ISOLATION AND CHARACTERIZATION OF A cDNA ENCODING A CHICKEN MIDDLE MOLECULAR WEIGHT NEUROFILAMENT PROTEIN (NF-M): EXPRESSION IN THE DEVELOPING EMBRYO

A Thesis

Submitted by

Paul A. Hill

to the Deparment of Cell Biology and Anatomy,

the Oregon Health Sciences University,

Portland, Oregon

in Partial Fullfilment of the Degree of Master of Science

July 1989

APPROVED

[redacted]

.....
(Professor in Charge of Thesis)

[redacted]

.....
(Chairman, Graduate Commitee)

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

ACKNOWLEDGEMENTS

I would like to acknowledge the help and guidance of the following people without which this work would never have been completed. Dr. Gary Ciment, my advisor and mentor. Dr. Serhan Alkan and Dr. Olivier Civelli for their work in preparing and screening the cDNA library, and the Ciment lab group for their support both technical and moral: Dr. Cindy Machida, Dr. Kate Stocker, Larry Sherman, and Sean Rees.

LIST OF ABBREVIATIONS

BCIP	5-Bromo-4-Chloro-3-Indolyl Phosphate
CNS	Central Nervous System
GFAP	Glial Fibrillary Acidic Protein
HMW	High Molecular Weight
IF	Intermediate Filament
IFP	Intermediate Filament Protein
IPTG	Isophenylthiogalactoside
kb	Kilo Base
kD	Kilo Dalton
MAP	Microtubule Associated Protein
NF	Neurofilament
NF-H	High Molecular Weight Neurofilament Protein
NF-L	Low Molecular Weight Neurofilament Protein
NF-M	Middle Molecular Weight Neurofilament Protein

NBT	Nitroblue Tetrazolium
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

LIST OF FIGURES AND TABLES

LIST OF FIGURES AND TABLES (cont.)

Appendix: Page

ABSTRACT

Intermediate filaments form a fibrillar network which makes up a major component of the cytoskeleton in all eukaryotic cells. Despite extensive biochemical and molecular characterization of their subunit proteins, no function has been directly demonstrated for this network. The isolation of the genes encoding these proteins will allow further studies aimed at understanding the function of this large and diverse class of cytoskeletal elements.

We have isolated a cDNA from chicken brain which appears to contain the entire coding region for a middle molecular weight neurofilament protein (NF-M). A lambda-gt11 expression library was prepared from 10 day embryonic chicken brains and screened with a polyclonal antiserum to the dephosphorylated form of chicken NF-M (Bennett et al., 1988). A cDNA of 2.9 kb was isolated and characterized. This cDNA hybridizes to a major transcript of approximately 3.5 kb expressed predominantly in the nervous system. Restriction mapping, DNA sequencing, and genomic southern blotting experiments all indicate that this is a longer version of a previously reported partial cDNA encoding part of the central rod domain and the carboxy terminal domain of the chicken NF-M protein (Zopf et. al., 1987). DNA sequence data and the putative translation product indicate that the cDNA reported here encodes the entire NF-M protein of 852 amino acids with a molecular weight of 93 kD. This protein exhibits considerable divergence from the mammalian protein.

In addition to expression in the nervous system, we have found transient expression of this gene in a number of non-neuronal tissues in the developing embryo. During embryonic days 3 to 10, expression was found in the developing heart (e3 to e8), in the developing limbs (e3 to e10), and in the liver (e9 and e10). These findings raise questions concerning both the function and tissue specificity of NF-M.

I INTRODUCTION

The bulk of the cytoskeleton in eukaryotic cells is composed of three classes of proteins: actin which forms microfilaments, tubulin which forms microtubules, and intermediate filament proteins (IFPs) which form intermediate filaments (IFs). Each of the three classes of proteins is encoded by a family of closely related but divergent genes (Engal et al., 1982; Cowan et al., 1985; Steinart and Parry, 1985). Both the actin and the microtubule networks are known to be quite dynamic and are involved in a number of cellular functions and responses such as cell division and movement. In contrast, the IF network is essentially inert under physiological conditions and no function has been directly demonstrated. Of the three classes, the IF genes show the most variation, but at the same time all IFPs retain a common structural motif (reviewed by Steinart and Parry, 1985). The tight regulation of the IF gene family results in each major cell type producing its own distinct form of IF (Table 1). With the exception of vimentin, the tissue specificity of the different classes of IFs has been found to be absolute.

A diagram of the basic structural motif of the intermediate filament proteins is shown in figure 1A. This consists of a central alpha-helical rod domain flanked by globular carboxy and amino terminal domains. The central rod domain is the most conserved feature of IF proteins. The levels of amino acid sequence identity in this region range from 50% between different IF types to above 90% between the same IF in different species. This domain has a molecular weight of approximately 40 kD for all IF proteins. The end domains are less conserved among different IF types, but do show high levels of conservation (ie: in the 90% range) when the same IFPs are compared from different species. The molecular weights of these domains range from around 5 kD up to 75 kD, and so account for the large variation in molecular weight of the different IF subunits. Since

these regions show the most variation among IF classes, it is assumed that they confer most of the tissue specific characteristics of a given IF type.

The basic subunit of the intermediate filaments (reviewed by Steinart and Parry, 1985) is thought to be a tetramer composed of a pair of dimers (see figure 1B). The dimer is composed of either two identical subunits (ie, for vimentin, desmin, GFAP) or in the case of the keratins, two different subunits (one of the type I and one of the type II keratins). The dimers have been proposed to form via coiled-coil interactions between the alpha-helical rod domains. Two dimers then form the tetramer, which polymerizes into a filament also with an alpha helical arrangement.

There are three pieces of evidence to suggest that the amino terminal domain is important in filament formation. Although the rod domain appears to be required for dimerization, this region cannot form filaments alone. This has been shown by producing a proteolytic fragment of desmin which lacks the end domains (Geisler et al., 1982). This fragment will not form filaments in vitro as will the intact molecule, but instead forms only short protofilaments. The notion that the end domains are important in filament formation is supported by additional studies in which the amino terminal end of vimentin was altered either by proteolytic digestion or chemical modification of arginine residues (Traub and Vorgias, 1983, 1984). Both modifications block polymerization in vitro. Finally, a monoclonal antibody which binds to the amino terminal domain of desmin will block in vitro polymerization completely when used at high concentrations and also causes the formation of unraveled and capped filaments at lower concentrations (Ip, 1984). The distribution of this antibody over the disrupted filaments indicates that the amino terminal domain is buried within the filament as polymerization takes place. These experiments all suggest that the amino terminal domains are important in filament formation.

As noted above, the function of intermediate filaments is not clear. The majority of the information regarding IFs has been gathered through microscopic analysis, such as studies examining the arrangement of filaments within the cytoplasm, the structures which

IFs associate with, and when the different types appear during development (Lazarides, 1980; Tapscott et al., 1981; Bennett et al., 1978). These types of observations as well as the extreme insolubility of IFs have led to models which suggest a role in determining cytoplasmic structure and cell morphology (Lazarides, 1980; Steinart and Parry, 1985). However, studies to date which have addressed IF function directly by disrupting the IF network have produced negative results. In these experiments there were no apparent changes in the shape or other properties of cells which had been injected with anti-IF antibodies to collapse the IF network, thus giving no indication of the role IFs play in the cell (Klymkowsky et al., 1981, 1983; Lin and Feramisco, 1981). This apparent lack of function and requirement for tissue specific diversity is unexplained, and is in contrast to the simple structural roles which are often proposed (e.g. cell shape). The precise biological function of IFs and the reason for generation of this diverse but conserved gene family is therefore unclear at this point.

Neurofilaments (NFs) are the intermediate filaments found specifically in neurons and have a number of characteristics which distinguish them from the other IF classes. In vertebrates these filaments are composed of three major subunits having molecular weights of approximately 70 kD (NF-L), 150 kD (NF-M), and 200 kD (NF-H) (Liem et. al., 1978). The genes or cDNAs for these proteins have been isolated and characterized for a number of species (Napolatino et al., 1987; Mack et al., 1988; Levy et al., 1987). From these studies it is apparent that the true molecular weights for these proteins are considerably less than those determined from migration on SDS-PAGE. The NF-M proteins are predicted to have a molecular weight in the range of 95 kD. The characterization of the mouse NF genes has indicated that this class of genes is distinct from the other classes of IFs at the level of intron placement and number (Levy et al., 1987; Lewis and Cowan, 1986). This unique gene structure has led Levy et al. to propose that the NF genes arose from the other IF genes by a retroviral transposition as a relatively recent event. On the protein level a uniquely high level of acidic residues has been reported for the carboxy terminal

domain of all three NF subunits for a number of species (Napolatino et al., 1987; Mack et al., 1988; Levy et al., 1987; Zopf et al., 1987; Geisler, et al., 1984). In each case this region has been reported to be composed of more than 50% acidic residues, the majority of which are glutamic acid. This region is not found in the other IF proteins and no function has been proposed for this distribution. As with the other IF components, the NF components seem to be regulated on a number of levels including tissue specificity, developmental stage, and extent of phosphorylation (Pachter and Liem, 1984: Carden et al., 1987; Shaw and Weber, 1982).

Immunoelectron microscopy has shown the arrangement of the subunits within neurofilaments to be such that NF-L forms the core of the filament. NF-M is found more randomly distributed surrounding the core, and NF-H appears between filaments, possibly as a crosslinker (Hirokawa et al., 1984). In contrast, desmin, vimentin, and glial filaments are all composed of a single major subunit, while the keratins absolutely require two distinct subunit types for polymerization (Hatzfeld and Franke, 1985).

In the chicken, immunocytochemical studies indicate that NF-M displays a distinctive pattern of expression when compared with the other NF proteins (Bennett and Dilullo, 1985). In the developing embryo NF-H is expressed only as neurons terminally differentiate, and although NF-L is expressed somewhat earlier, it has a similiar pattern of expression. In contrast, NF-M is expressed in a distinctive pattern in the CNS a full 48 hours before any neurons become postmitotic. In this same work NF-M expression was reported in a number of non-neuronal tissues. This pattern of expression contrasts with the structural and morphological roles which have been proposed (Hoffman et al., 1984). The polymerization studies mentioned above have led some to consider NF-M and NF-H as intermediate filament associated proteins (Pachter and Liem, 1984), since they are not required to form filaments. It has been demonstrated that the NF triplet, and NF-M in particular, will interact with microtubules via microtubule associated proteins (HMW MAPs) (Leterrier et al., 1981, 1982). This has led to the proposal that NF-M may be a point of in-

teraction between the NFs and other organelles. However, no such function has been directly demonstrated.

As noted above the DNA sequences for a number of mammalian NF proteins have been reported. By isolating the homologous gene for the chicken we will gain two advantages. First, from a developmental point of view the chicken embryo has been particularly well characterized and can be manipulated with greater ease than the mammalian systems. This is true in general, but even more so during the very early stages of development. Therefore, the developmental role and function of these proteins should be much more amenable to study in the chicken. Second, this will allow comparative studies of structure and function. In other words, if there are differences between the chicken and mammalian proteins one may be able to correlate structural and functional changes and so gain indications of structure/function relationships for various domains of this class of proteins. In the present work we have isolated a cDNA encoding a middle molecular weight NF protein. By studying its expression in the developing embryo we hope to gain further insight into the regulation and function of this large and diverse component of the eukaryotic cytoskeleton.

References

Bennett, G., Fellini, S., Croop, J., Otto, J. Byran, J, Holtzer, H. (1978): Differences Among 100A-Filaments From Different Cell Types. Proc. Nat. Acad. Sci. USA 75:4364-4368.

Bennett, G., Dilullo, C. (1985): Transient Expression of a Neurofilament Protein by Replicating Neuroepithelial Cells of the Embryonic Chick Brain. Dev. Biol. 107:107-127.

Bennett G., Hollander, B., Laskowska, D. (1988): Expression and Phosphorylation of the Mid-Sized Neurofilament Protein NF-M During Chick Spinal Cord Neurogenisis. J. Neurosci. Res. 21:376-390

Carden, M., Trojanowski, J., Schlaepfer, Lee, V. (1987): Two Stage Expression of Neurofilament Polypeptides During Rat Neurogenesis with Early Establishment of Adult Phosphorylation Patterns. J. Neurosci. 7:3489-3504.

Cowan, N. J. (1985): Tubulin Genes and the Diversity of Microtubule Function, p.36-60. In McLean, N., (ed.), Oxford Surveys on Eukaryotic Genes, vol 1, Oxford University Press, Oxford.

Engal, J., Gunning, P., Kedes, L. (1982): Human Actin Proteins are Encoded by a Multigene Family, p. 107-118. in Pearson, M., Epstein, H., (eds.), Muscle Development: Molecular and Cellular Control. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Geisler, N., Kaufmann. E., Weber, E. (1982): Proteinchemical Characterization of Three Structurally Distinct Domains Along the Protofilament of Desmin 10nm Filaments. Cell 30:277-286.

Geisler, N., Fischer, S., Vanderckhove, J., Plessmann, Weber, E. (1984): Hybrid Character of a Large Neurofilament Protein (NF-M): Intermediate Filament Type Sequence Followed by a Long and Acidic Carboxy-Terminal Extension. EMBO J. 3:2701-2706.

Hatzfeld, M., Franke, W. (1985): Pair Formation and Promiscuity of Cytokeratins: Formation In Vitro of Heterotypic Complexes and Intermediate Sized Filaments by Homologous and Heterologous Recombinations of Purified Polypeptides. J. Cell Biol. 101:1826-1841.

Hirokawa, N., Glicksman, M., Willard, M. (1984): Organization of Mammalian Neurofilament Polypeptides within the Neuronal Cytoskeleton. J. Cell Biol. :1523-1536

Hoffman, P., Griffen, J., Price, D. (1984): Control of Axonal Caliber by Neurofilament Transport. J. Cell Biol. 99:705-715.

Ip, W. (1988): Modulation of Desmin Filament Assembly By a Monoclonal Antibody. J. Cell Biol. 106:735-745.

Klymkowsky, M. (1981): Intermediate Filaments in 3T3 Cells Collapse After Intracellular Injection of a Monoclonal Anti-Intermediate Filament Antibody. Nature 291:249-251.

Klymkowsky, M., Miller, R., Lane, E. (1983): Morphology, Behavior, and Interaction of Cultured Epithelial Cells After the Antibody-Induced Disruption of Keratin Filament Organization. J. Cell Biol. :494-509.

Krohne, G., Wolin, S., McKeon, F., Franke, W., Kirschner, M. (1987): Nuclear Lamin L1 of Xenopus Laevis: cDNA Cloning, Amino Acid Sequence and Binding Specificity of a Mem-

ber of the Lamin B Subfamily. EMBO J. 6:3801-3808.

Lazerides, E. (1980): Intermediate Filaments as Mechanical Integrators of Cellular Space. Nature 283:249-256.

Leterrier, J., Liem, R., Shelanski, M. (1981): Preferential Phosphorylation of the 150,000 Molecular Weight Component of Neurofilament by a Cyclic AMP-Dependent, Microtubule-Associated Protein Kinase. J. Cell Biol. 90:755-760.

Leterrier, J., Liem, R., Shelanski, M. (1982): Interaction Between Neurofilament and Microtubule-Associated proteins: A Possible Mechanism for Intraorganellar Bridging. J. Cell Biol. 95:982-986.

Lewis, S., Cowan, N. (1986): Anomalous Placement of Introns in a Member of the Intermediate Filament Multigene Family: an Evolutionary Condrum. Mol. Cell. Biol. 6:1529-1534.

Levy, E., Liem, R., D'Eustachio, P., Cowan, N. (1987): Structure and Evolutionary Origin of The Gene Encoding Mouse NF-M, the Middle-Molecular-Mass Neurofilament Protein. FEBS 166:71-77.

Liem, R., Yen, S., Salomon, G., Shelanski, M. (1978): Intermediate Filaments In nervous Tissues. J. Cell Biol. 79:637-645.

Lin, J., Feramisco, J. (1981): Disruption of the In Vivo Distribution of the Intermediate Filaments in Fibroblasts Through the Microinjection of a Specific Monoclonal Antibody. Cell 24:185-193.

Mack, K., Currie, J., Soifer, D. (1988): A cDNA Coding for the Tail Region of the High Molecular Weight Rabbit Neurofilament Protein NF-H. J. Neurosci. Res. 20:129-136.

McKeon, F., Kirschner, M., Caput, D. (1986): Homologies in Both Primary and Secondary Structure Between Nuclear Envelope and Intermediate Filament Proteins. Nature 319:463-468.

Napolitano, E., Chin, S., Colmman, D., Liem, R. (1987): Complete Amino Acid Sequence and In Vitro Expression of Rat NF-M, The Middle Molecular Weight Neurofilament Protein. J. Neurosci. 7:2590-2599.

Pachter, J., Liem, R. (1984): The Differential Appearance of Neurofilament Triplet Polypeptides in the Developing Rat Optic Nerve. Dev. Biol. 103:200-210.

Shaw, G., Weber, K. (1982): Differential Expression of Neurofilament Triplet Proteins in Brain Development. Nature 298:277-279.

Steinart, P., Parry, D. (1985): Intermediate Filaments: Conformity and Diversity of Expression and Structure. Ann. Rev. Cell Biol. 1:41-65.

Tapscott, S., Bennett, G., Toyama, Y., Kleinbart, F., Holtzer, H. (1981): Intermediate Filament Proteins in the Developing Chick Spinal Cord. Dev. Biol. 86:40-54.

Traub, P., Vorgias, C. (1983): Involvment of the N-Terminal Polypeptide of Vimentin in the Formation of Intermediate Filaments. J. Cell Sci. 63:43-67.

Traub, P., Vorgias, C. (1984): Differential Effect of Arginine Modification with 1,2-

Cyclohexanedione on the Capacity of Vimentin and Desmin to Assemble into Intermediate Filaments and to Bind Nucleic Acids. J. Cell Sci. 65:1-20.

Zopf, D., Hermans-Borgmeyer, I., Gundelfinger, E., Betz, H. (1987): Identification of Gene Products Expressed in the Developing Chick Visual System: Characterization of A Middle Molecular Weight Neurofilament cDNA. Genes and Development 1:699-708.

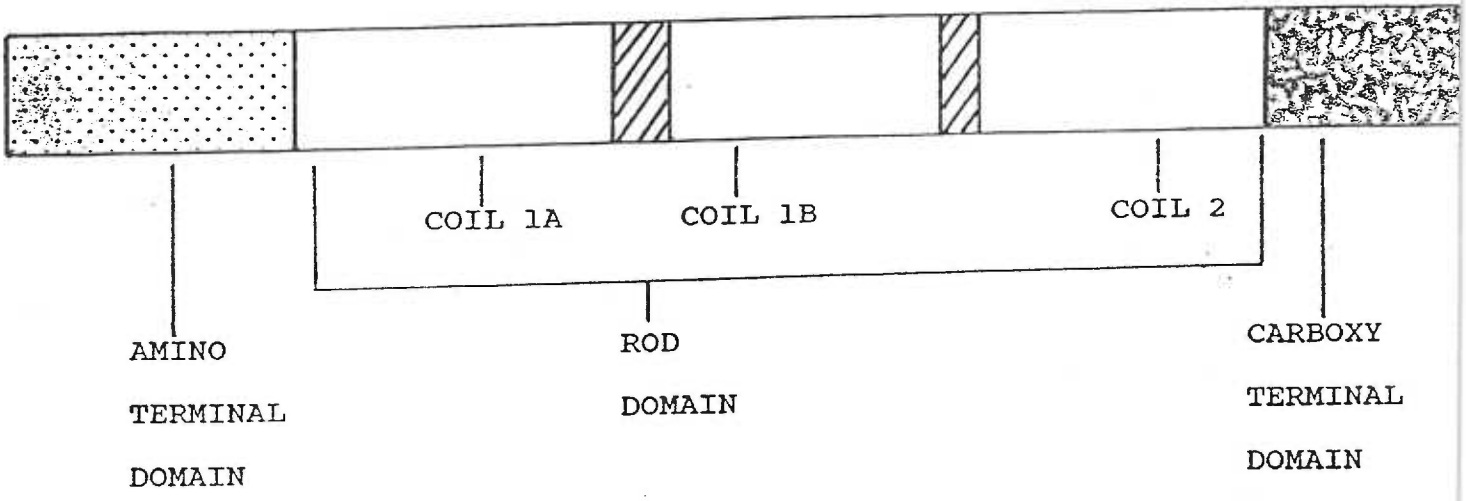
TABLE I: **Intermediate Filament Classification** - Intermediate filaments are grouped based on their primary sequence similarity (reveiwed by Steinart and Parry, 1985) The nuclear lamins have recently been found to share high levels of sequence homology with the IF genes and therefore have been proposed as a new class of intermediate filament proteins (Mckeon et. al., 1986; Krohne et.al., 1987).

TABLE I : INTERMEDIATE FILAMENT CLASSIFICATION

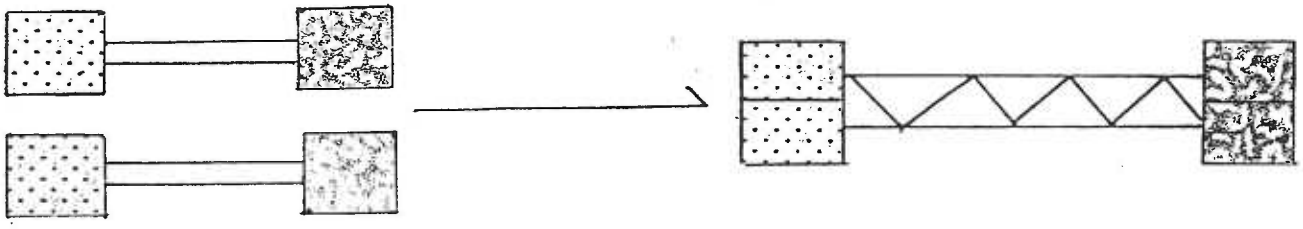
Class	Filament Type	Cell Type
I	Keratins	Epithelial
II	Keratins	Epithelial
III	Vimentin	Mesenchymal, Embryonic, Many cell types in culture
	Desmin	Muscle
	Glial Fibrillary Acidic Protein	Glia
IV	Neurofilaments	Neurons
V	Nuclear Lamins	-------

FIGURE 1: **Structural Domains and Polymerization of Intermediate Filament Proteins** - A schematic of the basic secondary structural domains is shown in (A). The cross hatched areas represent the non-alpha-helical hinge regions between the coil domains. Filament formation is shown in (B). Dimerization is thought to occur via interaction of the rod domains of the two subunits, which forms a higher order alpha-helical structure. The dimers then form a tetramer which polymerizes to form the filament. The orientation of the monomers in the dimer is thought to be such that the carboxy and amino terminal domains of one subunit interact with the same of the other (reveiwed by Steinart and Parry, 1985).

(A) STRUCTURAL DOMAINS OF INTERMEDIATE FILAMENT PROTEINS



(B) POLYMERIZAION OF INTERMEDIATE FILAMENT PROTEINS



MONOMERS

DIMERS

(ALPHA-HELICAL COILED-COIL)

100Å Filament

II MANUSCRIPT

ABSTRACT: We have isolated a cDNA from chicken brain which appears to contain the entire coding region for a middle molecular weight neurofilament protein (NF-M). A lambda-gt11 expression library was prepared from 10 day embryonic chicken brains and screened with a polyclonal antiserum to the dephosphorylated form of chicken NF-M (Bennett et al., 1988). A cDNA of 2.9 kb was isolated and characterized. This cDNA hybridizes to a major transcript of approximately 3.5 kb that is expressed predominantly in the nervous system. Restriction mapping, DNA sequencing, and genomic southern blotting experiments all indicate that this is a longer version of a previously reported partial cDNA encoding part of the central rod domain and the carboxy terminal domain of the chicken NF-M protein (Zopf et al., 1987). DNA sequence data and the putative translation product indicate that the cDNA reported here encodes the entire NF-M protein of 852 amino acids with a molecular weight of 93 kd. This protein exhibits considerable divergence from the mammalian protein.

In addition to expression in the nervous system, we have found transient expression of this gene in a number of non-neuronal tissues in the developing embryo. During embryonic days 3 to 10, expression was found in the developing heart (e3 to e8), in the developing limbs (e3 to e10), and in the liver (e9 and e10). These findings raise questions concerning both the function and tissue specificity of NF-M.

INTRODUCTION: Intermediate filaments (IFs) make up a major portion of the cytoskeleton in eukaryotic cells. The subunits which make up these filaments are encoded by a multigene family which is regulated to produce a distinct form of IF in each major cell type (reviewed by Steinart and Parry, 1985). At the same time these genes are conserved to maintain a common structural motif in all IFPs, which consists of a 40 kD alpha-helical rod domain flanked by variable amino and carboxy terminal domains.

Neurofilaments represent a distinct class of intermediate filaments and in mammals are composed of three subunits designated NF-L , NF-M, and NF-H based on their respective molecular weights of 70, 150 and 200 kD as estimated by SDS-PAGE (Liem et al., 1978; Hoffman and Lasek, 1975). However, these values have proven to be anomalously high and the true molecular weight of mammalian NF-Ms has been reported to be in the range of 95 kd (Napolitano et al., 1987; Levy et al., 1987). Immunocytochemical studies show that NF-L forms the core of the filament with NF-H possibly functioning to cross-link filaments, and NF-M is found randomly distributed along the NF-L core (Hirokawa et al., 1984). The role of NF-M in filament formation, if any, is not known. There is some evidence that NF-M may function as a link to other organelles such as microtubules (Leterrier, et al., 1981; 1982). The complete amino acid sequences for NF-Ms from a number of species have been determined and found to be quite highly conserved (Napolitano et al., 1987; Levy et al., 1987; Geisler and Weber, 1984). In all cases the expression of these genes is specifically localized to the nervous system. We have isolated a cDNA clone which appears to encode the entire chicken NF-M protein of 852 amino acids and 93 kD molecular weight. All data reported here are in agreement with a previously published report of a cDNA encoding the carboxy terminal half of the chicken NF-M protein (Zopf et al., 1987). In addition, we have found expression of this gene in a number of non-neuronal tissues in the developing embryo. Our results are in agreement with the work of Bennett and Dilullo (1985) in which they have reported NF-M immunoreactivity in the embryonic limb and heart of the developing chicken. These results raise questions as to the function and regulation of this protein and of the IF genes in general.

MATERIALS and METHODS:

Animals: Fertile chicken eggs (Gallus domesticus, White Leghorn variety) were purchased from the Poultry Sciences Dept., Oregon State University, Corvallis, OR. Eggs were

incubated at 38°C for the specified times. Zero time of incubation is when the eggs were placed in the incubator. Embryos were staged according to Hamburger and Hamilton (1951).

cDNA Library - Preparation and screening: Total RNA was isolated from 10 day embryonic chicken brain using the guanidinium-CsCl method and the poly A+ fraction was collected by oligo dT cellulose chromatography, as described by Maniatis et al. (1982). The poly A+ RNA was used to construct a cDNA library using a commercially available kit (Amersham corp.) following published methods (Davis et al., 1986). The resulting cDNAs were ligated to EcoRI linkers then packaged into lambda gt-11 using a commercially available kit (Stratagene).

The library was screened by first plating the bacteriophage on E. coli strain Y1090 at 42°C and incubating for 3 hours. The plates were then overlaid with an isophenylthiogalactoside (IPTG) soaked nitrocellulose filter (10mm IPTG in water followed by drying) and incubated at 37°C for an additional 3 hours. The filters were rinsed in TBST buffer (10mm Tris pH 8, 150mM NaCl, 0.05% Tween 20), and the dried filters were incubated in TBST with a rabbit polyclonal antiserum to dephosphorylated chicken NF-M (a gift from G. Bennett (1988)) followed by an alkaline phosphatase conjugated goat anti-rabbit IgG (Promega Inc.). The filters were then incubated in developing buffer (NBT and BCIP, 2:1 (v/v), (Promega)) until color appeared. Approximately 100,000 plaques were screened resulting in the isolation of four positive clones having inserts of 1.2, 2.0, 2.5, and 2.9 kb. The largest insert was selected for further characterization and ligated into the plasmid vector pBR322. This construct was designated p19NFM.

Northern and Southern blotting: For northern blotting, tissues were dissected from developing chick embryos and total RNA was isolated using the LiCl precipitation method (Auffray and Rougen, 1980). Samples of 10 to 20 ug were subjected to electrophoresis in

denaturing agarose gels (1.2% formaldehyde) and transferred to nylon membranes by capillary action. Blots were probed with the radiolabeled p19NFM insert, which had been labeled to specific activities of approximately 10^8cpm/ug by nick translation. Hybridization was carried out at 42°C in 5xSSC, 50% formamide, 1% SDS, 1X Denhardt's (see Maniatis et al., 1982), and 200ug/ml salmon sperm DNA. The final activity during hybridization was 10^6cpm/ml. Washes were at a final stringency of 55°C in 0.1X SSC. For Southern blots genomic DNA was isolated from 15 day embryonic chicken liver according to Maniatis et al. (1982). Samples of 20ug were digested with restriction enzymes, subjected to electrophoresis in 1% agarose gels, and transferred to nylon membranes by capillary action. The blots were probed with the radiolabeled p19NFM insert essentially as described for northern blotting above.

DNA Sequencing: Subfragments of the p19NFM insert were subcloned into M13 vectors and sequenced using the dideoxy chain termination method of Sanger et al. (1977).

RESULTS:

The restriction map of the p19NFM insert (designated c19NFM) is shown in figure 1B. As indicated this map aligns with the map of a partial chicken NF-M cDNA published by Zopf et.al. (1987) shown in figure 1A, but c19NFM extends an additional 1,034 bases pairs in the 5' direction. The DNA sequencing strategy is also given in figure 1B. The regions sequenced in the present work which overlap with the published sequence are also shown. Within these regions, there is complete agreement between c19NFM and the published sequence.

Figure 2 shows a composite of the entire coding region for the chicken NF-M protein (ie: the sequence published by Zopf et al. (1987) and the additional sequence data reported here) and the putative translation product of 852 amino acids. The resulting protein is compared in figure 3 with the sequence reported for the entire rat NF-M protein

(Napolitano et al., 1987). The alignment of the protein with the rat sequence, the total molecular weight of 93 kd, and a comparison of the nucleotide sequence surrounding the proposed initiator methionine codon with the consensus sequence for this region published by Kozak (1986), all suggest that c19NF-M encodes the entire chicken protein.

An alignment of the chicken and rat sequences at the DNA level (not shown) indicates 68% identity for the entire gene. The coding region within the conserved rod domain exhibits a value of 74% when analyzed separately. The sequences align quite well within the coding regions for the amino-terminal and rod domains with very few gaps resulting. However, the alignment within the carboxy-terminal domains is much less complete with large and frequent gaps. When the putative chicken protein is aligned with the rat protein, the level of identity is considerably lower and a suprising amount of divergence is evident. Comparison of the chicken and rat alpha-helical rod subdomains yields values from 36% to 76% identity. For the protein overall, 52% of the amino acids are identical in the alignment. These data are summarized in table I. There are two inconsistencies here. First, the homology at the nucleotide level would not be expected to be higher than at the amino acid level for a highly conserved protein. Second, the rod domains would be expected to have a higher level of homology than the protein as a whole. The most obvious explanation for these results is that the gene encoding c19NFM has diverged to a considerable extent from the common ancestral gene of mammals and birds. This means that either avian NF-M is quite different from the mammalian version or that birds have developed a second distinct version of this gene which encodes a novel neurofilament protein.

The tissue specificity and message size are demonstrated in figure 4. This cDNA hybridizes to a message of approximately 3.5 kb expressed predominately in the nervous system in the 10 day embryonic chicken. As a further characterization genomic Southern blotting was carried out. The result is shown in figure 5 and is in agreement with the previously reported NF-M cDNA (Zopf et al., 1987). In data not shown we have seen

expression of this message in both the neural tube and developing heart of 2 day embryos, while no expession is seen in the somites. To investigate the non-neuronal expression of c19NFM we have carried out a time course on the heart and various other tissues ranging from embryonic day 3 to embryonic day 10 of incubation. The results shown in figure 6 are suprising in that not only is c19NFM expressed in the developing heart from e3 to e8 (stages 19 to 34 of Hamburger and Hamilton) but also in the developing limbs from e4 to e10 (stages 24 to 36 of Hamburger and Hamilton), and in the liver on e9 and e10 (stages 35 and 36 of Hamburger and Hamilton). There also appears to be additional high molecular weight bands present in all positive lanes. However, these bands may represent unprocessed transcripts because all blots shown here were prepared with total RNA.

DISCUSSION: We report here the isolation and characterization of what we believe to be a full length cDNA encoding a chicken NF-M protein. Our results indicate we have cloned a longer version of the cDNA reported by Zopf et al. (1987). These workers have concluded that this is a NF-M clone based on the specific expression in the nervous system of the newly hatched chicken, the high level of acidic amino acids in the carboxy terminal domain characteristic of the NF proteins (Levy et al., 1987; Napolitano et al., 1987; Mack et al., 1988; Geisler et al., 1982), and the estimated molecular weight which is too low for NF-H and too high for NF-L. Additional evidence that this is the chicken NF-M gene is reported here in that the nucleotide sequence aligns with that of the rat gene and exhibits a high level of homology (alignment not shown), the putative translation product has a molecular weight in the expected range for NF-M (93kD), and the putative protein aligns overall (ie: the amino and carboxy terminal ends begin and end in alignment) with the rat protein. This cDNA appears to encode the entire protein based on the overall alignment (as defined above) with the rat sequence and by comparison of the nucleotide sequence surrounding the initiator methionine with that of the consensus sequence reported by Kozak

et al. (1986). This comparison gives a match in all positions except the G residue at position +4.

In contrast to the above results the alignment of the chick and rat proteins yields unexpectedly low levels of identity. These lie in the range of 50%. This value is what would be expected for a comparison of different classes of IFs. There are three possible explanations for this result. Either the avian and mammalian proteins have diverged considerably, or the gene from which the cDNA characterized here and by Zopf et al. (1987) is a novel form of NF-M which has arisen separately or diverged from the original chicken NF-M gene. Any of these possibilities is consistent with the unexpected expression of this gene in non-neuronal tissues discussed below, because this appears to be a divergent form of NF-M (when compared with the mammalian version) and would be expected to have distinct properties.

In addition to the nervous system we have found this gene to be expressed in the developing heart on days e2 to e8, the limbs on days e4 to e10, and the liver on days e9 and e10. Zopf et al. reported that this gene was expressed only in the nervous system of the post hatching (day P1) and adult chicken, meaning that the non-neuronal expression reported here can be considered transient and embryonic. For a number of reasons the expression of c19NF-M in these regions does not appear to be in neuronal cells. All of the elements of the peripheral nervous system are derived from the neural crest (see LeDouarin, 1982). This group of cells migrates from the developing neural tube to specific locations and differentiates to form a number of cell types. The expression of this gene in both the heart and the limb buds occurs before any of the neural crest cells will have migrated to these regions and differentiated. For example crest cells would not be expected to be in the vicinity if the heart until around the middle of day 2 of incubation (stage 12 of Hamburger and Hamilton) and would still be undifferentiated at this point. Therefore, at these early time points the cells expressing c19NF-M cannot be derived from the neural ectoderm. This conclusion is supported by the work of Bennett and Dilullo (1985) in which they

demonstrated NF-M immunoreactivity in both of these locations and were able to confirm the non-neuronal nature of NF-M-positive cells in the developing heart. Even at the later time points when the peripheral nervous system has formed none of these locations would be expected to contain enough neuronal cell bodies to account for the results.

If these cells are not from a neuronal lineage, then the expression of a neuronal marker could indicate that there are populations of cells in these areas which exhibit some sort of neuronal character. In the heart an obvious candidate is the conduction system. In support of this there has been a report of expression of neuronal markers (including NF-M) in the purkinje fibers of the rabbit heart (Gorza et al., 1988). However, there are no obvious candidates in the liver and limb, arguing against this explanation (see Patten, B., 1971).

One characteristic that these tissues do share is that the expression of this gene is restricted to the embryo. From this point of view the function of this gene could be one associated with the less differentiated state (as with the expression of vimentin by embryonic tissues) rather than a neuronal character. Other explanations are that the chicken NF-M gene simply is not as tissue specific as the mammalian version, or that IF genes in general are not regulated as tightly as has been believed. From any point of view these results raise interesting questions concerning the function and regulation of avian NF-M in particular, and of IFPs in general.

References

Auffray, C., Rougen, F. (1980): Purification of Mouse Immunoglobin Heavy Chain mRNAs. European J. Biochem. 107:303-307.

Bennett, G., Dilullo, C. (1985): Transient Expression of a Neurofilament Protein by Replicating Neuroepithelial Cells of the Embryonic Chick Brain. Dev. Biol. 107:107-127.

Bennett G., Hollander, B., Laskowska, D. (1988): Expression and Phosphorylation of the Mid-Sized Neurofilament Protein NF-M During Chick Spinal Cord Neurogenisis. J. Neurosci. Res. 21:376-390.

Davis, L., Dibner, M., Battey, J. (1986): Basic Methods in Molecular Biology. Elselvier Science, N. Y.

Geisler, N., Kaufmann. E., Weber, E. (1982): Proteinchemical Characterization of Three Structurally Distinct Domains Along the Protofilament of Desmin 10nm Filaments. Cell 30:277-286.

Geisler, N., Fischer, S., Vanderckhove, J., Plessmann, Weber, K. (1984): Hybrid Character of a Large Neurofilament Protein (NF-M): Intermediate Filament Type Sequence Followed by a Long and Acidic Carboxy-Terminal Extension. EMBO J. 3:2701-2706.

Gorza, L., Schiaffino, S., Vitadello, M. (1988): Heart Conduction System: A Neural Crest Derivative? Brain Res. 457:360-366.

Hamburger, V., Hamilton, H. (1951): A Series of Normal Stages in the Development of the

Chick Embryo. J. Morphology 88:49-92.

Hirokawa, N., Glicksman, M., Willard, M. (1984): Organization of Mammalian Neurofilament Polypeptides within the Neuronal Cytoskeleton. J. Cell Biol. :1523-1536

Hoffman, P., Lasek, R. (1975): The Slow Component of Axonal Transport. Identification of Major Structural Polypeptides of the Axon and Their Generality Among Mammalian Neurons. J. Cell Biol. 66:351-366.

Kozak, M. (1986): Point Mutations Define a Sequence Flanking the AUG Initiator Codon that Modulates Translation by Eukaryotic Ribosomes. Cell 44:283-292.

LeDouarin, N. (1982): The Neural Crest, Barlow, P., Green, P., Wylie, C., (eds.), Cambridge Univ. Press, Cambridge.

Leterrier, J., Liem, R., Shelanski, M. (1981): Preferential Phosphorylation of the 150,000 Molecular Weight Component of Neurofilament by a Cyclic AMP-Dependent, Microtubule-Associated Protein Kinase. J. Cell Biol. 90:755-760.

Leterrier, J., Liem, R., Shelanski, M. (1982): Interaction Between Neurofilament and Microtubule-Associated Proteins: A Possible Mechanism for Intraorganellar Bridging. J. Cell Biol. 95:982-986.

Liem, R., Yen, S., Salomon, G., Shelanski, M. (1978): Intermediate Filaments In Nervous Tissues. J. Cell Biol. 79:637-645.

Levy, E., Liem, R., D'Eustachio, P., Cowan, N. (1987): Structure and Evolutionary Origin of

the Gene Encoding Mouse NF-M, the Middle-Molecular-Mass Neurofilament Protein. FEBS 166:71-77.

Mack, K., Currie, J., Soifer, D. (1988): A cDNA Coding for the Tail Region of the High Molecular Weight Rabbit Neurofilament Protein NF-H. J of Neurosci. Res. 20:129-136.

Maniatis, T., Fritsch, E., Sambrook, J. (1982): Molecular Cloning-A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Napolitano, E., Chin, S., Colmman, D., Liem, R. (1987): Complete Amino Acid Sequence and In Vitro Expression of Rat NF-M, The Middle Molecular Weight Neurofilament Protein. J. Neurosci. 7:2590-2599.

Patten, B. (1971): Early Embryology of The Chick. McGraw-Hill, N. Y.

Sanger, F., Nicklen, S., Coulsen, A. (1977): Sequencing with Chain Terminating Inhibitors. Proc. Nat. Acad. Sci. USA 74:5463-5467.

Steinart, P., Parry, D. (1985): Intermediate Filaments: Conformity and Diversity of Expression and Structure. Ann. Rev. Cell Biol. 1:41-65.

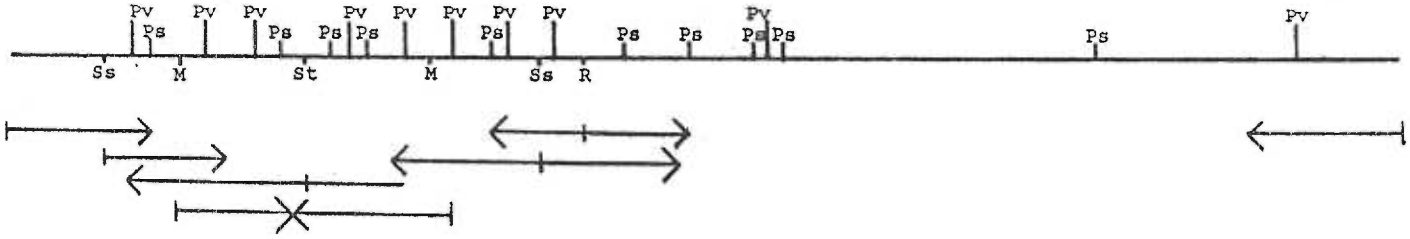
Zopf, D., Hermans-Borgmeyer, I., Gundelfinger, E., Betz, H. (1987): Identification of Gene Products Expressed in the Developing Chick Visual System: Characterization of A Middle Molecular Weight Neurofilament cDNA. Genes and Development 1:699-708.

FIGURE 1: **Restriction Map and Sequencing Strategy** - A map of the cDNA reported by Zopf et. al. (1987) is shown in (A). The Map of c19NFM showing the fragments and direction sequenced (arrows) is shown in (B). c19NFM contains an additional 1,034 bp at the 5' end and lacks 200bp of the 3' untranslated region when compared with the sequence reported in (A). Restriction enzymes are: MspI (M), PstI (Ps), PvuII (PV), EcoRV (R), SstI (Ss), and StuI (St).

(A)


Pv
Ps
Ps
Pv
Ps
Ps
Ps
Pv
Ss
R

(B)


Pv
Ps
Pv
Pv
Ps
Pv
Ps
Ps
Pv
Pv
Ps
Pv
Pv
Ps
Ps
Pv
Ps
Ps
Pv
Ss
M
St
M
Ss
R

⊢ = 100 bp

FIGURE 2: **Complete Nucleotide and Amino Acid Sequence for the Chicken NF-M Protein** - Shown is a composite of the sequence reported by Zopf et. al. (1987) with the additional sequence data reported here. The published sequence begins with amino acid 344 and is marked by the arrow. There are only 5 bp of 5' untranslated sequence and 400 bp of the 3' untranslated sequence in c19NFM. Also shown here is the additional 200 bp of 3' untranslated sequence previously published.

```
CCGCC ATG AGC TAC TCC ATG GAG CCC TTG GGA ACC CCT CGT ACC GCA GGT GAT GAC CGA
      M   S   Y   S   M   E   P   L   G   T   P   R   T   A   G   D   D   R
                                            10

GAC CCG CGC CAC GTA CAG CCG GGC TCA GCG CGT CCC CGT CCA GCG GCT TCC GTC
D   P   R   H   V   Q   P   G   S   A   R   P   R   P   A   A   S   V
    20                                        30

GCA GTC GTG TCG CGG GGC TCC GGC AGC ACC GTG TCC TCG TCC TAC AAA CGC ACC
A   V   V   S   R   G   S   G   S   T   V   S   S   S   Y   K   R   T
            40                                    50

AAC CTC GGG GCT CCG CGA ACC GCC TAC GGC TCC ACC GTG CTG AGC TCG CAG AGA
N   L   G   A   P   R   T   A   Y   G   S   T   V   L   S   S   Q   R
                    60                                70

CCT CGA CGT GAG CCA GTC CTG CTG CTG AAC GGG GCG GCC GAG CTG AAG CTG AGC
P   R   R   E   P   V   L   L   L   N   G   A   A   E   L   K   L   S
                        80                                    90

CGC TCC AAC GAG AAG GAG CAG CTG CAG GGG CTG AAC GAC CGT TTG CGG GGT ACA
R   S   N   E   K   E   Q   L   Q   G   L   N   D   R   L   R   G   T
                            100

TCG AGA AGG TGC ATT ACC TGG AGC AGC AGA ACA AGA GAT CAG CGA GCT GCG CGC
S   R   R   C   I   T   W   S   S   R   T   R   D   Q   R   A   A   R
    110                                   120

TGC GCA GAA GCA CGC CGG GAG GGC TCA GCT GGG GGA CGC CTA CGA GCA GGA GCT
C   A   E   A   R   R   E   G   S   A   G   G   R   L   R   A   G   A
            130                                   140

GCG GGA GCT GCG CGG GGC CCT GGA GCA GGT GAG CCA CGA GAA AGG CGC AGA TTC
A   G   A   A   R   G   P   G   A   G   E   P   R   E   R   R   R   F
                    150                               160

AGC TGG ACT CGG AGC ACA TCG GAG GAG GAG ATC CAG CGG CTG CGG GAG CGC TTC
S   W   T   R   S   T   S   E   E   E   I   Q   R   L   R   E   R   F
                        170                                   180

GAG GAT GAG GCC GGC TGC GAG ATG AGA CGG AGG CCA CCA TCG GCC CTG CGC AAG
E   D   E   A   G   C   E   M   R   R   R   P   P   S   A   L   R   K
                            190

GAG ATG GAG GAG GCC TCC CTG ATG CGG GCG GAC GTG GAC AAG AAG GTG CAG TCG
E   M   E   E   A   S   L   M   R   A   D   V   D   K   K   V   Q   S
    200                                   210

CTG CAG GAC GAG GTG CCT TCC TGC GGG GCA ACC ACG AGG AGG AGG TGG CCG AGC
L   Q   D   E   V   P   S   C   G   A   T   T   R   R   R   W   P   S
            220                                   230

TCT GGC CCA GCT GCA GGC TTC CCA CGC ACG GTG GAG AGG AAG GAC TAC CTG AAG
S   G   P   A   A   G   F   P   R   T   V   E   R   K   D   Y   L   K
                    240                               250

ACG GAC CTG ACC ACG GCC GTT GAA GGA GAT CCG CGC TCA GCT GGA GTG CCA GTC
T   D   L   T   T   A   V   E   G   D   P   R   S   A   G   V   P   V
                        260                                   270
```

```
CGA CCA CAA CAT GCA CCA GGC CAG GAG TGG TTC AAG TGC CCG TAC GCC AAG CTG
R   P   Q   H   A   P   G   Q   E   W   F   K   C   P   Y   A   K   L
                                    280

ACA GAG GCG GCG GAG CAG AAC AAG GAG GCC ATC CCG TCC GCC AAG GAG GAG ATC
T   E   A   A   E   Q   N   K   E   A   I   P   S   A   K   E   E   I
    290                                     300

GCC GAG TAC CCG GCG GCA GCT GCA GTC CAA GAG CAT CGA GCT GGA GTC GTG CGC
A   E   Y   P   A   A   A   A   V   Q   E   H   R   A   G   V   V   R
            310                                     320

GGC ACC AAG GAG TCG CTG GAG CGG CAG CTG AGC GAC ATC GAG GAG CGT CAC AAC
G   T   K   E   S   L   E   R   Q   L   S   D   I   E   E   R   H   N
                    330                                     340

AAC GAC CTC ACC ACC TAT CAG GAC ACG ATC CAT CAG CTG GAA AAT GAG CTC AGA
N   D   L   T   T   Y   Q   D   T   I   H   Q   L   E   N   E   L   R
                        350                                     360

GGA ACG AAG TGG GAG ATG GCA CGT CAT TTG AGG GAG TAC CAG GAT CTC CTC AAT
G   T   K   W   E   M   A   R   H   L   R   E   Y   Q   D   L   L   N
                                370

GTC AAG ATG GCC CTG GAT ATC GAA ATT GCT GCA TAC AGG AAG CTG CTG GAG GGT
V   K   M   A   L   D   I   E   I   A   A   Y   R   K   L   L   E   G
    380                                     390

GAG GAG ACA AGA TTC AGT GCC TTC TCT GGA AGC ATC ACT GGA CCC ATA TTC ACA
E   E   T   R   F   S   A   F   S   G   S   I   T   G   P   I   F   T
            400                                     410

CAC AGA CAA CCA TCG GTC ACA ATA GCA TCC ACT AAA ATA CAG AAA ACC AAA ATC
H   R   Q   P   S   V   T   I   A   S   T   K   I   Q   K   T   K   I
                    420                                     430

GAG CCA CCA AAG CTG AAG GTC CAG CAC AAA TTT GTA GAA GAA ATC ATT GAA GAG
E   P   P   K   L   K   V   Q   H   K   F   V   E   E   I   I   E   E
                        440                                     450

ACG AAA GTA GAG GAT GAG AAG TCT GAA ATG GAA GAT GCC CTC TCA GCC ATT GCA
T   K   V   E   D   E   K   S   E   M   E   D   A   L   S   A   I   A
                                460

GAA GAA ATG GCA GCA AAG GCT CAG GAG GAA GAA CAG GAG GAG GAA AAG GCA GAA
E   E   M   A   A   K   A   Q   E   E   E   Q   E   E   E   K   A   E
    470                                     480

GAA GAA GCT GTA GAG GAA GAA GCT GTT TCT GAG AAG GCT GCA GAA CAG GCA GCT
E   E   A   V   E   E   E   A   V   S   E   K   A   A   E   Q   A   A
            490                                     500

GAG GAA GAA GAG AAG GAG GAA GAA GAA GCA GAG GAG GAA GAA GCT GCA AAA TCA
E   E   E   E   K   E   E   E   E   A   E   E   E   E   A   A   K   S
                    510                                     520

GAC GCT GCA GAA GAA GGA GGC TCT AAA AAG GAA GAA ATA GAG GAA AAG GAA GAA
D   A   A   E   E   G   G   S   K   K   E   E   I   E   E   K   E   E
                        530                                     540
```

```
AGG GAG GAG GCT GAA GAA GAA GAA GCT GAA GCC AAG GGC AAA GCT GAA GAG GCA
R   E   E   A   E   E   E   E   A   E   A   K   G   K   A   E   E   A
                                    550

GGT GCA AAG GTA GAA AAA GTG AAA TCA CCT CCT GCA AAG TCA CCC CCT AAA TCC
G   A   K   V   E   K   V   K   S   P   P   A   K   S   P   P   K   S
    560                                     570

CCC CCT AAA TCC CCT GTA ACA GAG CAA GCC AAG GCC GTC CAG AAA GCA GCA GCA
P   P   K   S   P   V   T   E   Q   A   K   A   V   Q   K   A   A   A
            580                             -       590

GAG GTA GGA AAG GAT CAG AAA GCA GAG AAA GCT GCT GAG AAG GCA GCC AAG GAG
E   V   G   K   D   Q   K   A   E   K   A   A   E   K   A   A   K   E
                    600                                 610

GAG AAG GCA GCA TCC CCA GAG AAG CCG GCG ACA CCA AAG GTG ACC TCC CCG GAG
E   K   A   A   S   P   E   K   P   A   T   P   K   V   T   S   P   E
                        620                                     630

AAA CCA GCG ACT CCG GAG AAA CCA CCA ACC CCA GAG AAA GCG ATC ACC CCG GAG
K   P   A   T   P   E   K   P   P   T   P   E   K   A   I   T   P   E
                            640

AAG GTC CGT TCC CCA GAA AAA CCA ACA ACC CCG GAA AAA GTG GTG AGC CCA GAG
K   V   R   S   P   E   K   P   T   T   P   E   K   V   V   S   P   E
    650                                     660

AAA CCA GCA AGC CCA GAG AAG CCC CGA ACC CCA GAG AAA CCA GCA AGC CCC GAA
K   P   A   S   P   E   K   P   R   T   P   E   K   P   A   S   P   E
            670                                 680

AAA CCG GCA ACA CCA GAG AAG CCC CGC ACT CCT GAA AAG CCA GCG ACG CCG GAG
K   P   A   T   P   E   K   P   R   T   P   E   K   P   A   T   P   E
                    690                                 700

AAG CCC CGT TCT CCA GAG AAG CCA TCC TCC CCG CTC AAA GAT GAA AAG GCT GTG
K   P   R   S   P   E   K   P   S   S   P   L   K   D   E   K   A   V
                        710                                     720

GTG GAG GAG AGC ATC ACT GTC ACA AAG GTA ACA AAA GTC ACT GCA GAG GTG GAG
V   E   E   S   I   T   V   T   K   V   T   K   V   T   A   E   V   E
                            730

GTG TCG AAG GAA GCC AGG AAA GAA GAC ATT GCG GTG AAT GGT GAA GTG GAG GAG
V   S   K   E   A   R   K   E   D   I   A   V   N   G   E   V   E   E
    740                                     750

AAG AAG GAT GAG GCG AAG GAG AAG GAG GCT GAG GAG GAA GAG AAG GGC GTT GTC
K   K   D   E   A   K   E   K   E   A   E   E   E   E   K   G   V   V
            760                                 770

ACC AAT GGG CTC GAT GTG AGC CCC GTC GAT GAG AAG GGT GAG AAA GTT GTA GTA
T   N   G   L   D   V   S   P   V   D   E   K   G   E   K   V   V   V
                    780                                 790

ACC AAA AAA GCA GAG AAA ATC ACA AGT GAA GGA GGG GAC AGT ACT ACC ACG TAC
T   K   K   A   E   K   I   T   S   E   G   G   D   S   T   T   T   Y
                        800                                     810
```

```
ATC ACG AAG TCG GTG ACG GTC ACT CAG ACG GTG GAG GAA CAC GAA GAG AGC TTT
I   T   K   S   V   T   V   T   Q   T   V   E   E   H   E   E   S   F
                                    820

GAG GAG AAA TTG GTG TCC ACT AAG AAA GTG GAG AAA GTT ACT TCA CAT GCT GTA
E   E   K   L   V   S   T   K   K   V   E   K   V   T   S   H   A   V
  830                                       840

GTA AAA GAG ATT AAA GAA TGA ATTAAACTCAATTATAGCTAAATTAA AAAAAAATTAAAGAGCT
V   K   E   I   K   E   .
          850

TGGGTCGGTGCAAAAGGTTAAGCCATATGACAGCTGCAAAATGCATGTGAGTG ACGGCTTTAAAGCAGAAC
GGGTTCTCTCATGGAGGCTCCAGAGCACAGTATTTTACTTTTTTGTGCAATATAGGGGAGGGGGGGGGAATG
CATGCAGGCTCAAGATGTGCTCCCTCCACAGAGCTTGGGGAACTAAATAA AAAATAATAAAAAAAAAATAAT
AAAATACTAATAATAGTGCATGAGTTGAAATGTGCAAAGGAACCTTTTGAATTTCCTGAGCTGTTGGAGGGA
CGTATCTGAGGAACGACTTAAGATGTATTATGCAAAGAACCAACTGAGCAAAACCAACAAAACAGTACTAGA
ATATTCATGAGAACCAGAACTCTCCTAGCCTTAATAATAATAAAAAAAAAAAGCTACTTATAAGTAATTAT
GTTTACCTCACTGGTGCAATTAGGTGGACTTTTGCTCATGGGAGAACCTAGTTGACATGCACAGTACGCAACC
TTTTGTTTGTTTGATGTAAAACCGTCACAGCAGTTCTTGCAGAAAGTTACTTCACATGCTGTAGTAAAAGAGA
TTAAAGAATGA
```

FIGURE 3: **Alignment of Rat and Chicken Amino Acid Sequences** - The chicken sequence is shown on top (C) with the rat sequence below (R) (Napalitano et. al., 1987). Shown is a computer generated aligment of the complete amino acid sequences for both proteins with mismatches or gaps designated by an asterisk. The three rod domains are designated (underlined in rat sequence) as reported for the rat and porcine sequences (Geisler and Weber, 1984). The arrow marks the begining of the sequence published by Zopf et.al., (1987).

```
                10          16          26          36          46          56
C~MSYSMEPLGT P....RTAGD DRDPRHVQPG SARPRPAASV AVVSRGSGST VSSSYKRTNL
    ****  *  **** **** * ** ****   ** ****** ***     *           **
R~MSYTLDSLGN PSAYRRVPTE TRSSRSRVSG S..PSSGFRS QSWSRGSPST VSSSYKRSAL
                10          20          30          38          48          58

                66          76          79          87          97         107
GAPRTAYGST VLSSQRPRRE PVL.......  LLNG..AAEL KLSRSNEKEQ LQGLNDRLRG
*  **  * * *   *****  ** *******      ******                      *
.APALAYSSA MLSS....AE SSLDFSQSSS LLNGGSGGDY KLSRSNEKEQ LQGLNDRFAG
                67          73          83          93         103         113

                COIL IA
               117         127         137         147         155         160
TSRRCITWSS RTRDQRAARC AEARRAGSAG GRLRAGAAGA ARGPGA..GE PRERR.....
***** **** **** *****   ******** **  *** **  **** ***  *  * *****
....YIEKVH YLEQQNKEIE AE.......I HALRQKQASH AQLGDAYDQE IRELRATLEM
               119         129         132         142         152         162

                                    COIL IB
               164         174         184         191         199         209
......RFSW TRSTSEEETQ RLRERFEDEA GCEMRRR... ..PPSALRKE MEEASLMRAD
********** *****  ***   *    *   **** * *** ***** *  * *  * * *****
VNHEKAQVQL DSDHLEEDIH RLKERFEEEA ....RLRDDT EAAIRAVRKD IEESSMVKVE
               172         182         192         198         208         218

               219         229         234         244         254         264
VDKKVQSLQD EVPSCGATTR .....RRWPS SGPAAGFPRT VERKDYLKTD LTTAVEGDPR
*            **** **  ********** *********                ** *****
LDKKVQSLQD EV....AFLR SNHEEEVADL LAQIQASHIT VERKDYLKTD ISTA.LKEIR
               228         234         244         254         264         273

               273         283         293         303         313         323
SAGVPVRPQ. HAPGQEWFKC PYAKLTEAAE QNKEAIPSAK EEIAEYPAAA AVQEHRAGVV
 ******* * *****      *                      ****  *********
SQLECHSDQN MHQAEEWFKC RYAKLTEAAE QNKEAIRSAK EEIAEYRRQL QSKSIELESV
               283         293         303         313         323         333

                                    COIL 2
               333         343         353         363         373         383
RGTKESLERQ LSDIEERHNN DLTTYQDTIH QLENELRGTK WEMARHLRE YQDLLNVKMAL
                      *    **     *
RGTKESLERQ LSDIEERHNH DLSSYQDTIQ QLENELRGTK WEMARHLRE YQDLLNVKMAL
               343         353         363         373         383         393

               393         403         413         423         433         443
DIEIAAYRKL LEGEETRFSA FSGSITGPIF THRQPSVTIA STKIQKTKI EPPKLKVQHKF
                  * *          *           *   *   *   **
DIEIAAYRKL LEGEETFRST FSGSITGPLY THRQPSVTI. SSKIQKTKV EAFKLKVQHKF
               403         413         423         432         442         452
```

```
       453         463         473         483         487         492
VEEIIEETKV EDEKSKMKDA LSAIAEEMAA KAQEE.EQEE EKAEE..... .....EAVEE
           ** *     *     *     * *  * **    *  ***** *****  *
VEEIIEETKV EDEK.EMEDA LTVIAEELAA SAKEEKEEAE EKEEEPEVKS PVKSPEAKEE
       462         471         481         491         501         511

       502         412         522         532         542         552
AVSEKAAEQA AEEEEKEEEE AEEEEAAKSD AAKKGGSKKE EIEEKEEAEE AEEEEAEAKG
***  ***** *    **    *  ****    * **   *   ***  * * * *  * * *  *
EEGEK..... EEEEEGQEEE EEEDKGVKSD QAEEGGSEKE GSSEKDEGEQ EEEGETEAEG
       516         526         536         546         556         566

       562         572         582         592         602         612
KAEEAGAKVE KVSPPAKSPP AKSPPKSPVT EQAKAVQKAA AEVGKDQKAE KAAEKAAKEE
**   *  *   ********* *********  ** **** * * ** ****  **********
EGEEAEAKEE K......... .........KT E.GKVEEMAI KEEIKVELPE ..........
       576         577         579         588         598

       622         632         641         646         656         666
KAASPEKPAT PKVTSPEKPA TPEKPPTPE. KA.....ITP EKVRSPEKPT TPEKVVSPEK
****** ***  ***  **** ** ** *  *   ********    ***  ** *** * ** *
......KAKS PVPKSP.... VEEVKPKPEA KAGKDEQKEE EKV..EEK.. ...KEVAKES
       602         608         618         628         634         641

       676         685         695         705         715         725
PASPEKPRTP EKPASPEKP. ATPEKPRTPE KPATPEKPRS PEKPSSPLKD EKAVVEESIT
 **** ****   ****   * ** * ***** ****** **   * *******    *   *
P....KEEKV EK..KEEKPL DVPDK..... .....KKAES PVK....... EKA.VEEMIT
       647         655         660         665         668         677

       732                     736         737         746         756
VTKVTKV... .......... ......TAEV E......... .VSKEARKED IAVNGEVEEK
*  ** **** ********** *******  *  ********* **** **      *     *
ITKSVKKSLE KDTKEEKPQQ QEKVKEKAEE EGGSEEEVGD KSPQESKKED IAINGEVEGK
       687         697         707         717         727         737

       762         772         782         787         796         806
KDEAKE.... KEAEEEEKGV VTNGLDKSPV DE.KGE.... .KVVVTKKAE KITSEGGDST
** ** **** ****           *  * * *   **** *       *          **
EEEEQETQEK GSGQEEEKGV VTNGLDVSPA EEKKGEDRSD DKVVVTKKVE KITSEGGDGA
       747         757         767         777         787         797

       816         826         836         846         852
TTYITKSVTV TQTKVEEHEE SFEEKLVSTK KVEKVTSHAV VKEIKE
 *           *                       *    ****
TKYITKSVTV TQKVEEHEET FEEKLVSTKK VEKVTSHAIV KEVTQD
       807         817         827         837         839
```

TABLE I: **Comparison of Rat and Chicken Structural domains - Levels of identity are given for individual subdomains, domains, and the whole protein on both the amino acid and nucleotide levels. For the protein, the values are given with (w IF) and without (w/o IF) considering isofunctional changes.**

TABLE I : LEVELS OF IDENTITY BETWEEN CHICKEN AND RAT STRUCTURAL DOMAINS

Domain	Amino Acids Included	%Identity		
		Protein		DNA
		w/o IF	w IF	
Amino Terminal	1-100	45	52	71
Carboxy Terminal	401-852	58	61	64
Total Rod	101-400	57	62	74
Coil 1A	101-137	50	50	88
Coil 1B	150-235	36	44	78
Coil 2	257-400	76	78	87
Entire Protein	1-852	52	57	68

FIGURE 4: **Northern Analysis of 10 Day Embryonic Chick Tissues - Total RNA** from 10 day embryonic chick brain (B), liver (L), skeletal muscle (M), heart (H), and gut (G) was blotted and hybridized with radiolabeled c19NFM as described in materials and methods. Positions of the ribosomal RNA bands are shown as molecular weight markers. Exposure time was 51 hours.

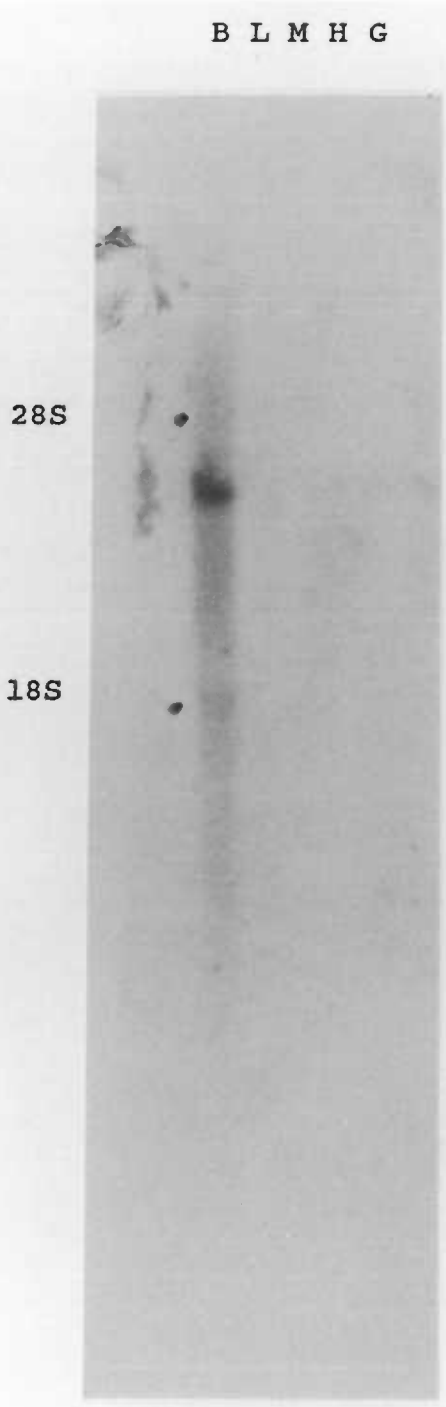

B L M H G
28S
18S

FIGURE 5: **Southern Analysis** - Genomic DNA (20ug) was digested with either SstI (S) or EcoRI (R) and probed with c19NFM as described in materials and methods. Molecular weight markers were visualized by ethidium bromide staining prior to blotting. Exposure time was 24 hours.

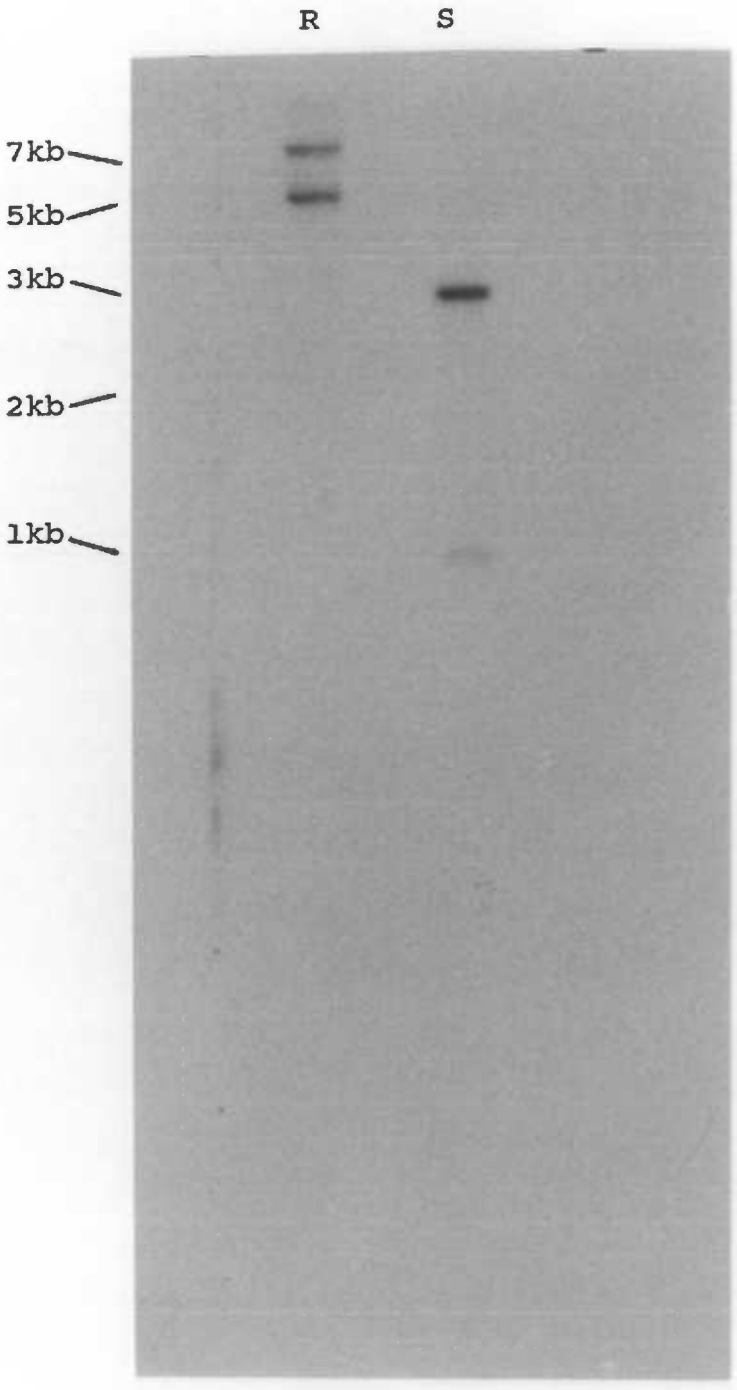
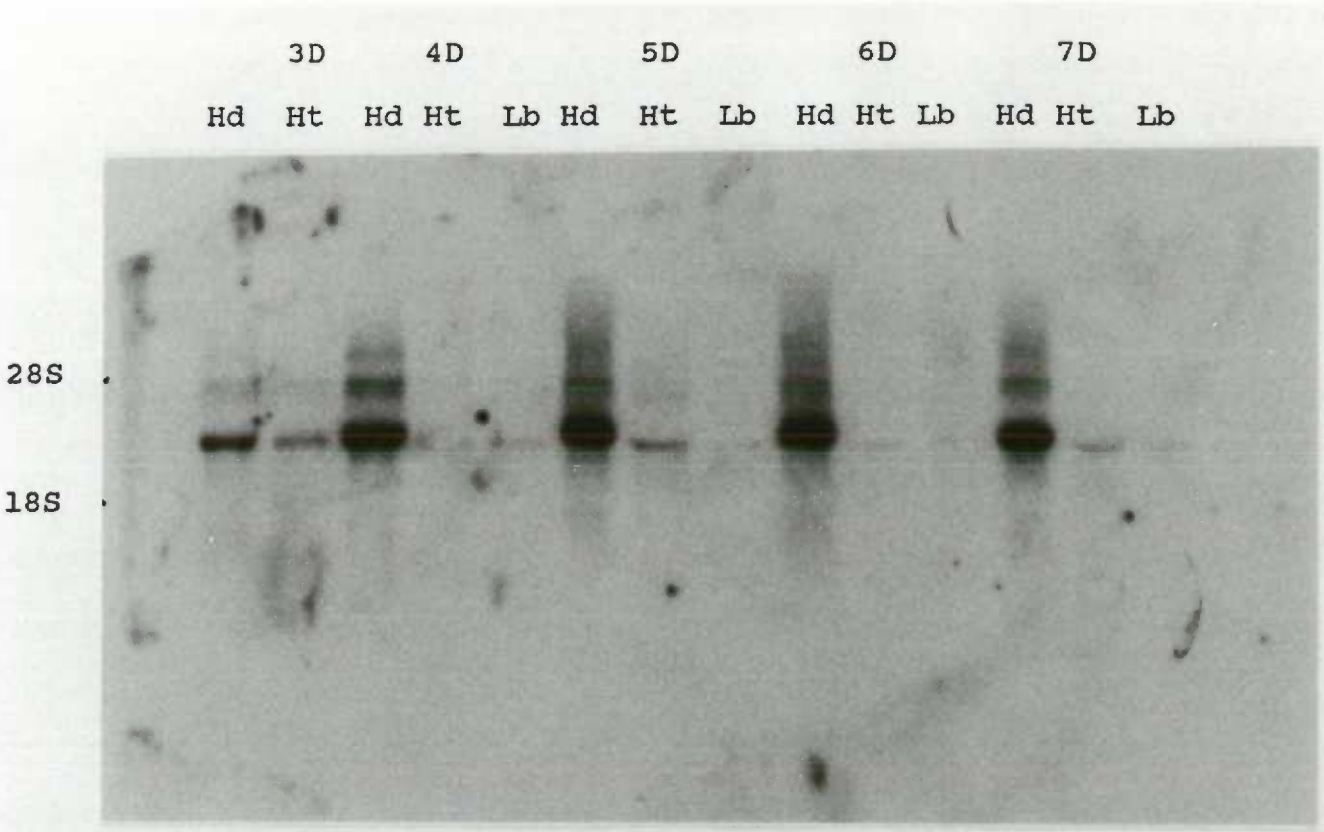
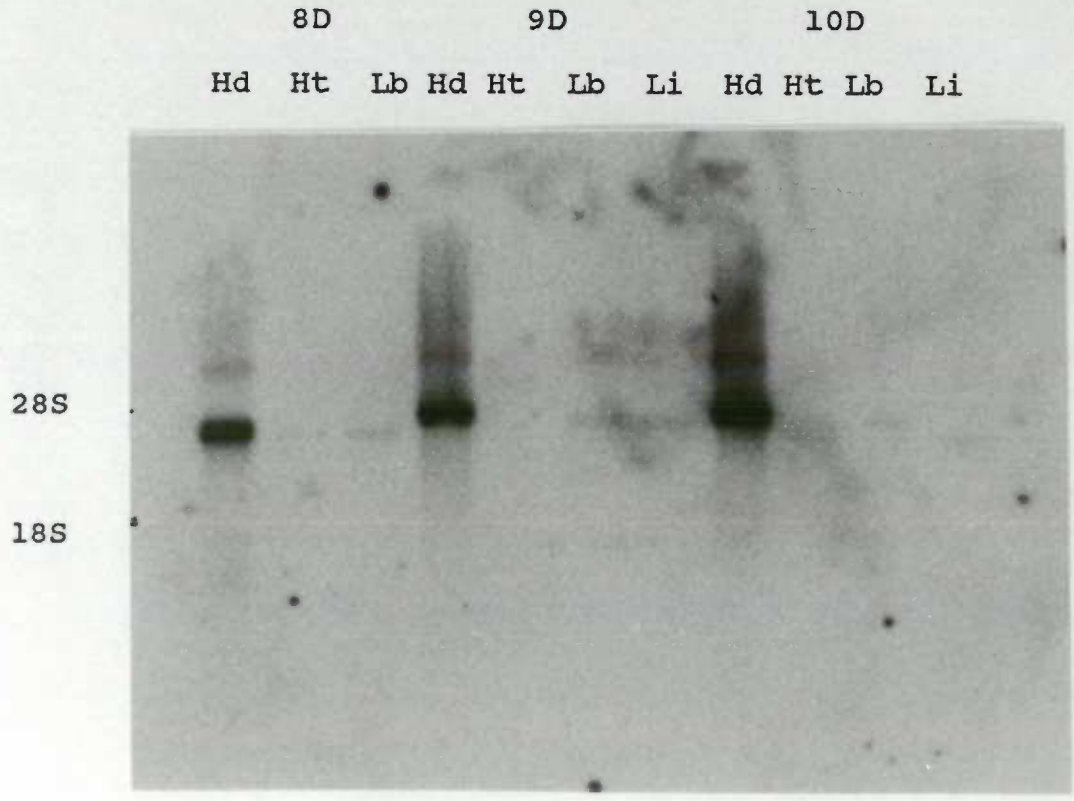

R
S
7kb
5kb
3kb
2kb
1kb

FIGURE 6: **Time Course of Expression of c19NFM** - Samples of 10ug of total RNA from Head (Hd), Heart (Ht), Limb (Lb), and Liver (Li) of 3 to 10 day embryos were blotted and probed as described in materials and methods. The positions of the ribosomal RNA bands are shown as molecular weight markers. Exposure time was seven days.


3D
4D
5D
6D
7D
Hd Ht Hd Ht Lb Hd Ht Lb Hd Ht Lb Hd Ht Lb
28S
18S


8D
9D
10D
Hd Ht Lb Hd Ht Lb Li Hd Ht Lb Li
28S
18S

III CONCLUSIONS and DISCUSSION

(A) Conclusions:

From our results the following conclusions can be drawn:

(1) We have isolated and characterized a cDNA (c19NFM) from 10 day embryonic chicken brain which appears to be a longer version of a cDNA previously reported by Zopf, et al. (1987) to encode the carboxy terminal 509 amino acids of a chicken middle molecular weight neurofilament protein.

(2) c19NFM appears to contain the entire coding region of this gene. Alignment of both the predicted protein and the DNA sequence (see Appendix I,figure 1) with the sequences published for the rat NF-M protein (Napolitano et al., 1987), the predicted molecular weight, and a comparison of the DNA sequence surrounding the putative initiator methionine codon with the consensus sequence reported for this region by Kozak (1986) (see Appendix, figure 2), all indicate our cDNA contains the entire coding region for this protein.

(3) The alignment of the predicted chicken amino acid sequence with that of the rat exhibits a level of identity of 52% indicating both conservation and divergence.

(4) The alignment of the chicken and rat DNA sequences exhibits a considerably higher level of identity of 68%, indicating a selective divergence between this protein and the mammalian counterpart.

(5) In addition to the nervous system we have found this gene to be expressed in the developing heart, limbs, and liver. Zopf et al. (1987) reported that this gene was expressed only in the nervous system in the post hatching chicken, meaning that the non-neuronal expression described here can be considered transient and embryonic. In addition there are a

number of strong indications that the cells expressing this gene outside the nervous system are not neuronal or neuronal-like cells. Therefore the function of this protein is not necessarily associated with a neuronal phenotype.

(6) Our results can be explained using models which involve either one or two genes which encode NF-M or NF-M-like proteins in the chicken. Although the published report indicates that there is a single NF-M gene in the chicken genome, multiple copies of a single gene or closely related copies of similar genes can not be eliminated.

(B) Discussion:

Intermediate filaments represent a large and diverse class of conserved cytoskeletal elements. Although they have been extensively studied there is no solid evidence of their function.

Both the expression pattern and the putative translation product of c19NFM indicate a divergence of function from the mammalian NF-M protein. In mammals NF-M is not thought to be involved in filament formation but may play a role in the interactions of IFs with organelles such as microtubules (Leterrier et.al., 1981; Leterrrier et.al., 1982). If this is the case, a possible model that would explain our results is one in which a divergent form of NF-M evolved which retains the properties which allow interactions with organelles (the "NF-M like" character), but is altered so that it can interact with intermediate filament types in other cells (the divergent character). At the same time new regulatory mechanisms would have to evolve which would result in the expression of this gene in other cell types. The final result would be a new NF-M like gene which is expressed during development in distinct populations of cells which require a specific interaction of their IF network (probably vimentin) with another organelle; in other words this gene wolud be an embryonic variant of NF-M. This type of a model is directly supported by the results of Bennett and Dilullo (1985), and although the work of Zopf et al. (1987) does not directly support this possibility it does not dispute our results on any point.

Bennett and Dilullo (1985) carried out studies similiar to ours by characterizing the expression of NF-M immunoreactivity in the developing chicken embryo. Their report includes two main points not addressed by our results. First, they studied the expression and localization of the protein rather than the transcript. This point allows us to eliminate the possibility that in our results the non-neuronal RNA expression we have seen is an abortive transcript that is not translated. At the same time our results confirm their studies by reducing the possibility of cross reactivity and background problems associated with an immunological approach. Because of the high level of conservation of the IFPs, the use of a polyclonal antiserum (as was used by Bennett and Dilullo) could exhibit cross reactivity with other IFPs. But, our results demonstrate a transcript size, in the same tissues in which they reported NF-M immunoreactivity, which has a molecular weight that is inconsistent with this possibility. The other possible point of confusion with an immunological approach is that in non-neuronal regions, neuronal projections begin to appear as some of these structures develop (for example the heart). These projections contain neurofilaments and so raise background levels. In our work we are studying the transcript, which has not been reported to be found in the axons (Zopf et al., 1987), eliminating this possibility. Second, Bennett and Dilullo were able to study NF-M immunoreactivity as early as the first day of incubation (stage 9 , Hamburger and Hamilton, 1951). The results of this part of their study shows immunoreactivity in a distinctive pattern in the early CNS, in cells which will eventually give rise to a number of cell types other than neurons. This result supports the possibility of an embryonic variant of NF-M. The only disagreement between our results and that of Bennett and Dilullo is in the time course of expression in the heart. In our case we see expression from day 2 of incubation (stage 12 of Hamburger and Hamilton) out to day 8 (stage 34 of Hamburger and Hamilton), but in their results the immunoreactivity is gone by the end of day 5 (stage 28 of Hamburger and Hamilton). However, this slight difference could be due to difference in the detection limits, and our overall results agree quite well with each other and form a convincing argument for tran-

sient embryonic expression of NF-M in the chicken embryo.

The report by Zopf et al. (1987) indicates this cDNA encodes a typical neurofilament protein. However, in their report they have concentrated on a later developmental stage than our studies and those of Bennett and Dilullo. Their cDNA was isolated from a library which had been enriched for transcripts derived from the optic lobe of post hatching day 1 (P1) chickens.

There are three aspects addressed by Zopf et al. which were not addressed by our work. These include the expression of this gene in later stages of development as well as the adult, analysis by in situ hybridization of the P1 chicken CNS, and northern analysis of poly A+ RNA samples rather than total RNA. The analysis of later stages and adult tissues has shown the expression of this gene to be specific to neurons, which we have concluded indicates that the non-neuronal expression we have seen is restricted to the embryo. Their in situ studies have demonstrated specific hybridization to cells in a number of regions in the CNS, all of which appeared to exhibit long projections or extensive arborization. These results also indicated that the message was located in the cell body and not in the axons. These results aid in eliminating the possibility that the RNA is transported down the axons, which would cause an apparent background expression in non-neuronal tissues as mentioned in the discussion of the work of Bennett and Dilullo above.

One point that might be interpreted as a disagreement between our work and that of Zopf et al. is the presence of high molecular weight (HMW) bands in our northern blots. There was no mention of such bands in their report. However, the only northern blot reported which includes both total RNA and developmental time points which we have studied cannot be directly compared with our results for two reasons. First, this blot was exposed for a relatively short time period (8 hours), and second it contained RNA isolated specifically from the optic lobe. Therefore, it is still possible that these HMW bands are unprocessed transcripts that would be eliminated by using poly A+ RNA in our experiments. But, if this is the case then these additional bands would be exhibiting the same

developmental regulation as the 3.5 kb transcript.

Other than the points mentioned, our results are in complete agreement with those of Zopf et al. (1987) (ie: DNA sequencing, message size, and genomic southern analysis), and indicate that we have both isolated the same cDNA. In addition to the confirmation of their results we have been able to describe the expression of this gene in the early embryo, and the pattern of embryonic expression appears to be quite different than in the adult chicken. Finally, our cDNA appears to contain the entire coding region for this protein. Therefore, the results of all three groups are in agreement, and extend the description of how this gene (or genes) is expressed from the first day of incubation to the adult chicken.

(C) Prospectus:

The cDNA which we have isolated allows us two advantages over other workers in the field. Our cDNA appears to encode the entire protein, and it is derived from the chicken, which is particularly well characterized and amenable to the study of development of the early embryo.

The first point which should be addressed in the further study of this cDNA, is whether or not it is derived from a gene which exists as a single unique copy, or as either multiple copies of the same gene or as one of a pair of closely related genes. A rapid method of gaining insight into this problem would be to analyze the remaining cDNAs which were isolated with c19NFM. These cDNAs all appear to hybridize to the same transcript (data not shown), and therefore any differences found in the restriction maps as compared with c19NFM would indicate two closely related genes. If this approach is unproductive the cDNA library could be screened again and this process repeated for any positive clones. The same approach could be used with a genomic library. A different method would be to use subfragments of c19NFM in more detailed southern and northern analysis which would involve variations in stringency, in order to determine if the presence of multiple bands in both northern and southern blots is due to multiple

transcripts or genes. If there appears to be a single copy of the gene this can be confirmed by direct quantification of the bands in a genomic southern blot by comparison with hybridization signal of a known single copy gene.

A number of further studies aimed at understanding the function and regulation of this gene (or genes) can be carried out using the full length cDNA. To study the function of this protein the cDNA can be used to produce overexpression, underexpression, ectopic expression, and expression of mutated forms of the protein. All of these could be accomplished in tissue culture by transient transfection of cells with a plasmid construct containing the cDNA under control of a regulated promoter. This would allow expression of this protein in a number of cell types which can be cultured from the chicken embryo (such as neural crest, heart, liver, and fibroblasts). Therefore the protein could be overexpressed in both neuronal and non-neuronal cells, and by producing an antisense transcript, underexpression could be produced in neuronal cells. Finally, specific domains could be altered by site specific mutation or truncation, and these mutated forms could be expressed in a similiar fashion and their effect on the cytoskeleton studied. In addition these same questions could be addressed in vivo by using a retroviral expression vector to produce expression of this cDNA in different cell types during development.

The first step in studying the regulation of this gene could be to use the 5' region of the cDNA to screen a genomic library and isolate the corresponding region of the genomic sequence in order to isolate the promoter regions. These regions would be sequenced and analyzed for known promoter elements. Any promoter regions isolated could then be placed in a plasmid construct in front of a reporter gene and transfected into cells of different tissues (as listed above) derived from the chicken embryo, in order to determine if different tissues use different cis-regulatory elements. In addition these cells could be grown in the presence or absence of various growth factors to begin to determine which factors regulate this gene in a given tissue.

Finally, these studies could be extended to include experiments aimed at defining

the proteins which bind to and regulate this gene. This would begin with band shift or protection studies of the promoter region using protein extracts from different tissues and developmental stages. In these studies restriction digests of a promoter region are incubated with protein extracts, and in the case of band shift experiments are then subjected to electrophoresis. Any fragment bound by a protein will exhibit a decrease in electrophoretic mobility. In protection studies the fragments are incubated with extracts then treated with DNAse before electrophoresis, so that unprotected fragments are degraded. From these types of experiments one can begin to determine what regulatory proteins are present in a specific stage or tissue, what specific sequences it is binding to, and what factors effect this binding.

By following these approaches, and using this cDNA as a tool, one can begin to study much more directly the function and regulation of this gene, the results of which would not only be useful in understanding the function and regulation of NF-M and the NFs, but also in understanding the same aspects of the intermediate filaments and their subunit proteins in general.

References

Bennett, G., Dilullo, C. (1985): Transient Expression of a Neurofilament Protein by Replicating Neuroepithelial Cells of the Embryonic Chick Brain. Dev. Biol. 107:107-127.

Hamburger, V., Hamilton, H. (1951): A Series of Normal Stages in the Development of the Chick Embryo. J. Morphology 88:49-92.

Kozak, M. (1986): Point Mutations Define a Sequence Flanking the AUG Initiator Codon that Modulates Translation by Eukaryotic Ribosomes. Cell 44:283-292.

Leterrier, J., Liem, R., Shelanski, M. (1981): Preferential Phosphorylation of the 150,000 Molecular Weight Component of Neurofilament by a Cyclic AMP-Dependent, Microtubule-Associated Protein Kinase. J. Cell Biol. 90:755-760.

Leterrier, J., Liem, R., Shelanski, M. (1982): Interaction Between Neurofilament and Microtubule-Associated Protein: A Possible Mechanism for Intraorganellar Bridging. J. Cell Biol. 95:982-986.

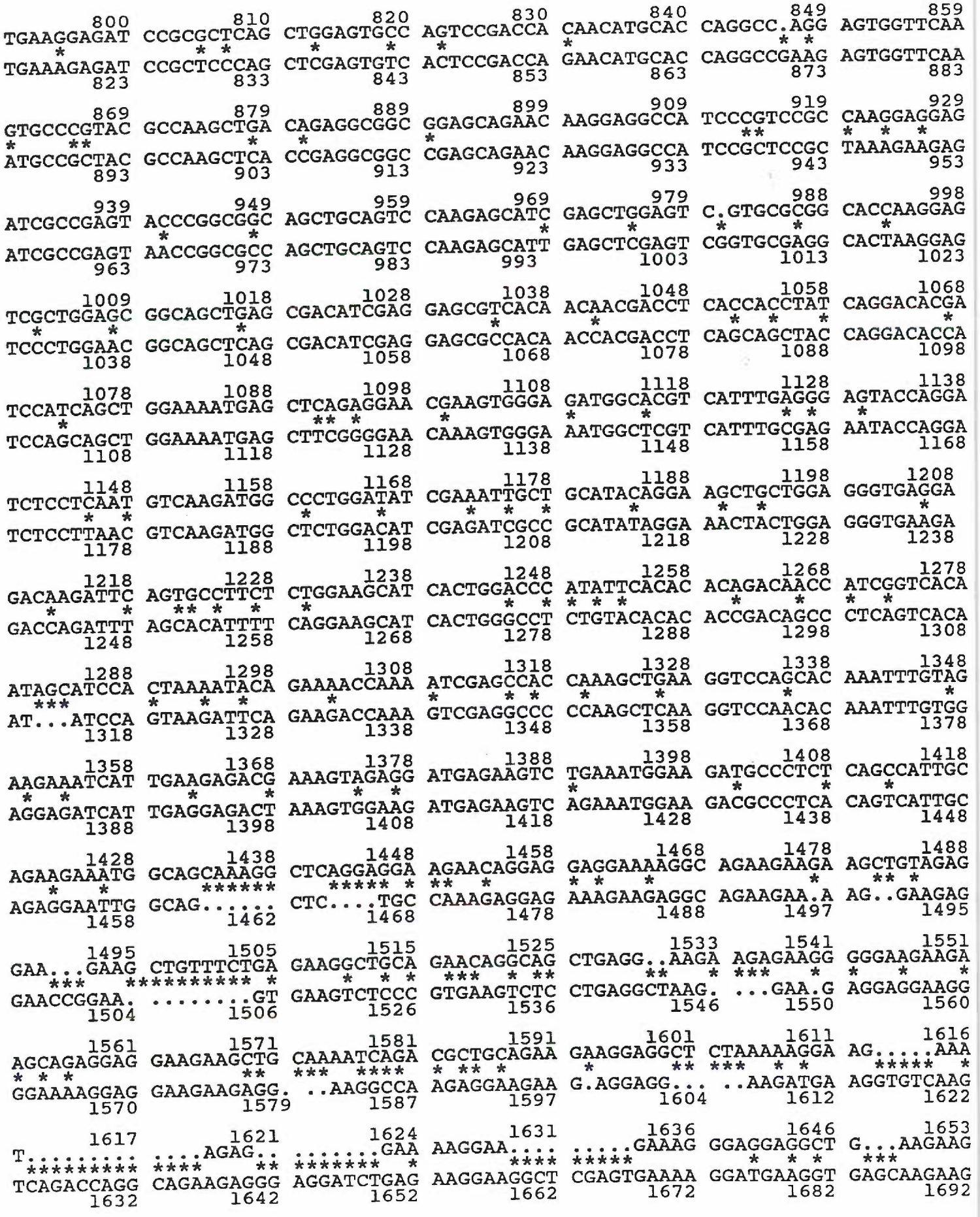
Napolitano, E., Chin, S., Colmman, D., Liem, R. (1987): Complete Amino Acid Sequence and In Vitro Expression of Rat NF-M, The Middle Molecular Weight Neurofilament Protein. J. Neurosci. 7:2590-2599.

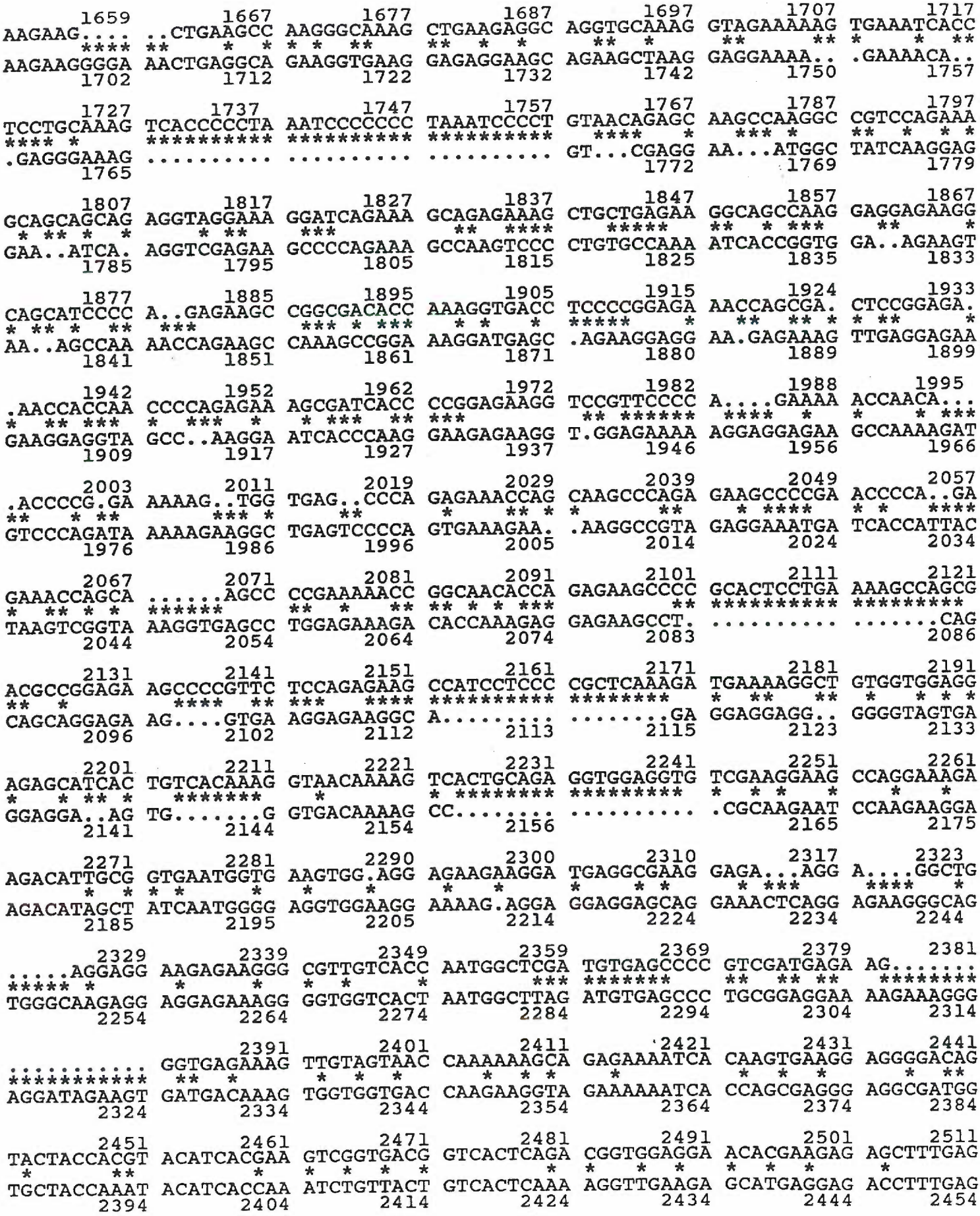
Zopf, D., Hermans-Borgmeyer, I., Gundelfinger, E., Betz, H. (1987): Identification of Gene Products Expressed in the Developing Chick Visual System: Characterization of A Middle Molecular Weight Neurofilament cDNA. Genes and Development 1:699-708.

APPENDIX:

FIGURE 1: **Alignment of Rat and Chicken Nucleotide Sequences** - A computer generated alignment for the entire coding region of the rat NF-M gene (Napolitano et.al., 1987) and chicken c19NFM is shown. The the sequence published by Zopf et.al. (1987) begins at nucleotide 1034. The initiation and termination codons are underlined for both sequences. The 3' untranslated region is not shown. Mismatches are designated by an asterisk.

```
                    7         17        26        35        43        52
CHICKEN:  CCGCC...AT GAGCTACTCC ATGGA.GCCC TTGGGAACC. ..CCTCGTAC C.GCAGGTGA
                  10        20        30        40        50        60
     RAT:  CCTCCAAGAT GAGCTACACG CTGGACTCGC TGGGCAACCC GTCCGCCTAC CGGCGCGTTC

          62        72        82        92       102       111       121
TGACCGAGAC CCGCGCCACG TACAGCCGGG CTCAGCGCGT CCCCGTCCAG CGGCTTCCG. TCGCAGTCGT
          70        80        90        99       109       119       129
CAACCGAGAC CCGGTCCAGC TTCAGTCGTG .TGAGCGGTT CCCCGTCCAG CGGCTTCCGC TCGCAGTCCT

         130       140       150       160       170       180       190
.GTCGCGGGG CTCCGGCAGC ACCGTGTCCT CGTCCTACAA ACGCACCAAC CTCGGGGCTC CGCGAACCGC
         139       149       159       169       179       186       196
GGTCCCGCGG CTCGCCCAGC ACCGTGTCCT CCTCCTACAA GCGCAGCGCG CTC...GCCC CGGCGCTCGC

         200       210       219       225       235       241       251
CTACGGCTCC ACCGTGCTGA GCTC.GCAGA GA....CCTC GACGTGAGCC AGTCCT.... GCTGCTGAAC
         206       216       226       236       246       256       266
CTACAGCTCG GCTATGCTCA GCTCGGCCGA GAGCAGCCTG GACTTCAGCC AGTCCTCTTC GCTGCTTAAC

         257       265       275       285       295       305       315
GGGG....CG GC..CGAGCT GAAGCTGAGC CGCTCCAACG AGAAGGAGCA GCTGCAGGGG CTGAACGACC
         276       286       296       306       316       326       336
GGCGGCTCCG GCGGCGACTA CAAGCTGTCC CGCTCAAACG AGAAAGAGCA GCTGCAGGGG CTGAACGACC

         324       334       344       354       363       368       378
GTTT.GCGGG GTACATCGAG AAGGTGCATT ACCTGGAGCA GCAGAACAA. GAGAT..... CAGCGAGCTG
         346       356       366       376       386       396       406
GTTTCGCCGG CTACATCGAG AAAGTGCACT ACTTGGAACA ACAGAACAAG GAGATCGAGG CAGAGATCCA

         387       397       407       417       427       437       447
CGCGCTGC.G CAGAAGCACG CCGGGAGGGC TCAGCTGGGG GACGCCTACG AGCAGGAGCT GCGGGAGCTG
         416       426       436       446       456       466       476
CGCGCTGCGG CAGAAGCAGG CCTCGCACGC CCAGCTGGGT GACGCTTACG ACCAGGAGAT CCGAGAGCTG

         457       467       477       487       497       507       517
CGCGGGGCCC TGGAGCAGGT GAGCCACGAG AAAGGCGCAG ATTCAGCTGG ACTCGGAG CACATCGGAG
         486       496       506       515       525       535       544
CGCGCCACCC TGGAGATGGT GAATCACGAG .AAGGCTCAA GTGCAGCTGG ACTCTGAT CAC.TTGGAG

         527       537       547       557       566       576       586
GAGGAGATCC AGCGGCTGCG GGAGCGCTTC GAGGATGAGG C.CGGCTGCG AGATGAGACG GAGGCCACCA
         554       564       574       584       594       604       614
GAAGACATCC ACCGGCTCAA GGAGCGCTTC GAGGAGGAGG CGCGGCTGCG GGACGACACG GAGGCTGCCA

         593       603       613       623       633       643       653
TC..GGCCCT GCGCAAGGAG ATGGAGGAGG CCTCCCTGAT GCGGGCGGAC GTGGACAAGA AGGTGCAGTC
         624       634       644       654       664       674       684
TCCGGGCGGT GCGCAAAGAC ATAGAGGAGT CGTCGATGGT TAAGGTGGAG CTGGACAAGA AGGTTCAGTC

         663       672       682       692       702       712       721
GCTGCAGGAC GAGGT.GCCT TCCTGCCTTC CTGCGGGGCA ACCACGAGGA GGAGGTGGCC GAGCT.CTGG
         694       704       714       724       734       744       754
GCTGCAGGAT GAGGTGGCCT TCCTGCCTTC CTGCGGAGCA ATCACGAAGA GGAGGTGGCC GACCTGCTGG

         731       741       750       760       770       780       790
CCCAGCTGCA GGCTTCCCAC .GCACGGTGG AGAGGAAGGA CTACCTGAAG ACGGACCTGA CCACGGCCGT
         764       774       784       794       804       814       813
CCCAGATCCA GGCGTCGCAC ATCACCGTAG AGCGCAAAGA CTACCTGAAG ACAGACATCT CCACTG.CGC
```





```
      2521       2531       2541       2551       2561       2571       2581
GAGAAATTGGT GTCCACTAAG AAAGTGGAGA AAGTTACTTC ACATGCTGTA GTAAAAGAGA TTA..AAGA
    **       *     *    *  *  *   *  *           **      *  *  *   * ***  *
GAGAAGCTGGT GTCAACTAAA AAGGTAGAAA AGGTCACTTC ACATGCCATA GTCAAGGAAG TCACCCAGG
      2464       2474       2484       2494       2504       2514       2524

2585
ATGA
*   *********************
GTGAATTAAA CTCAATTATA GCTAA
     2534       2544  2549
```

APPENDIX:

FIGURE 2: **Comparison of Kozak Sequences** - A comparison of the nucleotide sequence surrounding the initiator methionine codon of rat NF-M (Napolitano et.al., 1987) and c19NF-M with the consensus sequence reported by Kozak (1986) is shown. In the Kozak sequence position -3 can be either an A or a G residue, although A is reported to give the strongest initiator. The G at position +1 is also reported to be very important in determination of initiator strength. The base at position -6 in the chicken sequence was not present in the cDNA. Divergence from the consensus sequence is denoted by a (^).

```
RAT:        TCCAAGATGA
                ^  ^^   ^

c19NFM      ?CCGCCATGA
                     ^

                A
Kozak       GCC/CCATGG
                G
```