

11-28-2001

Construction and screening of A cDNA library for the C3 gene(s) of the Nurse Shark (Ginglymostoma Cirratum)

Janette Cristina Builes
Florida International University

DOI: 10.25148/etd.FI14051873

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Medical Sciences Commons](#)

Recommended Citation

Builes, Janette Cristina, "Construction and screening of A cDNA library for the C3 gene(s) of the Nurse Shark (Ginglymostoma Cirratum)" (2001). *FIU Electronic Theses and Dissertations*. 1940.
<https://digitalcommons.fiu.edu/etd/1940>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

CONSTRUCTION AND SCREENING OF A cDNA LIBRARY FOR THE C3 GENE(S)
OF THE NURSE SHARK (*GINGLYMOSTOMA CIRRATUM*)

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

MEDICAL LABORATORY SCIENCES

by

Janette Cristina Builes

2001

To: Dean Ronald M. Berkman
College of Health and Urban Affairs

This thesis, written by Janette Cristina Builes, and entitled Construction and Screening of a cDNA Library for the C3 Gene(s) of the Nurse Shark (*Ginglymostoma cirratum*), having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Janet A. Lineback

Charles H. Bigger

Sylvia L. Smith, Major Professor

Date of Defense: November 28, 2001

The thesis of Janette Cristina Builes is approved.

Dean Ronald M. Berkman
College of Health and Urban Affairs

Dean Douglas Wartzok
University Graduate School

Florida International University, 2001

©Copyright 2001 by Janette Cristina Builes

All Rights Reserved.

DEDICATION

I dedicate this thesis to my parents, Fabiola and Arturo Builes, and my sister Patricia Builes, who have been incredibly loving, supportive, patient, and understanding in all my educational endeavors. Without their belief in me this work would never have been accomplished. Thank you for everything you have done for me. To Andres Barreto a special thanks for all his advice, support, and love. You are a wonderful person.

ACKNOWLEDGMENTS

I would like to thank my mentor, Dr. Sylvia L. Smith, for giving me the opportunity to accomplish this research in her lab and for her belief in me and this project through her guidance, support, and patience. I also would like to thank the members of my committee, Dr. Janet A. Lineback and Dr. Charles H. Bigger, for their time and helpful suggestions in completing this thesis. A special mention and thanks go to Dr. Miki Nakao whose expertise guided and assisted me in constructing the cDNA library.

For the past few years, I have met several people in the MLS Program who have eased the stress of being a student through their jokes, kindness and support. From the lab I would like to thank past and present members Aline Jose, Ana Escobar, Barbara Webb, Betzy Gonzalez, Don-Ho Shin, Kyoko Nakamura, Luciana Amado, Margaret Contreras, Nichole Hinds, and Yanet Piriz. From outside the lab I would like to thank my family, Rosa Soto, Raul Bras, Jason Akl, Larry Dishaw, and my outside FIU friends. All of you have given me the added support to complete this thesis and I thank you and greatly appreciate your support. God bless you all.

ABSTRACT OF THE THESIS

CONSTRUCTION AND SCREENING OF A cDNA LIBRARY FOR THE C3 GENE(S) OF THE NURSE SHARK (*GINGLYMOSTOMA CIRRATUM*)

by

Janette Cristina Builes

Florida International University, 2001

Miami, Florida

Professor Sylvia L. Smith, Major Professor

Mammalian C3 is a complement protein which consists of an α chain (125kDa) and β chain (75kDa) held together by a disulfide bond. The α chain contains a conserved thiolester site which provides the molecule with opsonic properties. The protein is synthesized as a single pro-C3 molecule which is post-translationally modified. C3 genes have been identified in organisms from different phyla, however, the shark C3 gene remains to be cloned. Sequence data from the shark will contribute to understanding further the evolution of this key protein. To obtain additional sequence data for shark C3 genes a cDNA library was constructed and screened with a DIG-labeled C3 probe. Fifty clones were isolated and sequenced. Analysis identified four sequences that yielded positive alignments with C3 of a variety of organisms including human C3. Deduced amino acid sequence analysis confirmed a β/α cut site (RRRR), the CR3 and properdin binding sites, the catalytic histidine, and the reactive thiolester sequence. In the shark there are at least two C3-like genes as the gene sequence obtained is distinct from that previously described.

TABLE OF CONTENTS

CHAPTER	PAGE
I. Introduction	1
II. Background	4
The Mammalian Complement System (4)	
The Biological Role of C3 (9)	
C3 Binding Sites (12)	
C3 of Nonmammalian Species (15)	
C3 in the Nurse Shark (18)	
III. Materials and Methods	21
RNA Isolation from Nurse Shark Liver (21)	
Electrophoretic Analysis of RNA (21)	
mRNA Isolation (22)	
cDNA Synthesis (23)	
Construction of cDNA Library (25)	
Amplification of the Uni-ZAP XR Library (31)	
Lifting Plaques (32)	
Preparation of DIG-labeled C3 Probe (33)	
DIG Nucleic Acid Hybridization and Detection Kit (35)	
Positive Plaque Collection (37)	
Secondary Screening (38)	
In Vivo Excision Protocols Using ExAssist Helper Phage (38)	
Restriction Enzyme Digestion of C3 Confirmed Plasmids (40)	

	Design of Internal Primers (41)	
	Sequence Analysis and Alignments (41)	
IV.	Results	42
	Total RNA Isolation from Nurse Shark Liver (42)	
	Isolation of mRNA (42)	
	CDNA Synthesis (44)	
	Packaging Uni-ZAP XR Library into Extracts (49)	
	Amplification of Library and Screening of cDNA Library (52)	
	Plaque Lifts (52)	
	In Vivo Excision of the pBluescript Phagemid (56)	
	Sequencing Reactions (56)	
	Size Determination of Inserts (57)	
	Analysis of C3-Like Clones (67)	
V.	Discussion	82
VI.	References	89
VII.	Appendix	95

LIST OF TABLES

TABLE		PAGE
1	Primers used for obtaining nucleotide sequence of C3 clones	40
2	Determination of plaque forming units of library samples	51
3	Identification of C3-like clones isolated from the library	55
4	Percent identity between the deduced amino acid sequence	74
5	Homologous regions present I the nurse shark C3 protein	88

LIST OF FIGURES

FIGURE	PAGE
1 Schematic illustration of the classical, lectin and alternative pathways	2
2 Molecular structure of Pro-C3, native C3 and C3b	10
3 Deduced amino acid sequence of the thiolester region amplified	19
4 Electrophoretic analysis of total cellular RNA in a 1% agarose, 1xTBE gel	43
5 Electrophoretic analysis of cDNA synthesized from mRNA	45
6 Estimation of cDNA concentration in column fractions	47
7 Electrophoretic analysis of column fractions for cDNA samples A and B	48
8 Estimation of cDNA concentration in pooled fractions of samples A and B	50
9 Agarose electrophoresis of DIG-labeled probes	53
10 Alignment of nucleotide sequence data of C3-like clones	58
11 Alignment of nucleotide sequence data of C3-like clones	61
12 Electrophoretic analysis of restriction digests of Clones I, II, III, and IV	66
13 Nucleotide and deduced amino acid sequence of C3 clone P16A1.2/1	68
14 Nucleotide and deduced amino acid sequences of C3 clone P16A1.2/1	69
15 Nucleotide and deduced amino acid sequence obtained for clone C3	70
16 Nucleotide and deduced amino acid sequence of C3 clone P16A1.2/1	71
17 Nucleotide and deduced amino acid sequence of internal sequence data	72
18 Compilation of nurse shark C3 cDNA sequences	76
19 Alignment of deduced amino acid sequences of NSC3-I, II, and III	77
20 Alignment of amino acid sequence deduced from C3 clones I, II, and III	81

I. Introduction

The complement system is an important recognition and effector system of innate immunity (Jensen, et al., 1968; Day, et al., 1970; Reid, et al., 1981; Muler-Eberhard, et al., 1988). It is made up of several serum components that work in a series of reactions to attack invading pathogens. Components of the complement activation pathways are illustrated in figure 1. The third component (C3) is pivotal as it is central to the alternative, classical, and lectin pathways of complement activation. C3, along with C4, are the thiolester containing complement proteins that are genetically related to each other and to C5. All three proteins are believed to have evolved from a primitive α -2 macroglobulin-like molecule (Sottrup-Jensen, et al., 1985; Liszewski and Atkinson, 1993). C3 is synthesized as a single chain pro-C3 which is post-translationally modified to a 190-kD glycoprotein that consists of two disulfide-linked chains known as the α and β chains. The α chain contains a highly conserved thiolester site which provides the molecule with opsonic properties. C3, upon activation by C3 convertases, is cleaved into two fragments, C3a and C3b (figure 1). Both fragments have several important roles in immune defense. C3b can: (a) opsonize invading pathogens, (b) mediate phagocytosis, (c) bind B cells bearing CR1 receptor, and (d) bind to C4bC2a or C3bBb Classical Complement Pathway or Alternative Complement Pathway C3 convertases to form C5 convertase of the classical/lectin pathways or alternative pathway, respectively. C3a is a chemotactic and anaphylactic peptide. Studying the complement system of the nurse shark (a primitive elasmobranch) can help us understand how complement systems have evolved and their role as an immune modulator. Whereas mammals have a single C3 molecule, in lower vertebrates multiple isoforms of the molecule exist (Sunyer and Tort,

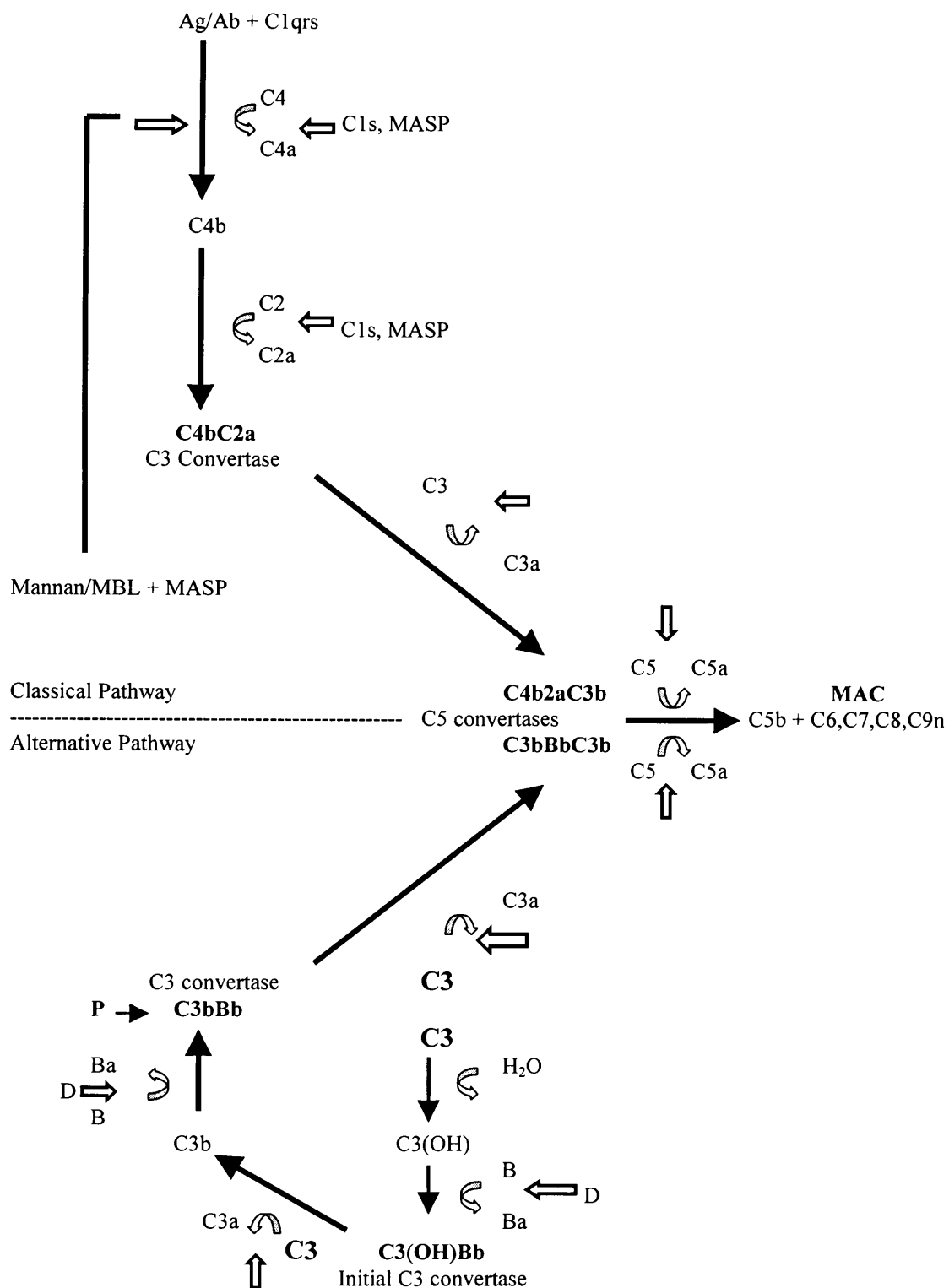


Figure 1. Schematic illustration of the classical, lectin and alternative complement activation pathways (see text for details) (bold straight arrows). Reactions catalyzed by serine proteases denoted by open arrows. Cleavage fragments of activated components shown by curved arrows.

et al., 1997; Nakao and Yano, 1998; Zarkadis, et al., 2001;) encoded by several genes. As C3 is a pivotal protein, this study proposes to focus on this component alone. Previous studies have yielded partial sequences for C3 homologues in the shark (Dodds and Smith, 1998; Lee, 1999). While only one biologically active form of the protein has been isolated from shark serum, gene cloning studies have yielded two partial cDNA clones which are 68% similar in deduced amino acid sequence indicating that in the shark at least two C3 genes exist. Furthermore, it appears, from earlier studies that one sequence, NsC31 lacks the catalytic histidine at position 1126 and in its place is a tyrosine (Lee, 1999). To understand the immunobiology of this molecule, it is necessary to obtain additional sequence data. The aim of the study is to construct and screen a cDNA library for clones that will provide additional nucleotide sequence data. The experimental approach will include (a) construction of a cDNA library, (b) screening of the library with a Digoxigenin labeled probe prepared from partial C3 cDNA clone (C3A4), and (c) isolation and sequencing of positive C3 clones. The sequence data obtained will permit phylogenetic analysis.

II. Background

The Mammalian Complement System

The mammalian complement system is defined as several plasma proteins that act in a series of coordinated reactions to bring about lysis and/or opsonization of pathogens (Bokisch, et al., 1969; Muller-Eberhard, et al., 1988). Three activation pathways exist in mammals (figure 1) that can be initiated in different ways, they are: the classical complement pathway (CCP) (Muller-Eberhard, et al., 1966; Muller-Eberhard, 1975), the lectin pathway (Sato, et al., 1994), and the alternative complement pathway (ACP) (Medicus, et al., 1976; Liszewski and Atkinson, 1993). The classical pathway can be activated by antigen-antibody complexes. The lectin pathway, which is considered to be the most primitive was, however, the most recent to be described. This pathway is initiated by the binding of a serum lectin, the mannan-binding lectin (MBL), to mannose-containing proteins or carbohydrates present on the surface of bacteria or viruses. The alternative pathway is not activated by antigen-antibody complexes, but rather it can be spontaneously activated by interaction with pathogens of microbial origin or complex molecules such as lipopolysaccharides, zymosan, nucleic acids, etc. The alternative pathway is normally in a low level state of activation in serum where it is activated by hydrolysis of the thioester bond of C3 leading to the formation of C3(H₂O), a C3b-like molecule. The three activation pathways lead to the formation of a C5 convertase (C4b2aC3b, C3bBbC3b) that initiates the terminal lytic pathway by cleaving C5 and activating C6 through C9 to form the Membrane Attack Complex (MAC) which, when inserted into membranes, brings about cell lysis (Green, et al., 1969; Humphrey, et al., 1969; Gotze, et al., 1970; Pangburn, et al., 1980; Bhakdi, 1983; Muller-Eberhard, 1986).

Thus, complement activation can be described as a cascade of reactions that includes proteolytic cleavage of some of the components with binding of activation fragments to other complement molecules forming multimolecular enzymatic complexes that bind to targets, bringing about lysis or opsonization.

The Classical Pathway of Complement Activation

The classical complement pathway involves four complement components that interact in a series of reactions to produce the classical pathway, they are in order of reaction sequence: C1, C4, C2, and C3. C1 is a calcium dependent multimeric complex with a molecular mass of 750 kD consisting of a single C1q molecule and associated serine protease tetramer, C1s-C1r-C1r-C1s. C1q has 18 polypeptide chains of three distinct types organized in a complex that resembles a bouquet of flowers. C1q binds to the Fc portion of antibody (a single molecule of IgM or two or more IgG molecules) bound to antigen complexes. IgM is more efficient than IgG in binding to C1q since only IgM one may be enough for binding. After binding to antibody C1q undergoes a conformational change which leads to the activation of the C1s-C1r-C1r-C1s tetramer. Activated C1s is an active serine protease that first cleaves the α -chain of C4 generating C4a and C4b, and then cleaves C2, which is a single polypeptide chain of 117 kD, when it associates with C4b to generate C2a and C2b with the formation of C3 convertase (C4bC2a) of the classical pathway. C4 is a 200 kD protein composed of three peptide chains (α , β , and γ) linked together by disulfide bonds and containing a thiolester bond in its α chain.

C4b when activated binds covalently to a target surface via its thiolester and acts as an opsonin. The C4bC2a enzymatic complex when formed remains bound to the

activating surface. If C4b does not bind as an opsonin or associate to form the convertase it is further cleaved by factor I and inactivated (iC4b) (Fearon, et al., 1977). C4b2a (C3 convertase) can cleave C3 leading to the generation of C3a and C3b. C3 is a 190 kDa protein made up of a 115 kDa α chain and an 75kD β chain linked together by disulfide bonds. The smaller fragment cleaved from C3 is C3a which is a moderately potent anaphylatoxin. The larger fragment is the metastable protein C3b that has the transient ability to attach covalently via its thioester bond to a nearby substrate via hydroxyl (OH) or amide (NH₂) groups. This membrane-bound C3b serves three different purposes: as an opsonin to facilitate phagocytosis by C3b receptor-bearing host cells, to associate with C4b2b to form the C5 convertase of CCP (C4b2aC3b), or to associate with factor B and trigger the alternative pathway of complement activation.

The C5 convertase (C4bC2bC3b) cleaves C5 which is a 190 kDa disulfide-linked heterodimer with homology to C3 and C4 but lacking an internal thioester bond in its α chain. Cleavage of C5 forms the potent anaphylatoxin, C5a, and the larger fragment C5b that remains close to the convertase on the microbial surface, thus initiating the assembly of the membrane attack complex (MAC) (Lambris, 1988).

The Lectin Pathway

As stated earlier, activation of this pathway occurs by the binding of mannans to the mannose-binding lectin (MBL) associated with two serine proteases known as the MBL-associated serum proteases (MASP). The MASP cleaves C4 and/or C3 to form C3 convertase. There are three isotypes of MASP: MASP-1, MASP-2, and MASP-3. MASP-1 cleaves C3 but not C4, whereas, MASP-2 cleaves C4 (Fishelson, 1991). When C3 is cleaved by the MASP it leads to amplification via the alternative pathway.

The Alternative Pathway

Initiation of the alternative pathway is unlike the previous two pathways described. Four plasma proteins are directly involved in the alternative pathway: C3, factors B and D, and properdin. The alternative pathway is also not antibody dependent, yet the presence of antibody aggregates can enhance its activation (Fishelson, 1991). This pathway can be directly activated by microorganisms or complex polymers, or indirectly via the Lectin Pathway described above. The alternative pathway is usually at a low level of constant activation since in serum the thioester bond of native C3 can be hydrolysed by water forming C3(H₂O) which structurally resembles C3b but is lacking an active attachment site. C3(H₂O) in the presence of appropriate activators can bind to Factor B in the fluid phase. Factor B is a single-chain, 93 kDa serine protease precursor. The binding of factor B to C3(H₂O) or C3b allows Factor D, a 25 kDa plasma serine protease, to cleave factor B at the single arginine-lysine bond releasing the small fragment Ba and leaving the larger Bb fragment attached to C3(H₂O) forming the initial fluid phase C3 convertase of the alternative pathway. This, in turn, cleaves additional molecules of C3, producing C3b that associates with factor B which is further cleaved to form the C3 convertase of the alternative pathway (C3bBb). C3bBb is labile but can be stabilized by the binding of properdin to the complex. Properdin is a 220 kD stabilizing tetrameric protein that extends the half-life of the convertase six to 10-fold by the membrane bound C3 convertase, C3bBb. C3bBb associates with a second molecule of C3b to form the C5 convertase (C3bBb3b) of the alternative pathway. This, in turn, cleaves C5, thereby initiating the terminal lytic pathway and the formation of the MAC. As an opsonin, C3b binds to activation surfaces via ester or amide linkages through the

thiolester bond. In the absence of such a reaction, C3b is inactivated by factor I (a serine protease) cleavage in the presence of regulatory molecules such as complement receptor 1 (CR1), factor H, membrane cofactor protein (MCP), and decay accelerating factor (DAF).

The Terminal Lytic Pathway

All three pathways lead to the terminal effector lytic pathway which involves the formation of the Membrane Attack Complex (MAC) (Green, et al., 1959; Humphrey, et al., 1969; Gotze, et al., 1970; Pangburn, et al., 1980; Bhakdi, 1983; Muller-Eberhard, 1986); The assembly of MAC involves activation of C6 through C9. Activation of these components, however, does not involve proteolytic cleavage of the proteins but rather conformational change, association and polymerization. C5b bound to target surfaces binds to C6, a 128 kDa single-chain protein which further binds C7, a 121 kDa single-chain protein (Lachmann, 1970; Arroyave, et al., 1973; Muller-Eberhard, 1975). C7 gives lipophilicity and membrane attachment capabilities to the complex formation. The C5b67 complex acts as a high-affinity membrane receptor for a C8 molecule (a 155 kDa trimer). The final stage of assembly consists of polymerization of C9, a 79 kD monomeric protein, that binds to the C5b-8 complex. Polymerization and insertion of C9 molecules forms transmembrane channels (holes) with subsequent lysis of target (Gotze, et al. 1970; Lachmann, et al., 1970; Kolbet, et al., 1972).

The different mechanisms by which the complement system can be activated has allowed the complement system to efficiently handle the invasion of pathogens and to mount an effective inflammatory response. The key step in complement activation is the formation of C3 and C5 convertases which cleave C3 and C5 at identical cut sites on the

α chain to generate fragments C3a and C3b, C5a and C5b, respectively. The end result of complement activation is the lysis and removal of pathogens or target cells. Components such as C3a, C4a, and C5a act as peptide mediators of inflammation and phagocyte recruitment (chemotaxis). These three components are small fragments known as anaphylatoxins. Anaphylatoxins induce a systemic inflammatory response that includes increase of vascular permeability, smooth muscle contraction, and chemotactic migration of inflammatory cells to the site of antigen activator deposit. Additionally, anaphylatoxins bind to receptors on the surfaces of mast cells and basophils to elicit the release of vasoactive compounds such as histamine and serotonin. Activation of complement also modulates the adaptive immune response via the activity of these active fragments generated from C3, C4, and C5.

The Biological Role of C3

Mammalian C3 has been thoroughly studied and much is known about its structure and biological function. In humans, C3 is the most abundant complement protein in serum (1.2mg/mL). It is the most versatile and multifunctional molecule of the complement system. More than 20 soluble and membrane bound proteins interact with C3 (Lamris, 1998). The C3 molecule is a glycoprotein composed of a 115 kDa α chain and a 75 kDa β chain, held together by a single disulfide bond (figure 2b). The 3D structure consists of two domains, a flat ellipsoid domain associated with a flat domain (Whitehead, et al., 1982).

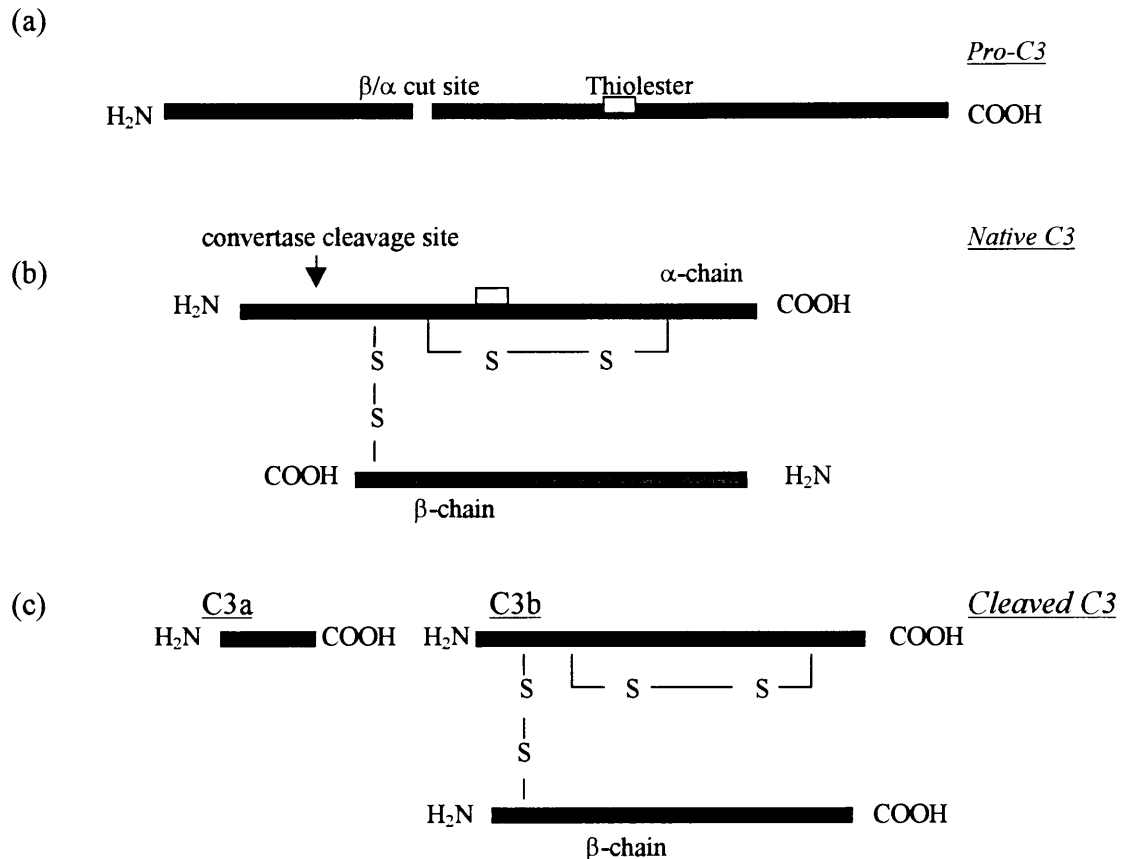


Figure 2. Molecular structure of pro-C3, native C3, and C3b. (a) Pro-C3 protein shown as a single chain with a β/α cut-site and thiolester site. (b) Native C3 following post translational modification of the single chain pro-C3 into the two-chain C3 structure composed of the α and β chains held together by an interchain disulfide bond. The site of activation of cleavage by C3 convertase is indicated by an arrow. (c) C3 cleavage products, C3a and C3b.

The activation of native C3 occurs by C3 convertase cleavage between the 77 and 78 (Arg-Ser) amino acid residues of the α chain. The released N-terminal fragment (C3a) is a small 77 amino acid residue peptide which is an anaphylatoxin (figure 2c). C3a consists of five carboxy-terminal amino acids that elicit the diverse biological effects on target cells. The cellular interaction is believed to be via a specific C3a receptor as not all cells respond to C3a. The large C3b fragment, made up of the α chain and the complete β chain, is also biologically active (figure 2c). It interacts with other complement components to produce enzymatic complexes, and is a major opsonic molecule that facilitates phagocytosis of target cells. C3b interacts with different cell receptors such as complement receptor type 1 (CR1), and can be further degraded by various control proteins (factor I, factor H) which regulate its activity.

Inactivation of C3b is brought about by the enzymatic activity of factor I in the presence of cofactors such as factor H in the ACP or C4b in the CCP. Factor I cleavage yields the inactive form, iC3b, by the excision of a 2 kDa peptide (C3f) between residues 1281-1282 (Arg-Ser) and 1298-1299 (Arg-Ser) (Lambris, 1998). A second factor I cleavage occurs at residues 932 and 933 (Arg-Glu) resulting in the release of C3dg while C3c remains bound to the target surface. C3dg can be degraded further with trypsin, elastase, or plasmin producing C3g (for review Lambris, et al., 1986). The various activation fragments of C3, i.e. C3b, iC3b, C3c, C3dg, C3c and C3g are all capable of further interaction with a variety of soluble and cell bound molecules that modulate the immune system (for review Fishelson, 1991; Lambris, 1988). C3a through binding to the C3a receptor (C3aR) can induce several cellular responses in cells expressing the

receptor, such as mast cells, neutrophils, monocytes, basophils, and macrophages (for review Lambris, 1988).

C3 Binding Sites

C3b has several different binding sites: the thioester site is present in the α chain between amino acid residues 1010 and 1013 (Cys and Gln). In native C3, the thioester is inaccessible and buried inside the C3 molecule. Upon cleavage by C3 convertase and the release of C3a, a conformational change occurs that exposes the reactive thioester bond which can react with the amine ($-\text{NH}_2$) and hydroxyl ($-\text{OH}$) groups present on proteins and carbohydrates of targets, or alternatively the thioester bond can be hydrolysed with water (Wetsel, et al., 1984; Fishelson, 1991, Dodds, et al., 1996). If a transacylation reaction occurs, the thioester bond covalently attaches the molecule to the surface of targets via amide or ester groups (Fishelson, 1991). When C3b binds to an activator surface, three things may occur: 1) surface-bound C3b behaves like an opsonin and mediates the phagocytosis of the target by polymorphonuclear leukocytes, monocytes, and macrophages bearing Complement Receptor Type 1 (CR1) receptors; 2) B-cells bearing CR1 receptors are activated; or 3) the covalently attached C3b forms the C5 convertase of either pathway (Sunyer, et al., 1997).

C3b has multiple binding sites for C5, properdin, factor H, factor B, factor I, C4 binding protein (C4bp), CR1, and the membrane cofactor protein (MCP). C3b can also interact with laminin, fibronectin, Herpes Simplex Virus gpC, Epstein-Barr Virus (EBV), and IL-2. The binding site for C5 has been localized to the β chain of C3b and it is the only ligand found to bind to the β chain of C3. Properdin binds to the C3bBb convertase and acts to stabilize the complex and to regulate the alternative pathway (Fearon, et al.,

1977; Reid, et al., 1986). The properdin binding site lies between residues 1402-1435 of C3 (Lambris, et al., 1986; Lambris, 1988; Fishelson, 1991).

Factor H competes with factor B binding for a site on C3b, and serves as a cofactor for factor I mediated cleavage of C3b (Fearon, 1977; Reid, 1986; Lambris, et al., 1986; Lambris, 1988; Fishelson, 1991). The C3 molecule has several binding sites for Factor H. One binding site on the α chain for factor H has been localized to the amino acid region between 745-754 which overlaps the binding region for CR1 (residues 749-790) (Fishelson, 1991). Two other sites for factor H binding can be found within the C3c region of C3b between 1209-1271 and one site has been identified as being discontinuous within the C3d region of residues 1187-1249 (Lambris, et al., 1986).

C3b binds factor B at a site between 749-790. The initial binding of factor B to C3(H₂O) occurs within the Ba region of factor B allowing it to be cleaved. Bb remains attached to C3(H₂O) and Ba is released. The C3(H₂O)Bb complex is the initial C3 convertase in the alternative pathway when additional molecules of C3 are cleaved. Factor B binds to C3b in the presence of magnesium ions to form the ACP C3 convertase after it is cleaved by factor D. The convertase is stabilized by properdin (Lambris, et al., 1986; Fishelson, 1991). It is believed that C3b may have two sites for binding factor B. One site in C3b has been located to be within amino acid residues 730-739 (Lambris, et al., 1986) and it has also been postulated that the Bb region of factor B may be within the C3d region (residues 933-942) (Lambris, et al., 1986). Although, the binding site for factor I has not been identified it appears that it can associate with C3b-H and thus close to binding sites for Factor H, CR1 and CR2 (Lambris, et al., 1986; Fishelson, 1991).

C3 Genetics

C3 is a member of a group of thiolester proteins which include C4, C5, and a noncomplement protein, α -2 macroglobulin (α 2MAC) (Swenson, et al., 1979; Howard, et al., 1980; Tacket, 1980; Campbell, et al., 1981). These components are believed to have arisen from an ancestral primordial gene by gene duplication (Sim, et al., 1981; Starkey, 1981; Sottrup-Jensen, et al., 1985; Nonaka, et al., 1992; Nonaka, et al., 1998). While C3, C4, and C5 share sequence homology they differ in their subunit structure. C3 and C5 are 2 chain molecules while C4 is a 3 chain molecule. C3 resembles C5 in subunit structure but the latter lacks the thiolester bond. All (including α -2 macroglobulin), however, arise from single genes and the single transcript yields the pro-protein which is post-translationally modified (Brade, et al., 1977; Hall and Colten, 1977; Swenson, 1980).

While the loci of complement components factor B, C2 and C4 are closely linked to the HLA-B locus on chromosome 6, the human C3 gene has been mapped outside the HLA-B region on chromosome 6 and has been mapped to chromosome 19. Family studies show that C3 is linked to the Lewis blood group that happens to be linked to myotonic dystrophy present on chromosome 19. Allelic variants of C3 do exist, but to date the number present in the human genome is unknown (Whitehead, et al., 1982).

The C3 gene contains 41 exons that range in size from 52 to 213 base pairs. Introns, which are spliced out during mRNA processing, range in length from 82 to 4400 base pairs. Human C3 gene intron sequences have the 5' and 3' splice donor and acceptor sequences followed by the usual GT/AG pattern. The transcription initiation sites are localized to two positions at 61 and 63 residues 5' from the start codon ATG.

TATA and CAAT boxes have been identified at –28 and –87 base positions respectively. The stop codon, TGA, is found at position 5050 and the poly-(A) tail at position 5077 (Vik, et al., 1991). Genomic DNA encoding human C3 produces a mRNA transcript molecule that codes for 1663 amino acid residues, prepro-C3 (Vik, et al., 1991).

C3 of Nonmammalian Species

C3-like molecules have been identified in the different species. C3 homologues or C3-like proteins have been reported in invertebrate species (Smith, et al., 1996; Al-Sharif, et al., 1998) and vertebrates in which complement or complement-like activity has been demonstrated (Sunyer, et al., 1980; Nonaka, 1981). In these organisms, C3 gene arrangement or protein structure is not identical, however, each possesses highly conserved sequences within the genes. One such conserved sequence is the thiolester bond sequence. Interestingly, the thioester bond is also present in the protease inhibitor α -2 macroglobulin (α 2Mac), which is not considered a complement related protein (Wetsel, et al., 1984; Nonaka, et al., 1992).

C3 related molecules have been described in invertebrate species. The first C3-like protein to be identified in an invertebrate was a molecule designated SpC3 expressed in the purple sea urchin, *Strongylocentrotus purpuratus* (Smith, et al., 1996; Al-Sharif, et al., 1998). Until recently, this was thought to be the most ancient invertebrate C3-like molecule, however recent studies have identified a C3 sequence in the coral (Dishaw, unpublished data).

C3 homologous proteins have been described for several vertebrate phyla, including two agnathan cyclostomes considered to be the most primitive vertebrates, the hagfish and lamprey. The hagfish (*Eptatretus burgeri*) C3 is a thioester-containing

protein that behaves like an opsonin, but interestingly has a three-chain structure after proteolytic cleavage of an irregular two-chain C3 structure made up of an $\alpha + \gamma$ chain and a β chain (Ishiguro, et al., 1992; Hanley, et al., 1992; Fujii, et al., 1995). Homology to human C3, and C4 was 30.5% and 27.9% respectively. This may imply that the hagfish C3 may have similar properties to both C3 and C4 of humans, but more so of C3 (Ishiguro, et al., 1992). Lamprey, *Lamprera japonica*, C3 is similar in subunit structure to mammalian C4 as it is a three-chain structure (containing α , β , and γ chains), yet its homology is highest with C3. Lamprey C3 does not mediate target cell lysis, but acts as a phagocytic factor. It has been proposed by Nonaka (1992) that the lamprey C3 may be the common ancestral molecule of C3, C4, and C5 in that C4 diverged from a three-chain C3, C3 lost the α - γ junction region becoming a two-subunit structure, and C5 diverged from a two-chain C3 and lost the thiolester bond (Nonaka, et al., 1984; Nonaka, et al., 1992)

Several isoforms of C3 have been reported for certain species of teleost (bony) fish (Nonaka and Yamaguchi, et al., 1981; Nonaka and Sakai, et al., 1981; Lambris, et al., 1993; Sunyer, et al., 1997; Nakao, et al., 2000; Zarkadis, et al., 2001). Four isoforms of C3 have been found in the rainbow trout and designated as C3-1, C3-2, C3-3, and C3-4, of which only three (C3-1, C3-3, and C3-4) are functional (Nonaka, et al., 1981; Lambris, et al., 1993; Sunyer, et al., 1996). These four isoforms differ in amino acid sequence identity and in their binding to complement activators and may be encoded by two different C3 genes (Lambris, et al., 1993; Nonaka, et al., 1998; Sunyer, et al., 1996; Zarkadis, et al., 2001). The common carp (*Cyprinus carp*) has been found to have five types of C3 known as C3-S, C3-Q1, C3-Q2, C3-H1, and C3-H2 (Nakao, et al., 2000) Of

these five types, C3-S, C3-Q1, and C3-Q2 are missing the catalytic histidine which is conserved in C3 of most animals analyzed to date and provides the thiolester with the ability to bind covalently to hydroxyl groups on target cells or to be hydrolyzed quickly. Although, these variants in the carp were missing the catalytic Histidine, some demonstrated high hemolytic activity suggesting that the thiolester's catalytic mechanism is not a determinant of C3 activity and molecules lacking the catalytic histidine play a functional role in the carp (Nakao, et al., 2000). The trout and the carp are tetraploids, which might explain the presence of different isoforms (Sunyer, 1996; Nakao, et al., 2000; Zarkadis, et al., 2001). It has been suggested that the C3 isoforms in trout and the carp increase the number of possible pathogens that can be recognized and thus enhance the protective effect of complement.

Another telost is the gilthead sea bream, *Sparus aurata*, which is a diploid and was also found to contain five isoforms of C3 (Sunyer, et al., 1997; Sunyer, et al., 1997). The sea bream C3s are believed to have evolved from the trout C3s and it is likely that one or more of the C3s may have duplicated, thus producing the five isoforms. The five isoforms in the sea bream have remained functional (Sunyer, et al., 1980; Sunyer, et al., 1997). These five isoforms in the seabream also increase the performance of their immune system since their specific immune response to pathogens based only on immunoglobulin M is a later response and not immediate (Sunyer, et al., 1997).

The reptile complement system has been studied in the cobra, *Naja naja kaouthia*. The C3 gene in the cobra encodes for a two-chain protein with an internal thioester and high homology with C3 molecules of other vertebrates (Fritzinger, et al., 1992). Three additional C3-related genes have been found in the cobra. The product of one of these

genes was found in the venom of the cobra and has been named Cobra Venom Factor (CVF) (Fritzing, et al., 1994). CVF is a complement activating protein and is functionally related to C3b in that it is able to associate with factor B and form a C3bBb-like convertase in the presence of factor B and D. The molecular structure of CVF is different from its complement component C3 in that it is composed of three chains and does not have a thioester site in the α chain. Two other mRNA transcripts have been identified in the cobra and found to be 90% homologous to cobra C3, as was CVF (Fritzing, et al., 1992; Sunyer, et al., 1998).

In aves, C3 has been identified in the chicken and Japanese quail (Kai, et al., 1983; Mavrodis, et al., 1995). Chicken C3 has 54% identity to human and has a two-chain structure similar to humans (Mavrodis, et al., 1995). The Japanese quail C3 was also found to resemble human C3 and to function as complement component C3 (Kai, et al., 1983). The murine C3 has also been identified and its β and α chain structures have high identity to human C3 β and α chains, 73% and 75%, respectively.

C3 in the Nurse Shark

C3 molecules have been described for the elasmobranch, *Ginglymostoma cirratum*. The N-terminal amino acid sequence of nurse shark C3 α and β chains display 28% and 50% homology with corresponding chains of human C3 (Dodds, et al., 1998). Previous studies in the nurse shark yielded two related, but distinct partial C3 clones, C3A4 (figure 3) and C3A7, which suggested that there may be more than one isoform of C3 in the nurse shark although only one functional C3 protein was isolated from shark serum (Dodds & Smith, 1998; Nonaka & Smith,). The two C3-like cDNA clones (Nakao & Smith, unpublished data) were used to design gene specific primers for 3' and 5' Rapid

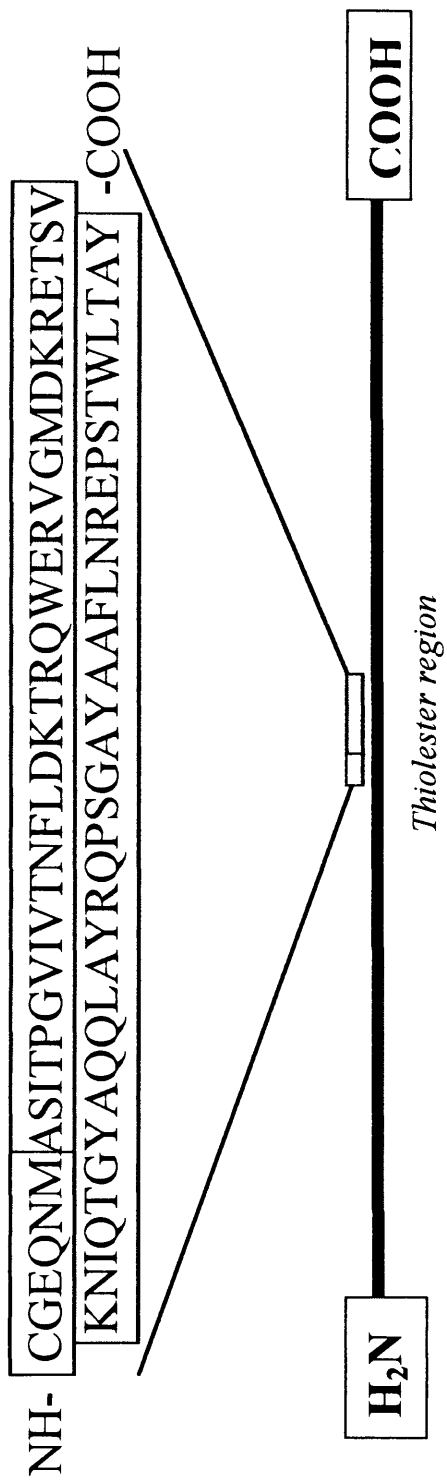


Figure 3. Deduced amino acid sequence of the thiolester region amplified by RT-PCR. Amplification yielded a 220bp product (C3A4) which was used as the DIG-labeled probe to screen the cDNA library.

Amplification of cDNA Ends (RACE) PCR. Several RACE products ranging in size from 800bp to 1800bp were generated. Deduced amino acid sequence generated from a 1800bp (NsC31) 3' RACE product clone showed the absence of the conserved catalytic histidine residue at position 1126 (Lee, 1999). As this residue is essential for optimal function of the thiolester, it is possible that the clone, represents sequence of a nonfunctional C3 protein since catalytic histidine is believed to be essential for thiolester binding to hydroxyl groups. This 3' RACE product was generated from gene specific primers designed from the sequence of clone C3A7. This observation suggests that as the two partial clones, C3A4 and C3A7, represent two genes encoding two different isoforms of C3 in the shark, and C3A7 represents a C3 gene which encodes a protein missing the critical histidine, then C3A4 must code for the functional protein that has been isolated from shark serum. To confirm this hypothesis the construction and screening of a cDNA library for C3 clones, using C3A4 as a probe, will provide additional gene sequence data that can determine the number of C3 isoforms in the nurse shark. Analysis of amino acid sequence data will also provide information on the biological significance of a C3-like molecule which lacks a catalytic histidine residue.

III. Materials and Methods

RNA Isolation from Nurse Shark Liver

Approximately 2.5g of liver from a nurse shark, *Ginglymostoma cirratum*, was added to 25ml of TRIzol (GibcoBRL Life Technologies). The sample was homogenized for 1 minute using a homogenizer. The homogenized sample was placed on ice and 0.2 ml of chloroform per 1ml of Trizol was added to the mixture, which was shaken vigorously for 45 seconds and incubated for 3 minutes at 25°C. The mixture was centrifuged at 3220xg for 40 minutes at 5°C, and the aqueous phase was withdrawn and transferred to a fresh tube. The RNA was precipitated from the aqueous mixture by adding isopropanol (0.5ml of Isopropyl alcohol per 1mL of Trizol). Sample was mixed gently, incubated at 25°C for 10 minutes and centrifuged at 3220xg for 40 minutes at 5°C. The RNA formed a gel like pellet after centrifugation. The supernatant was removed and the pellet washed once with 75% ethanol. Sample was mixed by vortexing for 5 seconds and centrifuged 3220xg for 20 minutes at 5°C. The supernatant was discarded and the RNA pellet dried for 10 minutes. The dried RNA pellet was dissolved in 250 µL of DEPC H₂O. A 500 fold dilution was made of the RNA sample and the concentration and purity of the RNA sample was determined spectrophotometrically by determining the OD₂₆₀/OD₂₈₀ ratio.

Electrophoretic Analysis of RNA

A 1% agarose gel in tris-borate/EDTA (TBE) was prepared by heating and dissolving 0.5 g of agarose in 50 mL of 1x TBE. The agarose was allowed to cool before adding ethidium bromide (10mg/ml stock solution) at a concentration of 0.5ug/ml. It was mixed

thoroughly, and poured into the electrophoresis chamber containing the comb in place and allowed to set for 45 minutes. Four 10 fold dilutions of the RNA sample were mixed with 6x Gel Loading Buffer in a final volume of 18 μ l. The agarose was immersed in 1x TBE and 12 μ l of sample RNA was loaded per well. An RNA standard was also run which provides visible banding at the 28S and 18S rRNA with a smear ranging from 400bp to 8kb. Electrophoresis was carried out at a constant voltage (80 volts) and terminated when the sample migrated passed the midpoint of the gel. The gel was stained with ethidium bromide and visualized over an ultra violet lamp.

mRNA Isolation

The QuickPrep mRNA purification Kit (Amersham Pharmacia) was used to isolate mRNA. Two samples of mRNA were prepared using 40 μ l and 80 μ l of total nurse shark liver RNA (12 μ g/ μ l concentration) and labeled A and B, respectively. The RNA was mixed with 1.5 ml of Extraction Buffer in a 15ml tube and 3mL of Elution Buffer was added to the extract and mixed thoroughly. Afterwards, the Elution Buffer was warmed and maintained at 65°C. An Oligo(dT)-Cellulose Spin Column was prepared (one column per sample) and inverted several times to resuspend the contents. The top and bottom closures of the column were removed and the column placed in a 15 mL centrifuge tube. Together they were centrifuged at 350xg for 2 minutes. The column was removed from the tube and the liquid remaining in the tube was discarded. The top and bottom closures of the spin column were replaced securely before placing the column upright in a rack. A portion (4mL) of RNA sample mixture was loaded onto the surface of the resin of the spin column which was then inverted for 10 minutes and centrifuged at 350xg for 2 minutes. The top closure was removed and the supernatant was discarded.

Three milliliters (3mL) of High Salt Buffer was applied to the top of the resin and after replacing the top the matrix was resuspended by tapping on the tube. The column was returned to the 15ml tube and together centrifuged at 350xg for 2 minutes. The supernatant decanted and the High Salt Buffer wash was repeated twice and after the last wash 3 mL of the Low Salt Buffer was applied to the column and the matrix resuspended. The column was centrifuged as before and the supernatant decanted and discarded. After removing the bottom closure of the column 3 mL of Low Salt Buffer was applied and the column centrifuged at 350xg for 2 minutes. The spin column was removed and placed into a fresh 15 mL tube and the bound poly(A)⁺-RNA was eluted with 250 µL of Elution Buffer prewarmed at 65°C. The column was centrifuged at 350xg for 2 minutes and the eluted poly(A)⁺-RNA was collected and transferred to a microcentrifuge tube. Elution with 250 µL of Elution Buffer was repeated two more times. At the end of the elution step approximately 750 µL of poly(A)⁺-RNA had been isolated. The OD_{260nm/280nm} readings of the poly(A)⁺-RNA preparation was measured to confirm the presence of RNA.

cDNA Synthesis

First Strand Synthesis

The cDNASynthesisKit (Stratagene) was used to synthesize cDNA. Prior to setting up the reaction all nonenzymatic first-strand components were thawed by placing in a 37°C H₂O bath. When completely thawed (approximately 40 minutes) each reagent tube was briefly vortexed and the contents spun down before preparing a reaction mixture of 50 µL (final volume) which contained 12.5 µL of reagents and enzymes and 37.5 µL

of mRNA and DEPC water. A control was set up with a final volume 37.5 μL that contained 25 μL of test RNA and 12.5 μL of DEPC water. In a microcentrifuge tube on ice the following reagents/ sample were added and mixed in the order listed:

10x 1 st strand buffer	5 μL
1 st strand methylnucleotide mix	3 μL
linker primer (oligo dT)	2 μL
RNAse Block Inhibitor	1 μL
mRNA	5 μg
DEPC H ₂ O	volume adjusted to obtain a total volume of 50 μL in the reaction mix

All sample reaction mixtures were mixed and spun down to collect contents. Equal volumes (0.75 μL) of MMLV-Reverse Transcriptase (50U/ μL) and Powerscript Reverse Transcriptase (100U/ μL) were added to each reaction mixture and incubated for 30 minutes at 37°C and then placed on ice.

Second Strand Synthesis

The following reaction mixture was prepared:

10x 2 nd Strand Buffer	20 μL
2 nd Strand dNTP mix	6 μL
H ₂ O	111 μL
RNAse H	2 μL
Polymerase I	11 μL
Template (first-strand synthesis as above)	50 μL

The contents were thoroughly mixed and incubated for 2.5 hours at 16°C. The temperature was maintained by preparing an ice bath in a styrofoam box that was kept constant at 16°C. The reaction mixtures were placed on ice after incubation and 23 µL of Blunting dNTP mix was added to each, followed by 2 µL Pfu DNA Polymerase. The reaction tubes were incubated for 30 seconds at 72°C and placed on ice. Only five (5) µL of each sample was electrophoresed on a 1% TAE gel and during electrophoresis 200 µL of PCI was added to the remaining contents in the sample reaction tubes and vortexed. The samples were then centrifuged at 21,000 x g for 2 minutes at 25°C. The aqueous phase was saved and 25 µL of 3M sodium acetate and 500 µL of ethanol was added to it. Precipitation of cDNA was allowed to occur overnight at -20°C. Precipitated samples were centrifuged at 15,000 rpm for one hour at 4°C. The supernatant was decanted and the pellet washed with 70% ethanol, centrifuged at 21,000 x g for 2 minutes at 4°C, and after decanting and discarding the supernatant, the pellet was dried at 37°C for 10 minutes. The pellet contained the cDNAs synthesized after second-strand synthesis.

Construction of cDNA Library

Ligating EcoRI Adapters

The cDNA pellet was resuspended in 8 µL of EcoRI Adapters and incubated for 30 minutes at 4°C. The following components were added to the cDNA tube containing the EcoRI Adapters and mixed:

10x Ligase Buffer	1µL
10mM nATP	1µL
T4 DNA Ligase (4U/µL)	1µL

The reaction tube was incubated overnight at 8°C in a cooling block. The ligase in the reaction mixture was inactivated by placing the sample in a 70°C water bath for 30 minutes. The components were spun down and the reaction cooled for 5 minutes at 25°C. The Adapter Ends were phosphorylated by adding the following reagents mixed:

10x Ligase Buffer	1μL
10mM rATP	2μL
Sterile H ₂ O	6μL
T4 Polynucleotide Kinase (10U/μL)	1μL

The reaction mixture was incubated for 30 minutes at 37°C. The kinase was heat inactivated for 30 minutes at 70°C and components centrifuged and equilibrated to room temperature for 5 minutes before digesting with XhoI by adding 28μL of XhoI Buffer and 3μL of XhoI (40U/μL). The reaction mixture was incubated for 1.5 hours at 37°C. The sample was precipitated by the addition of 5μL of 10x STE and 125μL of 100% ethanol (w/v). Following precipitation the sample was centrifuged at maximum speed, 15,000 rpm for 15 minutes at 4°C. The supernatant was discarded and the pellet dried in a 37°C incubator. The dried pellet was resuspended in 14μL of 1x STE Buffer. Loading dye (3.5μL) was added to the sample to determine size fractionation of digested sample by chromatographic separation.

In order to proceed with size fractionation a drip column was assembled. A 3mL syringe and a 21 ½ gauge needle per sample was used. A small piece of sterile cotton from a disposable glass pipet was cut in half so that it fit inside the syringe. The cotton was tightly packed into the syringe. The Sepharose CL-2B Gel Filtration Medium and

the 10x STE Buffer were equilibrated to room temperature. The 10x Stock of STE Buffer was diluted in sterile H₂O to prepare 50 mL of 1x STE Buffer. The syringe with the cotton plug was placed upright on a clamp. With a Pasteur pipette 2 mL of the 1x STE Buffer was added to the cotton plug. The STE was allowed to drip through the cotton plug and the flow through was disposed. Immediately, 2mL of the suspension of Sepharose CL-2B was added to the syringe by inserting the tip of the Pasteur pipette as far into the column as possible. When the resin settled, the Sepharose was continuously added until it was packed tightly and its height measured was equivalent to 1.2 mL in the syringe. The drip column was washed three times with 1.2 mL of STE buffer to ensure that the buffer flowed at a steady rate. The column was not permitted to dry out. When 50 μ L of STE buffer remained above the surface of the resin, the cDNA sample was immediately loaded using an automatic pipettor. The sample was gently released onto the surface of the column bed without disturbing the resin as it may affect the cDNA separation. Once the sample had entered the sepharose, buffer was applied to the column. Column fractions were collected in 20 microcentrifuge tubes until the dye front reached close to the bottom of the syringe. Approximately, 400 μ L (equivalent to 6 drops) of column eluate was collected per fraction. Collection of fractions was stopped when drops from the column became clear, that is, drops did not contain traces of loading dye. After fractionation, the cDNA in fractions was semi-quantitated by comparing its UV intensity to a diluted DNA marker (GibcoBRL) of known concentrations. This was accomplished by preparing six different dilutions of the marker: an initial 1/5 dilution in H₂O and then two fold (1/2) serial dilutions thereafter of the initial 1/5 diluted marker. An estimation of cDNA concentration was obtained as follows. First a sheet of saran

wrap was placed on the surface of the UV light illuminator, then a row of 2 μL drops of ethidium bromide (2 $\mu\text{g}/\mu\text{L}$) was placed on the saran wrap. To each ethidium bromide drop 2 μL of a single dilution was added and mixed, similarly each cDNA fraction was mixed with a 2 μL drop of ethidium bromide. A photograph was taken of the drop mix to determine whether and which fractions contained cDNA by observing the fluorescent intensity of the ethidium bromide in the size fractionated cDNA samples. The cDNA concentration of the samples was estimated by comparing fluorescence of sample to that of known marker dilutions. Visualization of the samples in this manner allowed selection of sample cDNA containing fractions that could be pooled. Also, the brighter the ethidium bromide fluorescence, the larger or more cDNA fragments were present in the fraction.

Agarose Gel Electrophoresis of Fractionated Samples

All fractions were electrophoresed and analyzed on a 1x TBE, 1% TAE agarose gel as follows: 5 μL of each sample fraction was mixed with 1 μL of 6x Gel Loading Buffer and added to each well. Electrophoresis was carried out for 45 minutes at 80 Volts and the ethidium bromide stained gel was photographed. From comparison of the fluorescence intensity of each sample fraction we could determine the presence of cDNA smears and identify the sample fractions which could be pooled together, as brighter fluorescence signaled the presence of more cDNAs. Samples selected to be pooled together were all combined into one microcentrifuge tube and approximate measurements of DNA concentrations were made using the saran wrap over the UV illuminator technique described above. To visualize cDNA concentrations, 1 μL of

ethidium bromide (2 $\mu\text{g/mL}$) was mixed with 1 μL of the pooled samples and photographed. cDNA concentrations were estimated by comparing the fluorescent intensity of the known marker concentrations to the intensity of that of the samples.

Revival of Stock Bacterial Cells from Library Construction Kit (Stratagene)

Splinters were scraped off from a frozen XL-1 Blue MRF' *Escherichia coli* glycerol stock with a wire loop and the bacteria placed into small vials. The splinters were suspended in a drop of sterile water and streaked onto an Luria's Broth (LB) agar plate. Plates were incubated at 37°C overnight. Following incubation plates were stored at 4°C and cells were subcultured (restreaked) periodically until used as host cells for cDNA phage library infection.

Ligating cDNA into the Uni-ZAP XR Vector

A control ligation was set up to ligate the test insert, pBR322, into the Uni-ZAP XR vector as follows. The components listed below were added in the order listed:

Uni-ZAP XR vector (1 μg)	1 μL
Test insert (0.4 μg)	1.6 μL
10x Ligase Buffer	0.5 μL
10mM rATP (pH 7.5)	0.5 μL
Water	0.9 μL
T4 DNA Ligase (4U/ μL)	0.5 μL (added at the end)

For the cDNA sample the following components were mixed:

Resuspended cDNA	2.5 μL
10x Ligase Buffer	0.5 μL

10mM rATP (pH 7.5)	0.5μl
Uni-ZAP XR vector (1μg/μl)	1.0μL
Water	0.5μL
T4 DNA Ligase (4U/μL)	1.0μL

All samples were incubated at 12°C overnight. In the interim, host bacteria were prepared by inoculating 6 mL of sterile NZY medium with a single isolated colony taken from XL-1 Blue Bacteria colonies streaked out and grown on LB agar the night prior to the start of the protocol. Cultures were grown at 30°C overnight with shaking.

Packaging ligated cDNA into Bacteriophage Extracts

Two packaging extracts were performed per sample, but one sample was processed at a time. The packaging extracts were quickly thawed and immediately 2.5 μL of ligated DNA was added. Samples were named A/1, A/2, B/1, and B/2 to correspond to the two original RNA samples A and B and the number of packaging extracts, #1 and 2. With a pipette tip the mixture was stirred by gently pipetting up and down ensuring no bubbles formed. The tube was centrifuged quickly to settle contents and incubated for exactly 110 minutes at 22°C. The packaging process was stopped by the addition of 500 μL of SM buffer, followed by 20 μL of chloroform. The tube was gently mixed and centrifuged briefly before samples were ready to be titrated. Samples were stored at 4°C when not in use

Plating and Determination of Titre

The bacterial cells were cultured in NZY medium according to ZAP-cDNA

Synthesis Kit (Stratagene) and centrifuged at 500x g for 10 minutes at room temperature. In the interim NZY top agar was liquified and maintained at a temperature between 40-50°C. Following centrifugation the supernatant was discarded and cell pellet resuspended, in half the original volume of 10 mM MgSO₄. Cells were adjusted to a density equivalent of an O.D.₆₀₀ of 0.5 by diluting with 10 mM of MgSO₄. To 200 µL of bacterial suspension (reading of 0.5 at O.D.₆₀₀) 1µL of cDNA library suspension from a total volume of 500 µL from samples A/1, A/2, B/1, and B/2 was added, mixed and incubated for 15 minutes at 37°C. Following incubation 3mL of NZY top agar was added and quickly mixed with the phage and bacteria mixture in a 15 mL polypropylene tube. The tube was rolled between hands to thoroughly mix contents which were then added directly to the middle of an NZY agar plate which in the interim was warmed to 37°C and kept in a styrofoam box to ensure that the NZY top agar did not solify before completely spreading across the plate. The inoculum was spread quickly over the surface of the plate with a sterile glass spreader, and any bubbles were removed using a sterile needle. Plates were allowed to set for 10 minutes and then incubated overnight at 37°C. The next day the number of plaques per sample were counted and the titre of the cDNA library was determined in plaque forming units (pfu) employing the following equation:

$$[\text{number of plaques (pfu)} \times \text{dilution factor/volume plated (}\mu\text{L)}]$$

The dilution factor was 500 and the volume plated was 1 µL. The size of a good representational primary library consists of 1×10^6 plaque forming units.

Amplification of the Uni-ZAP XR Library

Bacterial host strains were prepared as described previously. Cells were diluted

to an O.D.₆₀₀ of 0.5 in a total volume of 10 mL to be used for each 150mm plate. cDNA library sample B/2 with a volume of 500 µL was chosen for amplification as its pfu titre was 1.15×10^6 , and since a cDNA library titre of 1×10^6 pfu is considered optimal. Twenty-two aliquots of 600 µL of the bacterial culture standardized to an O.D.₆₀₀ of 0.5 were set up. To each 600 µL aliquot of bacterial cells 1 µL of cDNA library sample was mixed. Tubes containing this mixture of phage and host cells were incubated at 37°C for 15 minutes. NZY top agar (6.5 µL) was liquified and cooled to 48°C and mixed with the phage and host cells and quickly spread evenly onto the surface of 150 mm diameter NZY agar plates as described previously. After the NZY top agar solidified the plates were incubated at 37°C for 6-8 hours. Plaques were not allowed to grow larger than 1-2 mm in diameter.

Lifting Plaques

Using flat forceps a nitrocellulose membrane was placed onto the surface of each NZY agar plate (22 plates total) containing the amplified library. The membranes were left on the agar surface for 2 minutes to allow the transfer of the phage particles to the membrane. In order to later identify the positive plaques on membranes and agar the orientation of the membranes was done by making reference points on each membrane as follows. Using a sterile needle three assymetrical pricks were made through the membrane, just touching the surface of the top agar. The membranes were peeled off using the flat forceps and placed plaque-side up onto 3 MM paper or paper towels. The original plates were stored at 4°C. The membranes were air dried for 30 minutes during which time the following solutions (500 mL of each) were prepared:

(i) Denaturing Solution (0.1 N NaOH-1.5 M NaCl)

5 M NaOH	10 mL
5 M NaCl	150 mL
Sterile Water	340 mL

(ii) Neutralizing Solution [0.2 M Tris-HCl (pH 7.4) - 2xSSC

2 M Tris-HCl	50 mL
Sterile Water	400 mL

(iii) 2xSSC

20xSSC	50 mL
Sterile Water	450 mL

The three solutions were added to three separate plastic containers with lids. The dried membranes were added one at a time to the denaturing solutions so each first floated on top of the solution's surface and then gently immersed into the solution for 20-30 seconds. The membranes were briefly blotted on a 3MM paper or a paper towel and submerged into the neutralizing solution for 5 minutes. The membranes were again briefly blotted on a 3 MM paper towel and rinsed for 5 minutes in 2xSSC. The membranes were placed plaque-side up on paper towels and allowed to dry for 30 minutes. The DNA was crosslinked using a UV'Crosslinker at $700,000 \mu\text{J}\cdot\text{cm}^{-2}$ for the Hybond-N membranes used.

Preparation of DIG-labeled C3 Probe

The C3 cDNA probe C3A4 (figure 3) used for hybridization was non-radioactively labeled by incorporation of a digoxigenin-labeled nucleotide. Digoxigenin is a steroid hapten linked via a spacer arm to the corresponding nucleotide. The DNA

probe was generated using Klenow polymerase that randomly primed incorporation of DIG-labeled deoxyuridine triphosphate. Oligonucleotides were enzymatically labeled at their 3'-end with terminal transferase either by incorporation of a single DIG-labeled dideoxyuridine-triphosphate. Preparation of a DIG-labeled C3 probe was carried out using previously isolated and sequenced cDNA plasmid C3A4 (figure 3). The reaction mixture contained the following:

10x PCR buffer	5 μ l
PCR DIG mix	5 μ l
M13F primer (10 μ M)	2.5 μ l
M13R primer (10 μ M)	2.5 μ l
template plasmid (1ng/ μ l)	2 μ l
Sterile Water	33 μ l
Enzyme mix	0.75 μ l

All contents were mixed thoroughly and placed in a thermocycler. The thermocycling parameters optimal for making the C3 probe were:

95°C, 2 minutes	
95°C, 10 seconds	x 30
50°C, 10 seconds	
72°C, 1 minute	
4°C, ∞	

When amplification was complete the probe was analyzed by agarose gel electrophoresis to confirm that the probe was labeled accurately.

DIG Nucleic Acid Hybridization and Detection Kit

The following solutions were prepared prior to DIG Nucleic Acid Hybridization:

Prehybridization solution

20x SSC	150 mL
10% Blocking Reagent	60 mL
Sterile Water	390 mL
N-lauroylsarcosine (powder)	0.6 g
SDS (powder)	0.12g

500 mL of the prehybridization solution was preheated at 65°C. Residual 100mL was reserved and used for dilution of the DIG-labeled C3A4 probe. For the secondary screening this 100 mL was not needed because the diluted probe solution was recovered and used after the primary screening and stored at 20°C. The plaque-lift membranes were added into the prehybridization solution and incubated at 65°C for 2 hours in a hybridization bag for the primary screening.

The DIG labeled C3A4 probe was heated at 95°C for 10 minutes, chilled quickly on ice, diluted 1/1000 in fresh prehybridization solution. After 2 hours, the prehybridization solution was removed from the membranes and membranes were submerged completely in probe solution (prehybridization buffer with DIG labeled C3A4 probe) inside hybridization bags. Before sealing the hybridization bags all air bubbles were removed. Membranes were hybridized with the probe at 65°C overnight (16 hours). During the overnight incubation the following washing solutions were prepared:

2x Wash solution (2x SSC- 0.1% SDS)

20x SSC 100 ml

Water 900 ml

SDS powder 1g

0.1x Wash Solution (0.1xSSC- 0.1% SDS)

20x SSC 5 ml

Water 995 ml

SDS Powder 1g

The next morning, the membranes are removed and washed twice in 500 mL of preheated (65°C) 2x Wash Solution and incubated in 2x Wash Solution for 5 minutes at room temperature (25°C) with shaking. The probe solution was recovered from the bags and stored at -20°C. After 5 minutes, the membranes were placed into 500ml of 0.1x Wash Solution (heated to 65°C) for 30 minutes. The membranes were washed after 30 minutes in 500 mL of DIG buffer-2 (0.1 M Maleic Acid-NaOH Buffer, 0.15 M NaCl, pH 7.5) at room temperature (25°C) for 5 minutes with shaking, then incubated in 300 ml of DIG Buffer-2 (1% Blocking reagent in DIG buffer-1) at room temperature for 30-60 minutes. After this incubation, they were placed in 200 ml of DIG buffer-2 containing 40µl of anti-DIG-alkaline phosphatase at room temperature for 30 minutes. After washing the membranes three times with 500 ml DIG buffer-1 at room temperature for 15 minutes with shaking, they were incubated in 200 ml of DIG buffer-3 (0.1M Tris-HCl, pH 9.5, 0.1 M NaCl, 0.05 M MgCl₂) at room temperature for 5 minutes. The DIG Buffer-3 solution

was prepared by mixing:

1 M Tris-HCl (pH 9.5)	20 ml
5 M NaCl	4 ml
Sterile Water	166 ml
1 M MgCl ₂ (autoclaved)	10 ml

The membranes were incubated in 200 ml of NBT/BICP substrate solution until positive plaques developed a blue-purple color. The color development was stopped by immersing the membranes in TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 8.0). After 30 minutes, the membranes were dried on paper towels and stored in TE buffer at 4°C.

Positive Plaque Collection

Membranes were placed on top of a transparency sheet and covered with another transparency sheet with the plaque-side of the membrane uppermost it was traced onto the top transparency. The position of the needle marks were marked on to the transparencies and the positive plaque positions traced. The labeled transparency sheet was placed traced side up on a bench and its corresponding NZY agar plate containing the plaques was placed on the labeled sheet and adjusted to correspond to the needle marks taken from the membranes to those made on the agarose plate. The plates were oriented according to the traced needle marks of the transparency that corresponded to the original needle marks made on the top agar.

A 2 mL pipetman tip was cut at the point where a 4mm diameter can be formed. Using the tip, positive plaques were removed by inserting the tip into the top and bottom agar. Suction was used to draw out the agarose plug. The agarose plug containing the plaque was added into 500 µl of SM buffer containing 20 µl of chloroform. Samples

were vortexed 20 seconds and incubate overnight at 4°C to extract phage particles. This phage stock is stable at 4°C for 6 months. The phage isolated from this 1st screening was used for the secondary screening.

Secondary Screening

A second screening was performed to obtain phage containing only the cDNA that gave a positive result in the first screening. Dilutions of the resuspended plaque were made. The primary phage stock isolated from the agarose plugs after primary screening were diluted 1/1000, 1/5000, and 1/10,000 using SM Buffer. From these dilutions 1 µl of the phage stock was added to 200 µl of bacterial cell suspension equivalent to an O.D.₆₀₀ of 0.5. The cell mixture was incubated and plated as described above for the primary screening, the only difference was that for this secondary screening 100x15 mm plates were used and only 3ml of NZY top agar was needed. Plates were incubated overnight. Hybridization and detection was performed as above. Identified positive plaques were isolated by cutting out the agarose plug containing the plaques as described previously. A total of 50 plaques gave positive hybridization results to probe C3A4 and they were all isolated.

InVivo Excision Protocols Using ExAssist Helper Phage with SOLR Strain

The ExAssist helper phage and SOLR bacterial strain allows efficient excision of the pBluescript phagemid from the Uni-ZAP XR vector. The helper phage contains an amber mutation that prevents replication of the phage genome using the SOLR cells as the host strain. Only the excised phagemid will replicate in the host so that risk of coinfection with the helper phage does not occur. In order to excise the pBluescript phagemid from the Uni-ZAP XR vector the following protocol was followed. The 50

resuspended plaques were vortexed to release the phage particles and incubated for 1-2 hours at room temperature (25°C). Separate cultures of XL1-Blue MRF' and SOLR cells were grown in LB broth overnight at 30°C. The overnight cultured XL1-Blue MRF' and SOLR cells (1000xg) were centrifuged for 10 minutes and pellets resuspended to an O.D.₆₀₀ of 1.0 in 10 mM MgSO₄. For *in vivo* excision of the vector the following components were mixed in 15 mL polypropylene tube:

XL1-Blue MRF' cells at an OD₆₀₀ of 1.0 200μl

Phage stock (containing $>1 \times 10^6$ pfu/μl) 250μl

ExAssist helper phage ($>1 \times 10^6$ pfu/μl) 1μl

The sample mixtures were stirred gently and incubated at 37°C for 15 minutes. After incubation 3mL of LB broth was added and incubated 2.5-3 hours at 37°C with shaking. The sample was then be heated at 65-70°C for 20 minutes and centrifuged at 1000xg for 15 minutes. The supernatant was decanted into a sterile 15 mL tube and saved. This supernatant contained the excised pBluescript phagemid containing the cDNAs of interest and packed as filamentous phage particles.

Excised phagemids were plated by adding 200 μl of freshly grown SOLR cells at an O.D.₆₀₀ of 1.0 to two 1.5ml microcentrifuge tubes. Two different volumes of excised phagemid were used to transform bacterial cells and obtain cloned cDNA into a pBluescript vector. 100 μl of each phage supernatant was added to one set of microcentrifuge tubes and 10μl of the phage supernatant was added to the other set of tubes. The tubes were incubated at 37°C for 15 minutes. 200μl of the cell mixture was plated from each tube onto LB ampicillin plates (50μg/ml) and incubated overnight at

37°C. The next day, single colonies were visible on all plates and one colony from each plate was isolated with a toothpick and streaked onto LB ampicillin plates and inoculated into LB Ampicillin broth to isolate the plasmids using Promega's SV Wizard MiniPrep Kit. Cycle PCR sequencing using Big Dye Terminators and M13 primers forward and reverse (Table 1) was performed on all 50 clones.

Primer Name (sense or antisense)	Primer Sequence 5' to 3'
M13 Reverse (sense)	ggaaacagctatgaccatg
M13 Forward (antisense)	gtaaaacgacggccagt
Thiolester Primer	aggaattcggntgyggngarcacacnatg
Primer I/S (sense)	cctgataaaggaccagcaa
Primer I/AS (antisense)	tggcactgaacttctgttgg

Table 1. Primers used for obtaining nucleotide sequence of the C3 clones isolated from the cDNA library. M13 Reverse and Forward primers flanked the cloned insert at the 5' and 3' end. The thiolester primer was used to obtain sequence in that region. Primer I/S and Primer I/AS were designed using Primer 3 program to obtain internal sequence of the cloned C3 inserts.

Restriction Enzyme Digestion of C3 Confirmed Plasmids

Insert sizes were determined by digesting samples with two restriction enzymes that flanked the cDNA inserts on the vector. The two enzymes used were *EcoRI* and *Hind III*. Electrophoretic analysis on a 1x TBE, 1% Agarose gel of the digested plasmids along with a low DNA mass ladder was performed to visualize the digested clones. This confirmation allowed us to proceed with obtaining additional sequence data by demonstrating that our C3-like clones were longer than the approximate 1000 bp of nucleic acid sequence obtained.

Design of Internal Primers

For all sequencing materials (16A1.2/1, 16A1.2/2, 16A1.3/2, and 20A1.3/2), M13

Forward and Reverse primers were used since the primer templates flanked our C3 cDNA cloned inserts. After obtaining nucleotide sequence data, we decided to use only one of the samples from the 16A group, that is, 16A1.2/1. From this sequence two primers were made of which there was a sense primer (**Primer I/S**) and an antisense primer (**Primer I/AS**) using Primer3 Program (available at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) (table 1). Another primer we used was one that would recognize the thiolester region and corresponded to the conserved thiolester sequence CGEQNM (table 1).

Sequence Analysis and Alignments

Samples were sequenced using the ABI 377 DNA Sequencer and analyzed using the BLAST program from www.ncbi.nlm.nih.gov to identify whether the sequences of isolated clones aligned with known C3 sequences. Only 4 clones gave positive sequence alignments with C3 of several organisms. Nucleotide sequences were submitted to the BLAST program and their reading frame identified. DNA translations were carried out using GeneDoc software to identify the proteins. Alignments were carried out in the Clustal X program using its default parameters. Alignments were saved as multiple sequence format (MSF) for use in GeneDoc where sequences were further aligned and edited, and their similarity, identity, conservation, and physiochemical property similarities could be determined.

IV. Results

Total RNA Isolation from Nurse Shark Liver

Total RNA was isolated from 0.5g of nurse shark liver. The 260nm/280nm ratio was 1.570 and the concentration was 11.6µg/µl. The integrity of the RNA was analyzed electrophoretically on a 1x TBE agarose gel (figure 4). Each diluted RNA sample showed two prominent bands representing the 28S and 18S species of ribosomal RNA, along with a smear ranging from 400bp to 8kb representing a heterogenous family of mRNA. Each sample displayed an intense fluorescent band at the top of the lane which suggests the presence of large fragments of DNA which are unable to enter the gel during electrophoresis due to their large size. Alternatively, this may be due to an initial drop in pH of the sample caused by the TRIZOL reagent causing excess RNA to remain in the wells and not migrate across the gel.

Isolation of mRNA

Isolation of mRNA is necessary to provide the template for cDNA synthesis. The latter represents the genes expressed by the nurse shark liver and forms the cDNA library. Two separate samples of mRNA were prepared using different amounts of total RNA as starting material. The concentration of total RNA was 12 µg/µl and The protocol required approximately 500µg of total RNA. For isolation of mRNA, 40 µl, (sample A) and 80 µl (sample B), (i.e. 480 µg and 960 µg respectively) was used. Two different concentrations of RNA were used to ensure that one of the concentrations would be sufficient for the preparation of cDNA. The two samples were labeled A and B, representing the 40 µL and 80 µL, respectively, of total RNA used. The concentration of

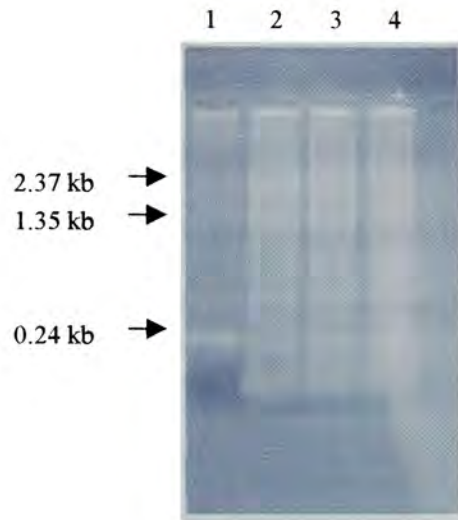


Figure 4. Electrophoretic analysis of total cellular RNA in a 1% agarose, 1 x TBE gel. RNA isolated from the nurse shark liver was diluted 100, 500, and 1000 fold. The 100, 500, and 1000 dilutions were run in lanes 2, 3, and 4, respectively. The gel was stained with ethidium bromide. The 18S and 28S ribosomal RNA bands of the RNA ladder run in lane 1, are indicated by arrows. All three diluted samples contained RNA. The smearing observed in the lanes represents mRNA in the sample. The additional discrete molecular weight bands may correspond to small rRNA or tRNA. The high molecular weight remaining in the well may be a result of high molecular weight RNA or DNA, or a result of overloading the gel with a high concentration of RNA which can not migrate through the gel properly.

the mRNA isolated was not determined spectrophotometrically because the procedure would have consumed one-third of the sample yield. Thus, the mRNA samples were used directly for cDNA synthesis without determining RNA concentration.

cDNA Synthesis

First and Second Strand Synthesis

The cDNA synthesis from the mRNA samples was performed using the Zap cDNA Synthesis Kit from Stratagene. First-strand synthesis was performed on both mRNA samples, A and B. The protocol required at least 5 µg of mRNA and a total volume of 37.5 µL mRNA could be used in the reaction mixture. Since quantitation of mRNA was not performed the total volume of mRNA permitted was used. From mRNA preparation A, 37.5 µl of sample was taken while 25µl of mRNA sample B was used. Results of electrophoretic analysis of products of first and second-strand synthesis using samples A and B are illustrated in figure 5. Lanes, two and three, containing cDNA for mRNA sample A and sample B, respectively, show a faint smear in cDNA samples. The faintness of the smear may be due to the small amount (5µl) of cDNA sample taken from the reaction mix for electrophoretic analysis. The presence of a cDNA smear indicated that the sample contained the range of cDNA sizes synthesized. The cDNA was precipitated, washed, and resuspended in 8µL in a mixture containing EcoRI adapters, ligase, and buffer.

Ligation of EcoRI Adapters and cDNA Size Fractionation

After ligation of *EcoRI* adapters to the cDNAs, size fractionation was performed to separate unligated adapters (small molecular size) from cDNAs ligated with adapters.

For each cDNA preparation (A and B) twenty column fractions were collected, each

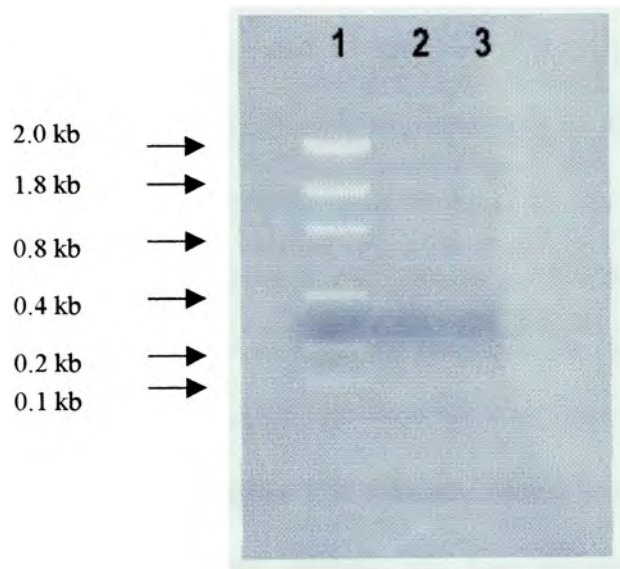


Figure 5. Electrophoretic analysis of cDNA synthesized from mRNA. cDNA samples, corresponding to mRNA preps A and B, were run in lanes 2 and 3 in 1% agarose, 1x TBE gel. Lane 1 contained the low mass DNA ladder (GibcoBRL) and bands corresponding to 2, 1.8, 1.0, 0.4 kb are indicated by arrows. Faint smearing can be observed in the sample lanes indicating successful first and second strand synthesis. The gel was stained with ethidium bromide.

consisting of 6 drops, approximately 400 μL /each drop. To estimate the concentration of cDNA in a 2 μL aliquot from each fraction was mixed and stained with 2 μL of ethidium bromide and spotted onto saran wrap spread over an ultra violet lamp. For comparison a low mass DNA ladder with concentration of 117.5 ng/ μL was serially diluted (double dilutions) from an initial 1/5 dilution. Drops of each marker dilution were placed in a row on the saran wrap and examined for intensity of fluorescence. In figure 6a, the dots represent marker concentrations (ng/ μL) corresponding to: 0, 0.734, 1.46, 2.9375, 5.875, 11.7, 23.5, and 117.5. In the middle panel of figure 6b, the two rows of spots increasing in fluorescence intensity from left to right are the 20 fractions collected from sample A drip column. The lower panel of figure 6c illustrates the two rows of spots representing 20 fractions collected from sample B which show increasing fluorescence from left to right. In the photograph a higher UV intensity corresponds to increasing sizes of the cDNAs. The fractions with less intensity may have present in them the *EcoRI* adapters and residual linker-primer that are much smaller in size than the cDNAs. It was not clear whether *EcoRI* adapter or residual linker-primer was in samples with less intensity because our blank, which contained ethidium bromide in water, demonstrated some UV fluorescence. For sample A, fraction #6 through fraction #12 showed higher levels of fluorescence and were pooled after electrophoretic analysis (figures 6 and 7). Similarly for sample B, fraction #26 through #32 showed higher fluorescence and were pooled.

To reliably confirm and identify the cDNA-containing fractions which could be pooled together, fractions were analyzed electrophoretically. All fractions were electrophoresed in 1x TAE agarose gel (1%) using 5 μL sample of each fraction (figure 7). Faint smears corresponding to cDNA could be observed in several lanes, which

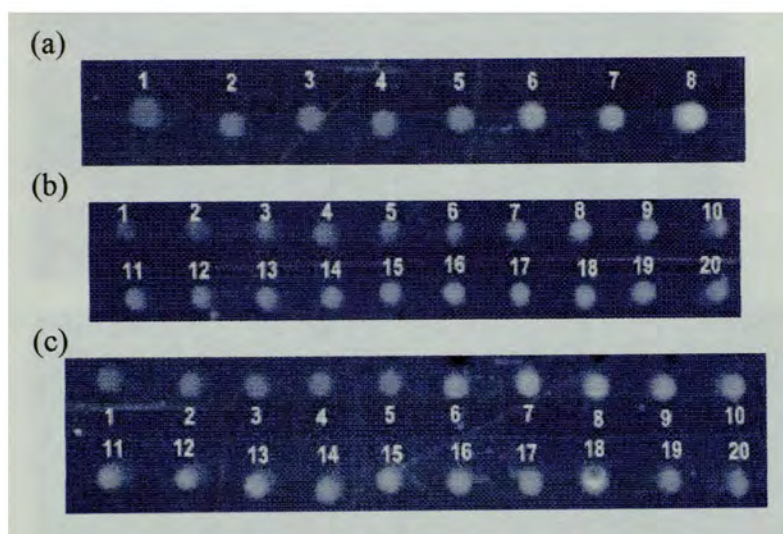


Figure 6. Estimation of cDNA concentration in column fractions. The top panel (a) shows dilutions of cDNA marker concentration ranging from zero (dot #1) only contains water) to 117.5 ng/ μ l (dot #8) (see text for details). An increasing concentration of cDNA is observed as intensity of fluorescence of the dots (ethidium bromide-stained). The middle (b) and lower (c) panel dots represent the 20 size fractions obtained for the two processed cDNA samples, A and B, respectively.

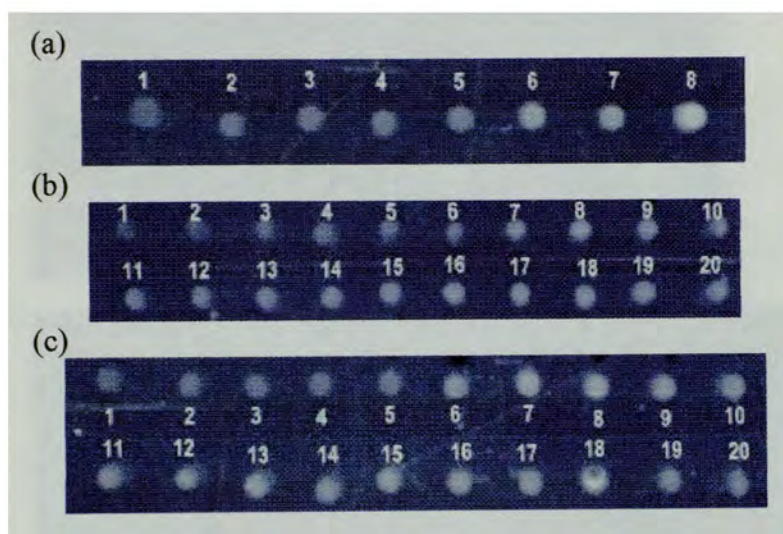


Figure 6. Estimation of cDNA concentration in column fractions. The top panel (a) shows dilutions of cDNA marker concentration ranging from zero (dot #1) only contains water) to 117.5 ng/ μ l (dot #8) (see text for details). An increasing concentration of cDNA is observed as intensity of fluorescence of the dots (ethidium bromide-stained). The middle (b) and lower (c) panel dots represent the 20 size fractions obtained for the two processed cDNA samples, A and B, respectively.

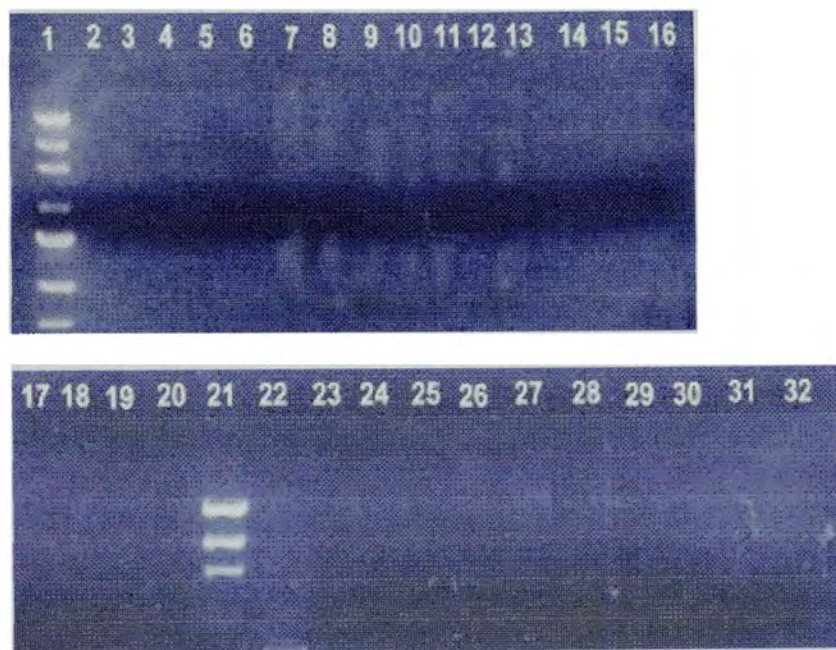


Figure 7. Electrophoretic analysis of column fractions for cDNA samples A and B. cDNA samples A and B were fractionated on a Sepharose CL-2B gel filtration column. Fractions were electrophoretically analyzed on a 1 x TAE, 1% agarose gel. Fractions # 1 through #20, from sample A were run in lanes #2 through #20, while fractions # 10 through #20, from sample B were run in lanes #22 through #32. Lanes 1 and 21 contained the low DNA mass ladder. Fractions #6 through #12 (lanes 7-13) of sample A and fractions #14 through #20 (lanes 24-32) of sample B showed presence of cDNA (faint smearing) and were pooled.

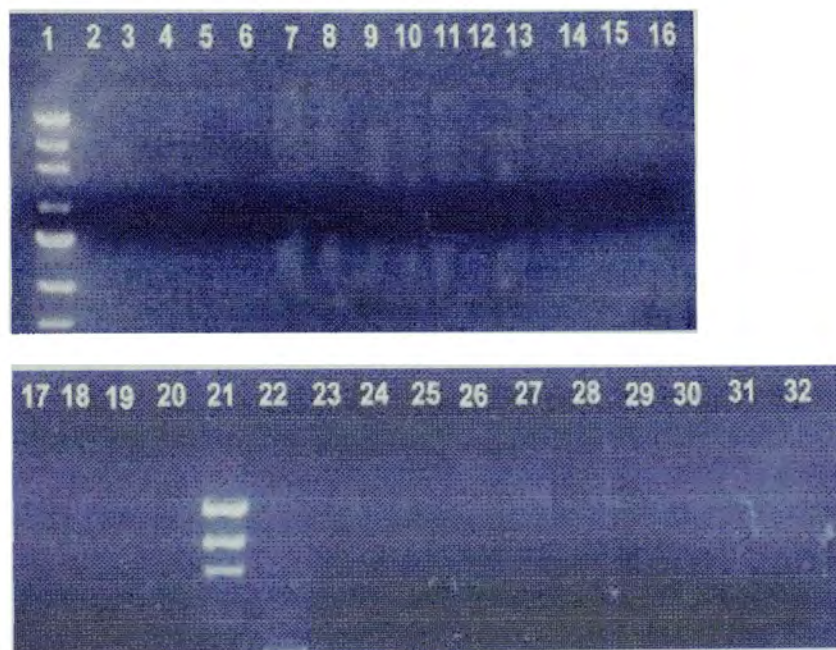


Figure 7. Electrophoretic analysis of column fractions for cDNA samples A and B. cDNA samples A and B were fractionated on a Sepharose CL-2B gel filtration column. Fractions were electrophoretically analyzed on a 1 x TAE, 1% agarose gel. Fractions # 1 through #20, from sample A were run in lanes #2 through #20, while fractions # 10 through #20, from sample B were run in lanes #22 through #32. Lanes 1 and 21 contained the low DNA mass ladder. Fractions #6 through #12 (lanes 7-13) of sample A and fractions #14 through #20 (lanes 24-32) of sample B showed presence of cDNA (faint smearing) and were pooled.

identified which fractions to pool together. Fractions #6 through #12 (lanes 7-13) of sample A were selected, and fractions #14 through #20 (lanes 24-32) from sample B were selected to be pooled. Fractions not selected were discarded. Another saran wrap spot test was performed after the appropriate fractions were pooled for each sample (figure 8). Four different concentrations of the low DNA mass ladder (GibcoBRL) were spotted corresponding to: 13ng/μl, 25ng/μl, 49ng/μl and 98ng/μl. A 1μL aliquot of pooled fractions of sample A and sample B were also spotted. Visual analysis estimated the concentration of cDNA in each sample to be approximately 50ng/μl for both samples A and B. cDNA inserts of sample A and sample B and the test insert were all ligated into the Uni-Zap XR Vector.

Packaging Uni-ZAP XR Library into Extracts, Plating and Titration

Packaging of cDNA Library into Extracts

The cDNA Uni-ZAP XR library was packaged into lambda phage extracts. Two extracts were used per sample (sample A/1, A/2, B/1 and B/2). After samples were packaged the initial phage titration was carried out on small size plates initially to determine the efficiency and plaque forming units (pfu). The total volume of each library constructed was 500μl. One microliter of each sample was used for the titration. The total titre in plaque forming units (PFU) was calculated using the following equation:

$$[\text{number of plaques (pfu)} \times \text{dilution factor/volume plated (1 } \mu\text{L)}]$$

After incubation of the phage sample with host bacteria, the plaques on each plate were counted (Table 2) and the pfu titre for each library sample determined. The optimal pfu for a cDNA library is considered to be 1×10^6 . Results from all four samples indicated

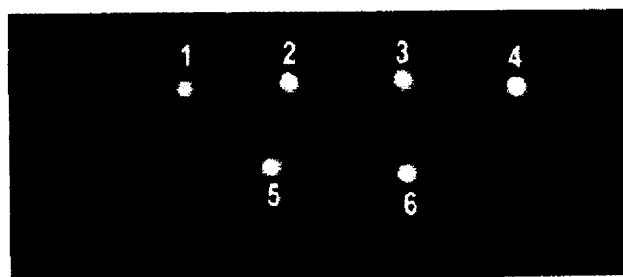


Figure 8. Estimation of cDNA concentration in pooled fractions of samples A and B. Dots 1-4 represent the four dilutions of the low mass DNA ladder (GibcoBRL): 13ng/ μ l (dot 1), 25ng/ μ l (dot 2), 49ng/ μ l (dot 3), and 98 ng/ μ l (dot 4). Dot 5 and 6 represent the pooled fractions for sample A and B, respectively.

cDNA Library	No. of Plaques	Plaque Forming Units
A	3568	1.78×10^6
A/1	2364	1.18×10^6
B	2872	1.43×10^6
B/1	2304	1.15×10^6

Table 2. Determination of plaque forming units of library samples. The number of plaques on each 100 x 15mm plate were counted and the number of plaque forming units of each cDNA library was calculated using the following equation:

$$\frac{(\text{number of plaques}) \times \text{dilution}}{\text{volume plated}}$$

(volume equals 1 μl and the packaging volume was 500 μl .)

good efficiency. Of the four samples, sample B/2, which contained a titre of 1.15×10^6 pfu was selected for further amplification and screening.

Amplification of Library and Screening of cDNA Library

Amplification of cDNA library sample B/2 was achieved following the protocol outlined in Materials and Methods. Twenty-two plates were prepared with NZY agar and layered with soft top NZY agar containing bacteria that had been incubated with 23 μ l of cDNA library extracts. After 8 hours of incubation a bacterial lawn was present with plaques spread throughout all the plates. All plates were suitable for performing plaque lifts that were then screened with the Digoxigenin labeled C3A4 probe.

Preparation of DIG Labeled Probes

Probes for screening cDNA library were labeled with non-radioactive digoxigenin using the DIG-PCR Kit (Amersham Pharmacia). Electrophoretic analysis of labelled and unlabelled probe is shown in figure 9 which illustrates profile of nonlabeled C3A4 (lane 1) and labeled C3A4 (lane 3) probe. Molecular size determinations were made by comparison with bands of a Low DNA Mass Ladder. The sizes of the unlabelled probe samples had an insert size of 220bp while labeled probe samples banded at approximately 400 bp because of the added size of label, digoxigenin.

Plaque Lifts

Plaque lifts were performed on all 22 plates. The plates were labeled P1-22. Two hybond N membranes (Amersham) were used per plate. Each membrane was labeled according to plate number and probe that would be used to screen for C3. The probe used were C3A4 and C3A7. However, because of the complexity of the project, only results from screening with probe C3A4 will be reported. The letter A identifies the

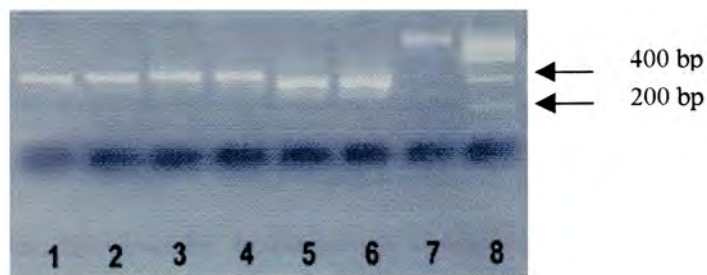


Figure 9. Agarose electrophoresis of DIG-labeled probes. Lanes 1 and 2 contain the 440bp DIG-labeled C3A4 PCR product. Lanes 3 and 4 contain the 440bp DIG-labeled C3A7 PCR product. Non-labeled PCR products for C3A4 and C3A7 can be seen in lanes 5 and 6, respectively, and have a molecular weight of 220. PCR products of the control template containing C3A4 is in lane 7. Low DNA Mass Ladder was run in lane 8, and size indicated by arrows.

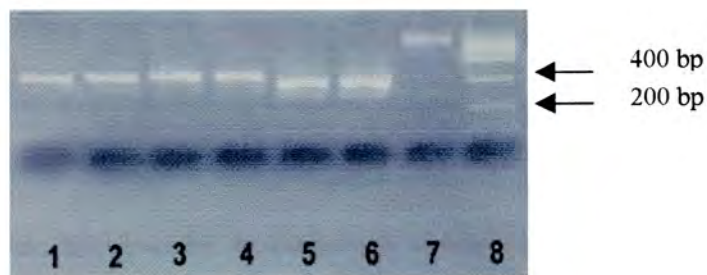


Figure 9. Agarose electrophoresis of DIG-labeled probes. Lanes 1 and 2 contain the 440bp DIG-labeled C3A4 PCR product. Lanes 3 and 4 contains the 440bp DIG-labeled C3A7 PCR product. Non-labeled PCR products for C3A4 and C3A7 can be seen in lanes 5 and 6, respectively, and have a molecular weight of 220. PCR products of the control template containing C3A4 is in lane 7. Low DNA Mass Ladder was run in lane 8, and size indicated by arrows..

plaque lifts that were screened with C3A4. Letter A corresponded to probe C3A4. For example, 1A denoted plaque lift collected from plate 1 and screened with probe A4, whereas, 1B denotes plaque lifts from plate 1 screened with C3A7 (not reported). Because of the complexity of the project, only results from screening with probe C3A4 will be reported.

Screening of Plaque Lifts

The primary screening of all plaque lifts was performed with the digoxigenin labeled C3A4 probe. Positive reactions were identified by the color reaction created by detection of the DIG labeled probe that hybridized to plaques containing potential C3 inserts. The color reaction created a bluish purple dot. There were many positive reactions, of which it was not possible to determine if they were all false positives; therefore, sixty samples were randomly selected from all the plates and collected.

Collection of samples required isolation and removal of the agarose plug containing the plaque that gave the positive reaction with probe. Since the distances between plaques were so small a secondary screening had to be done to isolate single phage plaques that had no contamination from neighboring plaques. Labeling of the plaques collected (as listed in table 3) was as follows: plate number, probe used for screening, and plaque number, that is, P4A2.2/1 corresponds to Plate #4, screened with C3A4 probe, yielding plaque #2.2/1.

Secondary Screening

Secondary screening was performed only on samples isolated using the A4 probe. The collected agarose plugs were vortexed in SM buffer and chloroform, and the phage particles released. A small amount of the phage sample was mixed with host bacteria,

Isolated and Sequenced Phagemids	Identified as C3	Isolated and Sequenced Phagemids	Identified as C3
P2A1.3/1	-	P4A2.3/2	-
P2A1.3/2	-	P5A2.1/1	-
P2A1.3/3	-	P5A2.1/2	-
P2A1.3/4	-	P7A2.2/1	-
P2A1.4/1	-	P7A2.2/2	-
P2A1.4/2	-	P7A7.2/1	-
P2A2.1/1	-	P7A7.2/2	-
P2A2.1/2	-	P7A7.3/1	-
P2A2.1/3	-	P7A7.3/2	-
P2A3.1/1	-	P9A2.2/1	-
P2A3.1/2	-	P9A2.2/2	-
P2A6.1/1	-	P10A2.1/2	-
P2A6.1/2	-	P16A1.2/1	+
P2A6.1/3	-	P16A1.2/2	+
P2A6.1/4	-	P16A1.3/1	-
P2A6.2/1	-	P16A1.3/2	+
P2A6.2/2	-	P16A1.4/1	-
P3A8.1/2	-	P20A1.1/1	-
P4A1.3/1	-	P20A1.1/2	-
P4A1.3/2	-	P20A1.2/1	-
P4A2.1/1	-	P20A1.2/2	-
P4A2.1/2	-	P20A1.3/1	-
P4A2.2/1	-	P20A1.3/2	+
P4A2.2/2	-	P20A2.1/1	-
P4A2.3/1	-	P20A2.1/2	-

Table 3. Identification of C3-like clones isolated from the library. Of the 48 samples excised only 50 gave phagemids that were isolated and then sequenced. After performing BLAST searches on NCBI only samples were identified as C3-like sequences: P16A1.2/1, P16A1.2/2, P16A1.3/2, and P20A1.3/2

plated on 110x15 mm size NZY agar plates, and incubated for secondary screening. After overnight incubation, plaque lifts were prepared using smaller Hybond-N membranes and then screened with the A4 DIG-labeled probe. The secondary screening led to the isolation of 58 plaques containing potential C3 cDNA inserts (Table 3). The labeling of these isolates consisted of adding the plaque number from the secondary screening.

In Vivo Excision of the pBluescript Phagemid from the Uni-ZAP XR Vector

The Uni-ZAP XR Vector allowed *in vivo* excision and recirculization of any cloned insert contained within the lambda vector to form phagemids containing the cloned insert. All 58 samples that gave positive signals after screening with hybridized Digoxigenin C3A4 probe was subject to excision and conversion to phagemids. All isolates formed several bacterial colonies containing the phagemids. Fifty (50) random colonies were isolated with a sterile toothpick onto a master plate containing LB agar and ampicillin (50µg/mL) and the toothpick containing the isolated colony was dipped into 3mL of LB and ampicillin broth from each sample and grown overnight at 37°C. The “phagemid” DNA was isolated from each bacterial culture using the methods described previously for DNA isolation using the Qiagen SV Miniprep Kit. The 50 DNA samples isolated were labeled as described except that a fourth number describing the colony number picked was added to the end of the series (Table 3) describing the cloned isolate.

Sequencing Reactions

All 50 phagemid samples purified were subject to cycle sequencing PCR using M13 forward and reverse primers (Table 1). A control sample was also prepared using the standard pGEM control that was supplied with the Big Dye Sequencing Kit. After

PCR and labeling of the samples for sequencing all samples were ethanol precipitated and washed. Samples were sequenced using ABI 377 Sequencer.

Samples were sequenced on 48 cm plates that provided us with sequences of more than 800 base pairs. Data was analyzed initially using the NCBI Blast database in which the DNA sequences are automatically converted into protein and then checked against all the amino acid sequences in the database. Of the 50 samples sequenced, four samples displayed similarity to C3 in other organisms, the samples were: 16A1.2/1, 16A1.2/2, 16A1.3/2, and 20A1.3/2 (Table 3). Sequence data from the sense and anti-sense direction was obtained (figure 10 and figure 11). After initial confirmation that these four samples were in fact shark C3-like clones sequence alignments using Clustal X (figure 11) and Genedoc computer programs were performed to determine similarity and identity between clones (figure 10 and figure 11). Clones P16A1.2/1, P16A1.2/2, and P16A1.3/2 had approximately 90% identity with each other and approximately 44% identity with clone 20A1.3/2 (figure 10 and 11). Furthermore, the size of the inserts (i.e. clones) was determined by electrophoretic analysis (figure 12).

Size Determination of Inserts

Insert sizes were determined by digesting samples with two restriction enzymes for which cleavage sites flanked the cDNA inserts on the vector (figure 12). The two enzymes used were *EcoRI* and *Hind III*. Electrophoretic analysis on a 1x TBE, 1% agarose gel of the digested plasmids along with a low DNA mass ladder displayed the fragment pattern for each sample (figure 12). Clones, P16A1.2/1 (Clone I), P16A1.2/2 (Clone II), and P16A1.3/2 (Clone III) had identical banding patterns when cut with *EcoRI* and *HindIII*, whereas P20A1.3/2 (Clone IV) had a different banding pattern and

(Figure 10 continued)

16A1.2/1For	:	AGCTTTTCAGAAAACTCTGCTAGGCTCTGACAGAGTCTGCGGGTATTTTGGCACTGAACTTCTGTGGCCA	:	426
16A1.2/2For	:	AGCTTTTCAGAAAACTCTGCTAGGCTCTGACAGAGTCTGCGGTATTTTGGCACTGAACTTCTGTGGCCA	:	428
16A1.3/2For	:	AGCTTTTCAGAAAACTCTGCTAGGCTCTGACAGAGTCTGCGGTATTTTGGCACTGAACTTCTGTGGCCA	:	455
20A1.3/2For	:	AGCTTTTCAGAAAACTCTGCTAGGCTCTGACAGAGTCTGCGGTATTTTGGCACTGAACTTCTGTGGCCA	:	433
16A1.2/1For	:	AAATTCCAACCCAGGTTCCATCTCCAAATCAGATAGTTGATCCTTT	:	502
16A1.2/2For	:	AAATTCCAACCCAGGTTCCATCTCCAAATCAGATAGTTGATCCTTT	:	504
16A1.3/2For	:	AAATTCCAACCCAGGTTCCATCTCCAAATCAGATAGTTGATCCTTT	:	531
20A1.3/2For	:	AAATTCCAACCCAGGTTCCATCTCCAAATCAGATAGTTGATCCTTT	:	508
16A1.2/1For	:	ATCAGTAGATAAATCATTTGTCCGAGTCCACATTAAATGTTCCCTTCACAG	:	578
16A1.2/2For	:	ATCAGTAGATAAATCATTTGTCCGAGTCCACATTAAATGTTCCCTTCACAG	:	580
16A1.3/2For	:	ATCAGTAGATAAATCATTTGTCCGAGTCCACATTAAATGTTCCCTTCACAG	:	607
20A1.3/2For	:	ATCAGTAGATAAATCATTTGTCCGAGTCCACATTAAATGTTCCCTTCACAG	:	583
16A1.2/1For	:	GCCATCCAATATGTCATCGCTT	:	654
16A1.2/2For	:	GCCATCCAATATGTCATCGCTT	:	656
16A1.3/2For	:	GCCATCCAATATGTCATCGCTT	:	683
20A1.3/2For	:	TCCATCCAATATGTCATCGCTT	:	657
16A1.2/1For	:	TTTTCCCTTCTTGTAAATG-CACATTTGTAACATAAATCTG	:	728
16A1.2/2For	:	TTTTCCCTTCTTGTAAATG-CACATTTGTAACATAAATCTG	:	731
16A1.3/2For	:	TTTTCCCTTCTTGTAAATG-CACATTTGTAACATAAATCTG	:	757
20A1.3/2For	:	TTTTCCCTTCTTGTAAATGTCACTTGTAAACATAAATCTG	:	732

(Figure 10 continued)

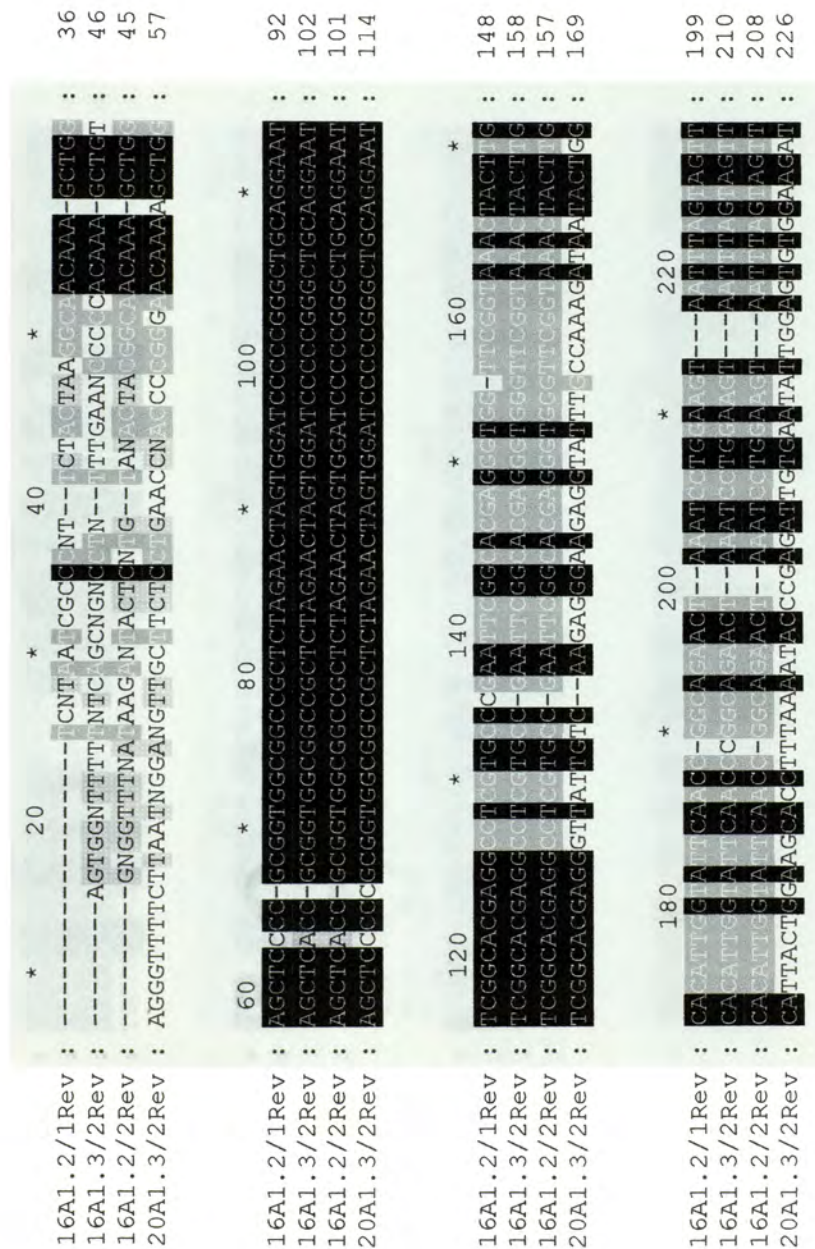
16A1.2/1For	AAGTGATATGG	GTTCA	GCA	CTTAACTG	ATAA	C	ACTTN	CTTNTGA	ATCTGCAACTT	CANCTTG	CAGA	801
16A1.2/2For	AAGTGATATGG	GTTCA	GCA	CTTAACTG	ATAA	C	ACTTT	CTTNTGNA	ATTTGCAAACTT	NACCTTG	CAGA	806
16A1.3/2For	AAGTGATATGG	GTTCA	GCA	CTTAACTG	ATAA	C	ACTT	CTTTTGA	ATCTGCAAACTT	NACCTTG	CAGA	828
20A1.3/2For	AAGTGATATGG	GTTCA	GCA	CTTAACTG	GATAT	CACTT	CTTTTGA	ATTTGCAAACTT	ACCTTTGG	CAGA	805	

16A1.2/1For	TTTGTCT	AA	CATGG	ACTG	TTTTCAT	CG	CAAT	TAA	AA	T	AG	A	T	CTCTGT	ATCA	AA	ATT	866
16A1.2/2For	TTTGTCT	AA	CATGG	CACTG	TTTTCAT	TG	CAATTA	TA	AA	TAG	G	AACTGT	TCTCTGT	TTTCA	AA	ATT	881	
16A1.3/2For	TTTGTG	AA	CATGG	ACTG	TTTTCAT	TCG	CAATTA	TA	AA	T	A	GCAACCT	TTCTCTGT	ATTC	AA	ATT	903	
20A1.3/2For	TTTGTG	AA	CATGG	CACTG	TTTTCAT	CG	CAATTTAA	AT	T	A	G	A	C	AGTCTCTGT	TTCA	AA	ATT	870

16A1.2/1For	CA	AAA	TTT	TAC	AG	AACTGG	CG	GAATAAG	CCAC	TTT	GA	AAATG	ATGT	CC	AAATCC	AA	CAA	931
16A1.2/2For	CA	AAA	TTT	TAC	CAA	AACTGGG	TGAATAAGG	CCAC	TTT	GA	AAATG	ATGT	CC	AAATCC	AA	CAA	956	
16A1.3/2For	CA	AAA	TTT	TAC	AAG	AGCTGGG	GGAATAAG	CCAC	TTT	GA	AAATG	ATGT	CC	AAATCC	AA	CAA	976	
20A1.3/2For	CA	AAA	TTT	TAC	CA	AGTGG	TGAATAAG	CCAC	TTT	GA	AAATG	ATGT	CC	AAATCC	AA	CAA	936	

16A1.2/1For	GT	TTTTTT	TA	AG	GG	GA	TCG	ATCC	AA	ATA	CAAT	CAT	NG	971				
16A1.2/2For	GG	TTTTTT	TA	TAAG	NG	GA	AA	CGA	ATN	CA	ATA	CAAT	CAAT	GG	TC	1004		
16A1.3/2For	GG	TTTTTT	CT	AA	GT	GT	GA	AA	T	CGA	ATN	CA	ATA	CAAT	CAAT	GG	TC	1025
20A1.3/2For	GT	TTTTTT	TA	AG	GG	GA	TCG	ATCC	AA	ATA	CAAT	CAT	NG	980				

Figure 11. The figure continues over several pages. Alignment of nucleotide sequence data of C3-like clones 16A1.2/1, 16A1.2/2, 16A1.3/2, and 20A1.3/2 obtained using M13 reverse primer.



(Figure 11 continued)

16A1.2/1Rev	:	CATCTCAGTCTCCGTAA-----TGACAATGACATGTCAGAACACGATATAAGTATTTT	:	252
16A1.3/2Rev	:	CATCTCAGTCTCCGTAA-----TGACAATGACATGTCAGAACACGATATAAGTATTTT	:	263
16A1.2/2Rev	:	CATCTCAGTCTCCGTAA-----TGACAATGACATGTCAGAACACGATATAAGTATTTT	:	261
20A1.3/2Rev	:	TGCTGCAAGCTACAAAGAAGCCAGCCATGTAAT-TACACCACTGAATTTGAAGTTA	:	282
16A1.2/1Rev	:	CAGC-TACATGTGTGATTAACAAAGGAAGATTTGCTAGTTTGACGACACGGAAGGG	:	308
16A1.3/2Rev	:	CAGC-TACATGTGTGATTAACAAAGGAAGATTTGCTAGTTTGACGACACGGAAGGG	:	319
16A1.2/2Rev	:	CAGC-TACATGTGTGATTAACAAAGGAAGATTTGCTAGTTTGACGACACGGAAGGG	:	317
20A1.3/2Rev	:	AGGAATATGTGTTACCAAGTTTGAAGTTACCTTCGAAACGGAAAGCAATTTTCT	:	339
16A1.2/1Rev	:	AAATTGGCCAAATTTGTGGTGAACATGTCTCAATCCAAACACTTCTTACTTTAATACCAAT	:	365
16A1.3/2Rev	:	AAATTGGCCAAATTTGTGGTGAACATGTCTCAATCCAAACACTTCTTACTTTAATACCAAT	:	376
16A1.2/2Rev	:	AAATTGGCCAAATTTGTGGTGAACATGTCTCAATCCAAACACTTCTTACTTTAATACCAAT	:	374
20A1.3/2Rev	:	ACGTGATGATTAATAGTTTGGAACT-TGCAGTCAACAGCAAGATTCACTTATGGAAAG	:	395
16A1.2/1Rev	:	CCTTCAGACTGTGTGTCATATTATACCTGTTAAGGGACACAGCAATAG-AAATGGTT	:	421
16A1.3/2Rev	:	CCTTCAGACTGTGTGTCATATTATACCTGTTAAGGGACACAGCAATAG-AAATGGTT	:	432
16A1.2/2Rev	:	CCTTCAGACTGTGTGTCATATTATACCTGTTAAGGGACACAGCAATAG-AAATGGTT	:	430
20A1.3/2Rev	:	CCTGTTGAGGGTCGTGCATTTATTATGTT-TGGAAATAATGAAGGATGCCGAAAGGAT	:	451
16A1.2/1Rev	:	GCTGATTCCTGT-ATGGGTTGATGTGAAGACAC-ATGCATGGGAAATTGATCTGT	:	476
16A1.3/2Rev	:	GCTGATTCCTGT-ATGGGTTGATGTGAAGACAC-ATGCATGGGAAATTGATCTGT	:	487
16A1.2/2Rev	:	GCTGATTCCTGT-ATGGGTTGATGTGAAGACAC-ATGCATGGGAAATTGATCTGT	:	485
20A1.3/2Rev	:	AGGAATTCATTCATCACTTCATCATATAGATTTACAGATGGCCAAAGGAAATGCTA-	:	507

(Figure 11 continued)

16A1.2/1Rev	: CTGCTGCAACGGAACAGTTAAAATAAGTTTCAAACTGGCAAAACAGTTTCAAC	: 533
16A1.3/2Rev	: CTGCTGCAACGGAACAGTTAAAATAAGTTTCAAACTGGCAAAACAGTTTCAAC	: 544
16A1.2/2Rev	: CTGCTGCAACGGAACAGTTAAAATAAGTTTCAAACTGGCAAAACAGTTTCAAC	: 542
20A1.3/2Rev	: CCATGACAAAGTGCA-ATGTTCAAAGCCAGCTTTCCCGAC-----A--TTTATGA-	: 553
	520 * 540 * 560 *	
16A1.2/1Rev	: TGAATAACAGGGGTGACTCTGCGGGCAACAGTTGCTCTCTGCT	

(Figure 11 continued)

16A1.2/1Rev :	GGTTACAAAGACGGCTTGACCTCAAGC--ACCAATGATA--CGTT--ACCTTGG	800
16A1.3/2Rev :	GGTTACAAAGACGGCTTG--CTTCAAGC--ACCAATGATA--CGTTGACCTTGG	814
16A1.2/2Rev :	GGTTACAAAGACGGCTTGACCTCAAGC--ACCAATGATA--CGTT--GACGTGG	814
20A1.3/2Rev :	AGT--CAATGGGGGACAGCCATTCAAAA--ACCCAACTGA--TG-GTC TTGGC	819
* 820 * 840 *		
16A1.2/1Rev :	GCAAG--GAGATTTATTGCAATGCTNAGGTGGCTTAANTGATTCAGGTACAAGCG	856
16A1.3/2Rev :	GCAAG--GAGATTT--TTGGATG--TTAAGTTGTTAAATGATTCAGGTACAAGCG	868
16A1.2/2Rev :	GCAAGCGATTTT TTGGATG--TTAAGTTGAGGCTTAATTTGATCA--GTAA--AAGCG	868
20A1.3/2Rev :	AAGGTTGACATTT--AAATACAGTTGGCGCTACAGTTCT-CCA GTAA--AAGTTG	873
* 860 * 880 * 900 *		
920	* 940 * 960	
16A1.2/1Rev :	GAAATATTTTAAATAAATGCTTCAACCTTNGGATTCAGACNTAANCCCTTGGGG	913
16A1.3/2Rev :	GAAATAATTTAAATAAATGCTTGCANCTTGGATGGACG AANCCCTTGGGG	924
16A1.2/2Rev :	AAAT--ATTTAAATAAATGCTTGCACGTT--GGATGAGACAACTCCAT--GGG	919
20A1.3/2Rev :	AAACA GTTGGCCAACTCTTCTTAATAATTCGGCTGGCCTTGCAAGTATGGGGGG	930
* 980 * 1000 * 1020		
16A1.2/1Rev :	GTATTNTTTTGGNCCGNNATGCAAGGNGNTAAACNCNNTNGGAAAC-TTG	969
16A1.3/2Rev :	GTATTNTTTTGCCTCCGAAACCAAGGAGATAAACCCTTACCCGGGAAAGAC-TTG	980
16A1.2/2Rev :	GTATTNTTTTGGCTGGC--AAACCAAGGGG--ATATAACCTTACCCGGGAA--CTG	969
20A1.3/2Rev :	GCCAACCTTA AAAACCT--AAGGGGGTTCCGTTAAATTCCTTGC CATGGTA	985

(Figure 11 continued)

```

16A1.2/1Rev : T T A A G G - T T G C C T T T C T - G G G C T T G G T G G G A N C C N C C T N C A A A A A T T T A G G A : 1024
16A1.3/2Rev : T T A A G C T T G C C T T T C T T G G G A T T G T G G G A N C C C C C T T C A A A A T T T A A G G : 1037
16A1.2/2Rev : T T A A C - T T G C - T T T C T - - G G A - - - T G C A A C C A C C C T C - - A A A A T T T N G G - - : 1015
20A1.3/2Rev : T T T A A A C C G G C A G A A C T T A A T T C - - - C T G G A G G A A T T T A A G A G T C A T T T T C A A - - : 1037

16A1.2/1Rev : T C T T C C N A A A A G G C A A A A T C N - - - - : 1044
16A1.3/2Rev : T T T G C A A A A G C A A A G T T C T C A : 1061
16A1.2/2Rev : - - - T T T C A A A G C A A G T C A - - - - : 1032
20A1.3/2Rev : T C T T C C G A A T G G A C C - - - - - - : 1054

```

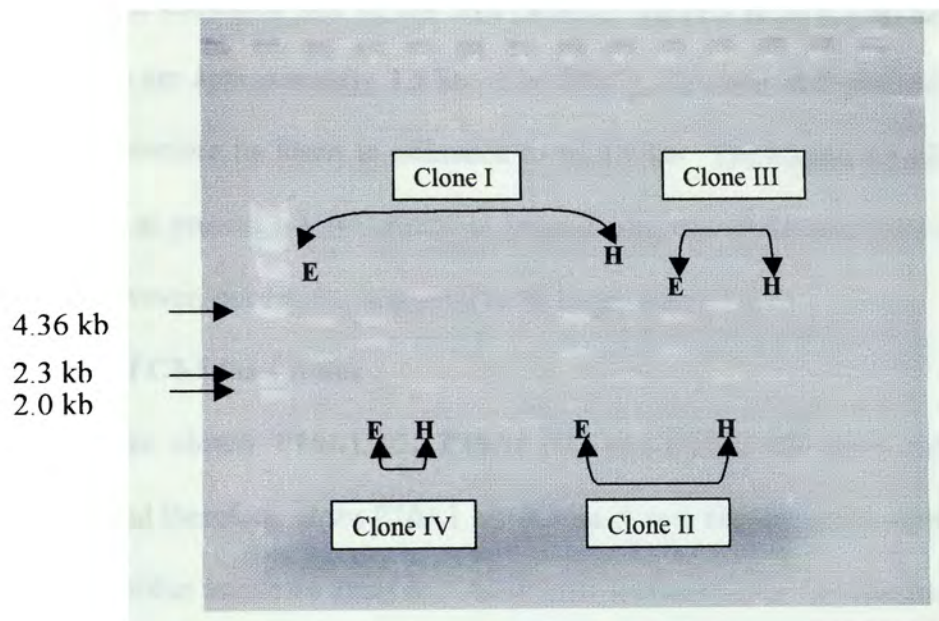



Figure 12. Electrophoretic analysis of restriction digests of Clones I, II, III, and IV. Clones I, II, III, and IV (16A1.2/1, 16A1.2/2, 16A1.3/2, and 20A1.3/2, respectively) were digested with restriction enzymes *EcoRI* (E) and *HindIII* (H). A λ DNA/*HindIII* fragment ladder was run on lane 1 and arrows indicate band size. Digests of Clones I, II, III, and IV were run in lanes 2 and 8, and 7 and 10, 9 and 11, and 3 and 4, respectively. Clones I, II, and III gave similar size restriction fragments (4.0 kb and 2.5 kb, with E, and 4.3 kb and 2.2 kb, with H), while restriction fragment pattern of Clone IV was different (4.2 kb and 3.3 kb, with E, and 5.3 kb and 2.2 kb, with H).

was larger in the total size of the insert (figure 12). Digestion with *EcoRI* produced two bands for the three P16A clones isolated from P16, while two bands at approximately 4.2 kb and 3.3 kb, were obtained for sample P20A1.3/2. Digestion with *HindIII* produced two fragments banding at 4.3 kb and 2.2 kb for the P16A clones (I, II, III) and for clone P20A1.3/2 (IV) bands formed at 5.3 kb and 2.2 kb. The size of the pBluescript insert is 2.9 kb. The estimated size of our 16A plasmids appears to be 6.5 kb meaning that our insert sizes are approximately 3.5 kb. The P20A1.3/2 clone plus plasmid was 7.5 kb in size and therefore its insert is estimated to be 4.5 kb. The human C3 cDNA is 5 kb in length and at present we are unable to estimate the size of the complete nurse shark C3 cDNA, however, our results suggest it to be larger than 3.5 kb.

Analysis of C3-Like Clones

The C3-like clones P16A1.2/1, P16A1.2/2, and P16A1.3/2 were considered to be identical and therefore, clone P16A1.2/1 (Clone I) was chosen as the representative of all three for further sequence analysis. Additional sequence data was obtained by designing gene specific primers, Primer I/S and Primer I/AS, that would provide us with additional sequence data for 5' and 3' ends of the inserts (Table 1). The thiolester primer sequence (CGEQNM) was used to obtain sequence data downstream of the thiolester region. A total of 2262 nucleotides were sequenced and the deduced amino acid sequence determined for 754 residues. Edited results of Clone I (P16A1.2/1) using primers M13 Reverse, Primer I/S, Thiolester Primer, M13 Forward, and Primer I/AS are illustrated in figures 13, 14, 15, 16 and 17, respectively, in which primer regions and biologically significant sites are denoted by boldface. The deduced amino acid sequences

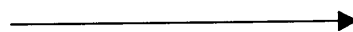
tcc ccc ggg ctg cag gaa ttc ggc acg agg cct cgt gcc cga att cgg cac gag ggt ggt tcg gta
S P G L Q E F G T R P R A R I R H E G G S V

aac tac ttg cac att ggt att caa ccg gca gaa cta aat cct gga agt aat tta gta gtt cat ctc agt
N Y L H I G I Q P A E L N P G S N L V V H L S

ctc cgt aat gac aat gca gat gtc cag aac cag ata aag tat ttc agc tac atg ttg ata aac aaa gga
L R N D N A D V Q N Q I K Y F S Y M L I N K G

aag att gtg gta gtt gga cga cag cga agg gaa att ggc caa att gtg gtg aac atg ctc att cca atc
K I V V V G R Q R R E I G Q I V V N M L I P I

Primer I/S

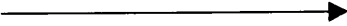


act tct gac tta ata cca tcc ttc aga ctg ctg gtc tat tac tac **ctg ata aag gga cca gca** ata gaa
T S D L I P S F R L L V Y Y Y L I K G P A I E

atg gtt gct gat tct gta tgg gtt gat gtg aaa gac aca tgc atg gga aaa ttg ata gtg tct gct gca
M V A D S V W V D V K D T C M G K L I V S A A

aac gaa cac gat gaa aat aag atc tac aaa cct ggc aaa cag ttt caa ctg aaa atc acg ggt gac
N E H D E N K I Y K P G K Q F Q L K I T G D

Figure 13. Nucleotide and deduced amino acid sequence of C3 clone P16A1.2/1 (Clone I) using M13 Reverse Primer.

Primer I/S


act tct gac tta ata cca tcc ttc aga ctg ctg gtc tat tac tac **ctg ata aag gga cca gca ata gaa**
T S D L I P S F R L L V Y Y Y L I K G P A I E

atg gtt gct gat tct gta tgg gtt gat gtg aaa gac aca tgc atg gga aaa ttg ata gtg tct gct gca
M V A D S V W V D V K D T C M G K L I V S A A

aac gaa cac gat gaa aat aag atc tac aaa cct ggc aaa cag ttt caa ctg aaa atc acg ggt gac
N E H D E N K I Y K P G K Q F Q L K I T G D

tct ggg gca aca gtt ggt ctc gtt gct gtg gat aaa gct gtg ttc gtg ttg aat aag aaa aat aaa tta
S G A T V G L V A V D K A V F V L N K K N K L

acg cag agc aag atc tgg gac att gtg gag aag aat gac atc ggc tgt acg gct gga agt ggc agt
T Q S K I W D I V E K N D I G C T A G S G S

gat gtt ctt ggt gtt ttc aca gat gca gga cta aca att gta aca agc agt aac ctc aag acg gct aca
D V L G V F T D A G L T I V T S S N L K T A T

β/ α cut site

aga aca gac ctg acg tgc aag cag cca atg aaa cgt aga cgt cgt gca acg aga ctt atg gat gtt
R T D L T C K Q P M K R R R R A T R L M D V

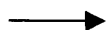
aag atg gct aaa ttg aat cag tac aag gcg aaa ata tta aat aaa tgc tgc aac gat gga atg aaa
K M A K L N Q Y K A K I L N K C C N D G M K

cga aac ccc atg ggg ttt tca tgt gca cgc aga gca agg aga ata aca caa cgg gaa gac tgt aaa
R N P M G F S C A R R A R R I T Q R E D C K

gct gcc ttc ctg gac tgt tgc aac cac gtc caa aaa ttt aag
A A F L D C C N H V Q K F K

Figure 14. Nucleotide and deduced amino acid sequences of C3 clone P16A1.2/1 (Clone I) using an internal gene specific sense primer (**Primer I/S**).

Thiolester Primer
(CGEQNM)



	TTG	GCT	TAT	CGN	CAA	CCC	CNT	GGT	GCT	TAT	GCA	GCA		
	L	A	Y	R	Q	P	X	G	A	Y	A	A		
TTT	CTA	AAT	CGA	GAA	CCG	AGC	ACG	TGG	TTG	ACT	GCT	TAT	GTG	GTG
F	L	N	R	E	P	S	T	W	L	T	A	Y	V	V
AAA	GTA	TTC	GCA	ATG	TCA	TTC	AAT	TTG	ATC	GTA	ATT	GAT	CCA	AAT
K	V	F	A	M	S	F	N	L	I	V	I	D	P	N
GTT	CTA	TGT	GGA	GCT	GCA	CAG	TGG	CTG	ACA	TTA	AAT	AAA	CAG	AAG
V	L	C	G	A	A	Q	W	L	T	L	N	K	Q	K
CCT	GAT	GGT	TCA	TTC	AAA	GAA	GAT	GCT	CCC	GTA	ATT	CAT	GGA	GAG
P	D	G	S	F	K	E	D	A	P	V	I	H	G	E
ATG	ATT	GGA	GGA	GTC	AAG	GGA	TCG	GAA	AGT	GAT	GCT	TCA	CTG	ACT
M	I	G	G	V	K	G	S	E	S	D	A	S	L	T
GCA	TTT	GTA	CTC	ATT	GCA	TTG	GTT	GAG	TCT	AAG	GCT	GTT	TGC	TCT
A	F	V	L	I	A	L	V	E	S	K	A	V	C	S
CCT	TAC	ATA	GAG	AGT	TAT	GAA	GAC	AGC	ATC	CAA	AAA	GCA	GGA	
P	Y	I	E	S	Y	E	D	S	I	Q	K	A	G	
CAA	TAC	CTG	AAC	AAT	CGA	TTT	AAT	AAT	CTG	AAG	AGA	ACA	TAC	CCT
Q	Y	L	N	N	R	F	N	N	L	K	R	T	Y	P
GTG	ACA	ATA	ACA	GCC	TAT	GCA	CTT	TCA	TTA	GTG	AAC	ATC	AAC	AAA
V	T	I	T	A	Y	A	L	S	L	V	N	I	N	K
CTG	GAC	ACC	CTG	ATG	ATG	TTT	GCT	TCA	GAA	GAC	CAA	ACA	TAC	TGG
L	D	T	L	M	M	F	A	S	E	D	Q	T	Y	W
GCA	GAT	GCT	GGC											
A	D	A	G											

Figure 15. Nucleotide and deduced amino acid sequence obtained for clone C3 clone

P16A1.2/1 (Clone I) using the internal sense thiolester primer (CGEQNM).

TTT	AAT	GTG	GAC	TCG	GAC	AAT	GAT	TAT	CTA	CTG	ATG	GGG	CAG
F	N	V	D	S	D	N	D	Y	L	L	M	G	Q
GGA	TCA	GAC	TTG	TGG	AGT	GAT	GGT	AAA	AGG	ATC	AAC	TAT	CTG
G	S	D	L	W	S	D	G	K	R	L	N	Y	L
ATT	GGA	GAT	GGA	ACC	TGG	TTG	GAA	TTT	TGG	CCA	ACA	GAA	GTT
I	G	D	G	T	W	L	E	F	W	P	T	E	V
CAG	TGC	CAA	AAT	CAA	AAA	TAC	CGC	AGA	CTC	TGT	CAG	AGC	CTA
Q	C	Q	N	Q	K	Y	R	R	L	C	Q	S	L
GCA	GAG	TTT	TCT	GAA	AGC	TTA	ATG	ATT	AAT	GGC	TGC	CCG	ACT
A	E	F	S	E	S	L	M	I	N	G	C	P	T
Stop													
Codon	3'Untranslated Region →												
TAA	TAT	TTT	TGT	TTC	ATA	AAT	AAT	CAG	TGT	TAT	TTA	CAT	CCA
-	Y	F	C	F	I	N	N	Q	C	Y	L	H	P
GTA	ATC	CAA	AAT	AAA	CAA	TGC	TTT	CAT	GTC	GAA	TGA	TGT	TGA
V	I	Q	N	K	Q	C	F	H	V	E	-	C	-
AAC	TGA	TTT	ATC	TTA	CCT	GTG	TAC	AAT	GTT	ACA	AGG	TAA	GCT
N	-	F	I	L	P	V	Y	N	V	T	R	-	A
CCT	GTA	CGG	TCA	CGT	CAG	ACA	AGT	GCA	AGA	CTC	AGA	ATC	TGA
P	V	R	S	R	Q	T	S	A	R	L	R	I	-
ACT	ACA	ACG	ACA	GTT	CTT	CGA	AAT	GTG	TTT	ATT	GCC	AAT	ATG
T	T	T	T	V	L	R	N	V	F	I	A	N	M
TAT	GGC	TAT	CTA	GTT	ACC	TTG	ACT	TCA	TGC	TAA	ATT	CAG	TAG
Y	G	Y	L	V	T	L	T	S	C	-	I	Q	-
PolyAdenylation													
Signal →													
ATA	CCG	CAA	TAA	AAA	TTA	AAC	CAT	TAA	AAA	AAA	AAA	← M13 Forward Primer	
I	P	Q	-	K	L	N	H	-	K	K	K		

Figure 16. Nucleotide and deduced amino acid sequence of C3 clone P16A1.2/1 (Clone I) using M13 Forward Primer, indicating the position of the 3' stop codon, the 3'Untranslated Region (3'-UTR), and the polyadenylation signal.

Figure 17. Nucleotide and deduced amino acid sequence of internal sequence data obtained using an antisense primer (Primer I/AS) using as template the cloned C3 insert C316A1.2/1 isolated from the cDNA library.

TCA	ATG	ACC	ATT	CCA	GAC	ATA	TCC	ATG	CTG	ACA	GGA	TTT	GCT
S	M	T	I	P	D	I	S	M	L	T	G	F	A
CCA	GAT	ATG	CAA	GAC	CTT	GAA	TTG	CTT	CAA	AAT	GGA	GTT	GAT
P	D	M	Q	D	L	E	L	L	Q	N	G	V	D
AAA	TAT	ATA	TCA	AGA	TTT	GAA	CTA	GAC	AGG	GCA	TTA	TCA	GCC
K	Y	I	S	R	F	E	L	D	R	A	L	S	A
CAA	GGA	TCA	TTG	ATT	GTT	TAT	TTG	GAT	TCG	ATT	TCA	CAC	TTA
Q	G	S	L	I	V	Y	L	D	S	I	S	H	L
GAA	GAA	ACT	TGT	TTC	GGA	TTT	AAA	GTA	CAT	CAA	TTC	TTC	AAA
E	E	T	C	F	G	F	K	V	H	Q	F	F	K
GTT	GGA	CTT	ATT	CAA	CCA	GCT	TCT	GTT	AAA	ATT	TAT	GAA	TAT
V	G	L	I	Q	P	A	S	V	K	I	Y	E	Y
TAT	GAT	ACA	GAG	AAC	AGT	TGC	ACT	AAA	TTC	TAT	AAT	GTC	GAT
Y	D	T	E	N	S	C	T	K	F	Y	N	V	D
GAA	AGC	AGT	GCC	ATG	CTT	AGC	AAA	ATC	TGT	CAA	GGT	GAA	GTT
E	S	S	A	M	L	S	K	I	C	Q	G	E	V
TGC	AGA	TGT	GCA	GAA	GGA	AGT	TGT	ATA	TCA	GTT	AAG	TTG	CCT
C	R	C	A	E	G	S	C	I	S	V	K	L	P
GAA	CAC	CAT	ATC	ACT	TAC	ATA	GAT	CGT	GAG	GAT	AAA	GCA	TGT
E	H	H	I	T	Y	I	D	R	E	D	K	A	C
GAC	CAT	GGA	ACA	GAT	TAT	GTT	TAC	AAA	GTG	ACA	TTT	AAC	AAG
D	H	G	T	D	Y	V	Y	K	V	T	F	N	K
AAG	GAA	AAA	AGA	GAC	AGT	TAC	ATA	TAC	TAC	GAA	ATA	GAG	GTT
K	E	K	R	D	S	Y	I	Y	Y	E	I	E	V
AAG	TCT	GTC	GTA	AAG	CAA	GGA	AGC	GAT	GAC	ATA	TTG	GAT	GGA
K	S	V	V	K	Q	G	S	D	D	I	L	D	G

(figure 17 continued)

CAA GTC AGA GAA CTA ATC ATT CAT AAC AAC TGT GAA GGA ACA
Q V R E L I I H N N C E G T

TTT AAT GTG GAC TCG GAC AAT GAT TAT CTA CTG ATG GGG CAG
F N V D S D N D Y L L M G Q

GGA TCA GAC TTG TGG AGT GAT GGT AAA AGG ATC AAC TAT CTG
G S D L W S D G K R L N Y L

PRIMER I/AS

5'-TGGCACTGAACTTCTGTTGG-3'



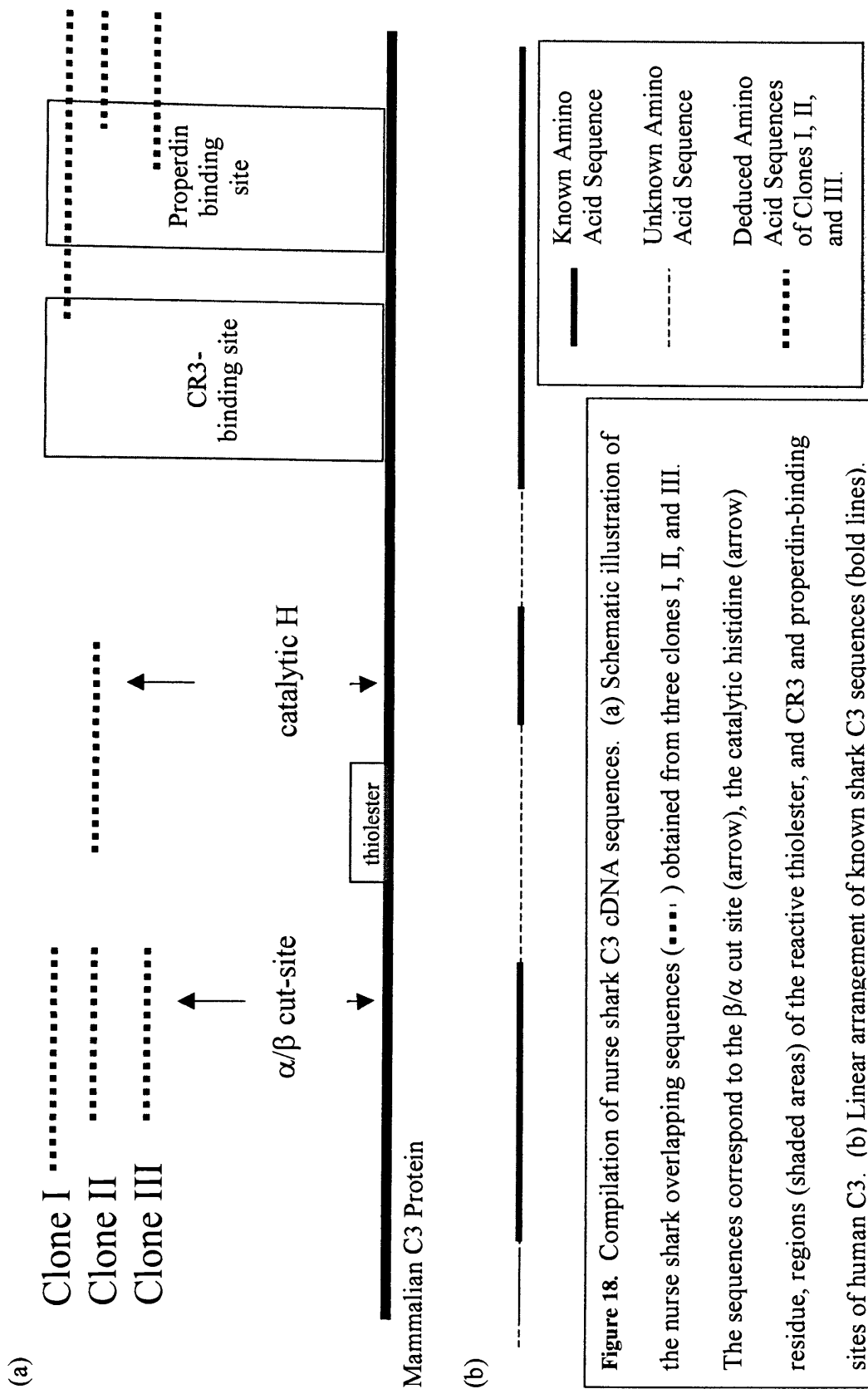
ATT GGA GAT GGA ACC TGG TTG GAA TTT TGG CCA ACA GAA GTT
I G D G T W L E F W P T E V

CAG TGC CAA
Q C Q

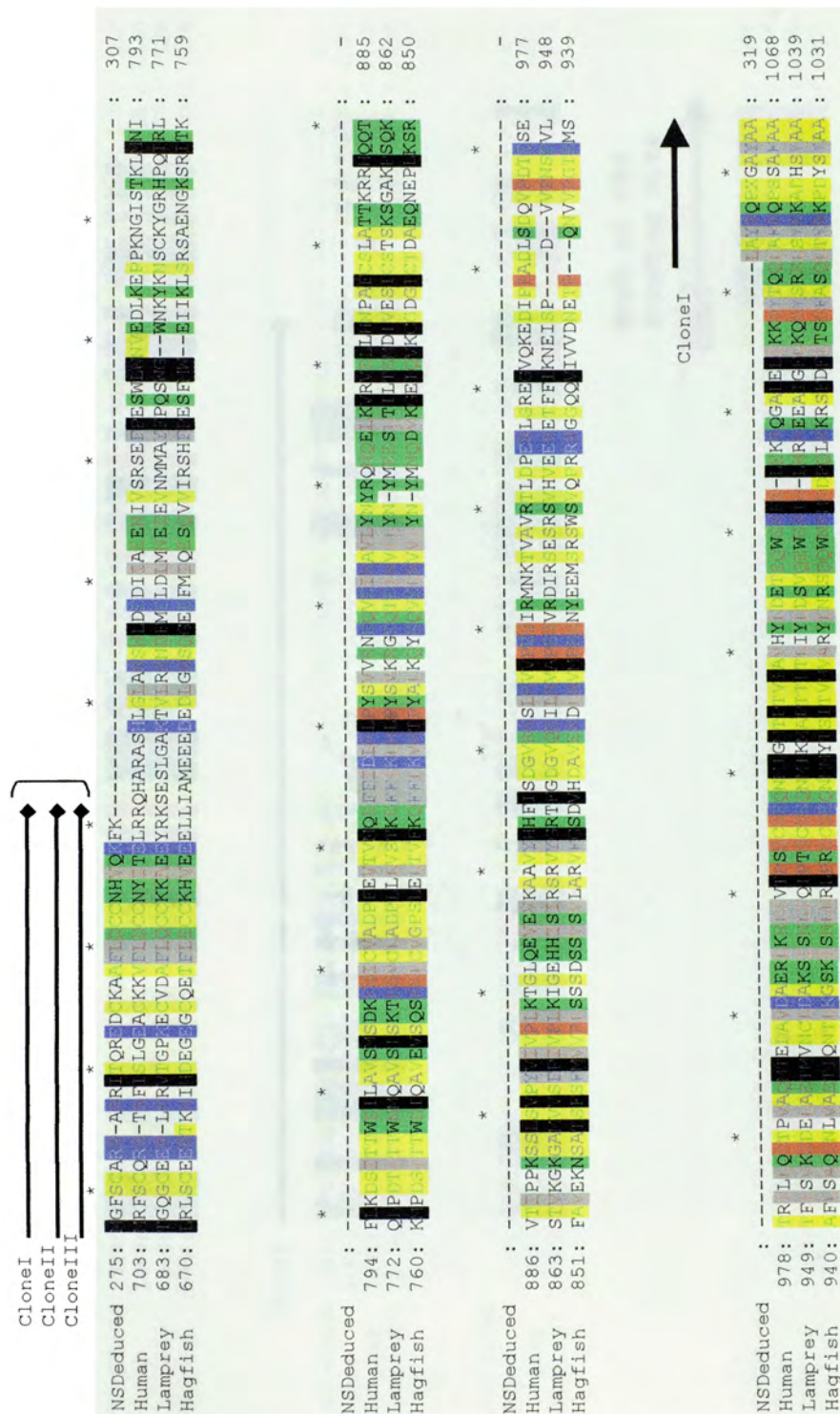
	Nurse Shark	Human	Chicken	Cobra	Rainbow Trout	Japanese Medaka	Carp C3-H1	Xenopus	Lamprey	Hagfish	Sea Urchin
Nurse Shark	100 100 100	50 53 43	49 50 44	51 45 43	40 45 36	38 45 32	43 48 35	46 45 43	31 36 31	28 35 27	24 34 27
Human		100 100 100	49 52 51	52 45 52	42 42 43	38 44 41	44 45 42	48 51 51	31 39 32	30 33 27	20 31 23
Chicken			100 100 100	53 59 51	44 46 40	36 49 39	49 50 51	47 56 49	34 37 31	33 35 27	19 33 21
Cobra				100 100 100	44 37 35	40 44 36	47 41 36	56 50 51	35 36 29	29 32 27	19 31 24
Rainbow Trout					100 100 100	51 54 41	52 56 50	40 43 37	29 33 29	26 33 26	19 26 24
Japanese Medaka						100 100 100	48 59 46	41 45 36	32 30 26	29 31 25	20 28 24
Carp C3-H1							100 100 100	39 51 38	31 37 28	28 37 24	18 32 21
Xenopus								100 100 100	33 35 32	29 35 28	20 30 24
Lamprey									100 100 100	30 35 28	21 29 24
Hagfish										100 100 100	19 32 19
Sea Urchin											100 100 100

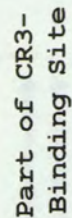
Table 4. Percent identity between the deduced amino acid sequence of the N-terminal region (N-term), thiolester region (TE), and C-terminal (C-term) sequence of the nurse shark (Ns) region with corresponding regions of C3 of several species. The three percentages in each box are as follows: N-terminal region (top), thiolester region (middle), and C-terminal region (bottom).

obtained for N-terminal, C-terminal, and thiolester (TE) region of nurse shark C3 clones were aligned and compared to corresponding regions of C3, or C3-like sequences of several vertebrate and invertebrate species (Table 4). Shark sequences showed the highest sequence identity with Cobra (Co) C3, Human (Hu) C3 and chicken (Ch) C3 for the N-terminal, TE region and C-terminal regions, respectively. Figure 18 displays a diagrammatic illustration of the biologically significant regions of human C3 which correspond to sequences that have been identified in the nurse shark from sequence data obtained from Clones I, II, and III. Using Clustal X and Genedoc software shark sequences were aligned with corresponding sequences of different species such as human, lamprey, and hagfish to show residues of similar physiochemical properties (figure 19). Significant similarity was observed. The deduced amino acid sequence of C3 I, II, and III composite was also aligned with deduced amino acid sequence of Clone C3A4, used as the probe to screen the library, and partial cDNA clone NsC31 which had previously been isolated and overlaps in sequence with partial cDNA clone C3A7, and which lacks the catalytic histidine residue (figure 20). An alignment of the N-terminal amino acid sequence (over 222 residues) of the α -chain of C3 isolated from shark serum with the deduced amino acid sequence of NsC3 clone show more than 95% identity (figure 20). This suggests that partial C3 cDNA C3A4 represents the shark gene that encodes for the functional C3 protein isolated from serum (Dodds, et al., 1998).

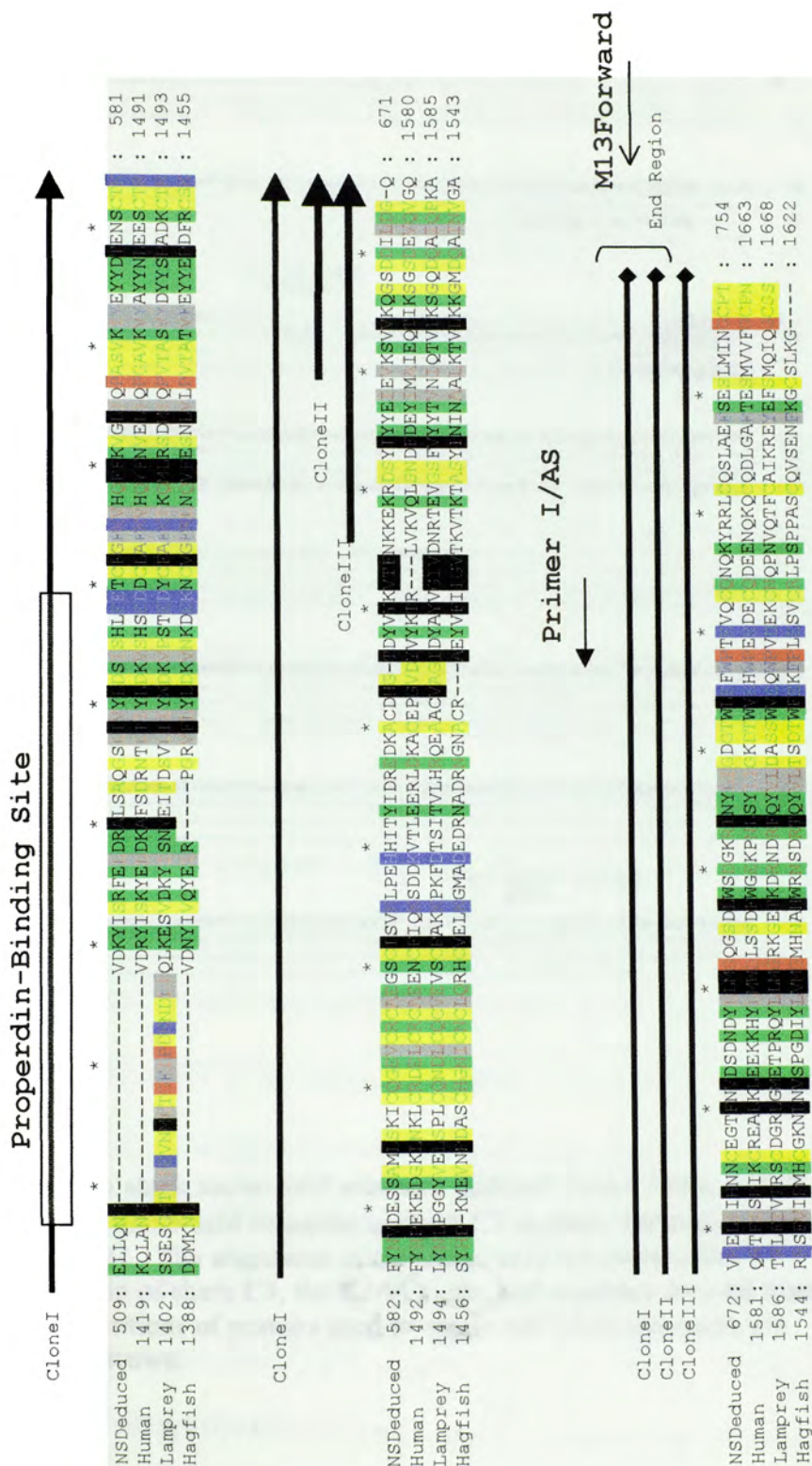


(Figure 19 continued)



Catalytic
Histidine

(Figure 19 continued)



M13REV
→

```

NSDeduced : -----SFGLCFEGTRFRPARIRHEGGSVNYLHIGIQFAELNFGSNLVVHLSLNDNDADVQNGIKYFSYMLINKGHIIVYGRIRREIQGLV : 64
C3A4       : ----- : -
C3ALPHA    : ----- : -
NSC31      : ----- : -

```

Primer I/S
→

```

NSDeduced : VNMLIFITSDLIFFSRLINYYLKGPALEMVADSVMVDNDTCMGKLVSAANEHDENNIYKFGKQFQLNITGDSGATVGLVAVDKAVFVLN : 177
C3A4       : ----- : -
C3ALPHA    : ----- : -
NSC31      : ----- : -

```

```

NSDeduced : KENKLTQSKIWDIVEKNDIGCTAGSGSDVLGVFTDAGLTIVTSSNLKTATRTDLTCKQFMKRRRRATRLMDVKMAKLNQYKAKILNKKCNDGM : 270
C3A4       : ----- : -
C3ALPHA    : -----ATXLMDEVMAKLNQYKAKILNK----- : 20
NSC31      : ----- : -

```

CGEQNM
→

```

NSDeduced : KKNFMGFSCARRARRITQREDCKAAFLDCGNHVQKFK----- A QFX : 215
C3A4       : -----CGEQNMASITFGVIVTNFLDKTRQWERVGMCKRETSVKNIQTGYAQQ A QPS : 55
C3ALPHA    : ----- : -
NSC31      : -----IAQGYAQQ T RDD : 16

```

```

NSDeduced : A NRE VVKVFAMSENLIVIDFNVLGGAQWLTLNKQKFDGSKEDAPVIHGEMIGGVKGSSESASLTAFVLIALVESKAVT : 408
C3A4       : A NRE ----- : 70
C3ALPHA    : ----- : -
NSC31      : S HIP VLKVLAMSSDLISIDVNHLCYAAWLINSKQLIEGRFKEDAPVLYLKTMTGGGLGDAENDAVLTAFLVLIAMFEADDX : 409

```

```

NSDeduced : SFYIESYEDSIQKAGQYLNRRFNLRNFKTYFVITAYASLVNINKNLDTIMXFASEDQTIWADAG----- : 471
C3A4       : ----- : -
C3ALPHA    : ----- : -
NSC31      : CSTXVENYKNSIAKAGDYLDKK----- : 131

```

```

NSDeduced : NSLFINVCAQYRXDSDSMTIFDISMLTGFAPDMQDLLELLQNGVDKYSISRFELDRALSAQGSILIVYLDSSISHLEETCFGFKVHQFFKVGLIQP : 564
C3A4       : ----- : -
C3ALPHA    : ----- : -
NSC31      : ----- : -

```

```

NSDeduced : ASVKIYEYDTSNCTKFINVDSSAMLSKIQGEVCRCAEGSCISVKLFEHHITYIDREDKACDHGTDVYKVTENKKEKRDYSIYYEIEVK : 657
C3A4       : ----- : -
C3ALPHA    : ----- : -
NSC31      : ----- : -

```

Primer I/AS
←

```

NSDeduced : SVVKQGSDDILDGQVRELIHNNCEGTFINVDSNDNYLLMGQGSDLWSDGKRINYLIGDGTWLEFWPTEVQCQNKYRRLCISLAEFSESIMIN : 751
C3A4       : ----- : -
C3ALPHA    : ----- : -
NSC31      : ----- : -

```

M13REV
←

```

NSDeduced : GCFT : 754
C3A4       : ---- : -
C3ALPHA    : ---- : -
NSC31      : ---- : -

```

Figure 20. Alignment of shark amino acid sequence deduced from C3 clones (I, II, and III) with N-terminal amino acid sequence of shark C3 α -chain, deduced amino acid of nsC3A4 and nsC31. The alignment of the amino acid sequence of the N-terminal of the alpha chain of shark C3, the C3A4 clone, and sequence derived from nsC31 is also shown. Positions of primers used to obtain nsC3A16 sequences are indicated by horizontal arrows.

V. Discussion

To comprehensively understand the evolution of a complex system such as complement it is essential that complement genes from species representing a variety of phyla, be cloned and sequences compared. In addition, protein sequence data of complement proteins also provide insight into the genetic and functional relationships. The shark, a primitive vertebrate, holds an important phylogenetic position for study of evolution of complement proteins, particular the thiolester proteins C3, C4, and C5 which are believed to have arisen by gene duplication (Sottrup-Jensen, et al., 1985). The point of divergence of these proteins from a primordial ancestor remains undetermined. Information from the shark will be a significant contribution to determining the origin of these proteins. Complement genes have been cloned, and complement proteins have been isolated and characterized from several species (Sunyer et al., 1980; Nonaka, et al., 1981; Kai, et al., 1983; Ishiguro, et al., 1992; Fritzinger, et al., 1994; Mavrodis, et al., 1995; Smith, et al., 1996; Sunyer, et al., 1997; Al-Sharif, et al., 1998; Nakao, et al., 2000). Characterization of shark complement genes is still in its infancy. Protein sequence data, however, is available for shark C3 and C4 homologues (Dodds, et al., 1998; Smith, 1998). These earlier studies provided N-terminal sequence data on the α and β chain of C3 and the α , β and γ chain of C4 of the shark. The shark C3 protein was biologically functional and contained an active thiolester as determined by its opsonic activity and incorporation of ^{14}C -methylamine (Dodds, et al., 1998). Previous studies on the nurse shark, using degenerate primers and employing RT-PCR yielded two distinct amplification products, that is, two partial cDNA clones, C3A4 and C3A7. Using gene specific primers and performing 5'RACE PCR a 1.8 kb PCR product was obtained and

cloned (nsC31). Sequence data obtained from nsC31 revealed that the sequence overlapped with the partial cDNA clone C3A7. Furthermore, it was determined that the crucial catalytic histidine at position 1126 was absent and was replaced with a tyrosine residue (figure 20).

As the sequence of clone nsC31 overlapped with that of C3A7 rather than with C3A4, and as the catalytic histidine (absent from nsC31) is important for the activation of the C3 protein (since it is believed to be involved in the thiolester binding of C3 to hydroxyl and amine groups) it seemed reasonable to assume that the partial cDNA clone C3A4 represents the gene that encodes for the functional shark C3 protein, previously isolated and characterized from serum. The isolation of biologically active C3 protein, the identification of two distinct C3 clones (C3A4 and C3A7), and the identification of NsC31 sequence lacking the critical histidine residue suggests that in the shark there are at least two C3 genes encoding two protein molecules both expressed in the shark with only one probably capable of being an active opsonic molecule. The C3A4 clone is believed to represent the gene encoding the biologically active protein whereas C3A7 is believed to represent a gene expressing a form of the protein that probably lacks an active thiolester, however, it may still be capable of other biological activities, such as release of anaphylatoxin fragment (C3a) upon cleavage of the α chain. While sequence corresponding to the C3 convertase cleavage (activation) site has not been obtained sequence data does suggest (from the conserved cysteine residues) that an active anaphylatoxin fragment is released upon C3 activation.

In this study C3A4 was used as a probe for screening the cDNA library since we were interested in cloning the gene encoding the biologically active thiolester containing

C3. The construction of the cDNA library was successful resulting in 1.15×10^{16} pfu. Plaque lifts were made of the amplified library and screened for the identification of potential C3-like clones using the C3A4 plasmid as a probe by labelling it with digoxigenin. Screening of the library with Digoxigenin labeled C3A4 probe resulted in the identification of four positive C3-like cDNA clones: 16A1.2-1 (I), 16A1.2-2 (II), 16A1.3-1 (III), and 20A1.3-2 (IV). Three of the four clones (16A1.2-1, 16A1.2-2, and 16A1.3-1) upon analysis exhibited 98% identity to each other and were considered to represent the same gene. Interestingly, these three similar clones came from the same master plate number 16. The nucleic acid sequences of the three similar clones, 16A1.2-1, 16A1.2-2, and 16A1.3-1 were translated to deduced amino acid sequence using Genedoc software. New primers were designed using the online Primer3 program (available at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) to obtain internal sequence of the C3 clones. Newly designed primers from the 5' and 3' ends and a degenerate primer made previously from the conserved thiolester region (CGEQNM) were used to obtain internal sequence data.

The fourth clone which had an insert size of 4.5 kb contained many substitutions and was significantly different in sequence from the remaining three C3 clones (figure 10 and figure 11). It is difficult to say if this clone represents a third C3 sequence (i.e. distinct from C3A7 and nsC31), that is as yet unidentified, or if it represents mutations that occurred in the synthesis of the library that led to this sequence being so different. BLAST analysis of this gene sequence showed sequence similarity only with Chicken C3 and with several different reading frames. Since this sequence was so different and because presently it has not been determined whether it represents contamination we

decided to focus on the analysis of the 3 similar clones (I, II, and III) that originated from plate 16.

A compilation of all sequence data obtained for the C3 clones was analyzed and alignments made. Alignments using Genedoc software determined that there was 48.7% identity and 66.3% similarity with human C3 over the regions compared. Lower level of identity and similarity was found to be with the hagfish C3, which was 30% and 49.6% respectively. A total of 754 deduced amino acid residues have been identified and the sequence data shows overlaps with the sequence of partial C3 cDNA clone, C3A4 (which was used to screen the library) (figure 3). Amino acid sequence for the N-terminal and C-terminal ends, and the thiolester region respectively show 50%, 43%, and 53% sequence identity with corresponding regions of human C3. The regions (using human C3 numbering system) span from amino acid residue: 432 to 736, 1057 to 1224, and 1382 to 1663.

The nurse shark C3 gene sequence has characteristics in common with the human C3 which include sequences representing several biologically active domains and indicating molecular subunit structure. The CR3 (Complement Receptor 3) binding domain through which binds iC3b binds to the receptor and the properdin binding site are similar in sequence and position to those found in human C3 (figure 18 and 19). The β/α cut site, RRRR, that is cleaved in the post-translational modification of the pro-C3 molecule into the α and β chains confirms its 2-chain structure (figure 14 and 19). Furthermore, the presence of this cut-site shows the molecule's closer relationship to C3 rather than the α -2 macroglobulin which lacks the β/α cut site, and is a smaller molecule consisting of a single chain. The catalytic histidine residue which previous studies have

shown to be missing in clone NsC31 (representing C3A7), is present in the sequence deduced from this study, indicating that the thiolester is most likely biologically functional in the protein encoded by C3 gene corresponding to clones I, II, and III. An active thiolester is present in the α -chain of the C3 protein isolated from nurse shark serum (Whitehead, et al., 1982; Fong, et al., 1990).

The total number of nucleotides sequenced from the three similar C3 clones (I, II, and III) isolated in this study so far is 2262 bp and the size of the clones is estimated to be approximately 3500 bp. We believe that the C3 clones isolated from the liver cDNA library do not represent clones of the entire C3 gene since its size is speculated to be approximately 5.0 kb, similar to that of human C3 gene. While the size of the nurse shark C3 gene(s) is estimated to be at 5 kb, 3.5 kb of the shark C3 gene was cloned from the cDNA library. Further studies should include the design and synthesis of new gene specific primers based on known sequence to obtain the remaining sequence of shark C3 gene. The complete sequence of the 5' end of the nurse shark gene was not obtained from our partial cDNA clones, however, with the 5' sequence data obtained, 5' RACE PCR can be employed to provide additional sequence data on the 5' end sequence. Alternatively, degenerate primers based on the protein sequence of the N-terminal of the shark C3 β -chain may also be used to perform RACE PCR that will permit obtaining the remainder of the C3 gene sequence, providing entire gene sequence data.

Problems encountered in this study arose when attempting to estimate the size of the cloned inserts. The size could not be determined using PCR since optimal conditions were not identified. When the clones were amplified in order to identify their sizes, upon electrophoretic analysis only 2kb products were observed. This did not correlate with the

amount of sequence data obtained which was over 2.2 kb. The reason for the discrepancy in size estimates obtained from performing PCR may be that optimal conditions for PCR were not provided for amplification of the large size of the clone. In this study, as an alternative we decided to use restriction enzymes to digest the insert from the vector and to estimate an approximate size of the clones based on restriction fragment size, which provided more reliable size data.

In this study, the isolation of cDNA sequences has now contributed more information on the sequence of at least one of the nurse shark C3 genes and has further established that there are at least two separate C3 genes in the shark. Since the entire C3 gene is estimated to be 5 kb and only sequence data collectively amounting to approximately 3 kb has been obtained, further studies dedicated to sequencing the entire cloned C3 gene to obtain the entire 5' end region will provide the entire nucleotide and deduced amino acid sequence. After identification of the entire gene sequence a complete phylogenetic analysis may be done that would provide more insight into the evolution of this pivotal protein of the complement system. The information derived from the present study on shark C3 gene(s) is summarized in Table 5.

Amino Acid Residues of Biologically Significant Domains of Human C3	Homologous Regions Present in the Nurse Shark C3 Protein *
# of C3 genes = 1	# of C3 genes = 2
1-22 leader peptide	
23-667 β chain	✓ 432-667
668-671 β/α cut-site (RRRR)	✓ 668-671
672-1663 α chain	✓ 672-735 ✓ 1057-1224 ✓ 1399-1663
672-748 C3a anaphylatoxin	✓ 672-735
955-1303 C3dg fragment	✓ 1057-1224
955-1001 C3g fragment	
1002-1303 C3d fragment	✓ 1057-1224
1303-1321 C3f fragment	
1010-1013 Thiolester domain	✓ based on probe sequence
1424-1456 Properdin binding site	✓ 1424-1456
1209-1271 Factor H-binding site	
749-790 Factor H, B and CR1 binding site	
1227-1236 CR2-binding site	
1383-1403 CR-3-binding site	✓ 1399-1403

* Based on deduced amino acid sequence data.

Table 5. Homologous regions present in the nurse shark C3 protein of biologically significant domains of C3 in humans.

VI. References

- Al-Sharif, W.Z., Sunyer, J.O., Lambris, J.D. and Smith, L.C. (1998) Sea urchin coelomocytes specifically express a homologue of the complement component C3. *Journal of Immunology*, **160**, 2983-2997.
- Bhakdi, S. and Trandum-Jensen, J. (1983) Membrane damage by complement. *Biochemistry et Biophysics Acta*, **737**, 343-372.
- Bokisch, V.A., Muller-Eberhard, H.J. and Cochrane, C.G. (1969) Isolation of a fragment of (C3a) of the third component of human complement containing anaphylatoxin and chemotactic activity and description of an anaphylatoxin inactivator of human serum. *Journal of Experimental Medicine*, **129**, 1109-1130.
- Day, N.K.B. Gewurz, H., Johannsen, R., Finstad, J. and Good, R.A. (1970) Complement and complement-like activity in lower vertebrates and invertebrates. *Journal of Experimental Medicine*, **132**, 941-950.
- Dodds, A.W. and Law, S.K.A. (1998) The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and α_2 -macroglobulin. *Immunological Reviews*, **166**, 15-26.
- Dodds, A.W., Smith, S.L., Levine, R.P. and Willis, A.C. (1998) Isolation and initial characterisation of complement components C3 and C4 of the nurse shark and the channel catfish. *Developmental and Comparative Immunology*, **22**, 207-216.
- Dodds, A.W., Ren, X., Willis, A.C. and Law, S.K.A. (1996) The reaction mechanism of the internal thioester in the human complement component C4. *Nature*, **379**, 177-179.
- Fishelson, Z., Pangburn, M.K. and Muller-Eberhard, H.J. (1984) Characterization of the initial C3 convertase of the alternative pathway of human complement. *Journal of Immunology*, **132**, 1430-1434.
- Fishelson, Z. (1991) Complement C3: A molecular mosaic of binding sites. *Molecular Immunity*, **28(4)**, 545-552.
- Fritzing, D.C., Petrella, E.C., Connelly, M.B., Bredehorst, R. and Vogel, C. (1992). Primary structure of cobra complement C3. *Journal of Immunology*, **149**, 3554-3562.
- Fritzing, D.C., Bredehorst, R. and Vogel, C. (1994) Molecular cloning and derived primary structure of cobra venom factor. *Proceedings of the National Academy of the Sciences: USA*, **91**, 12775-12779.

- Fujii, T., Nakamura, T., Sekizawa, A. and Tomonaga, S. (1992) Isolation and characterization of a protein from hagfish serum that is homologous to the third component of the mammalian complement system. *Journal of Immunology*, **148**, 117-123.
- Fujii, T., Nakamura, T. and Tomonaga, S. (1995) Component C3 of hagfish complement has a unique structure: identification of native C3 and its degradation products. *Molecular Immunology*, **32:9**, 633-642.
- Gotze, O. and Muller-Eberhard, H.J. (1970) Lysis of erythrocytes by complement in the absence of antibody. *Journal of Experimental Medicine*, **132**, 898-915.
- Green, H., Fleischer, R.A., Barrow, P. and Goldberg, B. (1959) The cytotoxic action of immune gamma globulin and complement on Krebs ascites tumor cell. II- Chemical studies. *Journal of Experimental Medicine*, **109**, 511-521.
- Hanley, P.J., Hook, J.W., Raftos, D.A., Gooley, A.A., Trent, R. and Raison, R.L. (1992) Hagfish humoral defense protein exhibits structural and functional homology with mammalian complement components. *Proceedings of the National Academy of the Sciences: USA*, **89**, 7910-7914.
- Harrison, R.A. and Lachmann, P.J. (1980) The physiological breakdown of the third component of human complement. *Molecular Immunology*, **17**, 9-20.
- Howard, J.B. (1980) Methylamine reaction and denaturation-dependent fragmentation of complement component 3: comparison with α -2 macroglobulin. *Journal of Biological Chemistry*, **255**, 7082-7084.
- Humphrey, J.H. and Dourmashkin, R.R. (1969) The lesions in cell membranes caused by complement. *Advances in Immunology*, **11**, 75-115.
- Ishiguro, H., Kobayashi, K., Suzuki, M., Titani, K., Tomonaga, S. And Kurosawa, Y. (1992) Isolation of a hagfish gene that encodes a complement component. *EMBO Journal*, **11**, 829-837.
- Jensen, J., Festa, E., Smith, D.S. and Cayer, D. (1981) The complement system of the nurse shark: Hemolytic and comparative characteristics. *Science*, **214**, 566-569.
- Kai, C., Yasuhiro, Y., Yamanouchi, K. And Okada, H. (1983) Isolation and identification of the third component of complement of Japanese quails. *Journal of Immunology*, **130:6**, 2814-2820.

- Lachmann, P.J. and Thompson, R.A. (1970) Reactive lysis: The complement-mediated lysis of unsensitized cells. II. The characterization of activated rector as C5, C6, and the participation of C8 and C9. *Journal of Experimental Medicine*, **131**, 643-657.
- Lambris, J.D. and Muller-Eberhard, H.J. (1986) The multifunctional role of C3: structural analysis of its interactions with physiological ligands. *Molecular Immunology*, **23**, 1237-1242.
- Lambris, J.D. (1988) The multifunctional role of C3, the third component of complement. *Immunology Today*, **9:12**, 387-393.
- Lambris, J.D., Lao, Z., Pang, J. and Alsenz, J. (1993) Third component of trout complement: cDNA cloning and conservation of functional sites. *Journal of Immunology*, **34:11**, 6123-6134.
- Lambris, J.D. (1993) The chemistry, biology and phylogeny of C3. *Complement Today: Complement Profiles*, **1**, 16-45.
- Liszewski, M.K. and Atkinson, J.P. (1993) The complement system. In *Fundamental Immunology, Third Edition*. Chapter 26. William E. Paul ed. New York: Raven Press, Ltd., 917-937.
- Lundwall, A., Wetsel, R.A., Domdey, H., Tack, B.F. and Fey, G.H. (1984) Structure of murine complement component C3: I. Nucleotide sequence of cloned complementary and genomic DNA coding for the β chain. *Journal of Biological Chemistry*, **259:22**, 13851-13856.
- Mavroidis, M., Sunyer, J.O. and Lambris, J.D. (1995) Isolation, primary structure, and evolution of the third component of chicken complement and evidence for a new member of the α_2 -macroglobulin family. *Journal of immunology*, **154**, 2164-2174.
- Medicus, R.G., Schreiber, R.D., Gotze, O. and Muller-Eberhard, H.J. (1976) A molecular concept of the properdin pathway. *Proceedings of the National Academy of the Sciences: USA*, **73**, 612-616.
- Nakao, M. and Yano T. (1998) Structural and functional identification of complement components of the bony fish, carp (*Cyprinus carpio*). *Immunology Reviews*, **166**, 27-38.

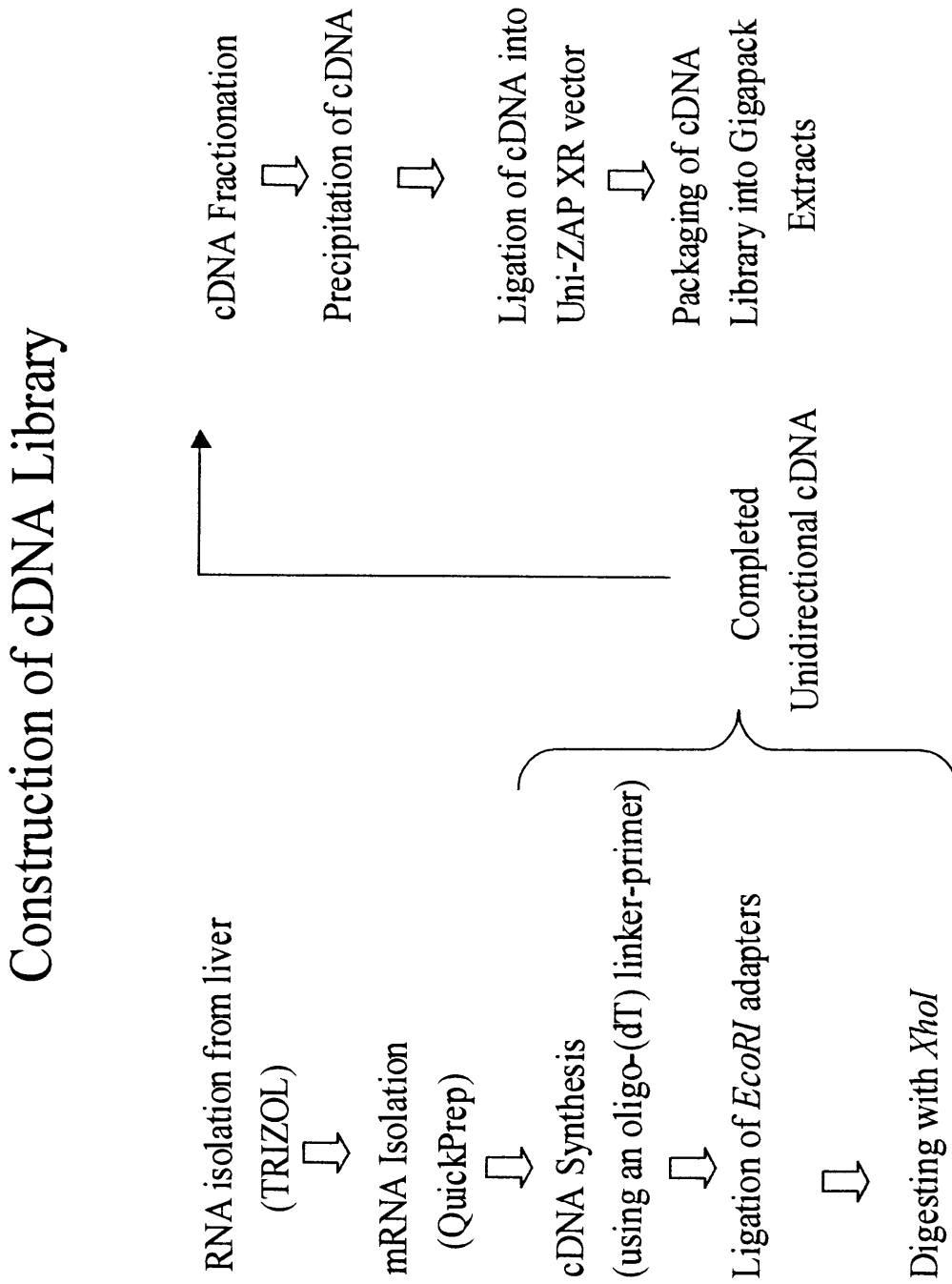
- Nakao, M., Mutsuro, J., Obo, R., Fujiki, K., Nonaka, M. And Yano, T. (2000) Molecular cloning and protein analysis of divergent forms of the complement component C3 from a bony fish, the common carp (*Cyprinus carpio*): presence of variants lacking the catalytic histidine. *European Journal of Immunology*, **30**, 858-866.
- Nonaka, M., Yamaguchi, N., Natsuume-Sakai, S. and Takahashi, M. (1981) The complement system of rainbow trout (*Salmo gairdneri*): I. Identification of the serum lytic system homologous to mammalian complement. *Journal of Immunology*, **126**:4, 1489-1494.
- Nonaka, M., Fujii, T., Kaidoh, T., Natsuume-Sakai, S., Nonaka, M., Yamaguchi, N. and Takahashi, M. (1984) Purification of a lamprey complement protein homologous to the third component of the mammalian complement system. *Journal of Immunology*, **133**:6, 3242-3249.
- Nonaka, M. and Takahashi, M. (1992) Complete complementary DNA sequence of the third component of complement of lamprey: Implication for the evolution of thioester containing proteins. *Journal of Immunology*, **148**, 3290-3295.
- Pangburn, M.K. and Muller-Eberhard, H.J. (1980) Relation of a putative thioester bond in C3 to activation of the alternative pathway and the binding of C3b to biological targets of complement. *Journal of Experimental Medicine*, **152**, 1102-1114.
- Pangburn, M.K. (1988) Initiation and activation of the alternative pathway of complement. In *Cytolytic lymphocytes and complement*, Vol. I. E. Podack ed. Florida: CRC Press, Inc., 41-52.
- Reid, K.B.M. (1986) Activation and control of the complement system. *Advances in Immunology*, **50**, 433-464.
- Reid, K.B.M. and Porter, R.R. (1981) The proteolytic activation systems of complement. *Annual Reviews of Biochemistry*, **50**, 433-464.
- Ross, G. and Jensen, J. (1973) The first component (C1n) of the complement system of the nurse shark (*Ginglymostoma cirratum*). *Journal of Immunology*, **111**:1, 175-182.
- Sim, R.B., Twose, T.M., Paterson, D.S. and Sim, E. (1981) The covalent-binding reaction of complement component C3. *Journal of Biochemistry*, **193**, 115-127.
- Smith, S.H. and Jensen, J.A. (1986) The second component (C2n) of the nurse shark complement system: purification, physico-chemical characterization and functional comparison with guinea pig C4. *Developmental and Comparative Immunology*, **10**, 191-206.

- Smith, S.L. (1998) Shark complement: An assessment. *Immunological Reviews*, **166**, 67-68.
- Sottrup-Jensen, L., Hansen, H.F., Mortensen, S.B., Peterson, T.E. and Magnusson, S. (1981). Sequence location of the reactive thiolester in human α_2 -macroglobulin. *FEBS Letters*, **123**, 145-148.
- Sottrup-Jensen, L., Stepanik, T.M., Kristensen, T., Lonblad, P.B., Jones, C.M., Wierzbicki, D.M., Magnusson, S., Domdey, H., Wetsel, R.A., Lundwall, A., Tack, B.F. and Fey, G.H. (1985) Common evolutionary origin of α_2 -macroglobulin and complement components C3 and C4. *Proceeding of the National Academy of the Sciences: USA*, **82**, 9-13.
- Starkey, P.M. (1981) Evolution of human α_2 -macroglobulin. *Acta. Biol. Med. Germ.*, **40**, 1555-1559.
- Sunyer, J.O., Zarkadis, I.K., Sahu, A. and Lambris, J.D. (1996) Multiple forms of complement C3 in trout that differ in binding to complement activators. *Proceedings of the National Academy of the Sciences: USA*, **93**, 8546-8551.
- Sunyer, J.O., Tort, L. and Lambris, J.D. (1997) Structural C3 diversity in fish: Characterization of five forms of C3 in the diploid fish *Sparus aurata*. *Journal of Immunology*, **158**, 2813-2821.
- Sunyer, J.O., Tort, L. and Lambris, J.D. (1997) Diversity of the third form of complement, C3, in fish: Functional characterization of five forms of C3 in the diploid fish *Sparus aurata*. *Biochemistry Journal*, **326**, 877-881.
- Sunyer, J.O. and Lambris, J.D. (1998) Evolution and diversity of the complement system of poikilothermic vertebrates. *Immunological Reviews*, **166**, 39-57.
- Swenson, R.P. and Howard, J.B. (1979) Characterization of alkylamine sensitive site in α_2 -macroglobulin. *Proceedings of the National Academy of Sciences: USA*, **76**, 4313-4316.
- Swenson, R.P. and Howard, J.B. (1980) Amino acid sequence of the tryptic peptide containing the alkylamine-reactive site from human α_2 -macroglobulin. *Journal of Biological Chemistry*, **255**, 8087-8091.
- Terado, Tokio, Smith, S.L., Nakanishi, T., Nonaka, M.I., Kimura, H. and Nonaka, M. (2001) Occurrence of structural specialization of the serine protease domain of complement factor B at the emergence of jawed vertebrates and adaptive immunity. *Immunogenetics*, **53**, 250-254.

- Vik, D.P., Amiguet, P., Moffat, G.J., Fey, M., Amiguet-Barras, F., Wetsel, R.A. and Tack, B.F. (1991) Structural features of the human C3 gene: Intron/exon organization, transcriptional start site, and promoter region sequence. *Biochemistry*, **30**, 1080-1085.
- Wetsel, R.A., Lundwall, A., Davidson, F., Gibson, T., Tack, B.F., and Fey, G.H. (1984) Structure of murine complement component C3: II. Nucleotide sequence of cloned complementary DNA coding for the α chain. *Journal of Biological Chemistry*, **259:22**, 13857-13862.
- Whitehead, A.S., Solomon, E., Chambers, S., Bodmer, W.F., Povey, S. and Fey, G. (1982) Assignment of the structural gene for the third component of human complement to chromosome 19. *Proceeding of the National Academy of the Sciences: USA*, **79**, 5021-5025.
- Fong, K.Y., Botto, M., Walport, M.J. and So, A.K. (1990) Genomic organization of human complement component C3. *Genomics*, **7**, 579-586.
- Zarkadis, I.K., Sarrias, M.R., Sfyroera, G., Sunyer, J.O. and Lambris, J.D. (2001) Cloning and structure of three rainbow trout C3 molecules: A plausible explanation for their functional diversity. *Developmental and Comparative Immunology*, **25**, 11-24.

VII. APPENDIX

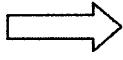
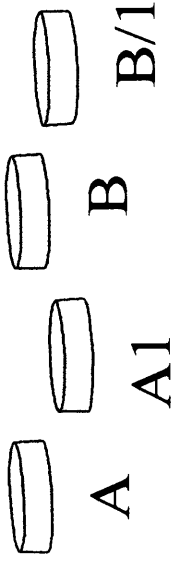
Summary of Materials and Methods



Titration, Amplification, and Primary Screening of Library

Plate and Titre Library

After mixing cDNA library with host bacteria



Amplification of cDNA library



Incubate at 37°C for 7.5 hours

Perform Plaquelifts from Amplified Library and Prepare for hybridization

↓

Label C3A4 probe using Digoxigenin PCR

↓

Hybridize with C3A4 Probe

↓

Detect Hybridization

Isolation and Sequencing of C3-Like cDNA Clones

Identify Plaques that hybridized with C3A4 probe



Extract Plaques as an



Agarose Plug



Perform Secondary Screening



In vivo excision of single clones



Transformed *E. coli* with excised clones gave positive hybridization

Mini-Prep to isolate cDNA clones



Big Dye-T Cycle Sequence using M13 F and M13R primers



Confirm C3-Like clones via Blast program from NCBI, perform multiple sequence alignments (ClustalX, Genedoc).



Digest samples w/*HindIII* and *EcoRI*



Design new primers, cycle sequencing, and sequence analysis as stated previously.