1995-11-01

Molecular Characterization of Mitofilin, a Novel, Mitochondrial, Coiled Coil Protein, and the Relationship Between Organism Complexity and Coiled Coil Protein-Mediated Structure: A Dissertation

Paul R. Odgren
University of Massachusetts Medical School

MOLECULAR CHARACTERIZATION OF MITOFILIN, A NOVEL, MITOCHONDRIAL, COILED COIL PROTEIN, AND THE RELATIONSHIP BETWEEN ORGANISM COMPLEXITY AND COILED COIL PROTEIN-MEDIATED STRUCTURE

A Dissertation Presented

by

Paul R. Odgren

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences

Worcester, Massachusetts

in Partial Fulfillment of the Requirements for the Degree of:

DOCTOR OF PHILOSOPHY IN MEDICAL SCIENCES

NOVEMBER, 1995

MOLECULAR CELL BIOLOGY

COPYRIGHT INFORMATION

Parts of this dissertation have appeared in separate publications:

Odgren, PR, Toukatly, G, Bangs, PL, Gilmore, R, Fey, EG. 1995. Molecular characterization of human mitofilin (HMP), a mitochondria-associated coiled coil protein. Submitted for publication, Fall 1995.

Odgren, PR, Harvie, LW Jr, Fey EG. 1995. The phylogenetic occurrence of coiled coil proteins: implications for tissue development in Metazoa via a coiled coil tissue matrix. In press, **Proteins: Structure, Function, and Genetics**.

MOLECULAR CHARACTERIZATION OF MITOFILIN, A NOVEL, MITOCHONDRIAL, COILED COIL PROTEIN, AND THE RELATIONSHIP BETWEEN ORGANISM COMPLEXITY AND COILED COIL PROTEIN-MEDIATED STRUCTURE

Dissertation Presented
by
Paul R. Odgren

Approved as to style and content by:

Janet L. Stein, Ph.D. Chair of Committee

Sandy C. Marks, Jr., DDS, Ph.D. Member of Committee

J. Reid Gilmore, Ph.D. Member of Committee

Jeanne Bentley Lawrence, Ph.D. Member of Committee

Donald S. Coffey, Ph.D. Member of Committee

Edward G. Fey, Ph.D Thesis Advisor

Gary S. Stein, Ph.D. On-Campus Advisor

Thomas B. Miller, Jr., Dean, Graduate School
of Biomedical Sciences

Department of Molecular Cell Biology
November, 1995

Dedication

This dissertation is dedicated to the memory of my father, John Albert Odgren. Though he did not live to see its completion, his curiosity and love for nature and its mysteries have inspired me throughout.

ACKNOWLEDGMENTS

None of us stands alone in our effort to contribute to scientific progress, especially graduate students, and I would like to express my gratitude to the many friends and colleagues who have helped me to develop as a scientist during the course of this work.

To my mentor, Dr. Edward G. Fey, I am deeply grateful for his vision, the depth of his generosity both scientific and personal, his wealth of ideas, his ability to see connections among seemingly fragmented areas of research, his unflagging support, and his light-handed guidance.

Without the computer expertise of Larry Harvie, it is hard to imagine how the work on coiled coil phylogenetics would have gotten beyond the stage of thought experiments. I express my thanks for his patience with my inexperience and for his ability to translate my goals almost instantly into executable routines.

I owe a special thank you to Dr. Jeanne Lawrence and members of her laboratory, especially John McNeil, for making available their cutting-edge expertise, equipment, and advice in fluorescence microscopy.

I am also indebted to Drs. Richard Melloni and Victoria Shaloub for their invaluable assistance with RNA-related work. Gary Toukatly at Matritech, Inc., earned many thanks for his cloning and sequencing efforts in the mitofilin work, as did Bob Szaro, also of Matritech, both for his expertise with proteins and antibodies and for his prize-winning homemade wine.

To all the members of the laboratory of Drs. Janet and Gary Stein, who "adopted" me near the completion of this work, I express my gratitude -- in particular to Dr. Andre van Wijnen, T.J. Last, and Dr. Mark Birnbaum, for advice and for many enthusiastic discussions of scientific matters, and Shirwin Pockwinse, for a million things. Jack Green and Liz Buffone were of great help in supplying cells for many experiments.

I am also greatly indebted to Peter L. Bangs for sharing his expertise in many areas of molecular biology, most especially the polymerase chain reaction; to Jeannette Landrie for her excellent photographic work; to Martha Keegan of BASF Bioresearch Corporation for her enthusiastic efforts in DNA sequencing; and to the members of my Research Advisory and Thesis Committees for sharing their time and expertise.

I wish also to express special gratitude to Dr. Sandy Marks, whose remarkable example proves that excellence as a human being and as a scientist can happily co-exist in a single individual.

Finally, to my wife Dorothy, who has done more than anyone has a right to expect from a friend, colleague, or spouse. I am so glad we found each other.

TABLE OF CONTENTS

Table of Contents, Chapter 2 Materials and Methods, Cont.

Table of Contents, Chapter 3, Results, Cont.

Table of Contents, Capter 3, cont.

Discussion

ABSTRACT

In the course of experiments designed to identify and characterize structural proteins of the nuclear matrix, one antibody was generated which recognized an extraction-resistant cytoplasmic protein. This antibody was used as the starting point in the cloning and molecular characterization of a novel protein of the inter-membrane space of the mitochondrion which has been named mitofilin. Mitofilin is expressed in all human cell types, and murine homologues also exist. Mitofilin associates only with mitochondria and not with other membrane-bounded organelles such as Golgi or endoplasmic reticulum. This observation has been confirmed both by biochemical fractionation and multi-label fluorescence microscopy. Recombinant mitofilin, purified to homogeneity by affinity chromatography and preparative electrophoresis, was used to raise second-generation antibodies. Results of Western blot and immunofluorescence microscopy experiments, identical to those obtained using the original monoclonal antibody, verify the cloning and biochemical characterization. The mitofilin polypeptide contains several regions which are predicted to interact by forming coiled coils; a mitochondrial targeting signal; and a hydrophobic, membrane-spanning domain. During the course of this work, a sequence match was found with a cDNA reported by Icho, et al (1994) for a mRNA preferentially expressed in heart muscle, which they have called HMP. Evidence is presented which contradicts those authors' contention that HMP is a kinesin-like motor protein.

In the course of these investigations, methods were developed to detect and quantitate the expression of solubilization-resistant proteins of the nuclear matrix and the nuclear matrix-intermediate filament scaffold. This was accomplished by combining SDS-PAGE, high sensitivity chemiluminescent Western blots, and scanning densitometry. Sensitivity in the picogram range was obtained, and reproducibility was assessed. For semi-quantitative measurements of protein expression in tissue samples, cell number was normalized by measurement of lamin B, the major protein of the nuclear envelope. Results of screening several cell and tissue types for the expression of mitofilin and for the nuclear matrix proteins NuMA, the nucleoporin tpr, and lamin B are presented. These preliminary data suggest a

potential connection of over-expression of NuMA, tpr, and mitofilin with ovarian carcinoma. In addition, quantitative analysis of mitofilin expression in a variety of human cell types was done using purified recombinant protein antigen as the standard.

The presence of coiled coil domains in these and other proteins associated with cellular sub-structures gave rise to the third area of investigation described here. Experimental observations of the nuclear matrix-intermediate filament scaffold (NMIF), a tissue-wide structure greatly enriched in coiled coil proteins, led to the following hypothesis: that the differentiated cell and tissue architecture which characterizes Metazoa has evolved through the propagation and selective expression of genes encoding a wide variety of coiled coil proteins, and the integration of the gene products into a tissue-wide matrix based on coiled coil interactions. This hypothesis was explored by computer searches of sequence data files. The GenBank phylogenetic sequence files were examined with a heptad repeat analysis program to assess the occurrence of coiled coil proteins, how heptad repeat domains are organized within these proteins, and what structural/functional categories they comprised. Of 102,007 proteins analyzed, 5.95% (6074) contained coiled coil domains; 1.26% (1289) contained "extended" (> 75 amino acid) domains. While the frequency of proteins containing coiled coils was surprisingly constant among all biota, extended coiled coil proteins were 4-fold more frequent in the animal kingdom, and may reflect early events in the divergence of plants and animals. Structure/function categories of extended coils also revealed phylogenetic differences. In pathogens and parasites, many extended coiled coil proteins are external and bind host proteins. In animals, the majority of extended coiled coil proteins were identified as constituents of two categories: 1) myosins and motors, or 2) components of the NMIF. This scaffold, produced by sequential extraction of epithelial monolayers *in situ*, contains only 1-2% of the cell mass while accurately retaining morphological features of living epithelium. The NMIF incorporates many proteins with extensive, interrupted coiled coil forming domains. The increased occurrence of this type of protein in Metazoa compared to plants or protists supports the hypothesis that a tissue-wide matrix of coiled coil interactions underlies metazoan differentiated cell and tissue structure.

INTRODUCTION AND LITERATURE REVIEW

Part 1: Overview of thesis sections: This thesis is divided into three areas of study. One concerns the molecular, biochemical, and cell biological characterization of a previously unknown mitochondrial protein, with emphasis on laboratory investigation and interpretation of experimental results. The investigations of this protein, called "mitofilin", have established its primary structure, its ubiquitous expression in human cell types, the existence of murine homologues, its exclusive localization with mitochondria, and the topology of its mitochondrial association.

The second body of work entailed the development of Western blot methodologies for comparative and quantitative analysis of solubilization-resistant proteins. Standard immuno-assay formats such as RIA or ELISA depend upon the availability of non-denatured antigen in solution, which makes measurement of proteins involved in extraction-resistant structures such as the nuclear matrix-intermediate filament scaffold (NMIF) problematic. By performing densitometric analyses of chemiluminescent Western blots, however, these problems were overcome, and assay sensitivities comparable to those of standard immunoassays have been obtained.

The structural stability of the NMIF is due in large measure to its high content of coiled coil proteins, and leads to the final area of work presented. This involved computer-based research on the GenBank sequence database in which a hypothesis concerning the molecular basis for tissue architecture was formulated and examined. The hypothesis proposes that a continuous, tissue-wide network of coiled coil-forming proteins underlies cell and tissue structure in the Metazoa. The occurrence of coiled coil protein sequences was analyzed phylogenetically in an effort to unify, on the one hand, long-standing ultrastructural and biochemical observations of the free-standing, minimal cell substructure, i.e., the NMIF, and, on the other hand, rapidly-accumulating sequence data showing this structure to be greatly enriched in coiled coil-forming proteins. The results suggest that while a subset of proteins with extensive coiled coil domains is essential to eukaryotic life, their widespread occurrence is a hallmark of the animal kingdom.

Part 2: Mitochondrial structure, function, and cytoskeletal interactions. Mitochondria are dynamic cytoplasmic organelles which contain and maintain their own genome, replicate semi-autonomously, bifurcate and re-fuse, and are actively transported through the cytoplasmic volume (reviewed in Attardi, 1988; Bereiter-Hahn et al., 1994). Mitochondria are thought to have evolved as endocytic symbionts by interaction of bacteria with primitive eukaryotes (reviewed in Gray, 1989) and are responsible for the production of the vast majority of the ATP made by the cell. It is thought that the ancestral bacteria had evolved metabolic pathways which enabled them to use the oxygen released into the atmosphere by other organisms to decompose pyruvate, the three-carbon product of glycolysis, completely to carbon dioxide and water. This series of enzymatic reactions includes the tricarboxylic acid cycle and oxidative phosphorylation. The ancient compartmentalization of metabolism is preserved to this day, with the reactions of glycolysis carried out by cytosolic enzymes and those of the tricarboxylic acid cycle and oxidative phosphorylation proceeding in the mitochondria. Cytoplasmic glycolytic reactions yield a net of only two moles of ATP per mole of glucose consumed. The tricarboxylic acid cycle and oxidative phosphorylation contribute an additional 34 moles net of ATP, for a final total of 36 moles ATP produced per mole of glucose consumed, an eighteen-fold gain over glycolysis (Stryer, 1988, p. 420).

Differing energy requirements and distinct cellular morphologies result in wide variations in the number and the appearance of mitochondria. The morphology of mitochondria ranges from the familiar spheroidal structures seen in hepatocytes to highly complex, anastomosing, filamentous networks many microns long in growing fibroblasts (Johnson et al. 1980; Amchenkova et al., 1988).

In humans the mitochondrial genome is a circular DNA molecule of 16,569 base pairs (Anderson et al., 1981). It encodes only fourteen protein constituents of mitochondria, including three cytochrome oxidase and seven NADH dehydrogenase subunits, plus two mitochondrial ribosomal RNA's and 22 tRNA molecules. The genetic code of the mitochondrion is slightly more degenerate than for nuclear genes, and there are corresponding differences between cytoplasmic and mitochondrial tRNA anti-codons (reviewed

by Attardi, 1984). Thus, the overwhelming majority of proteins required for mitochondrial replication, structure and motility, for ATP production, and for import and export are encoded by nuclear genes. Nuclear-encoded mitochondrial proteins are translated on free cytoplasmic ribosomes and then rapidly translocated to their proper mitochondrial destination (Schwarz and Neupert, 1994).

Mitochondria posses an elaborate double membrane system which complicates the import and proper assembly of the many protein sub-units needed for ATP production. Even in a simplified view, there are many possible destinations for mitochondrial proteins: 1) the outer membrane; 2) the inter-membrane space; 3) the inner membrane; 4) the internal mitochondrial matrix space (Schwarz and Neupert, 1994; Glick et al., 1992a,c; Hannavy et al., 1993). The inner membrane, across which is established the electrochemical gradient which drives oxidative phosphorylation, is usually highly convoluted, forming invaginations into the matrix called cristae, thus maximizing the surface area which accommodates these reactions. Morphologically distinct regions of close apposition of the inner and outer membranes have been identified as sites where protein translocation is initiated (Schleyer and Neupert, 1985; Pfanner et al., 1990). These "contact sites" serve as points of entry for imported proteins.

Short amino acid sequences have been identified which act as mitochondrial matrix targeting signals (van Loon et al., 1986; von Heijne, 1986; reviewed in Hartl et al., 1989; Attardi and Schatz, 1988). These sequences consist of short stretches (12-20 amino acids) in which positively charged residues are spaced about every four residues, so that they lie along one face of a helix. The remaining residues are a mixture of uncharged amino acids with serines and/or threonines. The signals tend to lie near the amino terminus of a polypeptide and are sufficient to target normally cytoplasmic proteins to the mitochondrial matrix when spliced onto their amino termini (van Loon et al., 1986).

The electrochemical potential across the mitochondrial membranes is required for targeting signals to initiate entry of proteins into the mitochondrion, and it is postulated that this may occur through the electrophoretic migration of the basic, amphiphilic targeting region through the membranes to the electron-rich interior. If a hydrophobic stretch of residues follows the matrix targeting sequence, it acts as

a stop-transfer signal, and the protein becomes associated with the inner membrane, the bulk of the polypeptide residing in the intermembrane space (van Loon et al., 1986; Glick et al., 1992a,c). Genes for many components of the protein recognition and transport machinery have been identified (reviewed in Hannavy et al., 1993; Schwarz and Neupert, 1994).

Recognition of the targeting sequence occurs in the absence of ATP, but transport requires energy (reviewed by Glick et al., 1992b; Schwarz and Neupert, 1994). Proteins must be unfolded to be transported, as evidenced by studies using dihydrofolate reductase (DHFR) fused to the cytochrome oxidase IV targeting sequence (Eilers and Schatz, 1986). Import of the fusion protein occurred normally if the DHFR moiety was unoccupied by substrate; however, when methotrexate, a folate antagonist was added to stabilize the folded conformation of DHFR, import was halted. Both cytoplasmic and mitochondrial heat shock proteins act as chaperones for retaining nascent polypeptides in the unfolded state during translocation, and for folding them properly once import has been accomplished (Hannavy et al., 1993; Schwarz and Neupert, 1994). The targeting sequences are usually removed in the matrix by a resident protease.

Because of the central role played by mitochondria in producing the ATP required for cellular processes, mutations which affect mitchondrial function have been implicated in disease processes which affect the major organ systems (reviewed by Johns, 1995; Poulton, 1992). The autonomous, cytoplasmic, asexual replication of mitochondria results in maternal and somatic cell inheritance patterns which can be informative as to the nature of mutations underlying clinical manifestations.

The electrochemical gradient across the mitochondrial membrane which drives ATP synthesis has been successfully exploited to selectively label mitochondria in living cells with fluorescent dyes (Johnson et al., 1980; 1981; Chen, 1989). That the labels depend specifically upon the gradient was demonstrated in experiments reported by Amchenkova et al. (1988), in which mitochondria so labeled were subjected to rupture by laser micro-irradiation of a spot ≤ 0.5 μm in diameter. The effect of the

resulting leak was the loss within 30 seconds of fluorescent signal from entire networks of interconnected mitochondria, but not from others in the same cells which were not connected.

Although much is known at the molecular level about mitochondrial metabolism and electron transport, there is still much to be learned about mitochondrial movement, morphogenesis, growth, transport machinery, and cytoskeletal interactions. Studies of yeast mutants have identified proteins which regulate mitochondrial morphology, possibly through interactions with the cytoskeleton, but the mechanisms underlying the effects of the mutations are not understood (Sogo and Yaffe, 1994; Burgess et al., 1994). While mitochondria have been co-localized with actin and intermediate filaments (Summerhayes et al., 1983; Mose-Larsen et al., 1982), several lines of evidence implicate microtubules as the cytoskeletal element along which mitochondria are translocated. These include ultrastructural analyses by conventional (Heggeness et al., 1978) and quick-freeze (Hirokawa et al., 1982) electron microscopic imaging as well as in vitro studies of reconstituted systems (Brady, 1985; Martz et al., 1984). A kinesin-like protein called KIF1B was recently described which functions as a mitochondria-specific motor protein in microtubule transport assays (Nangaku et al., 1994). Like other motor proteins, KIF1B is detergent-extractable; however, unlike other motors, it appears to act as a monomer, but how KIF1B interacts with mitochondria is currently unknown.

In the work described here, a monoclonal antibody to an unidentified protein antigen was used as the starting point in the cloning and characterization of the protein it recognizes. The investigation has identified the protein as a previously unknown component of the mitochondrial inter-membrane space. Because of its α-helical coiled coil content, its filamentous appearance in fluorescently labeled fibroblasts, and its mitochondrial localization, we have named the protein mitofilin. The cDNA sequence of mitofilin is identical to that recently published by Icho et al. (1994) for a mRNA preferentially transcribed in heart muscle, HMP (heart muscle protein).

Part 3: Measurement of expression of mitofilin and nuclear matrix proteins in cells and tissues. For nearly a century and a half, it has been known that malignant tissues show characteristic

alterations in cell and nuclear shape (see Papanicolau and Traut, 1943). Indeed, despite the tremendous strides of the past two or three decades in understanding molecular events accompanying oncogenesis, it is the careful characterization of deviations from normal cellular and nuclear architecture which remain to this day the primary basis for cancer diagnosis by pathologists. Thus, of the myriad molecular mechanisms which may underlie individual cases of malignancy, they have in common the disruption of normal cell, nuclear, and tissue form.

One approach toward understanding this defining feature of cancers is to investigate the expression of proteins known to participate in cell and nuclear structure, in particular proteins of the nuclear matrix (Berezney and Coffey, 1977; reviewed by Fey et al., 1991) and the nuclear matrix-intermediate filament scaffold (NMIF; Fey et al., 1984a-b). To this end, experiments were undertaken to examine ways of assessing the levels of some NMIF proteins in cultured cells and in patient tissues. The expression of many proteins of the NMIF varies with cell and tissue type, differentiation, and malignant transformation (see, for example, Albers and Fuchs, 1992; Fuchs and Weber, 1994; Fey and Penman, 1988; Stuurman et al., 1989; Getzenberg and Coffey, 1990; also, the next section of this Introduction and Chapter 4 of this thesis). Thus, the ability to measure accurately the expression of such proteins has value for asking basic questions regarding cell differentiation and malignancy and has potential clinical utility as a diagnostic and prognostic tool.

Quantitative analysis of protein expression in many biological specimens is accomplished routinely with immunological techniques such radio-immunoassay (RIA) or enzyme-linked immunosorbent assay (ELISA). The major limitation of these techniques is not their sensitivity, which generally extends to the low picogram range, but rather that the analyte must be soluble in non-denaturing conditions which permit antigen-antibody complexes to form (Gould and Marks, 1988). This prevents their use for measuring structural proteins such as constituents of the NMIF, which by definition resist non-denaturing extraction methods, or more generally for any extraction-resistant protein.

When analyzing samples of solid tissues, a widely used alternative to these quantitative tests is immunohistochemistry (Taylor and Cote, 1994). This approach provides the tremendous advantage of providing structural information as to the localization of a particular protein within the tissue; nevertheless it suffers the major disadvantages of being neither objective nor quantitative.

The recent development of highly sensitive, chemiluminescent Western blot technologies has added precision and sensitivity not available previously with chromogenic substrates (Bronstein et al., 1989; 1992; Albrecht et al., 1991). Methods which take advantage of these improvements are reported here for the measurement of nuclear matrix and other extraction-resistant proteins.

The development of several versions of these tests is described in Chapter 3. One problem in measuring materials extracted from tissues is determining the cellularity of the specimen, particularly in tumor specimens. Secreted or cytoplasmically accumulated proteinaceous materials can confound attempts to normalize using protein measurements, and aneuploidy makes DNA determinations less than ideal for this purpose. Here the use of lamin B measurement as a means of normalizing for cellularity is described. Lamin B, a member of the intermediate filament protein family, is the major protein of the nuclear envelope (reviewed by Gerace and Burke, 1988). It was chosen as the internal control for cellularity because, unlike lamins A or C, it is expressed in every mammalian cell type and at all developmental and differentiation stages examined (Manjula et al., 1994; Hytiroglou et al., 1993; Moss et al., 1993; Lanoix et al., 1992; reviewed by Clarke, 1992). By comparing lamin B signal to that for other proteins in Western blots of tissue extracts, the relative expression of proteins in the tissue can be measured in a semi-quantitative manner. The inter- and intra-assay variability of this technique was assessed, as well as its theoretical detection limits. Preliminary data are presented in which striking differences in protein expression between normal and cancerous ovarian tissue samples were noted.

In order to evaluate whether this assay could be made truly quantitative, mitofilin was chosen as a test case. Quantitative immuno-assays require the use of a known standard which reacts with antibody with the same dose-response relationship as the unknown samples (Szymanski and Odgren, 1984;

Szymanski et al., 1984). The method described in Chapter 3 utilizes highly purified recombinant mitofilin protein to generate standard curves from which the number of molecules of mitofilin per cell was determined in a variety of cell types.

Part 4: Coiled coil-mediated protein-protein interaction in cells and tissues. This section of the thesis explores a hypothesis concerning the molecular basis for tissue structure in Metazoa. The hypothesis proposes that the evolution of differentiated cytoarchitecture, which is the defining feature of higher animals, proceeds through variations in networks of proteins which interact via extended, interrupted coiled coil domains. This coiled coil tissue matrix represents an underlying feature of the nuclear matrix-intermediate filament scaffold (NMIF), a minimal structure, produced by chemical dissection in situ, which retains cell and tissue morphology. In this section, some of these issues are approached through computer-based research on the phylogenetic occurrence of proteins which interact to form α–helical coiled coils.

The α–helical coiled coil is a feature of protein structure with important roles as a mediator of specific and stable protein-protein association and as a building block for bundles, filaments, and higher-order structures (reviewed in Cohen and Parry, 1990; Conway and Parry, 1990). Periodicity at 0.5 to 0.55 nm in the X-ray diffraction patterns of wool, first observed in the 1930's by Astbury (described in Lehninger, 1976; Perutz, 1987), was due to the coiled coils in the α-keratins; these observations both informed the prediction of the α-helix by Pauling *et al.* (1951) and provided a foundation for future crystallographic studies. The basic structural unit of the coiled coil is the heptad repeat, in which predominantly non-polar amino acids occupy positions *a* and *d* within a repeat of the form $(abcdefg)_n$ (McLachlan and Stewart, 1975). This arrangement places hydrophobic residues along one face of the helix at each turn, thereby promoting its association with a complementary surface. Most commonly, two such right-handed α-helices then twist around one another to form a left-handed super helix in which ionic forces combine with the hydrophobic interactions to stabilize both helices. Some coiled coils are three- or four-stranded, but the heptad repeat is still the underlying motif (Cohen and Parry, 1990). The

precise axial register of the coiled coil and whether the strands run parallel or anti-parallel may be largely determined by charged residues at positions *e* and *g* (Conway and Parry, 1990).

Heptad repeat domains as short as 30 amino acids (i.e., with 8 or 9 residues along the contact surface) have been shown to dimerize with specificity as to register and directionality and to be stable in aqueous solution (O'Shea et al., 1989; Oas et al., 1990; Su et al., 1994). At the other end of the scale, the coiled coil domains of β–myosins continue virtually uninterrupted over a span of a thousand residues. Between these extremes are proteins with interrupted heptad repeat regions distributed along the length of the polypeptide chain. These include, among many others, the proteins of the intermediate filament family, in which globular domains (so-called "hinge regions") seem to be interspersed with the coiled coil regions. The expression of intermediate filament proteins is highly tissue specific, and in man they are encoded by some 40 or more different genes (reviewed in Albers and Fuchs, 1992). This correlation of intermediate filament protein expression with cell type suggests a relationship between their differential expression and the variety of phenotypic morphologies observed in metazoan organisms.

The work reported here was motivated by converging results from several areas of investigation. 1) In a series of reports by Fey et al. (1984a-c; 1986a-b; 1987), methodologies were developed to facilitate the study of proteins involved in cell and nuclear shape regulation. With no a priori assumptions as to the biochemical nature of the proteins, these authors used a combination of detergent and salt extractions and nuclease digestion to produce in situ the nuclear matrix-intermediate filament scaffold (NMIF). The NMIF is remarkable in that, after removal of 98-99% of the mass of cultured epithelial cells, including the actin- and tubulin-based cytoskeletal filament systems, many features of cell and tissue morphology are accurately retained. 2) Results of GenBank database screening reported by Lupas et al. (1991) revealed that the occurrence of extended coiled coil domains among all proteins is relatively rare. 3) As work in many laboratories has provided molecular information about individual components of the NMIF, a growing number have been found to share the common secondary structural feature of extensive, interrupted, coiled coil domains. This would indicate that the NMIF represents a selective and unusual

subset of total cellular proteins. 4) Many NMIF constituents, both the intermediate filaments mentioned above, and nuclear matrix proteins (Fey and Penman, 1988; Stuurman et al., 1989; Getzenberg and Coffey, 1990) have been shown to vary with cell type and differentiation. 5) Studies of mutations which perturb the proper associations of certain NMIF constituents produce phenotypes ranging from specific tissue defects, exemplified by the pioneering studies of Fuchs and collaborators on epidermal abnormalities resulting from keratin mutations (reviewed in Fuchs and Weber, 1994), to overall developmental failure caused by *D. melanogaster* spectrin dimerization mutants (Deng et al., 1995).

Taken together, these observations led to the hypothesis that a continuous network, or matrix, of coiled coil proteins underlies differentiated tissue morphology in Metazoa, maintaining the precise arrangement of cells, nuclei, and other cytological and histological features. This implies that the mutation, propagation, and selective expression of genes encoding coiled coil proteins was a required step in the evolution of metazoan organisms, since it is precisely the variations in cell and tissue structure which distinguish them from other phyla. To explore this hypothesis, an analysis of the occurrence of coiled coil proteins in the GenBank phylogenetic sequence files was undertaken. The general organization of the heptad repeat domains within proteins of the NMIF in comparison to other structural and motility-associated proteins was also analyzed.

CHAPTER 1. IDENTIFICATION AND MOLECULAR CHARACTERIZATION OF MITOFILIN

In this chapter, the cloning and initial characterization of the antigen recognized by a monoclonal antibody is described. The antibody, clone 302.47, was raised at Matritech, Inc., (Cambridge, MA) as part of ongoing efforts to identify nuclear matrix proteins. A cDNA clone was obtained from an expression library, and its sequence revealed it to encode a novel protein. The cDNA was subcloned and used to produce purified recombinant protein, which in turn was used to produce two additional antibodies, one monoclonal and one polyclonal, both highly specific. Northern and Western blot analyses were used to characterize the cellular mRNA and protein and to verify the clone.

MATERIALS AND METHODS:

Production of first monoclonal antibody

Nuclear matrix protein extraction for use as immonogen: Nuclear matrix proteins were isolated at Matritech, Inc. (Cambridge, MA) by the method of Fey and Penman (1988) from the human cervical tumor cell line Caski. Briefly, cells were extracted with non-denaturing detergents, digested with both RNase A and DNase I, and the chromatin removed by ammonium sulfate. The resulting NM-IF scaffolds were dissolved in disassembly buffer containing urea and β–mercaptoethanol. Intermediate filaments were re-polymerized by dialysis against a reducing KCl/imidazole buffer and were then removed by ultra-centrifugation. The supernatant, greatly enriched in nuclear matrix proteins, was used for the production of monoclonal antibodies.

Antibody production: Monoclonal antibodies were prepared by Matritech. Balb/c by J mice (Jackson Laboratory, Bar Harbor, ME) were injected intraperitoneally with purified Caski NM protein every 2 weeks for a total of 16 weeks. Spleen cells were fused with the SP2/O-Ag14 mouse myeloma line using the method of Koehler and Milstein (1975). Hybridomas producing antibodies that reacted with nuclear matrix proteins were cloned and grown as ascites. One clone, 302.47, was selected which reacted with a 90 and 91 kDa protein on immunoblots.

cDNA library screening: A cDNA clone was obtained from a lambda gt11 human carcinoma cDNA expression library (Stratagene, La Jolla CA.) by screening with the 302.47 monoclonal antibody. The positive clone contained a 2.25 kb insert which was subcloned into the pBluescript II plasmid (Stratagene). The resulting plasmid, pGT1, was sequenced directly and further subcloned to produce fusion protein (see below). pGT1 cDNA sequences were obtained using the dideoxy method of Sanger et al. (1977). Double-stranded sequencing was done utilizing the T3 and T7 primers and internal primers as needed. Four independent clones were obtained by further library screening using cDNA probes derived from the pGT1 sequences. The probe for this screening was a 728 base pair PCR product primed with mitofilin-specific primers. The resulting band was excised and purified from agarose gels and then random-prime labeled with digoxygenin (Boehringer Genius system). After hybridization with the labeled fragment, positive colonies were identified by screening with enzyme-conjugated anti-digoxyenin, as instructed by the manufacturer. DNA isolated from the phage clones was sequenced by automated cycle sequencing (Applied Biosystems, Foster City, CA), and the resulting sequence data was aligned with that obtained for the pGT1 sequence using the GeneWorks program (Intelligenetics Inc., Mountain View, CA). The predicted amino acid sequence was analyzed by protein motif algorithms using GeneWorks software (Intelligenetics); by a DOS-compatible translation (Odgren et al., 1995) of the Coils 2.1 coiled coil prediction program of Lupas et al. (1991); and by PSORT signal sequence analysis software available on the Internet.

Protein immunoblotting: SDS-PAGE was performed according to the method of Laemmli (1970) using 6.5 X 9.5 cm minigels. Coomassie staining of gels was done using colloidal Coomassie stain (Sigma). For protein transfer, gels were blotted onto Immobilon P PVDF membranes (Millipore, Bedford MA) using the semi-dry method of Kyhse-Anderson (1984). Some transferred proteins were stained with Coomassie by the method of LeGendre and Matsudaira (1989). For immunoblots, whether they were to be developed by chemiluminescent (Immun-Lite AMPPD, BioRad) or by chromogenic (NBT/BCIP)

substrates for alkaline phosphatase-conjugated secondary antibodies, the blocking, washing, and antibody dilution buffers were as instructed for the Immun-Lite chemiluminescent system.

The basic buffer for this procedure is 500 mM NaCl, 0.02% sodium azide, 20 mM Tris-Cl, pH 7.5 (TBS). For blocking, 5% nonfat dry milk was added to TBS. Following transfer, blots were blocked for at least 30 minutes at room temperature, then washed for 5 minutes. Washes were done with TBS plus 0.05% Tween-20 (TTBS), and antibodies were diluted in TTBS containing 1% nonfat dry milk. Incubation of blots with primary antibody was generally for one hour at room temperature, after which blots were washed three times with TTBS. Secondary antibodies conjugated to alkaline phosphatase were then added and incubated for about one hour at room temperature, followed by another three washes with TTBS, one with TBS to remove residual detergent, and then exposed to substrate for alkaline phosphatase. For detection of AMPPD chemiluminescent signals on PVDF membranes, exposure times to X-ray film varied between 15 minutes and four hours at room temperature or overnight at 4°C.

mRNA blotting: Total cellular RNA was prepared from ME-180 cells by the guanidinium isothiocyanate procedure (Chirgwin et al., 1979). Poly(A)+ selection by oligo(dT) cellulose chromatography was performed as described by Aviv and Leder (1972). Samples of total and poly(A) selected RNA (30 μg each) were solubilized in formaldehyde-urea-vanadate (FUV) buffer and subjected to agarose-formaldehyde gel electrophoresis as described by Melloni et al. (1992), after which the gel was blotted onto Zeta-Probe (Boehringer) nylon membrane. The probe for hybridization was generated by random-primed [α-^{32}P]dCTP labeling (High Prime, Boehringer) of a 728 base pair PCR product encompassing bases 528-1256 of the cDNA. Hybridization was carried out at 42°C overnight under standard conditions (Ausubel et al., 1990). Autoradiographs were exposed for 24 to 48 hours at -70°C on pre-flashed X-ray film. The size of the mRNA band was calculated in comparison to ribosomal RNA band positions.

Recombinant protein production and purification: A fusion protein was obtained using the insert from the pGT1 construct described above by subcloning into the pMAL-C vector (New England

Biolabs Inc., Beverly MA), expressing in E. coli strain TB1, and purifying the fusion protein by amylose affinity chromatography. All experimental details of the pMAL system were carried out according to the manufacturer's instructions. The ligation was such that the first Met codon in the pGT1 cDNA insert (Met 437, Figure 1A) followed immediately the Factor Xa protease cleavage site encoded by the vector. Ampicillin-resistant colonies of plasmid-containing TB1 bacteria were isolated, grown as overnight cultures, and seeded into glucose-containing rich broth. At log-phase growth, IPTG was added to induce expression of the fusion protein. Bacteria were pelleted and lysed by freeze/thaw and sonication. The bacterial lysate was cleared by centrifugation for 30 minutes at 10,000 X g, and the supernatant was applied to an amylose affinity column. The bound proteins were eluted with maltose, and the resulting eluate was concentrated by ultrafiltration (30,000 kDa molecular weight cutoff). Cleavage of the maltose binding moiety was accomplished by incubation with factor Xa, and the released maltose-binding protein was removed by a second amylose affinity chromatography step. The flow-through from this step contained many different sizes of mitofilin polypeptides, as assessed by Coomassie-stained gels and 302.47-probed immunoblots. Full-sized recombinant protein, free of low molecular weight fragments, was purified to homogeneity by preparative SDS-PAGE using a model 491 system (BioRad, Hercules, CA) with an 8% gel column. Fractions were assayed by analytical SDS-PAGE and immunoblot, and appropriate fractions were pooled and stored at -70°C.

Anti-recombinant protein antibody production: A second generation monoclonal, clone 2-8, was raised at Matritech, Inc., using recombinant protein as the immunogen following the same procedures as described above. One NZW rabbit was also immunized over a three-month period with a total of 40 μg of recombinant protein to produce a polyclonal antiserum (3584). Specificity and titer were assayed by immunoblot (above) and immunofluorescence (see below).

RESULTS

Amino acid and nucleotide sequence characteristics: A series of immunizations was carried out in which total nuclear matrix proteins, prepared by the method of Fey and Penman (1988), were used

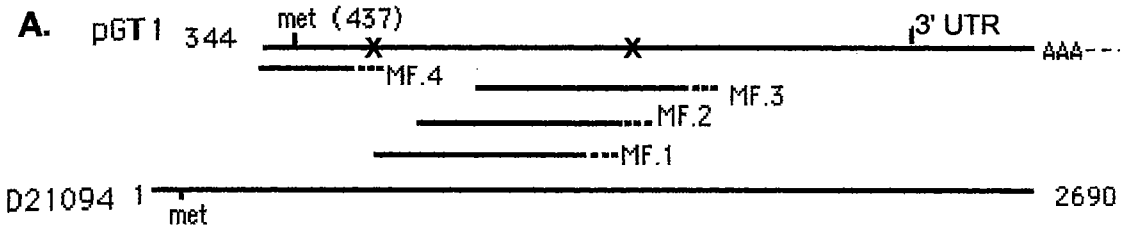
as the immunogen. While many of the resulting hybridomas recognized nuclear proteins, one clone, 302.47, recognized a protein which was present at extremely low levels in the nuclear matrix protein fraction. This antibody also showed an unusual, punctate cytoplasmic distribution in immunofluorescence studies of Triton-extracted cells (see below). Results of cDNA library screening and sequence analysis are shown schematically in Figure 1. Initial screening with 302.47 antibody of a lambda gt11 expression cDNA library from human breast carcinoma cells yielded one cDNA clone (Figure 1A, pGT1) with a 2.25 kb insert. Sequence analysis showed a single open reading frame with a 3' untranslated region, poly-adenylation signal, and a poly-A tail. Additional screening of cDNA libraries with either ^{32}P-labeled cDNA probes or with a digoxygenin-labeled cDNA PCR fragment yielded several independent clones. One of these, MF.4, while shorter, had an additional 60 nucleotides and 20 amino acids of open reading frame at the 5' end. The others (MF.1-.3) started further 3' than the original clone. Sequencing confirmed that all clones derived from various regions of the same cDNA. Additional screening both by hybridization and by PCR was carried out in an effort to obtain more 5' sequence, without success. During the course of this work, however, a match of over 99% was found in the GenBank database to sequence D21094, reported by Icho, et al. (1994) for a mRNA preferentially transcribed in growing heart muscle, which the authors called HMP (heart muscle protein). The HMP sequence included additional 5' coding sequence as well as a 5' untranslated region. The entire mitofilin cDNA sequence obtained in these investigations is shown aligned with that for the HMP sequence in Figure 2. Other than what appear to be occasional sequencing errors, they are identical.

Together, these results establish that mitofilin and HMP clones represent the same mRNA; however, experiments described below demonstrate that the protein product of this mRNA is present in all human cells tested, and further, that it is a constituent of mitochondria. Schematics of mitofilin cDNA clones are shown in Figure 1A aligned with the HMP sequence, and nucleotide positions are numbered according to the D21094 GenBank entry. Protein motif analyses revealed several important differences from those reported by Icho et al (1994). Those authors proposed that the protein contained an unusual

ATPase domain and also identified coiled coil regions based on the periodic disposition of non-polar amino acid residues in a heptad repeat pattern. While we have also identified α-helical coiled coil regions in the central portion of the protein, the coiled coil prediction program of Lupas, et al. (1991; Odgren, et al, 1995; Chapter 4 of this Thesis) identified shorter and slightly different domains which had extremely high probabilities (>99%) of coiled coil formation (Figure 1C).

Figure 1. A. Map of mitofilin cDNA clones aligned with the sequence for HMP. Clones pGT1 and MF.4 were obtained by human cDNA λgt11 expression library screening with monoclonal antibody 302.47. Clones MF.1-.3 were obtained by hybridization library screening with a digoxgenin-labeled PCR product encompassing the region between the two X's on pGT1. A ^{32}P-labeled probe made from this same region was used for Northern analysis. The insert of pGT1 was completely sequenced, and the other 4 clones were sequenced for several hundred bases to ensure that they represented genuine mitofilin cDNA. pGT1 begins at nucleotide 344 of D21904. The met (437) residue indicates the first amino acid in the mitofilin moiety of a pMal fusion construct used to produce recombinant mitofilin, the antigen used to generate additional antibodies. **B.** Coiled coil domains predicted by the Coils 2.1 program (Lupas et al., 1991). Regions with > 0.98 probability of coiled coil formation are shown with the number of the first residue of each line at the left. Amino acids occupying heptad positions *a* and *d* are shown underlined. Three stretches 5 to 6 heptads in length were identified. **C.** is a plot of coiled coil probability for the entire amino acid sequence, with amino acid number on the X-axis, and coiled coil probability on the Y-axis. **D.** Sequence and helical wheel diagram of the mitochondrial targeting signal of mitofilin. Beginning at amino acid 21 is the matrix targeting sequence (underlined) followed by the uncharged stop-translocation signal (italics). This combination of motifs results in the insertion of the amino terminus of a polypetide through the inner mitochondrial membrane. When the stop-translocation signal arrives in the membrane, transfer is halted and the bulk of the protein remains in the inter-membrane space. In the helical wheel diagram, the view down the helical axis shows the basic residues (bold-face) aligned along one face, with the remaining residues being either non-polar, or serines and threonines (open-face).

A.



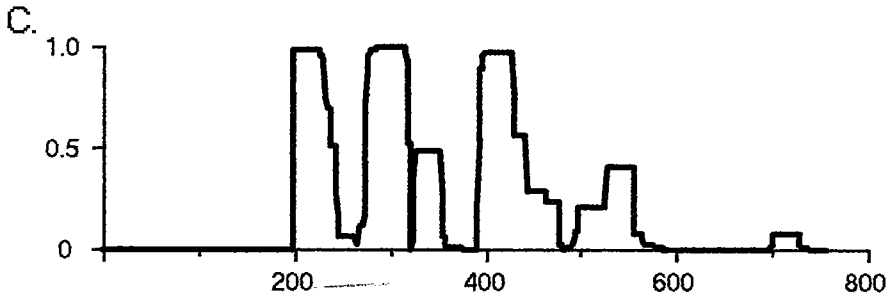
B.

203 AARLAQQ EKQEQVK IESLAKS LEDALRQ TASVTLQ

277 VEGALKER RKA VDEAADA LLKAKEE LEKMKSV IENAKKK EVAGAKP

392 DLADKLS TDDLNSL IAHAHRR IDQLNRE LAEQKAT EKQHITL

C.



D. 21 KFVLRPLRPC RRYSTSGSSG LTTGK*IAGAG LLFVGGGIGG TILY*AKWDSH

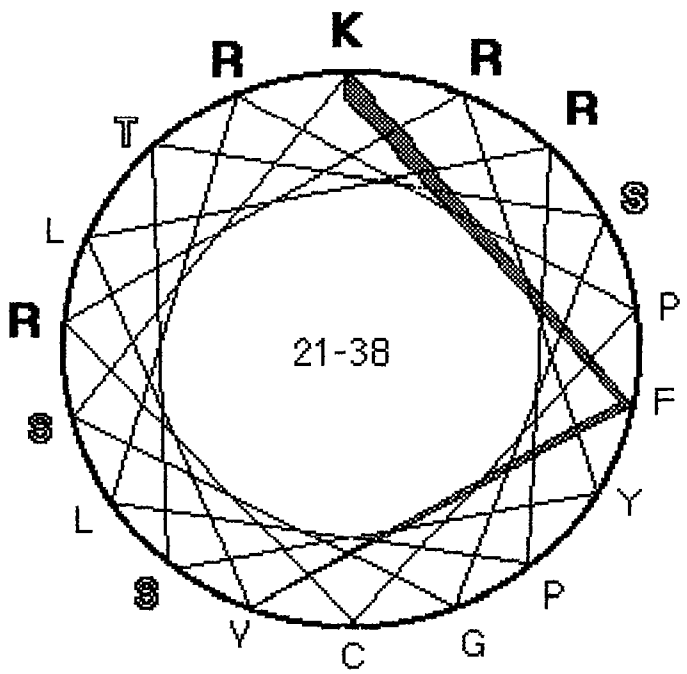


Figure 2. Mitofilin cDNA nucleotide sequence aligned with that of HMP (Icho et al., 1994; GenBank accession number D21094). Putative translation start ATG codon for HMP and the amino terminal ATG codon for recombinant mitofilin (met 437) are shown as raised text. 5' and 3' untranslated regions are shown in lower case. A polyadenylation signal (AATAAA) in the 3' untranslated region is underlined, as are mismatches between the two sequences. The kinds of differences seen suggest that many are simply occasional sequencing errors. The identity of the two sequences in their area of overlap is greater than 99%, and both encode an uninterrupted open reading frame.

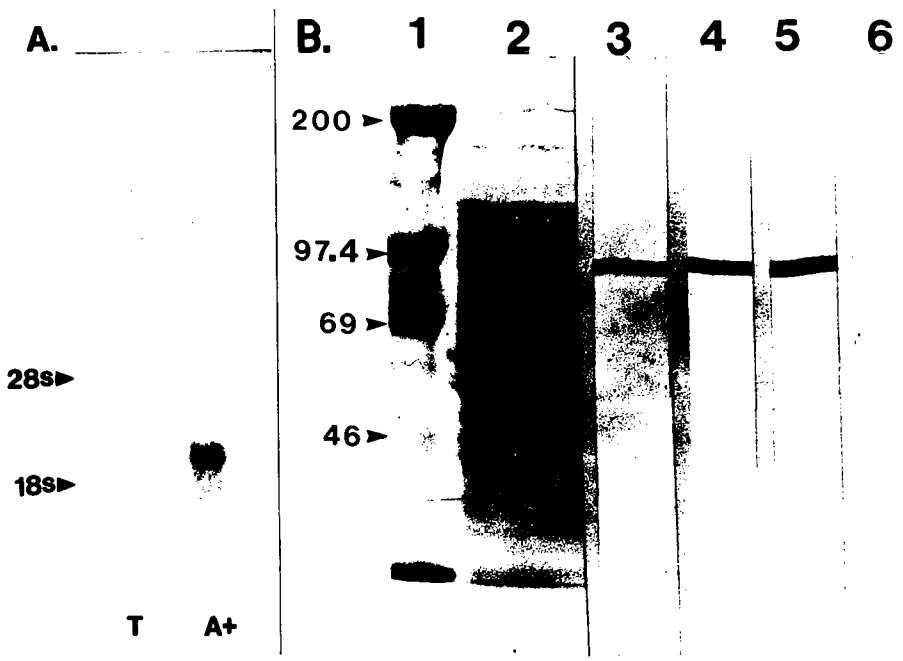
```
Mit. ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
HMP  cccggacaca cgcaagcacg cctccactta actcgcgccg ccgcggcagc tcgagtccac cagcagcgcc gtccgcttga ccgagATGCT   90

     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     GCGGGCCTGT CAGTTATCGG GTGTGACCGC CGCCGCCCAG AGTTGTCTCT GTGGGAAGTT TGTCCTCCGT CCATTGCGAC CATGCCGCAG   180
     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
     ATACTCTACT TCAGGCAGCT CTGGGTTGAC TACTGGCAAA ATTGCTGGAG CTGGCCTTTT GTTTGTTGGT GGAGGTATTG GTGGCACTAT   270

     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---GAGATGG TTCTTGGTCC    17
     CCTATATGCC AAATGGGATT CCCATTTCCG GGAAAGTGTA GAGAAAACCA TACCTTACTC AGACAAACTC TTCGAGATGG TTCTTGGTCC   360

     TGCAGCTTAT AATGTTCCAT TGCCAAAGAA ATCGATTCAG TCGGGTCCAC TAAAAATCTC TAGTGTATCA GAAGTAATGA AAGAATCTAA   107
     TGCAGCTTAT AATGTTCCAT TGCCAAAGAA ATCGATTCAG TCGGGTCCAC TAAAAATCTC TAGTGTATCA GAAGTAATGA AAGAATCTAA   450

     ACAGAATGCC TCACAACTCC AAAAACAAAA GGGAGATACT CCAGCTTCAG CAACAGCACC TACAGAAGCG GCTCAAATTA TTTCTGCAGC   197
     ACAGCCTGCC TCACAACTCC AAAAACAAAA GGGAGATACT CCAGCTTCAG CAACAGCACC TACAGAAGCG GCTCAAATTA TTTCTGCAGC   540

     AGGTGATACC CTGTCGGTCC CAGCCCCTGC AGTTCAGCCT GAGGAATCTT TAAAAACTGA TCACCCTGAA ATTGGTGAAG GAAAACCCAC   287
     AGGTGATACC CTGTCGGTCC CAGCCCCTGC AGTTCAGCCT GAGGAATCTT TAAAAACTGA TCACCCTGAA ATTGGTGAAG GAAAACCCAC   630

     ACCTGCACTT TC---AGAAG CATCCTCATC TTCTATAAGG GAGCGACCAC CTGAAGAAGT TGCAGCTCGC CTTGCACAAC AGGAAAAACA   374
     ACCTGCACTT TCAGAAGAAG CATCCTCATC TTCTATAAGG GAGCGACCAC CTGAAGAAGT TGCAGCTCGC CTTGCACAAC AGGAAAAACA   720

     AGAACAAGTT AAAATTGAGT CTCTAGCCAA GAGCTTAGAA GATGCTCTGA GGCAAACTGC AAGTGTCACT CTGCAGGCTA TTGCAGCTCA   464
     AGAACAAGTT AAAATTGAGT CTCTAGCCAA GAGCTTAGAA GATGCTCTGA GGCAAACTGC AAGTGTCACT CTGCAGGCTA TTGCAGCTCA   810

     GAATGCTGCG GTCCAGGCTG TCAATGCACA CTGGAACATA TTGAAAGCCG CCATGGACAA TTCTGAGATT GCAGGCGAGA AGAAATCTGC   554
     GAATGCTGCG GTCCAGGCTG TCAATGCACA CTCCAACATA TTGAAAGCCG CCATGGACAA TTCTGAGATT GCAGGCGAGA AGAAATCTGC   900

     TCAGTGGCGC ACAGTGGAGG GTGCATTGAA GGAACGCAGA AAGGCAGTAG ATGAAGCTGC CGATGCCCTT CTCAAAGCCA AAGAAGAGTT   644
     TCAGTGGCGC ACAGTGGAGG GTGCATTGAA GGAACGCAGA AAGGCAGTAG ATGAAGCTGC CGATGCCCTT CTCAAAGCCA AAGAAGAGTT   990

     AGAGAAGATG AAAAGTGTGA TTGAAAATGC AAAGAAAAAA GAGGTTGCTG GGGCCAAGCC TCATATAACT GCTGCAGAGG GTAAACTTCA   734
     AGAGAAGATG AAAAGTGTGA TTGAAAATGC AAAGAAAAAA GAGGTTGCTG GGGCCAAGCC TCATATAACT GCTGCAGAGG GTAAACTTCA  1080

     CAACATGATA GTTGATCTGG ATAATGTGGT CAAAAAGGTC CAAGCAGCTC AGTGTGAGGC TAAGGTTGTA TCTCAGTATC ATGAGCTGGT   824
     CAACATGATA GTTGATCTGG ATAATGTGGT CAAAAAGGTC CAAGCAGCTC AGTCTGAGGC TAAGGTTGTA TCTCAGTATC ATGAGCTGGT  1170

     GGTCCAAGCT CGGGATGACT TTAAACGAGA GCTGGACAGT ATTACTCCAG AAGTCCTTCC TGGGTGGAAA GGAATGAGTG TTTCAGACTT   914
     GGTCCAAGCT CGGGATGACT TTAAACGAGA GCTGGACAGT ATTACTCCAG AAGTCCTTCC TGGGTGGAAA GGAATGAGTG TTTCAGACTT  1260

     AGCTGACAAG CTCTCTACTG ATGATCTGAA CTCCCTCATT GCTCATGCAC ATCGTCGTAT TGATCAGCTG AACAGAGAGC TGGCAGAACA  1004
     AGCTGACAAG CTCTCTACTG ATGATCTGAA CTCCCTCATT GCTCATGCAC ATCGTCGTAT TGATCAGCTG AACAGAGAGC TGGCAGAACA  1350

     GAAGGCCACC GAAAAGCAGC ACATCACGTT AGCCTTGGAG AAACAAAAGC TGGAAGAAAA GCGGGCATTT GACTCTGCAG TAGCAAAAGC  1094
     GAAGGCCACC GAAAAGCAGC ACATCACGTT AGCCTTGGAG AAACAAAAGC TGGAAGAAAA GCGGGCATTT GACTCTGCAG TAGCAAAAGC  1440

     ATTAGAACAT CACAGAAGTG AAATACAGGC TGAACAGGAC AGAAAGATAG AAGAAGTCAG AGATGCCATG GAAAATGAAA TGAGAACCC-  1183
     ATTAGAACAT CACAGAAGTG AAATACAGGC TGAACAGGAC AGAAAGATAG AAGAAGTCAG AGATGCCATG GAAAATGAAA TGAGAACCCA  1530

     -CTTCGCCGA CA-GCAGCTG CCCACACTGA TCACTTGCGA GATGTCCTTA GGGTACAAGA ACAGGAATTG AAGTCTGAAT TTGAGCAGAA  1271
     CCTTCGCCGA CAGGCAGCTG CCCACACTGA TCACTTGCGA GATGTCCTTA GGGTACAAGA ACAGGAATTG AAGTCTGAAT TTGAGCAGAA  1620

     CCTGTCTGAG AAACTCTCTG AACAAGAATT ACAATTTCGT CGTCTCAGTC AAGAGCAAGT TGACAACTTT ACTCTGGATA TAAATACTGC  1361
     CCTGTCTGAG AAACTCTCTG AACAAGAATT ACAATTTCGT CGTCTCAGTC AAGAGCAAGT TGACAACTTT ACTCTGGATA TAAATACTGC  1710

     CTATGCCAGA CTCAGAGGAA TCGAACAGGC TGTTCAGAGC CATGCAGTTG CTGAAGAGGA AGCCAGAAAA GCCCACCAAC TCTGGCTTTC  1451
     CTATGCCAGA CTCAGAGGAA TCGAACAGGC TGTTCAGAGC CATGCAGTTG CTGAAGAGGA AGCCAGAAAA GCCCACCAAC TCTGGCTTTC  1800

     AGTGGAGGCA TTAAAGTACA GCATGAAGAC CTCATCTGCA GAAACACCTA CTATCCCGCT GGGTAGTGCG GTTGAGGCCA TCAAAGCCAA  1541
     AGTGGAGGCA TTAAAGTACA GCATGAAGAC CTCATCTGCA GAAACACCTA CTATCCCGCT GGGTAGTGCG GTTGAGGCCA TCAAAGCCAA  1890

     CTGTTCTGAT AATGTTTTCA CCCAAGCTTT AACCGCAGCT TACCCTCCAG AGTCCCTGAC CCGTGGGGTG TACAGTGAAG AGACCCTTAG  1631
     CTGTTCTGAT AATGAATTCA CCCAAGCTTT AACCGCAGCT ATCCCTCCAG AGTCCCTGAC CCGTGGGGTG TACAGTGAAG AGACCCTTAG  1980

     AGCCCGTTTC TATGCTGTTC AAAAACTGGC CCGAAGGGTA GCAATGATTG ATGAAACCAG AAATAGCTTG TACCAGTACT TCCTCTCCTA  1721
     AGCCCGTTTC TATGCTGTTC AAAAACTGGC CCGAAGGGTA GCAATGATTG ATGAAACCAG AAATAGCTTG TACCAGTACT TCCTCTCCTA  2070

     CCTACAGTCC CTGCTCCTAT TCCCACCTCA GCAACTGAAG CCGCCCCCAG AGCTCTGCCC TGAGGATATA AACACATTTA AATTACTGTC  1811
     CCTACAGTCC CTGCTCCTAT TCCCACCTCA GCAACTGAAG CCGCCCCCAG AGCTCTGCCC TGAGGATATA AACACATTTA AATTACTGTC  2160

     ATATGCTTCC TATTGCATTG AGCATGGTGA TCTGGAGCTA GCAGCAAAGT TTGTCAATCA GCTGAAGGGG GAATCCAGAC GAGTGGCACA  1901
     ATATGCTTCC TATTGCATTG AGCATGGTGA TCTGGAGCTA GCAGCAAAGT TTGTCAATCA GCTGAAGGGG GAATCCAGAC GAGTGGCACA  2250

     GGACTGGCTG AAGGAAGCCC GAATGACCCT AGAAACGAAA CAGATAGTGG AAATCCTGAC AGCATATGCC AGCGCCGTAG GAATAGGAAC  1991
     GGACTGGCTG AAGGAAGCCC GAATGACCCT AGAAACGAAA CAGATAGTGG AAATCCTGAC AGCATATGCC AGCGCCGTAG GAATAGGAAC  2340

     CACTCAGGTG CAGCCAGAGt gaggtttagg aagattttca taaagtcata tttcatgtca aaggaaatca gcagtgatag atgaagggtt  2081
     CACTCAGGTG CAGCCAGAGt gaggtttagg aagattttca taaagtcata tttcatgtca aaggaaatca gcagtgatag atgaagggtt  2430

     cgcagcgaga gtcccggact tgtctagaaa tgagcaggtt tacaagtact gttctaaatg ttaacacctg ttgcatttat attcttttca  2171
     cgcagcgaga gtcccggact tgtctagaaa tgagcaggtt tacaagtact gttctaaatg ttaacacctg ttgcatttat attcttttca  2520

     tttgctatca tgtcagtgaa cgccaggagt gctttctttg caacttgtgt aacatttttct gttttttcag gttttactga tgaggcttgt  2261
     tttgctatca tgtcagtgaa cgccaggagt gctttctttg caacttgtgt aacatttttct gttttttcag gttttactga tgaggcttgt  2610

     gaggccaatc aaaataatgt ttgtgatctc tactactgtt gattttgccc tcggagcaaa ctgaataaag caacaagatg aaaaaaaaaa  2351
     gaggccaatc aaaataatgt ttgtgatctc tactactgtt gattttgccc tcggagcaaa ctgaataaag caacaagatg ----------  2690
```

More importantly, near the amino terminus of the protein, a mitochondrial matrix targeting sequence is present (residues 21-38), followed by a hydrophobic stretch capable of acting as a stop transfer signal, as described in the Introduction. These are shown in Figure **1D**. In the helical wheel diagram, the basic residues in the target signal are seen to lie along one face, while uncharged or serines and threonines constitute the remaining residues. The "motor ATPase" domain described for HMP actually lacks the invariant lysine residue which is essential for nucleotide triphosphate binding and hydrolysis (Traut, 1994; Rapiejko and Gilmore, 1994). In fact, the putative motor domain actually lies within the stop-transfer region. This, along with results of biochemical fractionation, immunofluorescence, and protease access experiments described in Chapter 2, negates the speculation that mitofilin is a motor protein. Rather, it is a mitochondria-associated protein residing primarily in the intermembrane space of mitochondria.

Homology searches of GenBank have yielded very little in the way of potential functions for mitofilin. Nearly all the homologies identified were to short regions of myosin coiled coil domains. This helps to confirm the coiled coil regions identified with the Coils program, which uses a fundamentally different algorithm to search for coiled coil domains. The only other homologies were from part of the coiled coil nuclear matrix protein NuMA, from unidentified open reading frames obtained in genomic sequencing projects, or expressed sequence tags (not shown).

Protein and mRNA blot results: The same PCR-generated fragment used to screen cDNA libraries (above; Figure **1A**) was radio-labeled and used as a probe in Northern blot analysis of mitofilin mRNA. The results, shown in Figure **3A**, indicate that the protein is translated from a single mRNA of approximately 2.7 kb, in agreement with results reported for HMP. That the message is of low abundance in the ME-180 cervical carcinoma cell line used as the RNA source is indicated by the fact that the band was not detected in 30 μg of total RNA (lane T), and was detected when 30 μg of poly(A)-enriched RNA was

Figure 3. RNA (**A**) and protein (**B**) blot analyses of mitofilin. In **A**, total RNA (lane T) and poly(A)+ RNA (lane A+) from ME-180 cells (30 μg/lane) were prepared and blotted as described in the text. Positions of the 28S and 18S ribosomal RNA bands are indicated by arrows. The cDNA probe hybridized with a single mRNA band of approximately 2.7 kb that was not detected directly in the total RNA lane. The mRNA band indicates that mitofilin is translated from a single transcript of a size sufficient to encode a 90 kDa protein. **B** shows immunoblot identification of mitofilin. ME-180 total proteins were separated by SDS-PAGE on an 8% gel and blotted onto PVDF. Lanes 1 and 2 are Coomassie-stained markers and sample, respectively. Lane 3 was probed with monoclonal antibody 302.47, raised against total nuclear matrix proteins. Lanes 4 and 5 were probed with antibodies raised against recombinant mitofilin: monoclonal (clone 2-8) in lane 4 and rabbit polyclonal (3584) in lane 5. Lane 6 was probed with pre-immune serum from the same rabbit and at the same dilution as the immune serum. Antibodies were detected with species-specific goat anti-IgG conjugated to alkaline phosphatase and developed by chromogenic substrates. Mitofilin appears as a doublet at 90 and 91 kDa, with the 90 kDa form being predominant, and is detected specifically with all three antibodies.

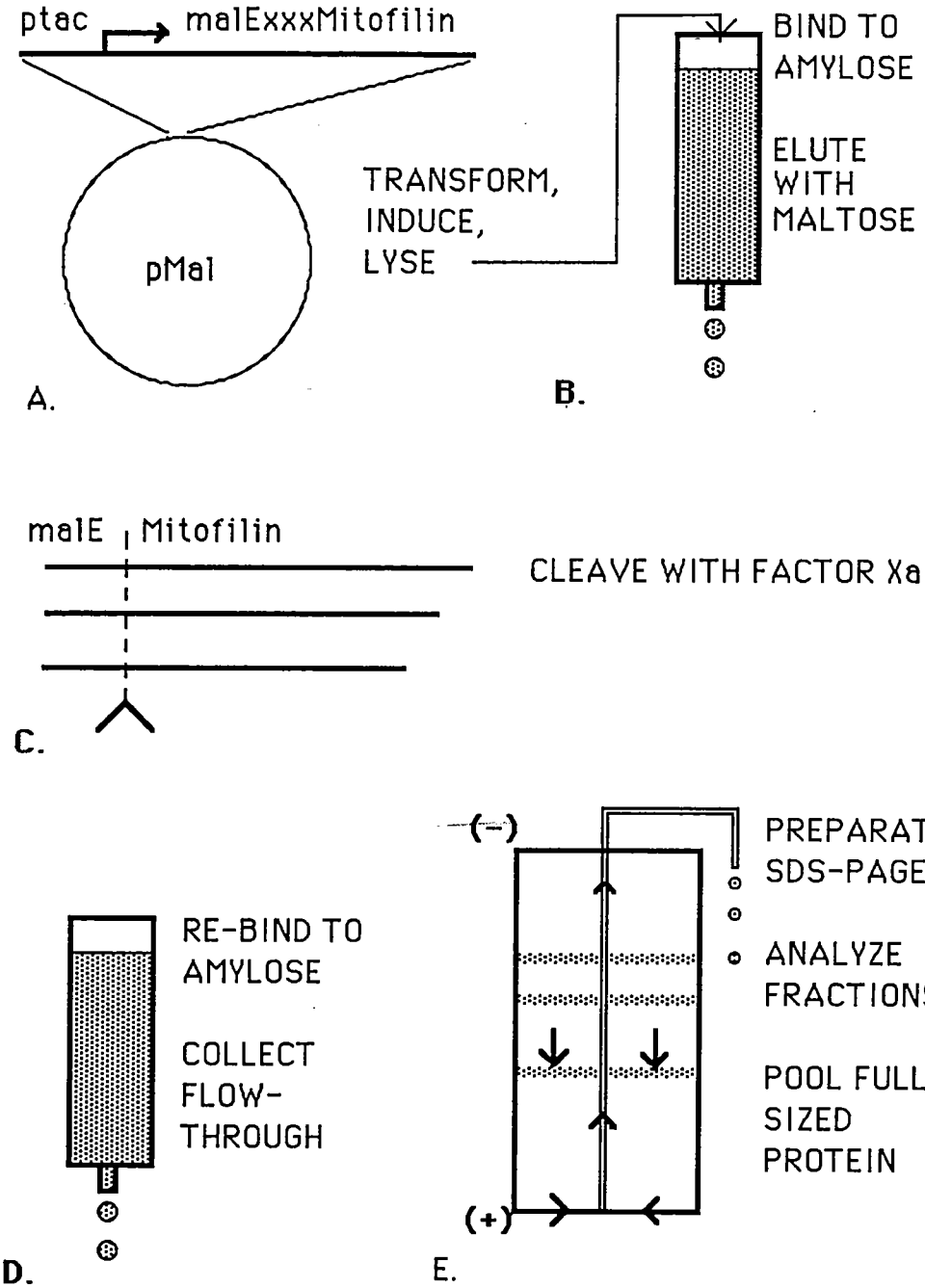

A.
28s
18s
T
A+
B.
1
2
3
4
5
6
200
97.4
69
46

loaded (lane A+). Detection even at this high loading required a 24-hour exposure to pre-flashed X-ray film. The results indicate that mitofilin is transcribed from a single mRNA.

The experimental design for the cloning and purification of recombinant mitofilin is shown in Figure 4. The insert from pGT1 was subcloned into the pMal C expression vector (New England Biolabs) and used to produce a fusion protein. The pMal construct places a 40 kDa bacterial maltose binding moiety (malE protein) at the amino terminus of the fusion protein, followed by a short peptide linker containing a cleavage site for the site-specific protease Factor Xa, and finally the mitofilin cDNA. The digestion and ligation of the cDNA were such that the Met (437) shown in Figures **1A** and **2** is at the amino terminus of the protein following cleavage, after which follows the rest of the coding region of pGT1 as well as the 3' untranslated region. Thus, the recombinant molecule consists of the carboxy-terminal 85% of the full-length cDNA-encoded protein and lacks the amino-terminal 117 amino acids.

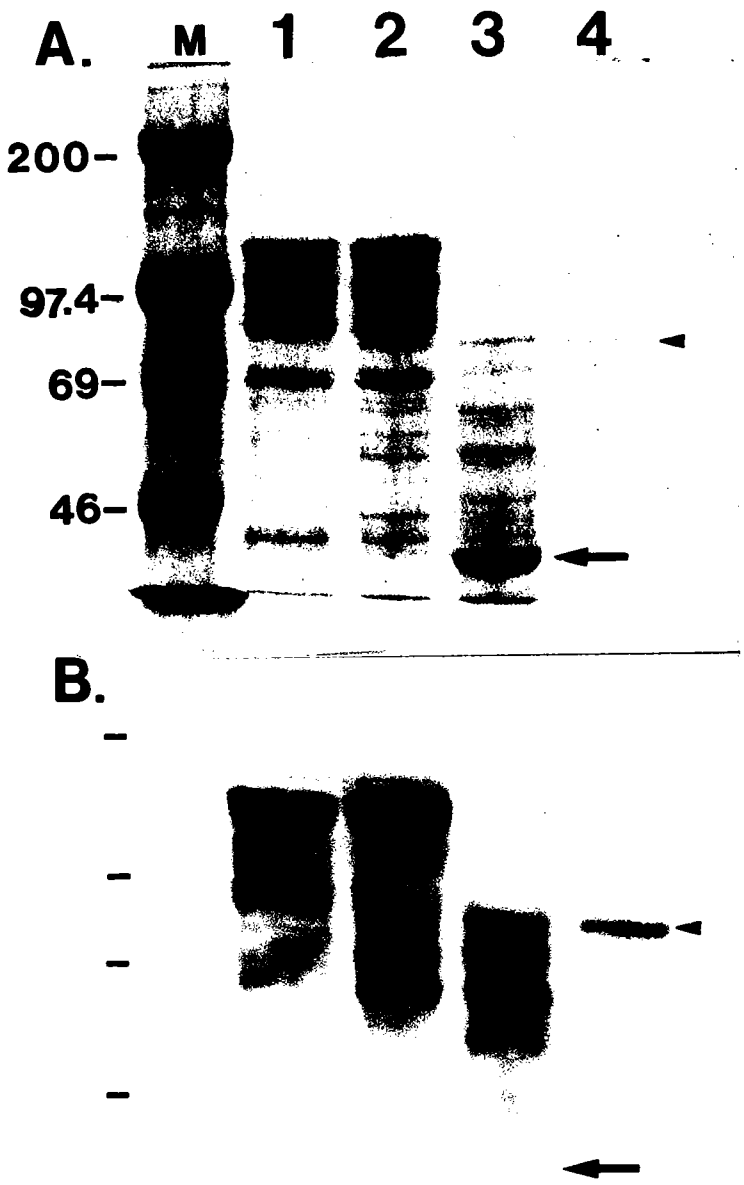
Figure 5 shows the purification of recombinant mitofilin from crude bacterial lysates containing the pMal fusion protein product. Amylose affinity chromatography resulted in an eluate (lane 1) which contained a wide range of molecular sizes of proteins bearing the 302.47 epitope. Whether these bands represent partial translation products, proteolytic fragments, or both, is uncertain. Induction conditions were varied in attempts to resolve this question, with both concentrations of the inducing agent IPTG and induction times tested over a wide range. The same, multiple bands were always observed, even with induction times as short as five minutes and lysis of bacteria directly into boiling SDS sample buffer (not shown). Lane 2 is an aliquot of the amylose column eluate after concentration by ultrafiltration. This concentrated material was subjected to proteolysis by Factor Xa (shown in lane 3), followed by re-chromatography on amylose. In this step, the maltose binding protein is bound to the column and the flow-through contains just the cDNA-encoded portion of the fusion protein. After proteolytic cleavage, the positions of bands seen both by Coomassie staining and by immunoblot probed with antibody 302.47 were shifted down by 40 kDa, indicating complete cleavage of the vector-encoded polypeptide.

Figure 4. Overview of the cloning, production, and purification of recombinant mitofilin. The pGT1 cDNA was inserted into the pMal expression vector such that its first met codon (Met 437, in Figures 1 and 2) was in frame with the carboxy-terminus of the protease cleavage site of the vector-encoded polypeptide (**A**). The protein was expressed in E. coli and purified by affinity chromatography on amylose (**B**), which yielded many fragments, shown schematically in (**C**). The vector-encoded portion of the protein was then cleaved with the site-specific protease factor Xa, after which a second round of amylose chromatography was carried out (**D**). The flow-through, which contained mitofilin cDNA-encoded polypeptides of varying lengths was then subjected to preparative SDS-PAGE to isolate the full-length recombinant protein (**E**).



PRODUCTION AND PURIFICATION OF RECOMBINANT MITOFILIN

Figure 5. Production and purification of recombinant mitofilin, analyzed by SDS-PAGE and immunoblotting. **A.** and **B.** are duplicate 8% gels of the bacterially produced malE:mitofilin fusion protein. **A.** is stained with Coomassie to visualize total proteins, and **B.** was blotted, probed with antibody 302.47, and developed by chemiluminescence. Lane M is molecular weight markers, with molecular weights indicated to the left of **A.**, and dashes indicating the same markers in **B.** Lanes 1 and 2 are the eluate from amylose affinity chromatography, used to purify the maltose-binding fusion protein from the crude bacterial lysate. Lane 1 is before, and lane 2 is after concentration by ultrafiltration. Note that many lower molecular weight forms containing the antigen were present in this eluate. In lane 3, the material shown in lane 2 has been further processed by proteolytic cleavage with factor Xa to remove the malE moiety from the amino terminus of the fusion protein. Bands have been shifted down by 40 kDa in lane 3, indicating complete removal of malE. The arrow in lane 3 shows the free malE protein, which is strongly Coomassie-stained in **A.**, but is not detected by the antibody in **B.** The malE moiety was removed by a second round of amylose chromatography, and the full-length recombinant protein was then purified to homogeneity, free of lower molecular weight polypeptides, by preparative SDS-PAGE. Fractions containing only the complete protein, as determined both by Coomassie staining and immunoblot, were pooled, and an aliquot of the pool is analyzed in lane 4. This pool was used as antigen to produce new antibodies to mitofilin.


A.
M
1
2
3
4
200–
97.4–
69–
46–
B.

The presence of fragments of various lengths necessitated further steps to isolate the full-length recombinant protein. This was accomplished by preparative SDS-PAGE using a BioRad model 491 Prep Cell. The design of this device far surpassed others which were tested in its capacity, ease of use, and resolution (data not shown). Fractions of bands eluted from the bottom of the gel were collected and analyzed by SDS-PAGE and Western blot. Fractions containing only the full-sized protein were pooled, aliquotted, and stored frozen at -70°C. The pooled product, free of lower molecular weight forms, shown in lane 4, was used as antigen to produce second-generation monoclonal and polyclonal antibodies. One monoclonal antibody (2-8) and one polyclonal rabbit antiserum (designated 3584) were raised.

When proteins from human cells were analyzed by Western blot techniques using either the original 302.47 antibody or the antibodies to recombinant mitofilin, the same 90 and 91 kDa doublet was observed (Figure **3B**). Pre-immune serum from rabbit 3584 gave no signal. As described below, these antibodies also give identical cytoplasmic signal when used in immunofluorescence experiments.

All antibodies recognize the same protein, confirm clone: When identifying a new protein based on expression screening, it is necessary to complete the circle of reasoning connecting the cDNA sequence with the genuine protein antigen. In these investigations, this was achieved through the production of the antibodies to recombinant mitofilin. The identical reactivities of all three antibodies with the closely-spaced 90 and 91 kDa doublet in Western blots, and their completely coincident labeling of morphologically distinct cytoplasmic structures in immunofluorescence (see below) confirm the identity of the clones. The two monoclonals do not bind to the same epitope on mitofilin, but rather recognize distinct antigenic sites. This was determined by immunoblot analyses both of the multiple bands seen in the bacterial expression product described above (see Figure **5**, lane 3) and of N-cholorosuccinimide cleaved recombinant protein. In both cases there were bands detected by each monoclonal which were not detected by the other (data not shown).

CHAPTER 2. BIOCHEMICAL AND CELL BIOLOGICAL ANALYSES OF MITOFILIN

Having discovered and verified that mAb 302.47 recognized a novel protein, experiments were undertaken to define possible functional roles for it in the life of the cell. This chapter describes the identification of mitofilin as a component of mitochondria. This is shown both by immunofluorescent multi-label experiments, in which mitofilin consistently co-localized with mitochondria, and by differential centrifugation, in which mitofilin co-purified with the mitochondrial fraction. Selective disruption of mitochondrial membranes followed by protease treatment indicated that mitofilin is present in the inter-membrane space of the mitochondria, a result supported by the identification of a mitochondrial matrix targeting signal and a stop-transfer membrane-spanning sequence near the amino-terminus of the protein sequence. The biochemical behavior of the mitofilin protein as a detergent-stable, salt extracted cell constituent suggests possible connections, be they direct or indirect, to the cytoskeleton.

MATERIALS AND METHODS:

Fluorescent labeling and microscopy of proteins and organelles:

Cell extraction and fixation: Human diploid fibroblasts WI38 and IMR-90, SV40-transformed IMR-90, and ME-180 human cervical carcinoma cells were grown in BME (IMR-90) or DMEM (WI38 and ME-180) medium supplemented with 10% fetal bovine serum on glass coverslips and washed twice with PBS. Non-extracted cells were fixed by three different methods. Some cells were fixed at room temperature for 15 minutes in PBS containing 3.7% formaldehyde. After fixation and washing away residual fixative, the cells were permeabilized with PBS containing 0.1% Triton X-100 for three minutes. In some experiments cells were fixed for 10 minutes in -20°C acetone, air dried, and re-hydrated in PBS. Some cells were fixed in -20°C absolute methanol for 10 minutes, after which the methanol was replaced in steps by PBS.

Cytoskeleton preparations, which retain actin and intermediate filaments but not microtubules, phospholipid membranes, or cytosol, were prepared by extracting three times for 1 minute with ice cold cytoskeleton buffer (CSK: 300 mM sucrose, 100 mM NaCl, 10 mM PIPES, 3 mM $MgCl_2$, 1 mM EGTA,

and 0.5 % Triton X-100, pH 6.8; Fey et al., 1984a,b) and fixing for 15 minutes on ice with 3.7% formaldehyde in CSK. Permeabilized cells containing intact microtubules were prepared either by fixation in -20°C methanol (Doxsey et al., 1994) or by extraction at room temperature with a microtubule stabilization buffer (MSB), a modification of that described by Mitchison and Kirschner (1984): 10 mM PIPES, 3 mM $MgCl_2$, 1 mM EGTA, and 0.5% Triton X-100. Following a 30- to 60-second extraction, cells were fixed at room temperature for 15 minutes in MSB containing 1% glutaraldehyde. After rinsing twice with MSB, glutaraldehyde active sites were quenched by three incubations of five minutes each in MSB containing 1% sodium borohydride.

Antibodies and fluorescent labeling of cells: Monoclonal and polyclonal antibodies to mitofilin were described above. Monoclonal antibody (clone 58k-9) to Golgi-associated 58k protein (Bloom and Brashear, 1989), and β–tubulin (clone TUB2.1) were from Sigma Chemical (St. Louis, MO). Monoclonal antibody to IP90 (calnexin), a resident protein of the lumen of the endoplasmic reticulum (Hostenbach et al., 1992), was kindly provided by Dr. Michael Brenner (Harvard University). Antibodies to other ER-associated proteins were also tested, including protein disulfide isomerase and Sec61, kindly provided by Dr. Reid Gilmore, UMMS. DiOC6 was from Kodak Chemical (Rochester, NY). Primary antibodies were detected with fluorochrome-labeled goat (Sigma) or donkey (Jackson Immunological Laboratories, Minneapolis, MN) anti-mouse or anti-rabbit IgG secondary antibodies which were cross-species absorbed. MitoTracker FM, phalloidin-rhodamine conjugate, and monoclonal antibody to human cytochrome oxidase subunit II (clone A-6404) were from Molecular Probes (Eugene, OR). Mitochondria were labeled in cells grown on cover slips by incubation at 37°C for 40 minutes in MitoTracker FM at a final concentration of 1 μM in medium, after which cells were fixed in formaldehyde and permeabilized as above. Following fixation, cells were washed with PBS and PBS with 0.5% bovine serum albumin (PBSA). Working dilutions of antibodies and optimal incubation times were determined empirically. Antibodies were diluted in PBSA, added to cover slips in multi-well trays, and incubated at 37°C. After removal of primary antibody followed by three five-minute washes in PBSA, secondary antibody was

added and incubated as above. In some double label experiments, primary antibodies from different species were combined in a single incubation, as were secondary antibodies labeled with different fluorochromes. Following the final antibody incubations, cover slips were washed once in PBSA containing 0.1% Triton X-100 and the DNA counterstain DAPI (14 μM), once with PBSA, twice with PBS, and then mounted with 90% glycerol containing phenylenediamine.

Selective disruption of cytoskeletal systems: SV40 transformed IMR-90 cells growing on cover slips were first exposed to MitoTracker FM (Molecular Probes) at a concentration of 1 μM in medium for 30 minutes at 37°C to label mitochondria. Actin filaments or microtubules were then de-polymerized for 30 minutes by adding cytochalasin D (final concentration 5 μg/ml; Taneja et al., 1992), or nocodazole (final concentration 10 μg/ml; Doxsey et al., 1994), respectively. De-polymerization was confirmed by staining with anti-β-tubulin or by fluorescent phalloidin. Duplicate cover slips were labeled with anti-mitofilin antibodies and detected with Texas Red- or rhodamine-conjugated secondary antibodies to assess co-localization with the green-fluorescing mitochondria.

Photomicrography: Cells labeled as above were visualized using a Zeiss Axiofot microscope equipped for epifluorescence with a 100X objective and single and multiple bandpass filter sets (Johnson et al., 1991). Images were photographed on Ektachrome film (400 ASA, Kodak, Rochester, NY), with exposure times between 2 and 60 seconds.

Sequential extraction of cellular proteins in vitro: All protein extraction solutions were supplemented with the following protease inhibitors (Bochringer Mannheim) at the concentrations indicated: bestatin, 40 μg/ml; pepstatin, 0.7 μg/ml; phenylmethylsulfonyl fluoride, 50 μg/ml; aprotinin, 2 μg/ml; leupeptin, 2 μg/ml; E64, 1 μg/ml. ME-180 cells were grown to approximately 80% confluence. Total cell proteins were solubilized in 8M urea, 2% Non-Idet P40, and 2% β-mercaptoethanol (UNM; Klose and Zeindl, 1984). The other cell protein fractions were obtained from duplicate dishes by the method of Fey et al. (1984a,b). The "soluble" fraction was obtained by extraction in ice cold CSK buffer, and "cytoskeleton" proteins by a subsequent extraction with CSK containing 250 mM ammonium sulfate.

After rinsing with CSK, the cells were scraped up into microfuge tubes, and DNA and RNA were digested for 20 minutes at room temperature by the addition of digestion buffer (DB: 300 mM sucrose, 50 mM NaCl, 10 mM PIPES, 3 mM $MgCl_2$, 1 mM EGTA, and 0.5 % Triton X-100, pH 6.8) containing 100 μg/ml each of DNase I and RNase A. Following digestion, the "chromatin" protein fraction was extracted by adding ammonium sulfate (250 mM) to the digestion buffer and collecting the supernatant after centrifugation for 5 minutes at 10,000 rcf. The resulting NMIF pellet was then solubilized in disassembly buffer (8 M urea, 20 mM MES, 1 mM EGTA, 0.1 mM $MgCl_2$, 1% β–mercaptoethanol, pH 6.6), dialyzed in the cold using 3500 molecular weight cut-off membrane (Spectra-Por, Spectrum, Los Angeles, CA) overnight vs. 3 changes of at least 1000 volumes of assembly buffer (0.15 M KCl, 25 mM imidazole hydrochloride, 5 mM $MgCl_2$, 2 mM dithiothreitol, 0.125 mM EGTA, pH 7.1; Zackroff et al., 1982), and then centrifuged at 100,000 rcf for 1 hour. The supernatant is enriched nuclear matrix proteins. The pellet, enriched in intermediate filament proteins, was re-solubilized in UNM. The soluble, cytoskeleton, and chromatin fractions were dialyzed against cold PBS. These, along with the nuclear matrix fraction, were precipitated by cold ethanol and dried under argon, after which they were re-dissolved in UNM. All fractions were aliquotted and stored at -70°C until use. Cell counts performed on trypsin-released cells from replicate plates of cells were used to normalize all fractions for cell number.

Mitochondria isolation by differential centrifugation. Mitochondria were isolated from HeLa cells grown in suspension according to standard differential centrifugation methods (O'Brien and Kalf, 1967; Carvalho Guerra, 1974). Cells were washed with PBS and an aliquot was taken to solubilize total cell proteins. This was done by brief homogenization in UNM as described above, followed by 5-minute centrifugation at 10,000 rcf at room temperature. Aliquots of the supernatant were snap frozen in dry ice-ethanol and stored at -70°C. All steps below were carried out at 0-2°C.

The remaining cells (range, 4-8 X 10^8, N = 6) were resuspended completely in 12 ml of hypo-osmotic buffer (0.15 M sucrose, 10 mM Tris, 1 mM EDTA, pH 7.5). After a two-to-five-minute incubation on ice, cells were dounce homogenized for 10 strokes with a glass pestle (A type), diluted with

an equal volume of hyper-osmotic buffer (same, but with 0.45 M sucrose) and transferred to two 15 ml Corex tubes (Corning Glass, Corning, NY). The tubes were then filled to a point 2 cm from the top with iso-osmotic buffer (IB; same as above, but with 0.3 M sucrose) and mixed by inversion. This was done to ensure reproducibility of the average distance from the center of the rotor, and thus the average centrifugal force (G_{av}) exerted on the tube contents. Tubes were then centrifuged at 1000 G_{av} for 10 minutes. The mitochondria-containing supernatant was carefully harvested into two new, 30 ml Corex tubes and diluted with IB to fill to 2 cm from top, as above. Any remaining supernatant was removed from the low-speed pellet. This first pellet was resuspended in PBS and an aliquot for the "nuclear" fraction was centrifuged at 12,000 G and then solubilized in UNM as for total proteins. The mitochondria-containing supernatant was centrifuged at 9000 G_{av} for 10 minutes, and the post-mitochondrial supernatant was harvested. The mitochondrial pellet was then resuspended in IB, transferred to two clean 15 ml Corex tubes, and washed twice more with IB at 9000 G_{av} for 10 min. An aliquot of the post-mitochondrial supernatant was precipitated by addition of trichloroacetic acid to a final concentration of 10% w/v, incubated on ice for 30 minutes, and centrifuged in a microfuge for 10 minutes at 14,000 rpm. The pellet was washed once with 70% ethanol and dried. Proteins from the low-speed pellet, the post mitochondrial supernatant, and the mitochondria were solubilized with UNM. Protein fractions were either analyzed immediately by SDS-PAGE and immunoblotting, or frozen, as above. Protein concentrations were measured by the Coomassie Plus micro-assay (Pierce Chemical, Rockford, IL) standardized with BSA.

Selective disruption of mitochondrial membranes and topological analysis by limited proteolysis: Mitochondria isolated as above were resuspended to 190-200 mg protein per ml either in IB, in osmotic shock buffer (OS: 1 volume of IB plus five volumes of 20 mM Tris, 1 mM EDTA, pH 7.5; Daum et al., 1982), or in IB containing 2% Triton X-100. OS mitochondria were rocked gently on ice for 30 minutes. Aliquots of intact, OS, and Triton-permeabilized mitochondria were then treated with trypsin (Sigma, type XIII, TPCK-treated) at concentrations of 0, 5, and 15 μg/ml for thirty minutes on ice. Reactions were stopped by adding Pefabloc SC (Boehringer) to a concentration of 12.5 μM (100-fold and

33-fold molar excess for the 5 and 15 μg trypsin doses, respectively). After a five-minute incubation on ice to allow the inhibitor to act, 6X SDS-PAGE sample buffer was added and the samples immediately placed in a boiling water bath for 10 minutes. After samples were loaded on gels for analysis, the remainder was snap frozen and stored at -70°C.

RESULTS

Mitofilin both co-localizes and co-purifies with mitochondria: Mitofilin-specific antibodies co-localize with fluorescently-labeled mitochondria in all cell types examined. This co-localization was observed reproducibly over the course of many experiments (N > 10). Figure 6 shows human cervical epithelial cells and fibroblasts double-labeled with MitoTracker FM dye (Molecular Probes) and anti-mitofilin antibodies. The morphology of mitochondria in these two cell types is quite distinct. In the fibroblasts (**6C** and **D**), mitochondria appear predominantly as an anastomosing, filamentous network, with some rod-like and vesicular forms, similar to those seen in the original publications by Chen and co-workers of mitochondria-specific fluorescent labeling with rhodamine 123 (see for example Johnson et al., 1980; 1981). In the cuboidal epithelial ME180 cells (**6A** and **B**), the mitochondria are smaller and less interconnected. In both cell types, the mitofilin and the MitoTracker signals coincide. Because the cytoplasm of fibroblasts extends as a thin layer, the morphology of mitochondria and other cell constituents is more easily seen in photographic reproductions; therefore the remaining images presented here are of fibroblasts, although similar results have been obtained with epithelial and other cell types.

Mitochondrial purification results in great enrichment of mitofilin protein: Figure 7 shows the experimental design used to isolate mitochondria from HeLa cells, and to selectively disrupt the mitochondria for topological studies. Cells were subjected to dounce homogenization and fractions were obtained by differential centrifugation according to standard methods, which have been shown by electron microscopy and biochemical analysis to yield a highly enriched and functional mitochondrial fraction (O'Brien and Kalf, 1967; Carvalho Guerra, 1974). SDS-PAGE and immunoblot analysis of the fractions obtained are shown in Figure 8. **8A** is a Coomassie-stained gel and **8B and C** are immunoblots of gels

loaded as in **A**. In **8C**, mitofilin is detected weakly in the total (lane T) and low-speed (1000 G_{av}) pellet fractions (lane N) when loaded at equal cell number. The low-speed pellet contains mostly nuclei, cytoskeletal elements, and dense nuclear-associated mitochondria and ER, while the supernatant contains a mixture of mitochondria (the densest remaining material), soluble cytoplasmic proteins, and vesicles derived from Golgi, ER, and other membrane-bounded organelles. This material was subjected to centrifugation at 9000 G_{av} for 10 minutes to produce a mitochondrial pellet and a post-mitochondrial supernatant. The post-mitochondrial supernatant (lane S) contained no detectable mitofilin, consistent with immunofluorescence results (below) which reveal no association of mitofilin with either ER or Golgi. The mitochondrial pellet was washed twice more at 9000 G_{av}, solubilized in SDS sample buffer, and loaded (lane M) at equal protein to lane T. Lane M shows that mitochondrial purification produced a great enrichment of mitofilin. In **B**, cytochrome oxidase II (COxII), detected by monoclonal antibody A-6404 (Molecular Probes), was used to confirm the enrichment of this mitochondria-specific protein in the washed pellet. COxII is encoded by the mitochondrial genome, is translated in the mitochondrial matrix, and is inserted into the inner mitochondrial membrane with both its amino- and carboxy-termini exposed to the inter-membrane space (reviewed by Capaldi, 1990). It is thus well-suited as a mitochondria-specific marker protein. At the loading used in Figure 6, the level of COxII in the total and nuclear pellet lanes was barely detectable even by high-sensitivity chemiluminescent immunoblot. In the mitochondrial pellet, however, the 26 kDa COxII band is greatly enriched. The same partitioning of mitofilin was observed in five separate experiments. The co-purification of mitofilin with COxII in the mitochondrial pellet, together with its specific co-localization with mitochondria in immunofluorescence experiments, establish that mitofilin is a previously unknown mitochondria-associated protein.

Figure 6. Co-localization of mitofilin with mitochondria. Mitochondria were labeled in living cells with a fluorescein-conjugated mitochondria-specific dye (MitoTracker FM), after which they were fixed with formaldehyde, permeabilized with Triton, and then incubated with anti-mitofilin antibodies followed by Texas Red-conjugated secondary antibodies (see Methods). **A.** and **B.** are human cervical carcinoma cells ME-180. **C.** and **D.** are SV40-transformed IMR90 human fibroblasts. In both cell types (as in all human cell types examined), the fluorescein mitochondria signal (**A** and **C**) co-localizes with anti-mitofilin antibodies. **B.** is monoclonal antibody 302.47 and **D.** is rabbit polyclonal 3584. Bar in A = 12 mm.

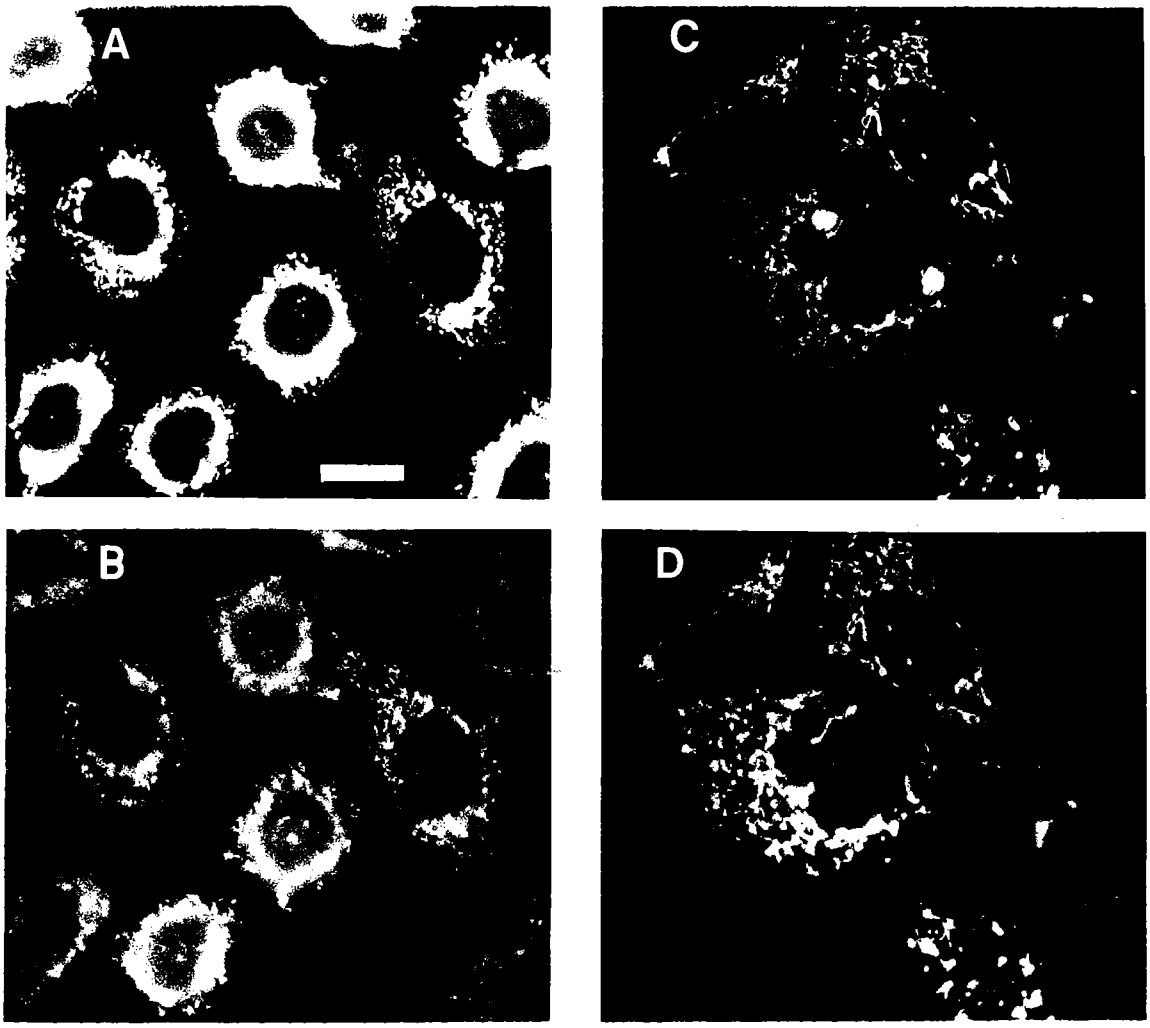
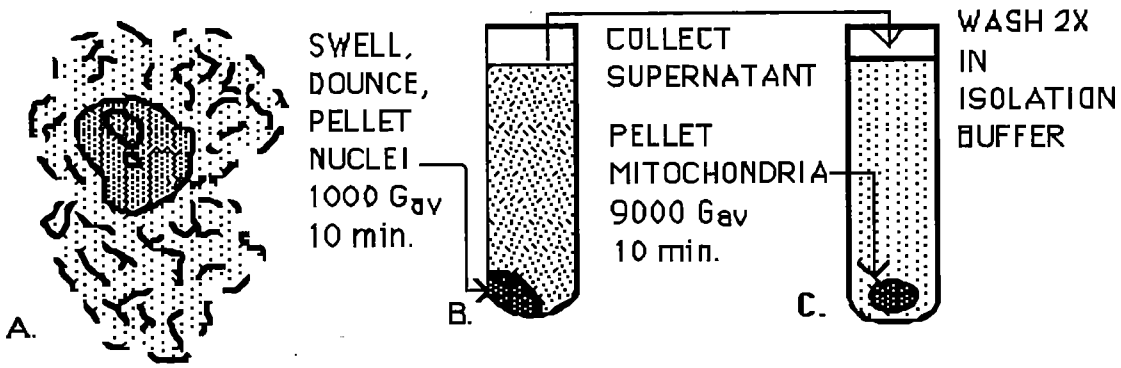
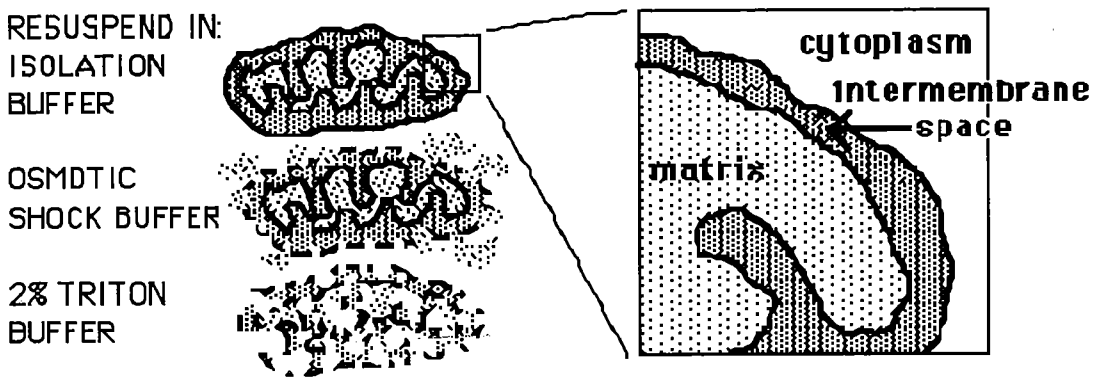

A
C
B
D

Figure 7. Diagram of experiments in which mitochondria were first isolated from HeLa suspension cells by differential centrifugation (top, **A-C**), and in some experiments were then subjected to selective disruption by osmotic shock or non-ionic detergent in order to test the accessibility of mitofilin protein to limited proteolytic degradation (bottom). Half-normal osmotic buffer induces fragility of the outer membrane (**A**) to facilitate cell rupture upon dounce homogenization. A low-speed spin (**B**) of the homogenate is used to pellet the heavy nuclei and other debris, while leaving a supernatant which contains mitochondria along with lighter organelles and soluble material. Finally, a higher speed spin of this supernatant (**C**) preferentially pellets the mitochondria. This post-mitochondrial supernatant is harvested for analysis, and the mitochondrial pellet is washed twice more with the isolation buffer. To infer topological information about mitofilin's location within mitochondria, selective permeabilization was carried out (bottom), and the preparations exposed to trypsin. The mitochondria in isolation buffer have intact inner and outer membranes. Osmotic shock induces swelling and rupture of the outer membrane, but not the inner membrane, allowing protease access to the contents of the inter-membrane space. Triton treatment solubilizes both membranes and permits protease access to the entire mitochondrial contents. The inset shows these compartments schematically.



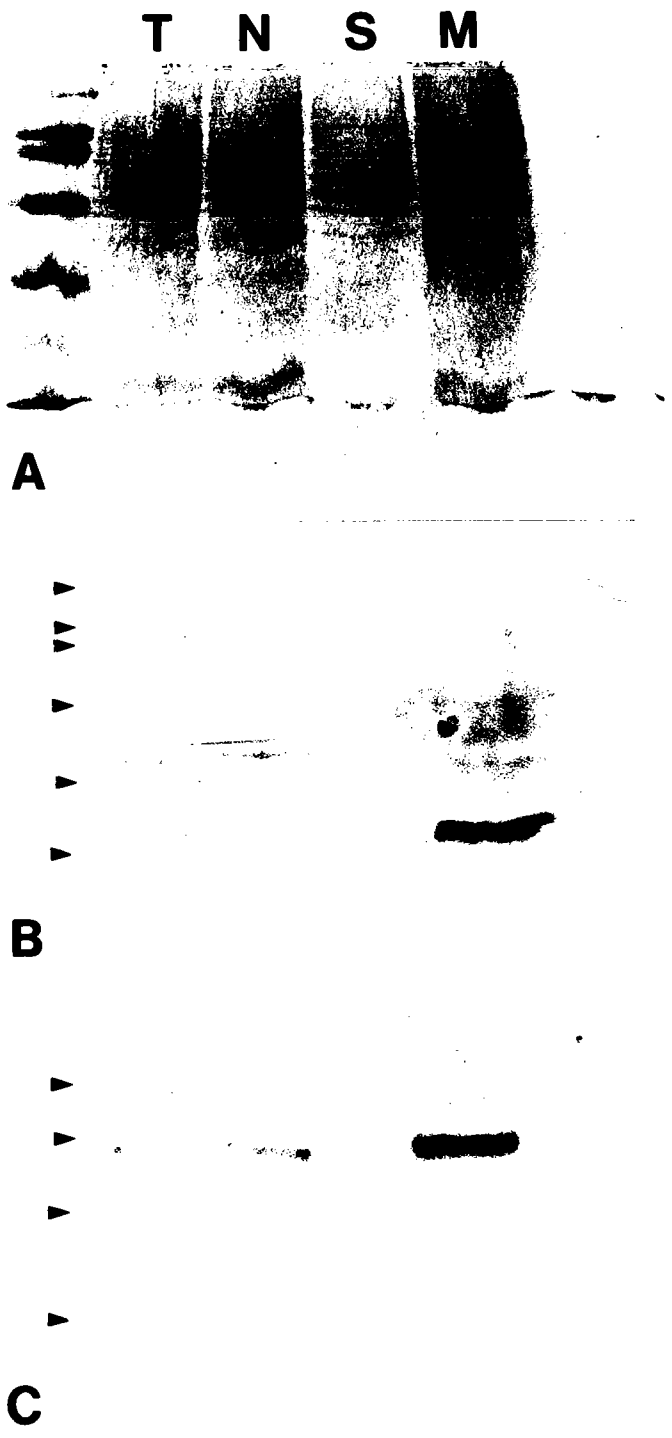
ISOLATION OF MITOCHONDRIA FROM HeLa CELLS



CHALLENGE WITH TRYPSIN, STOP DIGEST, ASSAY PROTEASE ACCESS BY WESTERN BLOT

TOPOLOGY OF MITOFILIN IN MITOCHONDRIA

Figure 8. Biochemical purification of mitofilin with the mitochondrial fraction, analyzed by SDS-PAGE. HeLa cells were grown in suspension, and mitochondria were purified by dounce homogenization and differential centrifugation. **A.** and **B.** are duplicate 12% SDS-PAGE gels of the fractions obtained. **A.** is stained with Coomassie, and **B.** is an immunoblot probed with monoclonal antibody to the mitochondrial marker cytochrome oxidase subunit II. Panel **C.** is an immunoblot of an 8% gel probed with mitofilin antibody 302.47. Weights of markers are indicated in panel **A.**, and their positions are indicated by arrowheads in **B.** and **C.**, with the 46 kDa marker being the lowest resolved in the 8% gel. Lane T represents total cell proteins obtained from an aliquot of intact cells. Lane N represents the nucleus-enriched pellet after homogenization and centrifugation at 1000 G_{av} for 10 minutes. Lane S represents the post-mitochondrial supernatant obtained after the first supernatant was re-centrifuged at 9000 G_{av} for 10 minutes. Lane M represents the 9000 G_{av} mitochondrial pellet after two further washes at the same speed. Lanes T, N, and S were loaded at equal cell numbers, and lane M was loaded at equal protein to lane T. In **B.**, the genuine 26 kDa band of cytochrome oxidase II is barely detectable in lanes T and N, even after extended exposure of the chemiluminescent immunoblot, while faint, non-specific bands are seen at other positions. As expected, lane M is greatly enriched in this enzyme. In **C.**, some mitofilin protein is seen in lanes T and N, none is detected in the post-mitochondrial supernatant, while in the mitochondrial pellet mitofilin is greatly enriched.

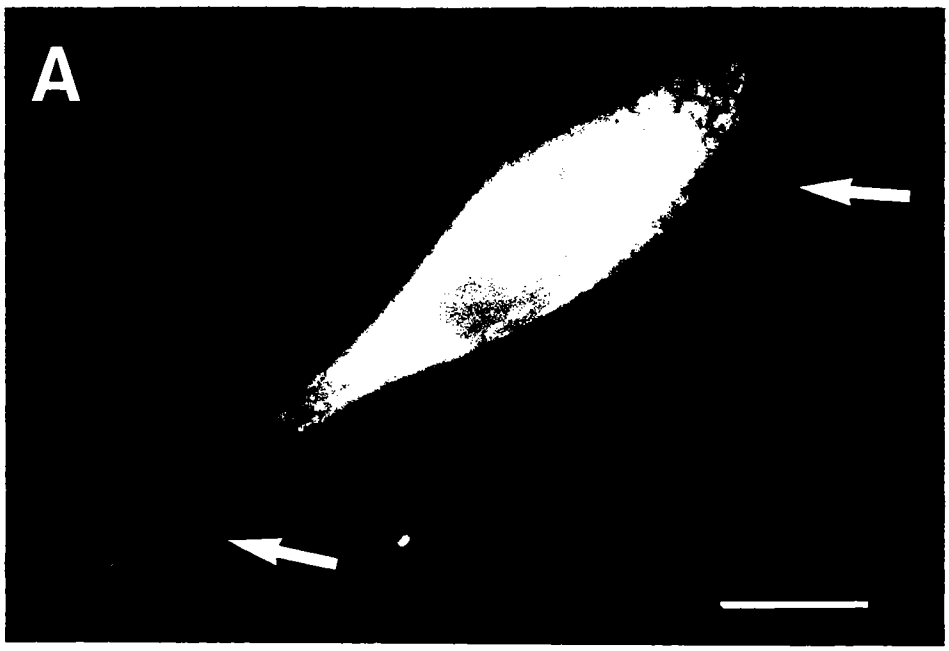

T
N
S
M
A
B
C

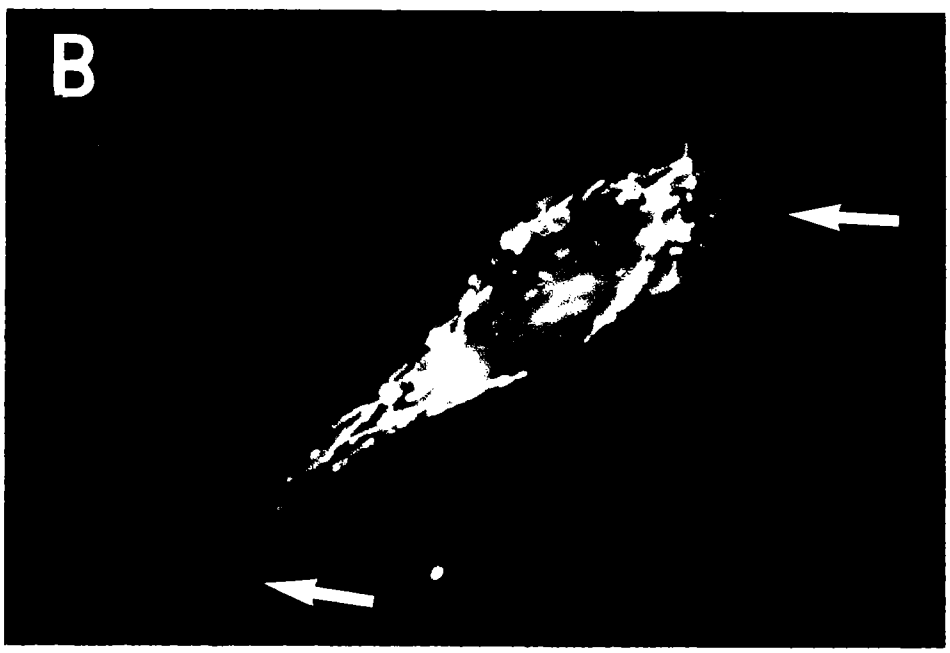
Co-localization is specific to mitochondria and not to ER or Golgi: Because the morphology of the major membrane-bounded organelle systems can appear similar to mitochondria in some cell types, double label immunofluorescence experiments were also performed with antibodies specific for Golgi and endoplasmic reticulum (ER) proteins (Figure 9). These results confirmed that mitofilin labeling was confined to the mitochondria. ER (**9A**) is shown labeled with a monoclonal antibody to calnexin, originally called IP90 (Hochstenbach et al., 1992; kindly provided by Dr. M. B. Brenner). This protein is a resident of the ER lumen, and its distribution was clearly distinguished from the mitofilin pattern (**9B**) in both its extent and morphological appearance. The ER occupies a larger area of the cell, and in the thinner, peripheral regions of the cytoplasm could be seen as a lacy, interconnected network. Mitofilin, on the other hand, was present in a smaller area of the cell, and the structures labeled were in general more rod-like and much less extensive than the ER. Similar results were obtained with antibodies to protein disulfide isomerase and to endoplasmic reticulum docking complex proteins (not sohwn), as well as with the fluorescent organic dye DiOC6 (Terasaki et al., 1984; Lee and Chen, 1988; 1993), which preferentially labels ER.

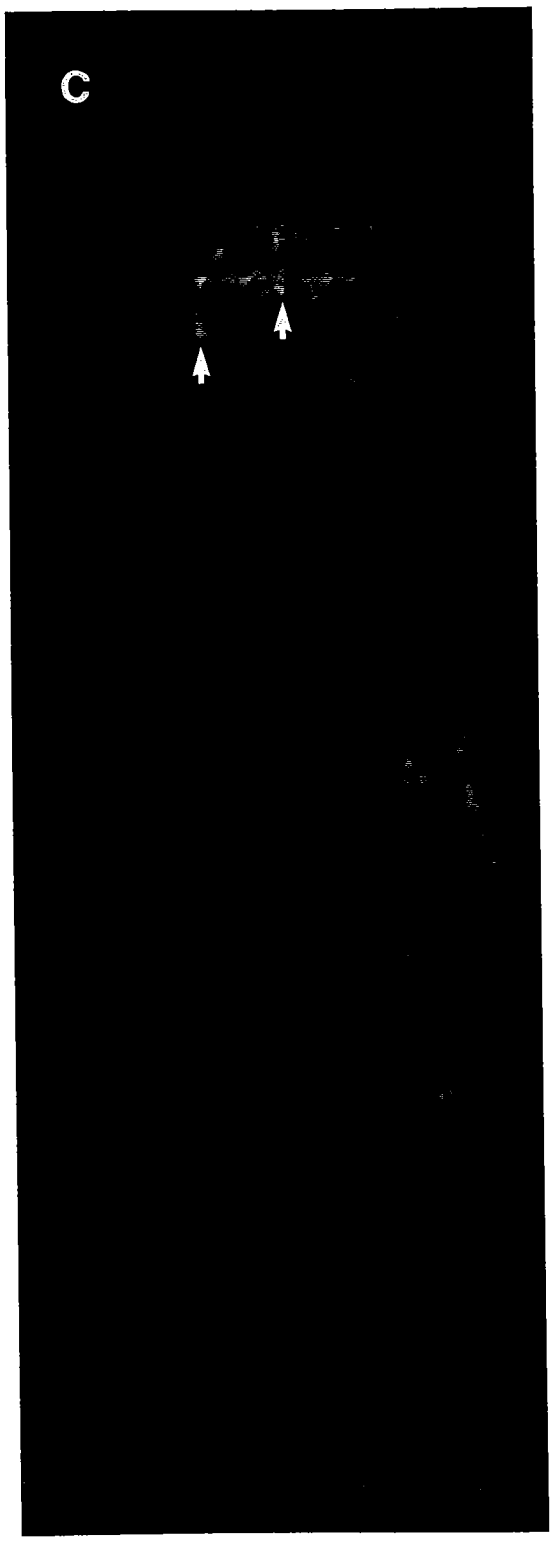
The separation of Golgi, labeled with antibody to the 58K Golgi-associated protein (Bloom and Brashear, 1989), from mitofilin signal is most wasily seen in color images (**9C**). Despite the apparent intertwining of the structures labeled by the antibodies, the red (mitofilin) and green (Golgi) signals are actually quite well separated, showing very little of the yellow signal which results from coincident red and green sources. The few yellow patches which do appear most likely originate from points where the two organelle systems are vertically stacked along the optical axis. This result was obtained in three replications of the experiment.

Figure 9. Mitofilin is associated only with mitochondria, and not with endoplasmic reticulum (**A** and **B**, opposite) or Golgi (**C**, next page). IMR90 cells were fixed in 3.7% formaldehyde and permeabilized with 0.1% Triton X-100 before labeling. In **A.** and **B.**, cells were double-labeled with monoclonal anti-IP90 (calnexin), a resident of the ER lumen, and polyclonal anti-mitofilin antibody 3584, respectively. The overall distribution of the two antigens is quite different, with the ER being much more finely articulated and occupying a greater extent of the cytoplasm. The mitochondria are distinctly more filamentous and rod-like and occupy a smaller area of the cytoplasm. The arrows denote peripheral, thinner regions of cytoplasm where mitochondria are virtually absent, but which clearly contain a lacy network of ER. **C.** shows IMR90 cells double-labeled with monoclonal antibody to the Golgi-associated 58K protein (green) and mitofilin 3584 (red), photographed with a double bandpass filter. Despite the fact that the two organelle systems appear to be closely interwoven, there is very little of the yellow signal (arrowheads) produced by coincident sources of red and green. The small amount of yellow signal probably results simply from vertical stacking of mitochondria and Golgi along the optical axis of the microscope. lumen, and its distribution was clearly distinguished from the mitofilin pattern (**9B**) in both its extent and morphological appearance. The ER occupied a larger area of the cell, and in the thinner, peripheral regions of the cytoplasm could be seen as a lacy, interconnected network. Mitofilin, on the other hand, was present in a smaller area of the cell, and the structures labeled were in general more rod-like and much less extensive than the ER. Similar results were obtained with antibodies to protein disulfide isomerase and to oligosaccharyl transferase (not shown), as well as with the fluorescent organic dye DiOC6 (Terasaki et al., 1984; Lee and Chen, 1988; 1993), which preferentially labels ER.

The separation of Golgi, labeled with antibody to the 58K Golgi-associated protein (Bloom and Brashear, 1989), from mitofilin signal is most easily seen in color images (**9C**). Despite the apparent intertwining of the structures labeled by the antibodies, the red (mitofilin) and green (Golgi) signals are actually quite well separated, showing very little of the yellow signal which results from coincident red and green sources. The few yellow patches which do appear most likely originate from points where the two organelle systems are vertically stacked along the optical axis. This result was obtained in three replications of the experiment.

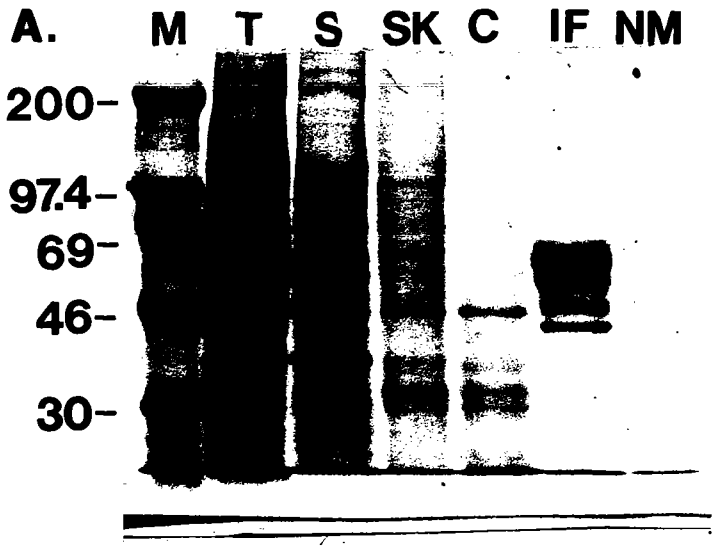

A

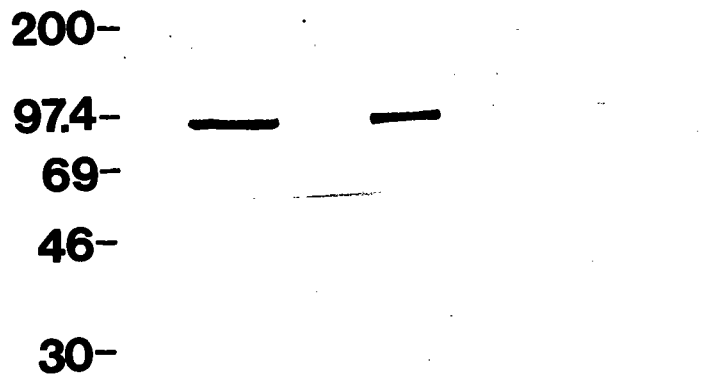

B

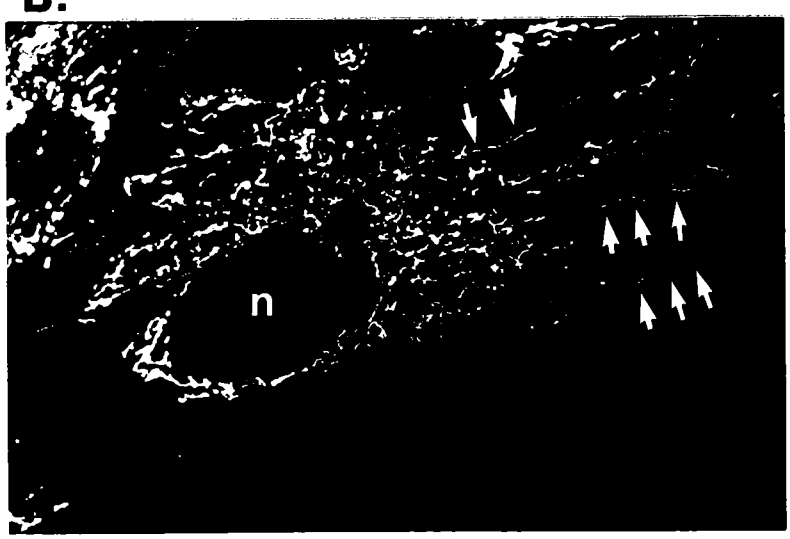

C

Resistance to non-ionic detergents: Analysis of protein fractions obtained by sequential cell extraction revealed that mitofilin is quantitatively retained after Triton X-100 treatment, but that subsequent salt extraction removes nearly all of the protein (Figure 10A). The lanes were loaded with cell extracts representing equal cell numbers. Total cell proteins, solubilized by 8M urea containing non-ionic detergent and disulfide reducing agents, are shown in lane T. This method has been shown to solubilize cellular proteins non-selectively (Klose and Zeindl, 1984); therefore, the bands seen in this lane represent the maximum mitofilin signal. Sequential extraction of cells was carried out as described by Fey et al. (1984a,b). CSK buffer, which contains 0.5% Triton X-100, extracted no detectable mitofilin protein. The protein obtained by this extraction (which has been shown to remove phospholipids, soluble cytoplasmic proteins, microtubules, and microtubule-associated proteins virtually quantitatively) is called the soluble fraction (lane S). Subsequent treatment of cells with CSK buffer supplemented with ammonium sulfate, which solubilizes the actin-based cytoskeleton and many other proteinaceous components, yields the cytoskeleton protein fraction. This treatment released nearly all the mitofilin protein from the remaining structures (lane SK). Following this extraction, nucleic acids were digested by treatment with DNase and RNase, after which another extraction was done with ammonium sulfate. The resulting chromatin protein fraction contained a small amount of mitofilin (lane C), as did the intermediate filament fraction (lane IF), which is obtained by first solubilizing the remaining proteins in a urea-containing buffer, then allowing IF's to re-polymerize and collecting them by ultra-centrifugation. The final supernatant remaining after IF removal is the nuclear matrix fraction, which is greatly enriched in structural nuclear proteins such as lamin B, NuMA, and certain nuclear pore complex proteins. This fraction (lane NM) contained no detectable mitofilin when loaded at equal cell number; however, when 10-fold greater cell equivalents were tested, the 90 and 91 kDa mitofilin doublet was faintly detected (not shown). These results were observed in three separate experiments. The presence of small amounts of mitofilin in the nuclear matrix fraction accounts for the original monoclonal antibody 302.47 being produced by immunizations with this fraction.

Figure 10. Biochemical partitioning of mitofilin in sequentially extracted cells, and the effect of non-ionic detergent on mitofilin distribution. In panel **A.**, sequential protein extracts of ME-180 cell are analyzed by duplicate SDS-PAGE gels stained by Coomassie (top) or blotted and probed with anti-mitofilin antibody 302.47 (bottom). Lanes were loaded with samples representing equal cell numbers. Extractions were carried out as described by Fey et al. (1984b; see Methods). Lane T is total cell proteins, the signal seen in lane T thus representing the maximum for the cell number loaded; lane S is soluble protein released by CSK buffer, which contains 0.5% Triton; lane SK is the cytoskeleton fraction obtained by re-extraction in CSK supplemented with 0.25M ammonium sulfate; lane Ch is the chromatin fraction; lane IF is the intermediate filament-enriched fraction; lane NM is the nuclear matrix-enriched fraction. Triton-CSK buffer extracted no detectable mitofilin protein (lane S). Subsequent treatment of cells with ammonium sulfate, which solubilizes the actin-based cytoskeleton and many other proteinaceous components, released nearly all of the mitofilin protein (lane SK).The chromatin and intermediate filament fractions both contained a small amount of mitofilin. The nuclear matrix fraction contained no detectable mitofilin when loaded at equal cell number (lane NM); however, when NM proteins were loaded at 10-fold greater cell equivalents, the 90 and 91 kDa mitofilin doublet was faintly visible (not shown). The presence of small amounts of mitofilin in the nuclear matrix fraction accounts for the original monoclonal antibody, 302.47, being produced by immunizations with this fraction. Panel **B.** shows immunofluorescence of human diploid fibroblast WI38 cells which were thrice extracted with cold CSK buffer, followed by fixation in CSK containing 3.7% formaldehyde. Labeling with antibody 302.47 shows that, despite the quantitative retention of mitofilin by cells in this treatment (see lanes S, panel **A**), the morphology of its distribution has been greatly altered. Instead of the continuous filament network seen in intact fibroblasts, the protein is present in a series of punctate bodies, some of which lie along filamentous trajectories (arrows).


A.
M
T
S
SK
C
IF
NM
200-
97.4-
69-
46-
30-


200-
97.4-
69-
46-
30-


B.
n

The effects of fixation and non-ionic detergent on mitofilin organization: The biochemical fractionation results above would seem to suggest that mitofilin is stable during extraction with detergent; however, immunofluorescence (Figure **10B**) shows that, at the structural level, CSK extraction has a marked effect. Despite the fact that all the mitofilin protein stays attached to structures remaining in the cell after this treatment, the protein is re-organized into an array of large punctate spots and short strands which are often arranged along filamentous trajectories. The "contact sites" between the inner and outer mitochondrial membranes have been shown to behave similarly in certain ways (reviewed in van der Klei et al., 1994). In rapid-freeze and freeze-substitution EM studies of whole cells, contact sites occupy most of the non-cristae area of the inner membrane, whereas in isolated mitochondria, they are confined to smaller regions; and in "mitoplasts", which are stripped of their outer membranes by digitonin extraction, the contact sites persist, and often retain small areas of the outer membrane. Images such as the one in **10B** imply that mitofilin is directly or indirectly attached to cytoskeletal elements.

In order to assess whether the structural disruption was due to effects of Triton X-100 or to the de-polymerization of microtubules, which occurs in this treatment, cells were examined after fixation by methods which retain microtubules intact. Extraction at room temperature with a Triton-containing microtubule stabilization buffer which is very similar to CSK (see Methods) preserves microtubule ultrastructure (Mitchison and Kirschner, 1984). This treatment affected mitofilin architecture in a manner indistinguishable from cold CSK, while microtubules remained intact (not shown), indicating that the disruption of the mitofilin organization by Triton occurs whether or not microtubules are also removed. Non-extracted cells which had been fixed at -20°C in either methanol or acetone were also examined. The effect of these organic solvents on the appearance of the mitofilin in fibroblasts was intermediate between PBS formaldehyde (Figures 6 and 9) and Triton (**10B**). That is, some was present as punctate bodies, but much of the mitofilin remained as filamentous structures in these cells (not shown).

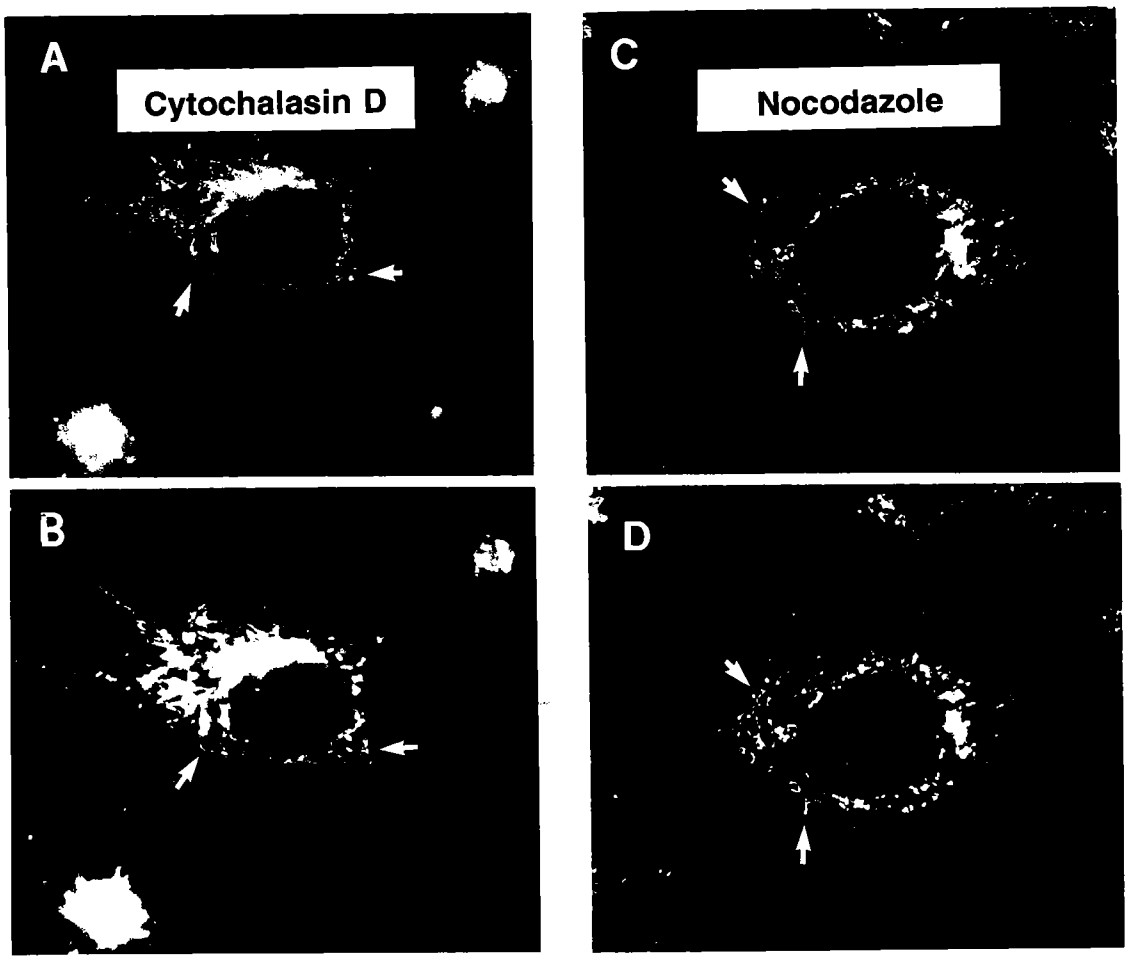
Association with mitochondria unaffected by actin or tubulin de-polymerization: The possibility that the association of mitofilin with mitochondria might be disrupted by the de-polymerization

of actin and tubulin was investigated for several reasons. First, it is well-established that mitochondria bind to and move along microtubules; and Nangaku et al. (1994) recently identified KIF1B, a mitochondria-specific microtubule-associated motor protein with three short coiled coil domains similar to those we noted in the mitofilin sequence (Figure 1C). We therefore wished to examine whether the association of mitofilin with mitochondria would persist in the absence of microtubules. And, as noted above (Figure 10), although Triton extraction leaves mitofilin behind quantitatively, it appears greatly altered in its morphology; moreover, subsequent treatment with 250 mM ammonium sulfate, which removes the actin cytoskeleton, also extracted mitofilin nearly quantitatively, suggesting a possible interaction between mitofilin and actin.

In immunofluorescence experiments in which either actin filaments (Figure 11A and B) or microtubules (Figure 11C and D) were disrupted with nocodazole or cytochalasin D, respectively, the co-localization of mitofilin with MitoTracker was unaffected. Duplicate cover slips included in the experiment and stained with anti-β–tubulin antibody or with phalloidin-rhodamine conjugate confirmed that de-polymerization had occurred (not shown). In the nocodazole-treated cells (11C and D), which lack microtubules, there was no readily discernible effect on either cellular or mitochondrial morphology by the thirty-minute time point at which cells were fixed, nor was the mitochondrial localization of mitofilin affected. Similarly, although actin de-polymerization had a marked effect on the morphology of the fibroblasts, most notably a rounding up and shrinking back of the cytoplasm toward the nucleus, the mitofilin signal remained coincident with mitochondria (11A and B). This was observed in two separate experiments, with multiple cover slips for each condition tested.

Homologue of mitofilin is present in mitochondria of murine cells: Western blots were performed on extracts of mouse 3T3 cells, rat osteoblast cells, and total proteins solubilized from rat liver tissue using antibodies 302.47, 2-8, and 3584. In addition, 3T3 cells were examined by immunofluorescence with all three antibodies. In both mouse and rat, a single band at approximately 90 kDa was seen in Western blots with 302.47 monoclonal and with 3584 rabbit polyclonal antibodies, while

Figure 11. Mitochondrial association of mitofilin is independent of actin filaments or microtubules. SV40 transformed IMR90 cells were incubated for 30 minutes in medium containing MitoTracker dye, after which either cytochalasin D, to depolymerize actin (**A** and **B**, 5 μg/ml final concentration), or nocodazole, to depolymerize microtubules (**C** and **D**, 10 μg/ml final concentration) , were added to the medium for an additional 30 minutes of incubation. **A** and **C** show MitoTracker FM (green) fluorescence, and **B** and **D** show anti-mitofilin antibody 3584 (red) signal. Nocodazole-treated cells showed no conspicuous alteration of overall morphology or of the appearance of the mitochondria compared to untreated cells. Cytochalasin D treatment induced collapse of the cytoplasm and rounding up of the cell body and nucleus, while leaving some thin cytoplasmic projections. Some conspicuous structures are denoted by arrows for comparison. Cells processed in parallel and stained for actin or tubulin confirmed the disruption of the filaments (not shown). Neither treatment disturbed the co-localization of mitofilin protein with mitochondria.


A
Cytochalasin D
C
Nocodazole
B
D

no band was seen with 2-8 monoclonal antibody. In immunofluorescence experiments, 3584 and 302.47 both stained the cytoplasmic mitochondrial network, whereas 2-8 did not. These results strongly suggest that mouse and rat homologues of mitofilin exist, although cloning and detailed molecular analysis of them has not been carried out. The lack of recognition by the 2-8 antibody confirms that the two monoclonals recognize different epitopes (see above). It also implies that mapping the 2-8 epitope should pinpoint a non-conserved domain of mitofilin. This gives 2-8 potential utility in expression studies of human mitofilin in murine cells for mapping functional domains.

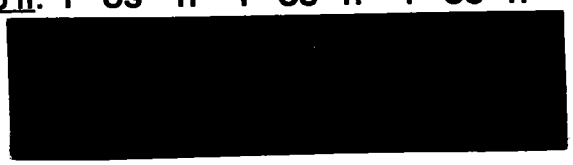
Mitofilin resides in the mitochondrial inter-membrane space: In order to assess the topology of mitofilin within mitochondrial compartments, aliquots of isolated mitochondria were exposed to trypsin after three different treatments (Figure 12). One aliquot was intact mitochondria resuspended in IB (Lanes I). A second aliquot was osmotically shocked (lanes OS) by exposure to 1/6 normal osmotic buffer for 30 minutes, as described by Daum et al. (1982). This treatment ruptures the outer mitochondrial membrane and allows access of protease to the intermembrane compartment. A third aliquot was treated with 2% Triton X-100 to solubilize both the inner and outer membranes and allow protease access to all compartments (lanes T). The samples were divided into 100 µl aliquots, each containing 190 µg/ml of mitochondrial protein. As indicated in the Figure 12, these were treated with final concentrations of trypsin of 0, 5, or 15 µg/ml and incubated on ice for 30 minutes. The digestions were stopped by adding Pefabloc, an irreversible serine protease inhibitor. Pefabloc was chosen because it is water-soluble to high concentrations, and it is stable in aqueous environments, unlike phenylmethylsulfonyl fluoride. It could therefore be added to great excess in order to stop the digestions completely before the mitochondrial preparations were solubilized in SDS-PAGE sample buffer.

In the intact mitochondria (lanes I), mitofilin was protected from proteolysis. In the osmotically shocked mitochondria (lanes OS), trypsin degraded mitofilin quite effectively. In the Triton-treated samples (lanes Tr), the highest dose of trypsin degraded mitofilin below detectable levels. Interestingly, the 5 µg/ml trypsin treatment yielded mainly a partial digestion product which ran approximately 9.6 kDa

smaller thanthe native protein. This same band can be seen faintly in other lanes (5 I; 5, OS; and 15, I; 15, OS), indicating preferential cleavage at an exposed hyper-sensitive site. Its relative accumulation in the 5 μg/ml-plus-Triton lane is likely due to increased competition for the limited enzyme by the detergent-released contents of the mitochondrial matrix, since just three-fold more trypsin was sufficient to digest mitofilin completely (lane 15, Tr). The fact the there was some degradation and hyper-sensitive site cleavage in the intact mitochondria exposed to 15 μg/ml trypsin could mean that the hyper-sensitive site is exposed on the mitochondrial surface. Alternatively, and more likely, it may simply reflect proteolysis-induced loss of outer membrane integrity.

Figure 12. Determination of mitofilin compartmentalization by sensitivity/resistance to limited trypsin digestion. Intact mitochondria (lanes I), osmotically shocked mitochondria (lanes OS), or mitochondria treated with 2% Triton (lanes T) were exposed to trypsin at 0, 5, or 15 μg/ml, as indicated, for 30 minutes at 0°C. Samples were separated by SDS-PAGE, blotted, and probed with mitofilin antibody 302.47. The left-hand three lanes show the starting level of mitofilin in the three mitochondria preparations. Osmotic shock, which permeabilizes the outer mitochondrial membrane, was sufficient to expose mitofilin to tryptic digestion at both enzyme concentrations. The lower band seen in several of the lanes indicates attack at a trypsin hyper-sensitive site. Interestingly, at the 5 μg/ml trypsin treatment of Triton-permeabilized mitochondria, this partial digestion product accumulated, perhaps due to competition for the enzyme from the released mitochondrial matrix proteins. Three-fold more enzyme was sufficient for complete digestion in lane Tr. In the intact mitochondria, mitofilin was largely protected.

Trypsin:	0			5 ug/ml			15 ug/ml		
Mitoch. Prep'n:	I	OS	Tr	I	OS	Tr	I	OS	Tr



I = Intact, OS = Osmotically Shocked, Tr = Triton X-100 Permeabilized

DISCUSSION

The purification of mitofilin protein in the mitochondrial pellet together with its co-localization with mitochondria by immunofluorescence establishes it as a previously unknown mitochondria-associated protein. The presence of a small amount of mitofilin in the nuclear matrix preparations used in the original immunizations accounts for the generation of the mitofilin-specific hybridoma 302.47, especially considering the high antigenicity of the recombinant protein (see Figure **3B**). The identical reactivities of both a monoclonal and a polyclonal antibody raised against highly purified recombinant protein verify the identity of the cDNA clone obtained with the original monoclonal antibody and confirm the Western blot and immunomicroscopic results obtained with mAb 302.47.

The cDNA sequence for all mitofilin clones obtai ned both by expression screening with antibody and by hybridization screening with mitofilin cDNA-derived probes all matched those of the D21094 HMP sequence (Icho et al., 1994) by >99%. Furthermore, a single 2.7 kb mRNA band was observed for mitofilin, as was seen for HMP. These findings demonstrate that mitofilin and HMP clones in fact represent the same mRNA.

All three of these antibodies have been used extensively in immunoblot studies of mitofilin expression in a variety of human cells and tissues. In addition to the fibroblasts and cervical epithelial lines described above, these included peripheral blood lymphocytes; erythrocytes; MG-63 osteosarcoma cells; A3-01 transformed lymphocytes; many normal and cancerous ovarian and cervical tissues; melanoma cells; MCF7 and Caski breast carcinoma lines; and six different ovarian carcinoma cell lines (see Chapter 3, Table 1). Excepting only the red blood cells (which lack mitochondria), the same 90 and 91 kDa mitofilin doublet was observed in every case.

The name HMP (heart muscle protein) used by Icho et al. (1994) reflects their observation that the mRNA was "preferentially transcribed" in growing heart muscle, which they demonstrated by Northern blot analysis of different tissues. Their report, however, contained no biochemical analysis of the

protein product of the mRNA. The experiments described here demonstrate that the mitofilin protein is expressed ubiquitously in all human cells possessing mitochondria; however, the greater abundance of mRNA in heart muscle is certainly consistent with the extremely high numbers of mitochondria in cardiac myocytes (Michaels et al., 1982).

Because of the reorganization of mitofilin which occurs upon extraction with cold Triton X-100 (see Figure 8), we investigated whether its mitochondrial association was affected by the removal of microtubules or actin filaments, and whether the re-organization was due primarily to the detergent. The drug-induced depolymerization of either filament system had no discernible effect on the disposition of mitofilin along the entire length of the mitochondria. When cells were Triton-treated under conditions which preserve microtubules, the same re-organization of the protein signal seen in cold CSK treatments occurred. The resistance to extraction implies that mitofilin is directly or indirectly attached to other cellular structures. The fact that it is retained quantitatively through three Triton extraction steps, with none detected in the supernatant, argues strongly against artifactual binding produced de novo by detergent exposure. The interrupted coiled coil domains of mitofilin very likely provide sites for its stable interaction with itself and/or with other proteins (Cohen and Parry, 1990). Such domains are most commonly found in microtubule-associated motor proteins, in structural proteins of the nuclear matrix-intermediate filament scaffold, or in subunits of certain holoenzymes (Odgren et al., 1995; Chapter 4 of this thesis).

In their description of the HMP cDNA sequence, Icho et al. (1994) noted coiled coil regions reminiscent of kinesin. In addition, they identified a region with some similarities to known nucleotide triphosphate-binding domains (residues 45-61: KIAGAGLLFVGGGTGGT). Together, these observations led them to conclude that the protein was a heart muscle-specific motor protein. The domain thought to be a potential NTP-binding motif, however, lacks the key lysine residue which is absolutely required for binding to occur (i.e., the penultimate G should be a K; Walker, et al., 1982; Traut, 1994; Rapiejko and Gilmore, 1994). Moreover, unlike all known motor proteins, mitofilin is not extracted in cold Triton-

containing buffers, nor is it present in supernatants of cell homogenates after the 100,000 G spins commonly used to isolate microtubule-associated motor proteins (see Figure 8). Only one mitochondria-specific motor protein has to date been identified, KIF1B (Nangaku et al., 1995), and its biochemical purification is similar to that of other motor proteins. Thus, despite the predicted amino acid sequence possessing certain kinesin-like characteristics, by other criteria mitofilin is unlikely to function as a motor. In fact, the putative NTP-binding domain lies within a region encompassing both the mitochondrial targeting and stop-transfer sequences shown in Figure 1D, from residues 21-65.

The amino acid sequence of mitofilin is thus informative in reconciling various experimental observations and formulating a model for the nature of its association with mitochondria. As has been shown by experiments using chimaeric proteins in which leader sequences from cytoplasmically-translated mitochondrial proteins are fused to cytoplasmic proteins, these leaders can act as matrix targeting signals (van Loon et al., 1986; Eilers and Schatz, 1986). Such targeting sequences, positioned near the amino terminus of a polypeptide, consist of three or four positively-charged residues which are spaced at about every fourth residue so as to align along one face of a helical stretch, with the remainder of the sequence consisting of a mixture of non-polar residues with serines and threonines. Further specification of the final destination is achieved if the matrix targeting signal is followed closely by a stretch of hydrophobic residues. In that case, translocation into the matrix proceeds until the hydrophobic stretch enters the lipid environment of the inner membrane and then stops. This leaves the protein anchored in the inner membrane with the bulk of the molecule residing in the inter-membrane space.

Mitofilin contains both a matrix targeting signal (residues 21-38) and a stop-translocation sequence (residues 46-65) near its amino terminus (Figure 1D). Analysis of the protein sequence by the PSORT program yields a probability of its residing in the intermembrane space of mitochondria of 82%. This result, together with the results of immunofluorescence and protease accessibility experiments described here, indicate that the targeting and stop signals are indeed utilized in the cell to place the protein in the inter-membrane mitochondrial space, anchored in the inner membrane.

The fact that the message encodes a mitochondria-specific protein explains the observation of its abundance in heart muscle, given the high numbers of mitochondria in cardiac myocytes (Michaels et al., 1982; Johnson, et al., 1980; Amchenkova, et al., 1988). Antibodies specific for the protein product of this mRNA, however, show that it is readily detected both by fluorescence microscopy and by immunoblot in every human cell and tissue type examined with the exception of erythrocytes (which lack mitochondria), and have consistently found the same 90 and 91 kDa doublet and specific mitochondrial localization. We conclude that mitofilin is a constituent of all human mitochondria. Antibodies 302.47 and 3584, but not 2-8, also recognize 90 kDa mouse and rat homologues and label mitochondria, implying that the protein is conserved, at least among mammals (not shown). Whether there is wider phylogenetic conservation of mitofilin remains to be investigated.

Our present understanding of the molecular interactions which regulate mitochondrial movement, transmembrane trafficking, and cytoskeletal interactions is far from complete. The precise role of the ubiquitous, conserved mitofilin protein in the life of the mitochondrion awaits further functional and structural analysis.

CHAPTER 3. ANALYSIS OF EXPRESSION OF MITOFILIN AND NUCLEAR MATRIX-INTERMEDIATE FILAMENT SCAFFOLD PROTEINS IN CELLS AND TISSUES

This chapter presents previously unpublished work in which methods were developed for the measurement of solubilization-resistant protein antigens in solid tumor samples. The proteins assayed were three nuclear matrix coiled coil proteins (lamin B, NuMA, and TPR) as well as mitofilin, whose identification and characterization is described in Chapters 1 and 2. Total proteins from the tissues were extracted and solubilized under strong, non-selective, denaturing conditions. The samples were then subjected to SDS-PAGE and analyzed by chemiluminescent Western blot. Prior to the availability of specific antibodies to this type of structural protein, quantitative analysis was limited to relative intensities of spots on silver-stained gels. While the data in Chapter 3 are preliminary, they represent proof of principle that such quantitation can be achieved routinely. They also mark a step forward in the measurement of specific nuclear matrix proteins in human tissue. While it is too early to draw conclusions as to potential clinical utility of the tests described here, the results do show promise.

MATERIALS AND METHODS

Non-selective solubilization of proteins from cells and tissues. The solution used for extracting cellular proteins non-selectively was 8M urea, 2% Nonidet P40, and 2% β-mercaptoethanol (Klose and Zeindl, 1984). Protease inhibitors were added in some experiments immediately prior to protein extraction. This consisted either of phenylmethylsulfonyl fluoride, added to 150 μg/ml final concentration, or of the following mixture of inhibitors at the final concentrations indicated: bestatin, 40 μg/ml; pepstatin, 0.7 μg/ml; phenylmethylsulfonyl fluoride, 50 μg/ml; aprotinin, 2 μg/ml; leupeptin, 2 μg/ml; E64, 1 μg/ml.

Adherent cell cultures were plated in replicate from master dilutions, grown in 100 mm round dishes, and one dish was trypsin-released and counted at the time of the experiment. To the replicate

dishes to be extracted, 0.5 ml of UNM was added per dish, the cells were scraped up into a microfuge tube, and DNA was sheared either by passing repeatedly through progressively smaller gauge hypodermic needles or by a brief (approximately 10-second) pulse with a hand-held homogenizer. Following shearing, samples were centrifuged for 5-10 minutes at 10,000 rcf in a microfuge and the supernatant was harvested.

Suspension cells were washed with PBS and an aliquot was counted. The remainder were pelleted at 1500 rcf for 10 minutes. Normal peripheral blood lymphocytes were purified by standard Ficoll-diatrizoate density centrifugation in the HLA typing laboratory at University of Massachusetts Hospital. Erythrocytes were obtained from the resulting pellet. UNM was added in a ratio of approximately 1 ml per 10^7 cells, and then processed as above.

Samples of female reproductive carcinomas and normal tissues were obtained from the Frozen Section room adjacent to the operating rooms at the University of Massachusetts Hospital. Tissue samples (approximately 1 cm^3) were either flash frozen at the time of dissection and stored at -70°C until extraction, or were processed immediately. Fresh tissue was minced to approximately 1 mm^3 and then solubilized in UNM with a hand-held homogenizer, as above, in about 5 ml per gram of tissue. Frozen tissue was cryo-sectioned in thick (20-40 μm) sections, thawed in UNM, and homogenized as above.

Protein concentrations were measured by the Coomassie Plus protein micro-assay (Pierce Chemical Co. Rockford, Il) standardized with BSA. The tolerance of this assay for UNM was found to be 5 μl in 2 ml total volume, so UNM was added to standards and unknowns to this final amount in all samples. Correlation coefficients obtained by linear regression analysis of standard curves routinely exceeded 0.99. Unknowns were tested at two concentrations in duplicate or triplicate and the results for each sample were averaged. The solubilized protein samples were analyzed immediately by SDS-PAGE and/or were aliquotted and stored at -70°C.

Development of quantitative and semi-quantitative protein immunoblot assays.

SDS-polyacrylamide gel electrophoresis. Disulfide-reducing SDS-PAGE was performed by the method of Laemmli (1970) using 6.5 X 9.5 X 0.08 cm minigels. Protein samples (above) were combined either with an equal volume of 2X sample buffer or 1/5 volume of 6X sample buffer, mixed vigorously by vortexing, placed in a boiling water bath for 10 minutes, and then centrifuged for about 1 minute in a microfuge before loading. To load samples with microliter accuracy and with no spill-over into adjacent lanes, a Hamilton syringe fitted with an HPLC sample injection needle was used. Between samples, the syringe was rinsed several times with SDS running buffer. Tests confirmed that this method produced no carry-over. Where necessary to maintain equal ionic conditions in lanes to minimize band spreading, a mixture of UNM and sample buffer was used to dilute samples so that equal volumes could be loaded in different lanes.

Desktop computer-based scanning densitometry. Dried gels, chemiluminescent immunoblot X-ray films, and NBT/BCIP-developed immunoblots were analyzed on a desktop computer (Macintosh, Apple Computer, Cupertino, CA) equipped with a flatbed scanner (OneScanner, Apple Computer) and with image capture (Ofoto, Apple Computer) and densitometry (ScanAnalysis, Biosoft Corp., Cambridge, UK) software. For maximum detection of blue Coomassie bands or purple NBT/BCIP bands, it was necessary to place a yellow or amber filter over the dried gel or blot. This was either a Cibachrome yellow 0.4 filter (Kodak), or simply a translucent amber plastic sheet obtained from a report cover. Both produced equivalent results.

Scans were done at 300 dpi (the maximum available on the equipment) using Ofoto, and the raw data were saved in TIFF file format with no compression. The Scan Analysis program was then used to select bands on the image and to integrate the signal present. Areas adjacent to the bands were selected and subtracted as background, and the resulting data were expressed as "net pixel area." These numerical values, which correspond to signal intensity, were used in all calculations and semi-quantitative comparisons.

Semi-dry electroblotting. SDS-PAGE gels were electroblotted onto Immobilon P PVDF membrane (Millipore) using the semi-dry method of Khyse-Anderson (1984). This involves building a "sandwich" of the gel, membrane, and Whatman 3MM paper soaked in transfer buffers on a metal plate electrode device (Panther model, Owl Scientific, Woburn MA). The layering was: cathode plate; 3 pieces of paper soaked in cathode buffer (25 mM Tris, 40 mM glycine, pH 9.4); gel; membrane; 2 pieces of paper soaked in anode buffer 2 (25 mM Tris/NaOH, pH 10.4); 1 piece soaked in anode buffer 1 (300 mM Tris/NaOH, pH 10.4); anode plate. Transfer was done with a current of 2.5 mA/cm^2. Time allowed for transfer was determined empirically for different acrylamide concentrations and protein molecular weights by Coomassie staining gels after transfer to verify complete transfer of proteins, as well as by visual inspection of the transfer of pre-stained protein molecular weight standards (Rainbow Markers, Amersham Life Sciences, Arlington Heights, IL). Typically, transfer times of 30-60 minutes were sufficient for complete transfer.

Immuno-development of blots. Blots were processed by the method recommended for the Immun-Lite system of BioRad (Hercules CA), as described in Chapter 1. Secondary antibody was anti-mouse IgG conjugated to alkaline phosphatase (BioRad). Murine monoclonal primary antibodies used in these experiments were clones 204.41, anti-NuMA; 302.47 and 2-8, anti-mitofilin; 101.B7, anti-lamin B; and 203.37, anti-tpr nucleoporin, all from Matritech (Cambridge, MA). Signal development was either by the chromogenic substrates NBT and BCIP or by the chemiluminescent substrate AMPPD (Immun-Lite, BioRad). For chemiluminescent signal detection, blots were exposed to X-ray film for periods ranging from 10 minutes to four hours at room temperature, or overnight at 4°C.

Slot blot measurement of assay sensitivity. Purified mouse IgG (Pierce Chemical) was serially diluted in PBS containing 0.1% BSA and applied to PVDF membranes using a vacuum slot blot device (Gibco). Dilutions of known amounts of IgG were loaded into the wells and were slowly vacuum blotted onto the membrane over three to five minutes. PBS was applied and pulled through to rinse unbound protein from the wells and membrane. The blot was then processed exactly as were electroblotted samples,

using alkaline phosphatase-conjugated anti-mouse IgG to detect the bound immunoglobulin, developing by both chemiluminescent and NBT/BCIP substrates, and analyzing densitometrically as described above.

Reproducibility of semi-quantitative Western blot. Four samples of ovarian carcinoma tissue were solubilized in UNM as described, aliquotted, and frozen at -70°C. In three separate experiments, the extracts were subjected to SDS-PAGE, electroblotted, probed with antibody to lamin B, and developed by chemiluminescence. In two experiments, a single extract was loaded in all 10 gel lanes and probed with anti-lamin B and anti-NuMA antibodies.

Coomassie-stained gels for quantitation of recombinant mitofilin. Aliquots of recombinant mitofilin (rMF), purified to homogeneity, were prepared as described in Chapter 1. In two separate experiments, SDS-PAGE was performed with different doses of both rMF and purified bovine serum albumin (BSA; Sigma # A-7030) loaded on the same gel. The BSA, used as the gravimetric standard for quantitation, was weighed on an analytical balance and the final volume of solution was adjusted at room temperature in a volumetric flask. The gels were fixed in 7% acetic acid, 40% methanol for 20 minutes, and then stained with colloidal Coomassie (Sigma). Following a twenty-second rinse in 10% acetic acid, 25% methanol, de-staining was done in 25% methanol until the background was clear, and the gels were vacuum dried on 3MM paper (Whatman). The dried gels were analyzed by scanning densitometry as described above. A standard curve of protein mass versus net pixel area was constructed for the BSA, and the concentration of rMF was calculated as the average of those values which fell within the range of the BSA standard curve. The resulting value for the concentration of rMF in the frozen aliquots was used to calculate the amount of mitofilin in Western blots of unknown samples.

Quantitative analysis of mitofilin expression in normal and tumor cells. Using data obtained from Coomassie-stained gels of rMF versus BSA (above), dilutions of rMF were loaded to provide a standard curve. Typically, these were 5, 3, 2, 1, and 0.5 ng of protein. The remaining lanes were loaded with duplicates of different dilutions of cell extracts representing known cell numbers. This was usually in the range of 20,000 to 150,000 cell equivalents per lane.

RESULTS

Absolute sensitivity of chemiluminescent detection of proteins. In order to test the theoretical limits of detection of protein antigens by chemiluminescent Western blotting, experiments were performedin which purified mouse IgG was applied to PVDF membranes in a slot blot format. After probing the blot with anti-mouse IgG-alkaline phosphatase conjugate, chemiluminescent signal was recorded on X-ray film and densitometric analysis was performed.

The results shown in Figure 13 indicate an extremely wide dynamic range covering five logs of detection, with even 10 picograms of IgG being detected. Longer exposures raised the 10 picogram signal to higher levels, but moved the high end of the curve outside the dynamic range of the film (not shown). It is also noteworthy that the response is far from linear. For example, doses differing by a factor of 10 produced signal differing by a factor of about 1.5 when comparing 1 μg versus 10 μg, while comparison of 1 μg with 0.1 μg showed a difference in signal of about seven-fold. The large deviation from linearity of dose/response parameters required designing assays for unknown samples which circumvent this limitation. This was accomplished by inclusion of dilutions of an external standard to which unknowns could be compared.

Reproducibility of assay values. Inter-assay and intra-assay variability were measured using antibodies to lamin B and NuMA. Intra-assay variation was assessed by loading a 10-lane gel with equal amounts of a single ovarian carcinoma protein extract and probing with a combination of anti-lamin B and anti-NuMA. One of these blots is shown in Figure 14. Densitometric analysis of the bands gave the following results (net pixel area):

Lamin B bands:

Mean = 2651; sd = ±375; cv = 14.1%; n = 10, range = 1993-3097

NuMA bands:

Mean = 2641; sd = ±572; cv = 21.7%; n = 10, range = 1653-3281

Ratios were calculated of NuMA:lamin B signal in each lane, as would be done in analysis of different unknown tissue samples. The results were as follows:

Mean ratio = 0.992; sd = ±0.147, cv = 14.8%, n = 10, range = 0.852-1.22

Figure 13. Plot of densitometric analysis of chemiluminescent detection of mouse IgG. Wells of a slot blot device were loaded with serial dilutions of purified mouse IgG (Pierce), probed with anti-mouse IgG conjugated to alkaline phosphatase, and supplied with chemiluminescent substrate. The signal was recorded on X-ray film and analyzed as described in the text. In this one-hour exposure, the dynamic range of IgG detection spanned five logs, from 10 pg to 1 μg. Response, however, was far from linear. For example, while the 10-fold difference in dose between 0.1 and 1 μg gave approximately 7 times the signal, 1 μg and 10 μg were only about 1.5-fold apart in signal. Note also that as little as 10 pg of IgG was detected. Longer exposures raised this signal to higher levels, but produced a leveling off at the upper end of the curve (not shown).

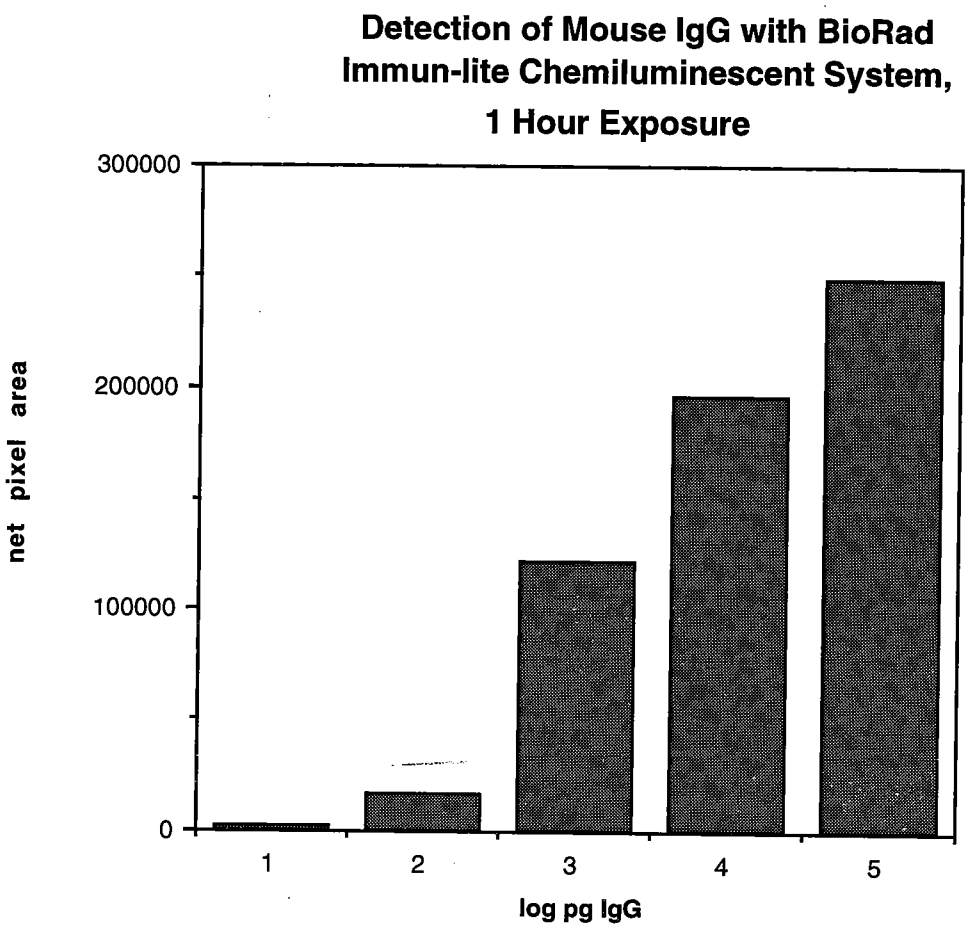
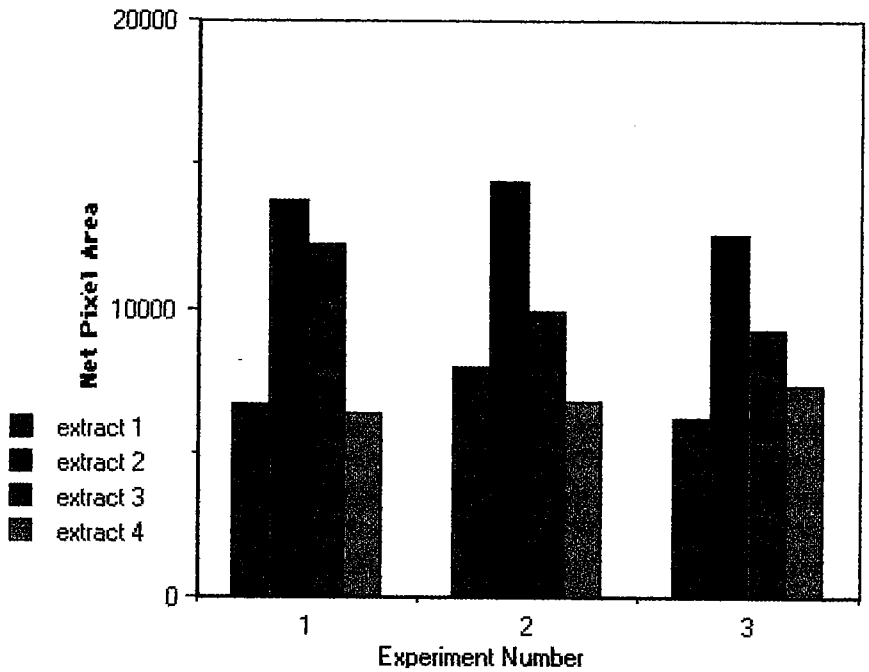

Detection of Mouse IgG with BioRad
Immun-lite Chemiluminescent System,
1 Hour Exposure
300000
200000
100000
0
net pixel area
1
2
3
4
5
log pg IgG

Figure 14. Intra-assay variability in chemiluminescent Western blots. All 10 lanes of the gel were loaded with equal volumes of a single ovarian carcinoma protein extract. The blot was probed with pooled monoclonal antibodies to lamin B and NuMA and secondary anti-mouse IgG alkaline phosphatase conjugate and developed by chemiluminescence. Results of densitometric analysis of the resulting bands are described in the text. The upper bands are NuMA, the lower are lamin B. Note that primary antibodies can be combined in one blot if molecular weight differences between the two proteins permit sufficient resolution on the gel.

Intra-assay variability of Western blot: Lamin B and NuMA in one ovarian carcinoma extract

Figure 15. Inter-assay variability of semi-quantitative Western blot for nuclear matrix proteins in tissue samples. SDS-PAGE gels of proteins extracted from four ovarian carcinoma tissues were blotted, probed with monoclonal antibody to lamin B, and developed by chemiluminescence. The plots show net pixel areas obtained in three experiments in which the same four extracts were tested. The magnitude of signals varied, in one case by about 20%. Note, however, that the order of peak heights remained constant.



Inter-Assay Variation of Semi-Quantitative Western Blots

The experiment was repeated, with essentially identical results (not shown).

To assess the degree to which day-to-day variations could affect results, a series of three experiments was carried out using aliquots of four extracts from ovarian carcinoma tissues. 10 μg of each extract was loaded per lane and the gels were blotted and probed with anti-lamin B antibody. The results are shown in Figure 15. While the signal for a given extract varied from gel to gel by a maximum of 18%, more importantly, for the purpose of relative quantitation of signals, the order of peak heights was the same in all three experiments.

Together, these evaluations of assay performance were interpreted as follows:

1: quantitative and semi-quantitative Western blots provide an extremely sensitive means of measuring proteins which cannot be tested by standard immunometric assays, while simultaneously providing molecular weight information on the analyte.

2: The relative amounts of proteins detected in different samples remains acceptably constant, despite day-to-day variations in signal magnitude.

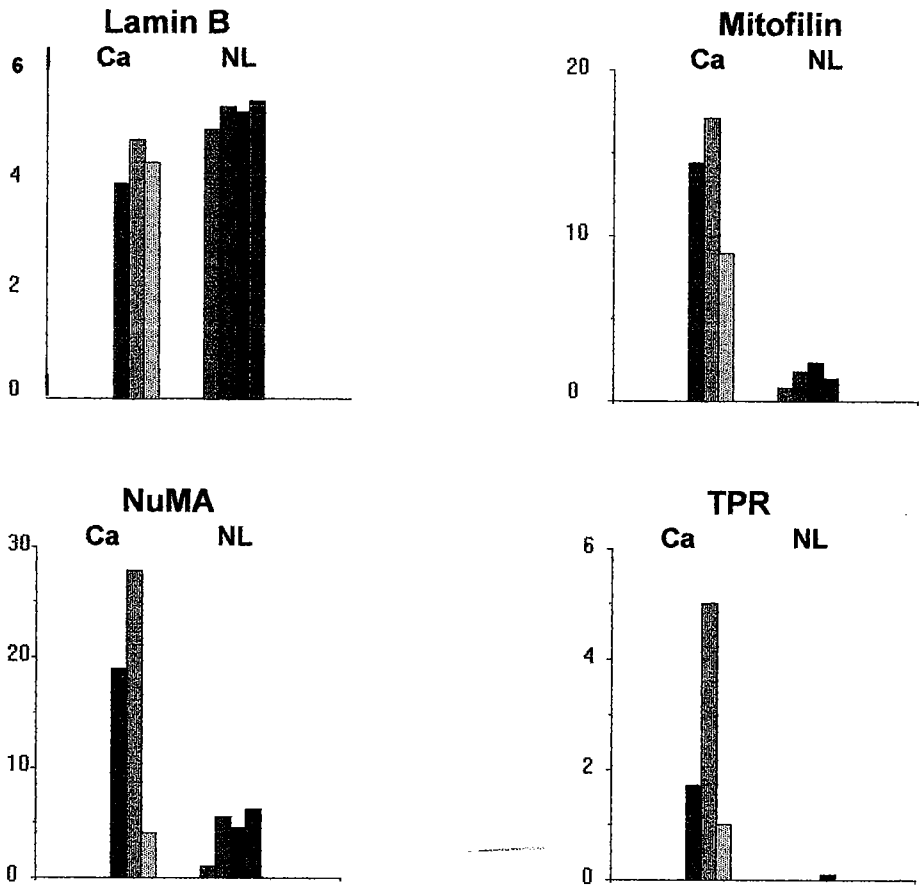
3: The intra-assay variability requires that differences between signals which are less than about 40% of their magnitudes be considered insignificant, unless replicates are tested.

4. The ratio of lamin B signal to that of other antigens can be an effective means of determining the relative cellularity of tissue samples and therefore the expression of other proteins on a per-cell basis.

5. Combining two primary antibodies which recognize different proteins provides useful data with half the number of gels, provided that the molecular weight differences are sufficiently resolved.

Semi-quantitative analysis of protein expression in ovarian carcinoma tissue samples: preliminary data. In a preliminary study, Western blots were performed on extracts of tissue from 3 ovarian carcinomas and 4 normal ovaries. All seven extracts were loaded on four gels and blotted. The blots were probed with antibodies to lamin B (clone 101.B7), NuMA (clone 204.41), mitofilin (302.47), and tpr (clone 203.37). The results shown in Figure 16 indicate that, while the lamin B signal was very similar in the normal and cancer samples, the other three proteins showed large differences. In cancer 3,

Figure 16. Western blots were performed comparing the total extracted protein from 3 ovarian tumors (left of each plot) and 4 normal ovaries (right). All seven samples were loaded in each of four gels. Western blots were probed with antibodies to lamin B (101.B7), mitofilin (302.47), NuMA (antibody 204.41) and nucleoporin tpr (203.37). Antibody reactivity was detected by chemiluminescence as described. Bar heights represent the net pixel area as determined by densitometry. Note that while lamin B signal was highly similar in all cases, the cancer tissues all had elevated levels of the other three proteins, excepting only NuMA in one case, which appeared to have normal levels.



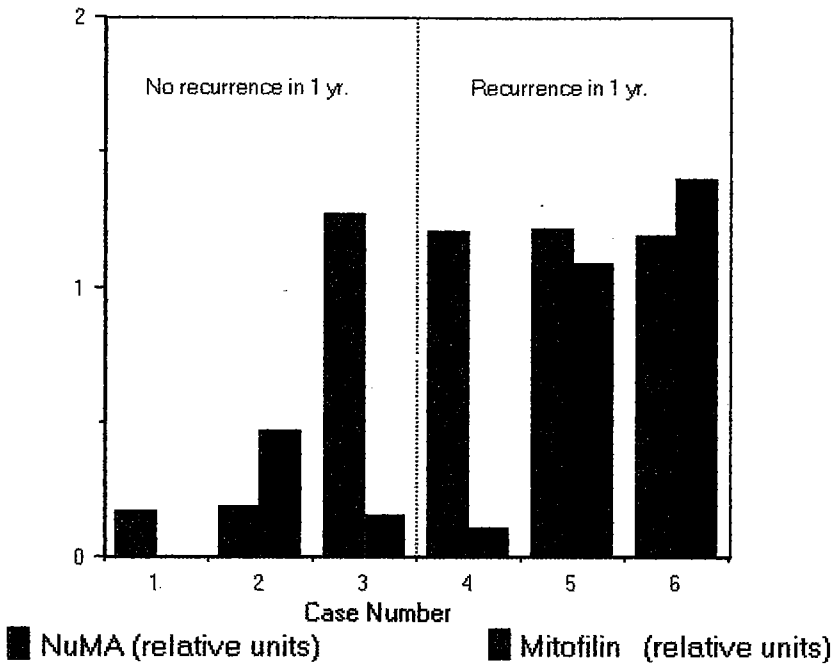
Semi-Quantitative Western Blots on Tissue Extracts from 3 Ovarian Cancers and 4 Normal Ovaries

only the NuMA signal was in the normal range. Cancers 1 and 2 were about five-fold higher for NuMA, and all three cancers had very elevated levels of mitofilin (about 10-fold) and tpr (about 200-fold). Thus, while the number ofcells analyzed in the samples was roughly equal, based on the amount of lamin B present, the other proteins suggest that cancer-specific over-expression may be taking place in the malignant cells.

In a separate preliminary study, semi-quantitative comparisons were done of blots probed with paired antibodies: lamin B + mitofilin, or lamin B + NuMA. In this way, ratios of signal could be obtained from the same blot. Six tissues obtained from the first surgery of ovarian cancer patients were analyzed in a blind study. The outcome of the cases was known to the surgeon but was not revealed until after the tests were run. Figure 17 shows plots of signal ratios for mitofilin:lamin B and NuMA:lamin B. The cases are categorized as to outcome, defined for ovarian cancers as recurrence within a year. The data are clearly preliminary, but they do indicate that with a suitable panel of antibodies, this method could prove useful in predicting responsiveness to treatment (see Discussion).

Determination of purified, recombinant mitofilin in preparative gel fractions. In order to determine the concentration of proteins in cell and tissue samples not just in relative terms, but in molecules per cell, it is necessary to have a known standard. Recombinant mitofilin, purified to homogeneity as described in Chapter 1, was used for this purpose. First, however, it was necessary to measure its concentration in the pooled fractions of purified protein. While there are many ways of measuring total protein concentration in solutions, each has its limitations (discussed in detail by Darbre, 1986). For example, OD_{280} measurements vary with the number and local environments of aromatic side chains of the amino acids. The nitrogen method gives reliable results, but requires milligram quantities of protein. The widely used dye-binding assays, for example, Coomassie blue, are sensitive to the choice of standard, since different protein molecules show different dose-response relationships. Also, they are affected by certain detergents or salts, especially in dilute protein solutions typical of those under investigation here. Mitofilin and many other proteins of interest to our laboratory require high salt,

Figure 17. A blind study of semi-quantitative Western blot comparisons of 6 ovarian carcinoma tissue samples for expression of mitofilin and NuMA. Samples were obtained from tissue removed upon first surgery. All six samples were loaded into 2 gels and blotted. The blots were probed with combined antibodies to lamin B and either mitofilin or NuMA. Bar heights correspond to the ratio of mitofilin or NuMA signal to that for lamin B in the same lane. Thus, they represent the relative amounts of protein on a per-cell basis. Clinical information is as follows: All cases except number 2 were papillary serous cystadenocarcinoma. Case 2 was mucinous. Case 1 is FIGO stage IIIc, grade 2 with no recurrence 19 months post 1st surgery. Case 2 is stage I grade 1 with no recurrence 10 months after completion of chemotherapy. Case 3 is stage II, grade 2 with no recurrence 16 months post first surgery. Case 4 is stage IIIb with positive second look surgery. Case 5 is stage IV, grade 3, with positive second look surgery. Case 6 is stage IIc, grade 2. This patient refused second look surgery, but according to the gynecologic oncologist, had other poor indicators. Inspection of the graph reveals 3 groups of patients: those with low levels of both mitofilin and NuMA, those with elevated levels of both proteins, and an intermediate category with only mitofilin elevated. The two patients with low levels responded to therapy. Of the two intermediate cases, 1 responded and 1 did not. For the two with elevated levels of both proteins, one did not respond to treatment and one refused second look surgery and so cannot be evaluated for outcome by the same criteria.



Semi-Quantitative Western Blots of 6 Ovarian Cancer Tissues

detergent, and sometimes strong denaturants to extract them from cells. In particular, recombinant mitofilin protein was purified by preparative SDS-PAGE, and so was in gel buffer containing SDS. This required alternative means of determining its concentration.

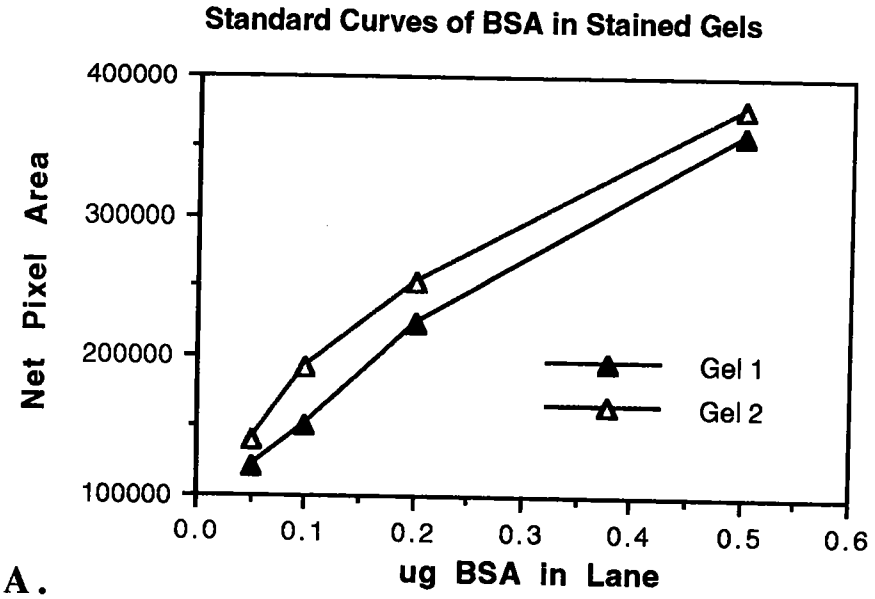
By combining SDS-PAGE with Coomassie protein staining and scanning densitometry, these problems were overcome. Purified bovine serum albumin was found suitable as the quantitative standard for these analyses since dose-response curves for it and for mitofilin were the same. Gels were run which were loaded with dilutions of both BSA and rMF. The purified BSA, Sigma type A-7030, gives a single band in SDS gels. A standard solution containing 10 mg per ml was made, and dilutions were prepared so that known weights could be loaded in gel lanes. Dilutions of rMF were also loaded on the gels, and after electrophoresis, the gels were fixed and stained with Coomassie dye, dried, and scanned. The net pixel area data were used to construct standard curves of signal per weight of BSA. The amount of rMF was then determined by comparison of several dilutions to the standard curve in two separate gels. Figure 18A shows the resulting standard curves, plotted as pixel area versus amount of BSA. The six values for the mitofilin dilutions which fell on the curve (one was too low on both gels) were used to calculate the amount of mitofilin in the undiluted sample. These were averaged, and the molecular weight of 71.5 kDa for the recombinant molecule was used for conversion to moles. The mean value for the undiluted mitofilin sample was 9.7 μg/ml (coefficient of variation = 8%, n = 6), or 136 nM, or 8.17×10^{13} molecules per ml:

$$\frac{(9.7\ \mu g/ml) \times 6.02 \times 10^{23}\ molecules/mole}{7.15 \times 10^{10}\ \mu g/mole} = 8.17 \times 10^{13}\ mol./ml$$

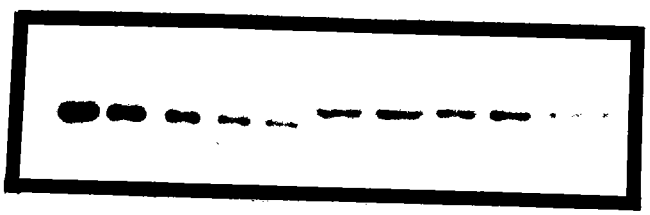
This value for mitofilin concentration in the pool was used for subsequent calculations in quantitative mitofilin immunoblot assays of cell extracts.

Determination of number of molecules of mitofilin expressed in cultured cells. Total protein extracts of a number of cell lines were prepared in UNM, as described in Methods. Gels were loaded with

Figure 18. A. BSA standard curves for quantitation of recombinant mitofilin. Two SDS-PAGE gels were loaded with dilutions of BSA and of purified recombinant mitofilin. Scanning densitometry was performed on the Coomassie-stained gels, and the resulting data for BSA were plotted as shown. The pixel values for mitofilin which fell within the range of the curve were used to calculate the concentration of mitofilin in the undiluted sample. **B.** Quantitative immunoblot of mitofilin in a cell line extract. The lanes at left were loaded with different amounts of rMF to generate a standard curve for densitometry. The right-hand six lanes were loaded in duplicate with different amounts of a total protein extract from ovarian carcinoma cell line CA OV3. The volumes loaded represented the number of cell equivalents shown below the lanes. Blots of this type were used to obtain the data presented in Table 1 on mitofilin expression in different cell types.



A.



5 3 2 1 0.5 5.3 2.7 1.3
rMitofilin std. CaOv3 Proteins
(ng/lane) (X 10^{-5} cells)

B. Quantitative Western blot of mitofilin in an ovarian carcinoma cell line.

duplicates of three dilutions of the extract, along with 5 lanes containing different amounts of rMF standard. A useful range of the curve was found empirically to be between 0.5 and 5 ng per lane. Gels were blotted and probed with anti-mitofilin antibodies and developed by chemiluminescence. The resulting films were scanned, and standard curves constructed exactly as for the Coomassie-stained BSA. Figure **18B** shows a typical resulting blot. In this way, the expression of mitofilin protein in a variety of human cells and cell lines was measured.

Mitofilin is expressed ubiquitously, but differentially, in human cells. The following table lists data derived from a series of experiments in which blots such as the one shown in Figure **18B** were analyzed by densitometry. Each value for molecules per cell is based on measurements of at least two concentrations of a culture cell extract measured in duplicate. Some measurements were done in more than one experiment, and the results for that cell type were averaged.

TABLE 1

CELL TYPE	TISSUE OF ORIGIN	AVG. MOL/CELL
ME-180	Cervical Carcinoma	150,000
CA OV3	Ovarian Carcinoma	42,000
Lymphocyte	Normal Peripheral Blood	20,000
Erythrocytes	Normal Peripheral Blood	None Detected
A-301	Transformed Lymphocyte	103,000
W1-38	Embryonic Lung Fibroblast	38,000
MG-63	Osteosarcoma	48,000

These results show there is wide variation of expression of mitofilin protein in different cells. At the time of testing, it had not yet been determined that mitofilin was mitochondrial protein; therefore, no comparisons were done involving cells with the highest levels of mitochondria, such as cardiac myocytes. Nevertheless, these results demonstrate the utility of the method for quantitative analysis of solubilization-resistant proteins in cell extracts. They also demonstrate that mitofilin appears to be expressed as a component of the mitochondria in all human cell types, with its levels perhaps reflecting the metabolic activity of cells.

DISCUSSION

Quantitative and semi-quantitative Western blots have potential applications both in basic research in cell biology and in clinical studies. For example, it has been shown by many workers that the expression of nuclear matrix proteins varies with cell type of origin, malignant transformation, and hormone-induced differentiation (Fey and Penman, 1988; Stuurman et al., 1989; Getzenberg and Coffey, 1990). These assessments were based primarily upon silver-stained or autoradiographic two-dimensional protein gels. Thus, proteins were characterized by their M_r and pI, but without knowledge of the identities of many of the low-abundance constituents of the protein mixtures. Having antibodies to specific components allows the next step to be taken, that is, quantitation of expression.

One difficulty in performing this type of analysis in tissue samples is the assessment of cellularity, and therefore of having a means of relating protein expression on a per-cell basis. In tumors, the mass of protein per cell can vary widely, depending upon the characteristics of the malignancy. Cells of mucinous tumors, for example, secrete large amounts of glycoproteins. Ploidy is also widely variable, making DNA determinations less useful than they are in normal tissues for the purpose of measuring cell number. In these investigations, comparative cellularity was determined by densitometric measurement of lamin B in Western blots. Lamin B was selected for several reasons discussed in the Introduction, including its ubiquitous expression in human cells. In addition, its expression per cell has not been shown

to vary; moreover, as both an intermediate filament protein family member and a nuclear matrix protein, it serves as a prototype for the class of proteins being investigated.

The preliminary data shown in Figures 16 and 17 demonstrate that there are measurable differences in the expression of known nuclear matrix proteins in patient-derived tissue samples. In Figure 16, for example, the lamin B signal of three ovarian cancer samples and four normal ovaries were nearly equal, and therefore the loadings represent roughly equal cell numbers. In contrast to lamin B, the nucleoporin tpr signal, measured by monoclonal antibody 203.37 (Bangs et al., 1995), was about two hundred-fold greater in all three cancers than in the normals, where it was barely detectable. There were also readily detected, though less dramatic, differences in mitofilin and NuMA expression. NuMA signal was much higher in two of three cancers (roughly 5-fold greater), and mitofilin was roughly 10-fold greater than normal in all three cancers.

In addition, the cases examined in Figure 17 suggest a possible relationship between the relative expression of NuMA and mitofilin with clinical responsiveness of ovarian cancers. Responsiveness was defined as no recurrence within a year upon second-look surgery, a criterion evaluated by the surgeons (Drs. Beth Nelson and Peter Rose) and pathologists (Dr. James Pullman and others) at U. Mass. Medical Center Hospital.

These early results will require much further evaluation to establish their diagnostic or prognostic value. The technical advances which continue to be made in both the chemistry of chemiluminescent substrates and in the capture of the emitted light means that even greater sensitivity can be expected in future versions of these tests. It does appear, however, that specific nuclear matrix proteins, and perhaps mitochondrial proteins, may prove useful as markers of proliferative potential in human epithelial-derived tissues. Basic investigations of the mechanisms relating proliferative capacity with nuclear matrix protein expression may provide new insights into carcinogenic mechanisms as well as potential points for clinical intervention.

CHAPTER 4. PHYLOGENETIC OCCURRENCE OF COILED COIL PROTEINS: IMPLICATIONS FOR TISSUE STRUCTURE IN METAZOA VIA A COILED COIL TISSUE MATRIX

As described in Chapter 1, reports from a variety of research areas led to the hypothesis that a continuous matrix of coiled coil protein interactions underlies tissue structure in Metazoa. In order to begin testing this hypothesis, all the phylogenetic protein sequence data in GenBank were analyzed for the occurrence and lengths of coiled coil domains. This represents the first time such an analysis has been undertaken, and it was made possible by the translation and customization of a computer program which identifies these domains. Modifications to the program outputs allowed comparisons to be made among the phylogenetic sequence files in terms of protein length, coiled coil domain length, and rates of occurrence. Careful analysis of the outputs made it possible to assign many of the long coiled coil proteins to just a few structural and functional categories. Several new observations arose from this analysis. First, short coiled coil domains occur with surprising constancy across all phyla. Second, the occurrence of extended coiled coil proteins is much greater in the animal kingdom than among other types of organisms. Finally, a large percentage of the extended coiled coil proteins found in Metazoa are components of the NMIF, and their coiled coil domains are usually highly interrupted, suggesting complex interactions. These findings are discussed in light of the coiled coil tissue matrix hypothesis.

MATERIALS AND METHODS:

Cell Extraction and Imaging. Human ME-180 cervical epithelial cells were grown in plastic dishes to about 70% confluence in DMEM with 10% fetal bovine serum. Digital images were obtained with a Nikon inverted microscope equipped with a 40X objective and a CCD camera. The live cells were imaged immediately after removal from the tissue culture incubator, after which they were subjected to sequential extraction on the microscope stage according to the method of Fey et al. (1984a), except that RNase was omitted from the digestion step (Fey et al., 1986). Briefly, membrane phospholipids, soluble

materials, microtubules, and microtubule-associated proteins were removed by three two-minute extractions with ice cold CSK buffer which contains 0.5% Triton X-100. DNA was digested with DNase I at room temperature for 20 minutes, and then chromatin, actin, and other salt-extractable material was removed with two changes of 250 mM ammonium sulfate in ice cold digestion buffer. The buffer was removed and replaced by cold CSK, and digital images of the unfixed NMIF were recorded. Other than initial exposure adjustments, no digital image reprocessing was done, and the photomicrographs in Figure 19 are reproductions of the original digital printouts.

PC-DOS-compatible version of coiled coil prediction program of Lupas et al. (1991). The Coils 2.1 programs, generously provided by A. Lupas (1991), were translated from Pascal for a Vax computer to C++ to run in a DOS environment on a personal computer. These programs scan a protein sequence with a sliding "window" to identify α–helical coiled coil regions. A weighted matrix of values for amino acids, derived from their utilization in the heptad repeats of myosins, tropomyosins and keratins is used to assign to each amino acid a numerical score, a derived probability (P) for coiled coil formation, and a letter (*a* through *g*) denoting its position within a heptad repeat. Two versions were compiled, one to run in batch mode for large-scale sequence database screening, and one designed for analysis of individual proteins. Counting functions were added which total the number of sequences searched, the number of coiled coil proteins found, the length of the proteins, and the total length of their coiled coil domains. Test runs verified that the score, probability, and heptad position outputs matched exactly those of the original programs. Both PC programs accept GenBank ASCII files, FASTA files, or text files as input.

The batch program searches a data file for sequences which meet or exceed a user-defined cutoff value for P, which was set at 0.9 to limit the output to the most unambiguous heptad repeats. Access to the sequence files was via GenBank flat file CD-ROM, release 84.0 (NCBI, Bethesda, MD). Two output files are generated, one containing FASTA versions of the "hit" sequences, and the other (the "report"), containing data on the search results. For each file analyzed, the report shows the number of protein sequences analyzed and the number of hits. The report then lists data for each hit: 1) the GenBank

definition line for unambiguous identification; 2) the number of amino acids in the sequence; 3) the number of amino acids for which $P \geq 0.9$; 3) the percent of residues lying in regions of $P \geq 0.9$ ("% coiled"); 4) the maximum value of P; and 5) a simple graph showing the position and extent of the coiled coil domains. Each character in the graphs represents about 28 amino acids. Vertical characters indicate $P \geq 0.9$, and horizontal characters designate regions where $P < 0.9$ (see Figure 20). The report is formatted as tab-delimited text to facilitate further analysis (see below).

The second PC-compatible program reports detailed numerical results of the analysis, with scores, probabilities and heptad positions for each residue. It also produces on-screen a graph of coiled coil probability for a visual inspection of the results. Typically, coiled coil proteins in the database were identified by screening with the batch program and then proteins of interest were analyzed in detail using the second program.

Analysis of Report Files: The report files generated by the batch program were analyzed using a standard spreadsheet program (Microsoft Excel 5.0). The proteins were sorted hierarchically by coiled coil length, protein length, and "% coiled" (an excerpt from the sorted primate report is shown in Figure 20). A second report was then generated for each GenBank file for proteins whose heptad repeat regions exceeded 75 amino acids, which here are designated "extended coiled coil proteins." This threshold excludes coiled coils of relatively simple dimerization domains and allowed proteins involved in more extended coiled coil interactions to be analyzed separately. Summary statistics were obtained for both reports (see Table 2). Statistical tests for significance of differences are not reported, since they are invalid for these data sets (see Results).

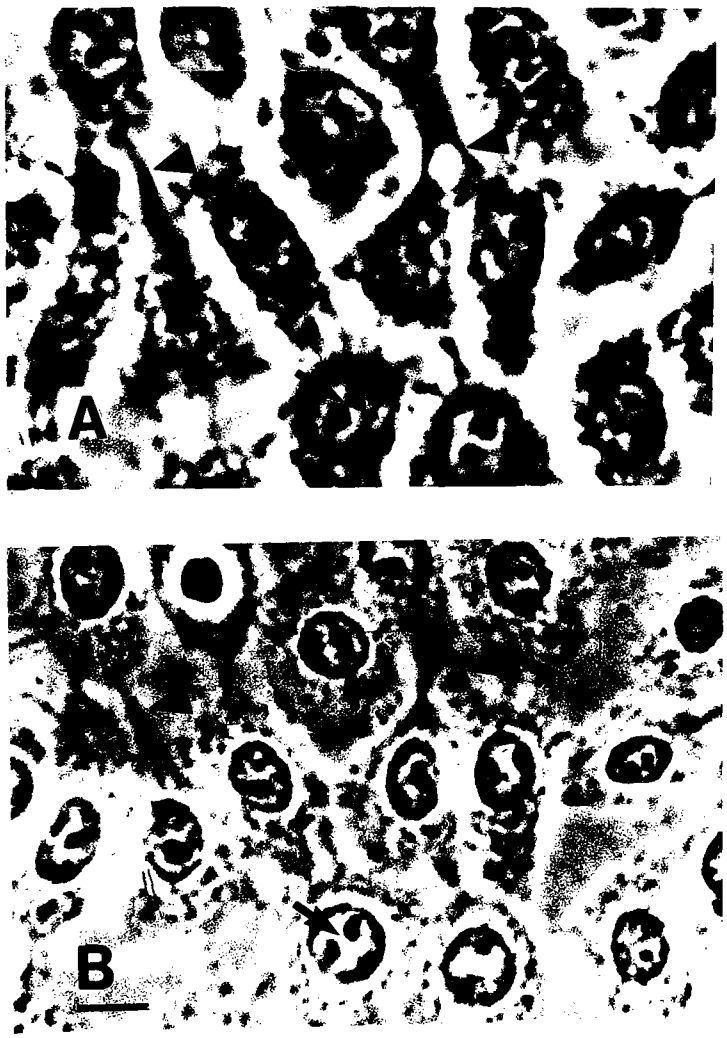
Structure/function classification of extended coiled coil proteins. GenBank submission data and references were scrutinized to subdivide further the extended coiled coil proteins in the bacterial, plant, primate, and invertebrate files into several structure/function categories. One category was created for bacterial proteins which occur on the outer surface of pathogenic bacteria and which are known to bind to host proteins and cell surfaces. The data for proteins from yeasts *S. cerevisiae* and *S. pombe*, which are

in the GenBank plant file, were also analyzed separately. The extended coiled coil proteins in the primate and invertebrate reports were assigned to one of two classifications if their identities in the GenBank submissions were unambiguous. The first category consisted of proteins involved in force production and motility, including all myosins, tropomyosins, troponins, paramyosins, and the microtubule-associated motor proteins kinesin and dynein. The second category consisted of proteins which have been shown experimentally to be selectively retained in NMIF preparations, either by Western blot, immunomicroscopy, or both. Several of the NMIF assignments are based on evidence from our laboratory (tpr, laminin, spectrin, and involucrin, data not shown). Other NMIF proteins included all intermediate filament proteins, the lamins, pericentrin, NuMA, and intermediate filament-associated proteins such as plectin, desmoplakin, and tricohyalin. The remaining proteins were designated as "other" if they have not been shown to belong to one of the two categories or if they were not sufficiently well defined in the GenBank submission references (for example, definitions such as "intermediate filament-like protein" were not considered sufficient for categorization as an NMIF protein). Specific proteins discussed in the present work are identified either by a literature reference or by GenBank accession numbers. FASTA files of individual proteins of interest were analyzed by the Coils program using a 28 amino acid window, and the position and probability score of each amino acid were plotted to visualize their coiled coil domain organization.

RESULTS

Retention of Epithelial Morphology by the NMIF. Preparation of NMIF *in situ* results in a selective enrichment in cytokeratins, lamins, desmosomal components, and other proteins with extended, interrupted coiled coil domains. Figure 19 demonstrates, at the level of phase-contrast microscopy, the morphological fidelity retained by the NMIF. Other studies have examined the ultrastructural details of the NMIF, the selective enrichment of intermediate filament and other coiled coil proteins, retention of the immunomicroscopic localization of specific NMIF proteins, and the removal of cell constituents such as phospholipids, DNA, and the actin and tubulin cytoskeletal filament systems (see for example Fey et

Figure 19. Structural fidelity of nuclear matrix-intermediate filament scaffold *in situ*, which is selectively enriched in certain coiled coil proteins. **A** and **B** are phase-contrast micrographs of the same field of ME-180 cervical epithelial cells. In **A**, the living cells are in tissue culture medium immediately after placement on the microscope stage. In **B**, the unfixed NMIF is shown after sequential detergent, DNase, and salt extraction (12,13). This protocol removes membrane systems and soluble elements, the actin and tubulin-based cytoskeleton systems, and chromatin. It represents roughly 5% of cell protein and 1-2% of starting cell mass. Note the faithful retention of cell size and shape and nuclear position within the cells, and that nucleoli (arrows) and cell-cell connections (arrowheads) are also preserved. Bar in **B** = 8 μm.


A
B

al., 1984a-c). Here it is shown for the first time that the overall organization of the NMIF is faithful to that seen in the same field of living cells. Many morphological features of the live epithelial layer seen in Figure **19A** are clearly discernible, after fractionation, in **19B**. Note the overall size, shape, and relative positions of the cells, their cytoplasmic connections to adjacent cells, the position of nuclei within the cells, and the retention of most of the visible nucleoli after the DNA has been removed. The fact that no fixation has been done underlines the stability, interconnectedness, and free-standing nature of this structure.

Frequency of Coiled Coils Among All Proteins in GenBank Files. The retention both of morphology and of many extended coiled coil proteins in the NMIF suggested a relationship between this group of proteins and variations in phenotypic morphogenesis. The algorithm described by Lupas et al. (1991) was successfully applied to the screening of gene sequence databases to identify coiled coil proteins; therefore, after translating to run in a PC-DOS environment and adding some new functions (see Methods), the same program was used to screen the nine phylogenetic sequence data files in GenBank release 84.0: primate, rodent, mammal, vertebrate, invertebrate, plant, bacterial, viral, and phage. Together these contained a total of 102,007 protein coding sequences. Each sequence appears in only one file, *e.g.*, proteins included in the primate and rodent sequence files do not re-appear in the mammal file, nor do mammalian sequences re-appear in the vertebrate file. An excerpt of the report for the primate file, after sorting by lengths of the coiled coil domains and of the proteins, is shown in Figure **20**. The reports contain numerical data on the search results (see Methods) which permitted comparisons of the files, and the simple graphs reveal the overall organization of coiled coil domains within each protein. In order to determine the prevalence of proteins with extended regions of coiled coil interaction, the data for proteins whose coiled coil domains totaled greater than 75 amino acids were analyzed separately. This cut-off excludes typical dimerization regions such as those of "leucine-zipper"-containing transcription factors. Table **2** summarizes the data obtained from these reports.

Figure 20. An excerpt from the Coils batch program report on the GenBank primate sequence file to illustrate the search output. The proteins have been sorted both by coiled coil domain length and by protein length, and the ten shown here contain the longest coiled coil domains in the file. The text at the top contains a summary of the search parameters, the number of proteins analyzed, the number of coiled coil-containing proteins found, and the fraction of all proteins which contain coiled coil domains. Each sequence is identified by its GenBank definition line. This is followed by a series of numerical values: 1) the amino acid length of the protein; 2) the number of residues which lie in the coiled coil domains above the 0.9 probability; 3) the percent of coiled coil residues, *i.e.*, 100 X (value 2/value 1); and 4) the maximum probability score in the protein. Following these data is a simple graphic which maps the coiled coil domains within the protein. Each character represents roughly 28 amino acids, with the vertical characters representing probabilities above the 0.9 cutoff and the horizontal representing those below. This format permitted rapid visual assessment of search results and protein organization and utilized minimal computer memory. Note the continuous coiled coil domains in the myosins, in contrast to those of the other proteins in this excerpt. NuMA and tpr are part of the NMIF, and they both contain extended, interrupted coiled coil forming regions. Whether giantin, p162, and CENP-E are NMIF proteins remains to be determined experimentally.

COILS version 1/14/95 using MTK matrix., no weights.
Probabilities >= 0.90 Input file is e:gbpri.seq.
Total proteins: 14115 Total coiled proteins: 809 Fraction: 0.0573

H.sapiens giantin mRNA. (X75304) 3259 2064 63.3 1

H.sapiens CENP-E mRNA. (Z15005) 2663 1319 49.5 1

H.sapiens mRNA for NuMA protein. (Z11583) 2115 1305 61.7 1

H.sapiens mRNA for NuMA protein. (Z11584) 2101 1228 58.4 1

H.sapiens tpr mRNA. (X66397) 2094 1202 57.4 1

H.sapiens p162 mRNA. (X78998) 1411 1162 82.4 1

Human nonmuscle myosin heavy chain (NMHC) mRNA, 3' end. (M31013) 1247 1059 84.9 1

Homo sapiens beta-myosin heavy chain (MYH7) gene, complete cds. (M57965 M30603 M30604 M30605 M57747 M57748 M57749) 1935 1056 54.6 1

Homo sapiens beta-myosin heavy chain (MYH7) mRNA, complete cds. (M58018) 1935 1056 54.6 1

Human beta-MHC mRNA for beta myosin heavy chain clone gtMHC-V. (X51591) 1151 1054 91.6 1

TABLE 2. Phylogenetic Occurrence and Domain Lengths of Coiled Coil Proteins

File	# of Proteins	# of Coiled Coil Proteins	% of Proteins Containing Coiled Coils	Mean length of Coiled Coil Proteins	Mean length of Coiled Coil Domains
Primate	14,115	809	5.73	757	125
>75 aa Coiled		281	1.99	837	283
Rodent	13,702	785	5.73	549	99
>75 aa Coiled		259	1.89	652	220
Mammal	3830	199	5.20	687	90
>75 aa Coiled		41	1.07	570	281
Vertebrate	4866	331	6.80	509	135
>75 aa Coiled		149	3.06	549	250
Invertebrate	8532	808	9.47	689	115
>75 aa Coiled		248	2.91	986	287
Plant	14,982	929	6.20	575	59
>75 aa Coiled		110	0.73	1076	216
Bacteria	22,388	1008	4.96	431	60
>75 aa Coiled		166	0.74	522	164
Phage	1859	94	5.06	334	39
>75 aa Coiled		6	0.32	518	111
Virus	17,733	1111	5.68	593	40
>75 aa Coiled		29	0.16	1172	161
TOTAL	**102,007**	6074			
>75 aa Coiled		1289			
Mean			**5.95%**	**581**	**81.8**
>75 aa Coiled			**1.26%**	**772**	**247**

TABLE 2. The occurrence and extent of coiled coil domains in GenBank phylogenetic protein sequence files (release 84.0). The first row of data for each file represents all coiled coil proteins in the file with $P \geq 0.9$, and the second row is for proteins with over 75 amino acids in coiled coil domains. The percentage of proteins containing any coiled coil domains is relatively constant across phyla; however the occurrence of extended coiled coil domains is much greater in animals than in the other sequence files. Standard statistical methods are not applicable for reasons discussed in the text.

Several new observations emerged from the analysis. First, of the 102,007 protein sequences analyzed, 6074 (5.95%) contained coiled coil domains meeting the minimum probability value of 0.9. Of these, the vast majority (78.8%) contained domains of 75 or fewer coiled coil-forming amino acids. Also, excepting only the invertebrate proteins, of which 9.47% had coiled coil domains, the rate of occurrence of these domains was remarkably constant across all phyla. Coiled coils occurred in the other eight sequence files at a rate of 5.9 $\pm$ 1% (see Table 2). It must be noted that the extension of this analysis to include statistical tests for significance of differences (such as t-tests, etc.) would be invalid. This is due to the nature of the sequence data in the files, which is a complex mixture of complete genomic sequences for certain viruses and phages, redundant reports from different groups of investigators, partial sequences, over-representation of some proteins of high abundance in cDNA libraries, and some near-random sequence data from the genomes of certain well-studied organisms such as *E. coli* and *S. cerevisiae*. These inherent properties of the data sets which prevent the measurement of statistical significance led us to explore other means of making comparisons among groups of proteins and organisms in regard their potential biological significance.

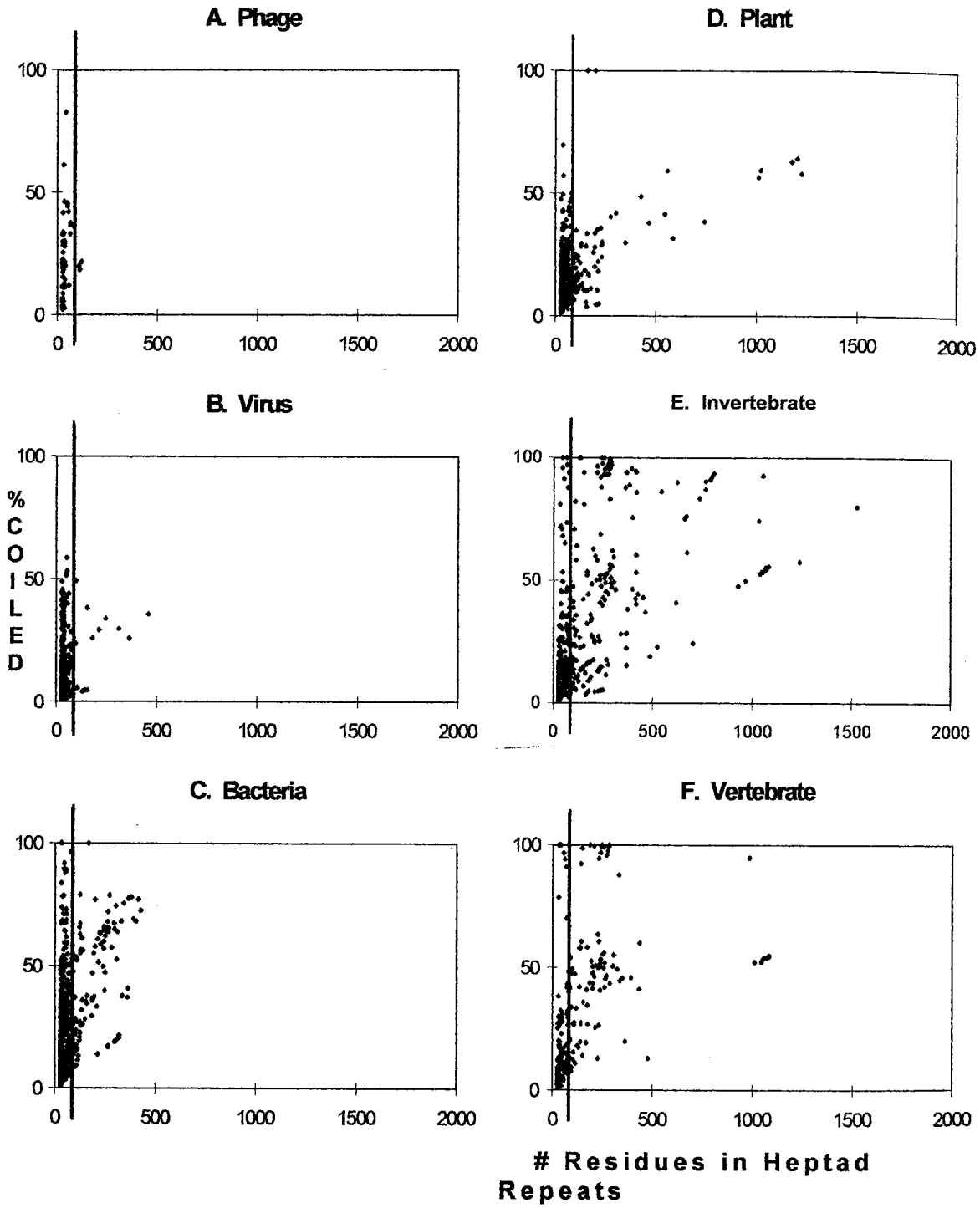
Extended Coiled Coils More Frequent Among Animal Proteins. Although our results showed that coiled coil domains occur with similar frequency in all organisms, when the analysis was restricted to those proteins whose coiled coil domains total greater than 75 amino acids, a striking phylogenetic difference emerged. The occurrence of extended coiled coils in the five animal kingdom files represented 2.18% (978/45,045) of total proteins, whereas in the other four files it was just 0.55% (311/56,962), or 4-fold less (Table 2). The average length of the coiled domains in animal protein sequences was 113 residues, over twice the average for non-animal proteins, which was 50 residues.

The low rate of occurrence of extended coiled coils in the non-animal files is not due simply to a shorter overall length of the coiled coil proteins in these phyla. For example, inspection of Table 2 shows that the average total length of coiled coil proteins in the plant and viral files is actually greater than that in both the rodent and the vertebrate files. Furthermore, the biases in the sequence files

described above do not explain the paucity of extended coiled coil proteins in at least some of the non-animal files. Especially notable is that, out of 17,733 viral proteins analyzed, only 29 (0.18%) had more than 75 amino acids lying within heptad repeat domains. In addition, the non-animal files contain not only many completely sequenced viruses and phages, but also bacteria, yeasts, and certain well-studied higher plants. It is unlikely that the sequencing of currently unknown *E. coli* or *S. cerevisiae* genes, for example, will contribute to a sudden increase in the fraction of known coiled coil proteins in their genomes. We therefore anticipate that the proportionality of the results should remain reasonably constant as more genes are sequenced; nevertheless, the formal possibilities do remain that there may be disproportionate numbers of undiscovered coiled coil proteins in plants, or that the fraction of coiled coil proteins in animals will decrease as more random sequences are obtained. These issues can be addressed by the periodic re-assessment of updated sequence files.

Comparisons of Phylogenetic Plots. The graphs in Figure 21 represent all the coiled coil proteins identified by the Coils program in the GenBank phylogenetic files, plotted on the same scale (note: the primate graph is extended by 64 amino acids to accommodate giantin, X75304, the first protein listed in Figure 20). The X value is the number of amino acids in a protein which have a probability of coiled coil formation ≥ 0.9, *i.e.*, the second value after the GenBank definition lines in Figure 20; the Y value is the percentage of the total amino acids in the protein which lie within these regions ("% coiled"), the third value after the definition lines. Each point on the graphs represents a protein sequence which met the probability cut-off of 0.9. The vast majority of these are seen clustered nearest the Y-axis and represent proteins with the shortest coiled coil domains, implying that by far the most frequent function for coiled coils is to mediate dimerization.

The vertical lines in Figure 21 which parallel the Y-axes are placed at 75 amino acids to demarcate the lower limit of what are considered in this analysis as extended coiled coil domains. It is evident that there is a rapid drop-off in the number of proteins as one scans each plot from left to right. The presence of incomplete coding sequences in the files affects the location of some data points, and


A. Phage
D. Plant
B. Virus
E. Invertebrate
C. Bacteria
F. Vertebrate
% COILED
100
50
0
0
500
1000
1500
2000
Residues in Heptad Repeats

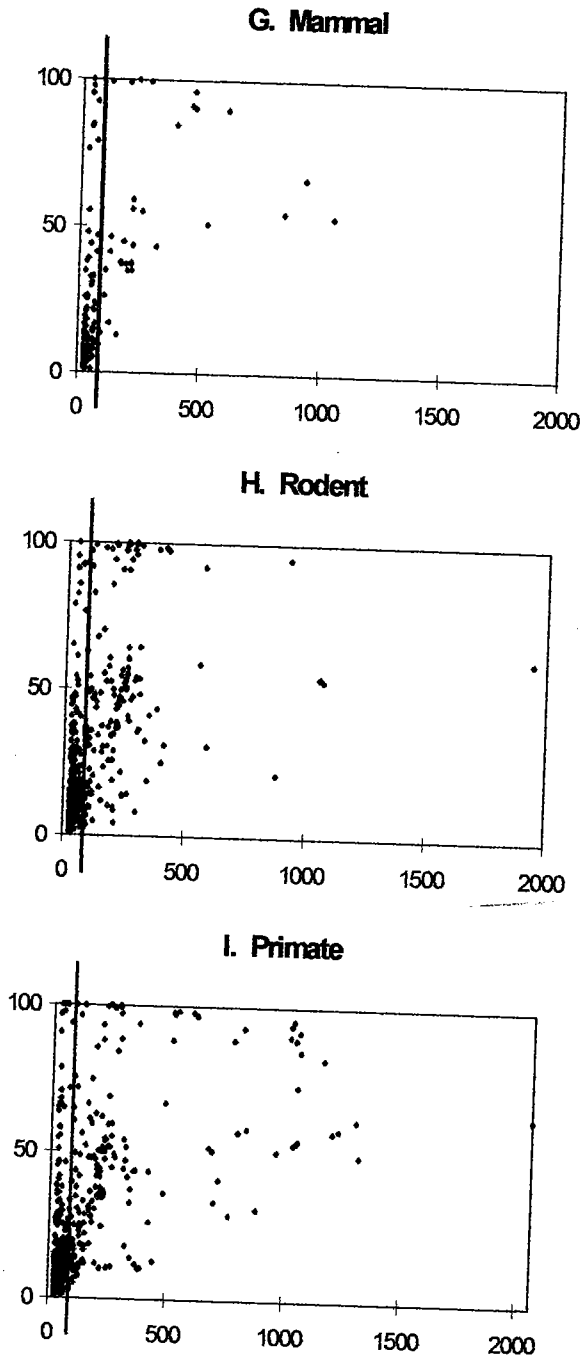


Figure 21. Plots, on the same scale, of all coiled coil proteins found in GenBank release 84.0 by the Coils batch program. The X-axis shows the number of amino acids which reside in heptad repeat domains, and the Y-axis shows the percentage of the protein's amino acids which lies within these domains. Each point on these graphs represents one protein in which a coiled coil probability ≥ 0.9 occurs, and together they represent all 6074 coiled coil proteins found in the phylogenetic sequence files. The majority of points are clustered in the region of the X-axis corresponding to 4 to 10 heptad repeats (28 to 70 amino acids). Simple dimerization domains, usually 28-50 amino acids long, undoubtedly account for most of these. The vertical line is placed at 75 amino acids to demarcate the lower bound of what are analyzed separately as "extended coiled coil" proteins. The animal sequence files all contained many examples of large proteins with extensive coiled coil forming regions (the mammal and vertebrate files have fewer sequence entries; see text). In plants and bacteria, domains were shorter and much less frequent, while in phages and viruses, extended domains were extremely rare.

duplicate submissions result in some superimpositions; nevertheless, differences between the animal sequences (Figure 21 E-I) and those of the other phyla (Figure 21 A-D) are apparent. In the animal file plots, particularly the primates, rodents, and invertebrates, many points representing proteins with very long coiled coil regions are spread out over the entire graph. In the mammal and vertebrate plots (Figure 21 G and F) there appear to be fewer extended coil proteins; however, these two files together contained only 8696 protein sequences, representing 19.3% of the 45,045 animal sequences (Table 2). They also contained 190 extended coiled coil proteins, or 19.4% of all such proteins found in animals, a result consistent with the view that extended coiled coil proteins occur with fairly constant frequency in the genomes of higher animals.

Analyses of Extended Coiled Coil Proteins:

1)Viral and Phage. The phage and virus plots (Figure 21 A and B) and data in Table 2 show how uncommon extended coiled coil domains are in these organisms, especially considering that not only did the viral sequence file contain 17,733 proteins, over 25% more than the primate file, but also that genomic sequences for many viruses are complete or near-complete. Duplicate submissions from different investigators and nearly identical proteins from related viruses result in fewer data points than the numbers listed in Table 2. Of the 29 extended coiled coil proteins, 15 (52%) were defined only as open reading frames found in the genomes of a variety of plant and animal viruses, and these were not investigated further. Three of the four proteins with the most residues in heptad repeat domains are A-type inclusion proteins from cowpox (D00319), mousepox (X69325), and orthopox (X69774) viruses, proteins which form a protective outer covering for virions when the host cell is killed (Moss, 1986). Three are from *vaccinia* DNA-dependent RNA polymerase (X57318; two in M61187). There are in addition a *variola* fusion protein (X67115), a G-protein-coupled receptor homologue from *Herpesvirus saimiri* (S76368), a capsid protein from bluetongue virus (D12532, D01183), a coat protein of peanut clump virus (L07269), and a zinc-finger protein from a *polyhedrosis* virus (M33604). Among the 1859 bacteriophage proteins analyzed, just 6 (0.32%) contained extended coiled coil domains. Of these, two are duplicate

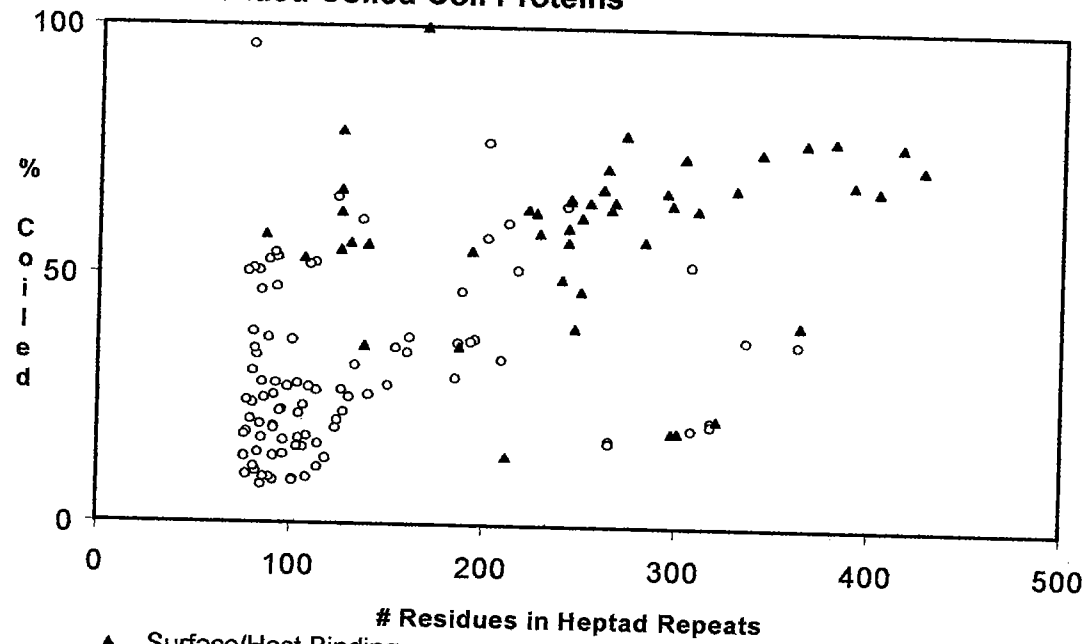
submissions of T4 55, alpha-gt/46 genes, and two are duplicates of T5 60.5-65.6% DNA. The other two extended coil phage proteins are in: 1) the BF23 17 and 18 sequence; and 2) P22 gp 10 and gp26 sequence.

In summary, proteins with extended coiled coil domains are almost vanishingly rare in viruses and phages, their domains are far shorter than those in other organisms, and the functions of only a few are known. Those are about equally divided among proteins which are on the outer surface, which interact with host RNA polymerase, or which form inclusion bodies in host cells.

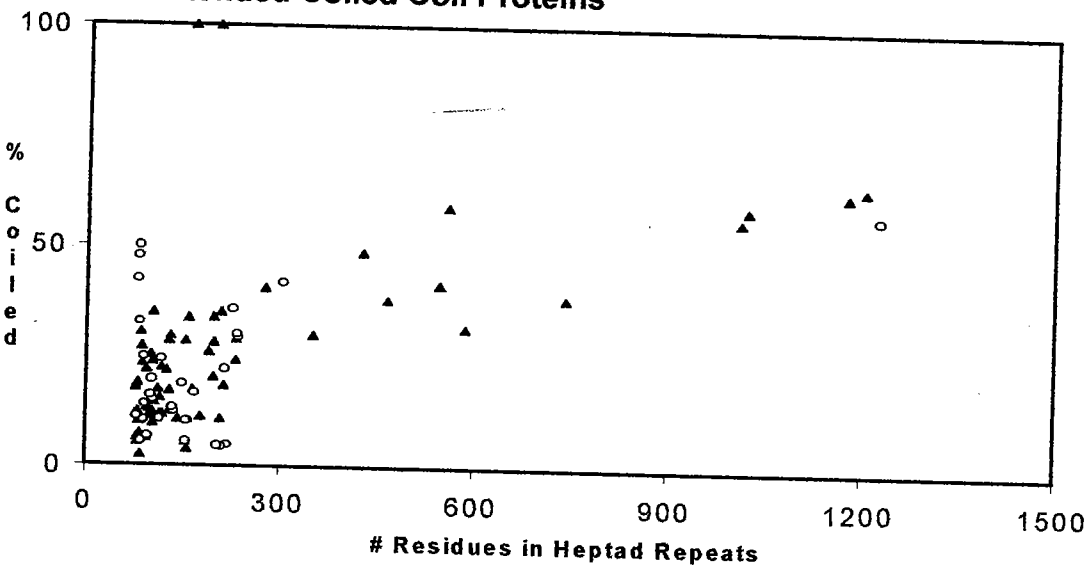
2) Bacteria. The bacteria plot (Figure **21C**) and data in Table **2** indicate that while extended coiled coil proteins do occur in prokaryotes, the domains are shorter overall than those in eukaryotes. The mean number of amino acids in heptad repeat regions of these proteins was 164 residues, and they occurred in only 166 proteins out of a total of 22,388 (0.74%). Examination of the report file and its submission references revealed that a large fraction of these extended coiled coil proteins shared an intriguing property, namely, they are antigenic surface proteins involved in the adhesion of pathogenic organisms to eukaryotic host cells and proteins. Figure **22A** shows a plot of bacterial extended coiled coil proteins after such classification. Many bacterial proteins with the highest heptad repeat contents fall into this category (see Table 2). These include the many variants of M proteins of *Streptococci* (Fischetti et al., 1988), and other proteins which bind plasminogen (Z32677), immunoglobulins (e.g., X69324, X15198, X73158), or other host proteins. In fact, the majority of the points seen in the upper region of Figure **22A** appear to result from the mutation and propagation of a single M protein gene (Scott et al., 1985). In addition to these host-binding proteins, there were, among the most extended coiled coil bacterial proteins, the *E. coli* chromosome partitioning protein mukB (D31701), and a number of surface antigens whose functions were not clearly identified. Another interesting bacterial protein was OMP alpha (X68276, S49394) from the ancient eubacterium *Thermotoga maritima*. This protein has been shown to form 7 nm rods which connect the outer membrane with inner cell body (Engel et al., 1992), a function reminiscent of the cytoplasmic filaments of eukaryotes. Overall, heptad repeat content of greater than 75

Figure 22. Bacterial and plant extended coiled coil proteins divided into categories (see Methods) and plotted as in Figure 3. **A** shows bacterial extended coiled coil proteins. Those known to occur on the surface of pathogens and to mediate binding to eukaryotic host proteins, including *Streptococcus* M proteins, are shown with closed triangles. Others are indicated by open circles. Note that many of the proteins with the most extensive coiled coil domains are in the surface/binding class of bacterial proteins. **B** shows the extended coiled coil proteins identified in the plant sequence file. Those from the yeasts *S. cerevisiae* and *S. pombe* are shown with closed triangles, and proteins from other plants are denoted with open circles. 70% of the 110 proteins represented in this plot are from yeast, and less than 25% are from higher plants.

A. Bacterial Extended Coiled Coil Proteins



B. Plant Extended Coiled Coil Proteins



amino acids occurs infrequently in bacterial proteins, and the domains are on average much shorter than those found in the animal kingdom. Very few of these seem to play structural roles, and many of those which do occur interact specifically with eukaryotic host proteins.

3) Plants, Including Yeasts and Other Fungi. The GenBank data are organized such that yeasts, other fungi, and molds are included in the plant sequence file. Figure **22B** is a plot of all plant extended coiled coil proteins, with yeast proteins indicated by a separate symbol, and Table **3** contains summary data from this analysis. Of the 110 extended coiled coil proteins identified among the 14,982 protein sequences in the plant file, 76 (69.1%) were from *S. cerevisiae*, including 11 of the 12 with the greatest heptad repeat content. The only non-yeast protein among these 12 was *D. discoidum* myosin (M14628, M11938). Among the 11 most extensively coiled coil yeast proteins were myosin type II (X53947), three submissions of a myosin-like protein (L01992, L01993), an integrin homologue (L03188), and USO1 (X54378), a protein involved in cytoskeleton attachment of endoplasmic reticulum and transport to distal secretory pathway components, all from *S. cerevisiae*. The rare proteins from photosynthetic plants which contained heptad repeat domains totaling 100 or more amino acids included 3 chloroplast-encoded open reading frames (X64616, S55421, X64615) from *O. odorata* (evening primrose), 4 kinesin- and dynein-like motor proteins, and *A. thaliana* myosin (Z28389). And, in sharp contrast to the animal myosins, the Arabidopsis myosin coiled coil domain constitutes only 157 amino acids of a total of 1520. The fact that so few extended coiled coil proteins have been identified in higher plants is consistent with the fact that cytoplasmic intermediate filaments have never been observed in these organisms, nor, obviously, do plants require the many coiled coil proteins associated with locomotion.

4) Invertebrates and Primates. The largest numbers of extended coiled coil proteins were found in the invertebrate and primate files. To investigate the potential biological significance of this result, we determined the contribution made by sets of proteins with known, related structure/function characteristics to this high rate of occurrence. The primate and invertebrate sequences with heptad repeat domains

File	# Proteins	Mean (SD) Length of Coiled Coil Proteins	Mean (SD) Length of Coiled Coil Domains
Bacteria			
Surface/Host Binding	33	377 (126)	242 (87)
Other	117	534 (336)	126 (66)
Plant			
Yeast	78	1037 (759)	234 (272)
Other Plants	32	1226 (1183)	173 (201)
Invertebrate			
NMIF	35	1108 (883)	247 (85)
Myosins and Motors	82	990 (1100)	447 (322)
Other	131	950 (686)	197 (201)
Primate			
NMIF	111	847 (709)	251 (212)
Myosins and Motors	58	714 (584)	483 (369)
Other	112	892 (738)	212 (281)

Table 3. Occurrence and Domain Lengths of Extended Coiled Coil Proteins Analyzed by Category

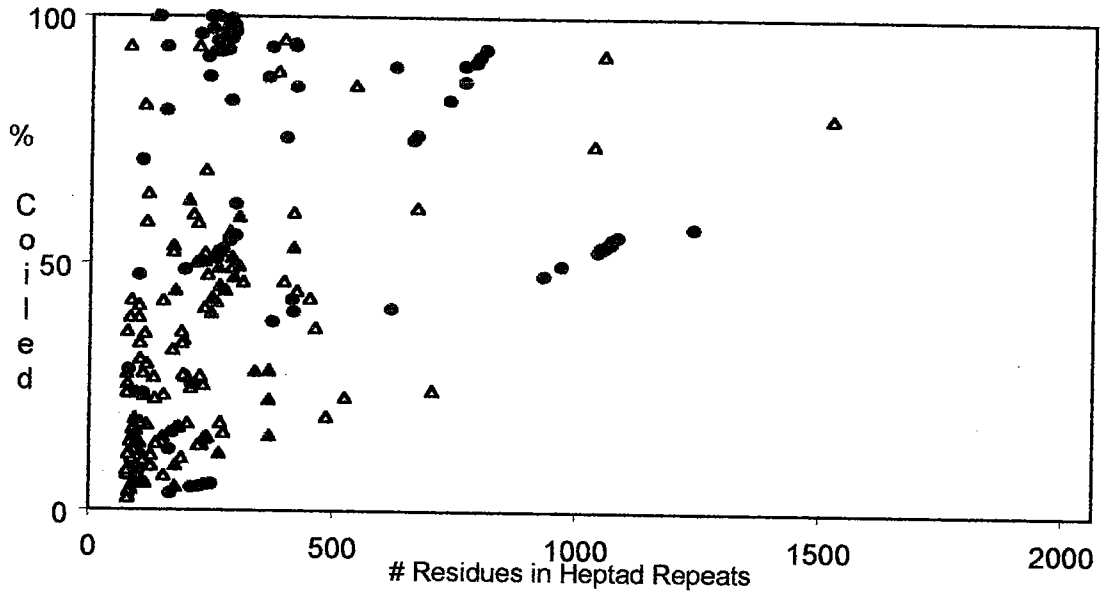
Extended coiled coil proteins (i.e., >75 amino acids with coiled coil probability scores equal to or greater than 0.9) after sub-classification as described in Methods. Bacterial proteins are categorized as Surface/Host Binding and Other. Proteins from the plant sequence file are sub-divided into Yeast and Other. The invertebrate and primate proteins are categorized as Myosins and Motors, NMIF, or Other. Bacterial proteins which occur on the surface of pathogens and mediate binding to host proteins and cells contain on average longer domains than other bacterial proteins. Yeasts encode the majority of extended coiled coil proteins found in the plant file. Nearly one third of invertebrate extended coiled coil proteins identified are myosins and motors, and only 13% are NMIF proteins. In primates, 40% are NMIF proteins and 21% are myosins and motors. The coiled coil domains in myosins and motor proteins are much longer, on average, than those in the other categories, and the lengths of domains in the different categories are quite similar in primates and invertebrates.

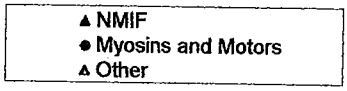
totaling over 75 amino acids were therefore further classified as "NMIF" or as "Myosins and Motors" (see Methods). These categories are mutually exclusive. The myosins and motors form a group related functionally through their involvement in force production and motility, properties not shared by NMIF proteins. Also, even though kinesin and dynein subunits have extensive and interrupted coiled coil domain organization similar to those seen in many NMIF proteins, they are readily extracted as soluble proteins in non-ionic detergents (Schroer et al., 1989) even when the microtubules to which they bind are first stabilized with taxol (Gill et al., 1991). This excludes their requirement for the morphological fidelity observed in NMIF preparations. In this analysis distinctions were not made as to the exact nature of the coiled coil interactions NMIF proteins may participate in, for example whether they are known to form two- versus three-stranded α–helical bundles. Rather, proteins are included together in the NMIF category because 1) they are selectively retained in NMIF preparations, and 2) the biochemical mechanism which underlies their helical associations is still based on the heptad repeat. As more functional and biochemical data become available for individual proteins in the "Other" category, assignment to the NMIF or Myosins and Motors categories may be warranted. The results of this classification scheme are plotted for the invertebrates in Figure **23A** and for the primates in **23B**, and summary statistics are shown in Table **3**.

Inspection of the plot in Figure **23A** shows the predominance of the myosins and motors among the known invertebrate extended coiled coil proteins. Of the 14 known proteins which contained heptad repeat domains totaling >1000 amino acids, 11 were myosins. Of the 248 extended coiled coil proteins identified in this file, 82 (33.1%) were myosins, tropomyosins, paramyosins, or subunits of the motor proteins kinesin or dynein. There were 35 known NMIF proteins (14.1%), and the remaining 131 (52.8%) were classified as Other, either because they are not part of the myosin or NMIF categories or because their identification was ambiguous. In the Other category were a number of interesting proteins from parasitic organisms which are in some respects similar to certain of the bacterial extended coiled coil proteins in that they are antigenic, occur on the surface of the organism, and mediate specific interaction with host cell proteins. The highest heptad repeat content among invertebrate proteins, 1526 residues out

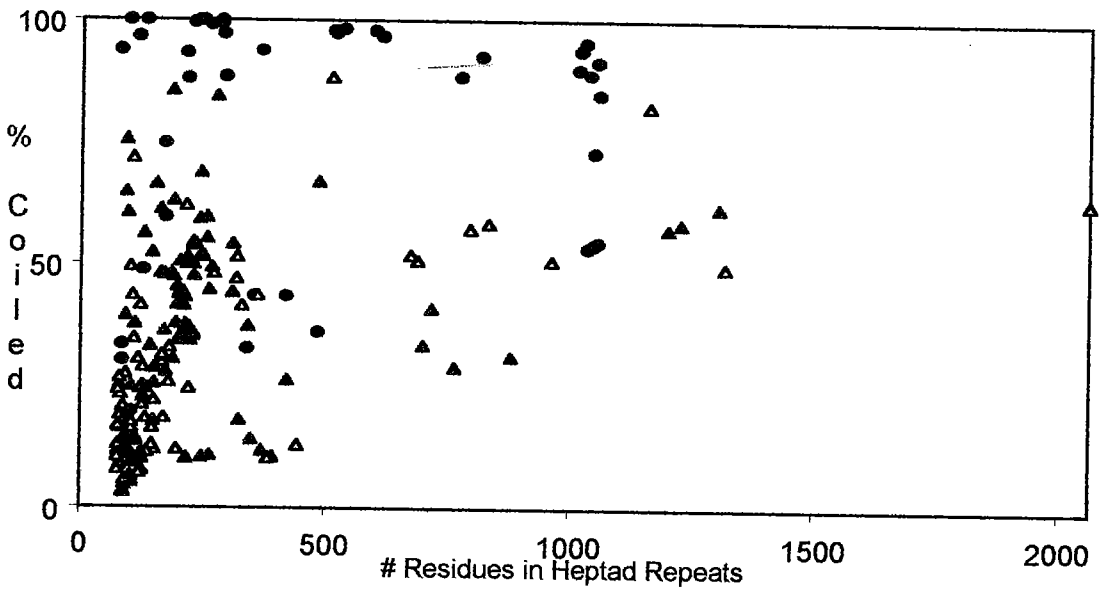
Figure 23. Invertebrate, **A**, and primate, **B**, extended coiled coil proteins are categorized (see Methods) and plotted as in Figure 3. Red triangles are NMIF proteins; blue circles are Myosins and Motors; other proteins, not currently known to belong to either of these groups, are open triangles. NMIF includes the intermediate filaments, certain intermediate filament-associated proteins, lamins, NuMA, tpr, and so on. The cluster of NMIF points centered near 250 residues, 50% coiled are intermediate filament proteins, including nuclear lamins. In **A**, cytoplasmic intermediate filament proteins are from metazoan invertebrates. In the primate plot, **B**, many NMIF proteins are seen to lie above and to the right of this cluster, where there are none in the invertebrate graph. Nearly all of the most extended coiled coil domains in the invertebrates are in myosins and motors from a wide variety of organisms, as are most of the proteins near the 100% regions of both graphs (mainly tropomyosins and paramyosins). In the invertebrate "other" category are many proteins with very extensive coiled coil domains located on the surface of parasites. They tend to be antigenic and to mediate binding to host cells, analogous to some found in bacteria. The coiled coil content of NMIF protein sequences ranged from about 75 to 1400 amino acids and from roughly 20% and 75% coiled.

A. Invertebrate Extended Coiled Coil Proteins





B. Primate Extended Coiled Coil Proteins



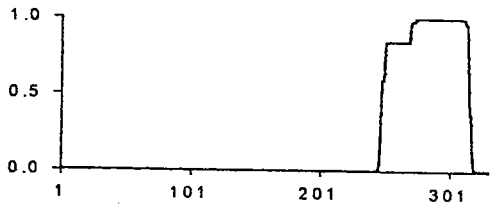
of 1909 total, was in the LSA-1 gene for liver stage antigen in the malarial parasite *Plasmodium falciparum* (X65203, S75010, S75012, S75014, S75016). The status of LSA-1 and many other such proteins in terms of the extended coiled coil protein classifications is currently unknown. The invertebrate proteins in the NMIF category were primarily intermediate filament proteins (including the lamins and neurofilament proteins), laminins, and spectrins from a variety of metazoan invertebrates, including *D. melanogaster*, *C. elegans*, and squid. The four NMIF proteins in Figure **23A** in the region of 350-400 coiled coil residues and 15-25% coiled are laminins and spectrin from *D. melanogaster*.

The primate extended coiled coil proteins are shown plotted according to the same classification scheme in Figure **23B**, and summary data are in Table **3**. The Myosins and Motors contained coiled coil domains which averaged 483 amino acids in length, much longer than those in the other categories, and overall, the domain lengths of all the primate categories were very similar to those of invertebrates (see Table **3**). The relative proportions in these categories, however, were different. Of the 281 extended coiled coil proteins identified in primates, 111 (39.5%) were constituents of the NMIF, 58 (20.6%) were Myosins and Motors, and the remaining 112 were Other. The intermediate filaments are seen centered in the same region as in the invertebrate plot. The highest coiled coil content identified in this study was in human giantin (X75304), a protein associated with the Golgi apparatus. Whether giantin is retained during NMIF fractionation has not been determined experimentally. Among the primate extended coiled coil proteins there are constituents of all the ultrastructural elements seen in NMIF preparations, from laminin in the extracellular matrix to NuMA in the interior nuclear matrix.

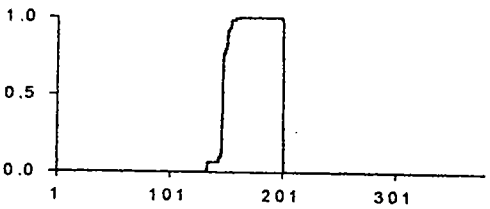
Domain organization of coiled coil proteins. As sequence data have become available for NMIF constituent proteins, the coiled coil content has often been reported as well. We noted the consistency with which major constituents of the NMIF contained heptad repeat domains which are extensive yet discontinuous, in contrast to myosin, for example, in which the coiled coil region in the tail is virtually uninterrupted for 1000 amino acids. In figures **24** and **25**, plots of individual proteins illustrate the extent of heptad repeat domains and their organization, with amino acid number plotted against coiled coil probability. Examples of different classes of proteins from a variety of organisms are shown. Short coiled

Figure 24. Plots of coiled coil probabilities of proteins of various categories. Plots are of individual proteins, with amino acid number on the X-axis and its coiled coil probability on the Y-axis. **A**, **B**, and **C** show examples of short coiled coil domains which mediate specific dimerization. The human proto-oncogene transcription factors fos (V01512) and jun (X51345), shown in **A** and **B**, contain the "leucine zipper" type of coiled coil domains in which leucine occupies positions *a* and *d* in the heptad repeats. **C**, the *E. coli* restriction/methylase Eco D1 (V00287, J01631) is an example of a dimerization domain from prokaryotes. Proteins with short domains, as shown here, occur in all phyla with surprisingly constant frequency, and represent nearly 80% of all coiled coil proteins. The next group represents extended coiled coil proteins from bacteria and parasites. **D** and **E** are examples of antigenic coiled coil proteins found on the surfaces of pathogenic bacteria and parasites, respectively, which mediate binding to host cell proteins. **D** is a *Streptococcus pyogenes* M protein, 24M (M19031). The M proteins mediate adhesion to pharyngeal epithelia and inhibit complement-dependent phagocytosis. **E** is a plot of the reticulocyte binding protein RBP1 of the malarial parasite *Plasmodium vivax* (M88098). **F** shows the *E. coli* mukB protein (X57550, M38402, M63930), which is required for chromosome segregation and is probably a motor protein (28, 29). **G** is a plot of the OMP alpha protein (X68276, S49394) from the eubacterium *Thermotoga maritima*, which has been shown to form 7 nm rods connecting the outer membrane to the inner cell body. Examples of the "myosins and motors" category are plotted as in Figure 6. **H**, human β–myosin heavy chain (M58018) and **I**, human tropomyosin 1 (M19267) illustrate the extensive and uninterrupted coiled coil domains in this class of molecules. **J**, *D. melanogaster* cytoplasmic dynein heavy chain (L23195), and **K**, human kinesin heavy (X65873) chain show the less extensive and interrupted coiled coil domain organization of these microtubule-associated motor complex subunits. The carboxy-terminal half of myosin is the coiled coil forming tail region which mediates the polymerization of myosin into filaments, while tropomyosin homodimers bind along the length of actin filaments. It is apparent that the motor protein subunits shown here have a more complex overall organization of coiled coil domains than either the DNA binding dimers (**A-C**) or the myosins. This is consistent with the complexity of their interactions, which include formation of holoenzyme complexes, binding to both microtubules and to a variety of transport substrates, ATP hydrolysis, and force production.

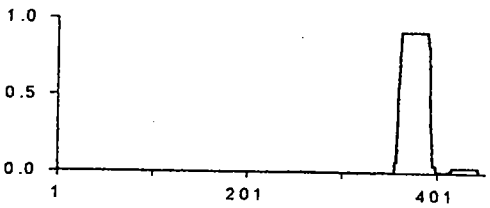
DIMERIZATION



A. JunB



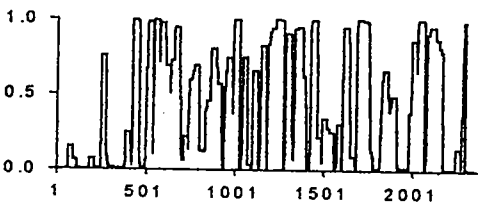
B. cfos



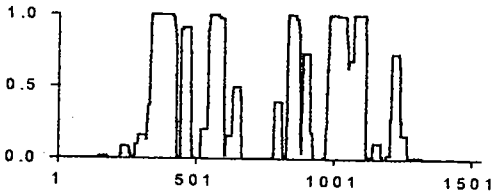
C. Eco D1

PATHOGENS AND PARASITES

D. Steptococcus 24M



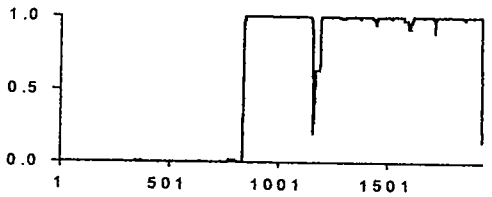
E. P. vivax reticulocyte binding protein



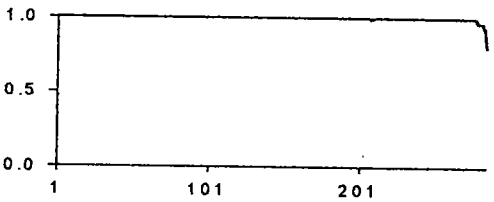
F. E. coli mukB

G. T. maritima Omp Alpha

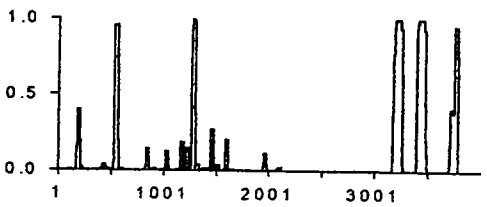
MYOSINS AND MOTORS



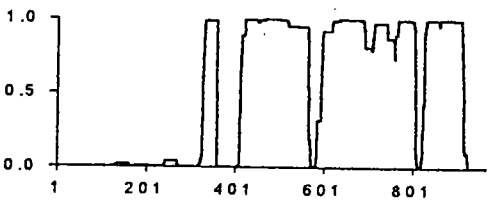
H. β-Myosin Heavy Chain



I. Tropomyosin 1



J. Dynein Heavy Chain



K. Kinesin Heavy Chain

coil domains which mediate dimerization are exemplified by the human "leucine zipper"-containing transcription factor dimerization pair cfos (V01512) and Jun B (X51345), and the *E. coli* restriction methylase Eco D1(V00287, J01631), shown in Figure **24A**, **B**, and **C**, respectively. This type of heptad repeat domain is by far the most common in nature, accounting for nearly 80% of all coiled coil proteins in the GenBank files (see Figure **21** and Table **2**).

Figure **24D-G** illustrate several interesting extended coiled coil proteins from bacteria and parasites. **D** shows *Streptococcus pyogenes* 24M protein (M19031). The M proteins, which typically have uninterrupted domains of heptad repeats, are present on the bacterial surface where they mediate binding to pharyngeal epithelia and inhibit complement-dependent phagocytosis (Fischetti et al., 1988). Shown in **E** is a reticulocyte binding protein of *Plasmodium vivax* (M88098), which exemplifies parasite proteins analogous to the M proteins in their antigenicity, surface occurrence, and specification of host target. Whether this large, complex protein also plays a role in merozite structure is currently unknown. **F** shows the *E. coli* mukB protein (X57550 M38402-M63930), originally identified through studies of mutants deficient in chromosome segregation, and thought to be a force-producing protein (Niki et al., 1991; Hiraga, 1993).**G** is a plot of the OMP alpha protein (X68276 S49394) from the eubacterium *Thermotoga maritima*, which has been shown to form 7 nm rods connecting the outer membrane to the inner cell body (Engel et al., 1992).

24H-K show examples of Myosins and Motors. These include human β–myosin heavy chain (**H**, M58018), tropomyosin (**I**, M19267), and sub-units of microtubule-associated motor proteins, dynein heavy chain from *D. melanogaster* (**J**, L23195) and human kinesin heavy chain (**K**, X65873). The continuous domains of the myosins contrast with the interrupted domains of the motor proteins. In fact, the coiled coil domain organization of some motor sub-units is strikingly similar to that of NMIF components (compare **K** with Figure **25**): however, unlike NMIF proteins, coiled coil motor proteins are extracted from cells in the soluble fraction and thus do not participate in the free-standing structure of the

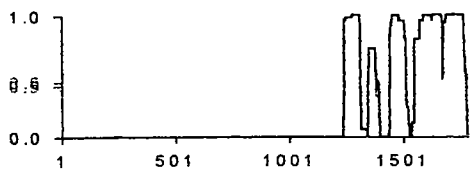
NMIF. The extractability of these proteins demonstrates that NMIF fractionation is not simply a way to "aggregate" non-specifically all proteins with similar biochemical characteristics, and it strengthens other lines of evidence that the NMIF represents structures which pre-exist in the living cell.

Figure 25 illustrates coiled coil proteins which participate in different domains of NMIF structure, from the extracellular matrix to the interphase nuclear matrix. **A** shows human laminin B1 (M61916, J02778), a component of the basement membrane. Laminins form hetero-trimeric triple-stranded coiled coils which interact with other extracellular and transmembrane proteins (Martin and Timpl, 1987). β–spectrin (**B**, human, J05500) lies immediately beneath the plasma membrane and contacts both other cytoskeletal elements and transmembrane proteins. It has been best-studied in the erythrocyte, where it connects transmembrane proteins with the cytoskeleton (Lux, 1979; Luna and Hitt, 1992). Mutations which impair spectrin dimerization and prevent the formation of a normal spectrin network have recently been shown to cause developmental failure in *Drosophila* (Deng et al., 1995). Human desmoplakin I (M77830), a major constituent of the desmosomal junctions between epithelial cells (Green et al., 1990), is shown in **C**. Intermediate filament proteins are represented in **D** and **E**, respectively, by human cytokeratin 18 (X12883, M26325) and vimentin (M14144). Intermediate filament-associated proteins are illustrated in **F** by rat plectin (X59601), a cytoplasmic protein involved in vimentin filament cross-linking (Wiche et al., 1993). Pericentrin, **G** (rat, U05823), a major component of the pericentriolar material at the microtubule organizing center (Doxsey et al., 1994), exhibits highly complex coiled coil domain organization. Constituents of the human nuclear pore-lamina complex are illustrated by the nucleoporin tpr (**H**, X66397) and lamin B (**I**, M34458). NuMA (**J**, Z11583), a component of the interior nuclear matrix, also participates in mitotic spindle formation during cell division. Note the extensive and interrupted nature of the coiled coil domains common to all these proteins. While similar in some respects to certain motor protein subunits (see Figure 24), the latter proteins are not NMIF components. Considered together, these data suggest that the NMIF

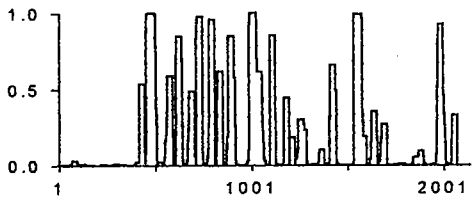
represents a highly complex, integrated, tissue-wide network whose underlying resilience, morphology, and resistance to extraction resides in large measure in multiple coiled coil associations among certain of its constituent proteins.

Figure 25. Examples of extended, interrupted, coiled coil domains in proteins which are NMIF constituents, arranged roughly in order from outside the cell into the interphase nuclear matrix. **A** shows human laminin I (M61916, J02778), a basement membrane component. **B**, β–spectrin (J05500), lies beneath the plasma membrane where it contacts both cytoskeletal and transmembrane proteins. **C**, desmoplakin I, (M77830), is a major constituent of the desmosome. Intermediate filament proteins are represented by **D**, cytokeratin 18 (X12883, M26325), and **E**, vimentin (M14144). **F** is the intermediate filament cross-linking protein plectin (X59601). **G**, pericentrin (U05823), is a major component of the pericentriolar material. Components of the nuclear pore-lamina complex are illustrated by **H**, the nucleoporin tpr (X66397), and **I**, lamin B (M34458). **J**, NuMA (Z11583), is a component of the interior nuclear matrix and the mitotic spindle. The plots of the larger proteins are highly complex, suggesting that their interactions may be correspondingly intricate.

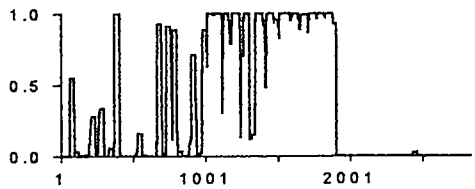
NMIF PROTEINS



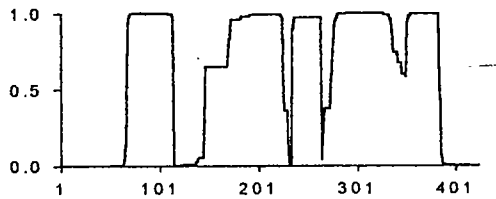
A. Laminin B1



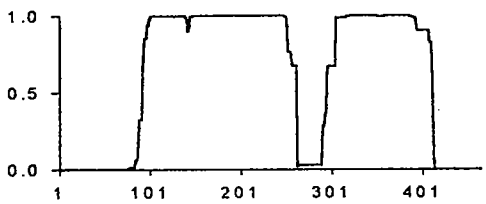
B. β-Spectrin



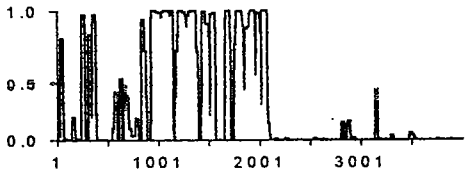
C. Desmoplakin I



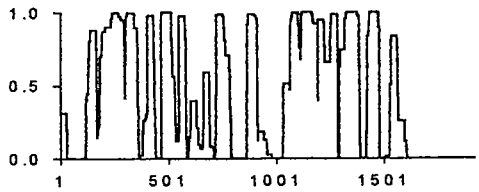
D. Cytokeratin 18



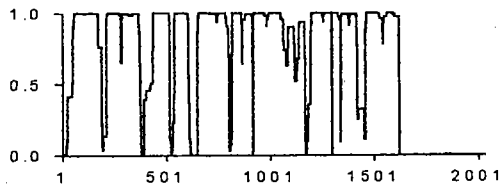
E. Vimentin



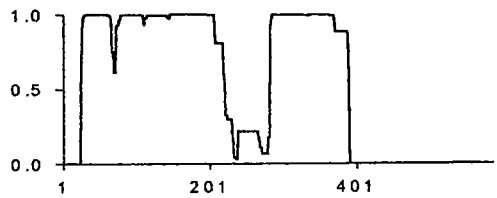
F. Plectin



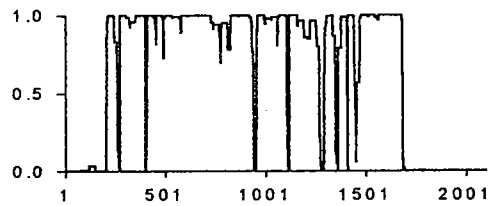
G. Pericentrin



H. TPR



I. Lamin B



J. NuMA

DISCUSSION

Coiled Coil Proteins in the Animal Kingdom: the Coiled Coil Tissue Matrix. To explore questions arising from an apparent relationship between coiled coil proteins and higher-order cell and tissue structure, the phylogenetic occurrence and relative lengths of coiled coil proteins were examined. Several new observations resulting from this analysis are presented here. First, the data suggest that the widespread occurrence of very extended coiled coil proteins is a hallmark the animal kingdom (see Figure 21 and Table 2). Moreover, only a portion of this high incidence can be attributed to the locomotion- and motility-associated proteins shared by Protozoa and Metazoa (Table 3 and Figure 23). What may in particular set apart metazoan organisms is the relationship between structural complexity and the number of the "NMIF"-type of proteins present, which are characterized by extensive and discontinuous heptad repeat domains (Figure 25). Heptad repeat-containing NMIF proteins are further distinguished from proteins with similar organization, such as microtubule-associated motors (24J and K), by their biochemical behavior in fractionation experiments. In fact, the extraction of the coiled coil motor proteins in non-ionic detergents provides an important counter-example to NMIF constituents: possession of interrupted, extended coiled coil domains is not sufficient for retention in the tissue-wide structures of the NMIF. This implies that the great enrichment of extended coiled coil proteins achieved by NMIF fractionation is selective and does not simply reflect biochemical similarity. Rather, it strongly suggests that proteins escape extraction by virtue of their incorporation into a continuous, pre-existing structure. Based upon its great enrichment in extended coiled coil proteins, its tissue-wide morphological fidelity, and its free-standing interconnectedness, this structure is designated the "coiled coil tissue matrix."

The proposal of a tissue-wide, interconnected, structural matrix is certainly not a new idea. Such models have been proposed, and studies of their relationship to signal transduction and differentiation have been pioneered by Ingber (Wang et al., 1993; commentary in Ingber, 1993) and Coffey (Isaacs et al., 1981; Pienta and Coffey, 1992) and their collaborators. Based upon the results reported in the present study, it would appear that extended coiled coil proteins of the NMIF are ideal candidates for constituents of such a tissue-wide matrix. First, they are present in all the major structural elements observed in NMIF

preparations, from the extracellular matrix to the nuclear interior (Figure 25), and many have been shown to vary with cell type. Second, it is shown here that their occurrence among proteins known thus far appears to correlate with organism complexity. Also, past uncertainties over the elusive function for intermediate filaments (Geiger, 1987; Paulin-Levasseur, 1992) have begun to be answered through recent studies such as those which clearly link keratin mutations to epidermal tissue defects (Chan et al., 1994), and which demonstrate that ectopic expression of epidermal keratin K14 in transgenic mouse liver has grave consequences for the morphology and function of that tissue (Albers et al., 1995). This suggests a vital role for intermediate filaments in tissue integrity and in intracellular organization. Finally, and perhaps most importantly, many workers have already demonstrated and mapped connectivity between individual coiled coil proteins which are NMIF constituents. For example, Georgatos and Blobel (1987a-b) demonstrated the interaction of vimentin with lamin B and with the plasma membrane of avian erythrocytes; Kouklis et al. (1994) have mapped the connectivity between cytokeratins and desmoplakin; Wiche et al. (1993) have identified the domains of plectin responsible for its association with vimentin intermediate filaments; and Lee, et al. (1993) showed that tricohyalin, another intermediate filament associated protein, associates in regular arrays with keratin intermediate filaments in the stratum granulosum of the epidermis. Taking one step back from these individual interactions provides a wider view in which interconnections among coiled coil proteins form an integrated, tissue-wide, stable matrix upon which other cell and tissue elements could be superimposed hierarchically. This is not to say that all of the hundreds of proteins obtained through an NMIF isolation are extended coiled coil proteins, but rather, that a coiled coil-based substratum provides the structural backbone with which non-coiled coil proteins associate.

While coiled coil-based connectivity can be traced with some confidence from the extracellular matrix to the nuclear lamina, connections across the nuclear envelope are still far from clear. The major extended coiled coil protein of the interior nuclear matrix presently known is NuMA (see Figure 20 and 25J). Sequence analysis by Parry (1994) has shown that NuMA, though capable of dimerizing as a two-stranded, parallel, coiled coil, does not contain the long-range periodicity needed for filament formation.

These predictions were confirmed by electron microscopic and biochemical studies of NuMA (Harborth et al., 1995). Furthermore, despite direct observations of filaments within the nuclear matrix, the available evidence indicates that such filaments contain RNA and/or RNP complexes (Fey et al., 1986; Lawrence et al., 1989; Xing and Lawrence, 1991). No purely proteinaceous filaments have to date been demonstrated inside the nucleus. Penman (1994) even speculates that the compartmentalization of chromatin into nuclei was an evolutionary response to the hazards imposed on DNA by the dynamic filaments and other structures of the cytoplasm. A major challenge will be to establish the mode of structural connectivity between the cytoplasm and the genome.

Divergence of Plants. In the plant file report, a striking deficiency of extended coiled coil proteins was noted, especially among the more complex higher plants. While the sequencing efforts in these organisms has been overall less intensive than in others, nevertheless, based on the ideas presented here, the prediction can be made that the proportion of their genomes encoding extended coiled coil proteins will remain lower than that seen in animals for several reasons. First, plants obviously do not require the many extended coiled coil proteins which participate in locomotion in animals, and their myosins have much shorter coiled coil domains. Moreover, the structural "backbone" of higher plants is deposited in the form of carbohydrate polymers, mainly cellulose, in the extracellular spaces. While the intermediate filament family is represented in green plants by the lamins (Minguez and Moreno Diaz de la Espina, 1993; Frederick et al., 1992), as is true for all eukaryotes, plants lack a cytoplasmic intermediate filament network. Higher animals must construct and maintain many highly specialized and localized organs and tissues in discreet regions of the body, each tissue utilizing a characteristic set of intermediate filament proteins. In plants, functions such as digestion, respiration, and photosynthesis are widely distributed throughout the organism. Also, a major factor which determines cell-to-cell organization in plants, the commitment to an axis along which to undergo cell division, is mediated by the microtubules (Lloyd, 1994). Thus, the deficiency of extended coiled coil proteins in plants may reflect very early evolutionary events in the divergence of plants and animals, not only concerning motility, but also whether coiled coil proteins regulate the morphology of cells and their assembly into tissues.

The incorporation of extended coiled coil proteins into NMIF structural elements common to all eukaryotes, for example the nuclear pore-lamina complex, implies that a minimal set of such proteins is encoded in all eukaryotic genomes. The extensive sequence data for *S. cerevisiae* could offer a means for future assessment of some of the ideas presented in the present work. Of the 929 coiled coil plant proteins identified here, 412 (44%) are yeast proteins. Of the 110 extended coiled coil proteins in the file, 76 (69%) are from *S. cerevisiae* alone. As sequence information for this organism nears completion, it will represent a less biased data set than is available for other eukaryotes, and it should be possible to determine the true fraction of the *S. cerevisiae* genome which encodes extended coiled coil proteins of the NMIF type. Given the relative structural simplicity of yeast, this fraction should represent a near-minimal set required by eukaryotes. The data reported here are consistent with the notion that the number of such proteins expressed by an organism correlates with the degree of differentiation it undergoes during its development. Animals would be expected, therefore, to surpass that minimum as required by their elaboration of higher order, coiled coil-incorporating tissue structures.

Extended Coiled Coils in Viruses and Prokaryotes. Among some 20,000 protein sequences in the virus and phage files, just 35 were in the extended coiled coil category, even fewer if one discounts duplicate submissions. This is despite the fact that the average length of viral proteins which contain short coiled coils is similar to that of other organisms (Table 2). It would appear that, having little need for higher-order structures or force production, viruses likewise have little need for extended coiled coil proteins. The rare exceptions usually involve either a holoenzyme assembly requirement, a host binding interaction, or construction of a protective structure (see Results). These findings in a sense constitute a phylogenetic negative control for the coiled coil tissue matrix hypothesis.

Next in order of complexity, the bacteria had a far greater number of extended coiled coil proteins than did viruses (Table 2 and Figure 22A); however, of these, a great many are not required for strictly "prokaryotic" purposes. That is, Figure 22A and Table 3 demonstrate that binding to cells and proteins of metazoan hosts is a major selective force driving the retention and propagation of genes for the most extensive coiled coil forming proteins currently known in bacteria, such as the M protein family of *S.*

pyogenes (Fischetti et al., 1988; Scott et al., 1985). Besides these, extended coiled coil proteins with essential functions do occur in prokaryotes. One example is *E. coli* mukB (Figure **24F**), a large, complex coiled coil protein required for chromosome segregation during cell division (Niki et al., 1991). Its organization is similar to eukaryotic kinesin. mukB binds nucleotide triphosphates and it is the first candidate for a bacterial motor protein (Hiraga, 1993). Along with FtsZ protein, mukB has been proposed to represent a bacterial functional homologue to eukaryotic proteins involved in cytokinesis (Norris et al., 1994). Overall, however, the present study shows that extended coiled coil proteins essential to the bacterial life cycle constitute only a very small fraction of their genomes.

Short Coiled Coils: Universal Modular Assembly Domains. Short coiled coil-forming domains were found with nearly equal frequency across phyla, from viruses to prokaryotes, from plants to primates. This implies that the ability to form specific protein dimers in a modular fashion through coiled coil interactions is an ancient, highly conserved, and successful evolutionary strategy. As such, further study of these dimerization domains should prove useful in taxonomy and molecular evolution. The ubiquity and success of this dimerization strategy may be due in part to the stable and self-sufficient nature of the coiled coil. Su et al. (1994) have shown that the melting temperature of coiled coil dimers of synthetic peptides increases by 34°C when the length is increased from 23 to 26 residues, and that further increases yield less dramatic gains in stability. Thus, the 28 residue domains identified here are near the minimum required for stability in the aqueous environment of the cell. The many and subtle effects of hetero- and homodimer formation on gene regulation have been described by others (for reviews, see Lee, 1992; Lamb and McKnight, 1991), but the universal, nearly constant occurrence of coiled coil domains as one means of achieving these effects has not been previously noted.

Future Prospects. Based upon the presence of extensive, interrupted coiled coil domains, it should be possible to predict with some success the selective retention of proteins through NMIF fractionation. It will be most interesting to determine whether NMIF fractionation results in retention of extended coiled coil proteins not yet studied in this regard. Several of these proteins are associated with organelles (for example human giantin) and could be mediating interaction with the NMIF. It may also be

possible to develop multivariate predictors of the participation of extended coiled coil protein in the NMIF. Sequence analysis algorithms which measure the "interruptedness" of coiled coil domains, or which discriminate between coiled coil motor proteins and NMIF proteins might form a useful starting point. As mentioned above, studies of *S. cerevisiae* may soon disclose a minimal set of extended coiled coil proteins required for eukaryotic life. Metazoan organisms should exceed this minimum in rough proportion to their tissue complexity.

This matrix model may be relevant to the relationship between cell shape and the regulation of proliferation. Folkman and Moscona (1979) first clearly demonstrated that proliferative signals were regulated in some fashion by cell shape. Fey and Penman (1984c; 1986) showed that transformation of cells is accompanied by characteristic morphological distortions of the NMIF. The model proposed here also implies that regulation of expression of extended coiled coil proteins in development may be integrated via feedback through the coiled coil tissue matrix. Work by Byrne et al. (1994) provides a potential example of such a mechanism. They showed that embryonic transcriptional activity of a human keratin promoter in transgenic mice correlates with variations in the embryonic origin of the underlying mesenchyme. Given those findings, together with the morphological information which resides in the coiled coil tissue matrix, it is plausible that certain cell cycle regulatory and differentiation signals are transduced via this network.

The work presented here attempts to extend observations of both heptad repeat-based coiled coil protein occurrence and the cell-type specificity of NMIF proteins to the phylogenetic level, and it suggests a unified biochemical basis for the major connections within the NMIF structure. The overall similarities in interrupted heptad repeat domain organization, as well as the enormous variety, of the extended coiled coil proteins of this tissue-wide matrix provides both a potential explanation and an underlying mechanism for the diverse cell forms in metazoan organisms.

Abbreviations: BSA: bovine serum albumin; **COxII**: cytochrome oxidase subunit II; **CSK**: cytoskeleton buffer, 300 mM sucrose, 100 mM NaCl, 10 mM PIPES, 3 mM $MgCl_2$, 1 mM EGTA, and 0.5 % Triton X-100, pH 6.8; **DAPI**: 4'-6-diamidino-2-phenylindole; **DB**: digestion buffer, 300 mM sucrose, 50 mM NaCl,

10 mM PIPES, 3 mM $MgCl_2$, 1 mM EGTA, 0.5 % Triton X-100, pH 6.8; **DHFR**: dihydrofolate reductase; **DiOC6**: 3,3'-dihexyloxacarbocyanine iodide; **ER**: endoplasmic reticulum; **FUV**: 7M urea, 0.5%SDS, 1X MOPS (4.18g/l MOPS, pH 7), 2.2% formaldehyde, 20 mM vanadyl ribonucleoside complex; G_{av}: the centrifugal force exterted upon a sample at a point midway between the top of the sample and the bottom; **HMP**: heart muscle protein; **IB**: iso-osmotic buffer, 0.3M sucrose, 10 mM Tris, 1 mM EDTA, pH 7.5; **IF**: intermediate filament; **IPTG**: isopropyl-β-D-thiogalactylpyranoside; M_r: relative electrophoretic mobility; **MSB**: microtubule stabilization buffer, 10 mM PIPES, 3 mM $MgCl_2$, 1 mM EGTA, and 0.5% Triton X-100; **NBT/BCIP**: nitroblue tetrazolium/ 5-bromo-4-chloro-3-indoyl phosphate; **OS**: osmotic shock buffer, 1 volume of IB plus 5 volumes of 20 mM Tris, 1 mM EDTA, pH 7.5; **NMIF**: nuclear matrix-intermediate filament scaffold; **PBS**: phosphate buffered saline; **PBSA**: PBS containing 0.5% BSA; **rcf**: radial centrifugal force, the force exerted on a sample at the point farthest from the center of the rotor; **rMF**: recombinant mitofilin, starting at met 437, as described in Chapter 1; **SDS-PAGE**: sodium dodecylsulfate polyacrylamide gel electrophoresis; **TBS**: Tris buffered saline, 500 mM NaCl, 0.02% sodium azide, 20 mM Tris/Cl, pH 7.5; **TTBS**: TBS plus 0.05% Tween-20

REFERENCES

Albers, K., Fuchs, E.. 1992. The molecular biology of intermediate filament proteins. Int. Rev. of Cytol. 134:243-279.

Albers, K. M., Davis, F.E., Perrone, T.N., Lee, E.Y., Liu, Y., Vore, M. 1995. Expression of an epidermal keratin protein in liver of transgenic mice causes structural and functional abnormalities. J. Cell Biol. 128:157-169.

Amchenkova, A. A., Bakeeva, L.E., Chentsov, Y.S., Skulachev, V.P., Zorov, D.B. 1988. Coupling membranes as energy-transmitting cables I. Filamentous mitochondria in fibroblasts and mitochondrial clusters in cardiomyocytes. J. Cell Biol. 107:481-495.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreirer, P.H. 1981. Sequence of human mitochondrial genome. Nature. 290:457-465.

Attardi, G. 1984. Animal mitochondrial DNA: an extreme example of genetic economy. Int. Rev. Cytol. 93:93-145.

Attardi, G., Schatz, G. 1988. Biogenesis of mitochondria. Ann. Rev. Cell Biol. 4:289-333.

Ausubel, F. M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. 1990. *Current Protocols in Molecular Biology*. Vol. 1. John Wiley & Sons, New York. Section 4.9.

Avise, J. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. Ann. Rev. Genet. 25:45-69.

Aviv, H., Leder, P. 1972. Purification of biologically active globin messenger RNA by chromatography on oligo-thymidylic acid-cellulose. Proc. Nat. Acad. Sci. USA 69:1408-1412.

Bangs, P.L., Sparks, C.A., Odgren P.R., Fey, E.G. 1995. The product of the oncogene activating gene *tpr* is a phosphorylated protein of the nuclear pore complex. In press. J. Cell. Bioch.

Bereiter-Hahn, J., Voth, M. 1994. Dynamics of mitochondria in living cells: shape changes, dislocations, fusion, and fission of mitochondria. Microsc. Res. Tech. 27:198-219.

Berezney, R., Coffey, D.S. 1977. Nuclear matrix. Isolation and characterization of a framework structure from rat liver nuclei. J. Cell Biol. 73:616-637.

Bloom, G. S., Brashear, T.A. 1989. A novel 58-kDa protein associates with the Golgi apparatus and microtubules. J. Biol. Chem. 264:16083-16092.

Brady, S. T. 1885. A novel brain ATPase with properties expected for the fast axonal transport motor. Nature. 317:73-75.

Burgess, S. M., Delannoy, M., Jensen, R.E. 1994. MMM1 encodes a mitochondrial outer membrane protein essential for establishing and maintaining the structure of yeast mitochondria. J. Cell Biol. 126:1375-1391.

Byrne, C., Tainsky, M., Fuchs, E. 1994. Programming gene expression in developing epidermis. Development. 120:2369-2383.

Capaldi, R. 1990. Structure and Function of Cytochrome c Oxidase. Ann. Rev. Biochem. 59:569-596.

Chan, Y., Anton-Lamprecht, I., Yu, Q.C., Jackel, A., Zabel, B., Ernst, J.P., Fuchs, E. 1994. A human keratin 14 "knockout": the absence of K14 leads to severe epidermolysis bullosa simplex and a function for an intermediate filament protein. Genes & Devel. 8:2574-2587.

Chen, L. B. 1989. Fluorescent labeling of mitochondria. Meth. Cell Biol. 29:103-123.

Chirgwin, J.J., Przbyla, A. E., Macdonald, R.J., Rutter, W.J. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochem. 18:5294-5299.

Cohen, C., Parry, D.A.D. 1990. Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. Proteins: Struc. Func. and Genet. 7:1-15.

Conway, J. F., Parry, D.A.D. 1990. Structural features in the heptad substructure and longer range repeats of two-stranded α-fibrous proteins. Int. J. Biol. Macromol. 12:328-334.

Darbre, A. 1986. Analytical methods. In, *Practical Protein Chemistry*. A. Darbre, editor. John Wiley and Sons, Chichester. 284-287.

Daum, G., Bohni, P.C., Schatz, G. 1982. Import of proteins into mitochondria. J. Biol. Chem. 257:13028-13033.

Deng, H., Lee, J.K., Goldstein, L.S.B., Branton, D. 1995. Drosophila development requires spectrin network formation. J. Cell Biol. 128(1&2):71-79.

Doxsey, S. J., Stein, P., Evans, L., Calarco, P.D., Kirschner M. 1994. Pericentrin, a highly conserved centrosome protein involved in microtubule organization. Cell 76:639-650.

Engel, A. M., Cejka, Z., Lupas, A., Lottspeich, F., Baumeister, W. 1992. Isolation and cloning of Omp alpha, a coiled coil protein spanning the periplasmic space of the ancestral eubacterium Thermotoga maritima. EMBO J. 11:4369-78.

Fey, E. G., Wan, K. M., Penman, S. 1984a. Epithelial cytoskeletal framework and nuclear matrix-intermediate filament scaffold: three-dimensional organization and protein composition. J. Cell Biol. 98:1973-1984.

Fey, E. G., Capco, D. G., Krochmalnic, G., Penman, S. 1984b. Epithelial structure revealed by chemical dissection and unembedded electron microscopy. J. Cell Biol. 99:203s-208s.

Fey, E. G., Penman, S. 1984c. Tumor promoters induce a specific morphological signature in the nuclear matrix-intermediate filament scaffold of Madin-Darby canine kidney (MDCK) cell colonies. Proc. Nat. Acad. Sci. USA. 81:4409-4413.

Fey, E. G., G. Krochmalnic, Penman, S. 1986. The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. J. Cell Biol. 102:1654-1665.

Fey, E. G., Penman, S. 1987. The protein composition and morphology of the nuclear matrix-intermediate filament scaffold reflect cell type. Prog. Clin. Biol. Res. 249:263-272.

Fey, E. G., Penman, S. 1988. Nuclear matrix proteins reflect cell type of origin in cultured human cells. Proc. Nat. Acad. Sci. USA. 85:121-125.

Fey, E. G., Bangs, P., Sparks, C., Odgren, P.R. 1991. The nuclear matrix: defining structural and functional roles. Crit. Rev. Euk. Gene Expr. 1:127-143.

Fischetti, V. A., Bessen, D. 1988. Effect of mucosal antibodies to M protein on colonization by group A Streptococci. In, *Molecular Mechanisms of Microbial Adhesion*. Switalski, L., Hook, M., Beachey, E., editors. Springer-Verlag, New York. 128-142.

Folkman, J., Moscona, A. 1979. Role of cell shape in growth control. Nature. 273:345-349.

Frederick, S. E., Mangan, M.E., Carey, J.B., Gruber, P.J. 1992. Intermediate filament antigens of 60 and 65 kDa in the nuclear matrix of plants: their detection and localization. Exp. Cell Res. 199:213-222.

Fuchs, E., Weber, K. 1994. Intermediate filaments: structure, dynamics, function, and disease. Ann. Rev. Biochem. 63:345-382.

Geiger, B. 1987. Intermediate filaments. Looking for a function. Nature. 329:392-393.

Georgatos, S. D., Blobel, G. 1987a. Lamin B constitutes an intermediate filament attachment site at the nuclear envelope. J. Cell Biol. 105:117-125.

Georgatos, S. D., Blobel, G. 1987b. Two distinct attachment sites for vimentin along the plasma membrane and the nuclear envelope in avian erythrocytes: a basis for a vectorial assembly of intermediate filaments. J. Cell Biol. 105:105-115.

Getzenberg, R. H., Coffey, D.S. 1990. Tissue specificity of the hormonal response in sex accessory tissues is associated with the nuclear matrix protein patterns. Mol. Endocr. 4:1336-1342.

Gill, S. R., Schroer, T.A., Szilak, I., Steuer, E.R., Sheetz, M.P., Cleveland, D.W. 1991. Dynactin, a conserved, ubiquitously expressed component of an activator of vesicle motility mediated by cytoplasmic dynein. J. Cell Biol. 115:1639-1650.

Glick, B., Brandt, A, Cunningham, K, Muller, S, Hallberg, RL, Schatz, G. 1992a. Cytochromes c1 and b2 are sorted to the intermembrane space of yeast mitochondria by a stop-transfer mechanism. Cell. 69:809-822.

Glick, B., Wachter, C., Schatz, G. 1992b. The energetics of protein import into mitochondria. Bioch. Biophys. Acta. 1101:249-251.

Glick, B., Beasley, E.M., Schatz, G. 1992c. Protein sorting in mitochondria. Trends Bioch. Sci. 17:453-463.

Gould, B., Marks, V. 1988. Recent developments in immunoassays. In, *Non-isotopic Immunoassay*, T. Ngo, editor. Plenum Press, New York. 3-21.

Green, K. J., Parry, D.A.D., Steinert, P.M., Virata, M.L.A., Wagner, R.M., Angst, B.D., Nilles, L.A. 1990. Structure of the human desmoplakins. J. Biol. Chem. 265:2603-2612.

Hannavy, K., Rospert, S., Schatz, G. 1993. Protein import into mitochondria: a paradigm for the translocation of polypeptides across membranes. Curr. Opin. Cell Biol. 5:694-700.

Harborth, J., Weber, K, Osborn, M. 1995. Epitope mapping and direct visualization of the parallel, in-register arrangement of the double-stranded coiled-coil in the NuMA protein. EMBO J. 14:2447-2460.

Hartl, F.-U., Pfanner, N., Nicholson, D.W., Neupert, W. 1989. Mitochondrial protein import. Bioch. Biophys. Acta. 988:1-45.

Heggeness, M. H., Simon, M., Singer, S.J. 1978. Association of mitochondria with microtubules in cultured cells. Proc. Nat. Acad. Sci. USA. 75:3863-3866.

Hiraga, S. 1993. Chromosome partitioning in E. coli. Curr. Opin. Genet. Devel. 3:789-801.

Hirokawa, N. 1882. Cross-linker system between neurofilaments, microtubules, and membranous organelles in frog axons revealed by the quick-freeze, deep-etching method. J. Cell Biol. 103:33-39.

Hochstenbach, F., David, V., Watkins, S., Brenner, M.B. 1992. Endoplasmic reticulum resident protein of 90 kilodaltons associates with the T- and B-cell antigen receptors and major histocompatibility complex antigens during their assembly. Proc. Nat. Acad. Sci. USA. 89:4734-4738.

Icho, T., Ikeda, T., Matsumoto, Y., Hanaoka, F., Kaji, K., Tsuchida, N. 1994. A novel human gene that is preferentially transcribed in heart muscle. Gene. 144:301-306.

Ingber, D. E. 1993. Cellular tensegrity: defining new rules of biological design that govern the cytoskeleton. J. Cell Sci. 104:613-627.

Isaacs, J. T., Barrack, E.R., Isaacs, W.B., Coffey, D.S. 1981. The relationship of cellular structure and function: the matrix system. Prog. Clin. Biol. Res. 75A:1-24.

Johns, D. R. 1995. Mitochondrial DNA and disease. New Eng. J. Med. 333:638-644.

Johnson, L. V., Walsh, M.L., Chen, L.B. 1980. Localization of mitochondria in living cells with rhodamine 123. Proc. Nat. Acad. Sci. USA. 77:990-994.

Johnson, L. V., Walsh, M.L., Bockus, B.J., Chen, L.B. 1981. Monitoring of relative mitochondrial membrane potential in living cells by fluorescence microscopy. J. Cell Biol. 88:526-535.

Johnson, C., McNeil, J.A., Carter, K.C., Lawrence, J.B. 1991. A simple, rapid technique for precise mapping of multiple sequences in two colors using a single optical filter set. Genet. Anal. Tech. Appl. 8:75-76.

Klose, J., and E. Zeindl. 1984. An attempt to resolve all the various proteins in a single human cell type by two-dimensional electrophoresis: I. Extraction of all cell proteins. Clin. Chem. 30:2014-2020.

Koehler, G., Milstein, C. 1975. Continuous culture of fused cells secreting antibody of predefined specificity. Nature. 256:495-497.

Kouklis, P. D., Hutton, E., Fuchs, E. 1994. Making a connection: direct binding between keratin intermediate filaments and desmosomal proteins. J. Cell Biol. 127:1049-1060.

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

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227:680-685.

Lamb, P., McKnight, S.L. 1991. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. Trends Bioch. Sci. 16:417-422.

Lawrence, J. B., Singer, R.H., Marselle, L.M. 1989. Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization. Cell. 57:493-502.

Lee, C., Chen, L.B. 1988. Dynamic behavior of endoplasmic reticulum in living cells. Cell. 54:37-46.

Lee, C., Chen, L.B. 1993. Motility and construction of the endoplasmic reticulum in living cells. Subcell. Biochem. 21:343-352.

Lee, K. A. 1992. Dimeric transcription factor families: it takes two to tango but who decides on partners and the venue? J. Cell Sci. 103:9-14.

Lee, S. C., Kim, I.G., Marekov, L.N., O'Keefe, E.J., Parry, D.A.D., Steinert, P.M. 1993. The structure of tricohyalin. Potential multiple roles as a functional EF-hand-like calcium-binding protein, a cornified cell envelope precursor, and an intermediate filament-associated (cross-linking) protein. J. Biol. Chem. 268:12164-12176.

LeGendre, N., Matsudaira, P. T. 1989. Purification of proteins and peptides by SDS-PAGE. In, *A Practical Guide to Protein and Peptide Purification for Microsequencing*. Matsudaira, P.T., editor. Academic Press, San Diego, CA. p. 57.

Leto, T. L., Marchesi, V. T. 1984. A structural model of human erythrocyte protein 4.1. J. Biol. Chem. 259:4603-4608.

Lloyd, C. 1994. Why should stationary plant cells have such dynamic microtubules? Mol. Biol. Cell. 5:1277-1280.

Luna, E. J., Hitt, A.L. 1992. Cytoskeleton-plasma membrane interactions. Science. 258:955-964.

Lupas, A., Van Dyke, M, and Stock, J. 1991. Predicting coiled coils from protein sequences. Science. 252:1162-1164.

Lux, S. E. 1979. Dissecting the red cell membrane skeleton. Nature. 281:426-429.

Martin, G. R., Timpl, R. 1987. Laminin and other basement membrane components. Ann. Rev. Cell Biol. 3:57-85.

Martz, D., Lasek, R.J., Brady, S.T., Allen, R.D. 1984. Mitochondrial motility in axons: membranous organelles may interact with the force generating system through multiple surface binding sites. Cell Motility. 4:89-101.

McLachlan, A. D., Stewart, M. 1975. Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J. Mol. Biol. 98:293-304.

Melloni, R. H., Estes, P.S., Howland, D.S., DeGennaro, L.J. 1992. A method for direct measurement of mRNA in discrete regions of mammalian brain. Anal. Biochem. 200:95-99.

Michaels, G., Hauswirth, W.W., Laipis, P.J. 1982. Mitochondrial DNA copy number in bovine oocytes and somatic cells. Dev. Biol. 94:246-251.

Minguez, A., Moreno Diaz de la Espina, S. 1993. Immunological characterization of lamins in the nuclear matrix of onion cells. J. Cell Sci. 106(pt 1):431-439.

Mitchison, T., Kirschner, M.W. 1984. Microtubule assembly nucleated by isolated centrosomes. Nature. 312:232-237.

Moss, B. 1986. Replication of poxviruses. In, *Fundamental Virology*. Fields, B. N., Knipe, D.M., editors. Raven Press, New York. p. 649.

Nangaku, M., Sato-Yoshitake, R., Okada, Y., Noda, Y., Takemura, R., Yamazaki, H., Hirokawa, N. 1994. KIF1B, a novel microtubule plus end-directed monomeric motor protein for transport of mitochondria. Cell. 79:1209-1220.

Niki, H., Jaffe, A., Imamura, R., Ogura, T., Hiraga, S. 1991. The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of E. coli. EMBO J. 10:183-193.

Norris, V., Ayala, J.A., Begg, K., Bouche, J.P., Bouloc, P., Boye, E., Canvin, J., Casaregola, S., Cozzone, A.J., Crooke, E., et al. 1994. Cell cycle control: prokaryotic solutions to eukaryotic problems? Jour. Theor. Biol. 168:227-230.

O'Brien, T. W., Kalf, G.F. 1967. Ribosomes from rat liver mitochondria I. Isolation procedure and contamination studies. J. Biol. Chem. 242:2172-2179.

O'Shea, E. K., Rutkowski, R., and Kim, P.S. 1989. Evidence that the leucine zipper is a coiled coil. Science. 243:538-542.

Oas, T. G., McIntosh, L.P., O'Shea, E.K., Dahlquist, F.W., and Kim, P.S. 1990. Secondary structure of a leucine zipper determined by nuclear magnetic resonance spectroscopy. Biochem. 29:2891-2894.

Odgren, P. R., Harvie, L.W. Jr., Fey, E.G. 1995. The phylogenetic occurrence of coiled coil proteins: implications for metazoan development via a coiled coil tissue matrix. Proteins: Structure, Function, and Genetics. (in press).

Parry, D. A. D. 1994. NuMA/centrophilin: sequence analysis of the coiled-coil rod domain. Biophys. J. 67:1203-1206.

Paulin-Levasseur, M. 1992. Intermediate filament proteins: many unanswered questions, few unquestioned answers. Bioch. Cell Biol. 70:842-848.

Pauling, L., Corey, R.B., Branson, H.R. 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc. Nat. Acad. Sci. USA. 37:205-211.

Penman, S. 1994. Why nuclei? J. Cell. Bioch. 55:1-3.

Pfanner, N., Rassow, J., Weinhues, U., Hergersberg, C., Sollner, T., Becker, K., Neupert, W. 1990. Contact sites between inner and outer membranes: structure and role in protein translocation into the mitochondria. Bioch. Biophys. Acta. 1018:239-242.

Pienta, K., Coffey, D.S. 1992. Nuclear-cytoskeletal interactions: evidence for physical connections between the nucleus and cell periphery and their alteration by transformation. J. Cell. Bioch. 49:357-365.

Poulton, J. 1992. Mitochondrial DNA and genetic disease. BioEssays. 14:763-768.

Rapiejko, P.J., Gilmore, R. 1994. Signal recognition and targeting of ribosomes to the endoplasmic reticulum by the signal recognition particle do not require GTP. Mol. Biol. of the Cell. 5:887-897.

Sago, L.F., Yaffe, M.P. 1994. Regulation of mitochondrial morphology and inheritance by Mdm10P, a protein of the mitochondrial outer membrane. J. Cell Biol. 126:1361-1373.

Sanger, F., Nicklen, S., Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA. 74:5463-5467.

Schleyer, M., Neupert, W. 1985. Transport of proteins into mitochondria: translocational intermediates spanning contact sites between outer and inner membranes. Cell. 43:339-350.

Schroer, T. A., Schnapp, B.J., Reese, T.S., Sheetz, M.P. 1989. The role of kinesin and other soluble factors in organelle movement along microtubules. J. Cell Biol. 107:1785-1792.

Schwarz, E., Neupert, W. 1994. Mitochondrial protein import: mechanisms, components, and energetics. Bioch. Biophys. Acta. 1187:270-274.

Scott, J. R., Pullman, W.M., Hollingshead, S.K., and Fischetti, V.A. 1985. Relationship of M protein genes in group A Streptococci. Proc. Nat. Acad. Sci. USA. 82:1822-1826.

Stryer, L. 1988. *Biochemistry*. W. H. Freeman, New York. 1089.

Stuurman, N., van Driel, R., de Jong, L., Meijne, A.M., van Renswoude, J. 1989. The protein composition of the nuclear matrix of murine P19 embryonic carcinoma cells is differentiation-stage dependent. Exp. Cell Res. 180:460-466.

Su, J. Y., Hodges, R.S., Kay, C.M. 1994. Effect of chain length on the formation and stability of synthetic peptide alpha-helical coiled coils. Biochem. 33:15501-15510.

Szymanski, I., Odgren, P.R. 1984. Measurement of fragments of the third component of human complement on erythrocytes by a new immunochemical method. Vox Sanguinis. 46:92-101.

Szymanski, I., Swanton, R.E., Odgren, P.R. 1984. Quantitation of the third component of complement bound to stored red cells. Transfusion. 24:194-197.

Taneja, K. L., Lifchitz, L.M., Fay, F.S., Singer, R.H. 1992. Poly(A) RNA codistribution with microfilaments: Evaluation by in situ hybridization and quantitative digital imaging microscopy. J. of Cell Biol. 119:1245-1260.

Terasaki, M., Song, J., Wong, J.R., Weiss, M.J., Chen, L.B. 1984. Localization of endoplasmic reticulum in living and glutaraldehyde-fixed cells with fluorescent dyes. Cell. 38:101-108.

Traut, T.W. 1994. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur. J. Biopchem. 222:9-19.

von Heijne, G. 1986. Mitochondrial targeting sequences may form amphiphilic helices. EMBO Journal. 5:1335-1342.

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1:945-951.

Wang, N., Butler, J.P., Ingber, D.E. 1993. Mechanotransduction across the cell surface and through the cytoskeleton. Science. 260:1124-1127.

Wiche, G., Gromov, D., Donovan, A., Castanon, M.J., Fuchs, E. 1993. Expression of plectin mutants indicates a role of COOH-terminal domain in intermediate filament association. J. Cell Biol. 121:607-619.

Xing, Y., Lawrence, J.B. 1991. Preservation of in vivo RNA distribution within the nuclear matrix demonstrated by in situ hybridization coupled with biochemical fractionation. J. Cell Biol. 112:1055-1063.

Zackroff, R. V., Idler, W.W., Steinert, P.M., Goldman, R.D. 1982. In vitro reconstitution of intermediate filaments from mammalian neurofilament triplet polypeptides. Proc. Nat. Acad. Sci. USA. 79:754-757.