined

secondary structure under the conditions used in the analysis(reviewed in Johnson, 1990). Based on ligand binding data (see Chapter 4), calcium ions may be required for the protein to adopt a defined structure. Calcium was not included in the analysis buffer because it would interfere with data collection in the appropriate wavelength ranges. Both of the recombinant Ig domains have CD spectra characteristic of beta sheets, as expected from their sequence (Johnson, 1990). The melting temperature of the 4 Ig domain construct was significantly higher than the middle 2 Ig domains (57 vs. 45 °C), which may be caused by the additional domains and expected disulfide bonds. The 2 Ig domain construct was also able to essentially refold after being heated to 81 °C.

Here, various domains of neogenin have been produced in two different heterologous protein expression systems, *Pichia pastoris* and *Escherichia coli*. The proteins have been purified to homogeneity, conclusively identified, and biochemically characterized. These recombinant proteins are able to be used in any biochemical or molecular biology experiment, including monoclonal antibody production or a biochemical search for binding partners to neogenin.

References

Barr, K. A., S. A. Hopkins, and K. Sreekrishna. (1992) *Pharm. Eng.* 12, 48-51.

Bebbington, C. R., G. Renner, S. Thomson, D. King, D. Abrams, and G. T. Yarranton. (1992) *Biotechnology* 10, 169-175.

Brankamp. R. G., K. Sreekrishna, P. L. Smith, D. T. Blankenship, and A. D. Cardin. (1995) *Prot. Express. Pur.* 6, 813-820.

Brent, R., and M. Ptashne. (1985) *Cell* 43, 729-736.

Brown, W. C., and J. L. Campbell. (1993) *Gene* 127, 99-103.

Cregg, J. M., T. S. Vedvick, and W. C. Raschke. (1993) *Biotechnology* 11, 905-910.

Davis, S. J., H. A. Ward, M. J. Puklavec, A. C. Willis, A. F. Williams, and A. N. Barclay. (1990) *J. Biol. Chem.* 265, 10410-10418.

Grinna, L. S., and J. F. Tschopp. (1989) *Yeast* 5, 107-115.

Harfst, E., A. P. Johnstone, and S. S. Nussey. (1992) *J. Mol. Endo.* 9, 227-236.

Hochuli, E., and S. Piesecki. (1992) *Methods: A Comp. to Meth. Enzym.* 4, 68-72.

Johnson, W. C. (1990) *Proteins* 7, 205-214.

Kidd, I. M., and V. C. Emery. (1993) *App. Biochem. Biotech.* 42, 137-159.

Kloc, M., B. Reddy, S. Crawford, and L. D. Etkin. (1991) *J. Biol. Chem.* 266, 8206-8212.

Lindner, P., B. Guth, C. Wulfing, C. Krebber, B. Steipe, F. Muller, and A. Pluckthun. (1992) *Methods: A Comp. to Meth. Enzym.* 4, 41-56.

Skerra, A., I. Pfitzinger, and A. Pluckthun. (1991) *Bio-technol.* 9, 273-278.

Studier, F. W., and B. A. Moffatt. (1986) *J. Mol. Bio.* 189, 113-130.

Rosenberg, A. H., B. N. Lade, D. S. Chui, S. W. Lin, J. J. Dunn, and F. W. Studier. (1987) *Gene* 56, 125-135.

Tschopp, J. F., P. F. Brust, J. M. Cregg, C. A. Stillman, and T. R. Gingeras. (1987) *Nuc. Acid. Res.* 15, 3859-3876.

Williams, A. F., and A. N. Barclay. (1988) *Ann. Rev. Immunol.* 6, 381-405.

Ziegler, F. D., F. Maley, and R. B. Trimble. (1988) *J. Biol. Chem.* 263, 6986-6992.

Chapter 3

Production of Monoclonal Antibodies Against the Ig Domains of Neogenin

Introduction and Strategy of Monoclonal Antibody Production

Antibodies are specific proteins of the immune system which react with a particular immunological epitope on a molecule. The immune system is capable of producing a huge number of antibodies that recognize different epitopes, with estimates ranging as high as 10^8 different possibilities (Harlow and Lane, 1982). This diversity is generated on the DNA level of the individual cell, where recombination events generate novel transcripts through DNA splicing that are not present in germline cells (Dreyer and Bennett, 1965). In nature, this amazing variety of protein production is designed to enable the body to recognize and tag foreign proteins and mark them for destruction.

Monoclonal antibodies are produced by cell lines derived from a single antibody producing cell that has been immortalized by fusing it with a B cell line. This population of antibody secreting cells can then be grown up and milligram quantities of specific antibody can be easily produced. In order to generate monoclonals against a particular protein, one typically injects the desired protein into a mouse (using various methods to enhance the immune response, discussed below) and allows the immune system of that mouse to generate B cells that are making antibodies against the protein. B cells from the mouse are then immortalized and cells producing the desired antibody are selected for subcloning (to remove nonproducing cells) and larger scale antibody production.

Monoclonals are one of the most powerful tools in the molecular biologist's arsenal. As specific reagents, they have been used to clone genes (Kayyem et al., 1992a), identify expression patterns of proteins (Kayyem et al., 1992b), inhibit cellular responses (Bovolenta et al., 1996), and purify proteins which interact with their target (Ozawa et al., 1989). Therefore, we wished to generate monoclonals against various Ig domains of neogenin in order to have specific reagents with which to study the molecule and eventually interfere with its function. We were specifically interested in generating IgG antibodies as they typically have the highest specific affinity for antigen and are the last type of antibody typically produced in the immune response. In order to increase the likelihood of generating IgGs, we wished to inject antigen over a long period of time (6 months) so that the immune system could select for antibodies with the highest affinity toward the injected protein. We used an injection protocol that would try to ensure slow release of the antigen (footpad injections as well as emulsification with Freund's Adjuvant), and also might help to increase the size and strength of the immune response.

Several other techniques can increase the specificity and strength of the immune response. By injecting purified protein, we hoped to make antibodies only against the desired protein. Secondly, by coupling the protein to a carrier (such as keyhole limpet hemocyanin) and by attaching a hapten (a small molecule such as biotin which is antigenic by itself and can increase antigenicity on molecules attached to it), we hoped to increase the strength of the immune response. Finally, by using a special mouse strain that carries a selectable marker near the heavy chain locus, we could select for cells that maintained the heavy chain locus which is occasionally lost since cell lines that are not producing antibody

have a selective growth advantage over those that are. By including these various techniques, we hope to maximize our chances of making antibodies against neogenin.

Materials and Methods

Monoclonal Antibody Production

Robertsonian 8.12 mice were injected with 60 μg of expressed neogenin Ig domains 1 thru 4, with a third of the protein (20 μg) biotinylated, one third coupled to KLH, and one third unmodified protein. Initial injections were subcutaneously in the back, tail, and rear footpads. The protein was diluted into 100 λ PBS and emulsified with 1 volume of Freund's Complete Adjuvant (Sigma). Three subsequent booster injections (subcutaneously in the back and tail) were repeated every 6 weeks using 60 μg protein distributed as above and emulsified in Freund's Incomplete Adjuvant. A final boost of 60 μg recombinant protein without emulsification was given three days prior to the fusion and was injected directly into the spleen.

All fusions were performed in the Caltech Hybridoma Facility essentially as described in Taggart and Samloff (1983). Supernatants from positive cells with good growth characteristics were screened on E7.5 chicken embryos. Horizontal sections (10 μm sectioned on a cryostat) containing the retina, optic chiasm, and tectum were used for the screening as described below. Positive clones were expanded and frozen in liquid nitrogen, while several bright positives were subcloned and rescreened.

Immunohistology

E7.5 chick embryos were fixed for 8 hours with gentle shaking in 4% paraformaldehyde, 0.1 M sodium phosphate buffer, pH 7.0, and the embryos were then soaked in 25% sucrose for several days to remove the paraformaldehyde. Heads were then embedded in tissue-tek (Miles), sectioned horizontally (10 μm) on a cryostat for retino-tectal sections containing retina, optic chiasm, and tectum, and were collected on subbed slides. Sections were then rehydrated briefly with PBS, blocked in 10% fetal calf serum (Gibco) in PBS, and then incubated in primary antibody supernatant for 1 hour. Slides were then washed 3 x 5 minutes in PBS and incubated in a 1:200 dilution of flourescein conjugated anti-mouse antibody (Cappel) in PBS + 10% fetal calf serum. Slides were then washed 3 x 5 minutes in PBS, fixed in 4% paraformaldehyde 2 min, mounted in glycerol, and coverslips were sealed over the sections. Slides were then analyzed on a Zeiss flourescent microscope and photographed.

Western Blots

Protein was run on standard polyacrylamide gels according to standard procedures. Gels were then blotted in Towbin Buffer (39 mM glycine, 48 mM tris base, 0.074% SDS, 20% methanol) at 150 mA overnight and blocked with 5% Carnation nonfat dry milk in PBS. The blots were then incubated in primary antibody diluted in milk+PBS for 1 hour then washed 3 x 10 minutes in PBS + 0.1% tween 20. Blots were then incubated in a 1:2000 dilution of AP-conjugated anti-mouse antibody (Amersham) for 1 hour, then washed 3 x 10 minutes in PBS + 0.1% tween 20 followed by 1 x 5 min PBS.

The blot was then briefly rinsed in color development solution (100 mM tris, pH 9.5, 100 mM sodium chloride, 5 mM magnesium chloride) to adjust the pH, and finally protein was visualized using 0.3 mg/mL nitro-blue tetrazolium (NBT) and 0.15 mg/mL 5-bromo-4-chloro- 3-indolyl phosphate (BCIP) in color development solution until good contrast was obtained.

Results

Monoclonal Antibody Production

The Ig domains of neogenin treated as described in materials and methods proved to be an excellent immunogen and the monoclonal antibody fusion was very successful. Of 384 plated wells, 312 (81%) contained growing cells. This averages approximately 1.5 clones per well, based on a poisson distribution which is thought to be a good estimator of clonality during these fusions (Harlow and Lane, 1982). Of the 312 wells with cells growing, 300 of them (96%) contained antibodies which stained in a pattern which resembled neogenin. The neogenin clones were separated arbitrarily into three different categories; strong, medium, and faint staining, based upon their appearance on tissue sections. 36 clones (12% of positive wells) expressed bright neogenin staining. 11 clones (30%) exhibited additional staining in addition to the neogenin pattern. 155 wells (52% of positives) showed medium strength neogenin staining, with 21% of those showing additional staining patterns. The remaining 109 clones (36%) showed weak neogenin staining, and 23 of the clones (21%) showed patterns in addition to neogenin.

Seven strong positive staining wells were chosen for subcloning. Two of the subclones were bright staining neogenin with additional patterns overlapping neogenin. In the subcloning step, these additional patterns separated clonally from neogenin patterns, suggesting that these were multiclones instead of neogenin antibodies recognizing a shared epitope with another molecule. Of the seven chosen for subcloning, five subclonings were successful in obtaining neogenin staining.

Immunohistology

All of the neogenin antibodies reacted in the pattern expected of neogenin on tissue sections. An example of this staining pattern in the retina is shown in figure 1. At E7.5 during chicken development (the stage of the sections), clone 4F6 stains nasal fibers and does not stain temporal axons. Inner plexiform layer axons are approximately equally stained on both sides of the retina, and the outer plexiform layer is starting to develop neogenin reactivity as well. Expression is also strong in the tectum (not shown), particularly around the posterior edges in the outermost layer (stratum opticum).

Western Blots

Several antibodies were also tested on western blots to begin examining the domain specificity of the antibodies. The results of this screening are shown in Figure 2. Antibodies recognize the intact form of neogenin from P0 chick brain lysates (Figure 2, lane 1). All tested positive wells cross reacted with the 4 Ig domain construct used as an

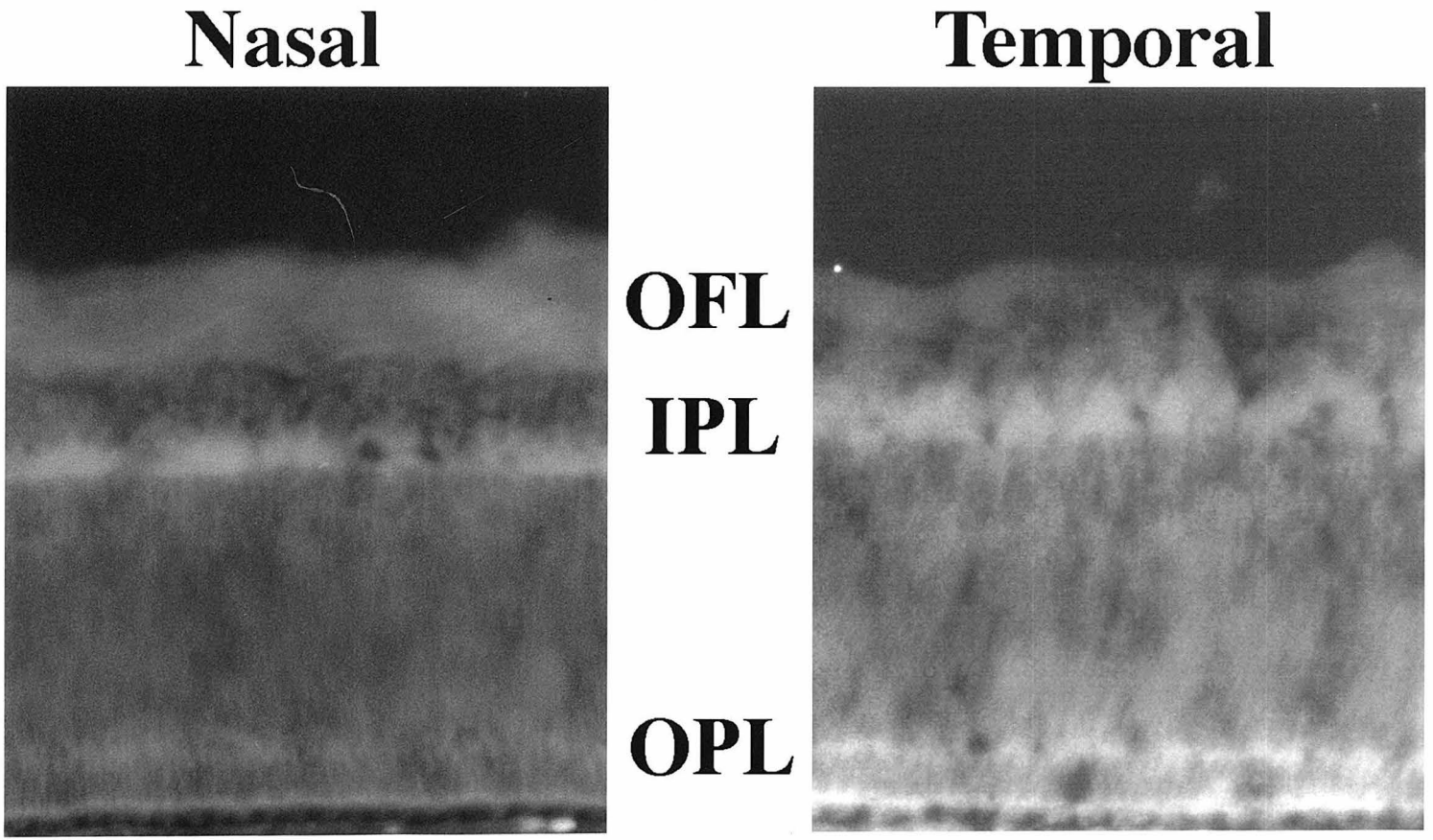


Figure 1: *E 7.5 Retina Tissue Staining* Figure 1 shows the staining of antibody 4F6 on E 7.5 nasal (left) and temporal (right) retina. Staining in the optic fiber layer (OFL) is clearly much more intense on nasal as opposed to temporal retina, which is a distinguishing characteristic of neogenin. Staining in the inner and outer plexiform layers (IPL and OPL, respectively) is slightly more intense on temporal as opposed to nasal retina. Weak staining of various fibers is an artifact due to high antibody concentration.

immunogen, as expected (shown in lanes 3 and 5). The 1G10 antibody, however, does not recognize the smaller second and third Ig domain construct while the 3E7 antibody recognizes both recombinant proteins. This shows that antibodies from this fusion recognize multiple epitopes on the expressed product and that at least some of these epitopes are found in the natural protein. Both 1G10 and 3E7 also recognize the approximately 36 kD fragment that corresponds to the 2nd through 4th Ig domains based on amino terminal sequencing. Because 1G10 does not react with the second and third Ig domain construct, it appears that it reacts with either the 4th Ig domain or the intradomain region between Ig3 and Ig4.

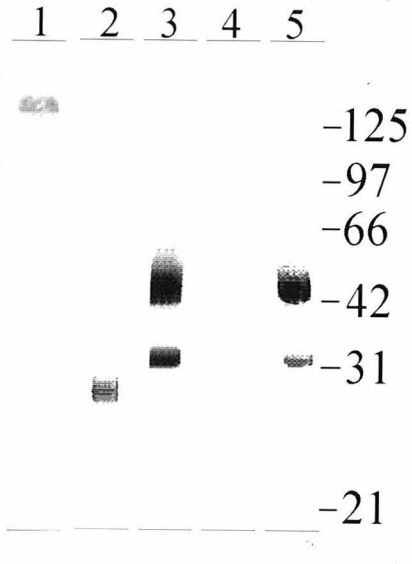


Figure 2: *Western Blot* Figure 2 shows the antigen reactivity of various antibodies on western blots. Lane 1-- 3E7 with P0 total brain lysate. Lane 2-- 3E7 with 2 Ig domain construct. Lane 3-- 3E7 with 4 Ig domain construct. Lane 4-- 1G10 with 2 Ig domain construct. Lane 5-- 1G10 with 4 Ig domain construct. 3E7 recognizes both expressed isoforms and native protein, while 1G10 does not recognize the smaller construct. Thus 3E7 recognizes an epitope in the second or third Ig domain, whereas 1G10 recognizes an epitope in the fourth Ig domain.

Discussion

Monoclonal antibodies have been raised against the 4 Ig domains of neogenin. This fusion was remarkably successful, generating 300 clones that react with native protein on tissue sections. There could be several potential reasons for this success. Immunization was carried out over a six month period with injection of antigen every six weeks during that period, which favors generation of high affinity IgG antibodies. In

addition, a fraction of the protein (1/3 of each injection) contained biotinylated material. Biotin could act as a hapten for generating an increased immune response. Finally, the immunization protocol (injecting protein into tail and footpad) (Kayyem et al., 1992b) and emulsification with Freund's adjuvant were used to ensure slow release of antigen. All three of these factors could have accounted for the apparent strength of the immune response. The high number of clones reactive on tissue sections also argue that the recombinant protein is predominantly native, as the vast majority of clones that contained growing cells react on tissue sections.

The staining pattern of antibodies found in this screen generally show the neogenin staining pattern. Many of the first supernatants show additional staining patterns overlaying that seen with the initial neogenin antibody, 10-22A8. Two strong neogenin antibodies that contained additional tissue staining were subcloned. These were selected due to the strength of these antibody patterns, and the goal was to determine whether the initial well was multi-clonal or a single antibody was recognizing an epitope that neogenin shares with other proteins. Subcloning showed that both of the selected clones were multi-clonal, as neogenin reactivity and additional staining patterns were separable in the subcloning experiment. In other cases, epitopes have been found that are shared between several family members which could be used to isolate new family members (Tonacchera et al., 1995). One potential problem with the fusion is that many of the initial wells show staining patterns in addition to neogenin. If the other staining pattern antibody producing cells have a selective growth advantage, it is possible that they could outcompete the

anti-neogenin producing cells and make subcloning much more difficult. If the inital plating had been done at a lower density, fewer multiclones would have been found.

These antibodies are also reactive on western blots. As shown in figure 2, antibodies raised in this screen react with a protein of the appropriate molecular weight from a brain lysate known to contain neogenin. All of the tested antibodies are also reactive with the 4 Ig domains used as antigen to raise the initial antibodies. Only some of the antibodies are reactive with the other expressed recombinant Ig construct containing the second and third Ig domains. This says that the antibodies recognize at least two (and probably many more) distinct epitopes from the Ig domain construct.

Monoclonal antibodies are one of the most powerful reagents available to study the function of a particular protein domain. Inhibitory antibodies have been used against several different molecules in the retino-tectal system, including NT-3 and N-Cadherin (Bovolenta et al., 1996; J. Sanes, personal communication). Monoclonal antibodies against neogenin may be particularly important due to neogenin's high sequence homology to the human neogenin homologue and similarity to the tumor supressor DCC. These antibodies may also be useful for studying neogenin in other species, as the high sequence identity between cloned neogenin homologues suggests that they should share a considerable number of epitopes.

References

Bovolenta, P., J.-M. Frade, E. Marti, M.-A. Rodriquez-Pena, Y.-A. Barde, and A. Rodriquez-Tebar. (1996) *J. Neurosci.* 16, 4402-4410.

Dreyer, W. J. and J. C. Bennett. (1965) *Proc. Natl. Acad. Sci. USA* 54, 864-869.

Harlow, E., and D. Lane. 1982. *Antibodies: A Laboratory Manual* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Kayyem, J. F., J. M. Roman, E. de la Rosa, U. Schwarz, and W. J. Dreyer. (1992a) *J. Cell Biol.* 118, 1259-1270.

Kayyem, J. F., J. M. Roman, Y. Vonboxberg, U. Schwarz, and W. J. Dreyer. (1992) *Eur. J. Biochem.* 208, 1-8.

Ozawa, M., H. Baribault, and R. Kemler. (1989) *EMBO J.* 8, 1711-1717.

Taggart, R. T., and I. M. Samloff. (1983) *Science* 219, 1228-1230.

Tonacchera, M., F. Cetani, S. Costagliola, L. Alcalde, R. Uibo, G. Vassart, and M. Ludgate. (1995) *Eur. J. Endo.* 132, 53-61.

Chapter 4

A Search for Protein Ligands to Neogenin

Introduction to Ligand Binding Studies

A number of different approaches have been used to gain insights into ligand binding and have been extensively reviewed in Phizicky and Fields (Phizicky and Fields, 1995). The general biochemical strategy is to expose the protein of interest to potential binding partners either individually or en mass and either mark or separate out those that do show an interaction. Some methods such as BiaCore (Panayotou et al., 1993; Schuster et al., 1993) and cell or covasphere binding experiments (Mauro et al., 1992; Brady-Kalnay et al., 1993) require essentially pure receptor and co-receptor in order to obtain an answer. With the Biacore device, however, quantitation of the binding constant is relatively straightforward and is becoming the method of choice for these types of studies. Since we do not know the nature of any proteins which bind to neogenin, we are restricted to using methods to examine proteins en masse. There are four major types of affinity approaches we considered to identifying proteins which interact with neogenin: affinity chromatography, antibody co-precipitation, expression library screening, and the yeast two-hybrid system.

Affinity chromatography is a technique based on coupling a ligand or binding partner onto a solid support (typically an activated agarose such as Pharmacia's Sepharose or Biorad's Affigel-10) and passing a protein solution through the column. Material that interacts with the covalently coupled group is retained or retarded on the column while

non-interacting proteins pass through with little or no retardation. The column is then washed to reduce non-specifically bound proteins and finally bound ligands are eluted from the column. Perhaps the most common affinity approach is immuno-precipitation, where an antibody is coupled to the solid support to separate out the antigen. Affinity chromatography is not limited to antibodies, however; proteins interacting with molecules such as RNA polymerase (Greenblatt and Li, 1981) or c-src (Weng et al., 1993) have identified NusA and paxillin respectively and have been analyzed in similar fashions for at least 15 years. In addition, the extracellular matrix molecule tenascin was coupled to an affinity support and specifically purified contactin from cells that bound to tenascin (Zisch et al., 1992), and an ankyrin affinity column was used to identify neurofascin as an ankyrin binding protein (Davis et al., 1993). One drawback to affinity chromatography is that it requires a large amount of protein to couple on the column. Typically columns are coupled with at least one milligram of protein per mL of solid support. For this approach to be successful, then, requires a good source of the desired ligand in order to obtain sufficient quantities for analysis.

Antibody co-precipitation is a technique related to affinity chromatography. An antibody against the protein for which binding partners are desired is coupled to a support as above. Protein lysate is flowed through the column and the antigen sticks to the antibody, and proteins that normally bind to the antigen hopefully remain bound under the chromatography conditions used. When the antigen is eluted, it and any proteins binding to it are then visualized in the eluted fractions. This method has been used in several different instances, including ras interaction with raf (Warne et al., 1993), the viral protein

E1A interacting with the tumor supressor Rb (Whyte et al., 1988), the interaction of the CD9 antigen with various integrins (M. Hadjiargyrou and P. Patterson, personal communication), and the purification of the three protein catenin complex from E-cadherin (Ozawa et al., 1989).

Protein expression library screening is another means of identifying interacting proteins. The general strategy is to plate a phage library which expresses a random cDNA insert from a tissue where a ligand is expected, and lift nitrocellulose filters with proteins from the lysed bacteria. A protein probe (which may be directly labelled or may be recognized by a labelled secondary probe) is then incubated with the filters, washed, and then detected using an appropriate technique. Expression library screening is commonly done using antibodies, and the proteins Bravo/NrCAM (Kayyem et al., 1992) and neogenin (Vielmetter et al., 1994) were isolated using this method in our laboratory. Researchers have also used labelled DNA as a probe to finding particular DNA binding proteins (Singh et al., 1989), as well as labelled Max and calmodulin to identify Mad (Ayer et al., 1993) and CAM kinase II (Sikela and Hahn, 1987), respectively.

The yeast two hybrid system (Fields and Song, 1989) is the newest of these four methods for identifying ligands. It utilizes the modular nature of some transcriptional activator genes which contain a DNA binding motif and a transcriptional activator region. A reporter gene is placed under the control of a given promoter. Any interaction which localizes the transcriptional activator domain in this region of the DNA will cause transcription of the reporter gene. The DNA binding domain is then fused in frame to the protein for which a desired ligand is desired, and the activation domain fused to a cDNA

library where a ligand is expected. If the separated DNA binding and transcriptional activator domains interact, then the activator is localized to the promoter region and reporter gene transcription can occur. Typically the gene of interest is cloned in frame with the DNA binding domain of GAL4, although other fusions including LexA of *E. coli* have been used. Typical reporter genes include β galactosidase, his4, or leu2, the last two acting as selectable markers which allow colonies to survive only if the gene is transcribed. This method is becoming increasingly popular and has been used to identify a number of different interactions, including the human guanine nucleotide exchanger Sos1 with the adaptor protein Grb2 (Chardin et al., 1993), the phosphatase PP1α2 with the tumor supressor protein Rb (Durfee et al., 1993), and the bHLH protein Mxi1 which interacts with Max (Zervos et al., 1993).

Our Experimental Ligand Identification Plan

We decided to pursue affinity chromatography, expression library screening, and antibody co-precipitation experiments to identify ligands. Affinity chromatography offers high sensitivity to ligands and affords the opportunity to obtain the bound protein. Expression library screening, on the other hand, generally requires higher affinity but yields a DNA clone which is more tractable to experimental analysis. Finally, antibody co-precipitation uses both ligand and interacting protein present in their native state of folding and post-translational modification which may be required for binding to occur. By pursuing these three methods in parallel, we hoped to obtain information about the extracellular and intracellular signalling properties of neogenin.

Materials and Methods

Protein Iodination

^{125}I was purchased from I.C.N. and used while still fresh. 2 iodobeads (Pierce) were washed in iodination buffer (0.1 M sodium phosphate buffer, pH 6.5) and blotted dry on a kim-wipe. 2 mCi of $Na^{125}I$ was added to 100 λ iodination buffer and reacted at room temperature for 5 minutes. 250 μg purified protein to be iodinated was then added to the buffer and reacted for 10 minutes at room temperature. The iodobeads were then removed from the reaction vessel to quench the reaction. Free ^{125}I was removed from the protein using a 2 mL Sephacryl S-200 column for gel filtration and collecting the void volume. ^{125}I incorporation was measured by blotting dilutions of iodinated protein onto PVDF filters, washing 4 times in TGI buffer (25 mM tris base titrated to pH 8.3 with 192 mM glycine + 20% methanol + 10 mM sodium iodide) and measuring γ radiation in a scintillation counter.

Affinity Blotting

Protein samples were separated on 8% and 12% SDS-PAGE gels and blotted overnight at 150 mA onto nitrocellulose in Towbin buffer. Filters were blocked in 5% Carnation nonfat dry milk for at least 30 minutes, and then incubated with $1x10^6$ cpm/mL of expressed protein iodinated with ^{125}I. Proteins were incubated overnight and washed 5 x 10 minutes in PBS + 0.1% tween-20, covered in plastic wrap, and exposed for

autoradiography using flashed Kodak XAR-1 film with intensifying screen for 20 to 60 hours at -70 °C. For determination of the influence of calcium on protein binding, blocking and incubation buffers contained 1% BSA in PBS supplemented with either 10 mM calcium chloride or 10 mM EDTA, and wash buffers contained PBS + 0.1% tween-20 supplemented with calcium chloride or EDTA as above.

Neogenin Coprecipitation

10-22A8 ascites were coupled to affigel-10 (Biorad) at 3 mg/mL and unreacted sites were blocked in 50 mM triethanolamine overnight. 200 mL P0 brain lysate (2.5% NP-40, 2.5% zwittergent, 140 mM sodium chloride, 10 mM HEPES, 4 mM EDTA, 0.02% sodium azide) containing 1 brain per 3 mL was passed over the antibody coupled column at 1 mL/min, and the column was washed in 15 mL AEB-1 (0.5% NP-40, 0.5% zwittergent, 140 mM sodium chloride, 20 mM tris, pH 8.0) and 1 mL fractions were collected. The column was then washed with 15 mL AEB-2 (0.5% NP-40, 0.5 M sodium chloride, 50 mM tris, pH 8.0) collecting 15 x 1 mL fractions, and 15 mL AEB-3 (0.1% NP-40, 0.5 M sodium chloride, 50 mM tris, pH 8.8) collecting 15 x 1 mL fractions. The neogenin protein was then eluted with AEB-4 (0.1% NP-40, 150 mM sodium chloride, 50 mM triethanolamine, pH 11.5) collecting 15 x 1 mL fractions.

Fractions were analyzed on 10% polyacrylamide gels and silver stained according to standard procedures. To identify fractions that specifically contained neogenin, fractions were also analyzed by western blots using 10-22A8 ascites diluted 1:2000 as the primary antibody.

Affinity Chromatography

Affigel-10 (Biorad) was prepared according to manufacturer's instructions. 1 mg purified protein (either 4 Ig domain construct, small intracellular isoform, or large intracellular isoform depending upon the experiment) was coupled to 1 mL affigel-10 at 4 °C for 6 hours and unreacted sites were blocked in 50 mM ethanolamine overnight. P0 or E8 brain lysate (1 brain/mL in same buffer as P0 lysate) was then passed over the protein column at 1 mL/min, followed by a 15 mL wash with AEB-1 collecting 15x 1 mL fractions. Proteins were then eluted by collecting 15 x 0.5 mL fractions of AEB-2, 15 x 0.5 mL fractions of AEB-3, and 15 x 0.5 mL fractions of AEB-4. Fractions were then analyzed on 10 or 12% polyacrylamide gels and silver stained according to standard protocols.

IMAC Chromatography

20 mL P0 brain lysate was mixed with 100 μg of either small or large expressed isoform of neogenin and incubated on a rotary shaker for 3 hours at 4 °C. The material was then passed over a 0.5 mL IDA-agarose (Pharmacia) column charged with Ni^{2+}. Protein was then washed from the column using 10 volumes of IMAC running buffer (1 M sodium chloride, 50 mM sodium phosphate, pH 8.0). The column was then washed with running buffer + 10 mM imidazole (collecting 3 fractions of 1 column volume each), running buffer + 50 mM imidazole (3 fractions of 1 column volume), running buffer + 200 mM imidazole (3 fractions of 1 column volume), and running buffer + 50 mM EDTA (3

fractions of 1 column volume). Fractions were then separated on an 8% SDS-PAGE gel and blotted onto nitrocellulose filters. Filters were blocked for 1 hour in 5% milk in PBS and then incubated in 5% milk PBS + 1.5 x 10^6 cpm/mL mixed labelled small and large isoforms of neogenin's intracellular domain overnight. Filters were then washed 5 x in PBS + 0.1% tween 20, exposed to pre-flashed Kodak XAR film with intensifying screen for 48 hours at -70 °C, and developed.

Library Affinity Screening

A fresh XL-1 bacterial culture was grown overnight in LB broth supplemented with tetracycline + 10 mM magnesium sulfate. 50 λ of the culture was then used to innoculate a 50 mL LB/tet + 10 mM magnesium sulfate culture that was grown to $O.D._{600}$= 0.6 at 37 °C. Bacteria were then spun down and washed twice in 10 mM magnesium sulfate and resuspended to $O.D._{600}$= 0.5.

24 NZY plates (15 cm.) were preheated to 42 °C for 2 hours. 5 x 10^4 phage (1.2 x 10^6 total number of phage screened) from a freshly titered chick brain library were then added to 0.6 mL of prepared XL-1 and incubated 15 minutes at 37 °C without shaking. While plates were incubated, top agar (LB broth + 1.5% agarose) was melted in a microwave and aliquotted to 8 mL per large plate and stored at 50 °C in a heating block until used. The infected bacteria were then quickly mixed with the top agar and spread evenly over the warm plate. The top agar was allowed to harden for 45 min. at room temperature and then transferred to 42 °C for 3.5 hrs.

One hour before the top agar was done incubating, nitrocellulose filters were carefully treated with 10 mM IPTG and air dried on blotting paper. Filters were then applied to the incubated plates and the orientation of the filters were marked. Plates were then incubated for 3 hrs at 37 °C and the first lift of filters was removed. The nitrocellulose filters for the second lift were treated with 10 mM IPTG 1 hour before the first lift was completed, and then they were carefully placed on the plates after the first filters were removed. After the filters were removed from the plates, they were immediately rinsed once in PBS + 0.1% tween 20 to remove sticky top agar and then blocked with PBS + 5% nonfat dry milk for 1 hour at 4 °C. Filters were then incubated in 10 mL of PBS + 1% nonfat dried milk + 1 x 10^6 counts of labelled protein per mL for 5 hours or overnight at 4 °C with constant shaking. Filters were then washed 3 x 15 mL PBS + 0.1% tween 20 for 10 minutes. Filters were then incubated for 24 - 48 hours using either flashed Kodak XAR-1 film and intensifying screen at -70 °C or Molecular Dynamics phosphorimager plates at room temperature to generate autorads.

Results

Affinity Blotting

Iodinated probes were incubated with P0 brain lysates and the results are shown in Figure 1. No protein bands are visible when incubated with labelled neogenin Ig domains I -IV (data not shown); however, two prominent bands at approximately 110 and 140 kD are brightly labelled using ^{125}I conjugated intracellular domain, and a weaker band at

approximately 200 kD is also labelled. Interestingly, the iodinated intracellular domain also sticks to the prestained molecular weight markers. Iodinated probe binding is fairly specific, as the labelled intracellular domain does not stain BSA, gamma globulin, or unstained molecular weight markers even in quantities up to 1 μg per band, nor are prominent bands in the P0 lysate labelled (data not shown).

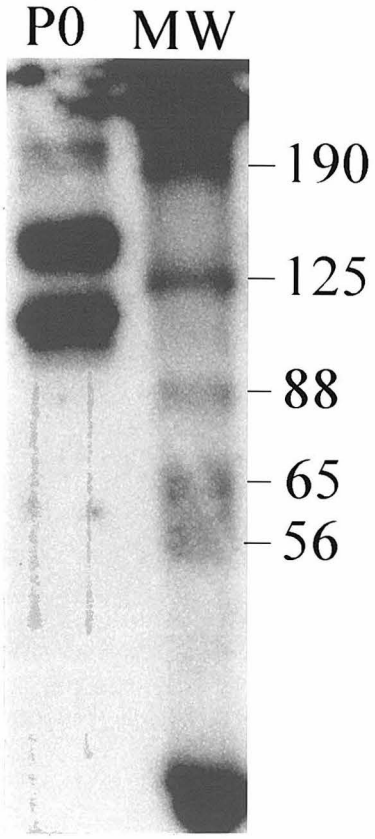


Figure 1: *Affinity Blot* Figure 1 shows an affinity blot using both isoforms of ^{125}I labelled intracellular domains. P0 is a brain lysate from P0 chickens. Three clear bands at approximately 200, 140 and 110 kD are visible. There is also a band at approximately 60 kD that is variable between different blots.

Figure 2 shows the same experiment where the iodinated proteins are incubated separately with and without the presence of divalent calcium ions (that were present in the initial experiments in the milk). Both isoforms of neogenin's intracellular domains interact with the same subset of proteins only in the presence of calcium. EDTA added to the incubation buffer prohibits binding. These protein bands appear slightly different than those found in other experiments, although the general features are the same. The less

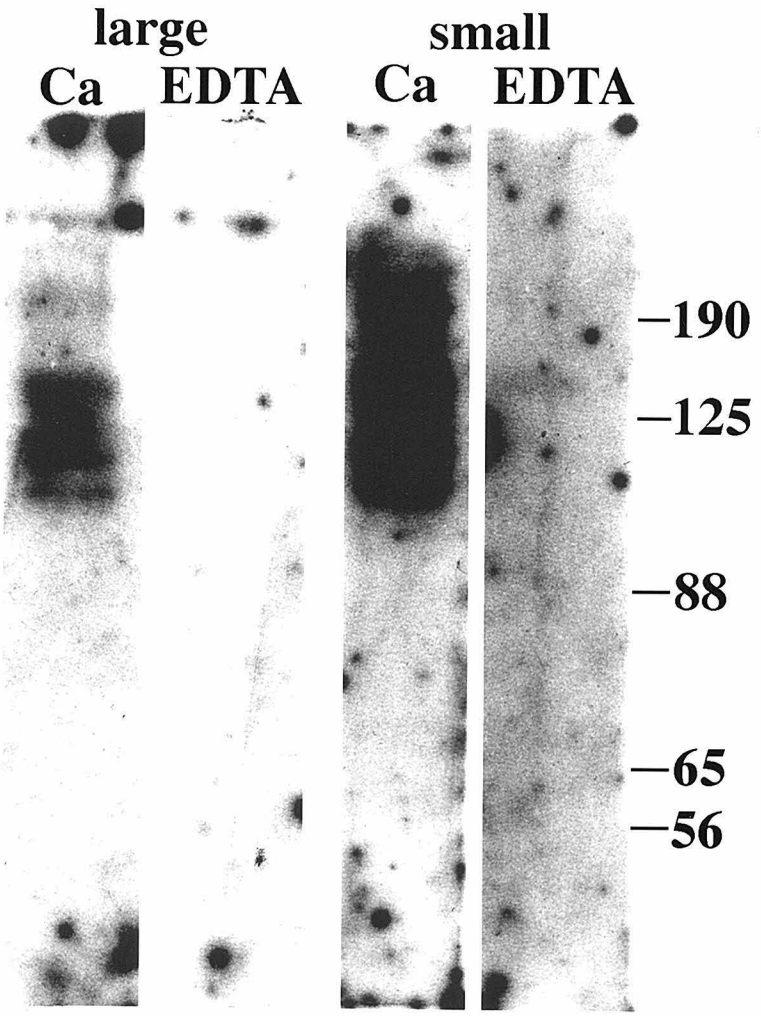


Figure 2: *Calcium Dependence of Affinity Blots* Figure 2 shows the calcium dependence of the affinity blotting bands. Both the large and small isoforms require calcium to bind the proteins, as addition of EDTA inhibits binding. The small isoform blot probably appears darker because more protein was used in the incubation buffer, although it is possible that the small domain has a higher affinity for ligands.

intense 200 kD band appears similar, but the 110 kD band appears to consist of what may be a doublet band.

Affinity Chromatography and Antibody Coprecipitation

Figure 3 shows the results of antibody co-precipitation and Figure 4 shows the affinity chromatography experiment using the small isoform of neogenin's intracellular domain (the large isoform results are virtually identical, data not shown). The 10-22A8 antibody originally used to isolate neogenin clearly gives a doublet of 190 kD and a proteolytic band at 160 kD, as expected (Vielmetter et al., 1994). No bands are visible in the 110 kD or 140 kD range. Affinity chromatography on the expressed small isoform of only neogenin's intracellular domain gives several bands of fairly low molecular weight, and unfortunately they appear to be bands that are prominent in the P0 brain lysate.

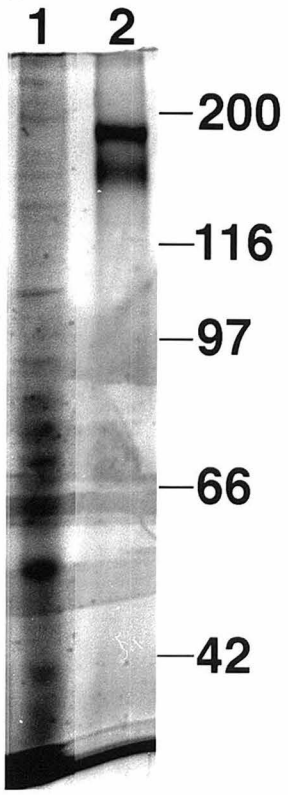


Figure 3: *Antibody Coprecipitation* Figure 3 shows the proteins precipitated using the 10-22 A8 antibody. Lane 1 is P0 brain lysate, and Lane 2 is an elution fraction from the antibody column. No proteins other than the expected neogenin bands at 190 and 160 kD are visible on the silver stained gel.

It is possible that the 200 kD band visible in the affinity blots could be the full length neogenin. In order to test this hypothesis and to see if any of the purified bands

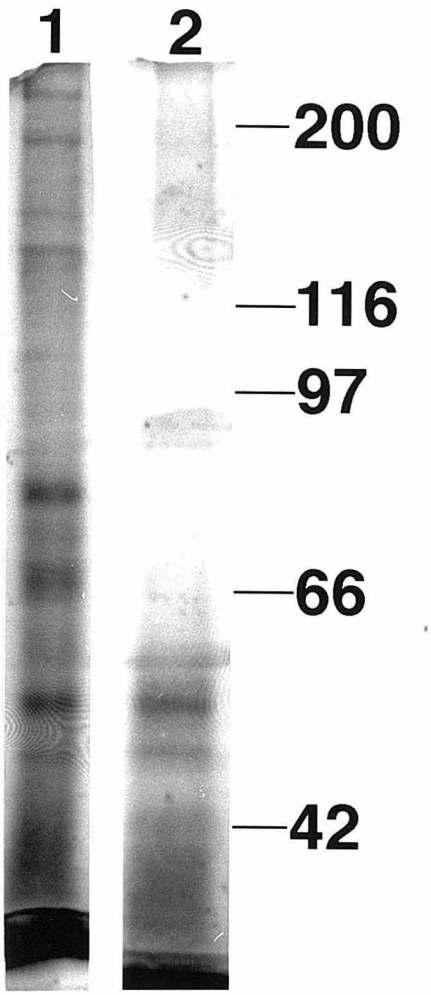


Figure 4: *Affinity Chromatography* Lane 1 shows a silver stained P0 brain lysate and lane 2 is a silver stained elution lane from a small intracellular isoform affinity column. The only proteins visible in the elution lane are also prominent in the P0 lysate lane, suggesting it is non-specific protein binding.

from coprecipitation or affinity chromatography react with the iodinated protein, these lanes were each blotted and incubated with labelled mixed intracellular domain. The results, shown in figure 5, clearly show several important facts. First, immunopurified neogenin does not bind to labelled intracellular domain, nor does the intracellular domain appear to dimerize to either isoform. Secondly, neither antibody coprecipitation nor affinity chromatography enrich for the 110 kD or 140 kD bands. Finally, none of the bands visible in the affinity chromatography lane appear to bind to the labelled probe.

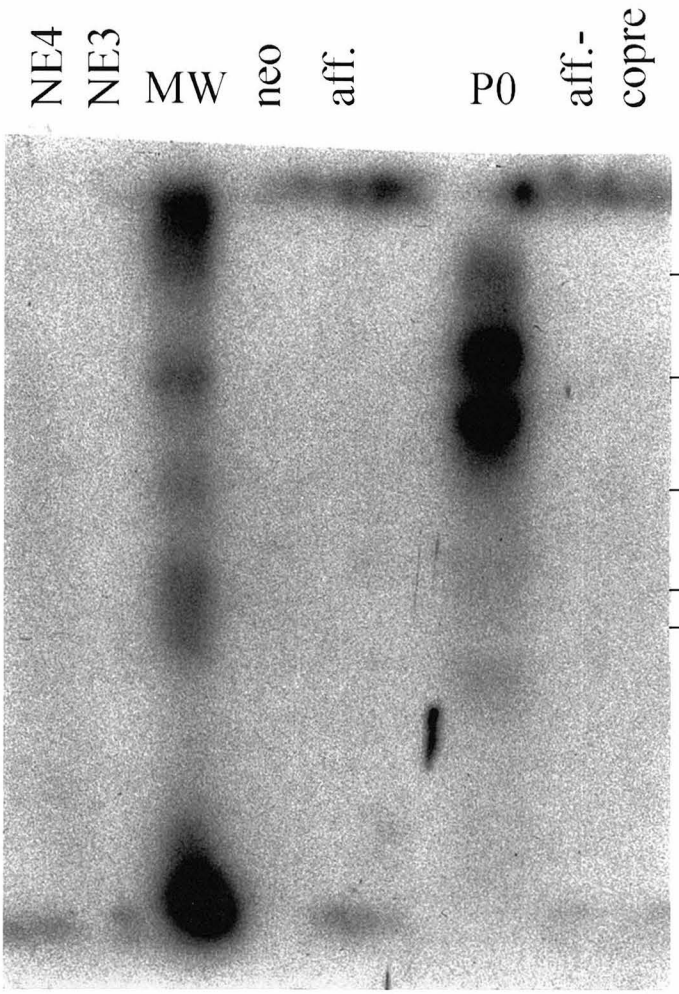


Figure 5: *Affinity Blot* Figure 5 shows an affinity blot, from left to right, of small isoform (NE4), large intracellular isoform (NE3), immunopurified neogenin, elution fraction from affinity chromatography (aff.), P0 brain lysate, a negative control affinity column (1 mg/mL γ globulin), and an elution fraction from antibody coprecipitation. Only the P0 brain lysate lane shows proteins that react with the labelled intracellular domain.

IMAC Affinity Chromatography

Covalently coupling the expressed proteins to a solid support through free amino groups could interfere with the active binding sites of the proteins. In order to test for this possibility, free domain was mixed with P0 brain lysate and purified using IMAC chromatography and included histidine tail. The results of the experiment are shown in Figure 6. Based on the presence of all previously identified reactive bands in the flow through fractions, it appears that there are no proteins which copurify with the expressed domains. The wash fraction of the small isoform purification contains a protein of

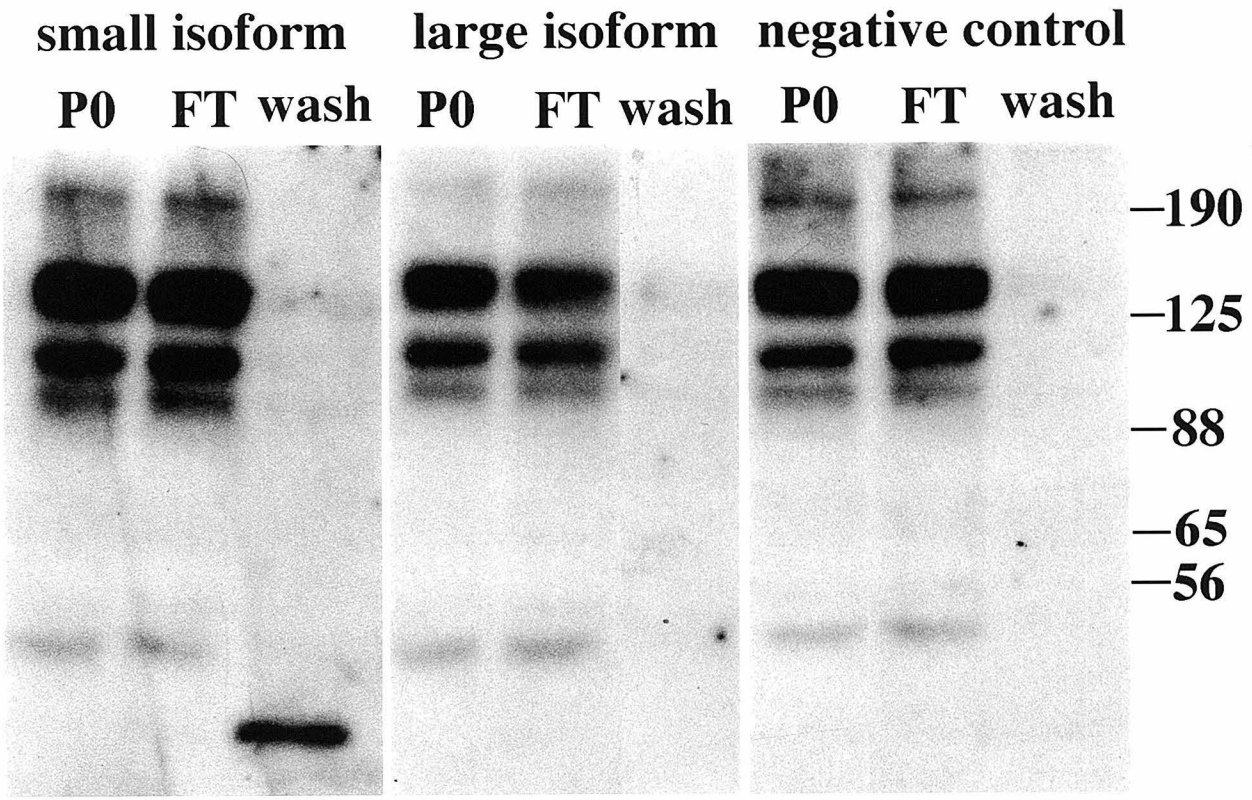


Figure 6: *IMAC Affinity Chromatography* Figure 6 shows the purification of bands reactive with labelled intracellular domain over a Ni^{2+} charged IMAC column with either small isoform, large isoform, or nothing added to the supernatant. The reactive bands are found entirely within the flow through material.

approximately 45 kD; however, this protein is most likely an artifact because it is not seen in P0 lysates and the wash fractions would not allow sufficient concentration of any ligand (a maximum of a factor of 4) to become visible only in that fraction and never seen in the P0 lysate lanes.

Antibody Co-precipitation and Affinity Chromatography with Calcium

Since P0 brain lysate was made in a buffer containing EDTA (see materials and methods), it is possible that the reason no binding protiens were identified using co-precipitation or affinity chromatography is that there was simply no calcium (and hence no binding) in the P0 lysate. Thus both experiments were repeated in the presence of 10

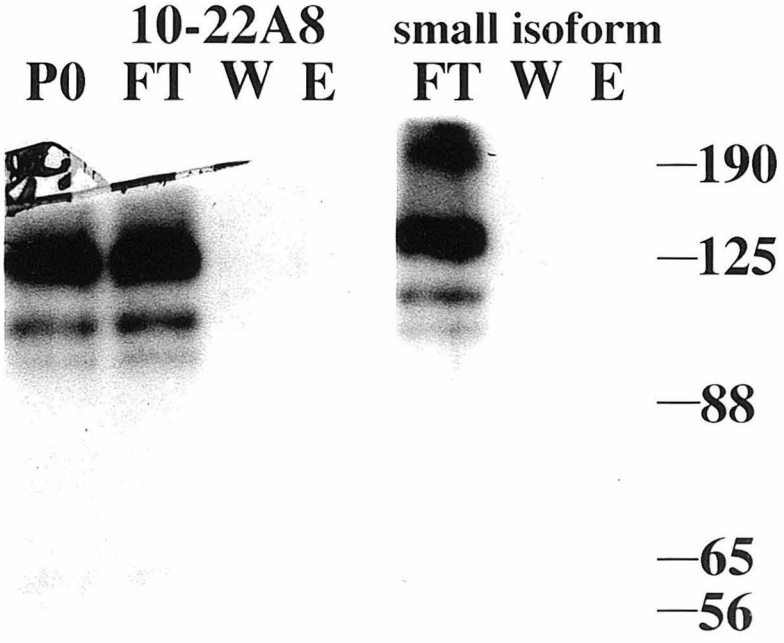


Figure 7: *Calcium Autorad* Figure 7 shows the labelled intracellular domain autorad on antibody precipitated material and affinity chromatography. P0 is a control brain lysate, FT stands for column flow-through, W stands for wash, and E for elution.

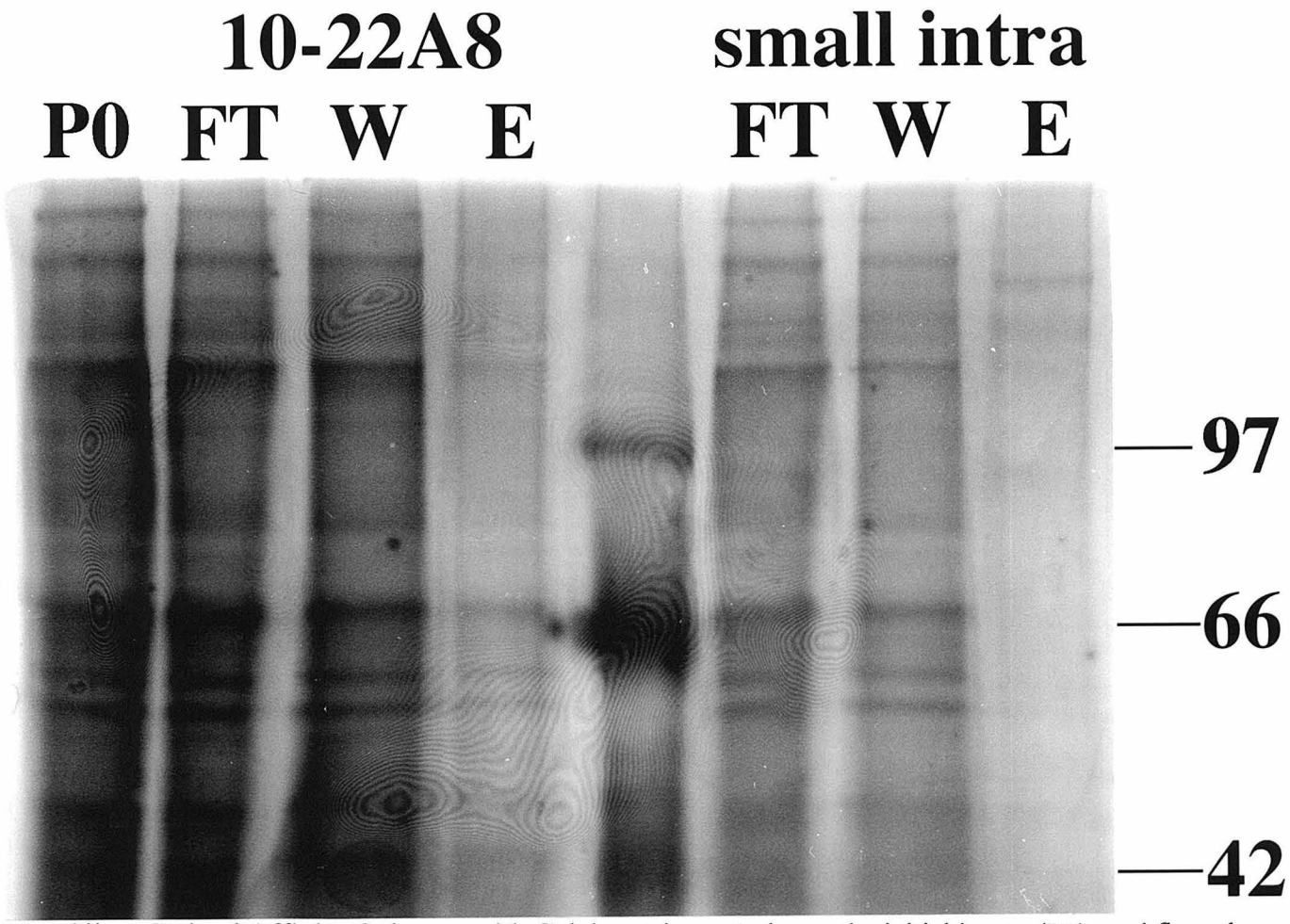


Figure 8: *Silver Stained Affinity Columns with Calcium* Figure 8 shows the initial lysate (P0), and flow through (FT), wash (W), and elution (E) fractions from an antibody co-precipitation experiment using 10-22A8 or affinity chromatography using the small isoform of neogenin's intracellular domain. No protein appears to be specifically retained in the affinity columns.

mM calcium and the results presented in Figure 7 (autorad) and Figure 8 (silver stain). These results are essentially the same as that found without calcium ions added.

Affinity Library Screening

As a final attempt to identify ligands to the expressed portions of neogenin (both extracellular Ig domains and intracellular domains), an E17 cerebellum library was screened using Ig domains and mixed intracellular domains. 1.2 million plaques (approximately 3x redundancy) were screened for each construct. Both Ig domains and intracellular domain filters had an extremely high background of spots due to apparently non-specific stickiness, but only four spots on the intracellular domain filters and zero spots on the Ig domain filters overlapped on the double lifts. The four intracellular domain spots were then picked and analyzed again, this time coming up negative.

Discussion

Based on the affinity blotting experiments, there appears to be perhaps four proteins that interact with the iodinated intracellular domain of neogenin. p200 and p50 are relatively faint bands, and p50 is particularly inconsistent as it appears in some experiments and not others. This variability does not appear to be based on either isoform dependence or calcium, but may be based on proteolysis or blotting conditions which were not adequately controlled. The main two bands, p110 and p140, are consistent and signficantly brighter than the two less intense bands. It is possible that the lower band may actually be a dimer, as in some experiments this band is resolved into two bands of slightly

different molecular weight. It is also possible that this is a proteolytic effect, as the doublet generally appears in experiments which take a greater length of time to complete. All four bands appear to be fairly specific, as prominent protein bands in the P0 brain lysate do not interact with the iodinated proteins but less prominent bands appear highly labelled.

Calcium ions appear critical for any binding interactions of the neogenin intracellular domain, as addition of EDTA to the incubation buffer inhibited binding to all bands equally. Buffer conditions therefore play a significant role in whether ligands can be viewed on affinity blots, and this is even more likely to be true in experiments designed to purify the binding proteins. The ligands also do not appear to be isoform specific, as both the large and the small expressed isoforms label the same subset of proteins on parallel blots. This argues that the alternative exon does not play a role in any of these binding interactions.

The calcium dependence of the interaction is a striking feature of the intracellular domain and could suggest a possible function for the intact neogenin protein. Presumably neogenin is activated by an extracellular binding event which allows the intracellular domain of neogenin to transmit a signal. The calcium requirement to bind to the identified bands suggests that a second signal, elevation of intracellular calcium, is also required for neogenin to interact with these proteins. The calcium ion requirement may be either for the expressed neogenin protein or the protein on the blot. Neogenin could thus serve as an integrator of two different signals that the cell receives. Neogenin would signal only if the extracellular domain bound an appropriate ligand, and a second event increased the

local concentration of calcium in the vicinity of the neogenin protein. This calcium could come from either intracellular stores or be obtained from outside the cell via a calcium selective ion channel (reviewed in Katz, 1996).

The protein most closely related to neogenin, DCC, has been shown to bind to netrin-1 in cell aggregation experiments (Marc Tessier-Lavigne, personal communication), and netrin mutants in *drosophila* have the same phenotype as a mutant called *frazzled*, yet another neogenin family member (Peter Kolodziej, personal communication). Unc-40 and unc-6, *C. elegans* homologues of neogenin and netrin, respecively, have also been shown to interact genetically in the control of axon outgrowth and guidance (Hedgecock et al., 1990). Therefore, it does not seem unreasonable to suggest that neogenin interacts with a protein similar to netrin, either a family member or perhaps a β2 laminin subunit which is related to netrin (Serafini et al., 1994). Experiments here to identify extracellular proteins that bind to the expressed Ig domains of neogenin failed to find any binding partners. While unfortunate, binding experiments in either high detergent buffers used to solubilize membrane proteins or bacterial lysates under expression screening conditions which may not allow proper refolding of heterologous proteins (especially secreted domains) do not rule out the possibility of finding proteins that interact with the extracellular domains of neogenin. Other conditions which may more closely resemble those in the cellular environment may allow such an interaction to be identified. It is also possible that neogenin requires a second subunit to bind its ligand, as is known for integrins (Reichart and Tomaselli, 1991).

As a putative tumor suppressor gene on the cell surface, signalling functions of DCC are of profound interest due to their role in controlling cell growth. Neogenin is the only protein currently in the database that has an intracellular domain related to that of DCC. Identification of a subset of proteins ranging from 50 to 200 kD that are able to bind to neogenin suggests that similiar or the same proteins may also interact with DCC. The eventual identification of these ligands may someday shed light on the signal transduction methods of neogenin and DCC and perhaps lead to a better insight as to how a cell surface molecule functions in resisting tumor formation.

References

Ayer, D. E., L. Kretzner, and R. N. Eisenman. (1993) *Cell* 72, 211-222.

Brady-Kalnay, S. M., A. J. Flint, and N. K. Tonks. (1993) *J. Cell Biol.* 122, 961-972.

Chardin, P., J. H. Camonis, N. W. Gale, L. V. Aelst, J. Schlessinger, M. H. Wigler, and D. Bar-Sagi. (1993) *Science* 260, 1338-1343.

Davis, J. Q., T. McLaughlin, and V. Bennett. (1993) *J. Cell Biol.* 121, 121-133.

Durfee, T., K. Becherer, R.- L. Chen, S. H. Yeh, Y. Yang, A. E. Kilburn, W. H. Lee, and S. J. Elledge. (1993) *Genes Dev.* 7, 555-569.

Fields, S., and O.-K. Song. (1989) *Nature* 340, 245-246.

Greenblatt, J., and J. Li. (1981) *J. Mol. Biol.* 147, 11-23.

Hedgecock, E. M., J. G. Colotti, and D. H. Hall. (1990) *Neuron* 2, 61-85.

Kayyem, J. F., J. M. Roman, E. de la Rosa, U. Schwarz, and W. J. Dreyer. (1992) *J. Cell Biol.* 118, 1259-1270.

Katz, A. M. (1996) *J. Am. Col. C.* 28, 522-529.

Mauro, V. P., L. A. Krushel, B. A. Cunningham, and G. M. Edelman. (1992) *J. Cell Biol.* 119, 191-200.

Ozawa, M., H. Baribault, and R. Kemler. (1989) *EMBO J.* 8, 1711-1717.

Panayotou, G., G. Gish, P. End, O. Truong, I. Gout, R. Dhand, M. J. Fry, I. Hiles, T. Pawson, and M. D. Waterfield. (1993) *Mol. Cell Biol.* 13, 3567-3576.

Phizicky, E M., and S. Fields. (1995) *Microbiol. Rev.* 59, 94-123.

Reichardt, L. F., and K. J. Tomaselli. (1991) *Annu. Rev. Neurosci.* 14, 531-570.

Schuster, S. C., R. V. Swanson, L. A. Alex, R. B. Bourret, and M. I. Simon. (1993) *Nature* 365, 343-346.

Serafini, T., T. E. Kennedy, M. J. Galko, C. Mirzayan, T. M. Jessel, and M. Tessier-Lavigne. (1994) *Cell* 78, 409-424.

Sikela, J. M., and W. E. Hahn. (1987) *Proc. Natl. Acad. Sci. USA* 84, 3038-3042.

Singh, H., R. G. Clerc, and J. H. LeBowitz. (1989) *Biotechniques* 7, 252-261.

Vielmetter, J., J. F. Kayyem, J. M. Roman, and W. J. Dreyer. (1994) *J. Cell Biol.* 127, 2009-2020.

Warne, P. H., P. R. Viciana, and J. Downward. (1993) *Nature* 364, 352-355.

Weng, Z., J. A. Taylor, C. E. Turner, J. S. Brugge, and C. Seidel-Dugan. (1993) *J. Biol. Chem.* 268, 14956-14963.

Whyte, P., K. J. Buchkovich, J. M. Horowitz, S. H. Friend, M. Raybuck, R. A. Weinberg, and E. Harlow. (1988) *Nature* 334, 124-129.

Zervos, A. S., J. Gyuris, and R. Brent. (1993) *Cell* 72, 223-232.

Zisch, A. H., L. D'Alessandri, B. Ranscht, R. Falchetto, K. H. Winterhalter, and L. Vaughan. (1992) *J. Cell Biol.* 119, 203-213.

Chapter 5

Summary and Future Directions

Heterologous protein expression is a very powerful technique for obtaining large quantities of any desired protein of interest. The choice of expression system seems to play a vital role in determining whether protein is produced in any given expression system. The immunoglobulin domains of neogenin and both of its alternatively spliced intracellular domains have been produced in *P. pastoris* and *E. coli* in relatively large yield and a purification method for the expressed constructs have been worked out to give essentially pure protein. These methods should prove useful for others in helping to choose either an expression system for a particular application, development of a purification scheme for proteins produced under similar conditions, or for others interested in the chick protein neogenin. Unfortunately, heterologous protein expression is still very much an experimental science, where the only way to find out if it will work is to test the given protein. It does not necessarily depend upon the type of protein being expressed, as the intracellular domain of Bravo is able to be produced in *Pichia* while the intracellular domain of neogenin cannot. However the immunoglobulin domains of Bravo failed to express in the yeast (R. Lane, personal communication) and the immunoglobulin domains of neogenin expressed well. Perhaps in the future studies can be conducted which will begin to explain why particular proteins cannot be expressed in a given system and other closely related proteins are produced well.

Numerous monoclonal antibodies have been raised against the expressed immunoglobulin domains of neogenin. The antibodies recognize at the very minimum two

different epitopes in the Ig domains of neogenin, and based on the extremely high number of antibodies generated and their various staining intensities, it appears that there are significantly more than two epitopes being recognized. These antibodies are powerful reagents and may be used for many different purposes. Monoclonal antibodies have been used to inhibit the function of particular proteins *in vivo*, as done during the development of the retina with antibodies against NT-3 (Bovolenta et al., 1996) or antibodies against N- cadherin during synaptic refinement in a retinotectal culture system (J. Sanes, personal communication). These monoclonals could be extremely useful for anybody interested in trying to use them to perturb the function of neogenin during development. If neogenin is actually involved in terminal differentiation and/or neurite outgrowth of retinal ganglion cells, perhaps one or more of the generated monoclonal antibodies would produce a phenotype if injected into embryonic chickens. A panel of neogenin antibodies could also be used to inhibit neogenin's function in an *in vitro* assay system, many of which are available for chicken retinas. By using neogenin antibodies to interfere with the expected result in a very specific assay, it may be possible to define a specific role for neogenin in the process of retinotectal development.

The monoclonal antibodies could also be used to study neogenin expression in organisms other than chickens. A human neogenin clone has been sequenced in the Dreyer lab (see appendix A-II), and is mostly identical to the chick protein. Neogenin has also been found in rat (M. Tessier-Lavigne, personal communication) and is also extremely similar to the chick protein. The ability to examine neogenin's expression in other organisms would be invaluable, as other systems can provide advantages which the chick

system does not possess. In addition, because of the relationship of neogenin to DCC, a putative human tumor suppressor molecule, neogenin's function in development and presence or absence in various types of tumors could be very interesting. Many of the currently identified tumor suppressor genes (including DCC) are deleted in multiple types of carcinomas (Knudson, 1993). For some tumors the presence or absence of particular cellular markers, often at the cell surface, have been correlated with the metastatic potential and clinical prognosis (Knudson, 1993). Examining neogenin in various types of tumors as well as its expression normally in development could be extremely useful and clinically relevant. Finally, there is preliminary evidence that a third protein related to neogenin exists (J. Korenberg, personal communication), suggesting that there may be a relatively large family of proteins with a similar structure and probably a similar function.

The expressed intracellular domains of neogenin have been used as a probe to identify proteins that bind to neogenin. When either the small or large isoforms of neogenin's intracellular domain are iodinated with the radioactive tracer ^{125}I and incubated on total protein blots, two prominent protein bands at 110 and 140 kD are labeled, along with a consistent weak 200 kD band and an inconsistent weak 60 kD band. These protein bands require divalent metal in order to see this binding, as addition of EDTA, a divalent metal chelating agent, abolishes binding to all of the proteins on the blot. This could be due to a calcium requirement of neogenin's intracellular domain or a requirement for the proteins on the blot. It is most likely a requirement of the labeled intracellular domain, as it is less likely that all 3 consistent bands on the blot would each independently require calcium to bind to neogenin.

These labeled domains of neogenin now provide a tool by which proteins interacting with neogenin may be followed and eventually purified. One approach is to separate the various proteins in a total brain lysate according to various physical properties, such as size, hydrophobic affinity, and ionic charge at a certain pH. These methods have been used for many years as a means of biochemically purifying any desired protein. However, in order to find the desired protein, an assay of some sort must be used to follow the proteins or activity of interest through the various purification and separation steps that lead eventually to a single or small number of purified proteins. This method is a brute force approach and requires many months of work, but it has proven repeatedly that it will produce results.

The question "What are the identity of proteins which interact with either the immunoglobulin domains or either alternative isoform of the intracellular domain of neogenin?" is probably the most interesting problem of the entire work. This question deals with the core of the molecular mechanism of how neogenin functions in normal development, and could very easily have ramifications for clinical and medical research involving the mechanisms of tumor growth and control of cell division. Attempts to identify interacting proteins using affinity chromatography, expression library screening, and antibody coprecipitation were ineffective for both extracellular and intracellular ligands to recombinant neogenin. However, there are several possible improvements that could be made in these attempts in the future that might circumvent the problems and difficulties of the brute force approach. Affinity purification approaches have proven to be effective against many proteins, particularly using antibodies coupled to a solid support

(such as used in Chapter 4). However, affinity approaches are also highly dependent upon the buffer in which the affinity experiments are carried out. In the case of the immunoglobulin domains, NP-40, Triton- X100, and the zwittergent/NP-40 buffer all proved ineffective. The zwittergent/NP-40 buffer, while excellent for purifying neogenin, was not necessarily a good choice for identifying ligands to the intracellular domain of neogenin.

There does not appear to be a single best buffer for this type of purification. It is an experimental field, where the only way to find a buffer that works is to try various combinations. For the immunoglobulin domains of neogenin, some detergent is probably necessary in the affinity buffer, as most cell surface proteins and many extracellular molecules are not soluble without assistance. Digitonin is a detergent which people have found to be particularly useful when looking for ligands, particularly in the immune system. Other people have used Triton- X100 or NP-40 effectively (which were not effective using the immunoglobulin domains of neogenin).

For the intracellular domains of neogenin, detergents may or may not be necessary to find proteins that interact. If neogenin interacts with soluble proteins in the cytoplasm, then detergents may not be necessary at all. They may definitely be harmful because they could occlude hydrophobic binding surfaces and prevent appropriate binding. If, however, neogenin's intracellular domain associates with a protein that is either a transmembrane molecule or tethered to the inside of the cell membrane via a hydrocarbon chain, detergent may be required to solubilize the lysate. If detergents are required, the choice of detergents is again best determined experimentally. Minimal detergents may be required if

the protein ligand is only weakly associated with the cell membrane. More powerful detergents would be necessary if the intracellular domain of neogenin interacts with the intracellular domain of a transmembrane co-receptor. In addition to the detergent concentration, it would be advisable to add additional calcium to all buffers looking for ligands to the intracellular domain of neogenin based upon the ability of EDTA to abolish labeling.

Ligands for the intracellular domain, however, can be assayed by their ability to bind to iodinated intracellular domains. This can serve as an excellent control for proteins that are specifically interacting with neogenin and exclude those proteins that interact nonspecifically with the support material. Because neogenin is related to the putative tumor suppressor molecule DCC, information about neogenin's mechanism of signal transduction may be useful for studying DCC's mode of action, as the two molecules have a significant amount of sequence similarity and are the two most closely related molecules in the known database.

References

Bovolenta, P., J.-M. Frade, E. Marti, M.-A. Rodriquez-Pena, Y.-A. Barde, and A. Rodriquez-Tebar. (1996) *J. Neurosci.* 16, 4402-4410.

Knudson, A. G. (1993) *Proc. Natl. Acad. Sci. USA* 90, 10914-10921.

Title: Expression and Purification of Protein Domains of Neogenin and Bravo/NrCAM Expressed in *P. Pastoris* and *E. coli* for Production of Monoclonal Antibodies

Authors:

Frank Miskevich, Robert P. Lane, Jost Vielmetter, and William J. Dreyer

Abstract

Cell surface receptors play a critical role in the ability of a cell to respond to its environment. These proteins are made up of individual domains that serve different functions, such as ligand binding or intracellular signaling. In order to generate monoclonal antibodies exclusively against the protein domains of interest, we expressed these domains in either *E. coli* or *Pichia pastoris*. Recombinant protein was then purified using both affinity and biochemical approaches, characterized to determine its properties, and used as an immunogen. Monoclonal antibodies were identified that reacted with the particular domains on both western blots as well as tissue sections. In multi-domain expressed proteins, antibodies against multiple domains were identified. These methods constitute a general approach useful to generate monoclonal antibodies against any protein domain of interest.

Introduction

Monoclonal antibodies are one of the most powerful specific reagents available today for molecular biologists. Although the technology for generating them has been available for many years (Kohler and Milstein, 1975), typically the antibodies have been generated against relatively impure populations of proteins. Many different techniques have been applied to increase the number of specific antibodies, including preselection of B lymphocytes with antigen (Casali et al., 1986), purification of cell surface proteins by biotinylation (Kayyem et al., 1992), and subtractive methods to find proteins different between two populations of cells (Suzue et al., 1990). While these methods have been successful in the past, immunodominant epitopes represent a significant challenge to generating monoclonal antibodies against rare proteins or less antigenic domains.

One way to circumvent these problems is to overexpress a protein or protein domain of interest in a heterologous system. This approach has many advantages over more traditional approaches. First, there is typically a large amount of protein available for immunization, which facilitates generation of monoclonal antibodies. Second, because the recombinant protein can be designed to include an affinity tag, purification can greatly reduce the complexity of the immunogen. Another advantage of using recombinant protein is that the immunogen can be generated directly from DNA sequences. Due to the proliferation of genomic sequencing efforts, the task of analyzing proteins and their functions using specific reagents will become increasingly more important compared to finding new genes. Finally, it is possible to express particular domains of proteins so that antibodies will be directed against a restricted subset of epitopes, which may be important in generating antibodies inhibiting particular functions of a protein.

The T7 protein expression system for *E coli* is one of the oldest and most powerful protein overexpression systems (Studier and Moffat, 1986; Rosenberg et al., 1987; Studier et al., 1990). A plasmid under the control of a T7 RNA polymerase promoter is transfected into a host cell. The host cell can either contain an inducible T7 RNA polymerase gene or be amenable to later infection by a virus carrying a T7 RNA polymerase (Studier and Moffat, 1986). The bacteria then are allowed to express the desired protein for a certain length of time. Because the T7 promoter is extremely strong, in many cases the protein is expressed at hundreds of milligrams per liter, which results in inclusion bodies that facilitates purification.

Pichia pastoris is a methylotrophic yeast that can, under the proper circumstances, generate massive quantities of recombinant protein (Cregg and Madden, 1988). Typically *Pichia* expression is carried out using the methanol inducible AOX1 promoter, a very powerful promoter than can generate up to 50% of the protein found in the yeast. The promoter is tightly regulated, so that there is little production in the absence of methanol and allows for normal growth to increase cell mass before inducing the expression of the protein. In addition, *Pichia* contains a conserved eukaryotic secretory apparatus that can be important for protein folding and greatly facilitates purification by removing the desired protein from abundant intracellular *Pichia* proteins. Many different types of proteins have been successfully produced in high yields using this system (Sreekrishna et al., 1989; Romanos et al., 1991; Siegel et al., 1990; Wagner et al., 1992).

We are interested in two particular cell surface proteins of the chicken visual system, neogenin and Bravo/NrCAM. These proteins are both transmembrane Ig

superfamily molecules (Vielmetter et al., 1994; Grumet et al., 1991), both are expressed dynamically in the retinotectal system during development, and both have well conserved human homologues (Vielmetter et al., submitted; Lane et al., 1996). Bravo/NrCAM is a cell adhesion molecule of the central nervous system and is most closely related to the cell adhesion molecules neurofascin (Volkmer et al., 1992), L1 (Moos et al., 1988), and NgCAM (Burgoon et al., 1991), each consisting of six Ig domains, five fibronectin type III repeats, a single transmembrane region, and an intracellular domain. Bravo/NrCAM's extracellular domain has been shown to bind heterophilically with axonin (Suter et al., 1995) as well as homophilically (Mauro et al., 1992), and its intracellular domain contains a motif that seems to be involved in tethering the protein to the cytoskeleton via ankyrin (Davis et al., 1993). Bravo/NrCAM is expressed on motor neurons and floor plate in the embryonic spinal cord (Moscoso and Sanes, 1995) and may be important in signaling functions associated with a calcium flux (Von Bohlen et al., 1992) and may be associated with a co-receptor on the same cell which contains a kinase domain (Williams et al., 1994).

Neogenin is a cell surface molecule that is dynamically expressed in both the retina and cerebellum during chicken development (Vielmetter et al., 1994). In the early retina, neogenin is expressed equally on both nasal and temporal retinal ganglion cell fibers as they exit to form the optic nerve. As the ganglion cells are finding their topographic targets in the tectum, temporal axons down-regulate the expression of neogenin first and nasal axons down-regulate the molecule slightly later. Neogenin is also strongly expressed in the posterior foliae of the cerebellum in the external granule cell layer in the region

where new cells are born. Interestingly, human medulloblastomas arise from the very posterior portion of the cerebellum approximately where chicken neogenin is expressed. Neogenin is most closely related to the tumor suppressor molecule deleted in colorectal cancer (DCC) that has been implicated in multiple different types of cancer (Cho and Fearon, 1995). Neogenin and DCC form a distinct subfamily of the vertebrate Ig superfamily that contains 4 Ig domains, 6 fibronectin type III repeats, a transmembrane domain, and an intracellular domain unique to this subfamily.

We wished to study individual regions of these molecules that seemed particularly likely to carry out important functions. The Ig domains of neogenin were chosen because these domains are known to mediate a wide variety of ligand binding interactions (Frei et al., 1992; Williams and Barclay, 1988). The fifth fibronectin repeat of Bravo/NrCAM was chosen for closer study because it is alternatively spliced in both chickens and humans. Finally, we also expressed the intracellular regions of both molecules because we are interested in how these cell surface proteins transduce signals across the cell membrane. We have used two of the recombinant expressed extracellular domains to raise monoclonal antibodies that react with native proteins on western blots and tissue sections; thus we have been able to generate antibodies against particular regions of proteins that are likely to be relevant for the functions of the molecules. These specific reagents can then be used to address the functions of the domains *in vitro* and *in vivo*.

Materials and Methods

Pichia Expression Vector Construction

A bacterial clone (NE1) isolated by screening a lambda-ZAP II expression library with the monoclonal antibody 10-22-A8 was used as a template to PCR amplify the Ig domain constructs for subsequent expression. The primers used were as follows:

Neo1.S- 5' TCT AGA GAA TTC GTA GTG AGA ACC TTC ACT C 3'

Neo4.A- 5' GC GGC CGC CCC GGG TTA ATG GTG ATG GTG GTG ATG TCG TGA AGT GGG AGG TAA TGT TG 3'

Neo2.S- 5' GGA TCC TTG TAA GCA GAA CAG 3'

Neo3.A- 5' CTG CAG TTA ATG GTG ATG GTG GTG ATG GCG GCC GCG CTT CAG AAA CTC AGG 3'

bravo-int.S 5'- GAA TTC AGG AGG AAT AAA GGT GGC A -3'

bravo-int.A 5'- GCGG CCGC TTA ATG GTG ATG GTG GTG ATG CAC AAA TGA ACT CAT GGC A-3'

PCR was carried out using Vent polymerase under the following conditions: 92 °C for 5 min., 72 °C 5 min. (enzyme added here--hot start), followed by 25 cycles of 92 °C 0.5 min., 56 °C 1 min., 72 °C 2 min., and then a 72 °C 5 min. extension. The PCR product was analyzed for homogeneity on an agarose gel, and the product was cloned using the TA Cloning Kit (Stratagene) according to the instructions.

Positive bacterial clones were selected by white-blue color selection, double checked by PCR using the conditions above, and plasmid prepared using Wizard minipreps. Plasmids were double digested with Not I and EcoRI to guarantee the

presence of the restriction sites, and the plasmids were sequenced by either manual or automated dye terminator sequencing. Positive clones with no PCR artifacts were then selected for expression.

Yeast expression plasmid pPIC-9K was prepared from XL-1 bacteria using Wizard minipreps and digested sequentially with Not I, followed by EcoRI. The double digested plasmid was gel purified and the DNA was then dephosphorylated with calf intestine alkaline phosphatase to avoid self ligation.

Insert DNA was prepared by digesting TA vector containing the cloned insert with EcoRI followed by Not I and gel purifying the insert. The ligation reactions were carried out using 1:1 and 3:1 molar ratios of insert to vector. The DNA was then spun down, washed with ethanol, dried, and resuspended in distilled water for electroporation in XL-1 bacteria. Positive colonies were selected using ampicillin resistance, and PCR was used to identify clones containing insert.

Pichia Transformation

A single colony of *Pichia pastoris* (strain GS115, a *his4* strain) was picked from an MD (1% glycerol, 400 μg/mL biotin, 13.4 g/L yeast nitrogen base w/o amino acids) plate supplemented with histidine and grown in 25 mL MD-his broth at 30 °C to an $O.D._{600}$= 0.6 to 1.0. The cells were then spun down and washed 3x with 25 mL ice cold distilled water and washed once in 25 mL 1.0 M sorbitol. The cells were finally resuspended in 250λ 1.0 M sorbitol and stored for up to 1 week at 4 °C.

Recombinant pPIC-9K plasmid was prepared and digested using 10 units of Bgl II enzyme at 37 °C for 2 hours, extracted once with an equal volume of equilibrated phenol:chloroform (1:1), and extracted with chloroform. Linearized plasmid was then precipitated and resuspended in 5 λ distilled water for electroporation.

40λ of electrocompetent yeast was placed in a 0.1 cm. electroporation cuvette along with 1λ (1 μg) of linearized plasmid on ice. Electroporation was carried out using 400 ohms resistance, 1.25 kV, 25 μF on a Biorad Gene Pulser; these conditions generally gave a time constant between 8.0 and 9.0. 900λ of ice cold 1.0 M sorbitol was immediately added to the suspension and incubated at room temperature for 15 minutes, shaking gently. The yeast suspension was then plated on 2 MD plates: 100 λ was placed on one plate, and the rest was briefly centrifuged and resuspended in 100 λ distilled water and plated on a second plate. The plates were incubated 2-4 days at 30 °C until colonies appeared.

Pichia Screening

Histidine independent colonies were picked into 150 λ distilled water in a microtiter plate. 2 λ of the suspension was spotted onto an MD plate, and 2 λ was spotted onto an MM plate in an identical position and incubated at 30 °C for 24 to 48 hours until clear differences between slow growing versus normal MM colonies were apparent.

Slow growing colonies were then examined by PCR to test for presence of the recombinant insert. Slow growing colonies were grown up in 1 mL MD broth and 100 λ

of yeast suspension was spun down in a microfuge. The yeast was then resuspended in TE buffer, pH 7.5 and the cell wall was digested with 50 U lyticase (Sigma) for 2 hours at RT. The presence of insert was tested using specific primers for the insert (the same ones used for cloning). PCR bands were then analyzed on agarose gels.

Methanol slow growing colonies were then screened for protein expression as in ref. Briefly, recombinant *Pichia* colonies were grown in 10 mL BMGY broth (20 g/L peptone, 10 g/L yeast extract, 1% glycerol, 400 μg/mL biotin, 13.4 g/L yeast nitrogen base, buffered with 100 mM potassium phosphate, pH 6.0) at 30 °C with rapid shaking for two days. Cells were centrifuged down and resuspended in 2 mL BMMY (BMGY with 1% methanol replacing the glycerol) for 2 days of induction at 30 °C with rapid shaking. Methanol (10 λ) was added after 1 day to replace that lost by evaporation. Yeast cells were spun down 10 minutes in a microfuge and supernatant was removed. 15 λ of supernatant was then analyzed on a coomassie stained SDS-PAGE gel. Positive expressing colonies gave protein bands which were clearly visible above background bands.

Expression of Recombinant Protein from Pichia

Expression was carried out generally using the procedure of Barr et al. (Barr et al., 1992). Briefly, frozen clones were grown 2 days in 10 mL MD broth at 30 °C to generate a starter culture for expression. The starter culture was used to inoculate a 2L culture of BMGY in a 6 L Erlenmeyer flask and shaken at 30 °C for 2 days to generate a high density culture. The yeast were then spun down and resuspended in 400 mL BMMY, transferred

to a 2 L Erlenmeyer flask and shaken at 30 °C for 2 days covered in 3 layers of sterile cheesecloth to guarantee adequate aeration. 2 mL of methanol were added after 1 day to replace that lost by evaporation.

The suspension was centrifuged at 4,000 g, 4 °C for 15 minutes, and the supernatant spun a second time for 1 hour at 10,000 g, 4 °C to remove cellular debris. The supernatant was then filtered through a 0.2 μm filter with glass fiber prefilter, and 1 λ/mL aprotinin and PMSF to 1 mM were added as protease inhibitors. The supernatant was then concentrated using a 76 mm YM10 10 kD or YM3 3 kD cutoff membrane (Amicon) to approximately 40 mL. The concentrated supernatant was then washed twice with 400 mL of 0.1 M ammonium bicarbonate and re-concentrated each time to less than 40 mL.

Construction of Bacterial Expression Clones

Chick neogenin (Genbank accession code U07644) clones containing both alternative splicing forms of the intracellular domain (NE3 and NE4) were used as starting material to generate PCR fragments for use in T7 promoter based bacterial expression systems. PCR primers were derived from the chicken neogenin or Bravo/Nr-CAM sequence, and appropriate restriction sites and an amino-terminal purification tag were added to the oligos.

neo-cyt.S- 5' GAA TTC CAT ATG TGC ACT CGT CGT ACC ACT- 3'

neo-cyt-his.A- 5' GCGG CCGC GGA TCC TTA GTG ATG ATG GTG GTG ATG TCG TGC TGT AGT GAT GGC ATT -3'

bravo-fib.S- 5' CAT ATG GTC CAA CCA CTT TAT CCA -3'

bravo-fib.A- 5' GGA TCC TTA TGG ACC TGT CTC AAA CAG -3'

PCR for neogenin was carried out using Vent polymerase under the following conditions: 92 °C for 5 min., 72 °C 5 min., followed by 25 cycles of 92 °C 0.5 min., 56 °C 1 min., 72 °C 2 min., and then a 72 °C 5 min. extension. Bravo/Nr-CAM PCR was identical except the annealing temperature was 55 °C. Cloning and analysis were essentially identical to that used for the *Pichia* vectors, except that the restriction enzymes used were NdeI and BamH1 to accommodate the different vectors (pET-3a or pT7SC, also prepared from XL-1 bacteria).

Bacterial Expression Conditions

Expression plasmids were then transformed into BL21-pLYS-S using calcium chloride/ heat shock transformation. Colonies were grown overnight on 2xYT/amp plates, and a single colony was picked into a 10 mL culture of 2xYT/chlor/amp. The culture was then grown to an $O.D_{600}$= 0.6 and induced with 2 mM (neogenin) or 0.5 mM (bravo) IPTG to begin transcription of the T7 RNA polymerase. Cultures were then grown at room temperature for 3 hours (bravo) or 24 hours (neogenin).

Bacteria producing neogenin's intracellular isoforms were spun down and resuspended at 1 mL lysis buffer (1% NP-40, 0.2 M sodium chloride, 20 mM tris, pH 7.5, 2 mM EDTA, 50 mM magnesium chloride, 5 mM manganese chloride) per 100 mL of initial culture. The resuspended bacteria were then frozen at -70 °C for at least 1 hour to lyse the cells. Lysate was then quickly thawed and 40 λ of 1 mg/mL DNAse I was added.

The lysate was incubated at room temperature until the viscosity of the solution was significantly diminished. The lysate was then centrifuged at 12,000 g for 15 minutes at 4 °C and the clear top solution was removed. The pellet was then washed with the same volume of lysis buffer, recentrifuged, and the supernatant added to the previous supernatant.

Bacteria producing bravo's fibronectin domain were pelleted at 5000g for 5 min. at 4 °C and resuspended in 2 mL lysis buffer (50 mM Tris pH 8.0, 25% sucrose, 1 mM EDTA, 0.5 mM PMSF, 2.5mg/ml lysozyme) per 250 mL culture. Cells are lysed for 30 min. on ice, at which time DNAse I is added to 20ug/ml, along with 25 ul 1.0M $MgCl_2$ and 2.5ul 1.0M $MnCl_2$, and the lysate is incubated at room temperature for 30 min. Two volumes of detergent buffer (0.2M NaCl, 20mM Tris pH7.5, 2 mM EDTA, 1% NP-40, 1% deoxycholic acid, 0.5 mM PMSF) were added, mixed vigorously, and the inclusion bodies were pelleted at 5000g for 10 min. at 4 °C. The inclusion body pellet was washed five times by resuspension in wash buffer (1 mM EDTA pH 8.0, 0.5% Triton X-100, 0.5 mM PMSF) and recentrifugation. The inclusion bodies are solubilized by 6M guanadinium hydrochloride in 50 mM Tris pH 8 and dialyzed against PBS for at least 6 hours.

IMAC Chromatography

The immobilized metal affinity chromatography (IMAC) column was prepared by taking 1 mL IDA-agarose (Pharmacia) and washing it with 10 volumes distilled water at 1mL/min. The agarose was charged with 2.5 mL 10% nickel chloride in distilled water

and washed with 20 volumes of distilled water. The column was then equilibrated with 10 volumes running buffer (1 M sodium chloride, 50 mM sodium phosphate, pH 8.0).

Protein supernatant was diluted 1:1 with 2x running buffer and loaded onto the column at 1 mL/min. The column was then washed with 20 volumes running buffer. Weakly bound proteins were removed from the column by washing with running buffer supplemented with 10 mM imidazole. Ten 0.5 column volume fractions were collected, followed by a 5 column volume long wash which was collected in batch. Protein was then eluted into 10 0.5 column volume fractions using running buffer plus 50 mM imidazole, and all remaining protein was removed by using running buffer plus 200 mM imidazole, again collecting 10 0.5 column volume fractions. Fractions were then analyzed by SDS-PAGE stained with coomassie blue.

Ion-Exchange Chromatography

IMAC fractions containing the protein of interest were then adjusted to low salt (< 50 mM sodium chloride) and 50 mM sodium phosphate buffer, pH 6.0. Protein was loaded onto a 1 mL Mono-S FPLC column (Pharmacia) at 1 mL/min and washed with 3 mL 50 mM sodium phosphate, pH 6.0. Sodium chloride concentration was increased to 1.0 M over 20 minutes using a linear gradient at a flow rate of 0.2 mL/min. Fractions containing protein were then identified by SDS-PAGE.

Hydrophobic Affinity Chromatography

As a final purification step, neogenin intracellular domain solutions were adjusted to 0.8 M ammonium sulfate, 50 mM sodium phosphate, pH 7.0 and loaded onto a 1 mL Phenyl Superose column (Pharmacia) at the recommended flow rate of 0.5 mL/min. Protein was eluted in a decreasing linear gradient of ammonium sulfate to 0 M, 50 mM sodium phosphate at 0.2 mL/min. Fractions were analyzed by SDS-PAGE.

N-Terminal Sequence Analysis

Recombinant protein was separated by SDS-PAGE and blotted onto Pro-Blott (Bio-Rad) PVDF membrane prepared according to manufacturer's recommendations, and proteins were blotted overnight at 150 mA in Towbin's buffer (39 mM glycine, 48 mM tris base, 0.074% SDS, 20% methanol). The blot was rinsed in distilled water and stained briefly in 0.05% coomassie blue in 45% methanol, 45% acetic acid for 5 min. The gel was then destained in 45% methanol, 45% acetic acid and thoroughly rinsed with distilled water. The desired protein bands were then excised from the PVDF filter and sequenced in the Caltech Protein/Peptide Microanalytical Laboratory using an ABI 373A protein sequencer.

Mab production

Protein was prepared as an immunogen according to protocols established in (Kayyem et al., 1992). 10 μg of KLH-coupled fibronectin domain or 60 μg of neogenin Ig domains 1 thru 4 (20 μg unmodified, 20 μg coupled to KLH, and 20 μg biotinylated) emulsified with Freund's Complete Adjuvant (Sigma) was injected subcutaneously on the

back and tail of anesthetized Robertsonian 8.12 mice (Jackson Labs). Three subsequent booster immunizations were given with the same amount of antigen (not coupled to KLH) emulsified in Freund's Incomplete Adjuvant (Sigma) in the same locations at 6 week intervals. A final boost was injected in the spleen three days prior to the fusion.

All fusions were performed in the Caltech Hybridoma Facility essentially as described in Taggart and Samloff (Taggart and Samloff, 1983). Supernatants from positive cells with good growth characteristics were screened on E7.5 chicken embryos. Horizontal sections (10 μm sectioned on a cryostat) containing the retina, optic chiasm, and tectum were used for the screening as described below. Positive clones were expanded and frozen in liquid nitrogen, while several bright positives were subcloned and rescreened.

Immunohistology

E7.5 chick embryos were fixed for 8 hours with gentle shaking in 4% paraformaldehyde, 0.1 M sodium phosphate buffer, pH 7.0 as in (Vielmetter, 1994). The embryos were then soaked in 25% sucrose for several days to remove the paraformaldehyde. Heads were embedded in tissue-tek (Miles), sectioned horizontally (10 μm) on a cryostat for retino-tectal sections containing retina, optic chiasm, and tectum, and were collected on subbed slides. Sections were then rehydrated briefly with PBS and incubated in primary antibody supernatant for 1 hour. They were then washed 3 x 5 minutes in PBS and incubated in flourescein conjugated anti-mouse antibody (Cappel) in PBS + 1% BSA. Slides were then fixed in 4% paraformaldehyde 2 min., washed 3 x 5

minutes in PBS, mounted in glycerol, and coverslips were sealed over the sections. Slides were analyzed on a Zeiss fluorescent microscope and photographed.

Western Blots

Protein was separated on standard polyacrylamide gels as described. Gels were then blotted as above and blocked with 5% Carnation nonfat dry milk in PBS. The blots were then incubated in primary antibody diluted in milk+PBS for 1 hour then washed 3 x 10 minutes in PBS + 0.1% tween 20. Blots were then incubated in AP-conjugated anti-mouse antibody (Amersham) for 1 hour, then washed 3 x 10 minutes in PBS + 0.1% tween 20 followed by 1 x 5 min. PBS. The blot was then briefly rinsed in color development solution (100 mM tris, pH 9.5, 100 mM sodium chloride, 5 mM magnesium chloride) to adjust the pH, and finally protein was visualized using 0.3 mg/mL nitro-blue tetrazolium (NBT) and 0.15 mg/mL 5-bromo- 4-chloro- 3-indolyl phosphate (BCIP) in color development solution until good contrast was obtained.

Results

Expression Strategy and Vector Construction

The 5' promoter of the pPIC-9K vector (figure 1a) and sequences further 3' to the AOX-1 locus (not shown) direct integration into the yeast genome generating the Mut^S phenotype (slow growth on methanol), which occurred in up to 30% of our his+ colonies, although the number of methanol slow colonies varied significantly depending upon the

insert. Colonies were screened in small shake tube cultures; and the number of expressing colonies at this stage varied, with 12 of 12 secreting the intracellular domain of bravo/NrCAM (not shown), 11 of 18 secreting the 2 Ig domain construct (figure 2), and 1 of 12 secreting the 4 Ig domain construct (not shown).

The fifth FNIII repeat of Bravo/NrCAM under the control of the bacterial T7 promoter was expressed in BL21-pLYS-S bacterial culture under IPTG induction. pLYS-S refers to an independent plasmid in the BL21 strain of bacteria which produces low levels of T7 lysozyme. T7 lysozyme is a natural inhibitor of T7 RNA polymerase and reduces T7 RNA polymerase activity in the absence of IPTG induction and has been shown to sometimes decrease the toxicity of certain gene products (Rosenberg et al., 1987). As shown in figure 4, virtually all of the overexpressed protein is found in the inclusion body fraction.

During the construction of the small isoform of neogenin in the pET-3a expression vector, only 1 of 20 colonies contained the full length expression insert adjacent to the T7 promoter. Other colonies contained no insert detectable by PCR or restriction digest, or contained various sized fragments that seemed to delete internal regions of the DNA. In the pET vector, the small isoform of neogenin expressed at low levels compared to the fifth fibronectin repeat of Bravo/NrCAM. Because the small isoform was poorly overexpressed, it did not form an inclusion body and therefore needed to be directly purified from the supernatant (figure 5).

The large isoform of neogenin's intracellular domain was produced in virtually undetectable amounts using the pET-3a vector (data not shown). In order to obtain this

protein, we used a related expression vector, pT7SC (31). Using this expression and control cassette (diagrammed in figure 1c), BL21-pLYS-S bacteria were able to produce the large isoform at detectable levels in the bacterial supernatant (figure 5).

Protein Purification

Purification from the *Pichia* supernatant was accomplished in a single chromatography step using immobilized metal affinity chromatography (IMAC). Protein was readily eluted under mild conditions by using imidazole as a competitive chelating agent (figure 3). The level of recombinant protein for all three *Pichia* expression constructs was approximately 20 mg of IMAC purified material per liter of BMGY starting culture. During the purification of the 4 Ig domain neogenin construct, there is a minor co-purifying 37 kD product. Since it is not present in purified fractions of other expression products, it is likely an N-terminal proteolytic fragment as it is purified on an IMAC column with a C-terminal his tail.

Purification of the inclusion body was also relatively simple and included several washes to yield essentially pure protein (figure 4). This protein was produced on a massive scale-- approximately 400 milligrams per liter of bacteria.

Because the neogenin isoforms did not form an inclusion body, purification of these proteins from the bacterial supernatants required three chromatography steps: IMAC chromatography, Mono-S cation exchange chromatography, and phenyl superose hydrophobic affinity chromatography (see figure 5). This yielded essentially pure protein that could then be used for further experiments. The final yield of these proteins was

approximately 5 mg/L of culture for the small neogenin isoform and 3 mg/L for the large product.

Amino-terminal protein sequencing

The amino-terminal sequence for our various expressed domains was determined and is shown in Table 1. Bacterial expression yields protein that begins essentially where it is expected based on the coding sequence. The first residue of the neogenin intracellular isoforms is not determined because it is a cysteine which must be alkylated before being visible in the protein sequencer. For the *Pichia* products, there appears to be a fair amount of variability in the cleavage of the alpha factor leader protein. The secreted Bravo intracellular domain begins immediately after two dibasic residues which resemble a leader peptide cleavage site. The 4 Ig domain construct begins at its expected position, and the 2nd and 3rd Ig domain protein begins 18 residues downstream of its expected starting position after the leader sequence. The smaller product seen as a minor component of the IMAC purified material is a heterogeneous proteolyzed product that corresponds to the region between the first and second Ig domains.

Monoclonal Antibody Production

Recombinant protein was injected into Robertsonian 8.12 mice. Antibodies were raised against two of the extracellular domain recombinant proteins, the fifth fibronectin repeat of bravo/NrCAM and the four Ig domains of neogenin.

A single positive antibody was generated against the fibronectin domain, while greater than 100 antibodies were found to be reactive to native neogenin on tissue sections. Representative tissue staining patterns for neogenin and bravo/NrCAM are shown in figure 6. These patterns are identical to those seen with the original antibodies (data not shown). The monoclonal antibodies were also used on western blots to show reactivity with expressed protein domains and full length neogenin (figure 7). Both Bravo/Nr-CAM and neogenin antibodies recognize the native protein from brain lysate as well as the immunogen. Moreover, some of the neogenin monoclonals react with different epitopes of the immunogen. 3E7 reacts with the immunogen as well as the smaller two Ig domain construct, saying that the epitope is in or around these two Ig domains. 1G10, however, is an antibody that reacts with the immunogen but not with the second and third Ig domain construct. Since it also reacts with the second through fourth Ig domain proteolytic fragment, it appears to react with an epitope in or around the 4th Ig domain. Thus the fusion against the 4 Ig domains of neogenin produced antibodies against multiple epitopes of the immunogen.

Discussion

The cell surface proteins neogenin and Bravo/NrCAM/NrCAM are two of the numerous Ig superfamily molecules that are believed to play a role in the development of the nervous system. In many ways, they are typical of this class of proteins in general. Both are very large proteins that contain multiple independent folding domains, with each

domain potentially involved in different functions. Each is also relatively scarce in the organism, requiring multiple brains to obtain scant micrograms of purified protein. Finally, both molecules contain an intracellular domain that is believed to transmit a signal to the cell (Vielmetter et al., 1994; Grumet et al., 1991).

We chose to express individual domains instead of full length proteins because we wanted to be able to tease apart the functions of the domains. We also wanted to optimize the chances of generating antibodies that would recognize particular regions of the proteins which would likely contain functions of interest. Numerous protein domains have been found to interact with particular ligands (Hynes, 1987; Rao et al., 1993), and recombinant protein expression followed by monoclonal antibody production can generate specific reagents that can interfere with that function.

We have expressed various domains of these molecules in two expression systems, *E. coli* and *Pichia pastoris*. The choice of expression systems is extremely important for producing recombinant protein, and even related proteins may not be expressed in a particular system. For instance, the immunoglobulin domains of Bravo/NrCAM failed to express in *Pichia* (R. Lane, personal communication) while the Ig domains of neogenin could be expressed at fairly high levels. The intracellular domain of Bravo was produced well in *Pichia*, but neogenin's intracellular domain was not produced (F. M.). The problem of generating a correct tertiary structure as well as appropriate post-translational modifications is also an important consideration. Both the fibronectin domain and the Ig domains contained sufficient native epitopes to generate monoclonals which react with

native protein, suggesting that at least portions of the expressed proteins are in a natural configuration.

Another problem occasionally encountered in heterologous expression is protein toxicity. Neogenin's intracellular domains appear to be toxic in *E. coli* according to several criteria, including the difficulty in obtaining an appropriate expression cassette, the long induction period to produce protein (24 hours vs. 3 hours for the fibronectin repeat--see materials and methods), the relatively poor yield of the intracellular domain compared to the fibronectin repeat, and the fact that the large isoform was only able to be produced when transcriptional repressors were included in the plasmid flanking the T7 promoter and terminator. Another possible approach would have been to use the pLYS-E bacteria (Studier, 1991) which express a higher level of the T7 lysozyme inhibitor to restrict uninduced expression even further, although we did not pursue this option.

Pichia avoids a number of problems typically associated with bacterial expression. Generally the product is secreted into the medium, which facilitates protein purification. For naturally secreted eukaryotic proteins, the processing steps in the yeast endoplasmic reticulum fairly closely resemble those found in mammalian cells. This could allow for expression of functional proteins if they require particular environments for folding or processing that are not present in bacteria. *Pichia* also has some advantages over mammalian expression systems including its inexpensive growth and induction media, the relative ease of generating a large cell mass for protein production and scale up, and the ease of purifying the protein from the culture medium.

IMAC chromatography has been shown to be useful in many different types of situations (Lindner et al., 1992; Hochule and Piesecki, 1992). It is extremely good for purifying proteins from the *Pichia* medium, obtaining almost homogeneous protein in a single purification step. Furthermore, a six histidine tail on a protein is unlikely to either alter the functional characteristics of the protein or generate a strong immune response. Other purification tags for expressed proteins often need to be removed because they are large or because they are highly immunogenic and are recognized by high affinity antibodies (Narayanan, 1994).

IMAC chromatography was less effective for bacterial lysates. Contaminating proteins which bound to the Ni^{2+} charged resin constituted a significant portion of the total isolated protein. For this reason we chose to further purify the intracellular domains using ion exchange chromatography and hydrophobic affinity chromatography. Both proteins eluted at nearly the same point in the sodium chloride gradient of cation exchange chromatography (0.67 M vs. 0.70 M), which is not surprising as they have a large amount of overlapping sequence. The final step in the purification protocol, hydrophobic affinity chromatography on phenyl superose yielded a fairly significant difference in their elution protocols. The large isoform elutes significantly later than the small isoform (0.35 M ammonium sulfate vs. 0.22 M), suggesting that the alternative exon encodes a hydrophilic stretch of amino acids. As the cores of folded domains are typically hydrophobic (Dill, 1990), this argues that these alternatively spliced amino acids are found on the surface of the protein exposed to the cytoplasm and could encode for a different function of the protein.

Amino terminal sequencing results proved conclusively that the correct proteins were expressed. Bacterially expressed proteins from the T7 promoter consistently showed removal of the initiator methionine, but the expressed product began with the first residue of the natural protein. Proteins secreted from *Pichia*, however, showed a fair amount of variability in their signal sequence cleavage. The 4 Ig domain construct included 4 residues added because of restriction sites, and then entered normally into the neogenin coding sequence. This protein also showed proteolytic processing between the first and second Ig domains. The secreted Bravo/Nr-CAM intracellular domain showed almost expected processing; the protein was cleaved after the first two expected arginine residues. The smaller 2 Ig domain construct from neogenin shows a large amount of N-terminal processing, as 18 residues are missing from the expected amino terminus. Expression conditions have been known to modify how proteins are processed in this yeast system (Brankamp et al., 1995), although we did not explore this possibility. It is also possible that proteases in the supernatant digest less tightly folded regions of the protein, so that the determined amino-terminal sequences represent the beginnings of a tightly folded structure, which would not be inconsistent with our observed cleavage sites.

Our monoclonal antibody method (Kayyem et al., 1992) is optimized for generating high affinity IgG antibodies. First, immunization is carried out over the period of several months with multiple protein injections spaced approximately every six weeks or more. This allows time for memory B-cells to mature and generate a strong immune response. Because protein was not a limiting factor in the immunizations, a large amount

of protein was injected to increase the magnitude of the immune response. Furthermore, because it was virtually pure, most of the antibodies should be against the desired protein.

Antibodies were raised against proteins produced in both bacteria and yeast expression systems. The fibronectin domain fusion produced an antibody which recognizes an epitope located near the cell membrane on a large molecule on tissue sections. It is clear that there are antibodies against a variety of epitopes on the neogenin domains from the western blots. Some antibodies recognize the second and third domains as well as the 4 domain construct, while others recognize only the larger recombinant protein. While not exhaustive, it is clear that there are multiple epitopes being recognized.

There are a number of factors that could account for this phenomenon. First, the biotinylation of a portion of the injected neogenin fragment could have enhanced its antigenicity, generating an increased number of antibodies. Secondly, the larger size of the neogenin domains could have been important due to the increased number of epitopes. Each of the 4 Ig domains is approximately the same size as the fibronectin type III repeat and adopts roughly the same type of protein fold (Main et al., 1992). Another possible explanation for the relative scarcity of fibronectin domain antibodies is the fact that the domain is found just outside the surface of the cell and antibody accessibility on tissue sections might be an issue. Bravo/Nr-CAM could associate with other proteins on the cell surface, making a dense matrix after tissue fixation and limiting the portions of the domain with available epitopes. As the antibodies were initially screened against tissue sections, epitopes that may be native could have been masked due to its location.

Protein expression coupled with monoclonal antibody production provides a powerful set of tools to study new genes which are being discovered at a phenomenal rate. As genomic DNA sequences for various organisms become available, specific reagents to study how a new protein fits into a holistic view of the cell will be very much in demand. Antibodies are excellent reagents for examining expression patterns of particular proteins, individual isoforms of a protein, and for inhibiting functions of extracellular proteins (Suter et al., 1995). By using the genomic sequence to identify protein coding regions and expressing domains of a molecule in a heterologous system, it is possible to make reagents that specifically target the protein of interest.

References

Barr, K. A., S. A. Hopkins, and K. Sreekrishna. (1992) *Pharm. Eng.* 12, 48-51.

Brankamp, R. G., K. Sreekrishna, P. L. Smith, D. T. Blankenship, and A. D. Cardin. (1995) *Prot. Exp. Pur.* 6, 813-820.

Brown, W. C., and J. L. Campbell. (1993) *Gene* 127, 99-103.

Burgoon, M. P., M. Grumet, V. Mauro, G. M. Edelman, and B. A. Cunningham. (1991) *J. Cell Biol.* 112, 1017-1029.

Casali, P., G. Inghirami, M. Nakamura, T. F. Davies, and A. L. Notkins. (1986) *Science* 234, 476-479.

Cho, K. R., and E. R. Fearon. (1995) *Eur. J. Cancer* 31A, 1055-1060.

Cregg, J. M., and K. R. Madden. (1988) *Devel. Indus. Micro.* 29, 33-41.

Davis, J. Q., T. McLaughlin, and V. Bennett. (1993) *J. Cell Biol.*, 121, 121-133.

Dill, K. A. (1990) *Biochem.* 29, 7133-7155.

Frei, T., F. von Bohlen und Halbach, W. Wille, and M. Schachner. (1992) *J. Cell Biol.*, 118, 177-194.

Grumet, M., V. Mauro, M. P. Burgoon, G. M. Edelman, and B. A Cunningham.. (1991) *J. Cell Biol.* 113, 1399-1412.

Hochuli, E., and S. Piesecki. (1992) *Methods: A Comp. to Meth. Enzym.* 4, 68-72.

Hynes, R. O. (1987) *Cell* 48, 549-554.

Kayyem, J. F., J. M. Roman, Y. Vonboxberg, U. Schwarz, and W. J. Dreyer. (1992) *Eur. J. Biochem.* 208, 1-8.

Kohler, G., and C. Milstein. (1975) *Nature* 256, 495-497.

Lane, R.P., X. N. Chen, K. Yamakawa, J. Vielmetter, J. R. Korenberg, and W. J. Dreyer., (1996) *Genomics* 35, 456-465.

Lindner, P., B. Guth, C. Wulfing, C. Krebber, B. Steipe, F. Muller, and A. Pluckthun. (1992) *Methods: A Comp. to Meth. Enzym.* 4, 41-56.

Main, A. L., T. S. Harvey, M. Baron, J. Boyd, and I. D. Campbell. (1992) *Cell* 71, 671-678.

Mauro, V. P., L. A. Krushel, B. A. Cunningham, and G. M. Edelman. (1992) *J. Cell Biol.* 119, 191-202.

Moos, M., R. Tacke, H. Scherer, D. Teplow, K. Fruh, and M. Schachner. (1988) *Nature* 334, 701-703.

Moscoso, L. M. and J. R. Sanes. (1995) *J. Comp. Neur.* 352, 321-334.

Narayanan, S. R. (1994) *J. Chrom.* 658, 237-258.

Rao, Y., X. F. Wu, P. Yip, J. Gariepy, and, C. H. Si. (1993) *J. Biol. Chem.* 268, 20630-20638.

Romanos, M. A., J. J. Clare, K. M. Beesley, F. B. Rayment, S. P. Ballantine, A. J. Makoff, G. Dougan, N. F. Fairweather, and I. G. Charles. (1991) *Vaccine* 9, 901-906.

Rosenberg, A. H., B. N. Lade, D. S. Chui, S. W. Lin, J. J. Dunn, and F. W. Studier. (1987) *Gene* 56, 125-135.

Siegel, R. S., R. G. Buckholz, G. P. Thill, and L. M. Wondrack. (1990) International Patent Application, Publication No. WO 90/10697.

Sreekrishna, K., L. Nelles, R. Potenz, J. Cruze, P. Mazzaferro, W. Fish, M. Fuke, K. Holden, D. Phelps, P. Wood, and K. Parker. (1989) *Biochem.* 28, 4117-4125.

Studier, F. W. (1991) *J. Mol. Bio.* 219, 37-44.

Studier, F. W., and B. A. Moffatt. (1986) *J. Mol. Bio.* 189, 113-130.

Studier, F., A. H. Rosenberg, J. J Dunn., and J. W. Dubendorff. (1990) *Meth. Enzym.* 185, 60-89.

Suter, D. M., G. E. Pollerberg, A. Buchstaller, R. J. Giger, W. J. Dreyer, and P. Sonderegger. (1995) *J. Cell Biol.* 131, 1067-1081.

Suzue T., Z. Kaprielian, and P. H. Patterson. (1990) *Neuron* 5, 421-431.

Taggart, R. T., and I. M. Samloff. (1983) *Science* 219, 1228-1230.

Vielmetter, J., X. Chang, F. Miskevich, R. P. Lane, K. Yamakawa, J. Korenberg, and W. J. Dreyer. Submitted.

Vielmetter, J., J. F. Kayyem, J. R. Roman, and W. J. Dreyer. (1994) *J. Cell Biol.*, 127, 2009-2020.

Volkmer, H., B. Hassel, J. M. Wolff, R. Frank, F. G. Rathjen. (1992) *J. Cell Biol.*, 118, 149-161.

Von Bohlen, F., J. Taylor, and M. Schachner. (1992) *Eur. J. Neurosci.* 4, 896-909.

Wagner, S. L., R. S. Siegel, T. S. Vedvick, W. C. Raschke, and W. E. van Nostrand. (1992) *Biochem. Biophys. Res. Commun.* 186, 1138-1145.

Williams, A. F., and A. N. Barclay. (1988) *Ann. Rev. Immunol.* 6, 381-405.

Williams, E. J., J. Furness, F. S. Walsh, and P. Doherty. (1994) *Neuron* 13, 583-594.

Figure Legends

Figure 1: Promoter and expression cassettes

Figure 1a represents the expression cassettes we used for the pPIC-9K vector for *Pichia pastoris*. The vector contains the 5' promoter region of the methanol inducible AOX-1 gene fused to the *S. cerevisiae* alpha factor mating type sequence to direct secretion into the medium. This is followed by a multi-cloning region into which we have inserted our domains of interest with a 6 histidine tail to aid in purification. The 3' AOX-1 region contains a transcriptional terminator to end the mRNA transcripts. 1b represents our expression cassettes for the bacterial pET-3a vector. It uses the T7 viral promoter followed by a polycloning site into which we have inserted DNA coding for either the fifth fibronectin repeat of Bravo/NrCAM or the small isoform of neogenin's intracellular

domain fused to a 6 histidine tail. This is followed by a viral terminator region to end the mRNA transcripts. 1c represents the pT7SC (SC standing for stringent control) expression cassette. It is very similar to the pET-3a vector elements, except that it contains T2 and T1 sequences from the rrnB operon that prevent read-through transcription of the gene or mRNA synthesis from cryptic promoter sites. It contains the larger isoform of the intracellular domain of neogenin and a 6 histidine tail in the polycloning site.

Figure 2: Yeast expression screening gel

Figure 2 shows 6 initial shake tube supernatants for the 2nd and 3rd Ig domain construct of neogenin. 15 microliters of unconcentrated supernatant were run on a 10% polyacrylamide gel and stained with coomassie blue. Four of the six lanes show the presence of a 29 kD band which is the size of the expressed product.

Figure 3: Yeast supernatant purification by IMAC chromatography

Lane 1 shows the supernatant of a culture producing the four Ig domain neogenin construct. Lane 2 is the flow through portion of the IMAC chromatography. Lanes 3, 4, and 5 represent 50 mM elution fractions for the 4 Ig domain construct of neogenin, the 2 Ig domain neogenin construct, and the intracellular domain of bravo, respectively.

Figure 4: Inclusion body preparation

Figure 4 shows the inclusion body purification of the bravo fibronectin domain. Lane 1 is the washed inclusion body pellet. Lane 2 shows the bacterial supernatant, and Lane 3 shows the crude bacterial lysate.

Figure 5: Neogenin intracellular domain purification

Figure 5 shows the purification of the small (lanes 1- 4) and large (lanes 6 -9) isoforms of neogenin from a bacterial lysate. Lanes 1 and 6 are total bacterial lysate, lanes 2 and 7 are a 50 mM imidazole elution fraction from IMAC chromatography, lanes 3 and 8 are the main peaks from mono-S ion exchange chromatography, lanes 4 and 9 are the major peak after phenyl superose hydrophobic affinity chromatography, and lane 5 is a molecular weight marker (containing Biorad low molecular weight standards of 97, 66, 42, 31, and 21 kD).

Figure 6: Tissue sections of Bravo/NrCAM and neogenin

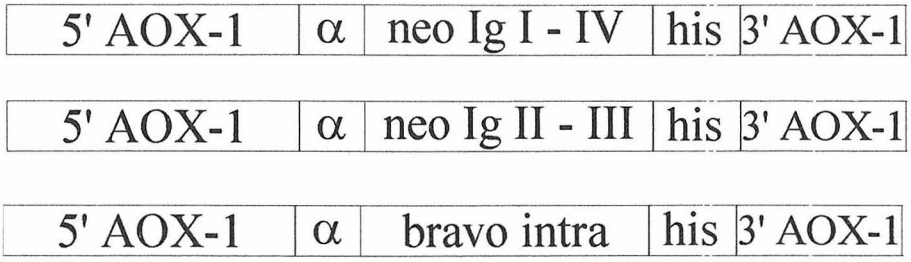
Figure 6a shows the tissue staining of a selected neogenin antibody (4F6). Notice that the nasal optic fiber layer is stained brightly whereas the temporal fibers are not. This is an identical staining pattern to what was seen using the first antibody to neogenin (data not shown). The inner plexiform layer is also reactive, although it is stained equally on both nasal and temporal sides. 6b shows the expression pattern of Bravo/NrCAM's alternatively spliced fifth fibronectin repeat, which is identical to that of the Bravo/NrCAM antibody (data not shown).

Figure 7: Western blots of selected antibodies

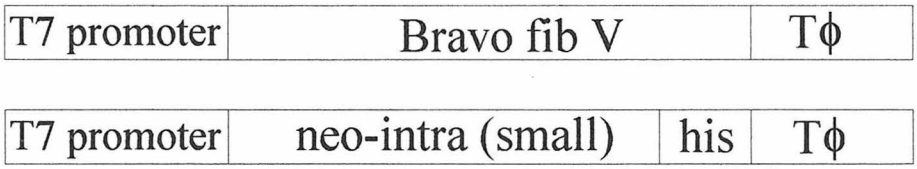
Figure 7 shows the antibody reactivity on western blots. The antibody against the fifth fibronectin repeat (lanes 2 and 5) recognizes an 80 kD band characteristic of an *in vivo* proteolyzed fragment of Bravo/NrCAM as well as the expressed domain. A bravo polyclonal antibody recognizes both the expressed fibronectin and intracellular domains (lanes 3 and 4). Neogenin antibodies (3E7) recognize the intact neogenin protein (lane 1) as well as both expressed Ig domain constructs (lanes 6 and 7). Some antibodies such as 1G10, however, recognize only the 4 Ig domain product (lanes 8 and 9).

Table 1: N-terminal amino acid sequencing results

1a

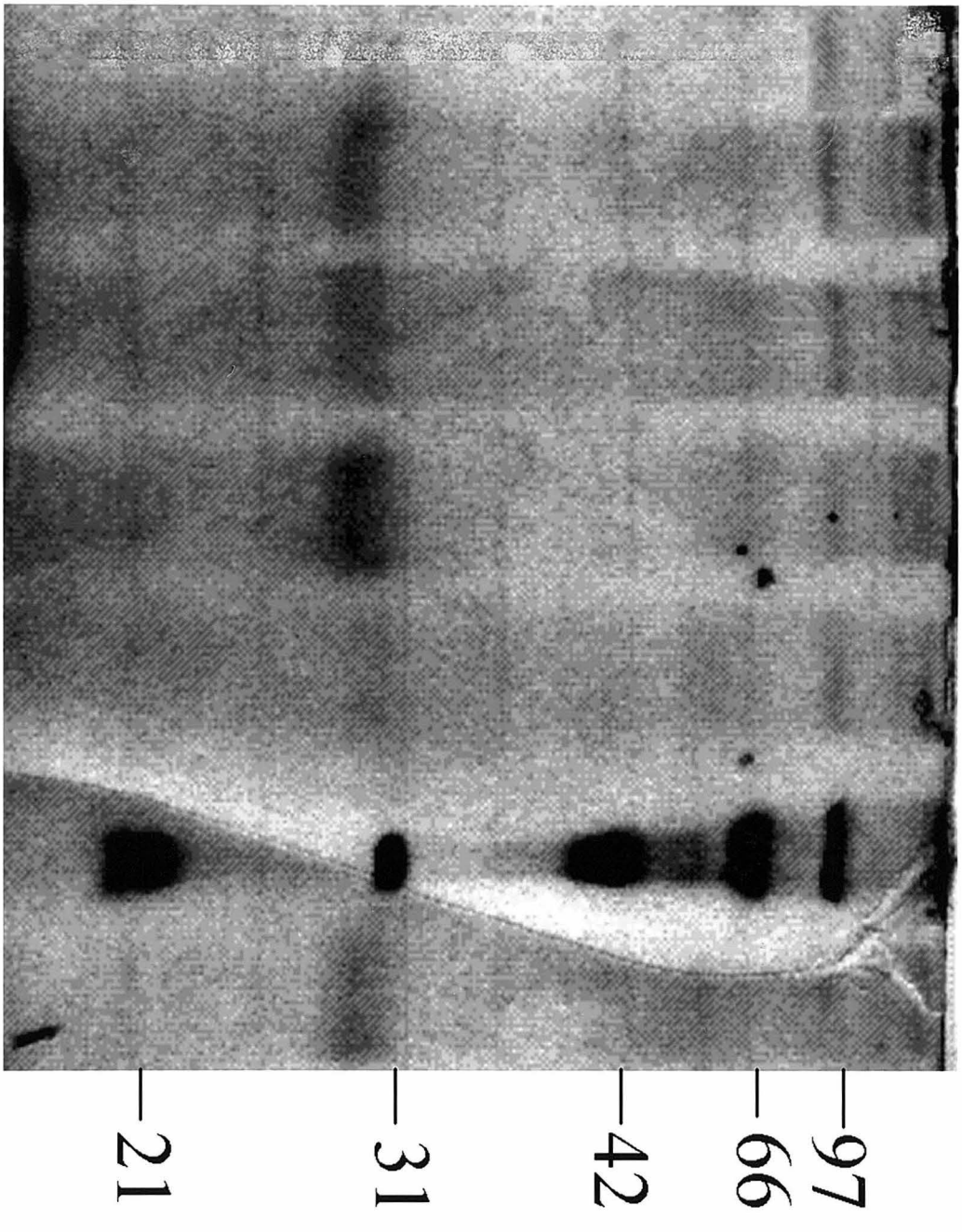

5' AOX-1
α
neo Ig I - IV
his
3' AOX-1
5' AOX-1
α
neo Ig II - III
his
3' AOX-1
5' AOX-1
α
bravo intra
his
3' AOX-1

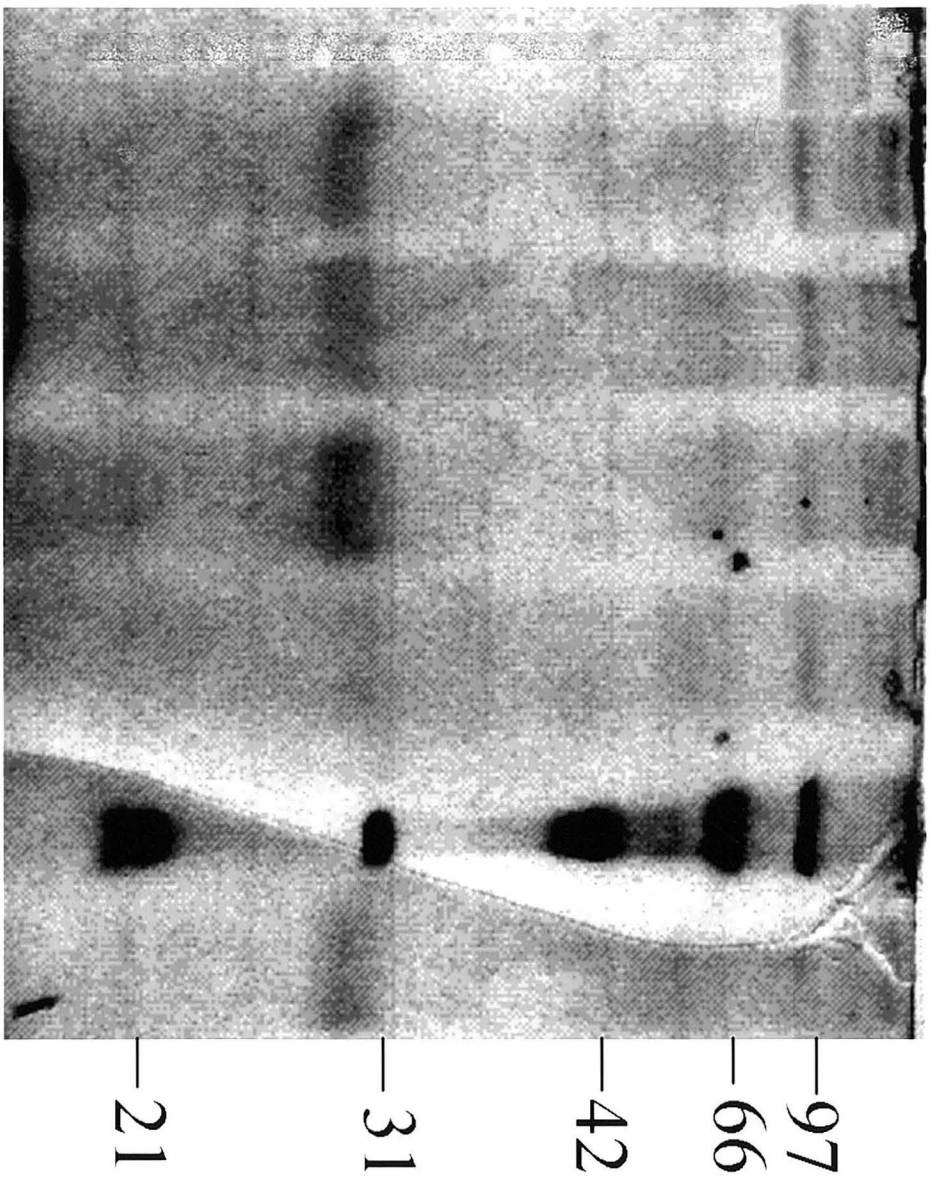
1b

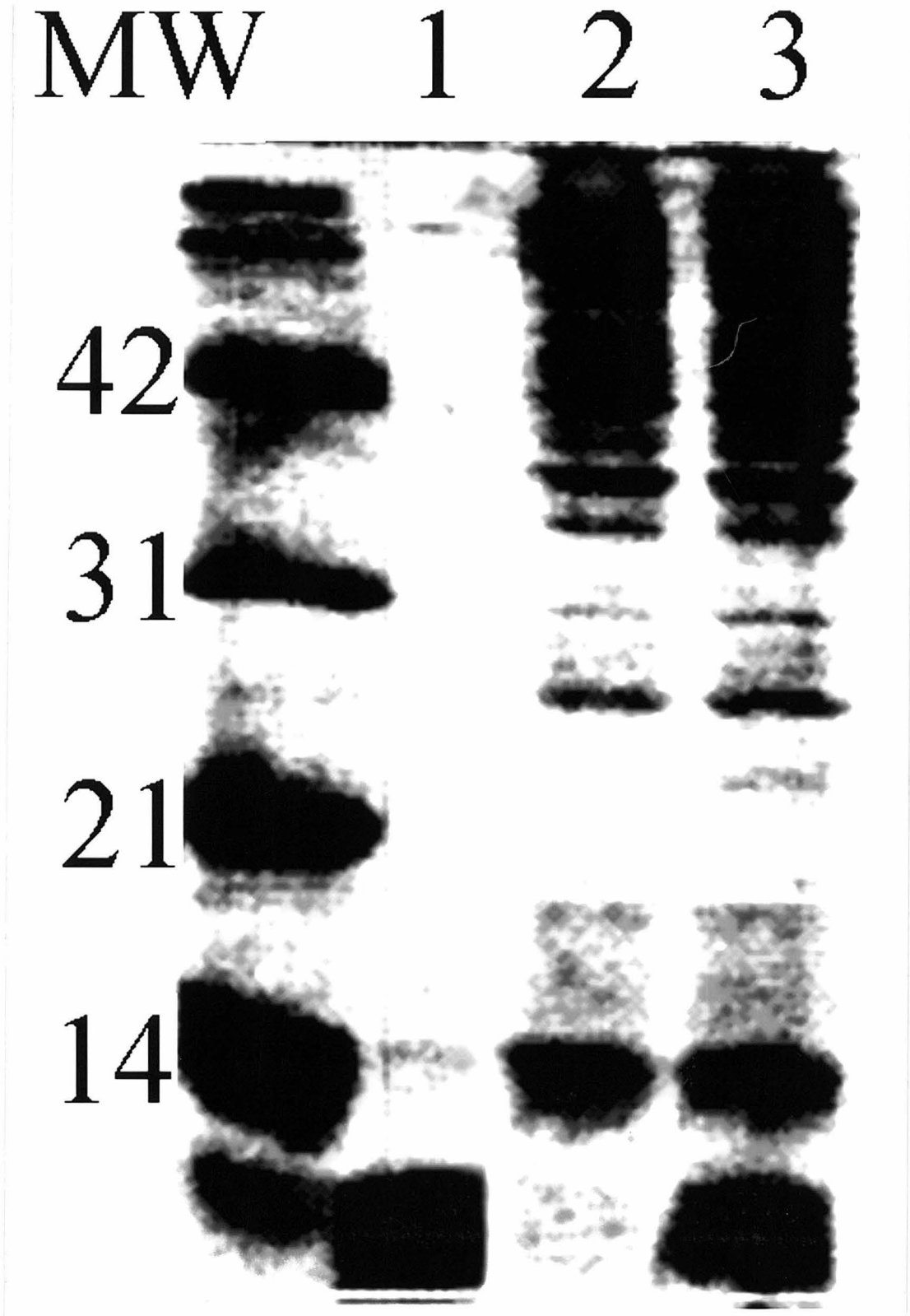

T7 promoter
Bravo fib V
Tϕ
T7 promoter
neo-intra (small)
his
Tϕ

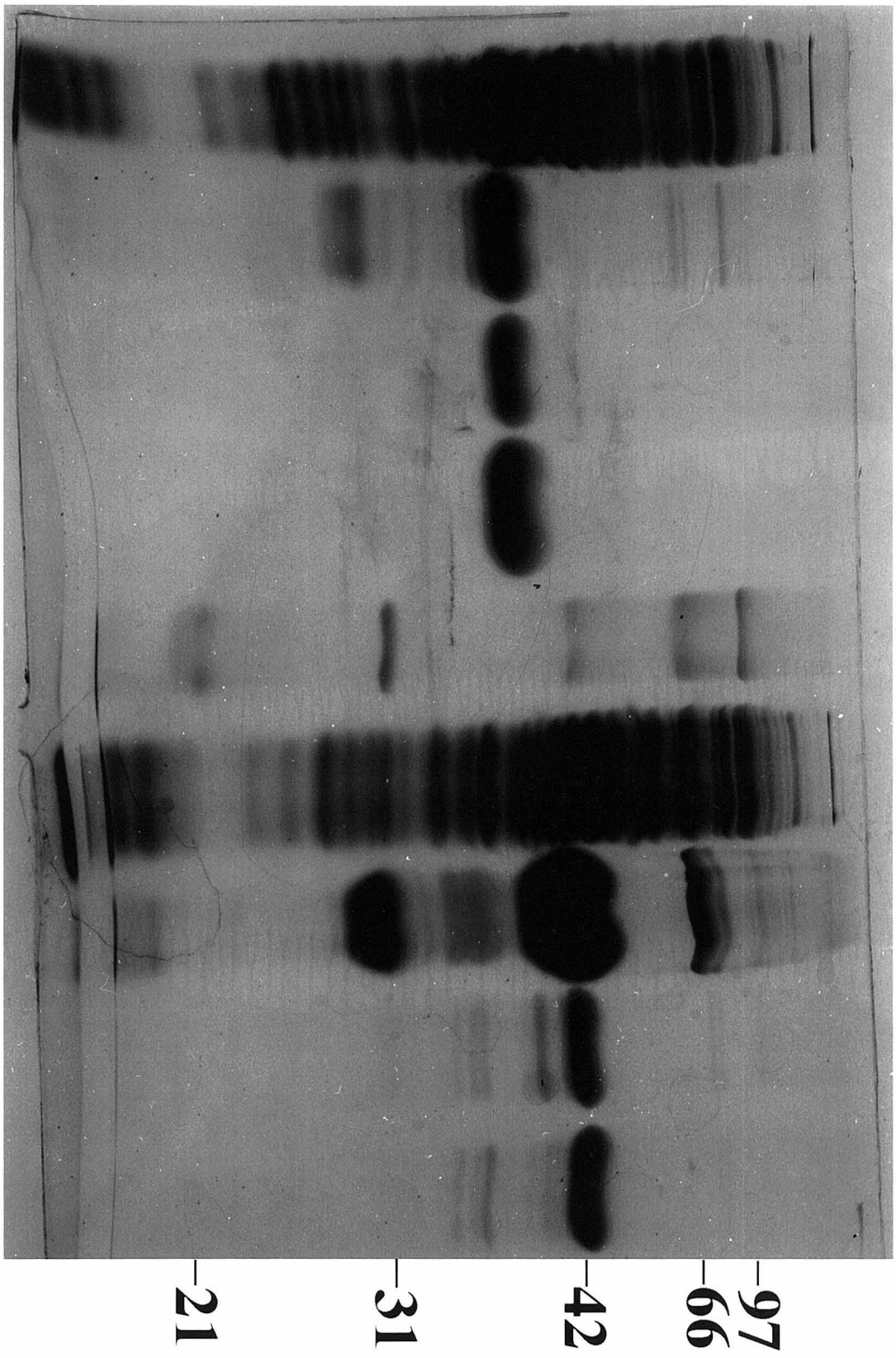
1c


T2
T1
T7 promoter
neo-intra (large)
his
Tϕ
T1
T2

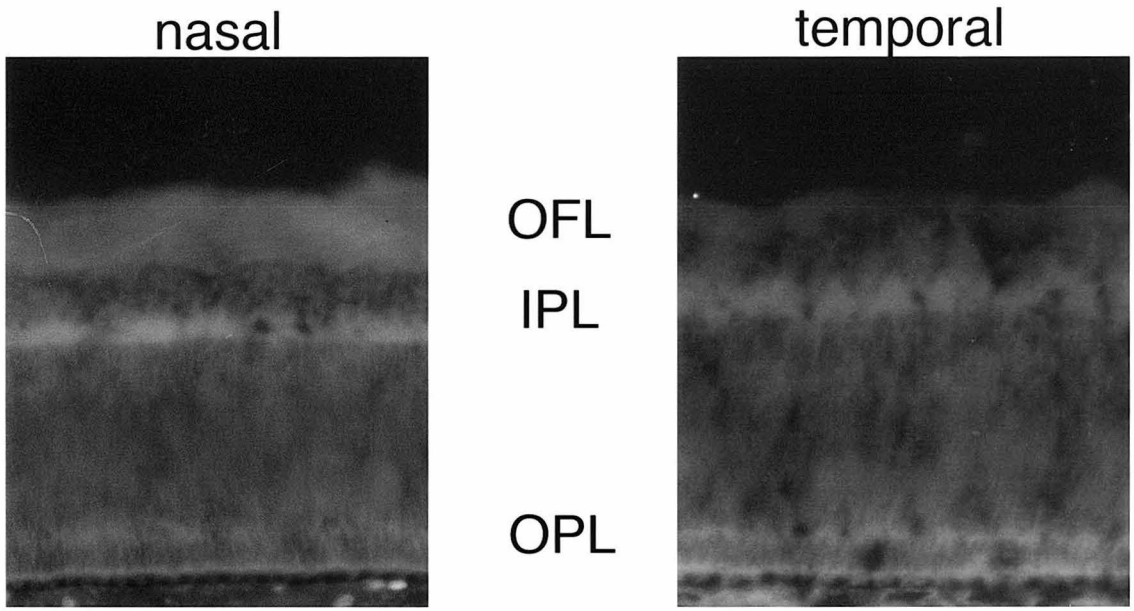

1
2
3
4
5
MW
6
97
66
42
31
21


97
66
42
31
21


MW
1
2
3
42
31
21
14


1
2
3
4
5
6
7
8
9
97
66
42
31
21

a



b

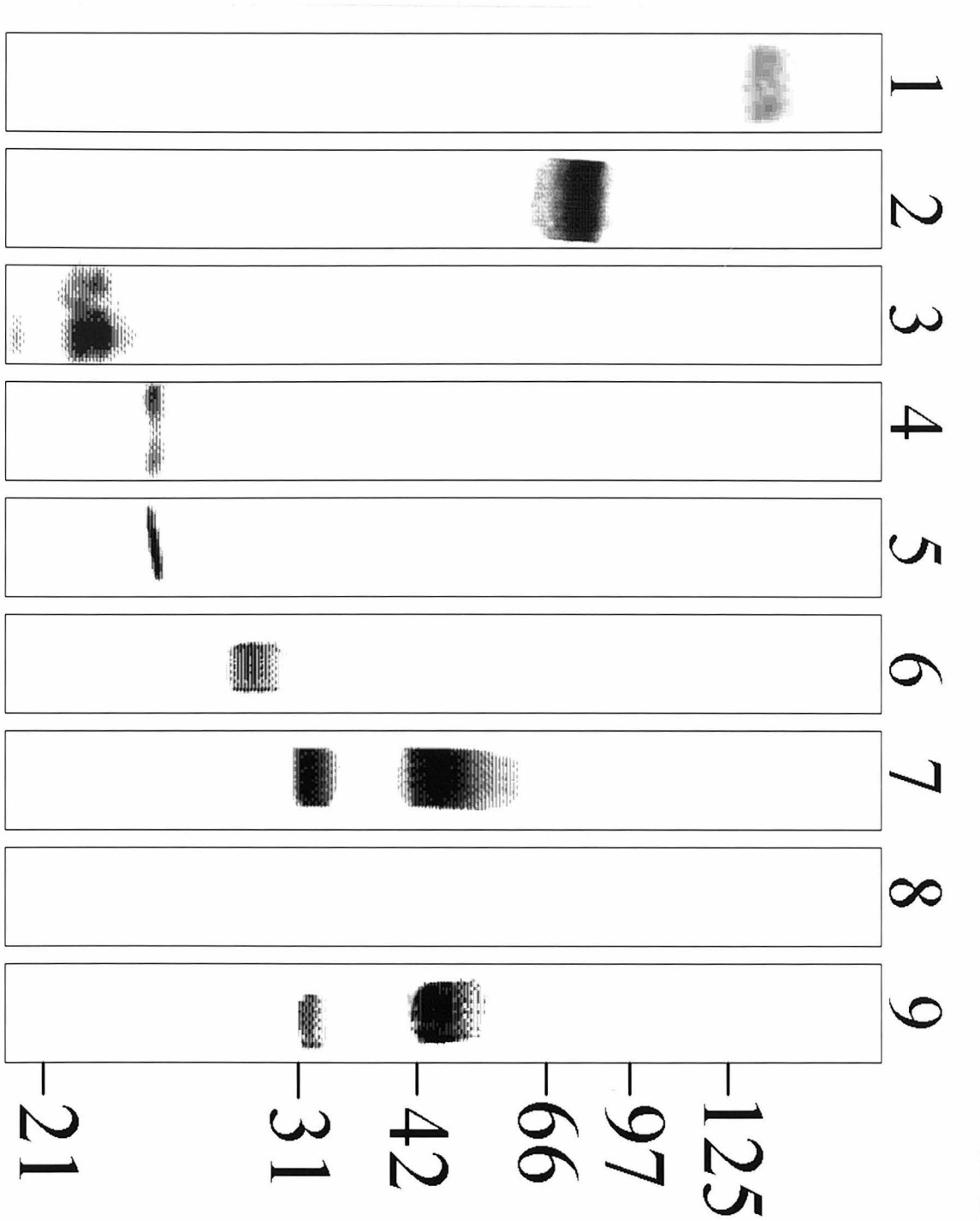

1
2
3
4
5
6
7
8
9
125
97
66
42
31
21

Table 1
N-terminal Sequences

Construct	Sequence
Bravo Fibronectin:	V-Q-P-L-Y-P-R-I-R-N
Bravo Intracellular:	N-K-G-G-K-Y-P-V
Neogenin small intracellular:	X-T-R-R-T-T
Neogenin large intracellular:	X-T-R-R-T-T
Neogenin Ig II-III:	S-N-P-E-L-S
Neogenin Ig I-IV:	Y-V-E-F-V-V-R-T-F-T-P-F
Neogenin Ig I-IV fragment:	G-L-P-R-F-T-S-Q
	T-V-A-G-L-P-R-F
	L-T-V-A-G-L-P-R

Human Neogenin a Member of the Ig Superfamily Shows Strong Evolutionary Conservation and is Related to the Tumor Suppressor Molecule DCC

Jost Vielmetter[1], Xiao-Ning Chen[2], Frank Miskevich[1], Robert P. Lane[1], Kazuhiro Yamakawa[2], Julie R. Korenberg[2], and William J. Dreyer[1]

[1]Division of Biology 156-29
California Institute of Technology
Pasadena, CA 91125

[2]Medical Genetics
Cedars-Sinai Medical Center
Davis Building
110 George Burns Rd.
Los Angeles, CA 90048

Address all correspondence to
Jost Vielmetter
Clinical Micro Sensors
101 Waverly Drive
Pasadena, CA 91105
Phone: (818) 395-4909
Fax: (818) 449-0756
e-mail: cmsjv@@wavenet.com

Running title:
Human Neogenin Sequence and Chromosomal Location

Subject Category: gene structure/function relationships

ABSTRACT

Neogenin was first identified in the chick embryo and, like a number of cell surface proteins of the Ig superfamily including N- CAM and L1 (generally called cell adhesion molecules or CAMs), is expressed on growing nerve cells in the developing nervous system of vertebrate embryos. Neogenin is also expressed in other embryonic tissues suggesting a more general role in developmental processes such as tissue growth regulation, cell-cell recognition, and cell migration. Neogenin, unlike the CAMs is closely related to a unique tumor suppressor candidate molecule, deleted in colorectal carcinoma (DCC). Like DCC, the neogenin protein consists of four immunoglobulin-like (Ig-Like) domains followed by six fibronectin type III-like domains, a transmembrane domain, and an intracellular domain. We now report the cloning and sequencing of cDNA clones coding for the human neogenin protein. Human neogenin shares 87% identity with its chicken homologue and like its chicken counterpart is expressed in at least two different isoforms derived from alternative splicing in the intracellular domain. Northern blot analysis reveals two mRNA species of about 5 kb and 7 kb. The chromosomal location of the human neogenin gene was determined as 15q22.3-23, using fluorescence in situ hybridization. It therefore maps in the vicinity of a locus associated with Bardet-Biedl syndrome. The identification of human neogenin and its chromosomal location provides a basis for studying human genetic disorders or diseases that may involve this protein.

INTRODUCTION

Cell surface proteins of the Immunoglobulin (Ig) superfamily have been linked to important functions in animal development such as cell adhesion, specific cell recognition, and cell signaling. In humans mutations of some of the genes of the Ig superfamily have been shown to cause congenital malformations such as hydrocephalus and/or hypoplasia of many parts of the nervous system leading to severe mental retardation in survivors (Jouet et al., 1994). Somatic mutations of other genes of the Ig superfamily have been implicated in abnormal cell growth and cancer (Cho and Fearon, 1995).

Neogenin is a member of the Ig superfamily and, together with the tumor suppressor molecule DCC, defines a subfamily with a distinct combination of structural motifs. Neogenin, like DCC, is composed of four Ig-like domains, six fibronectin type III (FNIII) repeats, a putative transmembrane domain, and an intracellular domain. The intracellular domains of both neogenin and DCC are similar to each other but show no similarity to any other known protein (Fearon et al., 1990; Hedrick et al., 1994; Vielmetter et al. 1994). Based on the sequence and structural similarity between neogenin and DCC, similar functions can be expected for both proteins.

Neogenin, like DCC, is believed to be involved in the transition from proliferation to terminal differentiation. Neogenin expression in the developing nervous system of the chicken embryo supports this notion and, in addition, suggests a role in early neurite outgrowth and projection (Vielmetter et al., 1994). In the retina, neogenin is visible on ganglion cell fibers as soon as they begin to extend their axons. Another example of early neogenin expression is its appearance in the posterior portion of the cerebellum, a region known to give rise to new neurons in cerebellum development (Ledouarin, 1993).

DCC has proven to be of functional importance for the maintenance of normal tissue differentiation in multiple tissues and its deletion has been shown to result in progression of various cancers (Hedrick et al., 1994). Its expression in embryonic tissues suggests important roles also in embryonic development in birds, frogs, and mice (Chuong et al., 1994; Piercall et al., 1994b; Cooper et al., 1995). DCC is most highly expressed in the nervous system, and *in vitro* neurite outgrowth experiments strongly suggests it is involved in this process (Lawlor and Narayanan, 1992; Pierceall et al., 1994a).

As a first step towards the exploration of functions of neogenin in humans and a possible link with genetic disorders, we cloned and determined the sequence of human neogenin and determined its expression in various tissues. We have also identified its chromosomal location where it maps close to the gene locus for Bardet-Biedel syndrome (Carmi et al., 1995a,b).

MATERIALS AND METHODS

Generation of human fetal cDNA library.

For the construction of a 14 week trisomy 21 fetal brain cDNA library we used a ZAP-cDNA synthesis kit (STRATAGENE) which is designed for construction of unidirectional cDNA libraries (Yamakawa *et al.*, 1995). Double stranded cDNA was synthesized from 5 ug trisomy 21 fetal brain polyA+ RNA using a oligo(dT)-XhoI linker primer with 5-methyl dCTP. EcoRI linkers were attached to the 5' ends and products were digested with EcoRI and XhoI, and cloned into the UNI-ZAP XR vector. The library was packaged using the GigapackII Gold packaging extract (STRAGAGENE). The number of independent clones was 1.1x106. The library was amplified once, and

blue-white color assay indicated that 99% of the clones have inserts with an average size of 1.9 kb (calculated from 14 clones).

Isolation and sequence analysis of human neogenin cDNA clones.

We identified the cDNA clone HFBEH15 (5' sequence entered in GENBANK under accession code T07322) through a database search of expressed sequence tag (EST) sequences in the GENBANK with the chicken neogenin sequence. Bacteria containing this plasmid clone (Bluescript SK) are publicly available and were obtained through the American Type Culture Collection (ATCC). Using the M13 "forward" (5'-GTTGTAAAACGACGGCCAGT-3') and "reverse" (5'-CAGGAAACAGCTATGACCAT-3') priming sites in the polylinker region of the Bluescript SK phagemid vector (STRATAGENE) we sequenced 400-500 bases of the 5-prime and 3-prime ends of the insert of HFBEH15. We used the 5-prime and 3-prime sequence information that we produced from this clone to design and synthesize the primers HNS1:(5'-CTGCGAGTAGAAACACAACCT-3') and HNA1:(5'-CTCAGAATGGAAGCCTGAACT-3') and generated a 1017 bp PCR fragment HNS1/HNA1 (position 1582 to 2598 in the completed nucleotide sequence) with HFBEH15-DNA serving as template. The HNS1/HNA1 PCR-fragment was labeled with [alpha-32P] dCTP (New England Nuclear) incorporation by random priming (Boehringer Mannheim), and used to screen the human fetal library. The library was plated at a titer of approximately 105 pfu/plate (106 total plaques). Replica filters (Hybond-N, Amersham) were lifted and hybridized at high stringency (50% formamide, 4x SSPE, 1% SDS, 0.5% Carnation Non-Fat Dry Milk, 0.1% 37 degrees shaking overnight); the filters were washed at increasing stringency (highest stringency: 0.1x SSC, 0.1% SDS, 65 degrees for 10

min.). 33 positive cDNA clones were isolated, and all were converted into pBluescript plasmids by *in vivo* excision (Stratagene). Clones HN9 containing the largest (approximately 7 kb) insert was selected for sequence analysis. Sequencing was performed using nested deletions from subcloned EcoRI and BamHI fragments of this clone generated with a "ExoIII/Mung Bean nuclease kit" (STRATAGENE) as templates. Sequencing was done using the cycle sequencing kit and dye-labeled M13 primers from Applied Biosystems Inc. (ABI) and the sequence was determined using the 373A automated DNA sequencing system from ABI. All sequences were analyzed and compared using the GCG system software package.

Northern blot analysis

An adult human RNA blot was obtained from Clonetech Laboratories, Inc., which contained blotted RNA populations from various tissues: heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas. The HNS1/HNA1 PCR-fragment was labeled with [alpha-32P] dCTP (New England Nuclear) incorporation by random priming (Boehringer Mannheim kit) and used as a probe to hybridize at high stringency (50% formamide, 5x SSPE, 2% SDS, 10x Denhardt's solution, 100ug/ml sheared salmon sperm DNA); the blot was washed at increasing stringency (0.1x SSC, 0.1% SDS, 65 degrees for 10 min.). The labeled autoradiograph was exposed for 28 hours.

Fluorescence in situ hybridization (FISH).

The HFBEH15 plasmid DNA was used as probe to map the neogenin gene to human chromosomes by fluorescence in situ hybridization (FISH). The probe was labeled

with biotinylated 14-dATP (GIBCO, BRL) using nick translation and was subsequently hybridized to metaphase chromosomes prepared from normal male peripheral blood lymphocytes using the bromodeoxy-uridine synchronization method (Korenberg and Chen, 1995). FISH was performed essentially according to the method described in Korenberg and Chen (1995). Hybridization solution contained 200 ng probe DNA, 5 ug Cot 1 DNA and 5 ug sonicated salmon sperm DNA per 10 ul hybridization solution (70% formamide, 10% dextran sulfate, 2x SSC). The DNA was pre-annealed at 37 °C for 15 minutes and then applied to denatured chromosome slides. Post-hybridization washes (4x) were performed at 44 °C for 5 minutes in 2x SSC and 50% formamide, followed by (4x) at 55 °C for 5 minutes in 1x SSC. Hybridized DNAs were detected by using avidin-conjugated fluorescein isothiocyanate (Vector Labs). To generate clear reverse bands, metaphase chromosomes were counterstained with chromomycin A3 followed by distamycin A (Korenberg and Chen, 1996). The image was captured by using the Photometrics cooled-CCD camera (CH250) and the BDS image analysis system (Oncor Imaging, Gaithersburg, MD).

RESULTS

Sequence analysis of human neogenin cDNA

33 independent cDNA clones encoding at least two alternatively spliced isoforms of human neogenin were isolated. One clone, HN9 with an approximately 7 kb insert, was selected for sequencing. This clone contains an open reading frame (ORF) with one possible ATG translation initiation codon at the 5' end at nucleotide position 1 and a stop codon at nucleotide position 4383. The predicted amino acid sequence begins with a

hydrophobic stretch suggesting that this is the signal peptide of the protein. No Kozak consensus translation start signal sequence is present (GCCGCC[AorG]CCATGG; Kozak, 1989); however, the alignment with chicken neogenin (Fig. 2) strongly suggests that methionine encoded by the ATG-codon at position 1 of the human neogenin cDNA is indeed the start methionine. This is based on the assumption that signal peptides are normally not much longer than 20-30 amino acids, and furthermore, the border between signal peptide and mature protein often coincides with a sudden transition from low or no homology in the signal peptide to a drastic increase in homology of the amino terminal region of the mature protein portion if species homologues are aligned (Vonheijne, 1990). We observe such a transition at amino acid 28 to 29 in the human neogenin protein sequence and its equivalent at amino acid 4 to 5 in the chicken neogenin protein sequence. We therefore also expect the signal cleavage sites for both the chicken and human neogenin homologues to be within close vicinity of these transitional amino acid positions (within ten amino acids upstream or downstream of this region).

Like its chicken homologue, human neogenin includes several motifs for structural domains (Fig. 2). There are four predicted N-terminal immunoglobulin (Ig)-like domains based on characteristically spaced cysteine core residues (underlined in Fig. 1 and 2) that are required for the properly folded beta-sheet domain structure of the Ig domains (Williams and Barclay, 1988; Vaughn and Bjorkman, 1996). These Ig-like domains are followed by six fibronectin type III (FNIII) repeats, which are predicted based on characteristically spaced tryptophans and tyrosines that are also structurally critical core residues (underlined in Fig. 1 and 2). A 23 amino acid hydrophobic stretch predicts a transmembrane spanning region (amino acids 1104-1126 boxed in Fig. 1) (Kyte and Doolittle, 1982), followed by a C-terminal 335 amino acid intracellular domain. The

cartoon in Fig. 2 illustrates this domain structure.

The extracellular domains contain at least eight N-linked putative glycosylation sites (indicated with bold italic letters in Fig. 1) that are conserved between human and chicken neogenin (Marshal, 1972). The intracellular domain contains at least 13 conserved potential serine or threonine phosphorylation sites (indicated with bold italic letters in Fig. 1) (Woodgett et al., 1986).

Homology between the chicken and human neogenin proteins

Overall, the amino acid sequence of the human neogenin protein is 87% identical to chicken neogenin, an extremely high degree of identity (Fig. 2). The structural topology of four Ig domains, six fibronectin type III repeats, transmembrane and intracellular domain is identical to chicken neogenin, and together these proteins are grouped as evolutionary close relatives in the Ig superfamily of cell adhesion molecules. The transmembrane and intracellular domains of the human and chicken neogenin proteins are 92% identical.

Several alternatively spliced isoforms have been identified for neogenin, including a potential alternative exon coding for 20 amino acids between residues 429 and 430 in the human neogenin protein. This exon is expected since the comparison between human and chicken neogenin sequences reveals a gap in this region of an otherwise highly conserved alignment (Fig. 2). This potentially alternatively used protein sequence may have interesting functional relevance since it contains several clustered proline residues that may convey a bent to this region, and since it is bound between the Ig domain and FNIII domain cluster, the presence or absence of the sequence in the two different isoforms may

modulate the overall protein structure in a significant way. Another alternatively spliced region has been identified in the intracellular domain affecting residues 1249 through 1301. The corresponding region in the chicken neogenin protein has been shown to contain an alternatively spliced region, and both isoforms with and without the alternative exon have been identified by PCR (Vielmetter et al., 1994). We confirmed by PCR and direct sequence analysis of several human cDNA clones that the splice junctions are conserved between chickens and humans (see underlined sequence in Fig. 1).

Northern blot analysis of human neogenin

A probe (between nucleotide positions 1582 and 2598) was generated by PCR and used to probe a pre-blotted Northern blot (Clonetech Laboratories) of RNA from various tissues: heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas. Two bands at approximately 7 kb and 5 kb were detected in most tissues examined except in placenta and heart tissue where the signals were extremely low or absent (Fig. 3).

Human neogenin maps to chromosomal region 15q22.3-23

The human neogenin gene was mapped to chromosome band 15q22.3-23 (Fig. 4). Two independent experiments were performed, and over 100 metaphase cells were evaluated. Signals were clearly detected on two chromatids of at least one chromosome band 15q22.3-23 in 45% of cells. No other chromosomal sites with consistent signals were detected in greater than 1.0% of cells. To date, one known syndrome maps to this chromosomal region, the so called Bardet-Biedl syndrome (Carmi et al., 1995 a,b). This syndrome is a recessive autosomal disorder and is characterized by obesity, mental

retardation, polydactyly, retinitis pigmentosa and hypogonadism. Interestingly, some of these phenotypes are compatible with roles of neogenin in the development and maintenance of tissues in the nervous system including the retina, and the gut epithelium. However, fine mapping of the neogenin gene has to confirm where it exactly localizes in the genome in relation to known genetic defects.

DISCUSSION

We have cloned the human homologue to the chicken cell surface protein neogenin. These species homologues share a high degree of sequence identity, with an overall identity of 87% and with 92% identity in the intracellular domain. The only other protein sharing significant homology with neogenin (53% identity) is the tumor suppressor candidate gene DCC (Cho and Fearon, 1995). The strong conservation between human and chicken neogenin proteins suggests that the protein will have a conserved extracellular ligand and use the same signal transduction pathway in the two organisms despite the hundred millions of years of evolution that separate them. Furthermore, it argues that this protein has very little room for changes in its amino acid sequence, with mutations presumably causing deleterious effects for the organism.

Humans and chickens also share several potential alternative exons in neogenin. Particularly striking is the exon in the intracellular domain (53 amino acids starting at position 1249 of the human sequence) and a 20 amino acid stretch between the immunoglobulin and fibronectin domains. The intracellular exon is rich in histidines (8 of 53 residues), four of which are clustered around position 1262. This histidine cluster could chelate a divalent metal ion such as found in some DNA binding proteins and

metalloproteases (Christianson, 1991). Another prevalent amino acid in this alternatively spliced exon is proline providing the potential for turns in the protein chain which may alter the overall structure significantly and affect either binding of a ligand or perhaps a signal transduction event mediated by neogenin. There are also two possible serine phosphorylation sites in the alternative exon and one site two residues downstream of the exon that could be affected by the alternatively spliced state of the protein. Serine phosphorylation is known to modulate the functions of receptors such as the insulin receptor (Chin et al., 1993). A prevalent motif found repeatedly in the intracellular domain of neogenin is the motif PXXP. This motif has recently been shown to be involved in the binding of srk homology type 3 (SH3) domains (Mayer and Eck, 1995). Five of these motifs are also found in DCC in homologous regions. The occurrence and conservation of these motifs indicate some potentially interesting interactions with proteins linked to signaling pathways.

Human neogenin was identified in many of the adult tissues examined by northern blotting, including brain, lung, liver, skeletal muscle, kidney, and pancreas. Clearly neogenin is not restricted to a single tissue, and even though it is most prominent in the brain, it is probably not performing a neuron-specific function. It is possible that neogenin is expressed in stem cells in these other tissues, which would be consistent with a predicted role in terminal differentiation and cessation of cell division. DCC has also been shown to be expressed in a variety of cell types, and it has been implicated by deletion studies of various tumor tissues and cell lines to act as a tumor supressor molecule in tissues other than colon where it was originally identified (Cho and Fearon, 1995). While no direct evidence linking neogenin to tumor suppression is known to us, the strong similarity between the two proteins suggest that they likely play related roles in the

organism.

As determined by FISH, neogenin maps to chromosomal bands 15q22.3-23. This location maps it close to the position of a locus involved in Bardet-Biedl syndrome (Carmi et al., 1995 a,b). Several of the effects (mental retardation, polydactyly, retinitis pigmentosa and hypogonadism) may involve errors in terminal differentiation. However, more detailed studies are needed to map neogenin between appropriate markers in this region of the chromosome. The cloning, sequencing, and mapping described here may provide the first step and new molecular probes to study human genetic diseases that are associated with the neogenin gene.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grant EY07725.

REFERENCES

Carmi, R., K. Elbedour, and E. M. Stone. (1995a). *Am. J. Medical Gentics* 59: 199-203.

Carmi, R., T. Rokhlina, A. E. Kwitekblack, K. Elbedour, D. Nishimura, E. M. Stone, and V. C. Sheffield. (1995b). *Human Mol. Genetics* 4: 9-13.

Chin, J. E., M. Dickens, J. M. Tavare, and R. A. Roth. (1993). *J. Biol. Chem.* 268: 6338-6347.

Cho, K. R. and E. R.Fearon (1995). *Europ. J. Cancer.* 31: 1055-1060.

Christianson, D. W. (1991). *Adv. Protein Chem.* 42: 281-355.

Cooper, H. M., P. Armes, J. Britto, J. Gad, and A. F. Wilks. (1995). *Oncogene* 11:

2243-2254.

Fearon, E. R., K. R. Cho, J. M. Nigro, S. E. Kern, J. W. Simons, J. M. Ruppert, S. R. Hamilton, A. C. Preisinger, G. Thomas, Kinzler, K. W., and Vogelstein, B. (1990). *Science* 247:49-56.

Hedrick, L., K. R. Cho, E. R. Fearon, T. Wu, K. W. Kinzler, and B. Vogelstein. 1994. *Genes & Dev.* 8:1174-1183.

Jouet, M., A. Rosenthal, G. Armstrong, J. MacFarlane, R. Stevenson, J. Paterson, A. Metzenberg, V. Ionasescu, K. Temple, and S. Kenwrick. (1994). *Nature Genetics* 7: 402-407.

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

Korenberg, J. R., and X.- N. Chen. (1996). *Cytogenet. Cell* 69: 196-200.

Kozak, M. (1989). *J. Cell Biol.* 108: 229-241.

Lawlor, K. G., and R. Narayanan. (1992). *Cell Growth & Differentiation* 3: 609-616.

Ledouarin, N. M. (1993). *TINS* 16:64-72.

Marshal, R. D. (1972). *Annu. Rev. Biochem.* 41: 673-707.

Mayer, B. J., and M. J. Eck. (1995). *Current Biol.* 5: 364-367

Needleman, S. B., and C. D. Wunsch. (1970). *J. Mol. Biol.* 48:443-453.

Pierceall, W. E., K. R. Cho, R. H. Getzenberg, M. A.Reale, L. Hedrick, and E. R. Fearon. (1994a). *J. Cell Biol.* 124:1017-1027.

Pierceall, W. E., M. A. Reale, A. F. Candia, C. V. E. Wright, K. R. Cho, and E. R.Fearon. (1994b). *Dev. Biol.* 166: 654-665.

Vaughn, D. E., and P. J. Bjorkman. (1996). *Neuron* 16: 261-273.

Vielmetter, J., J. F. Kayyem, J. M. Roman, and W. J. Dreyer. (1994). *J. Cell Biol.* 127:2009-2020.

Vonheijne, G. (1990). *J. Membrane Biol.* 115: 195-201.

Williams, A. F., and A. N. Barclay. (1988). *Annu. Rev. Immunol.* 6:381-405.

Woodgett, J. R., K. L. Gould, and T. Hunter. (1986). *Eur. J. Biochem.* 161: 177-184.

Yamakawa, K., S. Mitchell, R. Hubert, X.- N. Chen, S. Colbern, Y.- K. Huo, C. Gadomski, U.- J. Kim, and J. R. Korenberg. (1995). *Human Molecular Genetics* 4:709-716.

FIGURE LEGENDS

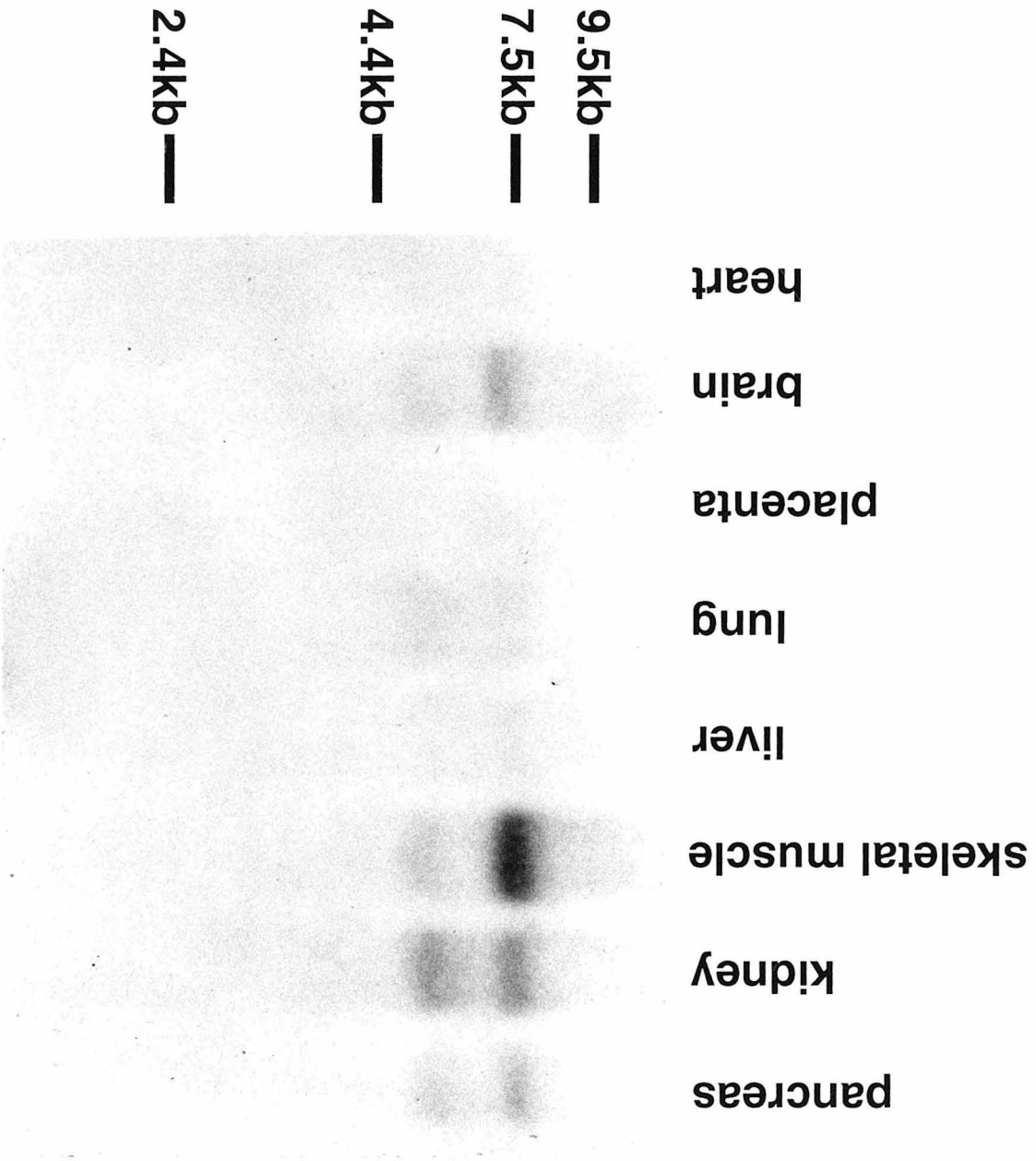
FIG. 1. Nucleotide sequence and deduced amino acid sequence of human neogenin. The open reading frame consists of 1461 amino acids. Significant residues of C in Ig-like domains and W and Y in fibronectin type III like domains are circled. Potential N-glycosylation sites in the intracellular and S or T phosphorylation sites in the extracellular domains are bold and italic. The putative transmembrane domain is boxed and the alternatively spliced exon in the intracellular domain is underlined. The nucleotide sequence has been submitted to GenBank under the accession code U72391.

FIG. 2. Alignment of the amino acids of the human and chicken neogenin proteins (Needleman and Wunsch, 1970). Line brakes indicate the boundaries between domains. The domains corresponding to each sequence portion are symbolized in the cartoon on the right and each domain is assigned a name and number (e.g., Ig-I stands for Ig-like domain I, FNIII-1 stands for fibronectin type III like domain 1). Core residues (C for Ig-lke domains and W and Y for FNIII-like domains) are underlined. Vertical dashes between human and chicken sequences indicate identity, double dots denote similarity, and single

dots indicate remote similarity between amino acids. The overall identity between both proteins is 87%.

FIG. 3. Northern blot analysis. A multiple tissue northern blot was probed with neogenin cDNA. Two mRNA transcripts at 5 and 7 kb are detected in variuos tissues.

FIG. 4. Fluorescence in situ hybridization (FISH). A. Hybridization of fluorescently labeled neogenin plasmid DNA probe reveals the neogenin locus on human chromosome 15q22.3-23. Chromosomes were visualized with reverse band staining using chromomycin A3 and distamycin. B. The ideogram of human chromosome 15 indicates the location of the human neogenin gene.


9.5kb
7.5kb
4.4kb
2.4kb
heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas