7-17-2002

Immunophylogenetic aspects of a Gorgonian coral

Larry J. Dishaw
Florida International University

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

IMMUNOPHYLOGENETIC ASPECTS OF A GORGONIAN CORAL

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGY

by

Larry J. Dishaw

2002

To: Dean Arthur W. Herriott
College of Arts and Sciences

This dissertation, written by Larry J. Dishaw, and entitled Immunophylogenetic Aspects of a Gorgonian Coral, having been approved in respect to style and intellectual content, is referred to you for judgment.

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

Victor Apanius

Timothy Collins

Lidia Kos

Sylvia L. Smith

Martin L. Tracey, Jr.

Charles H. Bigger, Major Professor

Date of defense: July 17, 2002

The dissertation of Larry J. Dishaw is approved.

Dean Arthur W. Herriott
College of Arts and Sciences

Dean Douglas Wartzok
University Graduate School

Florida International University, 2002

DEDICATION

To my daughter, Emma. Anything for you, always.

ACKNOWLEDGMENTS

I would like to first thank my committee members for their support and guidance. I am especially grateful to my major professor, Charles H. Bigger, for introducing me into the (sometimes) controversial field of comparative immunology. Our discussions on science and immune evolution has helped develop my scientific passions and future ambitions. My personal experiences as a graduate student at FIU has helped me develop a strong, independent character. For this, I appreciate the challenges which helped motivate my maturity. In addition, I am indebted to the many students and other colleagues I have had the great pleasure in meeting and interacting with here in FIU-Biology and at scientific meetings. I have developed what I hope will continue as long-term collaborations with some very special parties.

I would like to thank past and present members of my lab, whose sincere passion for science and good humor has made each day enjoyable and truly unforgettable. There are a few special friends (you know who you are) which, through countless discussions, have helped my development as a scientist and much of the thinking involved in this work. I would like to thank my wife and friend Jessica for her continued love, support, and patience. Jessica is a genuinely amazing person who is a remarkable pediatrician, yet remains a wonderful and dedicated mommy. Thank you.

A very special thanks to L. Scott Quackenbush, with whom I became a biologist.

In memory of my original mentor, L. P. Tosco. Because of you.

ABSTRACT OF THE DISSERTATION

IMMUNOPHYLOGENETIC ASPECTS OF A GORGONIAN CORAL

by

Larry J. Dishaw

Florida International University, 2002

Miami, Florida

Professor Charles H. Bigger, Major Professor

One goal of comparative immunology is to derive inferences about evolutionary pathways in the development of immune-defense systems. Almost 700 million years ago, a major divergence occurred in the phylogeny of animals, spitting all descendants into either the protostome or deuterostome (includes vertebrates) lineages. Genes have evolved independently along these lineages for that amount of time. Cnidarians originated before that divergence event, and can hold clues as to which immune response genes are homologous to both lineages. This work uses the gorgonian coral, *Swiftia exserta*, for two major reasons: 1) because of their phylogenetic position, corals are an important animal model in studies concerning the phylogeny of immune-response genes, and 2) nothing is known about the genes controlling immunocompetence in corals. The work described here has important implications in both innate and adaptive immunity.

The vertebrate complement system is a major component of innate immunity. C3 is a critical component of the three pathways of complement. Because of its opsonic properties, a C3-like protein is expected to have evolved early. However, currently available data suggests that complement-like components are unique to the

deuterostome lineage. This work describes the cloning and characterization of a C3-like gene from *S. exserta*. The deduced polypeptide sequence reveals conservation of multiple, functionally critical, sites while sharing physiochemical and structural properties with the complement components C3/C4/C5.

Antigen processing, via intracellular enzymatic proteasomes, is a major requirement of vertebrate adaptive immunity. These organelles have a catalytic core, through which pass intracellular proteins for degradation into peptides presentable to the immune system. LMP 7 is one component of the paralogous "immuno-proteasome". LMP 7 is a paralog of the ubiquitous LMP X, but is restricted to vertebrates. While LMP 7 is absent in the coral, this work describes a coral LMP X gene. Phylogenetic analyses, along with hydropathy profiling of a critical portion of the invertebrate and vertebrate paralogous genes, suggests that some invertebrates have two diverging LMP X genes. In some cases, one LMP X protein shares characteristics with vertebrate LMP 7. This work presents new evidence for how the LMP X and 7 genes evolved.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter1

Introduction and Background

Introduction

Evolution of immune defense

The hallmark events in the origin of muticellularity were the acquisition of the molecules that establish adhesion and communication with adjacent cells and the ability to protect self from non-self attack. As a first line of defense, metazoans acquired body plans with an outer (sometimes cellular) layer followed by specific mechanisms to selectively deal with microbes that had penetrated this barrier. The most ancient and conserved defense system would appear to be one controlled by cellular responses that include recognition and phagocytosis (Metchnikoff, 1905; Underhill & Ozinsky, 2002) and involve free, soluble molecules that neutralize and/or aid (agglutinins and opsonins) in the elimination of non-self materials. All living organisms display some form of immunocompetence (Bigger, 1984; Bigger, 1988; Burnet, 1970; Janeway & Medzhitov, 2002; Medawar, 1957; Rinkevich, 1996). Studies of immune system components, from a diverse array of extant organisms, reveal an assortment of both unique and shared mechanisms (Buss, 1982; Cooper *et al.*, 2002; Hildemann, 1981; Kasahara, 1998; Litman *et al.*, 1999; Nappi & Ottaviani, 2000; Salzet, 2001; Schluter *et al.*, 1994).

Molecular data on the diversity of immune defense mechanisms has only recently started to become available (Arala-Chaves & Sequeira, 2000; Cooper, 1996; Cooper *et al.*, 2002; Nappi & Ottaviani, 2000). Prior to this time, it has been very tempting for comparative immunologists to make or imply generalizations of vertebrate immune principles to invertebrate defense reactions (some examples include: Beck *et al.*, 1994; Ey & Jenkin, 1990; Ratcliffe, 1985; Rinkevich, 1996). Likewise, the inference of

homology between some associated immune components, based on function without real evidence of homology, has been sharply criticized (e.g., Klein, 1989; Klein, 1997) and has led to an ongoing debate concerning these issues (Beschin *et al.*, 2001; Cooper *et al.*, 1992; Hughes, 1998; Klein, 1989; Klein, 1997; Marchalonis & Schluter, 1990a). Evolution is a central theme in comparative immunology and unfortunately, many authors have erroneously described it as a linear event (e.g., evolving from extant invertebrates *to* vertebrates or from protostome *to* deuterostome).

Homology, convergence, and unique attributes.

Approximately 99% of all extinct and extant animals can be characterized as protostome or deuterostome invertebrates (Adoutte *et al.*, 2000; McMenamin & McMenamin, 1990; Nielson, 2001; Raff, 1996). The protostome and deuterostome lineages diverged about 670 million years ago (mya) (Doolittle *et al.*, 1996), so most modern extant phyla belong to one or the other (Adoutte *et al.*, 2000). Many components of immunity have been characterized in both lineages, and many comparisons to vertebrate immunity have been proposed as if the vertebrates (chordates) were at the pinnacle of evolution, along one major lineage. This ideology has led many authors to suggest that many functionally characterized invertebrate immunological phenomena and/or proteins are homologous to vertebrate forms (Beck *et al.*, 1994; Cooper, 1976) without any significant gene or protein sequence data.

The concept of homology between the two lineages can be valid only if the common ancestor, or extant members of phyla that diverged prior to the split, contain homologs of the genes of interest. Because we do not know what the hypothetical

ancestor's genome looked like, we are limited to extant members of phyla having diverged prior to this period (i.e., Porifera and Cnidaria) or inference of ancestral states. Unfortunately, because of the long divergence times between these phyla, and because many proteins of immunological nature are under varying functional and environmental constraints, significant divergence between homologous genes may still place identity and similarity values into what has been termed the "twilight zone" (Klein, 1997).

Some investigators have argued (Klein, 1989; Klein, 1997) and others have tested hypotheses, with molecular sequence data (Beschin *et al.*, 2001; Hughes, 1998), that many immune system genes from protostome and deuterostome animals are not homologous. These misleading inferences are based on data from a small number of immune response genes, and from a limited number of phyla. Other workers have investigated a separate array of immune system genes (including members of receptor families and signal transduction pathways) and have found significant evidence of homology among both lineages (Armstrong & Quigley, 1999; Bayne & Fryer, 1994; Feizi, 2000; Hoffman *et al.*, 1999; Magor & Vasta, 1998; Medzhitov & Janeway Jr., 2000; Muller, 2001; Muller *et al.*, 1999; Muller *et al.*, 2001; Ottaviani & Franceschi, 1997; Vasta *et al.*, 1996). Lack of homology and evidence of homology suggests that while some genes in both lineages share common ancestry others have been derived to confront similar pathogens and threats (convergence).

The Vendian and Cambrian periods were a relatively short period of evolutionary time (700-500mya), albeit producing some of the most significant events in the history of life (Fortey *et al.*, 1997; McMenamin & McMenamin, 1990; Nielson, 2001; Raff, 1996; Valentine *et al.*, 1991). All major metazoan phyla were produced during this era and

some significant and radical genome-wide events led to major separations in subsequent evolutionary pathways, which included diploblastic to triploblastic, radial to bilateral, acoelomate to coelomate, and protostome-deuterostome divergences (Cameron *et al.*, 1998; Martindale & Henry, 1998). The major lineages were rapidly established and consequently began to diverge as further genetic changes accumulated. Some of these changes were driven by outside genetic invasions (Andersson *et al.*, 2000; Kidwell & Lisch, 2000) which probably became more difficult once the germ-line was sequestered.

Most eukaryotic genes are composed of multiple exons interrupted by non-coding sequences (introns). This organization of the genome into "cassettes" (and/or exons) of coding sequences appears to have facilitated the explosive evolution of metazoans (e.g., see Kidwell & Lisch, 2000; Makalowski, 2000). Exon shuffling, recombination and rearrangement, duplication and divergence, along with conversion events allowed for the production of new proteins to fill a variety of pathways and meet the demands of increasing complexity. Genomic events such as these could provide, in a concerted fashion, new proteins for newly developing pathways in signaling, development, and immunity. Events such as these may also explain how some divergent proteins appear to share domains, motifs or other stretches of amino acids; because at some point in phylogeny some incomplete coding regions received a duplicate exon(s) from a donor gene. Divergent sequences, sharing a particular domain (especially a ligand-binding domain) could subsequently evolve (convergently) to meet similar environmental threats (Marchalonis & Schluter, 1990b; Marchalonis & Schluter, 1994). The great diversity of the fossil record most certainly suggests that these radical genomic changes were very common during the Vendian and Cambrian periods. Subsequently, the assembly of what

is now known as the vertebrate adaptive or combinatorial immune system appears to have involved similar abrupt changes to the genomes of the earliest jawed vertebrates (Agrawal *et al.*, 1998; Kasahara, 1998; Litman *et al.*, 1999; Marchalonis & Schluter, 1998).

Looking beyond vertebrates

Immune systems have been shaped over evolutionary history, greatly influenced by changes in animal genomes (such as genome or gene duplication e.g., Kasahara, 1998). Innate immune mechanisms, including both cellular and humoral elements, consist of a diverse arsenal of toxic intermediates, opsonic and/or agglutinating factors, and complex cellular interactions. Many of these are conserved in protostomes and deuterostomes (Cooper *et al.*, 2002; Nappi & Ottaviani, 2000). Invertebrates, from various phyla, are providing a historical perspective on the evolution of immune-protein families (Cooper *et al.*, 2002). Data of this kind may reveal the nature of the primordial precursors of jawed-vertebrate immunoglobulins, cytokines and recognition receptors in addition to unique invertebrate constituents. Both protostome and deuterostome invertebrates may also enlighten us as to alternative approaches to dealing with similar microbial threats and cellular anomalies (e.g., cancer, Montgomery *et al.*, 1994; Pestarino, 1994).

Misinterpretation of functional similarity

The majority of comparative immunology data, over the last four decades, have consisted of functional studies (e.g., protein characterization and bioassays). Molecular

biological approaches were not practical for many laboratories because DNA data, at the time, was of limited value. While classical protein isolation and characterization has provided significant contributions to the field, mistaken inferences, based on cross-reactive polyclonal antisera made to mammalian immune proteins, have demonstrated that these approaches can be misleading. A major example of this was the independent characterization of invertebrate "cytokines" by multiple laboratories (Beck & Habicht, 1996; Cohen & Haynes, 1990). More recent work, which utilized molecular sequence data, revealed that many of these genes are not homologous to their vertebrate counterparts, instead the proteins share cross-reactive lectin domains (Beschin *et al.*, 2001). In addition, gene sequence studies have now described vertebrate cytokine genes as arranged into exchangeable modules (Kallen *et al.*, 1999). These very significant findings lend support to the claim that independent evolution of genes/proteins to convergently deal with similar environmental and physiological circumstances or threats could have been facilitated by exon shuffling, rearrangement, and motif sharing events in the early establishment of immune response genes.

It has been proposed that various classes of proteins from the innate and vertebrate adaptive immune system may have evolved in this manner (Marchalonis & Schluter, 1990b). Phylogenetic studies of DNA sequences from these shared reactive sites are now starting to emerge (Beschin *et al.*, 2001), and will be essential to our understanding of immune system evolution. Genomic sequencing and mapping studies in humans have revealed chromosomal regions that appear to be hotspots for gene-duplication, recombination, and exon shuffling (Bailey *et al.*, 2002a; Bailey *et al.*, 2002b; Horvath *et al.*, 2001; Samonte & Eichler, 2002; van Geel *et al.*, 2002). This may be true

of other animal genomes as well. Creating or altering genes by these genomic mechanisms can be considered a major source of change upon which selection can act (i.e., adaptive evolution) (Bailey *et al.*, 2001; Eichler, 2001; Ji *et al.*, 2000; Lynch, 2002; Trask *et al.*, 1998).

These findings reinforce the need for collaborative functional studies and phylogenetic analysis of gene-sequence data. The advent of genome sequencing and increasingly accessible molecular biological techniques (that most labs can now utilize with ease) are permitting studies from a diverse array of organisms. Studies from a broad spectrum of invertebrates will afford clues to some of the most functionally important and relevant genes of the immune system, many of which were later co-opted into multiple, divergent, pathways of vertebrates (e.g., Mak & Simard, 1998; Miyazawa *et al.*, 2001). The increasing availability of genomic data is allowing us to evaluate complete gene sequences from multiple phyla simultaneously, where comparisons of introns and exons can be executed with ease.

Unfortunately, though, access to invertebrates has traditionally been limited to a few well-known and established models, which has mostly included insects. Most protocols for establishing invertebrate cell and tissue culture arose through modifications of those established in insects (e.g., Kuroda *et al.*, 1988). Recently, aquaculture interests, which include disease control in mollusks and arthropods, have produced a wealth of data for the field of comparative immunology (Arala-Chaves & Sequeira, 2000; Mothersill & Austin, 2000). Establishing reliable tissue culture and nucleic acid extraction techniques, though, continues to be hindered by the great diversity of issues associated with using marine and land invertebrates as laboratory models (Mothersill & Austin, 2000).

Previous attempts at producing molecular data from corals, for example, were hindered by difficulty in isolating high molecular weight nucleic acids of sufficient purity for downstream applications (Bundschuh, 1992). Based on their phylogenetic position, and what little we know about their alloimmune capabilities, there is a legitimate interest in using Cnidarians as models in immune phylogeny studies.

Corals (Cnidarians) as animal models

Functional analogies and sequence homologies in both the protostome and deuterostome lineages suggest that innate immune mechanisms most resemble the ancestral form of immunity (Janeway & Medzhitov, 2002; Nappi & Ottaviani, 2000; Salzet, 2001). Recent studies even suggest that innate immunity is required for the adaptive immune system to function (reviewed in Janeway, 2002). In efforts to map the origins of immunity, it is becoming increasingly important to compare phyla that diverged prior to the protostome-deuterostome split because one extant phylum, that existed prior to the split and prior to the Cambrian Period, is Cnidaria. This phylum now includes modern forms that probably diverged much later than the Cambrian. Cambrian fossil records suggest, though, that modern corals have retained much of the ancestral body architecture (McMenamin & McMenamin, 1990; Valentine *et al.*, 1991). The typical Cnidarian tissue contains cnidae (for stunning and killing prey), and consists of two tissue layers separated by a gelatinous mesoglea, through which amoeboid cells can travel. It lacks bilateral symmetry and a regular (polarized) embryonic cleavage program characteristic of all higher metazoans (Cameron *et al.*, 1998; Martindale & Henry, 1998). Thus, studies in Cnidarians may render the minimal gene and protein requirements for the

origins of tissue grade complexity, immunity, and other cellular-cooperation systems. It is expected that this work will contribute additional data to what is now being characterized in Porifera (sponges) (Muller, 2001; Muller *et al.*, 1999), a phylum predating true-tissue layered organization.

Sessile marine invertebrates, like corals, will provide intriguing information into the origins of transplantation-type immunity (Bigger, 1988; Buss & Green, 1985; Hildemann *et al.*, 1977; Rinkevich, 1996). Protection from infection is critical to sustain life, but protection from non-self invasion and overgrowth in the competition for space is, *unlike* in vertebrates, not an *artifact* of experimental systems (Buss, 1982; Buss & Shenk, 1990). Alloimmune-type recognition and specificity has been recorded *in situ* and successfully duplicated in laboratory conditions among tunicates (Urochordata) (Rinkevich, 1996), reef building corals (Cnidaria) (Hildemann *et al.*, 1975; Hildemann *et al.*, 1977; Jokiel & Bigger, 1994), soft corals (Cnidaria) (Bigger & Runyan, 1979; Rinkevich, 1996; Salter-Cid & Bigger, 1991; Theodor, 1970; Theodor, 1976; van Alstyne *et al.*, 1992), and sponges (Porifera) (Curtis *et al.*, 1982; Hildemann *et al.*, 1980; Johnston & Hildemann, 1982; Van De Vyver & Barbieux, 1983).

The means by which these animals protect the integrity of their bodies may afford important evidence into the origins and diversification of metazoan defense patterns (Hildemann, 1977; Leddy & Green, 1991). The phenomena of graft rejection (Salter-Cid & Bigger, 1991), along with the associated cellular events are now being described (Olano, 1993; Olano & Bigger, 2000). The genes and related molecular pathways controlling these events are essentially unknown. In addition, key considerations for using the coral, *Swiftia exserta,* as an animal model include: the relative ease involved in

acquiring the animals, the minimal legal requirements to be met, and the simplicity associated with caring for and maintaining the animals (see chapter 1 appendix).

Aim of the current study

The first goal in this work was to establish reliable methods of nucleic acid extraction to make molecular and phylogenetic analysis of genes feasible in this animal (chapters 2 and 3). Once a set of methods became available, they were applied to the endosymbiont-free gorgonian coral, *S. exserta*. In the attempts to characterize fundamental immune-associated genes from this animal, conserved homologs were pursued which appear to be essential to metazoan survival. I describe below, in the context of an introductory background, a component critical to vertebrate innate immunity (opsonic molecules and complement) that was investigated in this coral. This is followed by the description of a family of homologous genes, which was also pursued in this study, whose protein products make up the machinery responsible for protein degradation in cells and antigen processing in vertebrate adaptive immunity.

In addition, because of the long standing interest (e.g., Theodor, 1976) in understanding the mechanisms and genetics of allorecognition in these animals, a genetic fingerprinting approach was sought (Ch. 6) which would help categorize the corals by genetic relatedness. Short tandemly-repeated elements found in the genomes of most eukaryotes, microsatellites, were utilized in the development of a fingerprinting approach which could estimate genetic distance. The applicability of this relatedness-estimation to histoincompatibility studies, remains to be tested.

Opsonic molecules in innate immunity.

The acquisition of the ability to enhance phagocytosis by "tagging" foreign cells/ bodies (opsonization) was a major triumph in the evolution of immune defense mechanisms. Opsonized targets are eliminated much more efficiently by phagocytosis than free microbes, apoptotic bodies, or cancer cells (Dempsey *et al.*, 1996; Fearon & Locksley, 1996; Underhill & Ozinsky, 2002). Throughout phylogeny (in both the protostome and deuterostome lineages), the innate immune system has evolved (or acquired) multiple types of opsonins, which include lectins (and/or agglutinins), complement components, antibodies, and other serum proteins (Baldo *et al.*, 1977; Bayne & Fryer, 1994; Coombe & Parish, 1988; Drickamer & Taylor, 1993; Ey & Jenkin, 1990; Fearon & Locksley, 1996; Feizi, 2000; Levashina *et al.*, 2001; Marchalonis & Schluter, 1989; Nonaka *et al.*, 1999; Vasta *et al.*, 1996).

A major component of vertebrate innate immunity is the complement system, a collection of soluble serum proteins which, by at least three enzyme cascades (classical, alternative, and lectin), leads to the production and/or activation of a terminal product (membrane attack complex, MAC). The MAC can effectively lyse the membrane of most targets while some of the reaction intermediates are also involved in tagging microorganisms or other antigens for phagocytosis (Barrington *et al.*, 2001; Carroll, 1998). All three pathways lead to the activation of the third complement component (C3), which is a thiolester-containing protein (TEP). This is the central, and most critical component of complement, and a chief player in innate immunity (Carroll, 1998; Dempsey *et al.*, 1996; Sahu & Lambris, 2001).

Alpha 2-macroglobulin (A2M) and complement components C3, C4 and C5 are paralogous TEP proteins that are commonly believed to have diverged after the protostome and deuterostome split, so that C3, C4, and C5 are exclusive to the deuterostome lineage (Dodds & Law, 1998; Zarkadis *et al.*, 2001). Orthologous C3 proteins have been characterized from all vertebrate classes and species, along with paralogous C3-like proteins from two deuterostome invertebrate phyla, Echinodermata and Urochordata (Dodds & Law, 1998; Smith *et al.*, 1999). This family of proteins appears to have evolved with the major function of tagging microorganisms or immune complexes (or soluble antigen) for clearance via phagocytosis (Dempsey *et al.*, 1996; Levashina *et al.*, 2001). The origin of TEPs that could act as opsonins (Dodds & Law, 1998; Levashina *et al.*, 2001; Nonaka *et al.*, 1999) was a pivotal innate immune acquisition in early metazoan evolution (Dempsey *et al.*, 1996). Hence, it is hypothesized that a component of this opsonic system of proteins evolved very early on in phylogeny so that a primordial immune system would consist of, or have immediately acquired, an early member of the TEP family (see Chapter 4).

Antigen processing and presentation to the immune system.

Also fundamental in the evolution of adaptive immunocompetence was acquisition of the ability to process and display protein antigens to immunocytes. Antigen processing by the so-called "immunoproteasomes" to display peptides of intracellular origin on class I major histocompatibility molecules (MCH) is characteristic of mammalian adaptive immunity (Kloetzel, 2001; Shastri *et al.*, 2002). This establishes the concept of "self" with the immune system, so that infected cells (intracellular bacteria

or viruses) or cells that are apoptotic or cancerous can be eliminated by cytotoxic immunocytes while autoimmunity is essentially prevented (Shastri *et al.*, 2002). Regardless of the timing at which the machinery of adaptive immunity (composed of rearranging antigen receptors) was acquired by jawed vertebrates, the ability to process antigens, both of self and non-self origin, is essential to the survival of the host (Driscoll & Finley, 1992; Fehling *et al.*, 1994; Monaco & Nandi, 1995; Schoenhals *et al.*, 1999; Shastri *et al.*, 2002).

Proteasomes are organelles partly responsible for the cellular metabolism of proteins (Coux *et al.*, 1996; Kloetzel, 2001; Orlowski, 1990; Voges *et al.*, 1999). As such, they are ubiquitous. Beta subunits have been characterized, for example, in bacteria (Maupin-Furlow & Ferry, 1995; Zwickl *et al.*, 1992), yeast (Friedman *et al.*, 1992), some invertebrates (Pancer *et al.*, 1996; Takezaki *et al.*, 2002) and vertebrates (Clark *et al.*, 2000; Kandil *et al.*, 1996; Martinez & Monaco, 1991; Monaco & Nandi, 1995; Nonaka *et al.*, 2000). Vertebrates have adapted a unique, second "immunoproteasome" responsible for the generation of peptides for presentation to the adaptive immune system (Driscoll *et al.*, 1993; Gaczynska *et al.*, 1993; Shastri *et al.*, 2002). This immunoproteasome is assembled from paralogous copies of subunits belonging to the constitutive (housekeeping) form (Monaco, 1992; Monaco & Nandi, 1995).

The assembled structure appears to be more efficient in the generation of peptides for display on major histocompatibility complex (MHC) molecules (Fehling *et al.*, 1994; Griffin *et al.*, 1998; Khan *et al.*, 2001; Kingsbury *et al.*, 2000). It appears, though, that normal presentation of peptides representing housekeeping "self" proteins does not

require immunoproteasome assembly (Arnold *et al.*, 1992; Momburg *et al.*, 1992). It is only during times of infection that the interferon-γ inducible immunoproteasomes appear to replace the housekeeping proteasomes (Khan *et al.*, 2001; Kloetzel, 2001).

The point in phylogeny at which these paralogous subunits (i.e., LMP X / 7) were established and/or began to diverge has been difficult to determine (Hughes, 1997; Kandil *et al.*, 1996; Takezaki *et al.*, 2002) but LMP7 appears to be unique to the vertebrates. Phylogenetic analysis based on rates of nucleotide substitution, and calibrated with fossil data, has suggested that duplication and divergence occurred at about 600mya (Hughes, 1997). This estimate would imply the existence of two paralogous subunits, LMP X and 7-like, in deuterostome invertebrates and agnathans (jawless fish). No LMP 7-like gene has been characterized in organisms predating jawed-fish divergence. Therefore, there is legitimate interest in understanding which beta subunits exist in various invertebrates, and what subunits existed in organisms that diverged prior to the protostome and deuterostome split.

References:

Adoutte A., Balavoine G., Lartillot N., Lespinet O., Prud'homme B., and de Rosa R. (2000). Special Feature: The new animal phylogeny: Reliability and implications. *PNAS* **97:** 4453-4456.

Agrawal A., Eastman Q. M., and Schatz D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394:** 744-751.

Andersson J. O., Doolittle W. F., and Nesbe C. L. (2000). Are there bugs in our Genome? *Science* **292:** 1848-1850.

Arala-Chaves M., and Sequeira T. (2000). Is there any kind of adaptive immunity in invertebrates? *Aquaculture* **191:** 247-258.

Armstrong P. B., and Quigley J. P. (1999). Alpha2-macroglobulin: an evolutionarily conserved arm of the innate immune system. *Developmental and Comparative Immunology* **23:** 375-390.

Arnold D., Driscoll J., Androlewicz M., Hughes E., Cresswell P., and Spies T. (1992). Proteasome subunits encoded in the MHC are not generally required for the processing of peptides bound by MHC class I molecules. *Nature* **360:** 171-174.

Bailey J. A., Gu z., Clark R. A., Reinert K., Samonte R. V., Schwartz S., Adams M. D., Meyers E. W., Li P. W., and Eichler E. E. (2002a). Recent segmental duplications in the human genome. *Science* **297:** 1003-1007.

Bailey J. A., Yavor A. M., Massa H. F., Trask B. J., and Eichler E. E. (2001). Segmental duplications: organization and impact within the current human genome project assembly. *Genome Research* **11:** 1005-1017.

Bailey J. A., Yavor A. M., Viggiano L., Misceo D., Horvath J. E., Archidiacono N., Schwartz S., Rocchi M., and Eichler E. E. (2002b). Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *American Journal of Human Genetics* **70:** 38-100.

Baldo B. A., Uhlenbruck G., and Steinhausen G. (1977). Anti-Galactan Agglutinins from the Marine Sponge *Axinell polypoides*(Schmidt). *Biologisches Zentralblatt* **96:** 723-733.

Barrington R., Zhang M., Fischer M., and Carroll M. C. (2001). The role of complement in inflammation and adaptive immunity. *Immunological Reviews* **180:** 5-15.

Bayne C. J., and Fryer S. (1994). Phagocytosis and Invertebrate Opsonins in Relation to Parasitism. *In* "Primordial Immunity: Foundations for the Vertebrate Immune System" (G. Beck, G. S. Habicht, E. L. Cooper, and J. J. Marchalonis, Eds.), pp. 162-177, New York Academy of Science, New York.

Beck G., and Habicht G. S. (1996). Cytokines in Invertebrates. *In* "New Directions in Invertebrate Immunology" (K. Soderhall, S. Iwanaga, and G. R. Vasta, Eds.), pp. 131-154, SOS Publications, New Jersey.

Beck G., Habicht G. S., Cooper E. L., and Marchalonis J. J., Eds. (1994). "Primordial Immunity: Foundations for the Vertebrate Immune System," New York Academy of Sciences, New York.

Beschin A., Bilej M., Torreele E., and De Baetselier P. (2001). On the existence of cytokines in invertebrates. *Cellular and Molecular Life Sciences* **58:** 801-814.

Bigger C. H. (1984). Immunorecognition Among Invertebrates. *Developmental and Comparative Immunology* **3:** 29-34.

Bigger C. H. (1988). Historecognition and Immunocompetence in Selected Marine Invertebrates. *In* "Invertebrate Historecognition" (R. Grosberg, D. Hedgecock, and K. Nelson, Eds.), pp. 55-65, Plenum Press, New York.

Bigger C. H., and Runyan R. (1979). An In Situ Demonstration of Self-Recognition in Gorgonians. *Developmental and Comparative Immunology* **3:** 591-597.

Bundschuh J. (1992). Molecular studies in gorgonian alloimmunity: search for gene homologs of the immunoglobulin gene superfamily in *Swiftia exserta.*, pp. 114, Master's Thesis, Florida International University, Miami.

Burnet M. (1970). "Self and Not-Self," Melbourne University Press, Melbourne, Australia.

Buss L., and Green D. R. (1985). Histoincompatibility in vertebrates: the relic hypothesis. *Developmental and Comparative Immunology* **9:** 191-201.

Buss L. W. (1982). Somatic Cell Parasitism and the Evolution of Somatic Tissue Compatibility. *Proceedings from the National Academy of Science, USA* **79:** 5337-5341.

Buss L. W., and Shenk M. A. (1990). Hydroid Allorecognition Regulates Competition at Both the Level of the Colony and at the Level of the Cell Lineage. *In* "Defense Molecules: Proceedings of a UCLA Colloquium, Lake Tahoe, Ca" (J. J. Marchalonis, and C. L. Reinisch, Eds.), pp. 85-105, Whiley-Liss, New York.

Cameron R. A., Peterson K. J., and Davidson E. H. (1998). Developmental gene regulation and the evolution of large animal body plans. *American Zoologist* **38:** 609-320.

Carroll M. C. (1998). The role of complement and complement receptors in induction and regulation of immunity. *Annual Review of Immunology* **16:** 545-568.

Clark M. S., Pontarotti P., Gilles A., Kelly A., and Elgar G. (2000). Identification and characterization of a beta proteasome subunit cluster in the Japanese pufferfish (*Fugu rubripes*). *Journal of Immunology* **165:** 4446-4452.

Cohen N., and Haynes L. (1990). The Phylogenetic Conservation of Cytokines. *In* "Phylogenesis of Immune Function" (G. W. Warr, and N. Cohen, Eds.), pp. 241-268, CRC Press.

Coombe D. R., and Parish C. R. (1988). Sulfated Polysaccharide-Mediated Sponge Cell Aggregation: The Clue to Invertebrate Self/Nonself-Recognition? *In* "Invertebrate Historecognition" (R. Grosberg, D. Hedgecock, and K. Nelson, Eds.), Plenum Press, New York.

Cooper E. L. (1976). "Comparative Immunology," Prentice-Hall, Engelwood, NJ.

Cooper E. L., Ed. (1996). "Invertebrate Immune Responses: Cells and Molecular Products," Springer-Verlag, Berlin.

Cooper E. L., Kauschke E., and Cossarizza A. (2002). Digging for innate immunity since Darwin and Metchnikoff. *BioEssays* **24:** 319-333.

Cooper E. L., Rinkevich B., Uhlenbruck G., and Valembois P. (1992). Invertebrate Immunity: Another Viewpoint. *Scandanavian Journal of Immunology* **35:** 247-266.

Coux O., Tanaka K., and Goldberg A. L. (1996). Structure and functions of the 20S and 26S proteasomes. *Annual Review of Biochemistry* **65:** 801-847.

Curtis A. S. G., Kerr J., and Knowlton N. (1982). Graft rejection in sponges. *Transplantation* **33:** 127-133.

Dempsey P. W., Allison M. E. D., Akkaraju S., Goodnow C. C., and Fearon D. T. (1996). C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science* **271:** 348-350.

Dodds A. W., and Law S. K. A. (1998). The phylogeny and evolution of the thioester bond-containing proteins C3, C4, and alpha2-macroglobulin. *Immunological Reviews* **166:** 15-26.

Doolittle R. F., Feng D.-F., Tsang S., Cho G., and Little E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271:** 470-477.

Drickamer K., and Taylor M. E. (1993). Biology of Animal Lectins. *Annual Review of Cell Biology* **9:** 237-246.

Driscoll J., Brown M. G., Finley D., and Monaco J. J. (1993). MHC-linked LMP gene products specifically alter peptidase activities of the proteasome. *Nature* **365:** 262-264.

Driscoll J., and Finley D. (1992). A controlled breakdown: antigen processing and turnover of viral proteins. *Cell* **68:** 823-825.

Eichler E. E. (2001). Segmental duplications: what's missing, misassigned, and misassembled--and should we care? *Genome Research* **11:** 653-666.

Ey P. L., and Jenkin C. R. (1990). Molecular basis of self/non-self discrimination in the invertebrata. *In* "The Reticuloendothelial System: A comprehensive treatise" (N. Cohen, and M. M. Sigel, Eds.), Plenum Press, New York.

Fearon D. T., and Locksley R. M. (1996). The Instructive Role of Innate Immunity in the Acquired Immune Response. *Science* **272:** 50-54.

Fehling H. J., Swat W., Laplace C., Kuhn R., Rajewsky K., Muller U., and von Boehmer H. (1994). MHC class I expression in mice lacking the proteasome subunit LMP7. *Science* **265:** 1234-1237.

Feizi T. (2000). Carbohydrate-mediated recognition systems in innate immunity. *Immunological Reviews* **173:** 79-88.

Fortey R. A., Briggs D. E., and Wills M. A. (1997). The Cambrian evolutionary "explosion" recalibrated. *BioEssays* **19:** 429-434.

Friedman H., Goebel M., and Snyder M. P. (1992). A homolog of the proteasome-related RING10 gene is essential for cell growth in yeast. *Gene* **122:** 203-206.

Gaczynska M., Rock K. L., and Goldberg A. L. (1993). Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes. *Nature* **365:** 264-267.

Griffin T. A., Nandi D., Cruz M., Fehling H. J., Kaer L. V., Monaco J. J., and Colbert R. A. (1998). Immunoproteasome assembly: Cooperative incorporation of interferon gamma -inducible subunits. *Journal of Experimental Medicine* **187:** 97-104.

Hildemann W. H. (1977). Specific Immunorecognition by Histocompatibility Markers: The Original Polymorphic System of Immunoreactivity Characteristic of All Multicellular Animals. *Immunogenetics* **5:** 193-202.

Hildemann W. H. (1981). Immunophylogeny: From Sponges to Hagfish to Mice. *In* "Frontiers in Immunogenetics" (W. H. Hildemann, Ed.), pp. 3-19, Elsevier North Holland, Inc., Holland.

Hildemann W. H., Bigger C. H., Johnston I. S., and Jokiel P. L. (1980). Characteristics of transplantation immunity in the sponge, *Callyspongia diffusa. Transplantation* **30:** 362-367.

Hildemann W. H., Linthicum D. S., and Vann D. C. (1975). Transplantation and Immunoincompatibility Reactions Among Reef-Building Corals. *Immunogenetics* **2:** 269-284.

Hildemann W. H., Raison R. L., and Hull C. J. (1977). Immunocompetence in Corals: Issues of Specificity, Memory, and Mechanisms. *In* "Developmental Immunobiology" (J. B. Solomon, and J. D. Horton, Eds.), pp. 9-16, Elsevier/North-Holland Biomedical Press, Amsterdam.

Hoffman J. A., Kafatos F. C., Janeway C. A., and Ezekowitz R. A. (1999). Phylogenetic perspectives in innate immunity. *Science* **284:** 1313-1318.

Horvath J. E., Bailey J. A., Locke D. P., and Eichler E. E. (2001). Lessons from the human genome: transitions between euchromatin and heterochromatin. *Human Molecular Genetics* **10:** 2215-2223.

Hughes A. L. (1997). Evolution of the proteasome components. *Immunogenetics* **46:** 82-92.

Hughes A. L. (1998). Protein phylogenies provide evidence of a radical discontinuity between arthropod and vertebrate immune systems. *Immunogenetics* **47:** 283-296.

Janeway C. A., and Medzhitov R. (2002). Innate immune recognition. *Annual Review of Immunology* **20:** 197-216.

Ji Y., Eichler E. E., Schwartz S., and Nichols R. D. (2000). Structure of chromosomal duplicons and their role in mediating human genetic disorders. *Genome Research* **10:** 597-610.

Johnston I. S., and Hildemann W. H. (1982). Cellular Defense Systems of the Porifera. *In* "The Reticuloendothelial System" (N. Cohen, and M. M. Sigel, Eds.), pp. 37-57, Plenum Publishing Corp.

Jokiel P. L., and Bigger C. H. (1994). Aspects of Histocompatibility and Regeneration in the Solitary Reef Coral *Fungia scutaria*. *Biological Bulletin* **186:** 72-80.

Kallen K.-J., Grotzinger J., Leliervres E., Vollmer P., Aasland D., Renne C., Mullberg J., zum Buschenfelde K.-H. M., Gascan H., and Rose-John S. (1999). Receptor recognition sites of cytokines are organized as exchangeable modules. *The Journal of Biological Chemistry* **274:** 11859-11867.

Kandil E., Namikawa C., Nonaka M., Greenberg A. S., Flajnik M. F., Ishibashi T., and Kasahara M. (1996). Isolation of low molecular mass polypeptide cDNA clones: implications for the origin of MHC class I-restricted antigen presentation. *Journal of Immunology* **156:** 4245-4253.

Kasahara M. (1998). What Do the Paralogous Regions in the Genome Tell Us About the Origins of the Adaptive Immune System? *Immunological Reviews* **166:** 159-175.

Khan S., van den Broek M., Schwarz K., de Giuli R., Diener P.-A., and Groettrup M. (2001). Immunoproteasomes largely replace constituative proteasomes during antiviral and antibacterial immune responses in the liver. *Journal of Immunology* **167:** 6859-6868.

Kidwell M. G., and Lisch D. R. (2000). Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution* **55:** 1-24.

Kingsbury D. J., Griffin T. A., and Colbert R. A. (2000). Novel propeptide function in 20S proteasome assembly influences beta subunit composition. *Journal of Biological Chemistry* **275:** 24156-24162.

Klein J. (1989). Are Invertebrates Capable of Anticipatory Immune Responses? *Scandinavian Journal of Immunology* **29:** 499-505.

Klein J. (1997). Homology Between Immune Respones in Vertebrates and Invertebrates: Does it Exist? *Scandanavian Journal of Immunology* **46:** 558-564.

Kloetzel P.-M. (2001). Antigen processing by the proteasome. *Nature Reviews* **2:** 179-187.

Kuroda Y., Kurstak E., and Maramorosch K., Eds. (1988). "Invertebrate and Fish Tissue Culture," Japan Scientific Societies Press & Springer-Verlag, Tokyo & Berlin.

Leddy S. V., and Green D. R. (1991). Historecognition in the Cnidaria. *In* "Phylogenesis of Immune Function" (G. W. Warr, and N. Cohen, Eds.), pp. 103-116, CRC Press, Inc.

Levashina E. A., Moita L. F., Blandin S., Vriend G., Lagueux M., and Kafatos F. C. (2001). Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae. Cell* **104:** 709-718.

Litman G. W., Anderson M. K., and Rast J. P. (1999). Evolution of antigen binding receptors. *Annual Review of Immunology* **17:** 109-147.

Lynch M. (2002). Gene duplication and evolution. *Science* **297:** 945-946.

Magor K. E., and Vasta G. R. (1998). Ancestral Immunity Comes of Age. *Immunology Today* **19:** 54-56.

Mak T. W., and Simard J. J. L. (1998). "Handbook of Immune Response Genes," Plenum Press, New York.

Makalowski W. (2000). Genomic scrap yard: how genomes utilize all that junk. *Gene* **259:** 61-67.

Marchalonis J. J., and Schluter S. F. (1989). Immunoproteins in Evolution. *Developmental and Comparative Immunology* **13:** 285-301.

Marchalonis J. J., and Schluter S. F. (1990a). On the Relevance of Invertebrate Recognition and Defence Mechanisms to the Emergence of the Immune Response of Vertebrates. *Scandanavian Journal of Immunology* **32:** 13-20.

Marchalonis J. J., and Schluter S. F. (1990b). Origins of Immunoglobulins and Immune Recognition Molecules: Some Recognition Systems Emerged Early in Evolution; Others are Restricted to Particular Phyla. *BioScience* **40:** 758-768.

Marchalonis J. J., and Schluter S. F. (1994). Development of an Immune System. *In* "Primordial Immunity: Foundations for the Vertebrate Immune System" (G. Beck, E. L. Cooper, G. S. Habicht, and J. J. Marchalonis, Eds.), pp. 1-12, New York Academy of Science, New York.

Marchalonis J. J., and Schluter S. F. (1998). A Stochastic Model for the Rapid Emergence of Specific Vertebrate Immunity Incorporating Horizontal Transfer of Systems Enabling Duplication and Combinatorial Diversification. *Journal Of Theoretical Biology* **193:** 429-444.

Martindale M. Q., and Henry J. Q. (1998). The development of radial and biradial symmetry: the evolution of bilaterality. *American Zoologist* **38:** 672-684.

Martinez C. K., and Monaco J. J. (1991). Homology of proteasome subunits to a major histocompatibility complex-linked LMP gene. *Nature* **353:** 664-667.

Maupin-Furlow J. A., and Ferry J. G. (1995). A proteasome from the methanogenic archaeon Methanosarcina thermophila. *Journal of Biological Chemistry* **270:** 28617-28622.

McMenamin M. A. S., and McMenamin D. L. S. (1990). "The emergence of animals: the Cambrian breakthrough," Columbia University Press, New York.

Medawar P. (1957). "The Uniqueness of the Individual," Methuen Press, London.

Medzhitov R., and Janeway Jr. C. (2000). Innate immune recognition: mechanisms and pathways. *Immunological Reviews* **173:** 89-97.

Metchnikoff E. (1905). "Immunity in Infectious Diseases," Cambridge University Press, Cambridge.

Miyazawa S., Azumi K., and Nonaka M. (2001). Cloning and characterization of integrin alpha subunits from the solitary ascidian, *Halocynthia roretzi. Journal of Immunology* **166:** 1710-1715.

Momburg F., Ortiz-Navarrete V., Neefjes J., Coulmy E., van de Wal Y., Spits H., Powis S. J., Butcher G. W., Howard J. C., Walden P., and Hammerling G. J. (1992). Proteasome subunits encoded by the major histocompatibility complex are not essential for antigen presentation. *Nature* **360:** 174-177.

Monaco J. J. (1992). A molecular model of MHC class I-restricted antigen processing. *Immunology Today* **13:** 173-178.

Monaco J. J., and Nandi D. (1995). The genetics of proteasomes and antigen processing. *Annual Review of Genetics* **29:** 729-754.

Montgomery D. W., Shen G. K., Ulrich E. D., and Zukoski C. F. (1994). Immunomodulation by Didemnins: Invertebrate Marine Natural Products. *In* "Primordial Immunity: Foundations for the Vertebrate Immune System" (G. Beck, G. S. Habicht, E. L. Cooper, and J. J. Marchalonis, Eds.), pp. 301-314, New York Academy of Sciences, New York.

Mothersill C., and Austin B., Eds. (2000). "Aquatic Invertebrate Cell Culture," Springer-Verlag, Berlin.

Muller W. E. G. (2001). Review: How was the metazoan threshold crossed? The hypothetical Urmetazoa. *Comparative Biochemistry and Physiology* **Part A 129:** 433-460.

Muller W. E. G., Blumbach B., and Muller I. M. (1999). Evolution of the innate and adaptive immune systems. *Transplantation* **68:** 1215-1227.

Muller W. E. G., Schroder H. C., Skorokhod A., Bunz C., Muller I. M., and Grebenjuk V. A. (2001). Contribution of sponge genes to unravel the genome of the hypothetical ancestor of Metazoa (Urmetazoa). *Gene* **276:** 161-173.

Nappi A. J., and Ottaviani E. (2000). Cytotoxicity and cytotoxic molecules in invertebrates. *BioEssays* **22:** 469-480.

Nielson C. (2001). "Animal Evolution: Interrelationships of the living phyla," Oxford University Press, Oxford.

Nonaka M., Azumi K., Ji X., Namikawa-Yamada C., Sasaki M., Saiga H., Dodds A. W., Sekine H., Homma M. K., Matsushita M., Endo Y., and Fujita T. (1999). Opsonic complement component C3 in the solitary ascidian, *Halocynthia roretzi. Journal of Immunology* **162:** 387-391.

Nonaka M., Yamada-Namikawa C., Flajnik M. F., and Du Pasquier L. (2000). Trans-species polymorphism of the major histocompatibility complex-encoded proteasome subunit LMP7 in an amphibian genus, *Xenopus. Immunogenetics* **51:** 186-192.

Olano C. T. (1993). Cellular Aspects of Alloimmunity and Other Responses in the Gorgonian *Swiftia exserta*, Master's Thesis, Florida International University, Miami.

Olano C. T., and Bigger C. H. (2000). Phagocytic activities of the gorgonian coral, Swiftia exserta. *Journal of Invertebrate Pathology* **76:** 176-184.

Orlowski M. (1990). The multicatalytic proteinase complex, a major extralysosomal proteolytic system. *Biochemistry* **29:** 1028910297.

Ottaviani E., and Franceschi C. (1997). The invertebrate phagocytic immunocyte: clues to a common evolution of immune and neuroendocrine systems. *Immunology Today* **18:** 169-174.

Pancer Z., Scheffer U., Muller I., and Muller W. E. G. (1996). Cloning of sponge (*Geodia cydonium*) and tunicate (*Botryllus schlosseri*) proteasome subunit epsilon (PRCE): Implications about the vertebrate MHC-encoded homologue LMP7 (PRCC). *Biochemical and Biophysical Research Communications* **228:** 406-410.

Pestarino M. (1994). A Possible Immunomodulatory Role of Endozepine-like Peptides in a Tunicate. *In* "Primordial Immunity: Foundations for the Vertebrate Immune System" (G. Beck, G. S. Habicht, E. L. Cooper, and J. J. Marchalonis, Eds.), pp. 365-367, New York Academy of Science, New York.

Raff R. A. (1996). "The Shape of Life: Genes, Development, and the Evolution of Animal Form," The University of Chicago Press, Chicago.

Ratcliffe N. A. (1985). Invertebrate Immunity- A Primer for the Non-Specialist. *Immunology Letters* **10:** 253-270.

Rinkevich B. (1996). Immune responsiveness in marine invertebrates revisited: the concourse of puzzles. *In* "New Directions in Invertebrate Immunology" (K. Soderhall, S. Iwanaga, and G. R. Vasta, Eds.), pp. 55-90, SOS Publications, Fair Haven, NJ.

Sahu A., and Lambris J. D. (2001). Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunological Reviews* **180:** 35-48.

Salter-Cid L., and Bigger C. H. (1991). Alloimmunity in the Gorgonian Coral *Swiftia exserta*. *Biological Bulletin* **181:** 127-134.

Salzet M. (2001). Vertebrate innate immunity resembles a mosaic of invertebrate immune responses. *Trends in Immunology* **22:** 285-288.

Samonte R. V., and Eichler E. E. (2002). Segmental duplications and the evolution of the primate genome. *Nature Reviews in Genetics* **3:** 65-72.

Schluter S. F., Schroeder J., Wang E., and Marchalonis J. J. (1994). Recognition Molecules and Immunoglobulin Domains in Invertebrates. *In* "Primordial Immunity: Foundations for the Vertebrate Immune System" (G. Beck, G. S. Habicht, E. L. Cooper, and J. J. Marchalonis, Eds.), pp. 74-81, Annals of the New York Academy of Science, New York.

Schoenhals G. J., Krishna R. M., Grandea III A. G., Spies T., Peterson P. A., Yang Y., and Fruh K. (1999). Retention of empty MHC class I molecules by tapasin is essential to reconstitute antigen presentation in invertebrate cells. *The EMBO Journal* **18:** 743-753.

Shastri N., Schwab S., and Serwold T. (2002). Producing Nature's Gene-Chips: The Generation of Peptides for Display by MHC Class I Molecules. *Annual Review of Immunology* **20:** 463-493.

Smith L. C., Azumi K., and Nonaka M. (1999). Complement systems in invertebrates: The ancient alternative and lectin pathways. *Immunopharmacology* **42:** 107-120.

Takezaki N., Zaleska-Rutczynska Z., and Figueroa F. (2002). Sequencing of amphioxus *PSMB5/8* gene and phylogenetic position of agnathan sequences. *Gene* **282:** 179-187.

Theodor J. L. (1970). Distinction between "Self" and "Not-Self" in Lower Invertebrates. *Nature* **227:** 690-692.

Theodor J. L. (1976). Histo-incompatibility in a natural population of gorgonians. *Zoological Journal of the Linnean Society* **58:** 173-176.

Trask B. J., Massa H., Brand-Arpon V., Chan K., Friedman C., Nguyen O. T., Eichler E. E., van den Engh G., Rouquier S., Shizuya H., and Giorgi D. (1998). Large multi-chromosomal duplications encompass many members of the olfactory receptor gene family in the human genome. *Human Molecular Genetics* **7:** 2007-2020.

Underhill D. M., and Ozinsky A. (2002). Phagocytosis of Microbes: Complexity in Action. *Annual Review of Immunology* **20:** 825-852.

Valentine J. W., Awramik S. M., Signor P. W., and Sadler P. M. (1991). The biological explosion at the Precambrian-Cambrian boundary. *Evolutionary Biology* **25:** 279-356.

van Alstyne K. L., Wylie C. R., Paul V. J., and Meyer K. (1992). Antipredator Defenses in Tropical Pacific Soft Corals (Coelenterata: Alcyonacea). I. Sclerites as Defenses Against Generalist Carnivorous Fishes. *Biological Bulletin* **182:** 231-240.

Van De Vyver G., and Barbieux B. (1983). Cellular Aspects of Allograft Rejection in Marine Sponges of the Genus *Polymastia*. *The Journal of Experimental Zoology* **227:** 1-7.

van Geel M., Eichler E. E., Bech A. F., Shan Z., Haaf T., van der Maarel S. M., Frants R. R., and de Jong P. J. (2002). A cascade of complex subtelomeric duplications during the evolution of the homonoid and Old World monkey genomes. *American Journal of Human Genetics* **70:** 269-278.

Vasta G. R., Ahmed H., and Quesenberry M. S. (1996). Invertebrate C-Type Lectins and Pentraxins as Non-Self Recognition Molecules. *In* "New Directions in Invertebrate Immunology" (K. Soderhall, S. Iwanaga, and G. R. Vasta, Eds.), pp. 189-227, SOS Publications, New Jersey.

Voges D., Zwickl P., and Baumeister W. (1999). The 26S prosteasome: a molecular machine designed for controlled proteolysis. *Annual Review of Biochestry* **68:** 1015-1068.

Zarkadis I. K., Mastellos D., and Lambris J. D. (2001). Phylogenetic aspects of the complement system. *Developmental and Comparative Immunology* **25:** 745-762.

Zwickl P., Grziwa A., Puhler G., Dahlmann B., Lottspeich F., and Barmeister W. (1992). Primary structure of the Thermoplasma proteasome and its implications for the structure, function, and evolution of the multicatalytic proteinase. *Biochemistry* **31:** 964-972.

Chapter 2

Rapid and reliable coral DNA and RNA extraction procedures

Abstract

Isolation of intact and pure nucleic acids from invertebrates has been a major challenge for investigators. Equal to the diversity of invertebrates are the potential problems encountered by the profound variety in tissue types. Although numerous nucleic acid extraction procedures have been described over the years, most have been geared toward a tissue-specific concern, i.e. polysaccharide-rich or nuclease-rich. In our comparative immunology studies, we use a gorgonian coral whose tissue is mucus- and nuclease-rich. Conventional isolation procedures have proved inconsistent in providing pure and/or intact DNA and RNA primarily because conventional methods of isolation fail to protect the nucleic acids from the excessive amounts of nucleases. In this chapter, DNA and RNA isolation procedures based on commercial guanidine-salt-based solutions, are described, which quickly inactivate nucleases and destroy proteins for the effective isolation of intact and clean nucleic acids. The standard procedures have been modified considerably to deal with problematic tissues types. The methods described yield DNA and RNA of sufficient purity for most routine molecular biology applications, particularly for the study of gene expression and/or genetic fingerprinting.

Introduction

Recent topics and interests in comparative immunology (e.g., issues concerning homology) suggest that comprehension of the evolution of immunity will be greatly assisted by studying animal phyla that predate the protostome and deuterostome divergence in phylogeny (Cooper *et al.*, 2002; Hughes, 1998; Klein, 1989; Klein, 1995; Klein, 1997; Salzet, 2001). Highly specific and non-specific immunocompetence has been demonstrated in many invertebrate classes (Arala-Chaves & Sequeira, 2000; Cooper *et al.*, 2002; Cooper *et al.*, 1992; Hildemann, 1981) but the issues concerning the functional mechanisms and genetic similarities associated with these phenomena have sparked several enthusiastic debates (Arala-Chaves & Sequeira, 2000; Cooper *et al.*, 1992; Hughes, 1998; Klein, 1989; Marchalonis & Schluter, 1990). Although many immune system components from the divergent lineages appear to have evolved independently while converging on function (Beschin *et al.*, 2001; Hughes, 1998), some issues of true gene and exon homology can be addressed by studying invertebrates from various phyla. Cnidarians predate the protostome-deuterostome split, and because some immune defense reactions can be stimulated under laboratory conditions (Olano & Bigger, 2000; Salter-Cid & Bigger, 1991) corals are an appropriate animal model to address issues concerning gene homology in protostomes and deuterostomes.

Gorgonian corals are important inhabitants of tropical reefs and near shore environments. Population studies to determine genetic relatedness, phylogeny assessment, or genetic analysis of genes via genome studies require methods of isolating intact, high molecular weight DNA of high purity. Equally pure and intact RNA is a requirement for

gene expression studies or the cloning of gene family members (i.e., RT-PCR). Many shallow-water corals have tissues populated by endosymbiotic algae which can contaminate samples for molecular analysis.

Contaminating endosymbionts and their location within the tissues varies with species, and several methods have been developed by investigators to avoid them in nucleic acid preparation (such as the use of tissue immediately surrounding the central axis, extraction and isolation of nucleic acids from gametes or directly from swimming larvae, or the removal of the endosymbionts from lysed adult tissues (Lohuis *et al.*, 1990; Tom *et al.*, 1999). For most of our work, and for the purpose of this study, an endosymbiont-free deep-water gorgonian coral (*Swiftia exserta*) was used and, therefore, avoided the risk of foreign nucleic acid contamination from endosymbionts.

Molecular studies of corals are often hindered by difficulties in attaining "clean" nucleic acids that have not been degraded. This is because most corals contain high levels of nucleases in their tissues, are surrounded by polysaccharide-rich surface mucus, and by excessive tissue pigmentation. These factors interfere with extraction buffers which cannot immediately inactivate or destroy nucleases, resulting in highly degraded nucleic acids. Proteins, pigments, polysaccharides and other secondary metabolites often co-purify with the nucleic acids and thus interfere with subsequent enzymatic manipulations. Several methods (Ausubel *et al.*, 1997; Jones, 1953; Katterman & Schattuck, 1983; Kumar *et al.*, 1988; Lohuis *et al.*, 1990; Sambrook *et al.*, 1989) have been described that help circumvent many of these problems but involve many time consuming and tedious steps that may take several days before molecular analysis is possible (i.e., dialysis against Tris-EDTA buffers, or CsCl density gradient

centrifugation). Previously, our lab has had inconsistent results (unpublished data) with nucleic acids harvested via conventional methods for plant DNA extraction which utilize CTAB (hexadecyltrimethylammonium bromide, Fisher Scientific, Pittsburg, PA, USA) to prevent co-purification of polysaccharides with DNA (Doyle & Doyle, 1987; Katterman & Schattuck, 1983). These procedures, though effective at removing many polysaccharide (and other) contaminants (Stewart Jr. & Via, 1993), suffer from the inability to consistently protect the DNA during initial homogenization in tissues rich in nucleases. This work describes the rapid isolation of "pure" and intact high molecular weight genomic DNA and RNA from tissues of a soft coral utilizing one-step (guanidine-salt-based) extraction procedures (Chomczynski & Sacchi, 1987). The extraction protocol has been modified such that DNA and RNA is extracted within a few hours and is of acceptable quality for PCR, cloning, sequencing, and enzymatic manipulation.

Materials and Methods

DNA extraction using DNAzol™

Standard procedure with some modifications

Extraction of genomic DNA is performed using a commercially available guanidine-detergent, DNAzol (Molecular Research Center (Manufacturer's protocol, Cincinnati, Ohio, USA). This reagent is based on the one-step method that lyses tissues, rapidly inactivates nucleases, hydrolyzes RNA, and allows for the selective precipitation of DNA with ethanol (Chomczynski *et al.*, 1997). Several important modifications to the

manufacture's protocols have been made for the coral tissue. Following is a description of the standard protocol with some modifications.

Typically, 25-50 mg of tissue (or a small, 3-5 mm maximum, branch piece) is cut from the gorgonian coral colony, *Swiftia exserta*, and quickly rinsed in ice-cold (2-4°C) filtered sea water (which helps remove surface contamination and mucus). The tissue is then homogenized at room temperature in 1 ml of DNAzol in a 1.5 ml microfuge tube with a Kontes plastic disposable pestle (Fisher Scientific, Pittsburg, PA, USA). Homogenization is done with the least number of strokes to minimize mechnical shearing of the DNA. The homogenate is left standing at room temperature for 15 min and then spun at at16000 xg in a microcentrifuge at 4°C for 10 min. Centrifugation separates the insoluble cell/tissue debris and most proteoglycans/polysaccharides and RNA from the solublized lysate containing genomic DNA. The lysate is transferred to a new tube. Absolute ethanol (EtOH) at room temperature (RT) (500 µl) is then added to the lysate to precipitate the DNA. The solution is mixed by inversion, allowed to stand for 10 min (RT), and the DNA pelleted at 5000 xg (4°C) for 5 min. Centrifugation at high speed or for long periods of time should be avoided since it often results in the co-purification of contaminants while compacting the DNA pellet, making its resuspension more difficult. The DNA at this point may not be visible because it may adhere to the tube wall and not collect at the bottom. Additional or higher speed spins are not recommended to form a pellet. The DNA is first washed with a solution of 70% DNAzol:30% EtOH followed by a second wash with 70% EtOH (in water). In either step, a 1-2 min spin might be necessary if pellet comes loose. After complete removal of EtOH with a pipette tip, the

DNA is allowed to dry for about 10 min (by leaving tube caps open on the bench top, preferably in a clean-air hood).

Additional cleaning of the DNA can be performed by first resuspending in 100 µl of water (preheated to 65°C) and adding 5 µl of an RNase A solution (10 mg/ml; Amresco; Solon, Ohio, USA). The mixture is incubated for 20 min at 65°C.. The salt concentration is adjusted to 0.7M with 5M NaCl, followed by the addition of 65°C preheated 10% CTAB (10%CTAB:0.7M NaCl) to a final concentration of 2% (Ausubel *et al.*, 1997; Murray & Thompson, 1980). After an additional 10 min incubation at 65°C, the reaction is cleaned-up by extracting with an equal volume of tris-buffered (pH 8) phenol:chloroform (1:1) and then chloroform, each time mixing and incubating at RT for 2-5 min and spinning at full speed (4°C) for 2 min. The chloroform extraction should be repeated if the final interface is not clear. This CTAB step is added to help remove co-purified polysaccharides and proteoglycans from the DNA solution.

The DNA is then precipitated with 3 volumes of EtOH. If DNA does not become visible immediately upon mixing, it is stored at -80°C for 15 min. The DNA is pelleted by centrifugation for 5 min at 5,000-8,000 xg (4°C) and washed with 70% EtOH (to remove the salts). As described before, the tubes are briefly spun and slightly dried before the DNA is resuspended in 100 µl of nuclease-free sterile water. It is imperative that the DNA is allowed to dissolve completely, which may take a 15 min incubation at 65°C since genomic DNA is difficult to resuspend. An aliquot of the genomic DNA can be electrophoresed on a 0.8% TAE agarose gel to determine quantity, integrity and verify

absence of RNA. The pure (A260:280 >1.8) and high molecular weight DNA is ready for molecular applications (see figure 1a).

Extensive modification of standard procedure

In the following description, the standard procedures for using the DNAzol has been extensively modified, and essentially, the DNAzol becomes a grinding/ extraction buffer that is further purified with standard organic extractions prior to precipitation. The results throughout this dissertation are based largely on DNA extracted via this modified procedure.

DNAzol is prepared by adding polyvinylpyrrolindone (PVP) (2%, w/v) and 2-mercaptoethanol (2%, v/v) and heating for 10 min at 65°C to bring into solution. After solubilization of the PVP, Proteinase K is added (at 15 µl per ml, from 20 mg/ml stock) to the DNAzol extraction buffer. This buffer is ready for immediate use or can be kept for a few days at 4°C. As described above, the tissue is homogenized in 1ml of DNAzol (the modifications described here allow for extraction from tissue sizes 2-3x larger without effecting quality). The extraction is routinely scaled up by grinding the tissue in liquid nitrogen in a ceramic mortar and pestle and the ground tissue placed into a tube with DNAzol (up to 1ml of packed powdered tissue for each 10 ml of the extraction buffer). The homogenate is then rocked continuously (using a nutator) for 20 –30 min at RT. The homogenate is then transferred in 1 ml portions into 1.5 ml eppendorf-style tubes. Each 1 ml portion is extracted with 500 µl of phenol:chloroform:isoamyl (25:24:1) by mixing vigorously (by hand, do not vortex). This mixture is allowed to sit (with frequent mixing) for 10min at RT and then the phases are separated by centrifugation at 16000 xg. The upper aqueous phase is transferred and re-extracted as

before. One to two extractions of chloroform are then performed until the interface is clean. To the aqueous phase, an equal volume of RT absolute ethanol is added and the tube is inverted multiple times to precipitate the DNA (which should become visible). The tube is allowed to stand 5 min and spun as described above for 5 min at 5000 xg. The pellet is washed in 70% ethanol several times and allowed to dry.

The genomic DNA pellet is resuspended in 50-100 µl of nuclease-free water as described above. The DNA is treated with RNase A and cleaned with phenol:chloroform as described before. The resulting aqueous phase is extracted with chloroform. The aqueous phase is then separated into a new tube and one-half the volume of 7.5M ammonium acetate is added, followed by three volumes of ethanol to precipitate the DNA. Precipitation is performed at room temperature for about 10min. If DNA does not become visible, allow to precipitate at –80 for 20 min. Spin, wash, dry pellet and resuspend the DNA in 50-100 µl of nuclease-free water. Determine integrity and purity, and determine concentration as described above.

RNA extraction using TriReagent™

When extracting total RNA, we have found that the one-step extraction reagent based on acid phenol and guanidine thiocynate (available commercially as TriReagent [MRC, Cincinnati, Ohio, USA]) consistently provides RNA of exceptional purity and integrity from our corals. A small piece of tissue is homogenized, as described for DNA extraction, in 1 ml of TriReagent. Even though the TriReagent penetrates tissues almost immediately to inactivate nucleases, the tissues are homogenized as quickly as possible or the tubes kept on ice. After homogenization, the tubes are allowed to sit at room

temperature for 10 min to dissociate nucleoproteins. The tubes are then centrifuged at 16000 xg for 10 min to pellet cell debris and other insoluble components.

To the lysate, 100 µl of BCP (bromochloropropane; Sigma, St. Louis, MO, USA) is added and the tubes are vigorously mixed (by hand) and incubated at room temperature for 15 min. The tubes are then spun at full speed for 10 min and the aqueous phase recovered. A second organic extraction is performed to help remove polysaccharides and pigments by added 200 µl of chloroform and mixing. After a 5 min incubation, the tubes are spun for 5 min at full speed.

To isolate clean RNA from the aqueous phase, 250 µl of isopropanol and 250 µl of a high salt buffer (1.2M NaCl, 0.8M NaCitrate) are added and mixed. The high salt conditions excludes most polysaccharides from co-purifying with the RNA. The RNA is spun for 10 min at 12000 xg (note that the work area, equipment, buffers and reagents must all be RNase free, especially beyond this point in the procedure). The RNA pellet is then washed once or twice with 70% EtOH (by vortexing for several seconds). The pellet is collected by spinning at 16000 xg, dried for 2-5 min (RT) and then resuspended in 20-30 µl of 65°C preheated RNase-free-DEPC-treated water. The RNA is maintained at 65°C for 10 min to ensure full solubilization of the pellet before proceeding with other subsequent applications. Even though DNA contamination is almost non-existant, a Dnase extraction should be performed with 10 U of RNase-free DNase (Promega, Madison, WI, USA) for 20 min at 37°C. The reaction is extracted once with phenol:chloroform (3:1) (note: acid phenol works best) and the RNA is precipitated with 1/10 volume of 3M sodium acetate (pH 5) and 3 volumes of ethanol (-20° 2hrs, -80°

30min or dry ice for 10min). After collecting the RNA pellet (15 min full-speed spin), it is resuspended in 20 µl of RNase-free water and is ready for quantification and use.

Testing Purity of DNA

One microgram of high molecular weight genomic DNA from the coral was digested with the following restriction enzymes: Hind III, EcoRI, and Sau 3AI (Promega, Madison, WI). After overnight digestion at 37°C, the DNA was electrophoresed through a 1% TAE agarose gel and stained with ethidium bromide.

Genomic DNA was PCR-amplified with primers to a known region of a coral gene (data not shown) encoding a thiolester-containing protein we are currently studying (Dishaw *et al.*, 2000). In RT-PCR, these primers produce a 145 base pair (bp) product when amplified for 30 cycles with an annealing temperature of 55-60°C. At the genomic DNA level, an intron(s) separates the primers to produce a 1500 bp fragment.

Genomic DNA was also amplied with primers to 18s ribosomal DNA (using 18s rRNA primers; Ambion, Austin, Tx, USA), using 20 ng of template and 30 cycles of PCR (similar conditions used for RNA quantification studies).

DNA purity and integrity was also tested utilizing RAPD-PCR (Williams *et al.*, 1990). Briefly: 10 ng of genomic DNA was amplified in a 50 µl reaction volume with 20 pmol of primer (5'-CGGTCACTGT or 5'-CGGCCCCTGT). PCR conditions: 95°C for 5 min and 45 cycles of 94°C for 1 min, 38°C for 1 min, and 72 °C for 2min, followed by a 10 min extension at 72°C. After amplification, the reaction products were analyzed on a 1.5% TAE agarose gel and stained with ethidium bromide.

Testing Purity of RNA

Using total RNA and a standard RT-PCR protocol, the above mentioned coral gene-specific 145 bp PCR product was amplified. Briefly: 5 μg of total RNA was denatured at 80°C for 5 min and used as a template for cDNA synthesis using RNase H minus MMLV (Promega, Madison, WI, USA). First strand synthesis was primed with either oligo-dT(17) or the antisense gene-specific primer and incubated for one hour at 42°C as recommended by the manufacturer. After a 20 min incubation at 37°C with RNase H, 5 μl of first strand cDNA was used as the template for PCR and amplified with both sense and antisense primers for 30 cycles.

Using 1 μg of total RNA, first strand synthesis of cDNA was performed using the Smart PCR cDNA Synthesis Kit according to manufacturer's protocols (Clonetech; Palo Alto, California, USA). This kit takes small amounts of RNA and amplifies them using a oligo-GGG (Smart Oligo) anchored to the 3'-CCC ends of cDNA synthesized using RNase H minus MMLV. By having a place to anchor at both the 3' and 5' ends of double stranded cDNA, one can ideally amplify (using PCR) a library of mostly full length cDNAs. For an unknown animal model, this not only gives you the relative size range of cDNAs (because a smear is produced in an agarose gel) but can act as an indicator of the starting RNA purity.

Using an aliquot of the Smart-amplified uncloned cDNA, we used the same above-mentioned coral gene specific internal primers (2.9 kb from the 3' end of the mRNA) to amplify a 145 bp PCR product. Our sense and antisense gene specific primers (10 pmol each) were used in 30 cycles of PCR with an annealing temperature of 60°C.

After amplification, 15 µl was analyzed on a 2% TAE agarose gel and stained with ethidium bromide.

Results

DNA extraction and analysis

We have used DNAzol-extracted DNA with reproducible results for restriction analysis, construction of a microsatellite library, fluorescent-automated PCR fingerprinting, RAPD-based fingerprinting and automated sequencing (data not shown). Because of the guanidine-salt-based nature of DNAzol, genomic DNA from our nuclease-rich coral is consistently isolated intact and with a high-molecular weight (see figure 1a) without the added inconvenience of having to use liquid nitrogen. By controlling tissue size and speed and length of centrifugation (see materials and methods), we can eliminate most carried-over polysaccaride-like contaminants. The average yield of DNA from a small 25-50 mg piece of tissue is approximately 5-10 µg, which provides sufficient DNA for restriction anaylsis and numerous PCR reactions.

Restriction analysis of the DNA (see figure 1b) consistently results in fully digested genomic DNA in typical over-night digestions. Gene-specific amplification resulted in the expected 1.5 kb band using standard PCR conditions and standard Taq polymerase (Qiagen, Valencia, Ca, USA) with no optimization of conditions from RT-PCR (figure 1c). Amplification of 18s ribosomal DNA produced the expected 488 bp band with no optimization from conditions used to amplify the equivalent vertebrate gene (figure 1c). RAPD PCR-based fingerprinting is very sensitive to the quality of the

starting template (Williams *et al.*, 1990) and hence provides a useful method of determining DNA purity. Based on the reproducibility (data not shown) of our RAPD-PCR patterns (i.e., figure 1d) we conclude that consistently clean DNA from this coral can be attained with the DNAzol reagent, using the described protocols.

RNA extraction and analysis

Using total RNA (figure 2a) and standard RT-PCR conditions, we routinely produce our 145 bp coral gene-specific PCR product (figure 2b) whether first strand synthesis was primed by Oligo-dT or the antisense primer. To generate this PCR fragment with Oligo-dT primed cDNAs, at least a 3kb product must be generated by the RT reaction. We have been successful at using RACE (Rapid Amplification cDNA Ends) to amplify multiple portions of this gene and others (data not shown).

Using the Smart system (Clonetech), several PCR-amplified uncloned cDNA libraries have been produced from both total and messenger RNA. The libraries consistenly yield PCR fragments between 0.2-5kb (see figure 2c). This procedure requires the RNA to be of optimum integrity and purity for reverse transcriptase to efficiently reach the end of the mRNA (assuming minimal secondary structure) and for PCR to amplify both long and short cDNAs. From these libraries we were able to reproducibly generate our gene-specific 145 bp band (not shown, same as figure 2b).

Discussion

The methods above describe isolation of DNA and RNA from a gorgonian coral. Extraction reagents, based on guanidine salts, immediately and consistently inactivate nucleases to produce consistent yields of undegraded nucleic acids ready for routine molecular biology applications. Other methods for the isolation of genomic DNA from corals and other nuclease-rich sources have been described, however, in our hands they fail to provide consistently undegraded DNA because a strong nuclease-inhibiting reagent is not employed in the initial homogenization that works as fast or as effectively as guanidine salts. Some investigators (Lohuis *et al.*, 1990) have suggested that guanidine-based methods are unreliable for corals because too many contaminants are co-purified with the DNA. We have not found this to be a problem when employing the *extensive modification of standard procedure* for DNAzol. Even when using the *standard procedure with some modifications*, contaminant co-purification can be minimized by keeping the tissue:reagent ratio small (1:20). The speed and duration of centrifugation should also be closely controlled.

While many protocols have been published that claim to be DNA isolation methods or reagents for plants "for plants" or "for invertebrates," it would be incorrect to claim that one method is applicable to "all" organisms of a particular type (if only a few have been tested) without further protocol modification for optimization. The degree of difficulty is tissue- and species-specific whether for plants or mucus-rich invertebrates. For instance, this work describes guanidine-based methods that work on our gorgonian coral. Other species of gorgonians, for example, can be slightly more "stubborn," with denser tissues (requiring liquid nitrogen to pulverize) or tissue richer in mucus (and other

complex sugars) or polyphenolic compounds. Variation is also not uncommon within a species, since secondary metabolite production varies with stress, for example.

Several modifications to the above described protocol can be used when tissues become increasingly difficult to deal with. Extra co-purified polysaccharides that are carried over into the DNA precipitation can be dealt with using CTAB and chloroform extraction of the solubilized DNA solution (Ausubel *et al.*, 1997; Murray & Thompson, 1980) . As described above, this is done by increasing the salt of the solubilized DNA to 0.7M NaCl and adding 65°C preheated CTAB (10% CTAB/0.7M NaCl stock) to a final concentration of 2%. The DNA-CTAB mixture is chloroform extracted (equal volume) until no interface is noticeable. The DNA can then be precipitated and freshly resuspended. Because CTAB complexes to polysaccharides at high salt concentrations, it can be chloroform extracted and separated from the DNA. This CTAB step is included in the above described technique (see methods) because it is routinely used in our lab. But depending on the apparent purity of the DNA pellet during resuspension in water (contaminating polysaccharides do not go into solution well), the CTAB step can be skipped and the solubilized DNA cleaned with phenol:chloroform to remove the RNase. With the extensive modification procedure, many of these issues do not arise and hence it has become the preferred method for these studies.

RNA isolation rarely requires any deviation or optimization from the above described protocol. This is because the reagents and conditions which make for pure RNA isolation are less attractive to the co-purifying contaminants that are common in genomic DNA isolation. Additionally, the author has found that mRNA isolation (from total RNA) cleans even the most difficult total RNA preparations.

The above described guanidine-based methods are relatively easy, fast, and reliable and can be applied to a variety of nuclease-, pigment-, and polysaccharide-rich invertebrates with equal success. Nucleic acids are extracted with purity sufficient for almost all routine molecular applications. Reliable yet simplistic methods of isolating nucleic acids are invaluable in the study of invertebrates, which are animals commonly studied to pursue comparative immunology, phylogenetic, population biology, and developmental biology questions.

References

Arala-Chaves M., and Sequeira T. (2000). Is there any kind of adaptive immunity in invertebrates? *Aquaculture* **191:** 247-258.

Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Seidman J. G., Smith J. A., and Struhl K., Eds. (1997). "Short Protocols in Molecular Biology," Whiley, New York.

Beschin A., Bilej M., Torreele E., and De Baetselier P. (2001). On the existence of cytokines in invertebrates. *Cellular and Molecular Life Sciences* **58:** 801-814.

Chomczynski P., Mackey K., Drew R., and Wilfinger W. (1997). DNAzol: A reagent for the rapid isolation of genomic DNA. *BioTechniques* **22:** 550-553.

Chomczynski P., and Sacchi N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162:** 156-159.

Cooper E. L., Kauschke E., and Cossarizza A. (2002). Digging for innate immunity since Darwin and Metchnikoff. *BioEssays* **24:** 319-333.

Cooper E. L., Rinkevich B., Uhlenbruck G., and Valembois P. (1992). Invertebrate Immunity: Another Viewpoint. *Scandanavian Journal of Immunology* **35:** 247-266.

Dishaw L. J., Smith S. L., and Bigger C. H. (2000). Sequence analysis of partial cDNA clones from a primitive coral, encoding a thiolester-containing protein. *Developmental and Comparative Immunology* **24:** S23.

Doyle J. J., and Doyle J. L. (1987). A Rapid DNA Isolation Proceedure for Small Quantities of Fresh Leaf Tissue. *Phytochemical Bulletin* **19:** 11-15.

Hildemann W. H. (1981). Immunophylogeny: From Sponges to Hagfish to Mice. *In* "Frontiers in Immunogenetics" (W. H. Hildemann, Ed.), pp. 3-19, Elsevier North Holland, Inc., Holland.

Hughes A. L. (1998). Protein phylogenies provide evidence of a radical discontinuity between arthropod and vertebrate immune systems. *Immunogenetics* **47:** 283-296.

Jones A. S. (1953). The Isolation of Bacterial Nucleic Acids Using Cetyltrimethylammonium Bromide (CETAVLON). *Bichemica et Biophysica Acta* **10:** 607-612.

Katterman F. R. H., and Schattuck V. I. (1983). An Effective Method of DNA Isolation from the Mature Leaves of *Gossypium* Species That Contain Large Amounts of Phenolic Terpenoids and Tannins. *Perspective Biochemistry* **13:** 347-359.

Klein J. (1989). Are Invertebrates Capable of Anticipatory Immune Responses? *Scandinavian Journal of Immunology* **29:** 499-505.

Klein J. (1995). Ehrlich and Darwin: Homochauvinism in immunology. *Immunology and Cell Biology* **73:** 103-108.

Klein J. (1997). Homology Between Immune Respones in Vertebrates and Invertebrates: Does it Exist? *Scandanavian Journal of Immunology* **46:** 558-564.

Kumar S., Degnan B. M., Ross I. L., Hawkins C. J., and Lavin M. F. (1988). Isolation of DNA and RNA from Ascidians. *Marine Biology* **98:** 95-100.

Lohuis M. t., Alderslade P., and Miller D. J. (1990). Isolation and Cloning of DNA from Somatic Tissue of Soft Corals (Cnidaria: Octocorallia). *Marine Biology* **104:** 489-492.

Manufacturer's protocol M. (1997). TriReagent -RNA, DNA, Protein Isolation Reagent, Molecular Research Center, Inc., Cincinnati, OH.

Marchalonis J. J., and Schluter S. F. (1990). On the Relevance of Invertebrate Recognition and Defence Mechanisms to the Emergence of the Immune Response of Vertebrates. *Scandanavian Journal of Immunology* **32:** 13-20.

Murray M. G., and Thompson W. F. (1980). Rapid Isolation of High-Molecular-Weight Plant DNA. *Nucleic Acids Research* **8:** 4321-4325.

Olano C. T., and Bigger C. H. (2000). Phagocytic activities of the gorgonian coral, Swiftia exserta. *Journal of Invertebrate Pathology* **76:** 176-184.

Salter-Cid L., and Bigger C. H. (1991). Alloimmunity in the Gorgonian Coral *Swiftia exserta*. *Biological Bulletin* **181:** 127-134.

Salzet M. (2001). Vertebrate innate immunity resembles a mosaic of invertebrate immune responses. *Trends in Immunology* **22:** 285-288.

Sambrook J., Fritsch E. F., and Maniatis T. (1989). "Molecular Cloning," Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Stewart Jr. C. N., and Via L. E. (1993). A Rapid CTAB DNA Isolation Technique Useful for RAPD Fingerprinting and Other PCR Applications. *Biotechniques* **14:** 748-749.

Tom M., Douek J., Yankelevich I., Bosch T. C. G., and Rinkevich B. (1999). Molecular Characterization of the First Heat Shock Protein 70 from a Reef Coral. *Biochemical and Biophysical Research Communications* **262:** 103-108.

Williams J. G. K., Kubelik A. R., Livak K. J., Rafaski J. A., and Tingey S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18:** 6531-6535.

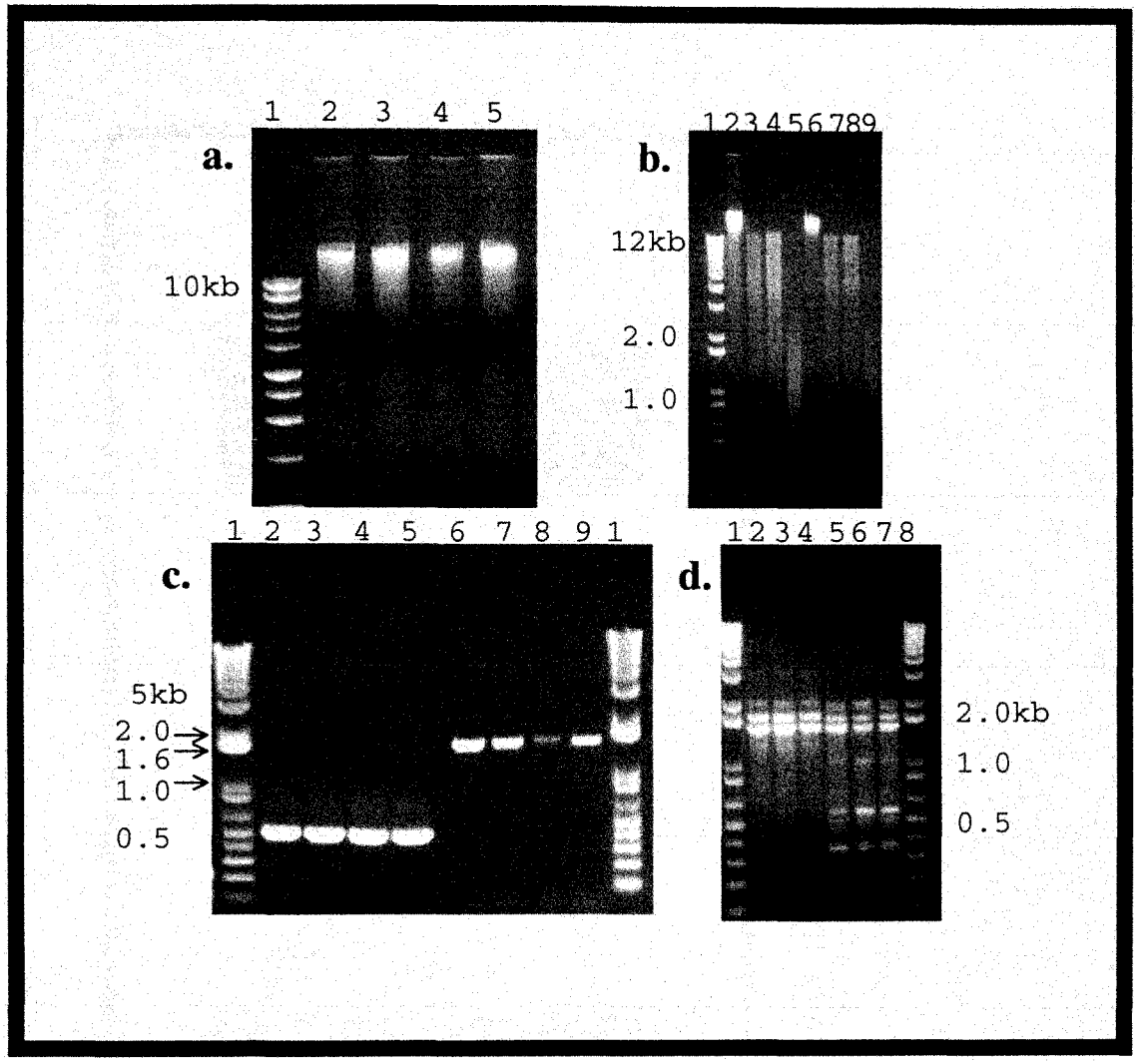


Figure 1. Total genomic DNA (a) extracted from four independent colonies of *Swiftia exserta*. Restriction enzyme digestion (b) of two DNA samples from two colonies. Digestion was performed overnight with Hind III, EcoRI, and Sau 3AI (lane 3-5 and 7-9, respectively). Four independently extracted DNA samples were used as template in PCR-reactions to amplify two different gene segments (c). A portion of the 18s ribosomal DNA gene (lane 2-5) and a portion of the coral thiolester-containing gene (lane 6-9). The RAPD-PCR approach of DNA fingerprinting was used as a easy method to verify purity and integrity of the DNA (d). The same DNA sample was amplified in three independent PCR reactions using either the RAPD 1 primer [5'-CGGTCACTGT; lane 2-4] or the RAPD 1 and RAPD 2 primer [5'-CGGCCCCTGT; lane 5-7] in combination. Reproducibly similar results were obtained from independent DNA extractions from various colonies (data not shown).

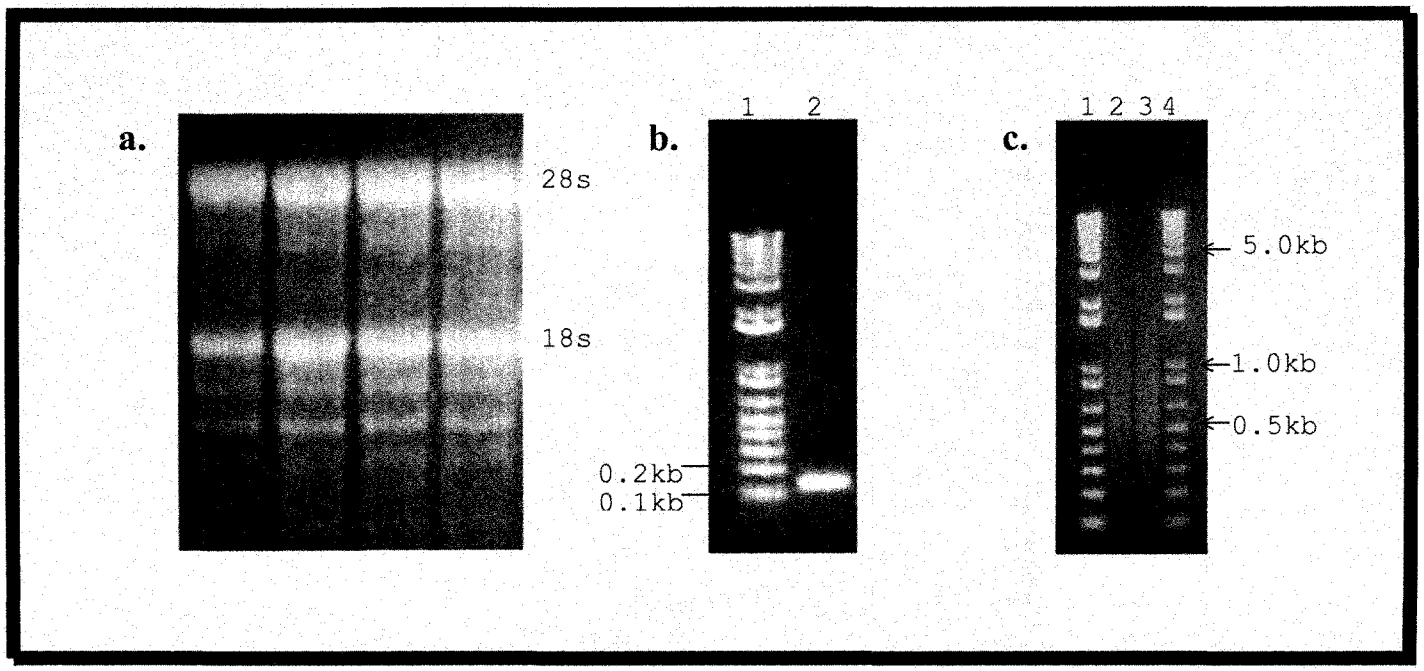


Figure 2. Total RNA isolation (a) from four different corals. RT-PCR results (b) generating the 145bp thiolester gene-specific product. Smart-cDNA PCR amplification results (c), first round (lane 2) and second round (lane 3) where 1ul was diluted and reamplified under the same conditions. Lane 1 (b) and lane 1 and 4 (c) are the same 1kb Plus Ladder (LifeTechnologies, Rockville, MD, USA).

Chapter 3

Applied Molecular Biological Methods

Extraction of nucleic acids

Extraction of both DNA and RNA is discussed in Chapter 2. Please see appendix for detailed protocols of the preferred optimized methods.

Generation of cDNA

Complementary DNAs (cDNA) were made from total or messenger RNA using Superscript II (Invitrogen, Carlsbad, CA) reverse transcriptase enzyme. For degenerate PCR or 5'RACE PCR (see below for both), cDNAs were prepared using the antisense primer designed for the gene of choice. The reverse transcriptase (RT) reaction follows the manufacturer's suggestions, except that for degenerate primer-produced cDNAs, more RNA (about 5 μg) was typically used. I have also found that excellent results can be attained by using 30-60 min at 42° followed by 55°C for 20 min, instead of the standard 1hour at 42°C. For 5' RACE, cDNAs were generally produced using Thermoscript RT enzyme instead. This enzyme was preferentially used for generating longer templates, or for difficult templates because the RT reaction can be done at 65°C. I also found that in some cases, long distance RACE products benefited from combining both enzymes in one reaction mixture and including a 30 min 65°C step. In any case, maintaining an RNAse-free environment is the only way to produce long cDNAs from intact RNA (see appendix for details of RT reaction).

Searching for conserved gene family members

PCR using Degenerate Primers

When interested in a particular protein because it serves a function of interest, one may want to design primers to search for the presence of this gene in their target organism. But because of the degeneracy of the genetic code, several codons can code for a particular amino acid when comparing the same protein across multiple taxa. Because of this, designing primers based on conserved regions of a polypeptide sequence is not straightforward. This is where degenerate primers are designed and used in PCR (Preston, 1996). There are many references and uses of degenerate primers in the literature, and many members of gene families have been cloned utilizing this highly effective PCR method.

In the work described in this dissertation, genes of interest were pursued in the following fashion. Protein sequences were downloaded from the appropriate databases, such as GenBank (http://www.ncbi.nlm.nih.gov/) or EMBL (European Molecular Biology Laboratory, http://www2.ebi.ac.uk/). The sequences from a diverse number of organisms with a broad phylogenetic spread were used to effectively design the degenerate primers. For example, if sequences were only available for mammals, degenerate primer design was avoided unless there was a high degree of confidence that the protein would exist in our animal model. In this case, the design was restricted to a region of the protein that serves critical structural or binding functions. Otherwise, sequences available for invertebrates, as well as vertebrates, provided a good indication of how conserved a particular region may be.

Sequences were downloaded locally to a PC and aligned using Clustal X (Thompson *et al.*, 1997). For further manipulation, sequence alignments were exported as .MSF files (see below) and opened in GeneDoc (Nicholas & Nicholas Jr., 1997). Conserved regions were shaded (allowing for conservative substitutions as well) and regions of high conservation were determined by eye. For example, in designing a degenerate primer of sufficient length, sites of interest were limited to 6-7 amino acids in length. This provided at least 18 bases for the primer. A perfectly conserved 3' (or C-terminal) -most amino acid, for example, was preferred in all cases. Primers were designed so that the 3' end was a 1^{st} or 2^{nd} codon position. In general, the 3'-most base, should never be degenerate; although I have successfully used primers with 3'-degeneracy. Two conserved sites were chosen along the length of the protein, so that the separated distance between the primers at the DNA level would produce a PCR product in the range of 150-350 bp, with 250 bp as optimal. Degenerate primers for a PCR product in excess of 500 bp is highly ambitious and rarely works because the kinetics of the reaction do not allow sufficient stringency.

Degenerate primers were designed with using the IUPAC/IUB code, each representing more than one DNA base. These include: R for A/G, K for G/T, S for G/C, B for G/C/T, H for A/C/T, N for A/C/T/G, Y for C/T, M for A/C, W for A/T, D for A/G/T or V for A/C/G. For example, aspartic acid (D) is coded for by two codons and hence the codon sequence in the primer will appear as GAY since the codon can be GAC or GAU. A recent review by Preston (Preston, 1996) provides a nice introduction to the science of using degenerate PCR. Highly degenerate primers, though, tend to produce significant background because the PCR is done at low temperatures and the primer mix

literally contains thousands of different primers. Because of this, and because of the existence of codon usage bias in many organisms and proteins (Wada *et al.*, 1990), I routinely realigned the region of interest at the DNA level and compared the codon sequence used. If a codon was preferentially used in all animals compared, then the degeneracy was reduced or removed completely from that codon.

PCR conditions can be any of the typically used formats, with the two major exceptions: annealing temperature is lower and the number of cycle is increased. Because the Tm is unknown and binding is desired at sites that may be slightly divergent, a lower temperature is typical. And because the actual template specific primer in the mix will actually be at very low concentration (because it is a mix of many primers), I routinely used much higher concentrations of the primers (5-10x more) and about 15 cycles above the standard 30. Without doing this, PCR products may be difficult to see, especially when a smeary background is present. In this work, degenerate PCR primers were used with Oligo-dT primed cDNAs or cDNAs that were primed by the antisense degenerate primer of interest. Standard conditions used were 95° for 5min, and 45 cycles of 95°for 1min, 37-42° for 1min and 72° for 1min, followed by a 15 min final extension at 72°. With the recent production of gradient thermal cyclers, degenerate PCR can now be done with a gradient annealing temperature of 37-55 ° to help eliminate background.

Designing Gene Specific Primers

All gene-specific primers for this work were produced with a MS-DOS version of the PRIMER DESIGNER program (ver 1.01; 1990, Scientific and Educational Software). More current Windows-based versions of this program are available, albeit, at a

considerable cost. The old versions, such as the one described here, work well because of their simplicity, and can occasionally be found free of charge. The program is available from the author if it cannot be found. Online versions of primer design programs are also available (i.e., Primer3, ver 2 at: http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)

Getting the full-length sequence of the partial gene products

Rapid Amplification of cDNA Ends-- RACE-PCR

For 3' and 5' RACE, the classic procedures of Frohman (Zhang & Frohman, 1997) were preferred to the commercial kits now available for RACE. There were several reasons for this. Classic RACE works well, and is quite reproducible, but tends to give many RACE products at the 5' end (see below for reason). It is also technically easy and requires very little extra material other than what is usually available in a molecular biology lab: Taq polymerase, TdT enzyme, dNTPs, and three extra primers (Q0, Q1, and Qt) which are easily made. But because of the background traditionally seen with classic RACE, particularly at the 5' end, modifications to the original technique have been proposed which were originally based on adaptor ligations to double stranded cDNAs (Chenchik *et al.*, 1996) or mRNA (Maruyama & Sugano, 1994; Shaefer, 1995). Many of these modifications have become commercially available as kits (i.e., Clontech, Ambion) for a hefty price. Most of the kits are composed of a few extra primers (or adaptor oligos) and enzymes. If the oligo sequences are known, and the user is comfortable with

the procedures, the kit is easily replaceable by purchasing the materials separately. A major modification to the classic 5' RACE procedures is the ligation of an adaptor to the 5' end of full-length mRNA , which becomes incorporated into the cDNA and can be used as a priming site in subsequent PCR reactions. Therefore, a major limitation of this technique (which I have tried with mixed success) is that if your target RNA is partially degraded at the 5' end (missing the cap sequence) or is too large a sequence for optimal PCR amplification, no product will result in the subsequent PCR. Unless DMSO is used, long sequences tend to have a lot of secondary structure and large 5' amplifications will not work. The enzyme used may also limit long template amplification; therefore a Pfu derivative, instead of Taq, may work better. The kits do not provide this information. Classic RACE, in contrast, provides smaller 5' RACE products but allows the user to progress along the template. The 5' anchored RACE approach appears to be beneficial for small genes and/or for determining the final 5' UTR sequence (of only several hundred bases).

RACE is composed of two parts, 3' and 5' extension (or amplification) of the sense sequence of the gene (mRNA) of interest. Most partial gene products amplified are somewhere in the middle of the gene's sequence. After orientation is determined (sense vs antisense), primers for RACE are established. For 3' RACE, we are interested in the sequence of the mRNA towards the poly-A tail. Therefore, we can establish cDNAs using an Oligo-dT primer that anchors on the poly-A tail of the gene. PCR amplification then proceeds using a modified primer at the oligo-dT end (Qt) and a sense primer designed to the already established sequence in the internal portion of the gene. For both

3' and 5' RACE, the Q0 and Q1 primers provide nested reactions to the Qt primer (see below).

In 5' RACE, the 5'end sequence of the gene (mRNA) is of interest. Usually, this portion of the gene is too far from the poly-A tail, so that the best results for RACE may be attained by priming within the already known sequence using an antisense primer to establish the cDNAs. In classic RACE, cDNAs extend as far as they can and then after the reverse transcriptase reaction is complete, the 3' end of the cDNAs (corresponding to the 5' end of the mRNA template) are poly-adenylated using dATP and TdT enzyme (terminal deoxynucleotidyl transferase). This procedure adds a poly-A tail to the cDNAs which allows one to anchor at this region with a modified Oligo-dT primer and the gene-specific antisense primer in the subsequent PCR. Because poly-adenylation occurs at the 3'end of all the cDNAs generated (including all the partial transcripts), multiple products are common in 5' RACE. But if all the products are real (corresponding to the gene of interest), then the obvious thing to do is to select the largest band available. In this work, I always performed RACE-PCR using two rounds of PCR, where in the second round, an aliquot of the 1st round products were reamplified using an internal (nested) primer. This eliminated most non-specific amplification products because only true race products contain the internal gene-specific priming site. See appendix for example protocol.

Primers mentioned above include, from (Zhang & Frohman, 1997):

Qt— 5' CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC(T17)-3

Q0—5' CCAGTGAGCAGAGTGACG

Q1—5' GAGGACTCGAGCTCAAGC

Screening of cDNA libraries

In this work, library screening was not used to clone gene products, because RACE was generally successful. But, several libraries were created by the author and are now available for screening or for use as a template in RACE-PCR reactions as well. cDNA libraries were created in Zap-lambda cloning vectors (Stratagene) and the packaged, amplified phage was stored at –80°C. Partial PCR products can be used to screen a library by infecting host bacterial cells with the phage, plating the cells and probing the lysed colonies with a labeled probe of the desired sequence. Positives are subcloned into a plasmid-type vector, transformed into a bacterial host, plated, and the positive colonies established (see Sambrook and Russell, 2001). From these clones, plasmid is extracted and sequenced to study the cloned cDNA products.

Additionally, aliquots of the cloned, packaged phage can also be used as a template for PCR to "RACE" out sequence (primarily from the 3' end of the gene) using a gene-specific primer and one anchoring primer (usually anchoring to the phage sequence itself). For cDNA library technology, background and instruction, the reader is referred to standard works (e.g., Cowell & Austin, 1997; Sambrook & Russell, 2001).

Assembling sequences

Once multiple RACE-PCR sequences were acquired, they were assembled primarily by eye and by utilizing the Find tool in Microsoft Word. This was easily done because the primer sequences were known, and it was expected that each over-lapping clone would overlap beginning with the priming site and that the intervening sequence would be identical. This procedure is only useful when the user is assembling clones in a step-wise manner after they have been collected. If dozens of overlapping clones are blindly sequenced, for example, then there are proprietary programs (i.e., AutoAssembler, part of the Perkin-Elmer Automated Sequencing Software Suite) that assemble the sequences into one overlapping reading frame. This is a common approach in high-throughput shotgun sequencing approaches, an approach not used in this work.

Reconfirming areas of ambiguity

For degenerate PCR, after each band of interest was cloned, 50-75 clones were routinely established, sequenced, and screened for the presence of different gene family members. This also allowed for confirmation of sequence ambiguities or artifacts that may have accumulated during PCR or cycle sequencing reactions. For gene-specific primer-amplified products, the cloned products' sequence was determined by sequencing 10 clones on average. This allows for the location of sequencing (or PCR introduced) artifacts, which are usually only found in a few of the clones. Any areas of ambiguity, for example areas where true polymorphism may be suspected, new gene specific primers were designed flanking the region of interest. This portion of the gene was then reamplified out, cloned, and the sequence determined from 10-20 clones. The appendix

for each appropriate chapter in this work contains a figure of the full-length gene sequences with the relevant primer sites highlighted.

Cloning of PCR products and sequencing

All PCR products were cloned into a TA-based pGem vector system using either Invitrogen's TOPO-Cloning kit (Carlsbad, CA, USA) or Promega's Pgem Teasy kit (Madison, WI, USA). All cloned or gel-purified products were sequenced on an automated sequencer (ABI 377, Perkin Elmer) using the BigDye Terminator kit (PE-Biosystems) for the dideoxyterminator cycle sequencing technology.

Confirmation by Northern and Southern Blot analysis

Northern blot analysis was utilized to confirm expression of cDNAs and cloned gene products, and to estimate size of full-length transcripts. Two versions of Northern blotting were performed: RNA-probed blots (Krumlauf, 1996) and DNA-probed blots (Sambrook & Russell, 2001). Since RNA-RNA hybridizations are extremely stable and difficult to remove, RNA-RNA hybridizations gave more background and the nylon membranes were almost impossible to strip and reuse. Unfortunately, some of the strongest signals are seen with RNA probes, but the extensive background can produce confusing results. Because of this, for Northern blotting, DNA-probed blots were preferred. All probes were radioactively labeled with ^{32}-P (d)NTPs (Amersham Biosciences).

The presence of a gene in the coral genome, along with its genomic organization, was studied using Southern Blotting techniques. ^{32}-P-(d)CTP was the preferred isotope and used in random priming reactions to label the probes (Amersham Biosciences), which were purified with G50 spin columns. High stringency phosphate-based hybridization buffers were preferred because they minimize background by allowing Southern hybridization to be performed at 65°C (Sambrook & Russell, 2001). See appendix for protocols for both Northerns and Southerns.

Phylogenetic analysis of gene products

All phylogenetic analyses of genes described in this work were performed primarily using the Windows based platform of the multi-use programs, Mega2 (Kumar *et al.*, 2001), PHYLIP (Felsenstein, 1995) , and PAUP 4 * (Swofford, 1998). Alignment of sequences, either DNA or protein, was performed with the Clustal X program (Thompson *et al.*, 1997). Alignment of DNA sequences that were even slightly divergent sometimes required increases in gap opening and extension penalties of at least 5 times the default value. Along these lines, when producing multiple sequence alignments for amino acid sequences belonging to multigene families, producing profile-type alignments first generally produces better results. Random input order of sequences is important to prevent biasing results in favor of sequence input order. Alternatively, when performing multiple sequence alignments (global alignments) of many large and divergent sequences, those that do not align correctly can be selected in the Clustal X program and realigned to the other members. This usually corrects uncertain alignments. Alignments

produced by profiles can also be realigned in global format to correct ambiguously aligned areas. It is important to keep in mind that the congruence and reliability of phylogenetic analyses is completely dependent on how accurate the initial alignments are.

All alignments were exported as .MSF files and opened for shading and editing in GeneDoc (Nicholas & Nicholas Jr., 1997), which also produces several statistical analyses and shades for physiochemical conservation and structural comparison. Additional sequence manipulations were also performed using the Sequence Manipulation Suite (Stothard, 2000). With the exception of PAUP, all of the above mentioned programs are available free of charge on various internet servers. Phylogenetic analysis is a comprehensive science which should be well understood by the user. Misrepresentation and misinterpretation of data commonly results from novice users of these programs who have little understanding of the implications of their results. Users new to phylogenetic analysis are strongly encouraged to read the program documentation files and the following sources as a reference point (Hillis *et al.*, 1996; Hughes, 1999; Li, 1997; Maddison & Maddison, 1992; Nei & Kumar, 2000).

***In Situ* hybridization**

In the course of this work, several *in situ* hybridization protocols were optimized to work in Cnidarians. Much of this inspiration came from previous work in our lab (Olano, 1993) and in my interest to localize expression of genes during allograft and wound healing events. Extensive experience was gained from my invited involvement in a side project working with a more difficult Cnidarian, a cubomedusan jellyfish

(Piatigorsky *et al.*, 2001). This work mostly consisted of frozen section and whole-mount *in situs* (based on modified techniques from Wilkinson & Nieto, 1993& L. Kos, personal communication) hybridized with dig-labeled RNA probes to localize expression patterns of eye crystallin proteins. These techniques have also successfully been applied to *Swiftia* (work in progress & manuscripts in preparation). Some protocols associated with fixation and parafin-embedding were also optimized (Darby, 2000; Presnell & Schreibman, 1997& K. Condon, personal communication) for use in *Swiftia*, primarily for antibody staining.

References

Chenchik A., Moqadam F., and Siebert P. D. (1996). A new method for full-length cDNA cloning by PCR. *In* "A Laboratory Guide to RNA: Isolation, Analysis, and Synthesis" (P. A. Krieg, Ed.), pp. 444, Wiley-Liss, New York.

Cowell I. G., and Austin C. A., Eds. (1997). "cDNA Library Protocols," Humana Press, Totowa, NJ.

Darby I. A., Ed. (2000). "*In Situ* Hybridization Protocols," Humana Press, Totowa, NJ.

Felsenstein J. (1995). PHYLIP: Phylogeny Inference Package, version 3.57c.

Hillis D. M., Moritz C., and Mable B. K. (1996). "Molecular Systematics," Sinauer Associates, Inc., Sunderland, MA.

Hughes A. L. (1999). "Adaptive Evolution of Genes and Genomes," Oxford University Press, New York.

Krumlauf R. (1996). Northern Blot Analysis. *In* "Basic DNA and RNA Protocols" (A. J. Harwood, Ed.), Humana Press, Totowa, NJ.

Kumar S., Tamura K., Jakobsen I. B., and Nei M. (2001). Mega2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics (submitted)*.

Li W.-H. (1997). "Molecular Evolution," Sinauer Associates, Inc., Sunderland, Mass.

Maddison W. P., and Maddison D. R. (1992). "MacClade: Analysis of Phylogeny and Character Evolution," Sinauer Associates, Inc., Sunderland, MA.

Maruyama K., and Sugano S. (1994). Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138:** 171-174.

Nei M., and Kumar S. (2000). "Molecular Evolution and Phylogenetics," Oxford University Press, Oxford.

Nicholas K. B., and Nicholas Jr. H. B. (1997). GeneDoc: a tool for annotating and editing multiple sequence alignments. *Distributed by author.*

Olano C. T. (1993). Cellular Aspects of Alloimmunity and Other Responses in the Gorgonian *Swiftia exserta*, Master's Thesis, Florida International University, Miami.

Piatigorsky J., Norman B., Dishaw L. J., Kos L., Horwitz J., Steinbach P. J., and Kozmik Z. (2001). J3-crystallin of the jellyfish lens: Similarity to saposins. *Proceedings of the National Academy of Sciences* **98:** 12362-12367.

Presnell J. K., and Schreibman M. P. (1997). "Humanson's Animal Tissue Techniques," The John's Hopkins University Press, Baltimore.

Preston G. M. (1996). Polymerase chain reaction with degenerate oligonucleotide primers to clone gene family members. *In* "Basic DNA and RNA Protocols" (A. J. Harwood, Ed.), pp. 514, Humana Press, Totowa, NJ.

Sambrook J., and Russell D. W. (2001). "Molecular Cloning: a laboratory manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shaefer B. (1995). Revolution in rapid amplification of cDNAs: new strategies for polymerase chain reaction of full-length cDNAs. *Analytical Biochemistry* **227:** 255-273.

Stothard P. (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28:** 1102-1104.

Swofford D. L. (1998). PAUP*: Phylogenetic analysis using parsimony (and other methods), Sinauer Associates, Sunderland, MA.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., and Higgins D. G. (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24:** 4876-4882.

Wada K.-N., Aota S.-I., Tsuchiya R., Ishibashi F., Gojobori T., and Ikemura T. (1990). Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Research* **18:** 2367-2411.

Wilkinson D. G., and Nieto M. A. (1993). Detection of Messenger RNA by in Situ Hybridization to Tissue Sections and Whole Mounts. *In* "Methods in Enzymology" (P. M. Wassarma, and M. L. DePamphilis, Eds.), Academic Press, San Diego.

Zhang Y., and Frohman M. A. (1997). Using Rapid Amplification of cDNA Ends (RACE) to Obtain Full-Length cDNAs. *In* "cDNA Library Protocols" (I. G. Cowell, and C. A. Austin, Eds.), pp. 61-88, Humana Press Inc, Totowa, NJ.

Chapter 4

Origins of the thiolester-containing protein (TEP) family in Cnidaria: Analysis of a complement C3-like gene from a coral.

Abstract

A full-length cDNA sequence of a C3-like (SeC3) gene has been cloned from the coral, *Swiftia exserta*. RT-PCR with degenerate primers initially yielded a 214bp product with sequence similarity to vertebrate C3 and A2M proteins. Northern blot analysis showed the gene to be approximately 6kb and expressed in normal, unstimulated tissue. Rapid amplification of cDNA ends (RACE) in the 3' and 5' direction yielded the complete cDNA sequence of 5.5 kb, with one open reading frame of 1728aa and two functionally relevant polypeptide cleavage sites. The deduced polypeptide contains a thiolester site, the C3-specific catalytic histidine, a complement-specific anaphylatoxin region, and two arginine-rich cleavage sites (suggesting a three chain structure). Preliminary Southern blotting experiments confirm the presence of SeC3 in the coral genome and may further suggest a complex genomic organization (multiple introns) or the existence of another TEP in the coral that cross-reacts with the probe. While functional studies remain to be carried out, physiochemical and structural properties based on deduced amino acid sequence, along with phylogenetic analyses, indicate it to be homologous to C3/C4/C5 with more specific similarity to C3. The data described here is the first evidence of a complement-related protein outside the deuterostome lineage. In addition, this work supports previous suggestions that the ancestor to the C3/C4/C5 group of TEPs was a C3-like protein. The existence of this gene in a Cnidarian suggests that an ancestral complement gene existed during the Pre-Cambrian era, when diploblasts and triploblasts diverged into modern metazoans. Furthermore, preliminary evidence for a second thiolester-containing protein (TEP) in the coral might be an indication that some

of the duplication events, giving rise to the paralogous TEP family, could have occurred very early in phylogeny.

Introduction

Innate immune mechanisms are the most ancient and versatile defense systems possessed by organisms (Janeway & Medzhitov, 2002; Salzet, 2001). They are often the first and only line of defense at the site of infection. A critical component of innate immunity, particularly in vertebrates, is the complement system, a collection of soluble serum proteins which by three enzyme activation cascades (classical, alternative, and lectin pathways) share a terminal lytic pathway to produce the membrane attack complex (MAC) and biologically active intermediates that serve as opsonins to tag micro-organisms or other target antigens for phagocytosis (Barrington *et al.*, 2001; Carroll, 1998). The MAC effectively lyses most cells by punching holes into the outer membranes. The third component of complement (C3), which is a thiolester-containing protein (TEP), is common (intersects) to all three pathways. This molecule is the central, and most critical component of complement activity, and quite possibly of innate immunity (Carroll, 1998; Dempsey *et al.*, 1996; Sahu & Lambris, 2001).

Alpha 2-macroglobulin (A2M), a non-complement-related thiolester-containing protease inhibitor, and complement components C3, C4 and C5 are paralogous genes believed to have diverged after the protostome and deuterostome split such that C3/C4/C5 are exclusive to the deuterostome lineage (Dodds & Law, 1998; Zarkadis *et al.*, 2001) (see Fig. 1). This family of proteins appears to have evolved with the major

function of tagging (opsonization) microorganisms or immune complexes (or soluble antigen) for clearance via phagocytosis (Dempsey *et al.*, 1996; Levashina *et al.*, 2001; Smith *et al.*, 1999). Opsonization occurs primarily through intermolecular covalent interactions via a thiolester site common to this family of proteins (the exception is C5, which lost its thiolester and was co-opted into the terminal lytic pathway) (Gadjeva *et al.*, 1998) (Fig. 2).

C3b and C4b, the major activation products of C3 and C4 respectively, are the key opsonic molecules generated by complement activation. Opsonized targets are then cleared by complement-receptor bearing phagocytes (Dempsey *et al.*, 1996). An original TEP that could act as an opsonin was a pivotal innate immune acquisition in early metazoan evolution (Dodds & Law, 1998; Levashina *et al.*, 2001; Nonaka *et al.*, 1999). Phagocytosis of opsonized macromolecules and/or cells plays a crucial role in removing harmful/unwanted substances and maintaining an organism's integrity (Dempsey *et al.*, 1996).

The finding of paired gene-duplication products as functioning components of the mammalian complement system is not uncommon (Campbell *et al.*, 1988; Zarkadis *et al.*, 2001) and suggests that the diverse complement systems of higher vertebrates evolved from co-opted diverging gene-duplication products of simpler systems (Jensen *et al.*, 1981) having a broader range of functions. Comparative studies aimed at understanding the phylogenetic origins of the complement system have primarily focused on the deuterostome lineage. Protostome invertebrates (*Drosophila*, *Anopheles*, and *C. elegans*) have been demonstrated to contain divergent A2M-like TEPs which are not considered analogous to vertebrate complement components (Levashina *et al.*, 2001).

BLAST searches of the sequenced *Drosophila* and *Caenorhabditis elegans* genome (BLAST, as Blastx, Blastn, and PHI-BLAST, http://www.ncbi.nlm.nih.gov/BLAST/; Drosophila Genome Project, http://www.fruitfly.org/; Flybase, http://flybase.bio.indiana.edu/; Sanger Center project, http://www.sanger.ac.uk/Projects/C_elegans/, and Washington University Genome Project, http://genome.wustl.edu/) further suggests the absence of any C3/C4/C5-like components and the presence of orthologous and paralogous (divergent copies) A2M-like members (LJD, personal observations). Therefore, formation of complement-like paralogous genes is believed to have occurred after the phylogenetic divergence of protostomes and deuterostomes (Dodds & Law, 1998; Smith *et al.*, 1999; Zarkadis *et al.*, 2001) (see Fig. 1).

To further understand the origins and evolution of the TEP family, a homologue was sought in an extant representative of an ancient phylum (Cnidaria) which predates the divergence of protostomes and deuterostomes (P-D). This work reports on the cloning of a C3-like cDNA (SeC3, Genbank accession No. AY186744) from the endosymbiont-free gorgonian soft coral, *Swiftia exserta*. Fossil evidence has suggested that corals existed prior to the Cambrian period (as early as 700mya) (Ayala *et al.*, 1998; Fortey *et al.*, 1997; Margulis & Schwartz, 1998; McMenamin & McMenamin, 1990; Raff, 1996; Valentine *et al.*, 1991) and recent protein-clock estimates places the P-D divergence at about 670mya (Doolittle *et al.*, 1996). The divergence of a C3-like gene now appears to have occurred sometime within the Pre-Cambrian, where diploblasts (Cnidarians) and triploblasts (P-D) diverged into a multitude of metazoan ancestors.

Materials and Methods

Collection and maintenance of animals

Swiftia exserta (Phylum Cnidaria, Class Anthozoa) was collected off the coast of Southeast Florida in approximately 20-30 m of water. The live animals were transferred to FIU where they were maintained in seawater aquaria (35-37 0/00; 21-23°C) with alternating light-dark cycles (14 & 10 hrs, respectively). The animals were fed freshly hatched brine shrimp (*Artemia* sp.) larvae every other day.

Isolation of RNA

Total RNA was isolated with TriReagent (Molecular Research Center, Cincinnati, OH, USA) using high salt precipitation as suggested by the manufacturer. RNA was stored as a pellet in 70% ethanol at -20°C or otherwise at -80°C until ready for use. Traces of genomic DNA were removed from the RNA using DNase I (Promega, Madison, WI) treatment.

cDNA synthesis and degenerate PCR

cDNA synthesis was performed with Superscript II or Thermoscript (5' RACE reactions) reverse transcriptases (Invitrogen, Carlsbad, CA, USA). For degenerate PCR, cDNAs were created in a degenerate primed reverse transcription (RT)-reaction using 5-10μg of total RNA in a 20 μl reaction with 400 μM of dNTP and Superscript II enzyme.

The RNA was initially melted in the presence of 250 pmol of degenerate antisense primer (5'- see below) at 80°C for 3min and quenched in an ice-water bath for 2min before the addition of the RT reaction mix. The RT reaction was incubated for 1 hour at 42°C. Five microliters of the RT reaction was used as template along with 250 pmol of each degenerate primer (AS- 5'-ACRTANGCNGTNAGCCANGT and S-5'-GNTGYGGNGARCARAAYATG) in a 50μl degenerate PCR reaction as follows: 95°C for 5 min and 45 cycles of 1 min at 95°C, 1 min at 42°C, and 1 min at 72°C, followed by a 10 min final extension at 72°C.

For 3'RACE (Zhang & Frohman, 1997), cDNA was created as follows: 1-2 μg of total RNA in a 20 μl reaction under standard reaction conditions, using Superscript II (Invitrogen). The RNA was melted in the presence of 20 pmol of RACE-modified (see below) oligodT primer (Qt) at 80°C for 3min and quenched in ice-water bath for 2min before the addition of the RT reaction mix. The RT reaction was incubated for 1 hour at 42°C and 15 min at 50°C and all reactions were stopped by incubating at 70°C for 5min and treating with RNase H for 20min. Two microliters of the diluted (2.5x) reaction was used as template for RACE PCR (see below).

For 5'RACE (Zhang & Frohman, 1997), cDNA was created as follows: 1-2 μg of total RNA in a 20 μl reaction under standard conditions, using Thermoscript (Invitrogen) RT enzyme. The RNA was melted as described above in the presence of 20 pmol of antisense gene-specific primer. The reactions were overlaid with a drop of mineral oil and were incubated at 65°C for 1hour. These reactions were stopped with a 5min incubation at 80°C, and incubated for 20min at 37°C in the presence of RNase H. The

reactions were precipitated with 0.5 vol of 7.5M ammonium acetate (NH4Oac) and 2.5vol of EtOH. The precipitated cDNAs were washed with 70% EtOH. The cDNAs were resuspended in 10μl of water and poly-adenylated at the 3'ends with 10U of terminal deoxynucleotidyl transferase (TdT) enzyme (Promega) in the presence of 4 μl of 1 μM dATP and 4 μl of 5x TdT buffer. Two micoliters of the diluted (2.5x) reaction was used as template for RACE-PCR (see below).

RACE-PCR and cloning of products

Rapid amplification of cDNA ends (RACE) was carried out according to the conventional (described as *Classic*) procedures (Zhang & Frohman, 1997). In 5'RACE, Thermoscript RT-polymerase (Invitrogen) was utilized with gene-specific antisense primers to prime the cDNA synthesis reaction. In the PCR steps of the 5'RACE, we used 1% DMSO to help facilitate the production of some of the more difficult regions of the gene. All RACE products were confirmed with nested PCR reactions and were gel purified (Qiagen) and cloned into TOPO-TA cloning vectors (Invitrogen).

Northern and Southern blot analysis

For Northern blot analysis, total RNA was extracted as described above and separated on a 1% formaldehyde gel and transferred to a positively charged nylon membrane (Hybond XL, Amersham Bioscience). Probes were generated either as riboprobes (Northern) (Krumlauf, 1996) or random primed reactions (Northern and

Southern blots)(Sambrook & Russell, 2001). Riboprobes were generated as run-off transcription reactions (with ^{32}P α-ATP) directly from the TOPO vectors essentially as recommended by the manufacturer's protocol for the T7/SP6 enzymes (Roche Biochemical). Northern hybridization using riboprobes followed previously described methods (Krumlauf, 1996). Random priming reactions were performed with the Mega Prime Labeling kit (Amersham Biosciences) using ^{32}P α-dCTP and SeC3-specific PCR products as templates. All hybridization conditions for the DNA-probed Northerns and Southerns followed established protocols using high stringency phosphate-based hybridization buffers (Sambrook & Russell, 2001) and are described in the Appendix. Hybridization occurred essentially at 60-65°C overnight in a buffer composed of 0.5M NaPO4 buffer, 1mM EDTA, 7% SDS and 1% BSA (w/v).

Five micrograms of genomic DNA was digested in the presence of EcoRI, PvuI, KpnI, SalI, HindIII, DraI, and Sau3AI (Promega) for 24hours. The digested DNA was run on a 0.7% TAE-agarose gel and transferred to a nylon membrane (Hybond XL) under alkaline conditions (Sambrook & Russell, 2001) and probed with a gel-purified random-primed PCR product as described above.

Assembly and analysis of cloned sequences

All cloned sequences consisted of overlapping RACE clones. As a result, assembling of the sequences at each step was, essentially, performed manually. Detail analysis of the sequences for all clones, which include sequencing of multiple transformed colonies, was performed by aligning the DNA sequence in Clustal X

(Thompson *et al.*, 1997). Verification of each product used to extend the sequence of SeC3 was determined by aligning the DNA sequence from at least 10 clones. Sequence Manipulation Suite (Stothard, 2000) and Genedoc (v2.5) (Nicholas & Nicholas Jr., 1997) were used to manipulate individual sequences and alignments.

Analysis of deduced amino acid sequence

All RACE-produced sequential extensions of the SeC3 sequence resulted in one deduced amino acid translation in one reading frame in the 5'-3' direction with no stop codons. The translated sequences within each of the other two alternative reading frames were always interrupted with multiple stop codons. Full length amino acid alignments were produced using TEP sequences available in the GenBank and Swiss-Prot Databases (table 1, and see Ch.4 appendix) with the *Clustal X* program and the Gonnet matrix (Gonnet *et al.*, 1992) under profile and global alignment conditions. Profile alignments were produced by aligning (in random taxonomic order) orthologous TEP proteins. All profiles were then aligned to each other, in random order. Representatives from the entire TEP family of sequences (N=45) were also aligned simultaneously using Global alignment parameters, and the alignments and phylogenetic analyses compared to results attained from profile alignments (see Results). Visual inspection, along with inspection by shared conserved physiochemical properties, was performed in the GeneDoc program so that any sequence(s) appearing to have been misaligned could be identified. Any sequence(s) appearing to be misaligned in some or multiple regions was selected in Clustal, gaps removed, and realigned against all other members. This typically corrected

most errors, though some minor errors (mismatches) were corrected by eye. For all alignments, the most reasonable results were produced by using Gap open penalties of 20 and extension penalties of 0.40- 1.0.

Using the alignments and known structural information about the TEP family members, the conservation of structural and functional sites were compared. Pairwise comparisons (alignments) were produced in calculating distance scores, percent identity and percent similarity using Mega2 (ver. 2.0)(Kumar *et al.*, 2001), GeneDoc, and Sequence Manipulation Suite. Secondary structure predictions were performed using the PSIPRED and PHD package (McGuffin *et al.*, 2000; Rost, 1996). The polypeptide sequence was also threaded through the Swiss-Modeling server to predict 3D structure utilizing the comparative modeling approach (Guex & Peitsch, 1997; Leach, 2001; Peitsch *et al.*, 2000) against the recently crystallized human C3d protein (Nagar *et al.*, 1998). This method utilizes multiple sequence alignments to predict a three-dimensional structure based upon the known structure of at least one or more proteins in the protein data bank (PDB). Secondary structural analyses were also performed using a similar comparative threading approach (McGuffin *et al.*, 2000; Rost, 1996).

Hydrophobicity profiling of SeC3

The human factor B and H and complement receptor I-III binding region, and the properdin binding region are regions unique to C3 proteins (Morley & Walport, 2000). Hydrophobic and hydrophilic characteristics of a protein can provide important information regarding its structural organization, its function with regards to substrate

interaction, and/or its antigenic character (Hoop & Woods, 1981; Kyte & Doolittle, 1982). Hydropathy profiles were produced by the Kyte and Doolittle method (Kyte & Doolittle, 1982). Pairwise sequences to be compared were aligned in *Clustal X* and hydropathy profiling performed with the program, *BioEdit* (Hall, 1999). Hydropathy profiling was employed to predict the presence or conservation of these two major regions in SeC3.

Sequences used for phylogenetic analysis

Forty five members of the TEP family were used in the subsequent phylogenetic analysis and comparisons. All sequence names and database accession numbers are available in table 1. In general though, the TEP family can be broken into 5-6 major paralogous groups. Alpha-2-macroglobulin (A2M) includes vertebrate and invertebrate forms, in addition to the A2M-like paralogous genes which include muriglobulins, alpha 1-inhibitors, endodermin, ovastatin, and pregnancy zone protein. The protostome TEPs, which include *Drosophila* TEP 1-4, mosquito TEP1, and *C.elegans* TEP1 are very similar to A2M proteins. They are most likely paralogous to A2M, but have been separated for almost 700my. Phylogenetic analysis usually clusters them as a sister group to A2M. Vertebrate C3, C4, and C5 are paralogous proteins to A2M. Invertebrate C3-like proteins are TEPs that are very similar to vertebrate C3, but are not orthologous. The invertebrate C3-like proteins most likely represent extant versions of the C3/C4/C5 ancestral sequence (Nonaka *et al.*, 1999). A major difference between the complement proteins (C3, C4, C5) and all the A2M-like proteins (including the divergent insect and

worm TEPs) is size. All A2M proteins are 200-250aa shorter at the C-terminal end than the complement proteins, which are all of similar size (Figure 2a). In mammals, this size difference corresponds to 4-5 missing exons at the C-terminal end of A2M. The deuterostome invertebrate C3-like proteins that have been characterized are all of similar size to the vertebrate complement forms (C3, C4, and C5). Another major difference is the specialization of the anaphylatoxin region (C3a, C4a, and C5a; Figure 2a) for a single reactive protease. This is in contrast to the highly polymorphic corresponding region of A2M, which is called a "bait region" and is reactive with many protease types (Armstrong & Quigley, 1999; Quigley & Armstrong, 1994)

Phylogenetic analysis of SeC3

Alignments were produced as described above using full-length TEP polypeptide sequences. Phylogenetic analysis was performed using the Minimum Evolution (ME) distance method (Kumar, 1996; Rzhetsky & Nei, 1993) with the Mega2 program (Kumar *et al.*, 2001) and pairwise deletion of gaps. All sequences used for the analyses can be found in Table 1. Phylogenetic trees were constructed on the basis of two amino acid distance methods: uncorrected proportion of difference (p-distance) and Poisson-corrected amino acid distance. When analyzing divergent genes, p-distance scores may be large and the resulting variance associated with the correction formula may become too high. In this instance, only uncorrected phylogenies would be reliable (Nei, 1991). In addition, Poisson-correction assumes equal rates of substitution among the length of the overall protein. Unfortunately, this is an unrealistic assumption, especially in the case

of highly divergent proteins. In this case, it would be more reasonable to predict that the rate of change is following some sort of heterogeneous gamma distribution. The gamma shape (with parameter,α) should be estimated for each protein family, but generally requires alignment of sequences that do not contain gaps or other ambiguities (Gu & Zhang, 1997). This, unfortunately, is difficult with sequences as large and divergent as the TEPs. Therefore, both corrected and uncorrected phylogenies were produced, and the resulting topologies (and their statistical significance) were compared.

Minimum evolution (ME) distance methods (with neighbor-joining as a heuristic search alternative) can out-perform other distance methods of phylogenetic tree reconstruction when comparing sequences from a homologous multigene family which have undergone a considerable amount of divergence (Li, 1997; Nei, 1991; Nei & Kumar, 2000; Rzhetsky & Nei, 1993). Some of this data, though, has been generated from simulation studies, which some authors have argued (Hillis *et al.*, 1996; Page & Holmes, 1998; Swofford *et al.*, 1996) do not represent real-life scenarios. In addition, it appears that the biases which produce better performance for the ME method disappear when sequence length increases (Swofford *et al.*, 1996).

Because the TEP proteins consist of large polypeptides, this condition was tested by performing phylogenetic analysis on the exact same data set using the Fitch-Margoliash distance method (Felsenstein, 1995). This was done in the Fitch (PHYLIP version 3.5) program using 100 bootstrap iterations and the Dayhoff model of protein sequence evolution. Random sequence addition, five jumble repetitions, and global rearrangement of sequences were performed for each round of analysis on the distance matrices of the bootstrapped data. The majority-rule bootstrap consensus tree was

produced in the program Consense (PHYLIP version 3.5) and viewed in TreeView (Page, 2001).

Analysis of polypeptide sequences, in general, can be more informative for highly divergent genes because achieving reliable DNA alignments is challenging while multiple substitutions can be difficult to account for (Hughes, 1998; Hughes, 1999). Likewise, when analyzing gene-family relationships, patterns of sequence similarity are more likely to be detected by protein level analysis (Mount, 2001).

In the present case, the TEP family is highly divergent, with identity scores less than 30% between the paralogous proteins, and coding for polypeptides over 1500 amino acids in length. This high degree of divergence creates a problem for character-based analyses which are sensitive to large amounts of homoplasy or hidden (unaccountable) substitutions (Felsenstein, 1978; Swofford *et al.*, 1996). Because of the large size of these proteins, it is possible, though, that the variance associated with the existence of homoplasy (if spread out randomly) will be low. As a preliminary test of this possibility, maximum parsimony (MP) analysis was conducted on the TEP family (same data as above) using global alignment (N=45) of proteins, and the PROTPARS program (Felsenstein, 1995). All sites were considered under the assumptions previously described by Felsenstein (1995; 1996), which include the number of steps required for each change in an amino acid replacement. Random addition of sequences with 5 jumble repetitions was performed on the bootstrapped data. The majority-rule bootstrap consensus tree was produced in the program Consense (PHYLIP version 3.5) and viewed in TreeView (Page, 2001).

In addition, maximum parsimony analysis was performed using the PAUP* 4.0b10 program (for comparison). A more rigorous analysis can often be produced (which in turn can produce a more parsimonious tree) using PAUP since it allows for more variation in analysis settings, assumptions, and other options. In general, for the complete global alignment protein data (N=45 taxa; >1700aa), a heuristic search of 100 bootstrap replicates was produced (starting tree produced by stepwise-addition; with 25 random addition sequence replicates at each round) using the tree-bisection-reconnection (TBR) branch swapping algorithm. As before, the trees were viewed and printed using the TreeView program. Uncorrected and corrected ME distance trees, along with MP trees for the same data set, were generated and the resulting topologies compared.

Reliability of internal nodes was determined by the bootstrapping method (Felsenstein, 1985) utilizing 100 (MP), 1000 and 10, 000 (ME) replicates. In phylogenetic analysis of all TEP members, gaps were treated in a pairwise deletion manner (based on pairwise comparisons). In the analysis of orthologous members (i.e., C3), gaps were treated by complete deletion (this data not shown).

The finding of a complement-related gene, SeC3, in a coral presents a difficult issue in rooting phylogenetic trees of the TEP family in general. Because no complement-like gene has ever been found outside of deuterostomes (protostomes only have A2M-like components) it has always been assumed that the root of the tree (and family) consists of an A2M-like ancestral gene. Therefore, all rooted phylogenies in the literature produce phylogenetic trees of the TEP family which are rooted with A2M. There are some very conserved structural differences between A2M and C3/C4/C5, which could be derived conditions in either A2M or the complement proteins. This study

has found that the coral appears to possess two TEP genes, yet only one has been sequenced to completion. Complete sequence information on the second TEP, which could help clarify its structural nature (as C3-like or A2M-like) is not yet available. It remains possible that the second TEP is similar to SeC3, and therefore the A2M-like condition is derived. If this were the case, then a similar duplication event may have occurred early in phylogeny (see Figure 9c) so that one of the copies diverged (and became secondarily modified; derived condition) into the A2M-types. If such were the case, the C3-like version was secondarily lost (either by a chromosomal deletion event or a gene conversion event to become A2M) in the protostome lineage.

In some gene families, the root of the tree is difficult to determine, as in the above case. To discern potential ancestral positions, rooting at the nodes of paralogous gene duplications (Donoghue & Mathews, 1998; Schwartz & Dayhoff, 1978) has been an area of great interest, particularly in rooting the tree-of-life (Brinkmann & Philippe, 1999). This approach theoretically fits the TEP family because (as can be seen in Fig. 9c) the complement components and the A2M-like genes are separated by one major duplication event in early phylogeny. The generation of two paralogous copies, early in phylogeny, allowed for the divergence of the A2M-like genes from the complement-like genes; however, as the nature of the ancestral condition is not yet known, it is difficult to predict which is the derived condition.

One of the required conditions for rooting gene families at a duplication event is that both copies of the paralogous gene forms should exist in both diverging lineages of the phylogenetic tree, so that the gene tree and species tree overlap. But in the case of the TEP family, no complement-like component are apparent in the protostome lineage of

organisms but appears in at least one phylum that predates the protostome-deuterostome split. This complicates the justification for rooting at this early duplication event. Because of this uncertainty, and until more data becomes available on the nature of some of these (and other) gene family members, the most logical and recommended method of tree display is the unrooted format (as in Figure 9a and 9b).

The recent addition of TEP gene and protein sequence data from the mosquito (through the genome sequencing project and work by Christophides *et al.*, 2002), *Drosophila* (described above, fruit-fly databases), and human CD109 (Lin *et al.*, 2002) has produced significant insight into the evolution of this gene family. The incorporation of these data into current phylogenetic analyses, which includes vigorous testing of multiple root hypotheses, is helping to elucidate the root of the TEP gene family. Because of the significant interest associated with these findings, the data will not be discussed further (manuscript in preparation). But there is now sufficient evidence to suggest that, as depicted in Figure 9d, the rooted tree will produce similar relationships between the paralogous complement components and A2M-like genes.

Results

Cloning of initial degenerate PCR product

Using degenerate primers (modified from Nonaka & Takahashi, 1992) and RT-PCR, two PCR products were isolated (Fig. 3a) in the approximate size range expected for the corresponding area of the TEP family. The band of approximately 220bp (Fig. 3a) was gel purified and cloned into a TA-cloning vector (Invitrogen) and sequenced. The

cloned and sequenced product was a 214 bp cDNA. The translated cDNA contained a thiolester site and sequence similarity to the TEP family of proteins in the corresponding region. Identity and similarity comparisons of this deduced amino acid sequence yielded 47|66 % with HuA2M, and 38|58 %, 41|60 %, 25|49 % with HuC3, HuC4, and HuC5 respectively. Comparisons to other TEP sequences produced similar scores. In this report, one of the cloned cDNAs was pursued, but preliminary analysis of other cloned sequences indicates the presence of at least one more TEP in the coral.

Northern and Southern blot analysis

To estimate the size of the coral TEP sequence, total RNA was isolated and Northern blot analysis was performed (Krumlauf, 1996; Sambrook & Russell, 2001) (Fig. 3b). Using either riboprobe or random primed PCR products as probes (using clone SeC32-35 or SeC3-3'R2, respectively), it was confirmed (several times) that the coral gene is expressed at relatively low concentrations, in normal unstimulated tissues, and was approximately 5.8-6kb in size (see Fig. 3b). Initial attempts to cross-hybridize the probes to Northern-blotted mouse and rat total RNA produced negative results (data not shown). This lack of cross hybridization to mammalian RNA rules out the possibility that the coral sequence has resulted from mammalian contamination. In addition, the full-length sequence of the coral TEP (SeC3) confirms that the original Northern-based estimation of size was correct. The assembled sequence appears to be a few hundred bases shorter only because the full length 5'-UTR sequence of SeC3 has yet to be cloned.

To confirm the presence of SeC3 in the genomic DNA of the coral, Southern blotting was performed (Fig. 3c & d) (see Appendix). Southern blotting, by using a combination of probes from the cDNA sequence, can be used to estimate gene copy number and the relative complexity of the gene (this family of genes contains an unusually high number of introns, see Morley, 2000). The banding pattern associated with the initial blots has confirmed the presence of SeC3 in the coral genome, and appears to further suggest a complex genomic organization consisting of multiple intron interruptions (see lane 1 of Fig 3c, cut with HindIII). A similar pattern in Southern blotting from bony fish (using HindIII as well) confirms the presence of multiple C3 genes (Nakao *et al.*, 2000). Because of this, gene copy number is more difficult to estimate unless the patterns from multiple Southern blots (using different probes) are compared.

In the Southern blots shown here, the probe used was a 762bp cDNA from SeC3 which corresponds to a region in human C4 that spans 7 exons and 6 introns and includes the α-γ cleavage site, that is absent vertebrate C3 proteins. The HindIII restriction site may be common within the introns, because the entire cDNA sequence of SeC3 contains only one recognition site, which lies within the probe's sequence. This would have to be confirmed by cloning and sequencing the intervening intronic regions.

Using the RT-PCR approach

Conventional methods of assembling full-length genes involve screening cDNA libraries from the animal. However, cDNA libraries carry the risk of containing

contaminating RNA from an outside source. This becomes a serious issue with marine invertebrates suspected of possessing tissue-resident endosymbionts. Histological analysis with light and electron microscopy has established (C. H. Bigger, C. Olano, and I. Spence, unpublished; W. Goldberg, personal communication) that *Swiftia* does not contain endosymbionts. In this study, to avoid the risk that contaminating RNA might be present in our libraries, the sequence of this gene was completed utilizing RT-PCR (as RACE). Each Northern blot was produced with total RNA extracted at independent or separate times. Each RT-PCR RACE reaction was performed with a new set of nested-paired primers and freshly isolated (from random colonies) RNA. Library screening, if necessary, is still a valuable asset as long as the products are later confirmed using RT-PCR on freshly isolated RNA.

Various sets of the SeC3 primers were tested in PCR reactions utilizing the following sources as a template: the sea water in which the coral is maintained and brine shrimp nauplii (*Artemia sp.*) RNA (cDNA) and genomic DNA. The *Artemia* was a potential source of contaminating RNA because it is the main diet for *Swiftia* in our aquaria. All PCR reactions failed to produce any amplified product with all primer pairs tested. PCR amplification using total RNA from *Artemia* and the initial degenerate PCR primers did, however, isolate partial cDNAs for two TEPs with significant homology to *Drosophila* TEP1 and TEP2 (data not shown, unpublished data). Neither of the *Artemia* sequences, nonetheless, contains significant similarity to SeC3. They are distinctly different and cannot be considered a cross-contaminating source of sequence. All RACE clones produced for SeC3 were overlapping with the preexisting sequence and because of

the simplicity of PCR-based methods, all reactions were confirmed while maintaining proper controls at each sequential step.

Primary structure of SeC3 sequence

Classic RACE (Zhang & Frohman, 1997) was utilized to generate cDNAs corresponding to the full length sequence of SeC3. A total of 10 overlapping RACE products were produced resulting in a 5488bp cDNA sequence with a deduced amino acid sequence of 1728aa in one open reading frame (Fig. 4).

A very important distinction between A2M proteins and the complement components C3, C4, and C5 is size. All A2M and A2M-like divergent paralogs are similar in size and 200-250 aa shorter than the complement components. All deuterostome invertebrate C3-like proteins that have been characterized are similar in size to vertebrate C3, yet all TEPs characterized in protostome invertebrates are similar to A2M. Interestingly significant, this report describes a coral TEP (SeC3) that is similar in size, and shares overall physical and chemical characteristics with vertebrate C3.

The conserved thiolester site for SeC3 can be found at positions 1024-1028 and the C3-specific reactive histidine (VI**H**QEM) at position 1140 (see Fig. 4 and 5). SeC3 contains two putative cleavage locations which would process the pro-molecule into a three chain structure (see Fig. 5 and Fig. 2b). The predicted (un-glycosylated) sizes of the individual chains of human C4A (a 3-chain TEP) are 74kDa, 86kDa, and 32kDa for beta, alpha, and gamma chains, respectively (Morley & Walport, 2000). In SeC3 five putative N-glycosylation sites are predicted at positions 162, 216, 700, 804, and 1256,

four in the beta chain and one in the alpha chain. SeC3 contains 35 cysteines, 4 in the beta chain, 15 in the alpha chain, and 16 in the putative gamma chain (most of which are conserved in C3 and C4). Although many of the cysteines align to corresponding conserved cysteines of vertebrate C3, some from SeC3 do not. This may imply a slightly different folding pattern with the formation of unique functional or binding sites.

C3-convertase cleavage of vertebrate C3 results in the active form, C3b, while releasing the anaphylatoxin peptide, C3a (see Fig. 2b, 5, and 6a). This produces an immediate conformation change in C3b and brings the catalytic histidine in direct contact with the thiolester site. The activated thiolester-containing protein, C3b, reacts in an immediate covalent fashion with the target (Gadjeva *et al.*, 1998).

The C3a peptide spans 65-70 amino acids and contains 6 cysteine residues (Fig. 5 and 6a) which are organized in a conserved fashion and presumably give it a characteristic fold and its anaphylotoxin activity. This organization of the cysteines is well conserved in the coral, and includes paired cysteines (-CC-) found at both ends of the C3a region (Fig. 5). The signature cleavage motif for vertebrate C3a is –LAR/S and is a conserved sequence because it is also a receptor-binding site for the peptide (Sahu & Lambris, 2001). A putative cleavage site, –RTR/S can be found in the corresponding region of SeC3.

There is sequence conservation immediately (C-terminal) following the C3a region (see Fig. 6a). This region (region 749-790 in HuC3), corresponds to the reactive area for vertebrate C3 interaction with Factor H, B, and CR1,2, and 3 (see below, section on Hydrophobicity Profiling). The first Factor I cleavage site (at position 955 in HuC3),

which generates the C3dg fragment, is missing in the coral (Fig. 6a), and is also missing in amphioxus and urchin (see Ch. 4 Appendix alignment).

Hydrophobicity profiling

In an attempt to predict the conservation of at least one major receptor/ligand binding region on SeC3, hydrophobicity (hydropathy) profiling was performed. Hydropathy profiles were produced and compared on corresponding regions of a major receptor-binding region (Factor B & H, CR1-3) for vertebrate C3. This binding region can be found C-terminal to the C3a anaphylatoxin region (position 748-790, HuC3), position ~731-782 in SeC3, which in mammalian C3 corresponds to a binding region for complement receptor (CR) 1, 2, and 3 and for Factors B and H.

TEPs (including SeC3) were aligned against HuC3, and the region corresponding to (or aligning with) the above described binding site was used for hydropathy analysis. Hydropathy profiles of SeC3 vs. the corresponding region in HuC3, HuC4, HuC5 and HuA2M, were generated (Fig. 7a). The fB+fH+CR1-3 region only exists in vertebrate C3 proteins. Results from the hydropathy profiles suggest that the chemical nature of the corresponding region in SeC3 is similar to that of the HuC3 fB+fH+CR1-3 region, and not to the other paralogous proteins compared (Fig. 7a).

The properdin binding region is also a unique regulatory site for C3 proteins. This binding site was also analyzed with hydropathy profiling. The corresponding region of the SeC3 protein, in this case, does not appear to be similar to HuC3 properdin-binding site (Fig. 7b). Like C4, the chemical nature of this region does not appear to favor

interaction with a properdin-like molecule. This may suggest that the properdin-binding site evolved after modern/contemporary C3 diverged from the C3/C4/C5 ancestor.

Structural organization of deduced polypeptide

SeC3 contains two cleavage sites, corresponding to those that generate the alpha, beta, and gamma chain in mammalian C4 and lamprey C3 (e.g. see Fig. 2a &b). This suggests that SeC3 contains a three chain structural fold similar to C4. The beta-alpha cleavage site (RKRR) is conserved at position 665. This generates the alpha and beta chain in processed C3 proteins. A puzzling issue, though, is that the two cysteines which link the beta and alpha chain in deuterostome C3s are missing in the coral. In SeC3 an alternative binding site may exist involving two non-conserved pairs of cysteines, each separated by two amino acids, and found on both chains (Fig. 6a and Ch4 Appendix). One is present at the N-terminal end of the beta chain (position 15-19) and the other near the C-terminal end of the alpha chain (position 1221-1225). This interaction would maintain the orientation of the alpha and beta chains similar to that seen in mammalian C3. The first set of these cysteines lies immediately following the leader peptide and, hence, casts some doubt as to whether this predicted interaction is valid.

The second cleavage site, mentioned above, can be found in the same conserved location at which the C4 and the lamprey C3 alpha-gamma site exists (see Fig 5 and 6a) . Interestingly, the coral contains two putative cleavage sites in this region, at position 1385 (RARR) and at position 1439 (RRCR). If cleavage were to occur at both locations, a 74aa product would be freed that is particularly interesting. This 74 aa stretch within

the putative cleavage region between the alpha-gamma chains is arginine (n=14) and lysine (n=13) rich and contains five prolines and three cysteine residues (Fig. 5 and 6a). The lysine-arginine rich region may represent a relic of the events that brought R residues into close proximity while assuring cleavage by keeping the region highly hydrophilic and exposed. Hence, the intriguing nature of this alpha-gamma cleavage region may imply something about its function and/or how it was established. The cysteines associated with gamma chain binding to the alpha chain are completely conserved in SeC3.

Deduced secondary structure and 3D comparative modeling

As discussed above, SeC3 contains most of the conserved cysteines associated with C3/C4/C5 proteins, yet a few are unique to the coral and may confer some level of distinction to SeC3. Full-length alignment shading for conserved physiochemical properties, though, suggests that SeC3 shares significant structural properties with not just the TEP family in general, but with vertebrate C3/C4/C5 in particular (Fig. 6b and 6c; Ch4 Appendix). This was confirmed by careful inspection and comparison of full length pro-molecule hydropathy profiles (data not shown). Consequently, the SeC3 protein sequence was submitted along with other TEP family members to the *Protein Structure Prediction Server* (McGuffin *et al.*, 2000; Rost, 1996). Results from multiple members of the TEP family suggests that overall secondary structure is highly conserved. The most important exception lies in the C-terminal region of C3, a region that is absent from A2M. The presence of this region is a major defining characteristic of complement

proteins C3, C4, and C5. This region is highly conserved between SeC3 and mammalian C3. The abundance of helical structures in the C3d region of C3 is also highly conserved in the corresponding region of all TEP homologues (Fig 8, C & D).

High conservation of secondary structure allows for the prediction of three-dimensional structure using the comparative modeling approach (Leach, 2001). Submission of SeC3 polypeptide sequence to the *SwissModel Server* (Guex & Peitsch, 1997; Peitsch *et al.*, 2000) resulted in a predicted conserved structure conforming to that of the X-ray crystal structure of human C3d (Nagar *et al.*, 1998) in the Protein Data Base (PDB)(see Fig. 8). All residues determined to be exposed in a contiguous patch on the surface of the C3d molecule are conserved in SeC3, along with the conservation of residues determined to be buried in the models (see (Nagar *et al.*, 1998)). This conservation of structure strongly suggests conserved functional properties at least at the level of the thiolester binding region (see Fig. 8, A & B).

Phylogenetic analysis of SeC3

Full-length amino acid alignments were produced using the *Clustal X* program. SeC3 was aligned against members of the TEP family (N = 45) using global or multiple alignment parameters. Global alignments produced reliable results, as long as the resultant alignment was scanned carefully for mismatched regions (Ch. 4 Appendix). This is feasible because a considerable amount of structural and functional information is available for the TEP family, and those regions are expected to align well because they are typically well conserved (Sahu & Lambris, 2001). As a consequence, results (which

were essentially identical) attained with profile alignment of the paralogous groups will not be discussed further. In generating all calculated sequence comparisons (distance computation) (Table 2), paired alignments of the sequences in question were produced.

The TEPs are a conserved family of large paralogous proteins. This is supported by the observation that most methods of phylogenetic reconstruction produce very similar topologies (see Fig. 9 - 12). The vertebrate complement components are well resolved and produce almost identical topologies with both distance and character-based methods. In all methods used, the coral TEP sequence, SeC3, clusters with the deuterostome invertebrate C3-like proteins, which form a sister taxa to the vertebrate complement components.

Some simulation studies have suggested that as the distance between sequences increases, the variance associated with corrected-distance analysis also increases so that only uncorrected p-distances are reliable (Nei, 1991). Both uncorrected (Fig. 9a) and corrected (Fig. 9b) bootstrapped distance trees were produced, under the minimum evolution criteria, and the resultant topologies compared. The previously described (Nei, 1991) bias associated with corrected distances (in analyzing divergent sequences) was not apparent in these results. Both corrected and uncorrected analysis produced identical topologies. This suggests that divergent sequences, when large enough, are affected less by the variances associated with correction-formulas (Swofford *et al.*, 1996).

Evaluating results from distance and character-based analyses

Rzhetsky and Nei (1992) have argued that minimum evolution (ME) methods outperform other methods of distance analysis, such as the Fitch-Margoliash (FM) least-

squares method (Felsenstein, 1995; Fitch & Margoliash, 1967), when analyzing divergent sequences. The justification for these results, though, is limited to short computer generated sequences. In addition, some authors have argued that these biases can further be reduced if negative branch lengths are not allowed (Felsenstein, 1995; Swofford *et al.*, 1996). The TEP family warrants analysis using both methods because it is a considerably divergent family of very large polypeptide sequences.

Phylogenetic analysis of the TEP family using the ME method was performed using both uncorrected and corrected distances (Figs. 9a-d). The results for the ME method are displayed in three ways: unrooted (preferred), rooting at hypothetical duplication event (appearing as mid-point rooting), and rooted with an outgroup. The unrooted option is preferred because the ancestral condition of this family is as yet undetermined. When studying multigene families, rooting at nodes representing an ancient duplication event is not uncommon (Page & Holmes, 1998). Consideration for this approach stems from the observation that in phylogenetic analysis of this data, the midpoint of the two longest branches usually roots the tree at the duplication event that gave rise to A2M and the complement component ancestor (C3/C4/C5), see Fig. 9c. But as discussed in the methods section, this approach is not justifiable because a complement-like gene does not exist (or is not apparent) in the protostome lineage. This would prevent overlap between the species and gene trees under consideration. Alternatively, the root of the family may be identified if sufficient sequence data becomes available, which may help determine the original ancestral similarities between A2M-like genes and complement genes. As mentioned before (in the methods section) we now have new data and work in progress to suggest that the root of the tree may soon be

revealed. In short, the rooted phylogeny may retain some of the branching patterns seen when the tree is rooted with some of the arthropod TEPs (Fig. 9d). Concurrent phylogenetic analysis using the FM method was produced and displayed using equal terminal branch lengths (for topological comparisons only) (Fig. 10).

Minimum evolution places insect TEPs (divergent A2Ms) as a sister group to vertebrate A2M. Interestingly, *Limulus* A2M always clusters with vertebrate A2M, even though the horseshoe crab is a protostome (arthropod). Although there appears to be no *Drosophila* or *C.elegans* sequences that would cluster with A2M the way that the *Limulus* protein does, multiple proteins have been characterized in protostomes that, based on functional and partial sequence data, appear to be *bona-fide* A2M. These data suggest the presence of A2M in multiple species of gastropods, bivalves, cephalopods, and crustaceans (see recent review by Armstrong & Quigley, 1999). Unfortunately, sequence data was not available for inclusion in this report's analyses. Although functional data from the *Drosophila* and worm TEPs is lacking, one (or more) of the fruit fly TEPs may be, for example, a divergent functional form of A2M. These data, however, suggest that the paralogous copies of insect and worm TEPs resulted from duplication events within the protostome lineage and that the A2M gene (and its functional constraints) existed prior to the protostome-deuterostome split.

Additional analysis of the ME trees reveals relationships supporting previous work (Nonaka *et al.*, 1999; Smith *et al.*, 1999) suggesting that the deuterostome invertebrate C3-like proteins are a sister group to the complement components 3, 4, and 5. Contrary to common expectations, there is now data in a Cnidarian suggesting that a similar gene (SeC3 in this report) existed much earlier in phylogeny so that a gene

encoding a protein with C3-like characteristics is not unique to the deuterostome lingeage. The *Halocynthia* C3-like sequence is interesting because the longer branch length suggests that it is more divergent, and functional studies are revealing its diverse nature (Nonaka *et al.*, 1999). Alternatively, this tunicate C3-like gene may not be orthologous to the other invertebrate C3-like genes.

The resultant unrooted FM distance tree is topologically similar to the ME trees (see Fig. 10). The major difference is that the mosquito TEP sequence lies as an outgroup sister taxa (68% bootstrap support) to the invertebrate C3-like genes. The congruence of the same data was further demonstrated by the production of a similar MP bootstrap tree (Fig. 11). In this case, though, DrosMCR and *C.elegans* TEP form outgroup sister taxa to the complement components rather than to the other insect TEPs. Under more vigorous analysis options, though, MP can produce a tree with almost identical topology as the ME tree (Fig. 12). The overall outcome suggests that the FM and ME distance methods, along with the MP method, can provide similar results in phylogenetic analysis of divergent protein sequences of sufficient length. It is difficult to determine the reliability of the clustering pattern of the mosquito TEP sequence (FM tree) or the DrosMCR and *C. elegans* TEP (MP tree, in Fig.11), but because they are firmly placed within the insect TEP group in the ME trees, their placement in the FM (fig. 10) and MP (fig. 11) trees may be incorrect.

Discussion

This report describes the molecular cloning of the first TEP outside of the protostome and deuterostome lineages. It is also the first complement-like component outside of the deuterostome. This coral protein, SeC3, is more similar to C3 than other members of the TEP family. This work supports the previous claim that the ancestor to the C3/C4/C5 paralogous proteins had C3-like characteristics before the duplication and divergence events took place. RT-PCR was utilized to clone the entire cDNA sequence of SeC3 from this coral (as described above). Of special interest, in relation to the evolution of this family, is the preliminary observation that the coral appears to have at least one more TEP protein, as yet to be characterized (LJD, unpublished data).

The coral sequence is more similar to the complement components based on the following observations: 1) the coral sequence shows higher overall identity to C3/C4/C5 than to A2M (table 2), 2) phylogenetic analyses supports this observation (Fig. 9-12), 3) the coral sequence contains conserved physiochemical and structural properties unique to the C3/C4/C5 lineage (Fig. 2a, 6b & 6c), 4) the coral deduced amino acid sequence contains regions characteristic for C3, and to some extent, C4 sequences (as can be seen by careful analysis of the full-length alignments), 5) the coral sequence shares similarity in the extended C-terminal region of the complement components, a characteristic not found in A2M proteins (Fig. 2a), and 6) the coral sequence appears to share the properties associated with a major receptor/ligand binding site in C3 (Fig. 7a) proteins.

Pairwise identity and similarity calculations (see table 2) would suggest that the relationship of the coral sequence to C3, C4, and C5 is unresolved. But this observation

may be due to shared pleisiomorphic characteristics between vertebrate C3, C4, and C5 and the coral sequence (SeC3). The SeC3 cDNA and deduced amino acid sequence is similar in size to C3, C4, and C5 and larger than any A2M sequence characterized. SeC3 contains a C-terminal region characteristic of C3/C4/C5 (see Fig. 2a, 6a-c; Ch. 4 Appendix).

All C3s, including the urchin and tunicate C3-like proteins, are two chain proteins after post-translational modification. The only exception to date is lamprey C3 (Nonaka, 1994), which contains the second cleavage site (including the extended 40-55aa region) characteristic of the alpha-gamma cleavage site in C4 proteins. Post-translational modification produces a three chain structure similar to C4 (see Fig. 2b). This feature is apparently conserved in cyclostomes because hagfish C3 has the conserved extended region where the cleavage site can be found (between the beta and alpha chain), but it appears to have been inactivated by a substitution event (RRRR→ RRRQ). Recently, *Amphioxus* C3-like sequence has become available in the GenBank database. Although not discussed by the authors (Suzuki *et al.*, 2002), *Amphioxus* C3 also has the characteristic extended region with a putative cleavage site (-RAIR-). Potentially, this would also produce a three chain molecule. Collectively, these data suggest that the three chain structure cannot be considered unique to C4. These observations provide support to the claim that a three chain configuration is an ancestral trait (Nonaka, 1994; Nonaka & Takahashi, 1992). The claim is further supported by the finding of the putative alpha-gamma cleavage site in SeC3 (Fig. 5 & 6a).

As discussed before, the cysteines associated with binding of the beta chain to the alpha chain are absent in the coral sequence (Fig. 6a) making SeC3 the only characterized

C3-like sequence presenting this condition. Interestingly, the required C at position 809 of the alpha chain is replaced by a G, which is characteristic of A2M, that contains a G at this position. However, A2M lacks the beta-alpha cleavage site and the subsequent two-chain structure. It should be noted, though, that this G can be converted to a C by a one base substitution (in the first codon position). On the beta chain, SeC3 contains a L at position (556) in place of the required C. A2M sequences have a CL in this position. The coral sequence is EL, a characteristic shared by the *Drosophila* TEP4 sequence. Vertebrate A2M already contains a C at the putative beta chain location, so that a simple substitution of the G to a C in the alpha chain provides binding sites for the two chains in those rare A2Ms that have retained (or gained) the beta-alpha cleavage site (e.g., Carp and Lamprey A2M).

The β-α chain interaction presents a novel and interesting scenario. The two cysteines involved in linking the two chains together are the only cysteines not conserved in SeC3 (see Fig. 6a). This would imply that the two chains associate in a different fashion or that the β-chain is released and is not a part of the processed protein (see Fig. 13). The latter case seems unlikely for the following reason. The coral β-chain is highly conserved with the corresponding region of C3/C4/C5, whose β-chain co-evolved with the structural constraints associated with its function (bound to the α-chain) . If SeC3's β-chain is involved in a separate function, one would expect that this region would have been modified to some extent by selection. An alternative condition would resemble gene sharing (Wistow & Piatigorsky, 1987) in eye crystallins, where divergent functions of the same protein in two parts of the body have not altered its primary structure.

Hypothetically, the lack of the β-chain would leave the N-terminal region of the α-chain exposed for immediate protease cleavage (releasing the C3a peptide) and activation of SeC3. This can, in turn, lead to immediate opsonic binding to nearby products.

The data presented here supports previous predictions that the ancestor to C3/C4/C5 was C3-like (Nonaka *et al.*, 1999). These results specifically suggest that the ancestor to vertebrate C3/C4/C5 existed prior to the divergence of protostomes and deuterostomes. Preliminary data in our lab shows that at least one more TEP exists in the coral and suggests that some of the duplication events giving rise to the paralogous TEP family also predates the Cambrian period. A hypothetical model is presented (Fig. 13), where the ancestral TEP sequence was C3-like and that a gene duplication event preceded the protostome-deuterostome split, producing a copy which could then diverge into the paralogous A2M genes. For reasons yet unclear, the protostome lineage appears to have lost the C3-like TEP immediately following its divergence from deuterostomes. This work strongly suggests the need for a re-evaluation of our knowledge of the origins and evolution of the TEP family of proteins.

References

Armstrong P. B., and Quigley J. P. (1999). Alpha2-macroglobulin: an evolutionarily conserved arm of the innate immune system. *Developmental and Comparative Immunology* **23:** 375-390.

Ayala F. J., Rzhetsky A., and Ayala F. J. (1998). Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. *Proceedings of the National Academy of Sciences* **95:** 606-611.

Barrington R., Zhang M., Fischer M., and Carroll M. C. (2001). The role of complement in inflammation and adaptive immunity. *Immunological Reviews* **180:** 5-15.

Brinkmann H., and Philippe H. (1999). Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Molecular Biology and Evolution* **16:** 817-825.

Campbell R. D., Law S. K. A., Reid K. B. M., and Sim R. B. (1988). Structure, organization, and regulation of the complement genes. *Annual Review of Immunology* **6:** 161-195.

Carroll M. C. (1998). The role of complement and complement receptors in induction and regulation of immunity. *Annual Review of Immunology* **16:** 545-568.

Christophides G. K., Zdobnov E., Barillas-Mury C., Birney E., Blandin S., Blass C., Brey P. T., and al. e. (2002). Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298:** 159-165.

Dempsey P. W., Allison M. E. D., Akkaraju S., Goodnow C. C., and Fearon D. T. (1996). C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science* **271:** 348-350.

Dodds A. W., and Law S. K. A. (1998). The phylogeny and evolution of the thioester bond-containing proteins C3, C4, and alpha2-macroglobulin. *Immunological Reviews* **166:** 15-26.

Donoghue M. J., and Mathews S. (1998). Duplicate genes and the root of angiosperms, with an example using phytochrome sequences. *Molecular Phylogenetics and Evolution* **9:** 489-500.

Doolittle R. F., Feng D.-F., Tsang S., Cho G., and Little E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271:** 470-477.

Felsenstein J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27:** 401-410.

Felsenstein J. (1985). Confidence limits on phylogenetics: an approach using the bootstrap. *Evolution* **39:** 783-791.

Felsenstein J. (1995). PHYLIP: Phylogeny Inference Package, version 3.57c.

Felsenstein J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *In* "Methods in Enzymology: Computer Methods for Macromolecular Sequence Analysis" (R. F. Doolittle, Ed.), pp. 418-427, Academic Press, San Diego.

Fitch W. M., and Margoliash E. (1967). Construction of phylogenetic trees. *Science* **155:** 279-284.

Fortey R. A., Briggs D. E., and Wills M. A. (1997). The Cambrian evolutionary "explosion" recalibrated. *BioEssays* **19:** 429-434.

Gadjeva M., Dodds A. W., Taniguchi-Sidle A., Willis A. C., Isenman D. E., and Law S. K. A. (1998). The covalent binding reaction of complement component C3. *Journal of Immunology* **161:** 985-990.

Gonnet G. H., Cohen M. A., and Benner S. A. (1992). Exhaustive matching of the entire protein sequence database. *Science* **256:** 1443-1445.

Gu X., and Zhang J. (1997). A simple method for estimating the parameter of substitution rate variation among sites. *Molecular Biology and Evolution* **14:** 1106-1113.

Guex N., and Peitsch M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modelling. *Electrophoresis* **18:** 2714-2723.

Hall T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95-98.

Hillis D. M., Mable B. K., and Moritz C. (1996). Applications of Molecular Systematics: The State of the Field and a Look to the Future. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), pp. 655, Sinauer Associates, Inc., Sunderland, Ma.

Hoop T. P., and Woods K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences* **78:** 3824-3829.

Hughes A. L. (1998). Protein phylogenies provide evidence of a radical discontinuity between arthropod and vertebrate immune systems. *Immunogenetics* **47:** 283-296.

Hughes A. L. (1999). "Adaptive Evolution of Genes and Genomes," Oxford University Press, New York.

Janeway C. A., and Medzhitov R. (2002). Innate immune recognition. *Annual Review of Immunology* **20:** 197-216.

Jensen J. A., Festa E., Smith D. S., and Cayer M. (1981). The complement system of the nurse shark: hemolytic and comparative characteristics. *Science* **214:** 566-569.

Krumlauf R. (1996). Northern Blot Analysis. *In* "Basic DNA and RNA Protocols" (A. J. Harwood, Ed.), Humana Press, Totowa, NJ.

Kumar S. (1996). A stepwise algorithm for finding minimum evolution trees. *Molecular Biology and Evolution* **13:** 584-593.

Kumar S., Tamura K., Jakobsen I. B., and Nei M. (2001). Mega2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics (submitted).*

Kyte J., and Doolittle R. F. (1982). A Simple Method for Displaying the Hydrophobic Character of a Protein. *Journal of Molecular Biology* **157:** 105-142.

Leach A. R. (2001). "Molecular Modelling: principles and applications," Prentice Hall, London.

Levashina E. A., Moita L. F., Blandin S., Vriend G., Lagueux M., and Kafatos F. C. (2001). Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae. Cell* **104:** 709-718.

Li W.-H. (1997). "Molecular Evolution," Sinauer Associates, Inc., Sunderland, Mass.

Lin M., Sutherland D. R., Horsfall W., Totty N., Yeo E., Nayar R., Wu X.-F., and Schuh A. C. (2002). Cell surface antigen CD109 is a novel member of the alpha2-macroglobulin/C3, C4, C5 family of thioester-containing proteins. *Blood* **99:** 1683-1691.

Margulis L., and Schwartz K. V. (1998). "Five Kingdoms: An illustrated guide to the phyla of life on earth," W.H. Freeman and Company, New York.

McGuffin L. J., Bryson K., and Jones D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16:** 404-405.

McMenamin M. A. S., and McMenamin D. L. S. (1990). "The emergence of animals: the Cambrian breakthrough," Columbia University Press, New York.

Morley B. J., and Walport M. J., Eds. (2000). "The Complement *FactsBook*," Academic Press, New York.

Mount D. W. (2001). "Bioinformatics: Sequence and Genome Analysis," Cold Spring Harbor Press, Cold Spring Harbor.

Nagar B., Jones R. G., Diefenbach R. J., Isenman D. E., and Rini J. M. (1998). X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* **280:** 1277-1281.

Nakao M., Musuro 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.

Nei M. (1991). Relative efficiencies of different tree making methods for molecular data. *In* "Recent advances in phylogenetic studies of DNA sequences" (M. M. Miyamoto, and J. L. Cracraft, Eds.), pp. 133-147, Oxford University Press, Oxford.

Nei M., and Kumar S. (2000). "Molecular Evolution and Phylogenetics," Oxford University Press, Oxford.

Nicholas K. B., and Nicholas Jr. H. B. (1997). GeneDoc: a tool for annotating and editing multiple sequence alignments. *Distributed by author.*

Nonaka M. (1994). Molecular analysis of the lamprey complement system. *Fish & Shellfish Immunology* **4:** 437-446.

Nonaka M., Azumi K., Ji X., Namikawa-Yamada C., Sasaki M., Saiga H., Dodds A. W., Sekine H., Homma M. K., Matsushita M., Endo Y., and Fujita T. (1999). Opsonic complement component C3 in the solitary ascidian, *Halocynthia roretzi. Journal of Immunology* **162:** 387-391.

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.

Page R. D. M. (2001). TreeView ver 1.6.5, Distributed by author.

Page R. D. M., and Holmes E. C. (1998). "Molecular Evolution: A Phylogenetic Approach," Blackwell Science, Oxford.

Peitsch M. C., Schwede T., and Guex N. (2000). Automated protein modelling- the proteome in 3D. *Pharmacogenomics* **1:** 257-266.

Quigley J. P., and Armstrong P. B. (1994). Invertebrate alpha2-Macroglobulin: Structure-Function and the Ancient Thiol Ester Bond. *In* "Primodial Immunity: Foundations for the Vertebrate Immune System" (G. Beck, G. S. Habicht, E. L. Cooper, and J. J. Marchalonis, Eds.), pp. 131-145, New York Academy of Science, New York.

Raff R. A. (1996). "The Shape of Life: Genes, Development, and the Evolution of Animal Form," The University of Chicago Press, Chicago.

Rost B. (1996). PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods in Enzymology* **266:** 525-539.

Rzhetsky A., and Nei M. (1992). Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *Journal of Molecular Evolution* **35:** 367-375.

Rzhetsky A., and Nei M. (1993). Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution* **10:** 1073-1095.

Sahu A., and Lambris J. D. (2001). Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunological Reviews* **180:** 35-48.

Salzet M. (2001). Vertebrate innate immunity resembles a mosaic of invertebrate immune responses. *Trends in Immunology* **22:** 285-288.

Sambrook J., and Russell D. W. (2001). "Molecular Cloning: a laboratory manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schwartz R. M., and Dayhoff M. O. (1978). Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* **199:** 395-403.

Smith L. C., Azumi K., and Nonaka M. (1999). Complement systems in invertebrates: The ancient alternative and lectin pathways. *Immunopharmacology* **42:** 107-120.

Stothard P. (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28:** 1102-1104.

Suzuki M. M., Satoh N., and Nonaka M. (2002). C6-like and C3-like molecules from the cephalochordate, amphioxus, suggests a cytolytic complement system in invertebrates. *Journal of Molecular Evolution* **54:** 671-679.

Swofford D. L., Olsen G. J., Waddell P. J., and Hillis D. M. (1996). Phylogenetic Inference. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), pp. 407-514, Sinauer Associates, Inc., Sunderland, MA, USA.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., and Higgins D. G. (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24:** 4876-4882.

Valentine J. W., Awramik S. M., Signor P. W., and Sadler P. M. (1991). The biological explosion at the Precambrian-Cambrian boundary. *Evolutionary Biology* **25:** 279-356.

Wistow G., and Piatigorsky J. (1987). Recruitment of enzymes as lens structural proteins. *Science* **236:** 1554-1556.

Zarkadis I. K., Mastellos D., and Lambris J. D. (2001). Phylogenetic aspects of the complement system. *Developmental and Comparative Immunology* **25:** 745-762.

Zhang Y., and Frohman M. A. (1997). Using Rapid Amplification of cDNA Ends (RACE) to Obtain Full-Length cDNAs. *In* "cDNA Library Protocols" (I. G. Cowell, and C. A. Austin, Eds.), pp. 61-88, Humana Press Inc, Totowa, NJ.

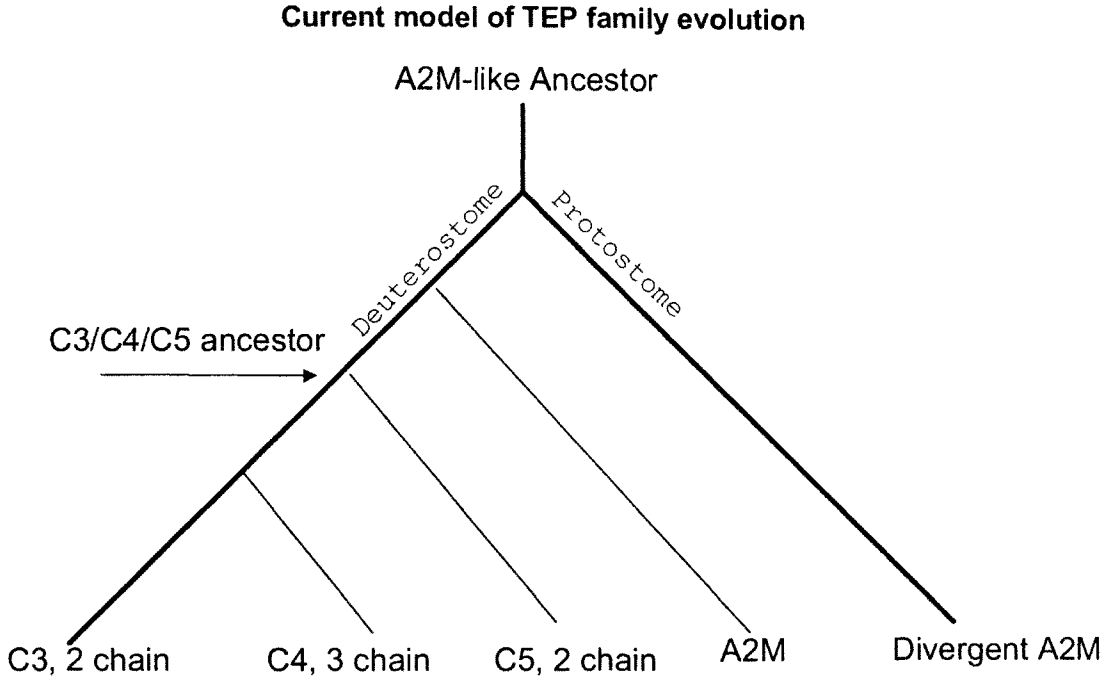


Figure 1. Currently accepted model of TEP family evolution. Notice that the most parsimonious prediction is that the second cleavage site was gained once, found in the three chain C4.

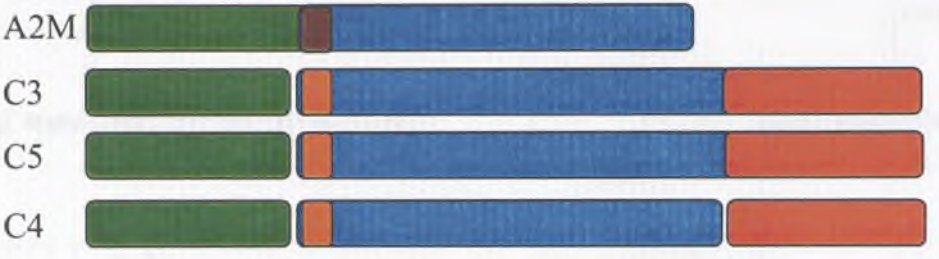


Figure 2a. Schematic representation of primary structural relationships between human TEP proteins. Color indicates homologous regions. Space between bars indicates post-translational cleavage site. A2M is a polymer of single chain polypeptides; C3 and C5 are two chain proteins; and C4 is a three chain protein. represents beta chain in C3, C4, and C5; represents alpha chain; represents the C-terminal region unique to C3, C4, and C5 and represents the gamma chain in human C4; represents the polymorphic A2M bait region; represents the homologous C3a, C4a, and C5a anaphylatoxin region; represents the homologous thiolester site, lost in C5 due to a substitution event early in its evolution.

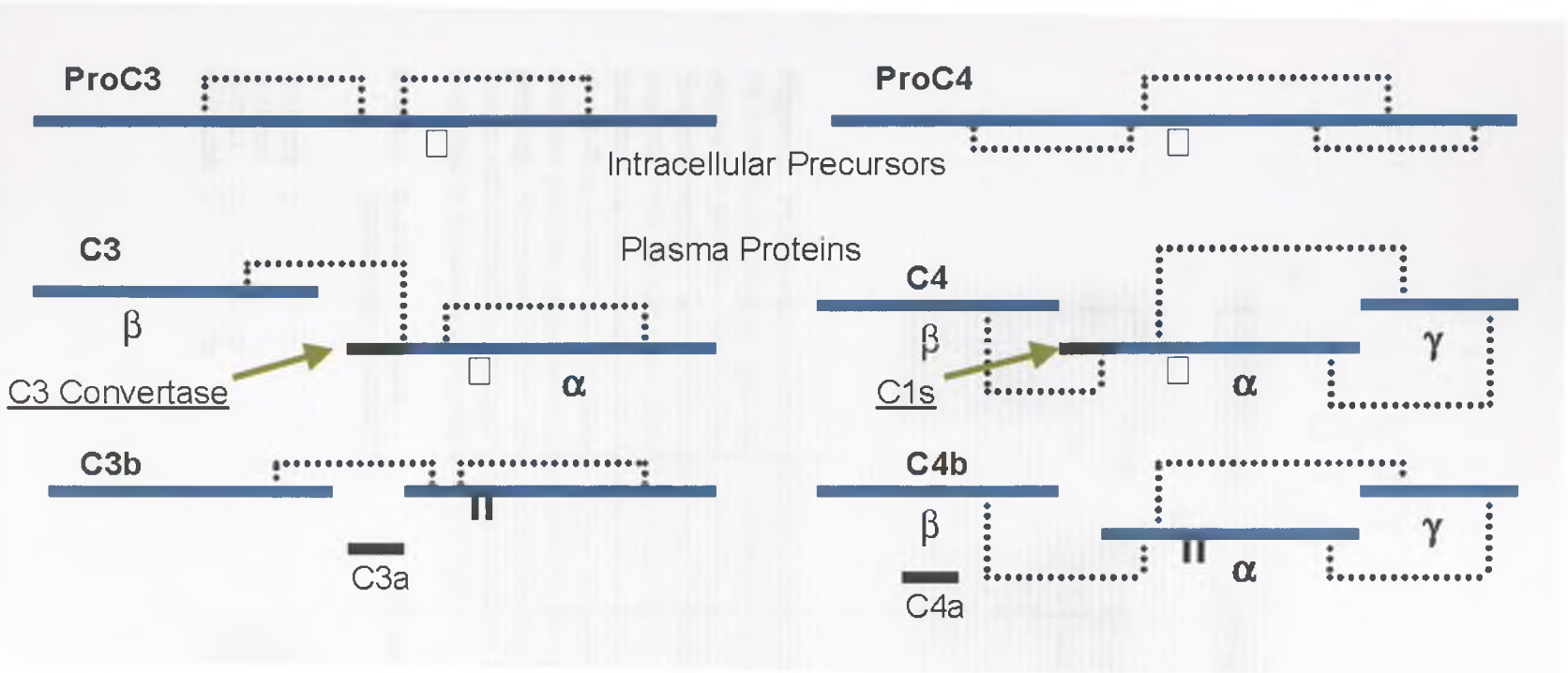


Figure2. Structure of C3 and C4 as deduced from biochemical analysis of human C3 and C4. Note that human C4 is a three chain protein and C3 is a two chain protein.

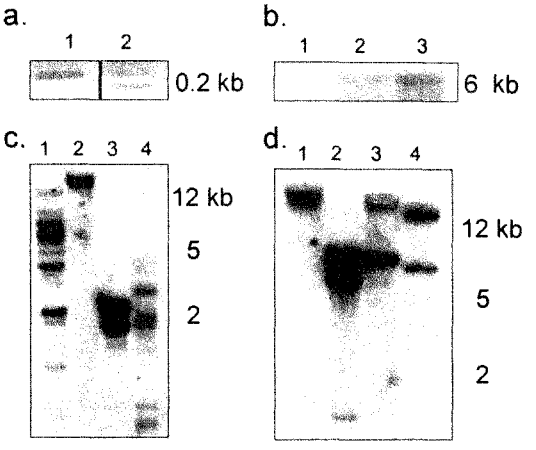


Figure 3. PCR product cloning, Northern and Southern blotting data. a. Degenerate PCR results for *Swiftia*, right lane. Left lane is rat positive control using the same degenerate primers. The top band, right lane, was excised, gel purified and cloned. Sequence analysis indicated it to be a true A2M-like protein. b. Northern blotting suggested that the gene was about 6kb transcribed. Northern blotting also suggested that the gene was constitutively expressed at low levels, since lane three consists of about 40ug of total RNA (lane two, 20ug). Southern blotting (c & d) suggests that a complex genomic organization (many intron interruptions) seen in the vertebrate paralogs. For example, lane 1 of c, the probed region must contain one or more introns rich in Hind III cleavage sites (see below). This genomic organization is expected in the ancestral sequence of the TEPs.

Enzyme	#cut in probe region	#cut in entire cDNA
1) HindIII	1	1
2) Uncut		
3) Dra I	1	2
4) Sau3AI	7	28

Enzyme	#cut in probe region	#cut in entire cDNA
1) Uncut		
2) PvuI	1	2
3) Kpn	0	0
4) Sal I	1	1

Figure 4. Full-length sequence and translation for SeC3. Primers used in cloning RACE PCR products follow this sequence.

```
(-101)gtgctgaagccaaacaacttccgcactctgtgaagtcaatataatccatattattcacaatatctcttacatacagttgttggc
      ggacctacttgcaagac
  1 M  K  M  L  R  A  L  I  G  F  A  L  L  L  C  L  N  Q  C  Y
  1 ATGAAGATGCTTCGAGCCTTGATTGGCTTTGCTTTGCTCTTATGTTTGAACCAATGTTAT
 21 A  A  K  Y  F  I  A  A  P  N  L  L  R  V  G  V  E  E  T  V
 61 GCGGCGAAGTATTTCATCGCTGCTCCAAATCTGCTCCGTGTTGGTGTTGAAGAAACAGTA
 41 S  I  A  V  F  D  V  N  V  D  V  N  V  Q  L  A  L  Q  D  F
121 TCGATTGCCGTGTTTGATGTAAACGTTGATGTCAACGTCCAATTGGCTCTTCAAGATTTT
 61 P  N  R  R  K  T  F  S  Q  V  S  G  N  V  R  A  Q  Q  P  G
181 CCAAACAGGAGGAAGACGTTTTCTCAAGTTTCTGGAAACGTTAGAGCACAACAACCTGGT
 81 I  L  K  I  K  V  N  A  K  D  L  H  D  Q  Q  S  L  D  K  Q
241 ATTTTGAAAATAAAGGtGAATGCCAAAGATCTTCACGACCAGCAATCTCTGGACAAACAA
101 Y  V  Y  L  I  A  S  S  S  T  A  G  F  Q  F  R  D  E  I  K
301 TATGTCTACCTGATAGCAAGCTCAAGCACAGCTGGCTTTCAATTCAGAGACGAAATCAAA
121 I  L  V  S  Y  R  S  A  M  V  F  I  Q  T  D  K  P  I  Y  N
361 ATTCTGGTCAGTTATCGAAGTGCCATGGTCTTTATTCAGACCGATAAACCAATATACAAT
141 P  G  Q  T  V  N  L  R  V  V  P  L  S  L  D  L  K  A  S  V
421 CCTGGACAGACAGTCAATCTGCGAGTGGTTCCTCTCAGTCTTGATTTGAAGGCATCCGTC
161 D  N  V  T  I  E  V  M  N  P  Q  G  I  R  V  E  R  W  S  N
481 GATAATGTAACAATAGAAGTTATGAATCCTCAAGGTATTCgCGTGGAGAGATGGAGCAAC
181 L  N  T  K  A  G  F  F  S  R  R  L  D  L  S  E  N  V  L  L
541 CTGAATACAAAAGCAGGCTTTTTCTCACGTCGTTTGGATTTATCGGAAAACGTTTTGCTT
201 G  L  W  T  I  S  A  L  Y  G  H  G  K  V  Q  N  A  S  I  Q
601 GGCTTGTGGACCATTAGTGCCCTGTATGGCCATGGGAAAGTGCAAAACGCCTCGATACAA
221 F  E  V  R  K  Y  V  L  P  T  F  S  V  K  L  K  G  P  S  Y
661 TTTGAAGTTCGAAAATATGTGTTGCCAACGTTCTCTGTGAAATTGAAaGGACCATCCTAC
241 I  L  E  S  D  P  S  I  T  I  K  V  T  S  K  Y  T  Y  G  K
721 aTTTTAGAAAGTGACCCGTCGATCACAATAAAAGTTACATCAAAGTACACGTATGGTAAG
261 A  V  I  G  S  V  R  V  N  L  A  V  L  D  D  A  G  K  V  E
781 GCTGTCATTGGGTCAGTTCGAGTGAATCTTGCCGTCCTTGATGATGCTGGTAAAGTTGAA
281 R  F  S  T  S  I  H  T  L  R  N  G  E  A  D  V  I  V  S  T
841 AGATTTAGTACTTCCATACACACGCTACGTAACGGAGAAGCCGATGTTATCGTATCGACA
301 D  L  L  K  A  H  A  K  I  P  W  F  P  D  G  K  R  L  V  I
901 GACCTACTGAAGGCACATGCTAAGATTCCGTGGTTTCCTGATGGCAAGCGTCTAGTTATC
321 E  A  K  V  I  E  Q  A  T  G  H  E  E  K  A  L  D  N  T  I
961 GAAGCTAAAGTTATTGAACAAGCAACAGGACACGAAGAAAAGGCTTTGGACAATACGATA
341 Y  F  T  N  T  P  L  K  I  S  F  K  R  S  P  R  F  F  K  P
1021 TACTTCACAAACACTCCTCTGAAGATCAGCTTTAAGAGATCACCAAGATTTTTCAAACCT
361 G  V  P  F  E  I  K  V  D  V  K  Y  M  N  G  Q  P  A  N  E
1081 GGAGTCCCGTTTGAAATTAAGGTGGATGTAAAGTATATGAATGGACAGCCAGCAAACGAG
381 I  P  I  Q  I  D  A  K  T  N  D  G  T  V  V  R  E  R  L  A
1141 ATTCCCATTCAAATTGATGCAAAAACCAACGATGGAACAGTTGTGCGAGAACGCCTGGCT
401 A  G  Q  V  G  G  D  K  T  N  E  L  G  H  G  R  F  V  V  D
1201 GCCGGGCAAGTCGGTGGAGACAAAACGAACGAACTTGGTCACGGAAGATTTGtGGTTGAT
421 I  P  K  T  F  T  I  A  H  L  V  V  K  V  R  A  T  I  S  Q
1261 ATTCCCAAAACGTTTACCATAGCACATTTGGTTGTTAAAGTCCGTGCGACGATCAGTCAA
```

```
 441 G  G  K  D  I  I  S  E  G  R  F  Q  P  S  K  Y  R  S  S  G
1321 GGAGGAAAAGATATCATATCAGAAGGAAGATTCCAGCCGTCAAAGTACAGATCCAGtGGC
 461 N  N  Y  L  F  V  R  F  L  T  K  P  K  V  G  Q  T  V  D  A
1381 AATAATTATTTGTTCGtTCGTTTCCTAACCAAACCTAAAGTTGGACAAACTGTTGATGCA
 481 E  A  F  A  L  S  E  G  K  P  N  S  L  T  Y  M  V  I  A  N
1441 gAGGCGTTTGCTCTTTCTGAAGGAAAACCAAATTCACTGACTTATATGGTCATCGCAAAT
 501 G  K  V  V  F  Q  G  Q  I  N  R  D  L  G  V  L  T  T  V  R
1501 GGCAAGGTCGTGTTTCAAGGTCAAATCAACAGAGACCTTGGTGTGCTAACAACAGTCAGG
 521 I  R  V  T  S  A  M  I  P  Q  A  R  F  V  A  Y  Y  R  V  N
1561 ATTCGAGTGACCTCAGCAATGATTCCCCAGGCAAGATTTGTAGCTTATTATCGCGTGAAT
 541 N  E  L  V  A  D  S  T  I  M  E  V  E  E  E  L  P  N  Q  V
1621 AATGAACTGGTTGCTGaCAGCACCATCATGGAAGTGGAAGAAGAATTGCCCAATCAGGTC
 561 S  F  F  G  D  Q  H  S  Q  K  I  P  G  D  S  H  A  I  T  I
1681 TCATTTTTCGGGGACCAACATTCCCAGAAAATACCAGGCGATTCgCATGcGATTACAATA
 581 Q  S  S  P  H  S  N  V  G  I  L  A  V  D  Q  S  V  Y  L  L
1741 CAAAGCAGTCCACATTCCaACGTTGGTATACTGGCTGTAGACCAAAGTGTTTATTTGTTG
 601 R  N  D  K  H  L  T  S  D  E  V  Y  K  R  M  K  S  H  D  L
1801 CGAAATGATAAACATCTCaCTAGTGATGAGGTGTATAAAAGGATGAAATCCCACGACCTG
 621 G  C  G  S  G  A  G  A  D  N  K  D  V  L  N  R  G  G  L  A
1861 GGGTGTGGTTCGGGAgCAGGCgCGGACAACAAAGATGTTTTAAATCGTGGTGGTCTTgCG
 641 V  M  T  T  I  N  N  L  K  T  D  T  R  A  E  Y  S  C  A  A
1921 GTAATGACAaCTATCAATAATCTTAAGACAGATACCCGAgCAGAATATTCGTGTGCGGCT
 661 D  G  K  R  K  R  R  S  T  D  A  S  V  D  P  Q  C  C  I  L
1981 GATGgAAAGAGAAAAAGGCGAAGCaCAGATGCCTCTGTTGACCCGCAATGCTGTAtACTT
 681 G  E  D  L  D  P  A  T  C  L  V  R  A  M  K  F  S  V  S  N
2041 GGCGAAGAtCtTGATCCGGCAACATGTCttgTTCGAGCAATGAAATTTTCTGTTTCAAAT
 701 I  S  S  S  F  H  S  L  D  A  C  I  I  E  F  Y  K  C  C  Y
2101 ATATCATCATCATTCCATTCGCTTGACGCATGCATCATAGAGTTCTATAAATGTTGTTAC
 721 R  K  F  E  M  D  W  R  T  R  S  G  E  I  A  I  P  N  N  V
2161 AGGAAGTTTGAAATGGATTGGAGGACTCGATCTGGGGAGATTGCCATTCCAAATAATGtG
 741 L  D  E  L  P  F  E  D  E  E  I  L  K  L  T  L  D  E  A  Q
2221 TTGGATGAACTTCCTTTTGAAGATGAAGAAATTCTTAAGTTGACTTTGGATGAAGCACAA
 761 V  R  T  N  F  P  E  T  W  L  Y  E  H  M  K  A  D  K  D  G
2281 GTGCGAACAAATTTTCCCGAGACGTGGTTGTACGAACATATGAAAGCTGACAAAGACGGT
 781 R  V  S  F  R  V  T  V  P  D  T  I  T  T  W  I  M  Q  A  I
2341 CGTGTTTCGTTCCGTGTTACAGTACCAGATACGATCACCACTTGGATCATGCAAGCCATC
 801 A  V  S  N  T  T  G  F  G  L  T  P  P  F  N  L  K  A  F  K
2401 GCCGTTTCAAATACGACAGGATTTGGTTTAACTCCGCCTTTCAACTTGAAAGCCTTTAAG
 821 S  F  F  V  S  L  K  L  P  Y  S  A  Q  R  G  E  Q  V  S  V
2461 TCTTTcTTCGTTTCCTTGaAACTGCCTTACTCAGCACAGCGTGGcGAACAAGTCTCcGTG
 841 I  A  T  V  F  N  Y  K  D  Q  A  E  M  V  R  I  Y  L  F  K
2521 ATAGCTACCGTTTTCAACTATAAAGACCAAGCCGAaATGGTCAGAATTTATCTCTTCAAG
 861 K  P  N  D  D  F  C  T  Y  S  N  Y  G  S  G  S  S  L  Y  E
2581 AAGCCAAACGACGATTTCTGTAcGTATTCAAATTACGGCTCGGGCAGCTCACTTTATGAA
 881 V  L  V  D  A  H  G  A  T  S  V  S  F  P  I  V  P  T  E  L
2641 GTTCTAGTTGATGCTCATGGCGcgACTTCCGTATCTTTCCCTATTGTTCCCACTGAACTT
 901 G  D  I  P  I  Q  V  K  I  I  S  R  N  F  D  N  D  G  E  Q
2701 GGGgATATTCCCATCCAGGTCAAGATTATTTCAAGAAATTTTGACAATGATGGTGAaCAA
 921 R  I  L  K  V  V  P  E  G  I  E  R  R  E  T  H  S  V  V  L
2761 CGAATATTAAAAGTGGTGCCCGAGGGTATTGAAAGACGAGAAACTCATTCAGTGGTTCTA
 941 D  P  L  D  V  L  R  D  P  S  D  A  K  P  S  A  A  P  T  T
2821 GATCCATTAGATGTCTTGCGGGATCCATCaGATGCCAAACCAAGTGCAGCACCGACGACT
 961 P  S  K  I  Q  S  S  P  K  G  N  G  E  Q  N  N  R  L  S  L
2881 CCATCGAAAATACAGTCGTCGCCAAAGGGAAATGGAGAACAGAACAATCGACTGAGCCTC
 981 K  L  P  K  S  A  I  P  E  S  E  Y  A  M  L  T  V  I  G  T
```

```
2941 AAACTTCCTAAATCTGCCATCCCTGAATCAGAGTACGCCATGCTCACAGTGATCGGCACC
1001 L  I  G  P  S  V  S  N  I  I  G  G  R  G  L  D  S  I  I  K
3001 CTCATCGGCCCATCTGTATCGAACATCATTGGCGGTCGAGGACTGGATTCTATCATCAAG
1021 M  P  T  G  C  G  E  Q  T  M  L  K  L  A  P  N  V  F  V  F
3061 ATGCCTACGGGTTGTGGTGAGCAAACTATGTTGAAACTCGCCCCCAATGTGTTTGTATTC
1041 N  Y  L  R  S  T  K  Q  V  T  Q  Q  I  E  A  T  A  F  N  F
3121 AACTATCTGAGAAGCACCAAGCAGGTCACACAACAGATTGAAGCAACTGCGTTCAATTTT
1061 I  R  S  G  Y  Q  R  E  L  N  Y  R  R  S  D  N  S  F  S  A
3181 ATACGGTCTGGCtATCAACGTGAGTTGAACTATCGTAGAAGTGATAATTCGTTCaGTGCG
1081 F  G  N  S  R  A  G  S  T  W  L  T  A  F  V  I  K  T  F  C
3241 TTTGGAAACAGCAGAGCTGGAAGTACTTGGCTTACAGCGTTTGTCATCAAGACATTCTGC
1101 A  I  K  K  L  D  G  I  D  I  D  Q  N  V  I  N  T  A  I  N
3301 GCGATTAAAAAACTCGACGGAATAGATATTGATCAGAATGTGATCAAcACAGCAATTAaC
1121 W  L  S  S  R  Q  R  A  D  G  A  I  S  E  S  S  P  V  I  H
3361 tGGTTGTcGTCaCgACAGcGtgcTGAtGGTGCTAtATCAgAAAGtaGTCCTGTTatACAT
1141 Q  E  M  N  G  D  I  T  G  D  I  A  M  T  A  Y  V  V  T  A
3421 CAAgaAATGAATGGTGATaTTacTGGtGAcAtAGCAATGACAGCATaCGTTgTTaCAGcG
1161 F  L  E  C  E  S  V  A  P  N  S  V  Q  T  V  K  R  A  V  A
3481 TTCCTTgAATgcGAgAGTgTCGCaCCAAATTCTGTCCAAaCTgTGAAACGcGCCGtGGCA
1181 Y  L  E  N  M  Q  P  N  V  G  R  V  Y  V  K  A  V  I  A  Y
3541 TaCTTGGAGAACATGCaGCCAAATgTCGgCCGTGTTTaCGTAAAGgCTGTGATTGCAtAC
1201 A  L  A  L  A  D  S  P  L  E  V  K  R  Q  S  R  T  V  E  Q
3601 GCTTTGGCaTTAgCCGATaGTCCTcTtGAAGtTAAaCGCCAaTCAAGAAcTGTTGAACAG
1221 C  S  L  L  C  R  Q  E  H  R  R  Y  W  H  R  R  S  G  G  N
3661 TGCTCGTTATtATGCAgGCAAGAaCACcGCcGTTACTGGCACcGaCGGTCAGGGGGTAAT
1241 A  I  E  P  S  K  R  T  S  Y  A  L  A  N  T  N  G  S  E  Q
3721 GCTATCGAaCcGTCGAAACGAaCATCTTATGCaCTtGCTAACACAAATGGTtCTGAACAG
1261 A  W  L  R  R  S  Y  R  C  L  A  D  G  T  K  R  G  G  G  G
3781 GCGTGGTTACGCAGGAGCTATCGTTGTCTGGCTGACGGAACAAAGAGAGGTGGCGGTGGG
1281 F  I  S  T  Q  D  T  C  V  A  L  Q  A  L  A  A  Y  S  E  K
3841 TTcATTTCAACTcAGGAtACATGCGTGGCTCTACAAGCgTTGGCTGCCTACaGCGAGAAA
1301 T  G  G  D  Q  M  D  L  R  I  E  V  S  T  D  G  D  Y  K  K
3901 ACTGGaGGAgATCAAATGGaCCTTCGCATTGAAGtCtCTACagACGGagATTACAaGAAg
1321 T  L  I  V  N  Q  K  N  A  L  V  Q  Q  Q  L  D  I  S  S  L
3961 ACTTTgATaGtGAaCCagAAAAATGCCCTAgTGCAgCaaCaGTTagaTATTTCgaGCCTG
1341 I  G  D  E  L  F  I  K  T  K  G  S  G  V  A  Q  L  Q  V  E
4021 atCggAGAtgAACTgTTcaTTaAgACAAAaGGTtCGGGAGTAGCTCAACTACAGGTCGAG
1361 T  R  Y  N  T  P  P  T  E  K  E  V  C  Q  F  D  L  R  V  I
4081 ACCAGATACAACaCTCCGCCGACGGAGAAAGAAGTGTGCCAGTTCGATTTAAGAGTGATC
1381 T  I  E  R  A  R  R  M  Y  D  Q  P  I  N  D  A  P  K  P  T
4141 ACCATTGAGCGAGCGAGAAGGATGTACGACCAGCCGATCAACGACGCACCAAAACCGACG
1401 K  A  P  K  K  K  K  N  R  P  G  K  G  K  G  R  K  R  N  R
4201 AAAGCGCCAAAGAAAAAGAAAAATAGGCCCGGGAAGGGAAAGGGCAGAAAGCGCAACAGA
1421 N  K  K  C  R  R  K  N  G  R  R  C  S  G  C  K  G  R  R  C
4261 AACAAGAAATGTCGACGAAAGAATGGACGCAGGTGCAGTGGCTGCAAAGGAAGAAGATGT
1441 R  K  P  K  P  T  T  A  A  P  Q  V  T  T  R  P  P  P  E  D
4321 AgAAAGCCCAAGCCCaCgaCAGCTGCTCCACAAGTTaCCaCGcGTCCACCCCCTGAAGAT
1461 G  P  V  P  N  S  V  S  I  K  I  C  T  R  F  K  K  A  G  A
4381 GGACCTGTGCCCAACTCGGTATCGATCAAAATCTGTACCAGGTTTAAGAAAGCAGGGGCC
1481 S  A  G  M  S  I  I  D  V  G  I  L  T  G  F  S  V  K  Q  E
4441 AGTGCCGGgATGTCGATCATTGATGTTGGTATTCTCACCGGATTTAGCGTTAAACAAGAA
1501 S  L  V  E  L  Q  E  K  V  K  P  G  I  S  K  F  E  I  S  D
4501 AGCTTAGTTGAGCTTCAAGAGAAAGTGAAGCCTGGTATTTCGAAATTCGAGATCTCGGAT
1521 R  H  A  I  L  Y  I  D  E  I  P  S  D  H  E  L  C  F  N  L
4561 CGCCATGCAATTCTATACATCGATGAAATACCAAGTGATCACGAATTATGTTTCAACTTG
```

```
1541 E  L  T  R  D  F  S  V  G  I  V  Q  P  V  P  V  T  V  Y  D
4621 GAACTAACGAGAGATTTTTCAGTCGGCATTGTTCAGCCTGTACCGGTAACTGTGTATGAC
1561 Y  Y  E  P  D  N  K  C  T  K  F  Y  G  P  E  P  N  S  L  L
4681 TATTATGAACCAGATAACAAATGCACAAAGTTCTATGGACCCGAACCAAACAGTCTCTTA
1581 N  L  A  T  C  E  H  D  T  C  K  C  A  L  D  K  C  S  S  C
4741 AACTTGGCTACGTGcGAGCACGACACCTGCAAATGTGCCTTAGACAAATGTTCCTCGTGC
1601 K  T  S  D  D  S  A  V  V  K  G  L  F  C  T  T  Y  D  Y  A
4801 AAGACATCCGATGATTCTGCTGTTGTAAAGGGACTGTTCTGCACAACATACGATTATGCC
1621 F  K  G  K  L  L  I  I  D  E  E  D  Q  W  L  H  L  T  F  E
4861 TTTAAAGGAAAATTACTAATAATCGACGAGGAAGATCAATGGCTTCATCTCACGTTTGAA
1641 V  V  E  V  Y  K  E  S  V  T  K  K  I  T  K  K  T  A  R  I
4921 GTCGTTGAAGTTTATAAAGAAAGCGTAACCAAAAAGATAACCAAAAAAACAGCACGTATC
1661 V  Y  S  K  K  I  S  C  D  C  P  V  F  A  G  K  I  D  R  H
4981 GTGTATTCGAAAAAAATCAGTTGCGACTGCCCCGTGTTTGCTGGCAAAATTGACCGCCAT
1681 F  L  I  M  G  K  D  V  G  L  R  G  S  S  K  V  V  L  G  H
5041 TTCCTTATTATGGGAAAGGACGTTGGTCTTCGGGGATCCAGCAAAGTTGTCTTGGGTCAC
1701 N  V  F  V  K  E  W  P  M  N  D  P  V  D  F  F  K  K  F  V
5101 AATGTGTTTGTCAAAGAATGGCCAATGAACGATCCGGTAGATTTCTTCAAGAAGTTCGTA
1721 R  L  L  R  K  D  G  C  *
5161 AGGCTTTTGAGAAAGGACGGTTGCTGA
```

atgaatgtattggcaatctgtcttggttaccacgagtaaagcagattcatatgactagaaaatatatagaaag
taaaccacaccatctgtgtctactttagaatgatttaaggtcattaatttgcaatatagccgttataaacgtgta
aattagctaaaatggaaatatatgtagatttttgcaaaaaaaaaaaaaaaa

Figure 4. continued.

Primers used for SeC3 cloning.

Initial degenerate PCR product:

Sense: (C3-thiosense)—5'GNTGYGGNGARCARAAYATG

Antisense: (C3-thioantisense)—5'ACRTANGCNGTNAGCCANGT

Cloned product pursued in RACE amplification of entire gene was: clone C3(2-35).

C3(2-35)S: 5'CTATCTGAGAAGCACCAAGC
C3(3'SN2): 5'GATACAACACTCCGCCGACG
C3(3'SN1): 5'CCATTGAGCGAGCGAGAAGG
C3(3'Send):5'GGCCAATGAACGATCCGGTA

← 3'RACE PIMERS

C3(Sa/g): 5'GATCACCATTGAGCGAGCGA
C3(ASa/g): 5'CGCGTGGTAACTTGTGGAGC

← CONFIRM α-γ CLEAVAGE SITE

C3(2-35)AS1: 5'TGTACTTCCAGCTCTGCTGT
C3(2-35)ASN: 5'CAACTCACGTTGATAGCCAG
C3-AS3: 5'GGCAGCCAACGCTTGTAGAG
C3-AS4: 5'CAGTCCTCGACCGCCAATGA
C3-AS4N: 5'GAGTCGTCGGTGCTGCACTT
C3-AS5: 5'GATACGGAAGTCGCGCCATG
C3-AS5N: 5'CACGGAGACTTGTTCGCCAC
C3-AS6: 5'GAGACTTGTTCGCCACGCTG
C3-AS6N: 5'GGCGATGGCTTGCATGATCC
C3-AS7: 5'GATGCATGCGTCAAGCGAAT
C3-AS7N: 5'CAAGACATGTTGCCGGATCA
C3-AS8: 5'CCGTGACCAAGTTCGTTCGT
C3-AS8N: 5'CGCACAACTGTTCCATCGTT
C3-AS8N2: 5'CGTTTGCTGGCTGTCCATTC
C3-AS9: 5'CACAGAGAACGTTGGCAACA
C3-AS9N: 5'CACTAATGGGTCCCACAAGC
C3-AS10: 5'GGCACTTCGATAACTGACCA
C3-AS10N: 5'CAGAGATTGCTGGTCGTGAA

← 5'RACE PRIMERS

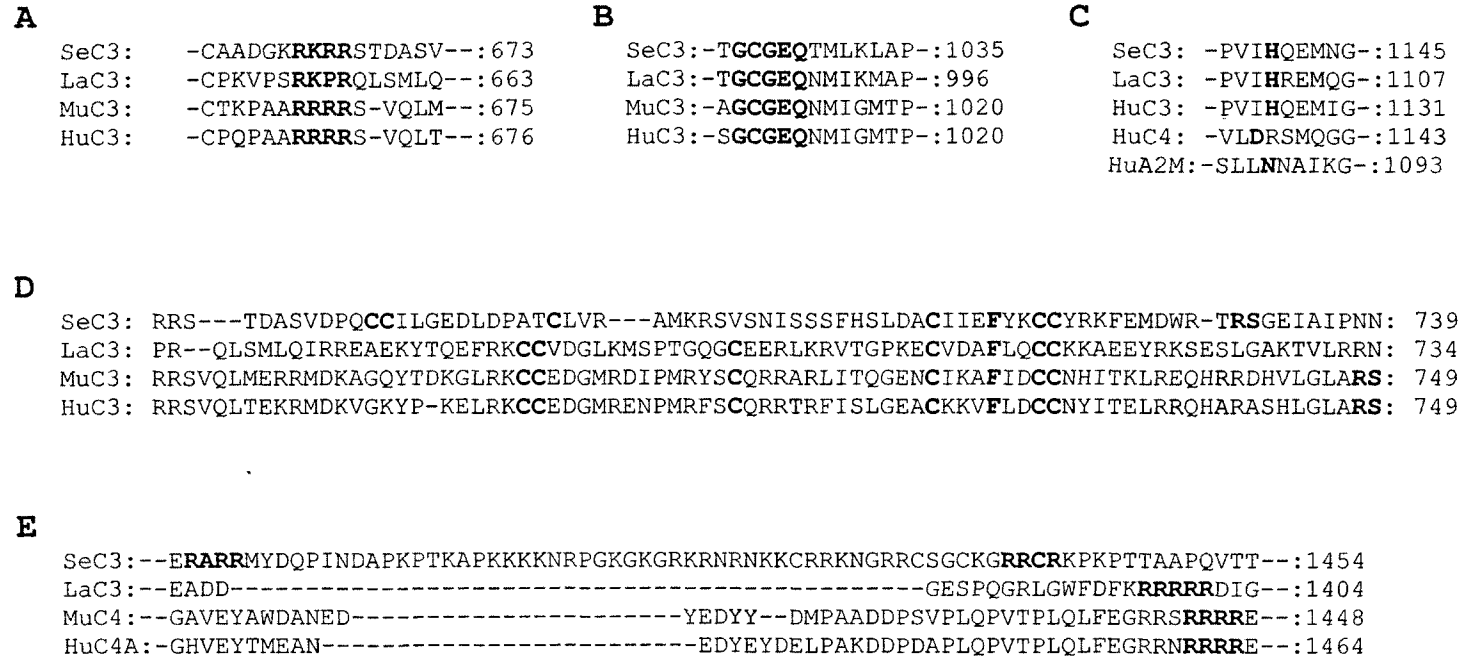


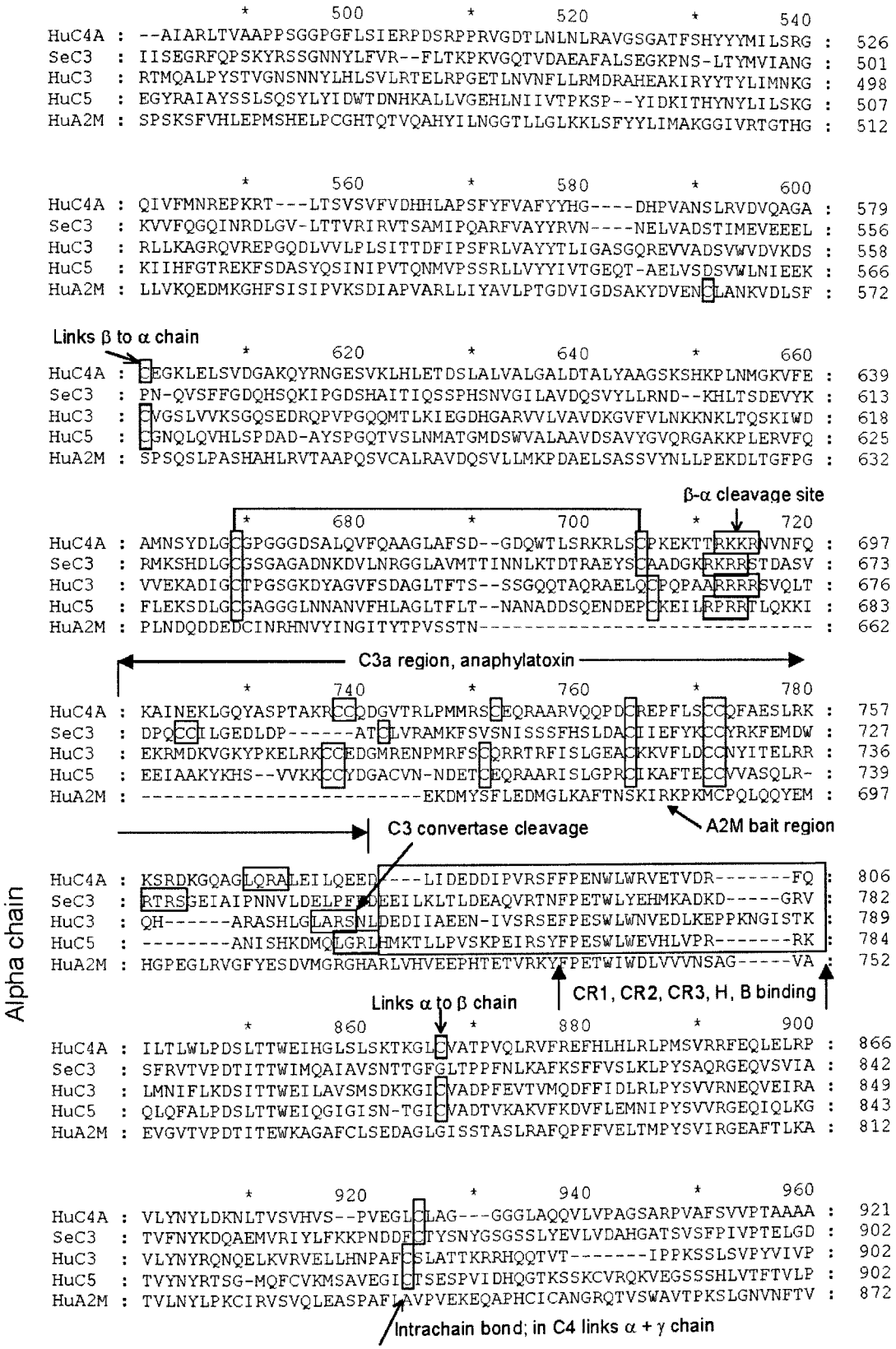
Figure 5. Sequence analysis and comparison of functional sites of interest. Residues of interest are in bold face. A. SeC3 contains putative beta-alpha cleavage site. B. Thiolester region of SeC3 and other C3 proteins. C. Region of catalytic residue, downstream of thiolester site. In C3, the catalytic residue is usually histidine, whereas it is usually arginine in C4 and asparagine in A2M. D. C3a anaphylatoxin region, note presence of the typical six cysteines found in vertebrate C3. E. SeC3 contains a putative alpha-gamma cleavage site, a characteristic of C4 and in Lamprey C3. The putative three chain structure may have been characteristic of the ancestor. Note that SeC3 contains two potential cleavage sites in this area, where if both are cleaved, a 74 amino acid peptide with very unique characteristics is generate

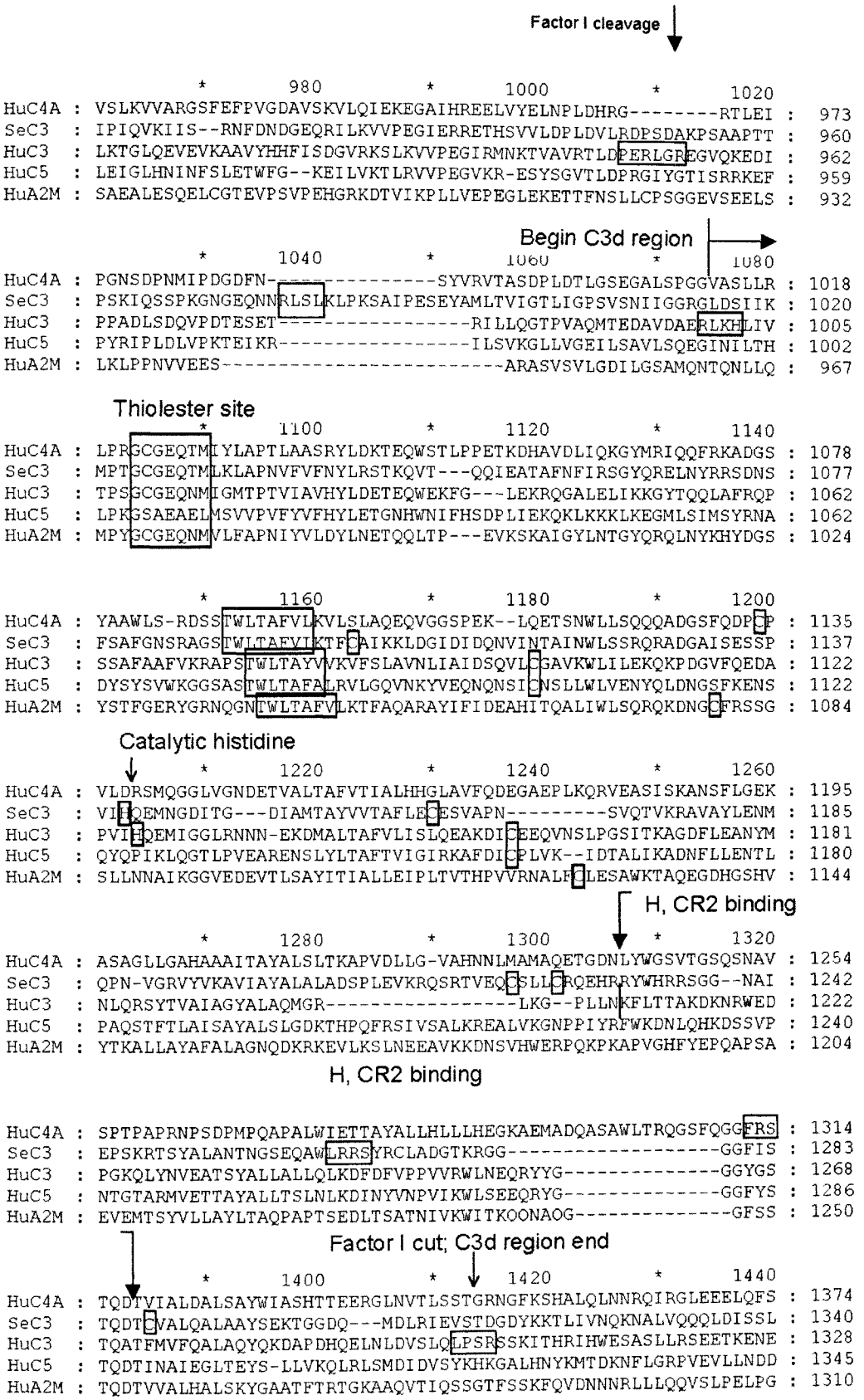
Figure 6a. Full-length sequence alignment of SeC3 and Human C4A, C3, C5 and A2M. All major reactive sites, receptor binding sites, and cysteines are boxed or highlighted. Where applicable, the different chains and the respective cleavage sites have been labeled. Labeling of sites is based on what is known from human C3 from functional and biochemical studies (Sahu & Lambris, 2001; Morley & Walport, 2000).

Figure 6a. Full-length alignment with areas of interest highlighted.

```
                  *         20          *         40          *         60
HuC4A  : TRSAPRAASWLEDPREVRSVCLSATFFTLSLQKPRLLLFSPSVVHLGVPLSVGVQLQDVP :  60
SeC3   : ----------MKMLRALIGFALLLCLN--QCYAAKYFIAAPNLLRVGVEETVSIAVFDVN :  48
HuC3   : ---------MGPTSGPSLLLLLLTHLPLALG-SPMYSIITPNILRLESEETMVLEAHDAQ :  50
HuC5   : ------------MGLLGILCFLIFLGKTWGQEQTYVISAPKIFRVGASENIVIQVYGYT :  47
HuA2M  : ------------------------MGKNKLLHPSLVLLLLVLLPTDASVSGKPQYMVLV :  35

                  *         80          *         100         *         120
HuC4A  : RGQVVKGSVFLRNPSRNNVPCSPKVDFTLSSERDFALLSLQVPLKDAKSCGLHQLLRGPE :  120
SeC3   : -VDVNVQLALQDFPNRRKTFSQVSGNVRAQQP---GILKIKVNAKDLHDQ---QSLDKQY :  101
HuC3   : GDVPVTVTVHDFPGKKLVLSSEKTVLTPATNHMGNVTFTIPANREFKSEKGRNKFVTVQA :  110
HuC5   : EAFDATISIKSYPDKKFSYSSGHVHLSSENKFQNSAILTIQP-KQLPGGQNPVSYVYLEV :  106
HuA2M  : PSLLHTETTEKGCVLLSYLNETVTVSASLESVRGNRSLFTDLEAENDVLHCVAFAVPKSS :  95

                  *         140         *         160         *         180
HuC4A  : VQLVAHSPWLKDSLSRTTNIQGINLLFSSRRGHLFLQTDQPIYNPGQRVRYRVFALDQKM :  180
SeC3   : VYLIASS-----STAGFQFRDEIKILVSYRSAMVFIQTDKPIYNPGQTVNLRVVPLSLDL :  156
HuC3   : TFGTQVV--------------EKVVLVSLQSGYLFIQTDKTIYTPGSTVLYRIFTVNHKL :  156
HuC5   : VSKHFSK--------------SKRMPITYDNGFLFIHTDKPVYTPDQSVKVRVYSLNDDL :  152
HuA2M  : SNEEVMFLTVQVKGPTQEFKKRTTVMVKNEDSLVFVQTDKSIYKPGQTVKFRVVSMDENF :  155

                  *         200         *         220         *         240
HuC4A  : RPSTDTITVMVENSHGLRVRKKE-VYMPSS--IFQDDFVIPDISEPGTWKISARFSDGLE :  237
SeC3   : KASVDNVTIEVMNPQGIRVERWSNLNTKAG--FFSRRLDLSENVLLGLWTISALYGHGKV :  214
HuC3   : LPVGRTVMVNIENPEGIPVKQDSLSSQNQLGVLPLS-WDIPELVNMGQWKIRAYYENSPQ :  215
HuC5   : KPAKRETVLTFIDPEGS--EVDMVEEIDHIGIISFPDFKIPSNPRYGMWTIKAKYKEDFS :  210
HuA2M  : HPLNELIPLVYIQDPKGNRIAQWQSFQLEGG---LKQFSFPLSSEPFQGSYKVVVQKKSG :  212

                  *         260         *         280         *         300
HuC4A  : SNSSTQFEVKKYVLPNFEVKITPGKPYILTVPGHLDEMQLDIQARYIYGKPVQGVAYVRF :  297
SeC3   : QNASIQFEVRKYVLPTFSVKLK-GPSYILESD---PSITIKVTSKYTYGKAVIGSVRVNL :  270
HuC3   : QVFSTEFEVKEYVLPSFEVIVEPTEKFYYIYNEKGLEVTITARFLYG-KKV-EGTAFVIF :  273
HuC5   : TTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKNFKNFEITIKARYFYN-KVVTEADVYITF :  269
HuA2M  : GRTEHPFTVEEFVLPKFEVQVTVPKIITILEEEMNVSVCGLYTYGKPVPGHVTVSICRKY :  272

                  *         320         *         340         *         360
HuC4A  : GLLDEDGKKTFFRGLESQTKLVNGQSHISLSKAEFQDALEKLNMGITDLQGLRLYVAAAI :  357
SeC3   : AVLDDAGKVERFS--TSIHTLRNGEADVIVSTDLLKAHAKIP----WFPDGKRLVIEAKV :  324
HuC3   : GIQD---GEQRISLPESLKRIPIEDGSGEVVLSRKVLLDGVQNLRAEDLVGKSLYVSATV :  330
HuC5   : GIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELSYYSLEDLNNKYLYIAVTV :  329
HuA2M  : SDASDCHGEDSQAFCEKFSGQLNSHGCFYQQVKTKVFQLKRKEYEMKLHTEAQIQEEGTV :  332

                  *         380         *         400         *         420
HuC4A  : IESPGGEMEEAELTSWYFVSSPFSLDLSKTKRHLVPGAPFLLQALVREMSGSPASGIPVK :  417
SeC3   : IEQATGHEEKALDNTIYFTNTPLKISFKRSPRFFKPGVPFEIKVDVKYMNGQPANEIPIQ :  384
HuC3   : ILHSGSDMVQAERSGIPIVTSPYQIHFTKTPKYFKPGMPFDLMVFVTNPDGSPAYRVPVA :  390
HuC5   : IESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDSLDQLVGGVPVI :  389
HuA2M  : VELTGRQSSEITRTITKLSFVKVDSHFRQGIPFFGQVRLVDGKGVPIPNKVIFIRGNEAN :  392

                  *         440         *         460         *         480
HuC4A  : VSATVSSPGSVPEAQDIQQN----TDGSGQVSIPIIIPQTISELQLSVSAGSPHP----- :  468
SeC3   : IDAKTNDGTVVRERLAAGQVGGDKTNELGHGRFVVDIPKTFTIAHLVVKVRATISQGGKD :  444
HuC3   : VQGEDTVQSLTQ-----------GDGVAKLSINTHPSQKPLSITVRTKKQELSEAEQAT :  438
HuC5   : LNAQTIDVNQETSDLDPSKSVTRVDDGVASFVLNLPSGVTVLEFNVKTDAPDLPEENQAR :  449
HuA2M  : YYSNATTDEHGLVQFSINTTNVMGTSLTVRVNYKDRSPCYGYQWVSEEHEEAHHTAYLVF :  452
```

Beta chain


Links β to α chain
β-α cleavage site
C3a region, anaphylatoxin
C3 convertase cleavage
A2M bait region
Alpha chain
Links α to β chain
CR1, CR2, CR3, H, B binding
Intrachain bond; in C4 links α + γ chain


Factor I cleavage
Begin C3d region
Thiolester site
Catalytic histidine
H, CR2 binding
H, CR2 binding
Factor I cut; C3d region end

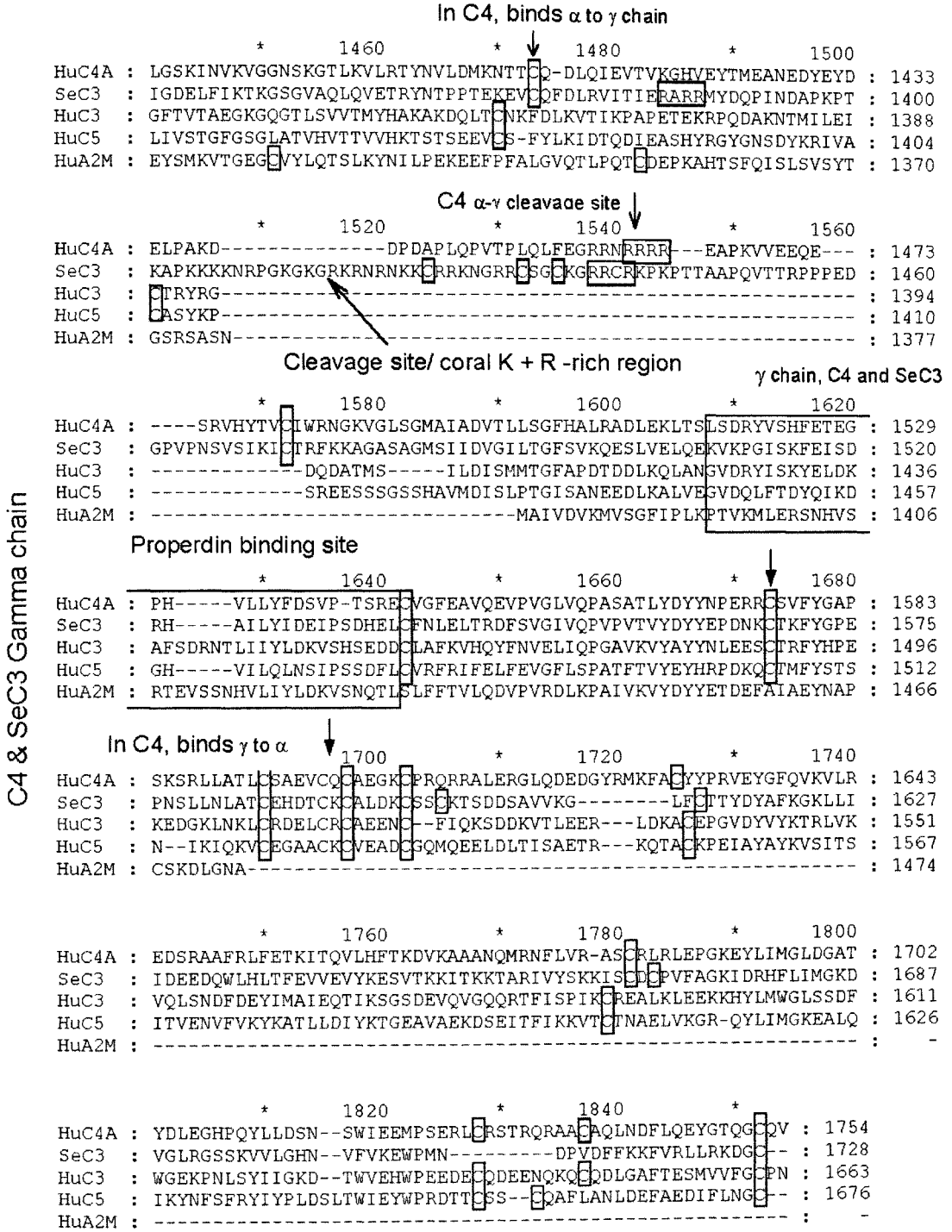
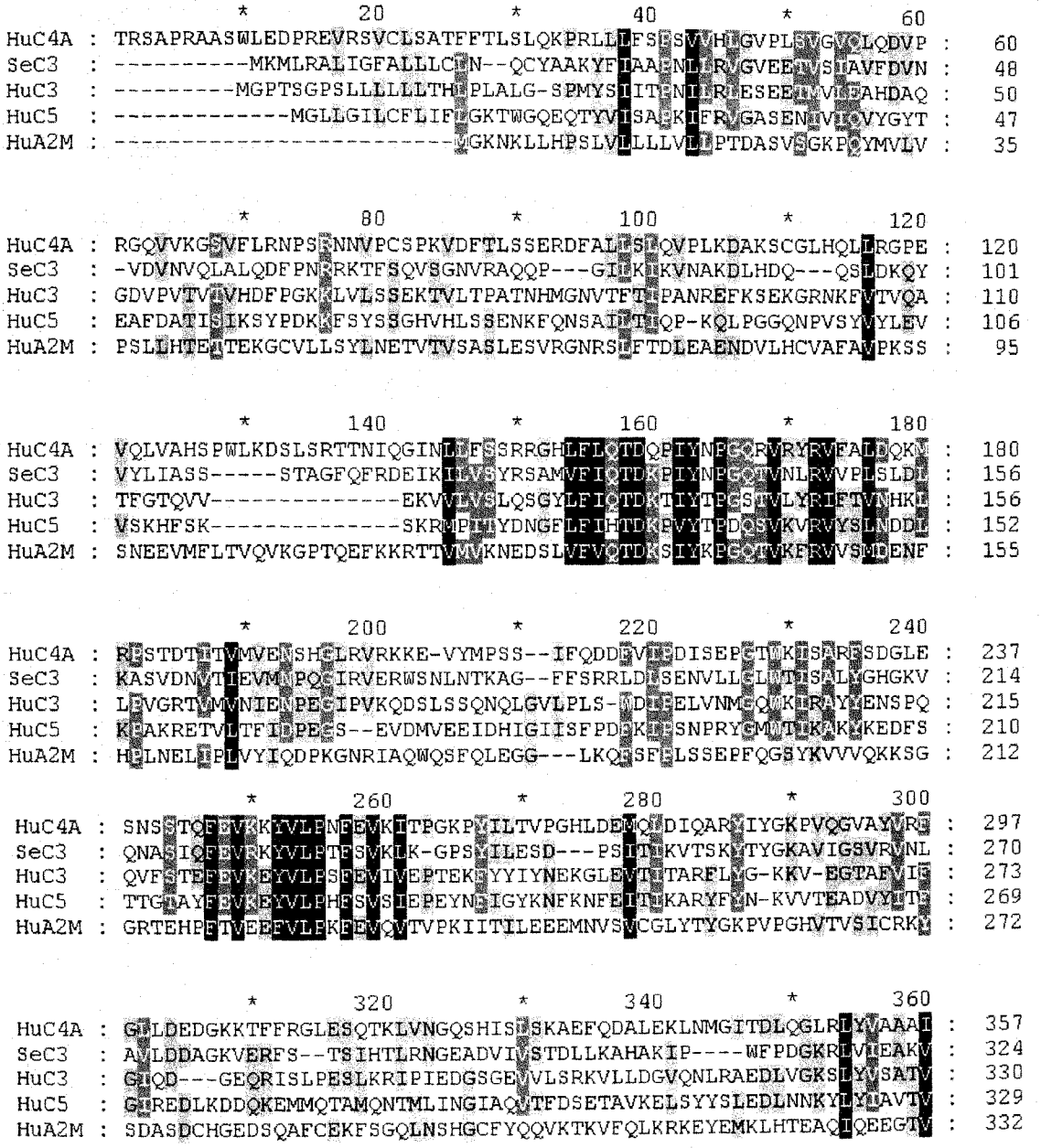
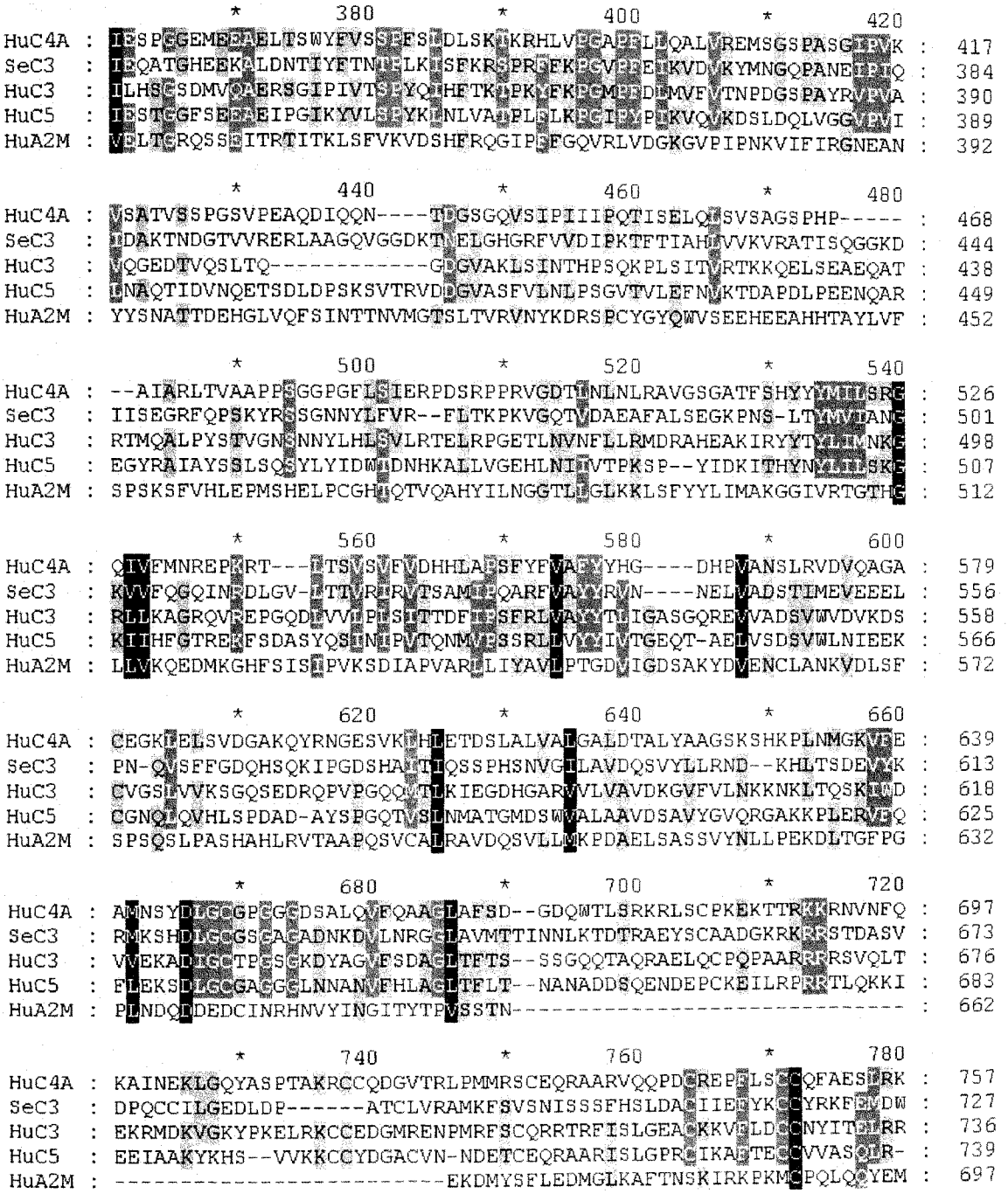
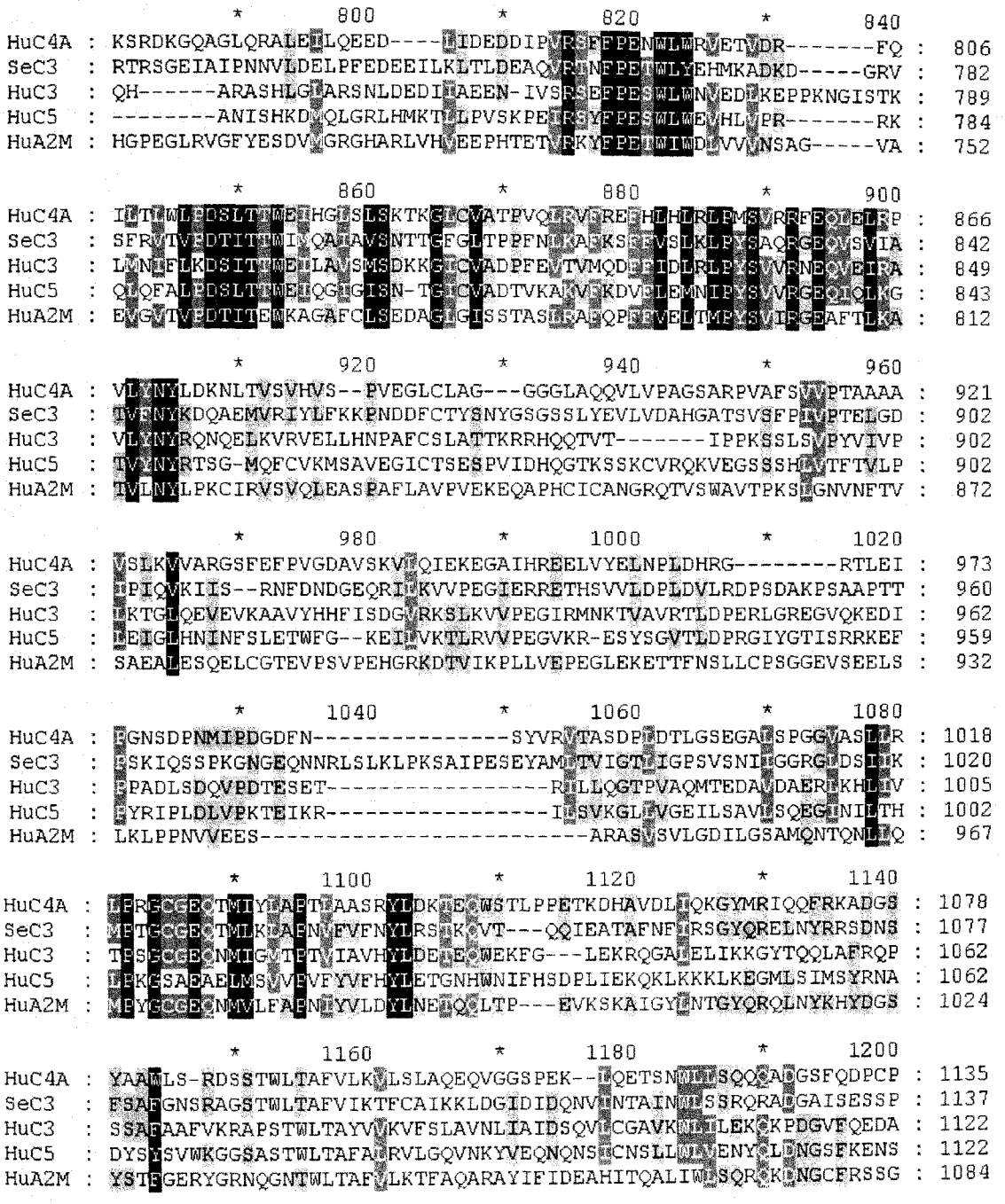
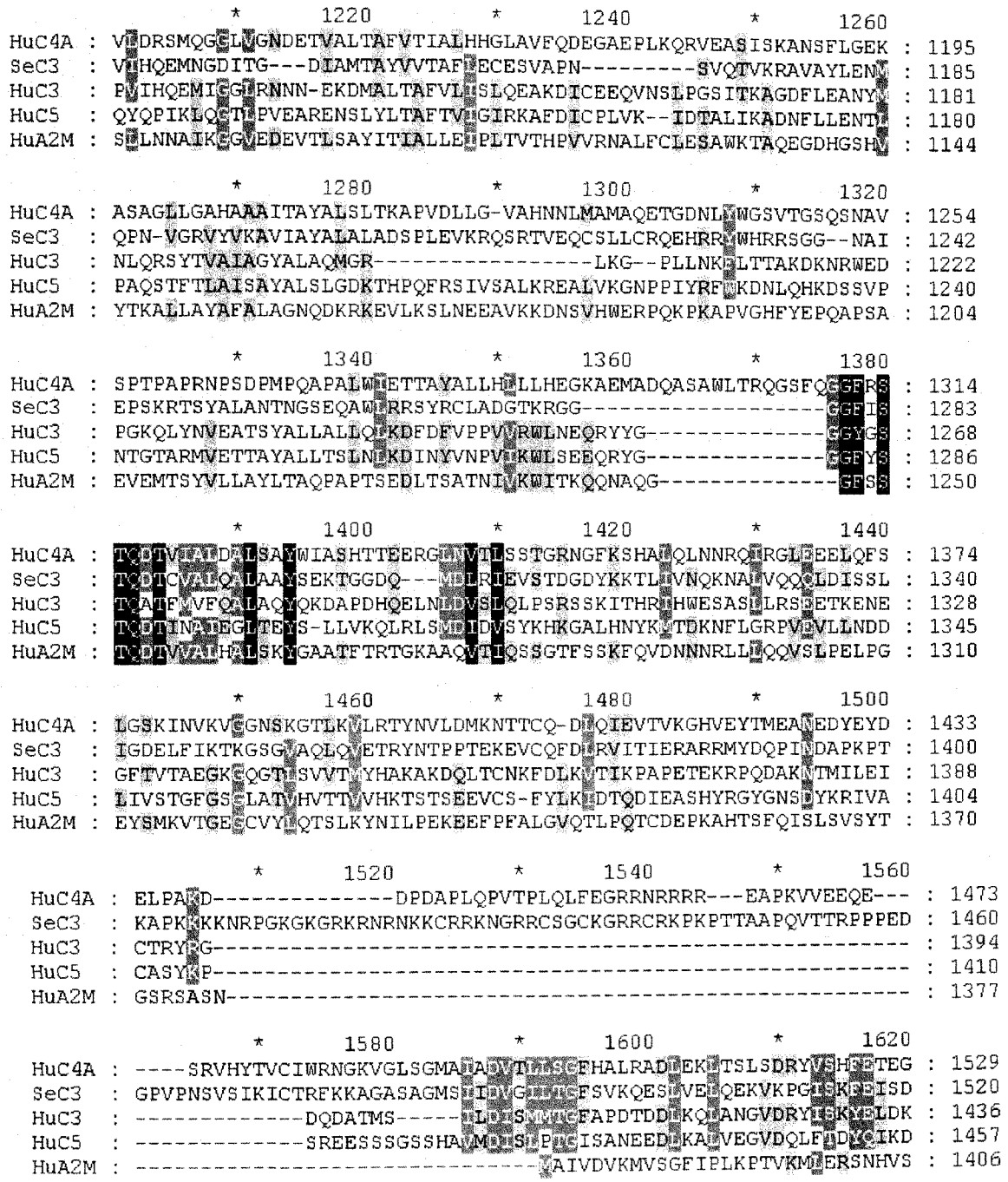

In C4, binds α to γ chain
C4 α-γ cleavage site
Cleavage site/ coral K + R -rich region
γ chain, C4 and SeC3
Properdin binding site
In C4, binds γ to α
C4 & SeC3 Gamma chain

Figure 6b. Same alignment as in 6a, but shaded for conservative residues. Vertical shading corresponds to conservative amino acids in that position of the alignment. Darker colors equal higher conservation among all sequences. Conservative substitutions allowed; shading includes conserved physiochemical properties.









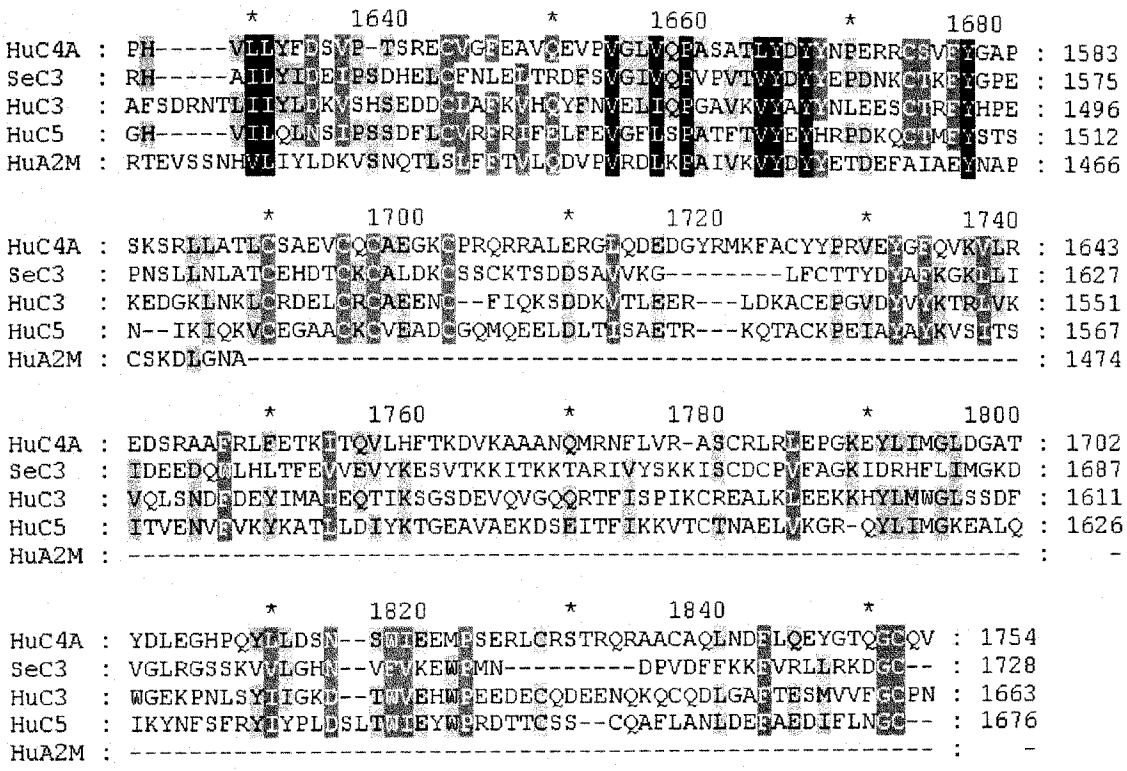
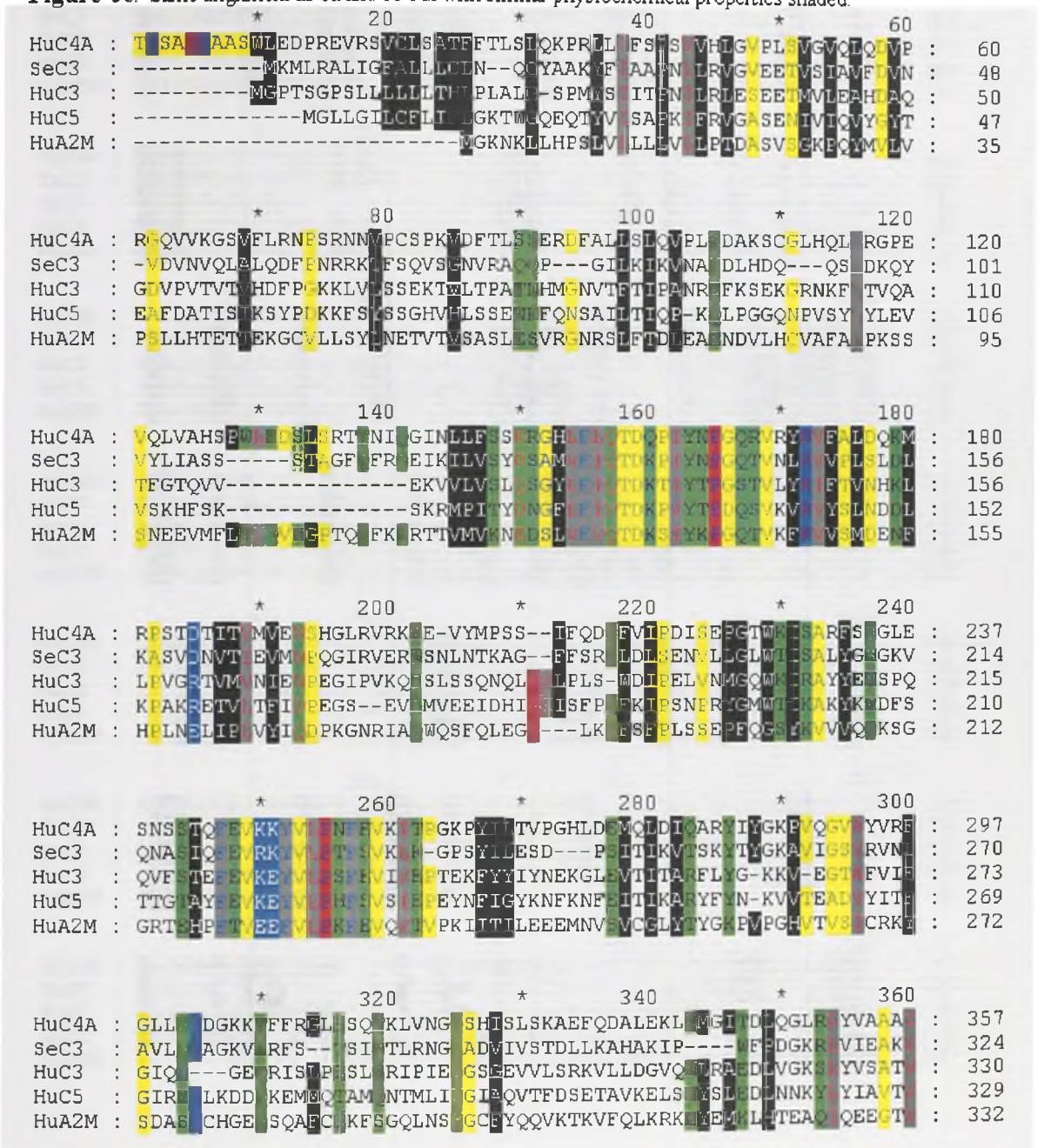
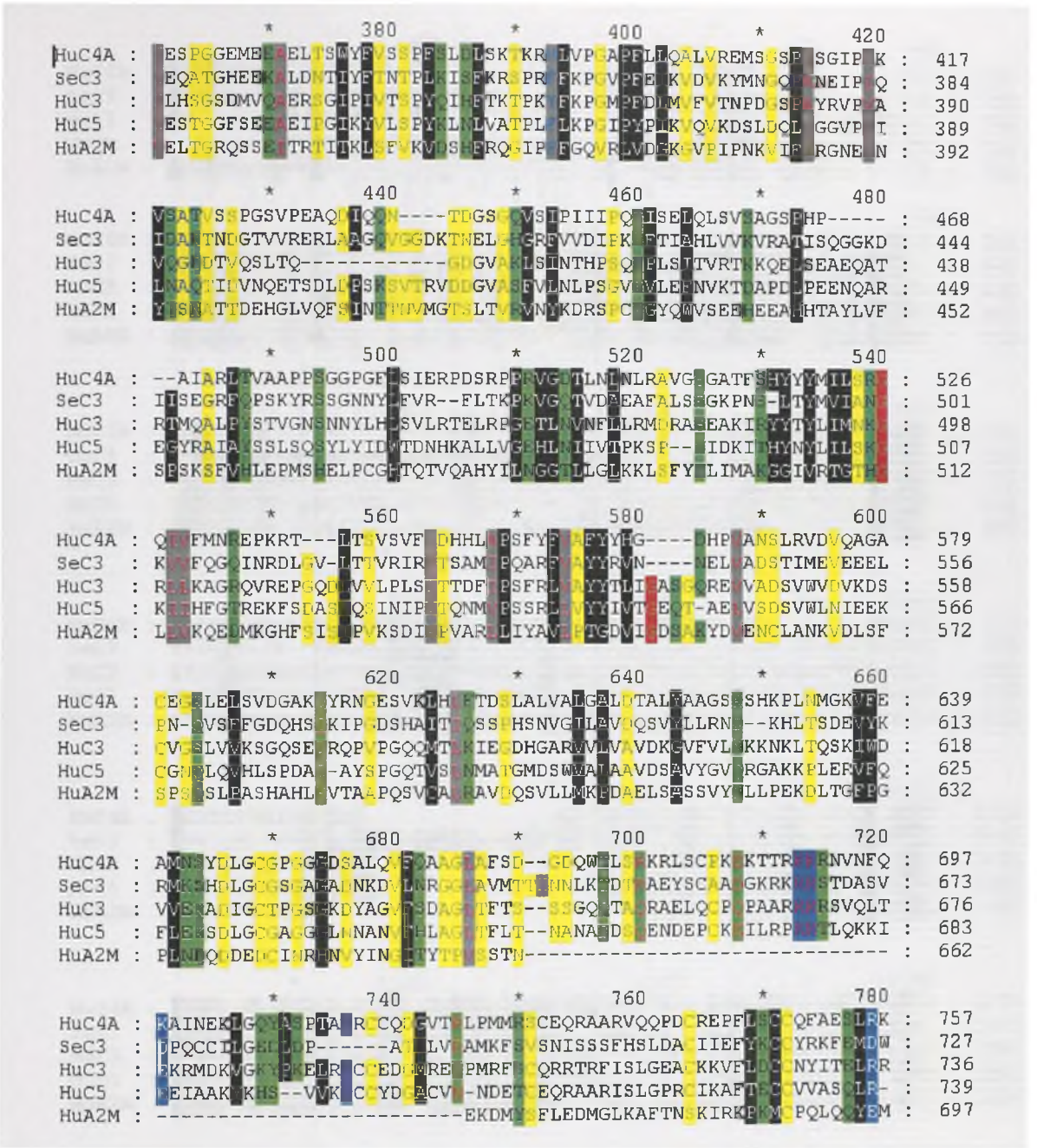
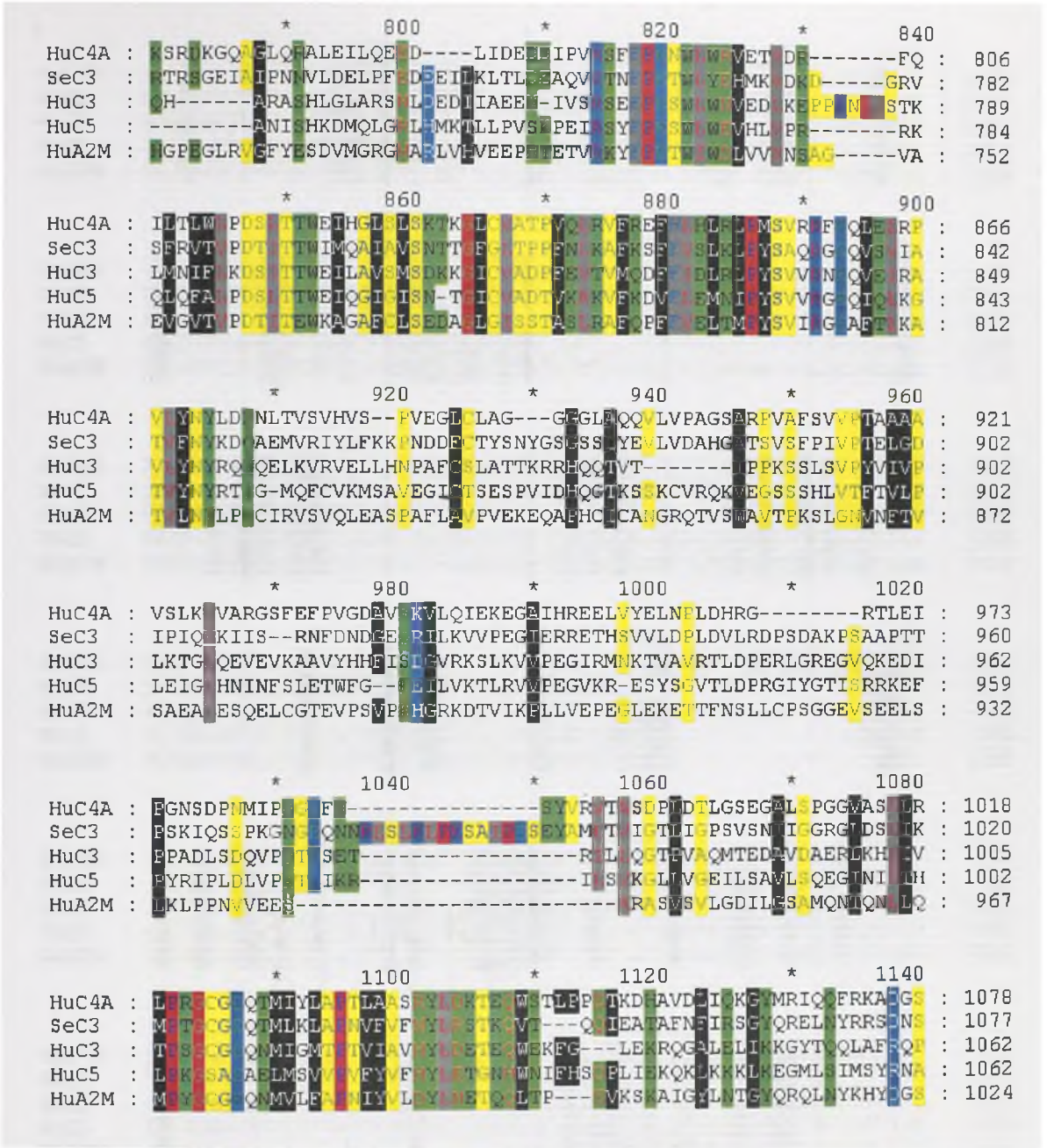
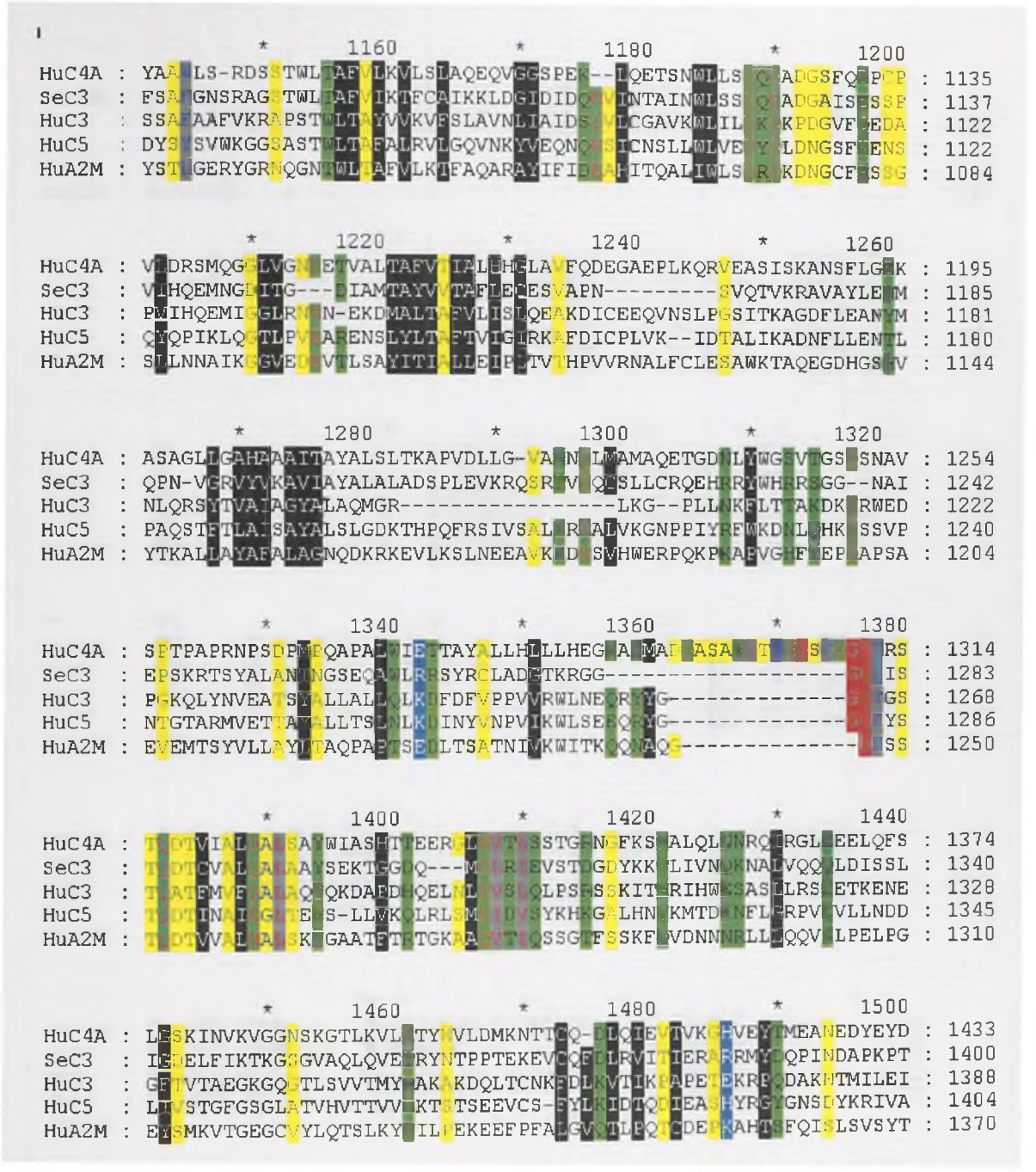

1640
1660
1680
1700
1720
1740
1760
1780
1800
1820
1840
HuC4A
SeC3
HuC3
HuC5
HuA2M

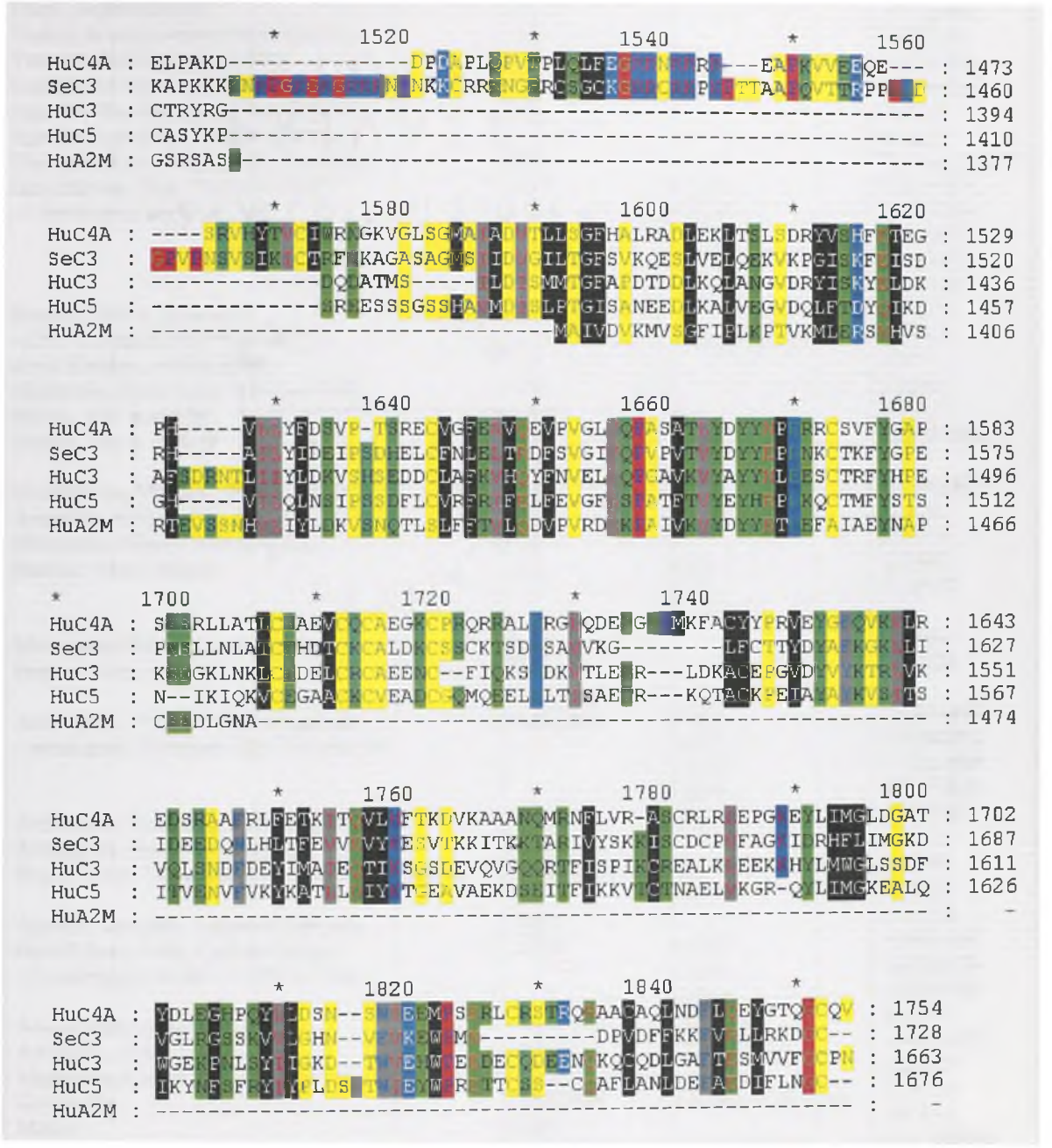
Figure 6c. Same alignment as 6a and 6b but with similar physiochemical properties shaded.











Animal	Gene	Dupl. Paralog	Accession #
Coral, *Swiftia exserta*	SeC3		AY186744
Urchin, *Strongylocentrotus purpuratus*	C3		AF025526
Tunicate, *Halocynthia roretzi*	C3		AB006864
Cephalochrodate (Amphioxus) *Branchiostoma belcheri*	C3		AB050668
Agnatha, Hagfish, *Eptatretus burgeri*	C3		Z11595
Agnatha, Lamprey, *Lampreta japonica*	C3		D10087
Chondrichthyes, Dogfish, *Triakis scyllia*	C3		M. Nonaka, Unpub.
Ostrichthyes, Carp, *Cyprinus carpio*	C3	C3-H1	AB016211
--5 paralogous copies in carp.		C3-H2	AB016212
		C3-S	AB016213
		C3-Q1	AB016214
		C3-Q2	AB016215
Reptilia, Cobra, *Naja naja*	C3		Q01833
--CVF is a paralogous copy of C3.		Venom factor,CVF	U09969
Aves, Chicken, *Gallus gallus*	C3		I50711
Mammalia, Guinea pig, *Cavia porcella*	C3		P12387
Mouse, *Mus musculus*	C3		P01027
Human, *Homo sapiens*	C3		NM_000064
Ostrichthyes, Medaka, *Oryzias latipes*	C4		BAA92287
Amphibia, African frog, *Xenopus laveis*	C4		D78003
Mammalia, Mouse, *Mus musculus*	C4		P01029
Human, *Homo sapiens*	C4	C4A	K02403
		C4B	U24578
Mammalia, Mouse, *Mus musculus*	C5		P06684
Human, *Homo sapiens*	C5		M57729
Arthropoda, *Drosophila melanogaster*	A2M-like	TEP1	AAF53490
--paralogous, divergent A2M-like proteins		TEP2	CAB87808
		TEP3	CAB87809
		TEP4	AAF53826
Arthropoda, Mosquito, *Anopheles gambiae*	A2M-like	TEP1	AF291654
Arthropoda, Horseshoe crab, *Limulus sp.*	A2M		D83196
Round worm, Nematoda, *Caenorhabditis elegans*	A2M-like	TEP1	Z82090
		TEP2	Z75527
Agnatha, Lamprey, *Lampreta japonica*	A2M		D13567
Ostrichthyes, Carp, *Cyprinus carpio*	A2M	A2M1	AB026128
--3 paralogous copies of A2M in Carp		A2M2	AB026129
		A2M3	AB026130
Aves, Chicken, *Gallus gallus*	A2M-like	Ovastatin	X78801
Amphibia, *Xenopus laevis*	A2M-like	Endodermin	AAB51432
Mammalia, Guinea pig, *Cavia porcella*	A2M		D84338
Guinea Pig	A2M-like	GP-Murinoglobulin	D84339
Mouse	A2M		Q61838
Mouse	A2M-like	Murinoglobulin	NM_008646
Rat, *Rattus norvegicus*	A2M		NM_012488
Rat	A2M-like	Alpha-1-inhibitorIII	J03552
Human	A2M		NM_000014
Human	A2M-like	Preganancy zone protein,HuPZP	NM_002864

Table 1. Database accession numbers of TEP sequences used throughout this study.

Organism-Gene	Accession #	I(%) *	S(%) **	Distance ± SE¶	p-Dist. (± 0.01)
PWM: C3 (n=16)§	--	24	44	1.300 ± 0.041	0.731
UrchinC3	AF025526	26	45	1.270 ± 0.040	0.719
TunicateC3	AB006964	18	40	1.635 ± 0.050	0.805
AmphioxusC3	AB050668	32	52	1.084 ± 0.034	0.662
LampreyC3	D10087	25	45	1.296 ± 0.041	0.726
CarpC3-H1	AB016211	25	45	1.256 ± 0.040	0.715
ChickenC3	I50711	25	45	1.288 ± 0.041	0.724
MouseC3	P01027	24	44	1.313 ± 0.041	0.731
HumanC3	NM_000064	24	43	1.322 ± 0.041	0.733
PWM: C4 (n=5)§	--	23	44	1.350 ± 0.042	0.760
Xenopus C4	D78003	24	45	1.361 ± 0.042	0.744
Mouse C4	P01029	23	43	1.320 ± 0.042	0.748
HumanC4A	K02403	24	43	1.320 ± 0.041	0.733
PWM: C5 (n=2)	--	22	44	1.404 ± 0.044	0.754
Mouse C5	P06684	22	44	1.415 ± 0.044	0.757
HumanC5	M57729	22	44	1.392 ± 0.043	0.751
PWM: A2M (n=22)§	--	20	39	1.376 ± 0.046	0.780
Drosphila TEP1	AAF53490	17	34	1.379 ± 0.050	0.748
Limulus A2M	D83196	21	39	1.365 ± 0.045	0.745
Lamprey A2M	D13567	20	41	1.432 ± 0.047	0.726
Mouse A2M	Q61838	21	40	1.373 ± 0.045	0.747
HumanA2M	NM_000014	21	39	1.380 ± 0.045	0.741

Table 2: Pairwise amino acid comparisons of SeC3 vs TEP family members; only some shown. Individual pairwise alignments were performed for all calculations. Only one *C.elegans* TEP sequence was used because both are almost identical. Only available full-length sequences used. *Identities (I) are calculated as the percentage of identical amino acids per column/position in the alignments. **Similarity (S) was calculated as the percentage of identical plus similar residues, which are conservative substitutions (maintaining physiochemical properties) and were designated as KRH, DE, NQSTY, GAVLIFMW, C, and P. §Pairwise means were derived from averaging results from pairwise alignments of *n* number of sequences. Only a sub-sample of results are shown, table of all sequence calculations can be obtained from the authors. ¶ Poisson corrected distance scores (Mega 2, Kumar et al., 2001), ± standard error, were calculated for all pairwise comparisons. Proportion of difference (p-distance) calculations, uncorrected for multiple substitutions, were performed by the analytical method (Mega 2). Abreviations used: PWM, pairwise mean.

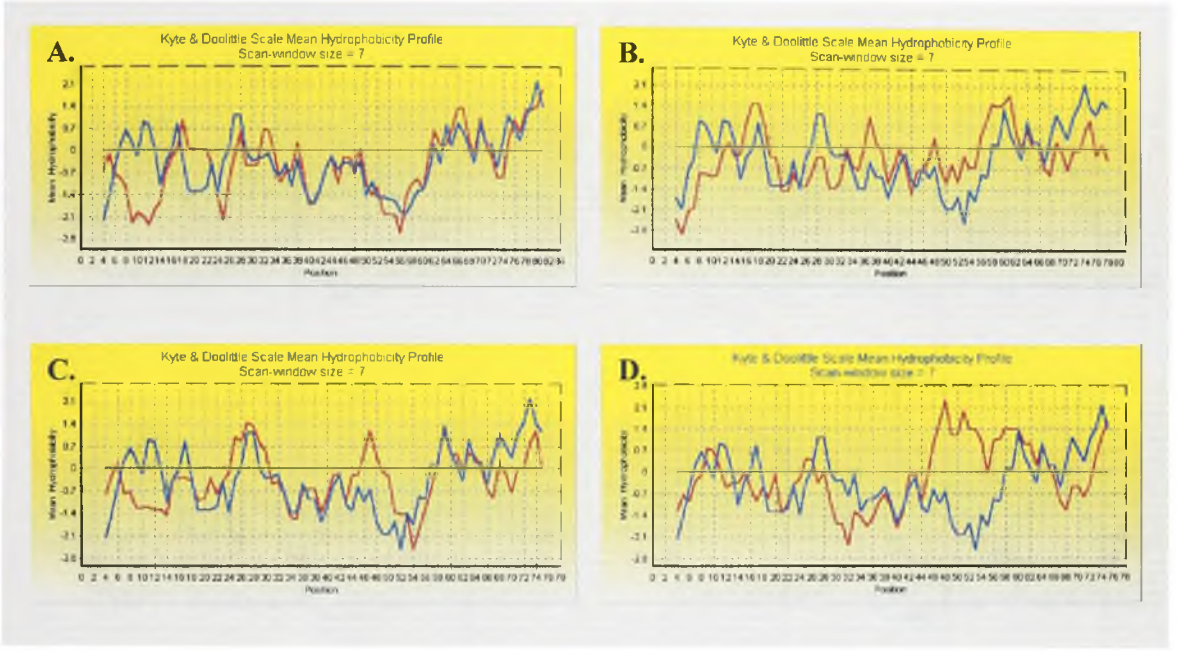


Figure 7a. Kyte and Doolittle hydrophobicity profiling of the corresponding region for the Factor B and H, and complement receptor I-III specific binding site on activated Human C3b; N-terminal 85 amino acids of the alpha chain. **A**. SeC3 and HuC3 in the N-terminal region of C3b alpha chain, major binding site ranges from position 20-84. Human sequence is red in all cases. **B.** SeC3 and corresponding region of HuC4A. **C.** SeC3 and corresponding region of HuC5. **D.** SeC3 and the corresponding region of HuA2M. Corresponding regions in C4A, C5, A2M and SeC3 were determined by alignment to the Human C3b alpha chain sequence. The sequence aligned to the appropriate region of HuC3b was determined to be the corresponding region.

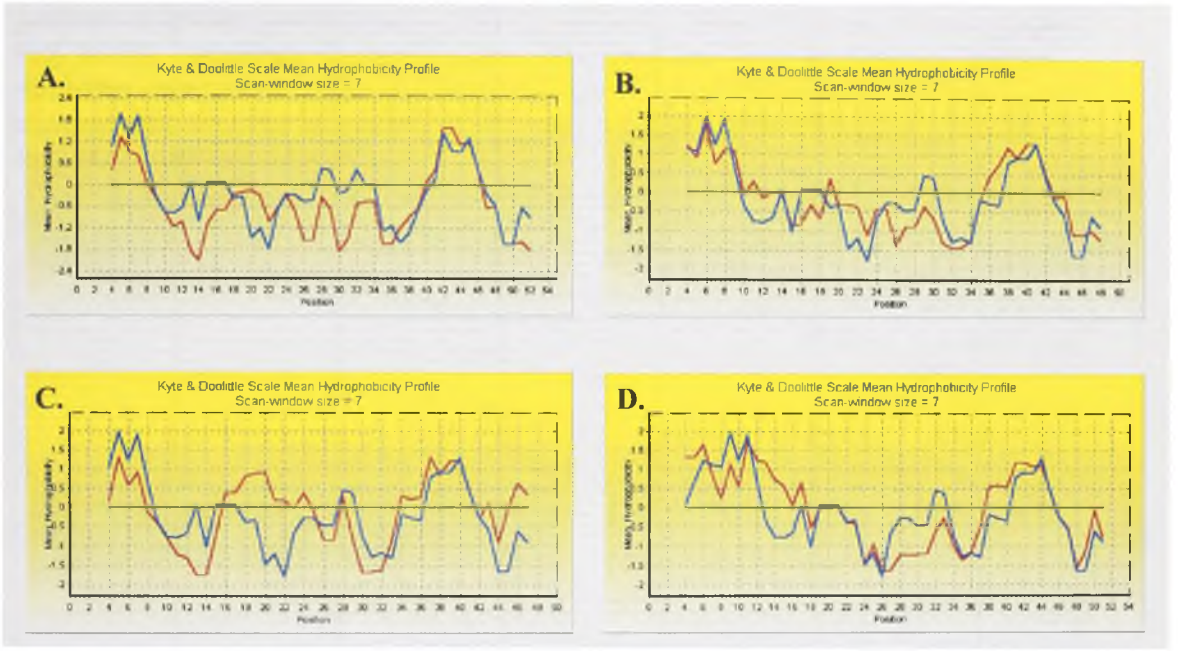
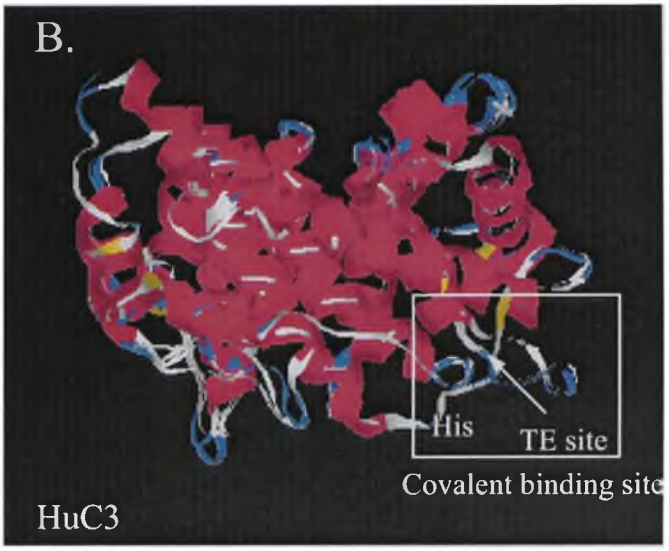
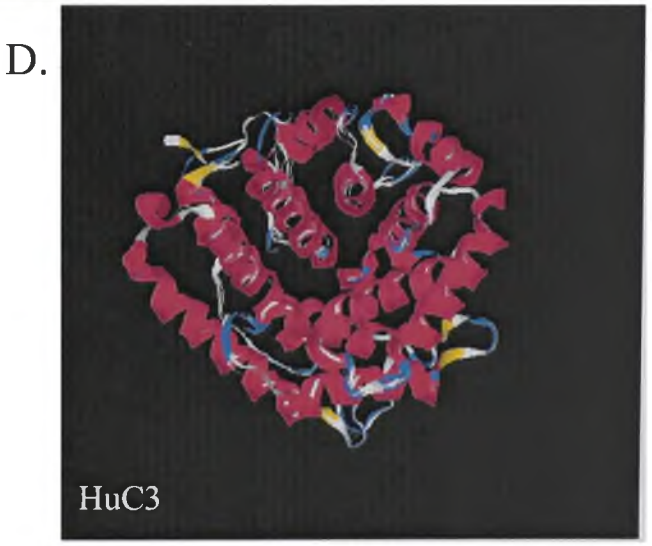


Figure 7b. Kyte and Doolittle hydrophobicity profiling of the corresponding region of the properdin binding site on activated Human C3b. Properdin binding is not present on human C4, C5, or A2M. **A.** SeC3 and HuC3 in the corresponding region of the properdin binding site. **B.** SeC3 and HuC4A in the corresponding region of the properdin binding site of Human C3b. **C.** SeC3 and HuC5 in the corresponding region of the properdin binding site of Human C3b. **D.** SeC3 and HuA2M in the corresponding region of the properdin binding site of Human C3b. The corresponding region for the properdin binding site of each other gene was determined by alignment to the Human C3 protein sequence. The sequence aligned to the properdin binding site was determined to be the corresponding region.

Figure 8a. Predicted three-dimensional structure of the C3d region of SeC3, using the crystallized human C3d molecule (Nagar et al., 1996) and the comparative modeling approach. The overall structure of the C3d is predicted to be conserved, as is the relative position of the buried thiolester-site and the catalytic histidine between SeC3d (A) and HuC3d (B). Secondary structure is also highly conserved (verified with comparative threading approaches; see Rost, 1096 and McGuffin et al., 2000) as can be seen by the characteristic complex helical backbone composed of two sets of six parallel helices in SeC3d (C) and HuC3d (D).

Figure 8.


B.
His
TE site
Covalent binding site
HuC3


D.
HuC3

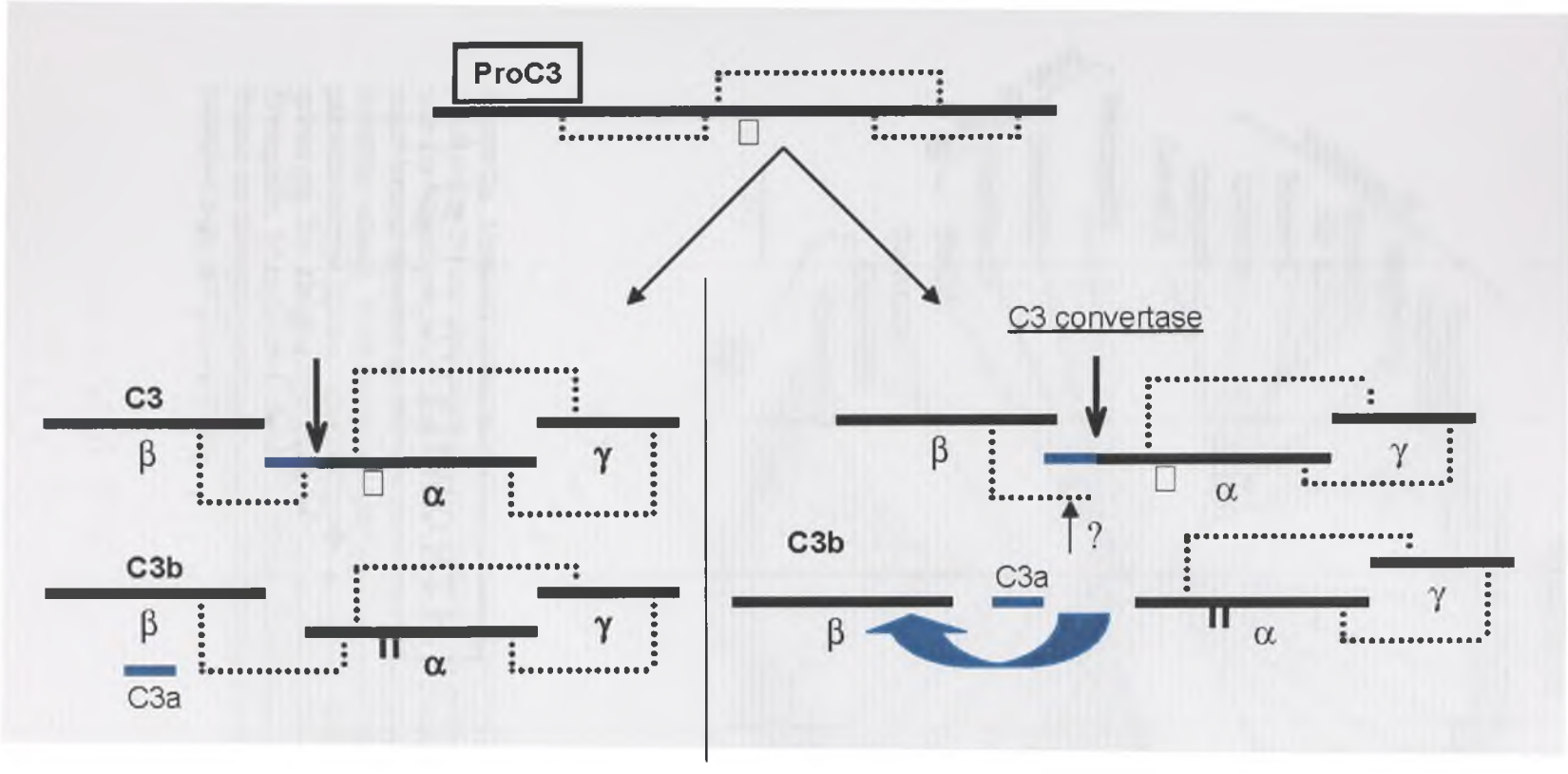


Figure 8b. Possible chain structure of SeC3 based on what is known from Human C3 and C4. In a three chain molecule (as in the case of mammalian C4), the structure on the left is what results after post-translational modification. In SeC3, unless the beta chain associates differently (in a novel fashion), it is possible that the chain comes off and performs a different function elsewhere. See text.

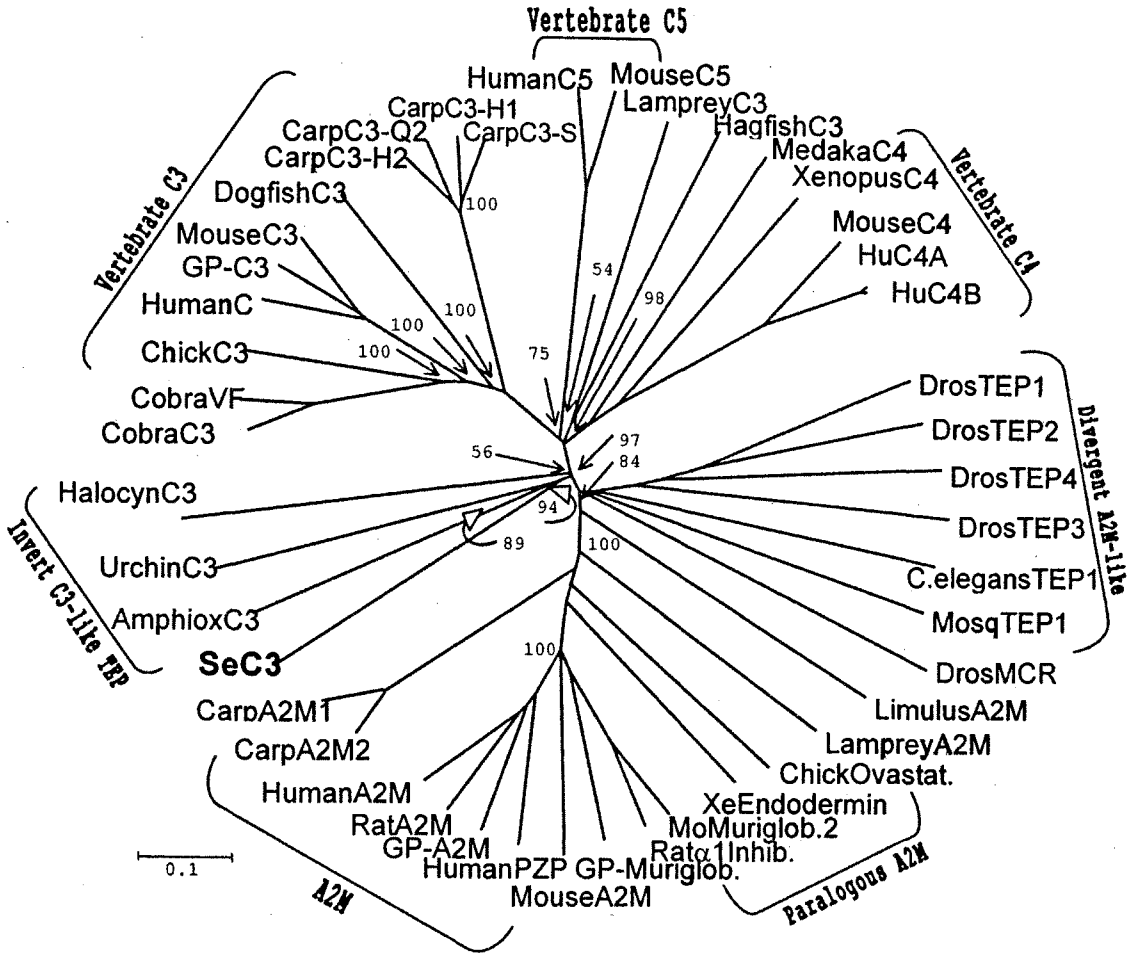


Figure 9a. Unrooted minimum-evolution bootstrapped distance tree (10,000 replicates) produced by the uncorrected proportion of differences method (p-distance). Tree produced with the Mega2 program. Major groups are labeled, and statistical support of some of the major internal branches are shown as percentage of bootstrap replicates (see Fig. 9b for other bootstrap values). N=45 sequences, globally aligned in Clustal X, and gaps treated in a pairwise deletion fashion. Abreviations are as follows: A2M- alpha 2-macroglobulin, GP- guinea pig, Xe- *Xenopus*, PZP-pregnancy zone protein, VF- cobra venom factor and Dros- *Drosophila*. Muriglobulin, Alpha-1-Inhibitor, Endodermin, Ovastatin and Pregnancy Zone Protein are divergent paralogous copies of A2M unique to vertebrates and CVF is a divergent paralgous copy of C3 in the Cobra.

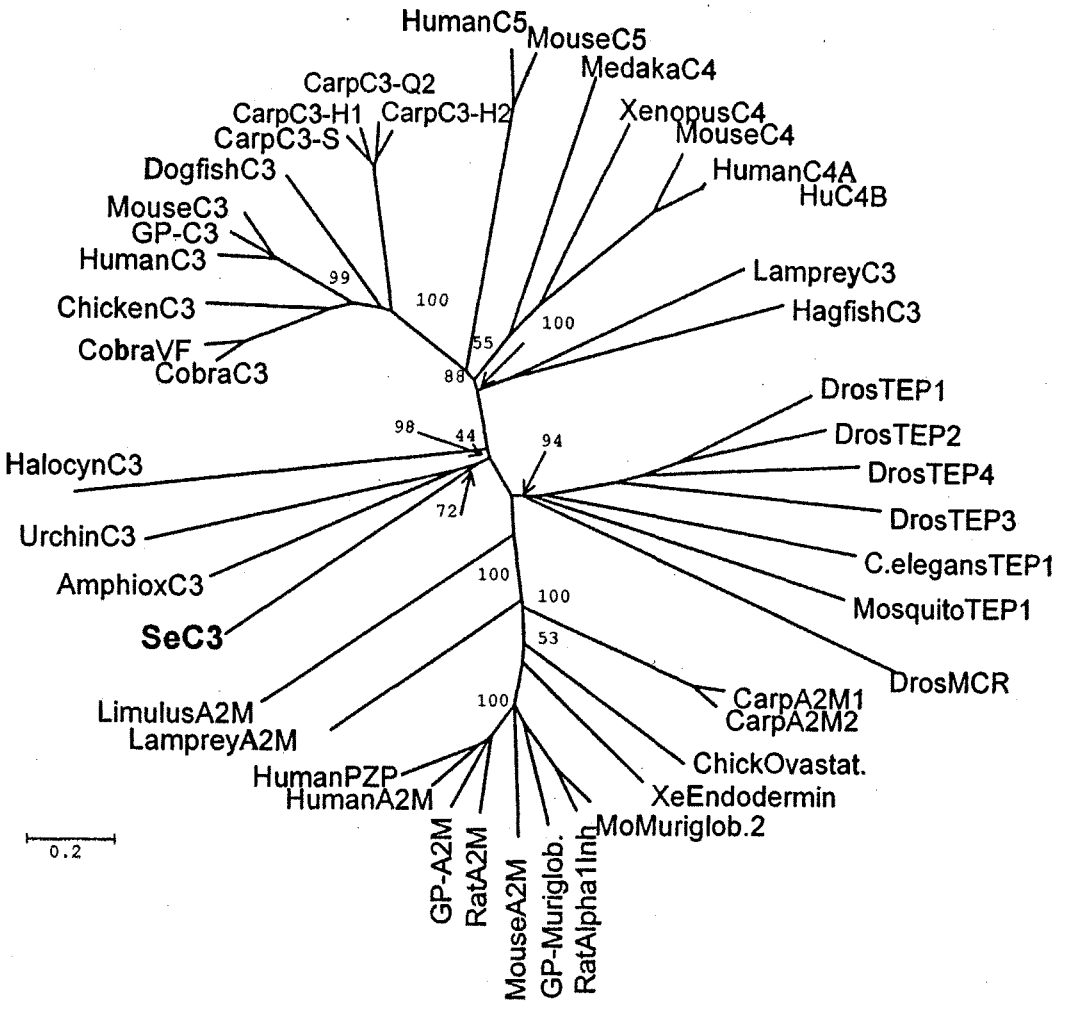


Figure 9b. Unrooted minimum-evolution bootstrapped distance tree (10,000 replicates) produced by the Poisson-correction distance. Tree produced with the Mega2 program. Statistical support of some of the major internal branches are shown as percentage of bootstrap replicates (see Fig. 10b for other bootstrap values). N=45 sequences, globally aligned in Clustal X, and gaps treated in a pairwise deletion fashion. Abreviations are a follows: A2M- alpha 2-macroglobulin, GP- guinea pig, Xe- xenopus, PZP-pregnancy zone protein, VF- cobra venom factor and Dros- *Drosophila*. Muriglobulin, Alpha-1-Inhibitor, Endodermin, Ovastatin and Pregnancy Zone Protein are divergent paralogous copies of A2M unique to vertebrates and CVF is a divergent paralgous copy of C3 in the Cobra.

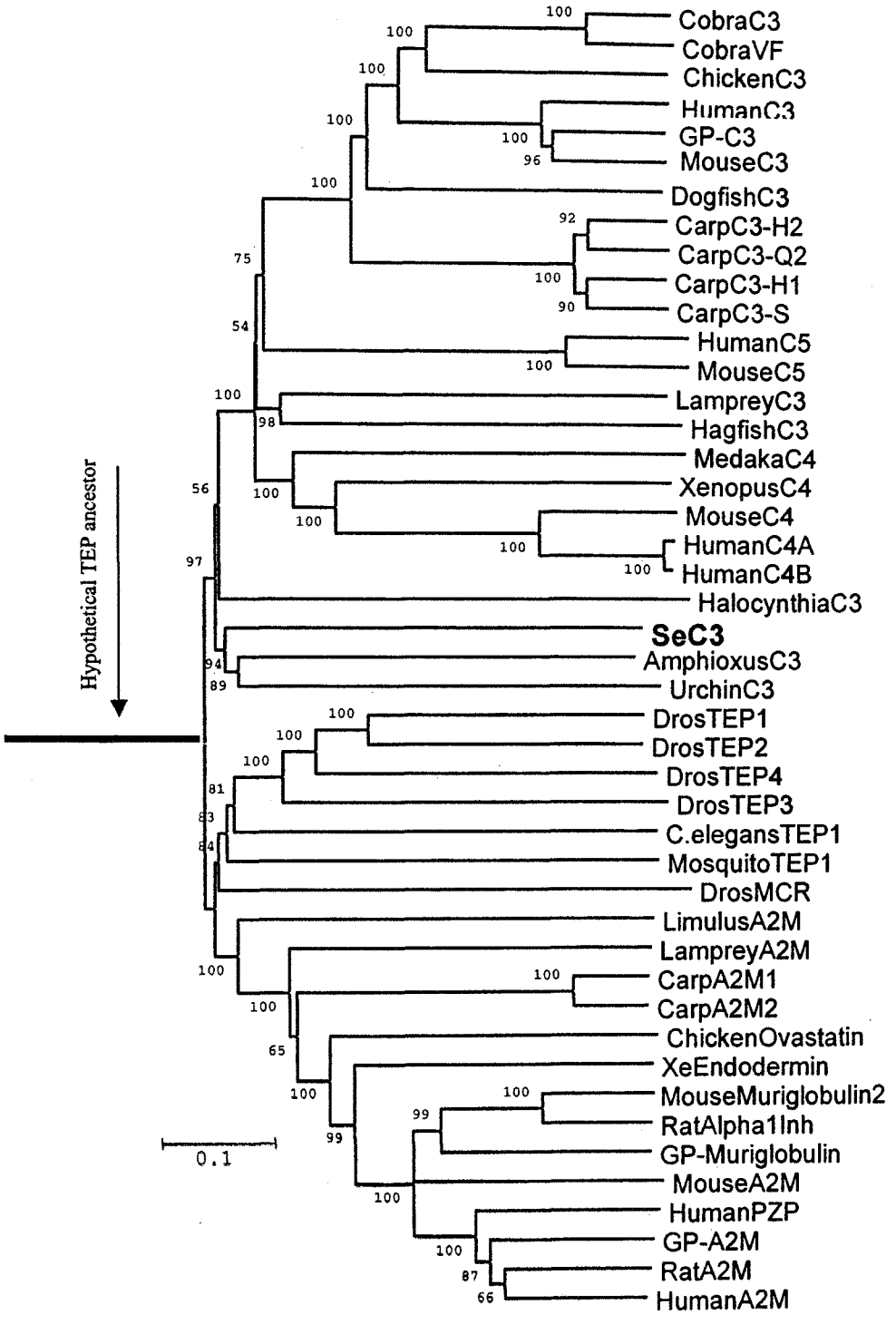


Figure 9c. Minimum evolution bootstrapped distance tree from Fig. 9a, rooted at the midpoint. Statistical support of the internal branches are shown as percentage of bootstrap replicates (10,000). N=45 sequences, globally aligned in Clustal X, and gaps treated in a pairwise deletion fashion.

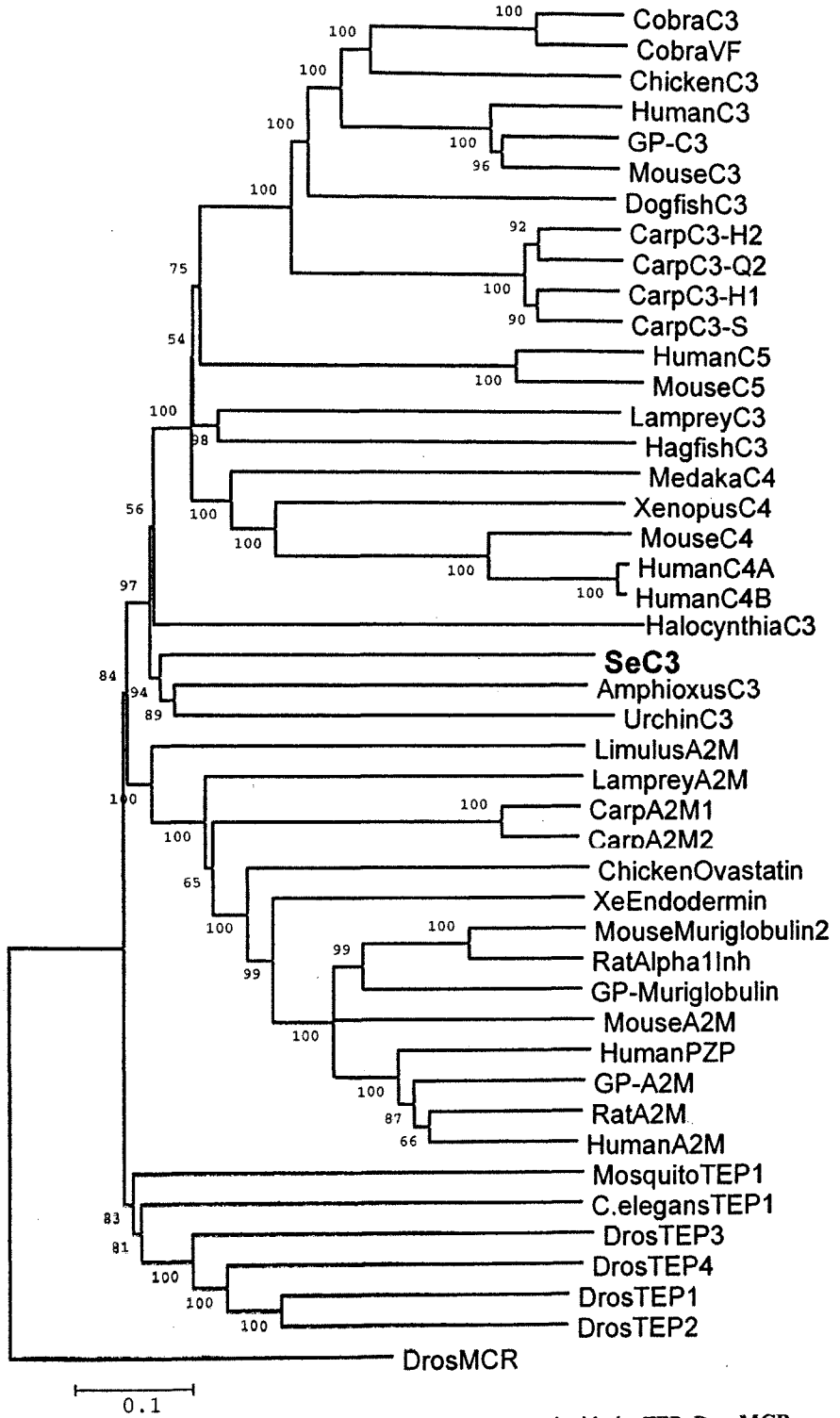


Figure 9d. Minimum evolution bootstrapped distance tree from Fig. 9a, rooted with the TEP, **DrosMCR**. Statistical support of the internal branches are shown as percentage of bootstrap replicates (10,000). N=45 sequences, globally aligned in Clustal X, and gaps treated in a pairwise deletion fashion.

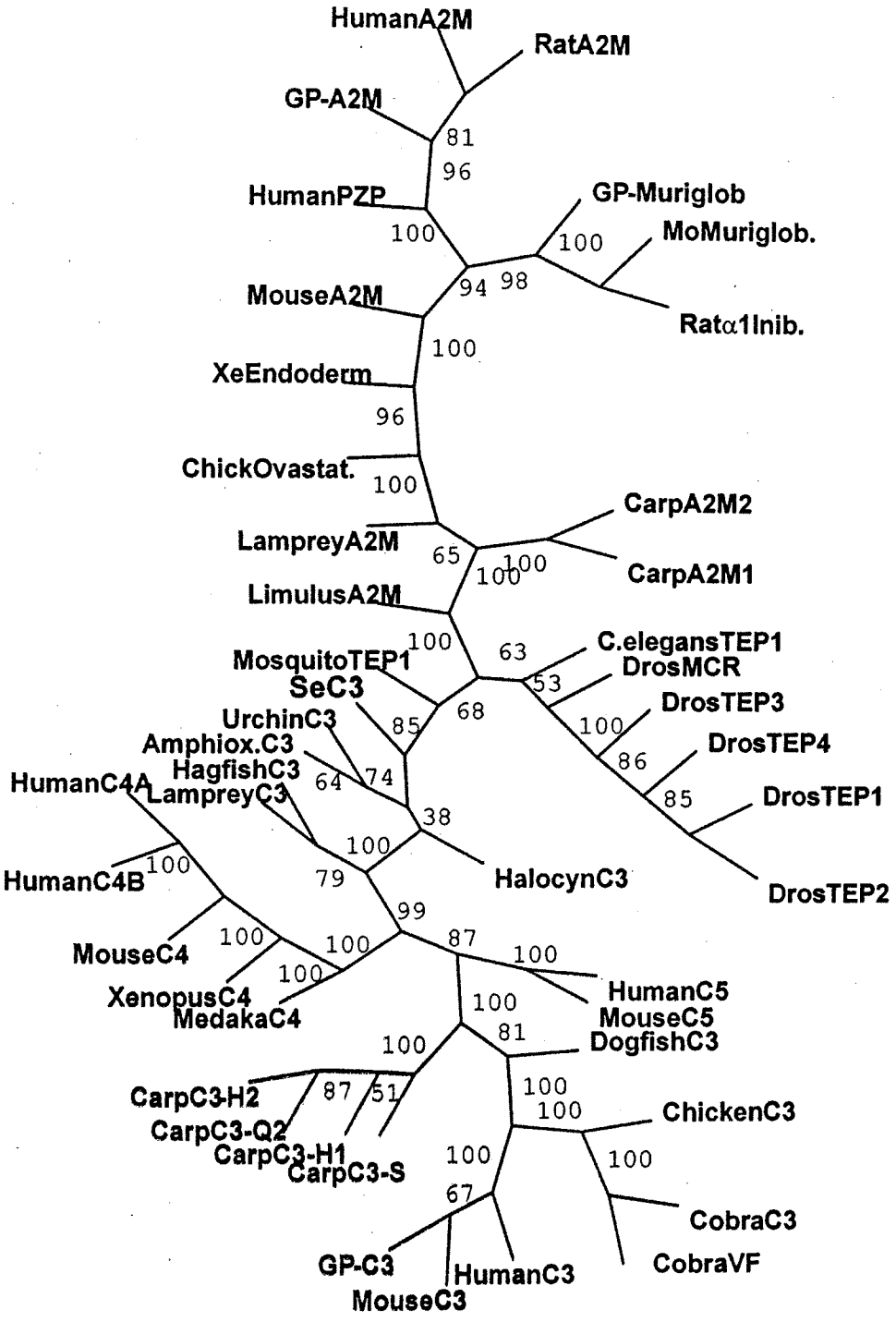


Figure 10. Unrooted Fitch-Margoliash least-squares bootstrap consensus distance tree. Statistical support of the internal branches are shown as percentage of bootstrap replicates (100 sets). N=45 sequences globally aligned in Clustal X.

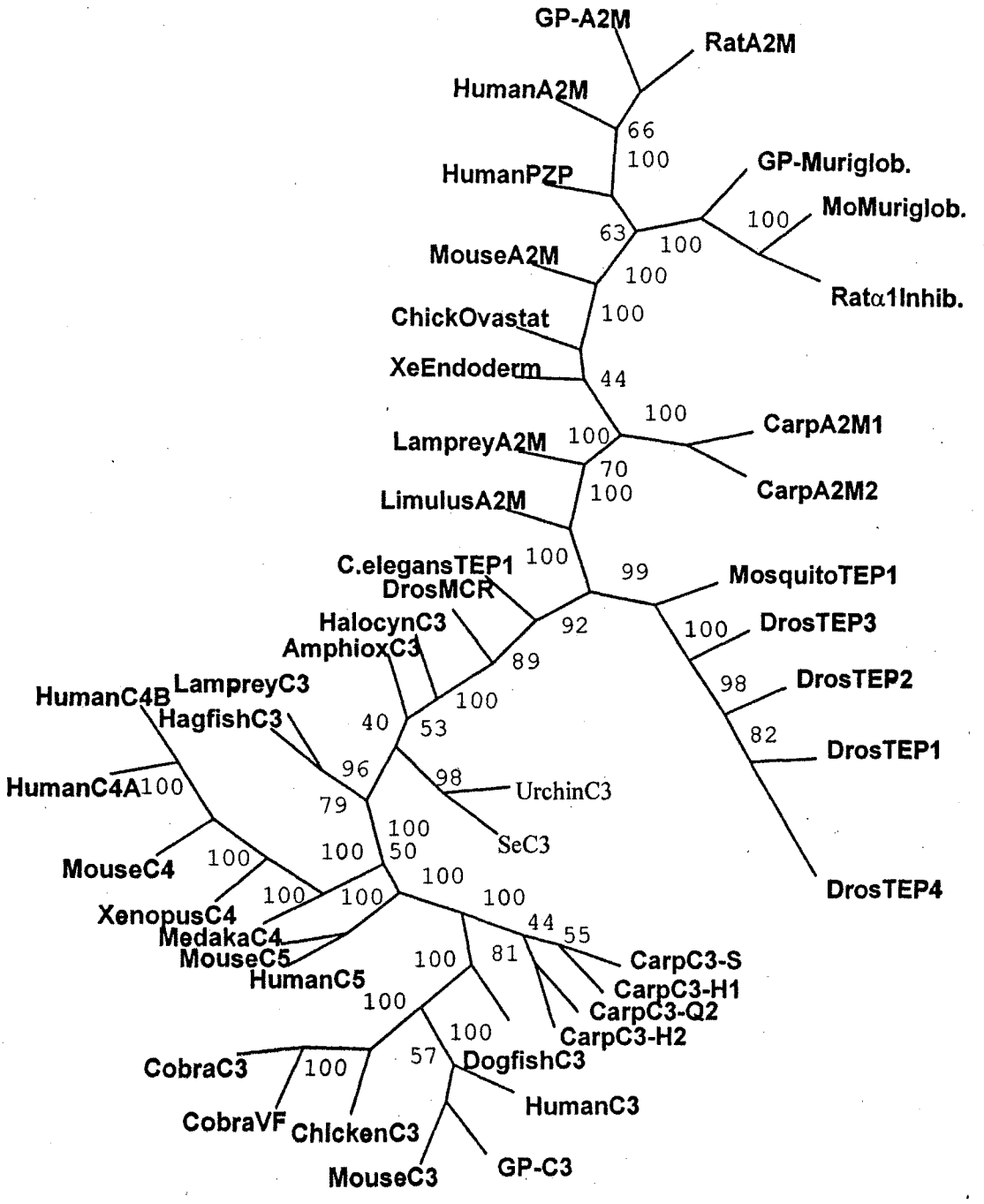


Figure 11. Unrooted Maximum Parsimony bootstrap consensus tree. Statistical support of the internal branches are shown as percentage of bootstrap replicates (500 sets). N=45 sequences, globally aligned in Clustal X.

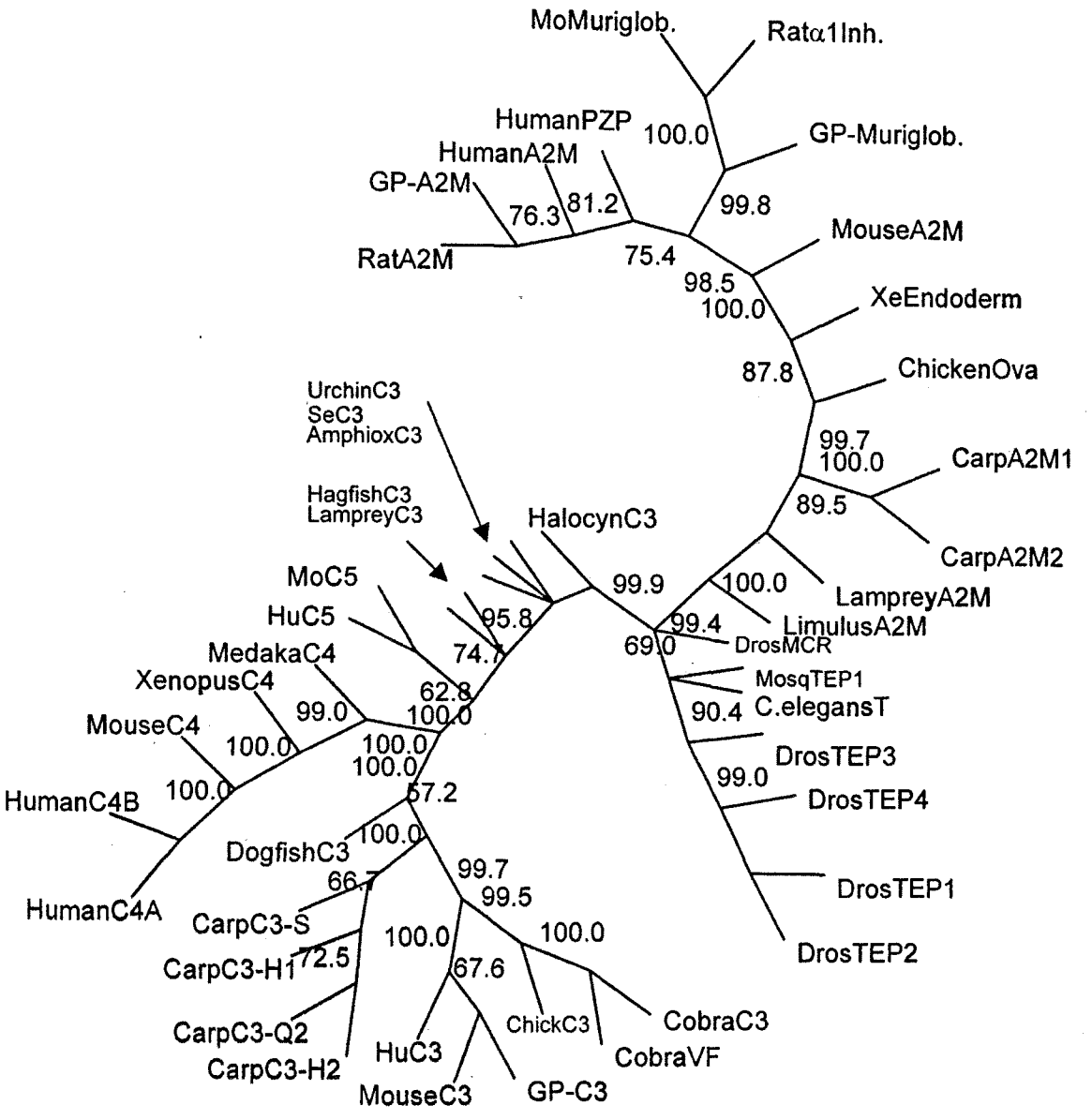


Figure 12. Unrooted Maximum Parsimony bootstrap consensus tree (100 replicates) generated in Paup *4.0b10. The data was analyzed with 25 random addition sequence replicates at each round, using steepest descent, and the tree-bisection-reconnection (TBR) branch swapping algorithm. N=45 sequences, same alignment using in figures 9-12. The topology of this tree is very similar to that seen using ME distance methods. Tree-length= 27679; CI=0.58; HI=0.42; RI=0.58 and RC=0.34.

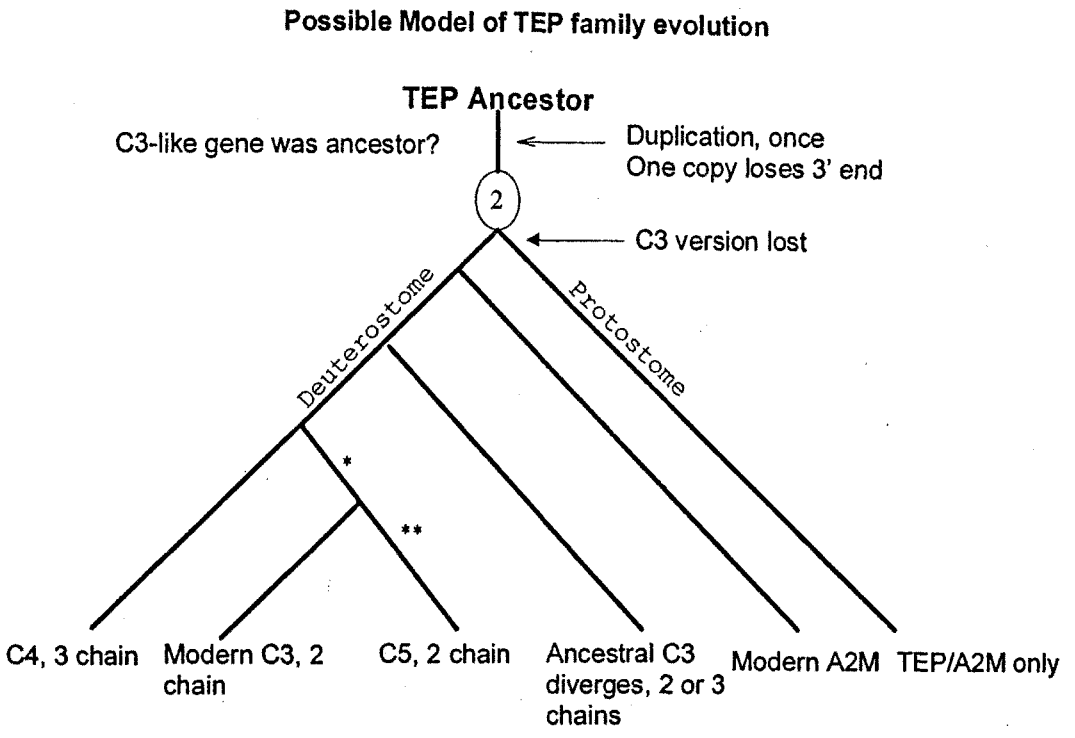


Figure 13. A new model proposed for the evolution of the TEP family based on the data presented in this study (see text). According to the model model, the ancestral TEP protein had C3-like structural characteristics and a duplication event to create a diverging paralog (A2M-like) occurred prior to the protostome-deuterostome split. Two copies existed before the split, while one of them lost the last 5-6 exons (C-terminal part of polypeptide). After the P-D divergence event, the C3-like ancestor was lost from the protostome lineage (probably through a chromosomal deletion event or a gene conversion event in the protostome ancestor). The second, truncated, TEP copy prevailed and continues to exist in modern protostomes as an A2M-like opsonin and non-specific protease inhibitor. In the deuterostome lineage, the truncated paralog became A2M-like as well, and the C3-like three chain TEP became the ancestral molecule to modern C3/C4/C5. The first duplication event split the ancestral C3 (still found today: coral, urchin, tunicate, and amphioxus) from the ancestral C3/C4/C5 molecule. A second duplication event split the three chain modern C4 from the C3/C5 ancestor. The C3/C5 ancestor lost the second cleavage site (*) before the duplication event. Duplication gives rise to two, two-chain proteins, one diverging into modern C3 and the other diverging into modern C5. Modern C5 loses its thiolester site (**) and is recruited into the terminal lytic pathway of complement and diverges further from C4 and C3.

Chapter 5

Molecular cloning of coral LMPX and implications for the evolution of the proteasome.

Abstract

Proteasomes are organelles partly responsible for the cellular metabolism of proteins. Vertebrates have adapted a unique, second "immunoproteasome" responsible for the generation of peptides presentable to the adaptive immune system. This immunoproteasome is assembled from paralogous copies of beta subunits belonging to the constitutive, housekeeping form. The assembled structure appears to be much more efficient in the generation of peptides for display on major histocompatibility complex (MHC) molecules. The point in phylogeny at which these paralogous subunits were established has been difficult to determine. To further understand the evolution of the immunoproteasomal subunits, a specific paralogous pair, LMP X/7, was pursued in a phylum whose divergence predates the phylogenetic divergence of protostomes and deuterostomes. This report describes an LMP X gene homologue in an endosymbiont-free gorgonian coral, *Swiftia exserta*. Phylogenetic analysis, along with hydrophobicity profiling of the N-terminal propeptide sequence of the coral LMP X and other invertebrate and agnathan sequences, suggests that more than one copy of LMP X may exist in invertebrates and that one of those copies may have evolved to function in a similar nature to its paralogous counterpart in vertebrate adaptive immunity, LMP7. In addition, this data may justify a re-investigation of jawless fish (agnathans) and all non-vertebrates for paralogous copies of LMPX.

Introduction

The 26S constitutive, ubiquitin, and ATP-dependent, proteasome is responsible for the generation of peptides in the cellular metabolism of proteins (Kloetzel, 2001). The 20S component is the inner, catalytic, core of this machinery and it consists of external alpha and internal beta subunits called low molecular mass polypeptides (LMP) or **p**roteasome **a**lpha/**b**eta subunits (PSMA/B). In eukayotes, there are multiple paralogous copies of each subunit, diverging from a common bacterial ancestry. For example, LMP X, Y, and Z are beta subunit components of the 20S, which appear to exist in a linked organization in the eukaryotic genome. In mammalian lymphocytes stimulated by INF-γ, LMP 7, 2, and MECL1 replace LMP X, Y, and Z, respectively, and form the immunoproteasome. Two of the subunits, LMP2 and LMP7 are linked to the MHC class II region in humans, mice and rats and have co-evolved with the adaptive immune system (Shastri *et al.*, 2002).

The alpha and beta subunits contain a N-terminal propeptide region (~60-80 amino acids), which in beta subunits is autocatalytically cleaved preceding their cooperative assembly within the proteasome. The precise function of the INF-γ-inducible proteins is not entirely clear, but as part of the immunoproteasome they appear to enhance MHC-specific, non-lysosomally derived, peptides for presentation to the adaptive immune system (Belich *et al.*, 1994; Driscoll & Finley, 1992; Tanaka & Kasahara, 1998). The propeptide of LMPX and 7 is particularly important for proper proteasome assembly, and appears to regulate which type of proteasome will assemble (Kingsbury *et al.*, 2000). LMP7 knockout mice have been shown to be inefficient in processing peptides for MHC class I display (Hehling *et al.*, 1994). Humans express two

versions of LMP7 proteins (e1 and e2), the result of alternative splicing of the first exon (responsible for the propeptide). The e1 isotype is rarely found in precipitated immunoproteasomes and suggests that beta subunit incorporation is dependent on proper propeptide properties (Fruh *et al.*, 1992; Griffin *et al.*, 1998). The properly assembled immunoproteasome is more efficient in generating a higher diversity of peptides for presentation to the immune system (Belich *et al.*, 1994; Driscoll & Finley, 1992; Griffin *et al.*, 1998; Kingsbury *et al.*, 2000; Tanaka & Kasahara, 1998).

Recent phylogenetic analysis of the proteasome genes has indicated that the alpha and beta subunits diverged prior to the divergence of eukaryotes from archaebacteria (Hughes, 1997). Utilizing fossil divergence times and nonsynonymous nucleotide substitution rates, Hughes (1997) estimated that the duplication event separating LMP X and LMP7 occurred about 600 million years ago. This time period is very close to (just after) the protostome-deuterostome split in phylogeny, and is much older than the divergence of jawed vertebrates (~450mya). This would suggest that an LMP 7-like gene should be present in the deuterostome invertebrates, such as echinoderms, uro- and hemi-chordates, and in jawless fish (agnathans).

In the attempt to understand the phylogeny of LMP X and 7, a representative gene homologue was pursued in a Cnidarian, which is a phylum that diverged prior to the protostome-deuterostome split. At least two beta subunit genes have been isolated from the gorgonian coral, *Swiftia exserta.* In this report, an LMP X homologue is described from this coral. Although LMP7 has not been described from agnathans, our preliminary phylogenetic analysis appears to suggest that at the time of hagfish and lamprey divergence, a newly evolving paralog had already been established. Our results warrant a

re-investigation into agnathans and deuterostome invertebrates where a paralogous copy may exist which has failed to become fixed as LMP7, since the functional constraints associated with vertebrate (adaptive) immunity are lacking.

Materials and Methods

Animals

The gorgonian soft coral, *Swiftia exserta* (Phylum Cnidaria, Class Anthozoa), was collected off the east coast of Florida (USA) and maintained in the laboratory as previously described (Salter-Cid & Bigger, 1991). RNA was extracted using TriReagent, under high-salt precipitation conditions as recommended by the manufacturer for tissues rich in polysaccharides and other contaminants (Molecular Research Center, Manufacturer's protocol, USA).

Reverse transcriptase-polymerase chain reaction amplification and cloning.

DNA-free RNA was reverse-transcribed (Superscript II, Invitrogen, CA, USA) using a modified oligo-dT primer and the Smart II oligo (Clonetech). The Smart cDNAs were amplified using PCR under slightly modified conditions recommended by the manufacturer. This method creates a library of enriched full-length cDNA-PCR products representing a good majority of expressed genes. Using degenerate PCR conditions (see Ch. 3 and Preston, 1996), previously described primers (Kandil *et al.*, 1996) and 0.5ul of the Smart cDNA library, a PCR product in the expected size range was purified and cloned (TOPO TA cloning kit, Invitrogen, Ca, USA). Sequence analysis of the product

indicated that it was highly similar to the corresponding region of LMPX from other animals.

Rapid amplification of cDNA ends (RACE)-PCR to clone full-length gene sequence

Gene-specific primers were designed for 5' and 3' RACE. For 3' RACE, gene-specific sense and oligo-dT primers were used along with the Smart cDNAs as template. For 5' RACE, gene-specific antisense primers were designed and used in conjunction with the Smart II oligo to the amplify 5'end of the gene. The final 5' RACE sequence was determined by using *classic* RACE (Zhang & Frohman, 1997) procedures, new antisense primers, and new RNA. Three overlapping race products produced the entire cDNA sequence for the coral LMPX homologue.

Nucleotide sequence analysis

Nucleotide sequence analysis was performed for all clones using the Big Dye dideoxynucleotide sequencing technology (version 2.0; PE-Biosystems) and an ABI 377 DNA sequencer (Perkin-Elmer). At least 10 clones were sequenced for each RACE product. The full-length cDNA sequence was produced by overlapping RACE products, assembled by eye, and analyzed using the Sequence Manipulation Suite (Stothard, 2000) and GeneDoc (Nicholas & Nicholas Jr., 1997). *Clustal X* (Thompson *et al.*, 1997) was used to align sequences for primary sequence confirmation (from multiple RACE clones) and for phylogenetic analysis. All alignments were produced under global alignment parameters.

Phylogenetic analysis

Aligned sequences, at both the amino acid and DNA level, were subjected to phylogenetic analysis using the Mega (v.2) program (Kumar *et al.*, 2001), the PHYLIP ver.3.5 package (Felsenstein, 1995), and PAUP* ver. 4.0b8 and 4.0b10 (Swofford, 1998). Full-length protein sequences, including LMPX and LMP7, were analyzed following global alignment in *Clustal X* (gap open penalty=20; gap extension penalty =0.40) (Thompson *et al.*, 1997). Most analyses, though, consisted of removing the N-terminal, non-conserved region, and only including the remaining ~200 aa corresponding to the highly conserved (ancestral) portion of the beta subunit (C-terminal region). This region is highly conserved, and appears to be evolving at a constant (albeit, slow) rate such that it may have clock-like behavior (equal substitution rates in all lineages). This hypothesis will be tested. The C-terminal region is also very easy to align at the DNA level and allows one to produce a codon-specific alignment and analysis.

Phylogenetic analysis of full-length protein sequences, and protein sequences with propeptides removed (C-terminal region only), was performed using the Minimum Evolution method (Kumar, 1996; Rzhetsky & Nei, 1993) and the Mega2 program (Kumar *et al.*, 2001). This was performed under a Poisson-corrected model (tree search by neighbor joining) of amino acid substitution and pairwise deletion of gaps. A Poisson-corrected distance corrects for multiple substitutions at each amino acid site, but assumes equal substitution rates among sites and equal amino acid frequencies (Nei & Kumar, 2000). Statistical significance of the internal nodes was tested using the bootstrap technique (Felsenstein, 1985) with 10,000 replications. Phylogenetic analysis

of the protein sequences (C-terminal region) was also performed under the maximum parsimony (MP) criteria (Fitch, 1971) using 500 bootstrap replicates and the tree-bisection-rearrangement (TBR) branch swapping algorithm with random addition of sequences and at least 10 repetitions at each round (PAUP 4.0* program).

The DNA sequences from LMP X and LMP 7 paralogous subunits were aligned in the corresponding C-terminal region (see Figure 2) consisting of 553 characters or 184 codons. The resulting alignment was analyzed in the program ModelTest version 3.06 (Posada & Crandall, 1998). This program uses maximum likelihood (ML) and likelihood ratio tests (LRT) to determine which phylogenetic model of nucleotide substitution best fits the data (56 models under consideration). Based on the LMP DNA sequence data, ModelTest recommended two models of nucleotide substitution. Hierarchical Likelihood Ratio Tests (hLRTs) indicates the data best fits (logL = -6820.13) the Tamura-Nei model of nucleotide substitution (Tamura & Nei, 1993) with a proportion of invariable sites (I) as 0.3161 following a gamma distribution (G) of 1.3192 (TrNef+I+G). This model assumes equal base frequencies and a substitution rate matrix [Rmat : (A-C)= 1.0000, (A-G)= 2.1079, (A-T)= 1.000, (C-G)= 1.000, (C-T)= 3.1156, (G-T)= 1.0000].

The Akaike Information Criterion (minimal theoretical information criterion, AIC) test, which does not require or assume nested models, indicates the data best fits (logL = -6793.39) the General Time Reversible (GTR) model of nucleotide substitution (Rodriguez *et al.*, 1990) with a proportion of invariable sites (I) as 0.3029 following a gamma distribution (G) of 1.1950 (GTR+I+G). The model appropriate to this data assumes unequal base frequencies [A=0.2349, C=0.2610, G=0.2948, T=0.2093] and a substitution rate matrix [Rmat: (A-C)= 2.0866, (A-G)= 2.7184, (A-T)= 2.1268, (C-G)=

0.7128, (C-T)= 5.0136, (G-T)= 1.0000]. For the purposes of this study, the GTR model was chosen and incorporated into the maximum likelihood criterion (Felsenstein, 1981; Huelsenbeck & Crandell, 1997) of the PAUP 4.0* program. The significance of the tree's branching pattern was determined by the bootstrap method (100 repetitions). The starting tree for this analysis was determined by stepwise addition and random addition of sequences (10 replicates, N=20 taxa) using the branch swapping algorithm of tree-bisection-reconnection (TBR).

The paralogous genes, LMP X and 7, are highly conserved and *appear* to be evolving at a slow but constant rate of nucleotide substitution. Because of this, the molecular clock hypothesis was tested using maximum likelihood (ML) and the likelihood ratio test (LRT) (Felsenstein, 1995; Huelsenbeck & Crandell, 1997; Huelsenbeck & Rannala, 1997). Under this method, the likelihood values of the ML trees are compared with and without the clock assumption. Then, the significance of this difference is tested using the LRT statistic and n-2 degrees of freedom (where n= the number of taxa).

Hydrophobicity Profiling

Hydrophobic and hydrophilic characteristics (hydropathy) of a protein can provide important information regarding its structural organization, its function in regards to substrate interaction, and/or its antigenic character (Hoop & Woods, 1981; Kyte & Doolittle, 1982). Hydrophobicity profiles were generated by the Kyte and Doolittle method (Kyte & Doolittle, 1982). Pairwise sequences to be compared were aligned in *Clustal X* and hydropathy profiles generated with the program, *BioEdit* (Hall, 1999).

This method was used in the attempt to determine the chemical nature of the N-terminal propeptide of LMPX and LMP7. The propeptide is the major region responsible for correct integration of beta subunits into either a ubiquitous- (LMPX) or immunoproteasome (LMP7).

Results

Coral LMPX sequence (SeLMPX)

SeLMPX was cloned, and the complete sequence determined, by assembling overlapping RACE products. These were generated after an initial degenerate RT-PCR product was isolated from normal, un-induced coral tissue. The initial PCR product was 174bp and was determined to be an LMPX-like cDNA. Three rounds of RACE were utilized to clone the full length cDNA sequence from overlapping PCR products. The SeLMPX sequence described in this report is just over 970bp and codes for a 268aa beta proteasome subunit (GenBank accession no. XXXX) (Fig. 1). This protein is composed of a 63aa N-terminal propeptide with the characteristic histidine (position 63) at the autocatalytic cleavage site, H/GTTT characteristic for both LMP X and 7. Sequence alignment with other beta proteasome subunits (Fig. 2) shows the SeLMPX sequence to contain those amino acids along the length of the 20S region which are a characteristic of LMPX or PSMB5 proteins. Identity and similarity calculations provide little information on the relatedness of these molecules (i.e., SeLMPX is 57% and 58% identical to Human LMPX and LMP7c, respectively). Highly conserved proteins such as these should be analyzed with molecular phylogenetics at both the amino acid *and* DNA level since the

DNA will reveal more historical information on the nature of evolutionary change (substitution).

Phylogenetic analysis

Phylogenetic analysis of proteins with and without the propeptide region

The PSMB5 and PSMB8 (LMP X and LMP7) proteins contain a common catalytic core which is highly conserved with bacterial 20S beta proteasome subunit proteins (Rivett, 1993; Zwickl *et al.*, 1992) and suggests a common origin for this 20S region. Full length protein alignments using 50 PSMB sequences (Fig 2) confirms that this catalytic core is common in all eukaryotic beta proteasome subunit genes (Hughes, 1997). Full-length protein sequences (with and without the N-terminal propeptides) of related beta subunits were analyzed by the minimum evolution method (Rzhetsky & Nei, 1993), under a Poisson-corrected distance model (Fig 3a and 3b). LMP X and 7 form sister clades with a monophyletic origin suggesting duplication from a common ancestor. YeastPRG1 clusters within LMPX and suggests that LMPX proteins existed prior to metazoan divergence. It is also of particular interest that in both trees, with and without the propeptide, the agnathan LMP X clusters with invertebrate LMP X rather than the vertebrate orthologs.

Other PSMB protein members were included in this alignment for the purposes of resolving the ingroup monophyletic relationship of LMP X and 7. These included members of the paralogous genes, LMP Y and 2; along with members of the LMP 3 or N3 genes. LMP Y and 2, like X and 7, are a pair of paralogous genes, which after γ-interferon immune stimulation, form the immunoproteasome by replacement of the

housekeeping forms, Y and X with 2 and 7, respectively (Monaco, 1992). The finding of LMP Y in yeast and plants suggests that, as with X and 7, Y resembles the ancestral form which existed prior to metazoan divergence. N3 subunits have not been studied extensively, as such only a few members have been characterized (Thomson & Rivett, 1993), but appears to be a beta subunit of the housekeeping proteasome. Until the cloning of the coral N3-homolog, SeN3 (not a topic of this chapter), only one invertebrate N3 sequence was available (*Drosophila* N3) which has not been studied or characterized (*Drosophila* genomic sequence data). Extensive phylogenetic analysis of alpha and beta proteasome subunits has shown that the divergence of these two gene families occurred prior to the split between eukaryotes and archaebacteria (Hughes, 1997).

In phylogenetic analysis of LMP X and 7 (using only C-terminal sequence), the yeastPRG1 is the most divergent (over 2x the branch length of the 2nd longest branch, sponge LMPX) (Fig. 3c). To prevent biases associated with long branches, the yeast sequence was removed from the subsequent DNA analysis (see below). It is also of special interest that in Fig. 3c, the agnathan LMPX sequences cluster as an outgroup to LMP7 with low (54%) bootstrap support. The agnathan sequences have been found by other authors (Takezaki *et al.*, 2002) to cluster as an outgroup to LMP7 and have suggested that the functional constraints on agnathan LMPX is similar to that of vertebrate LMP7. This hypothesis, though, has not been tested at the functional level.

Furthermore, when the other beta subunit genes (LMP Y, 2, and N3) are not used as outgroups LMPX no longer forms a (weakly supported) monophyletic clade as in Fig. 3a and 3b (Fig. 3c). This is confirmed by character-based heuristic searches using maximum parsimony (Fig. 3d). Therefore, without the use of outgroups, the branching

patterns of LMP X appear as paraphyletic. Inasmuch, if the two *Drosophila* sequences, DrosProBeta5 and DrosB5, are to be considered PSMB5 (LMPX-like) sequences (Fig. 3a & 3b), then indeed LMPX is paraphyletic. The true nature of these two *Drosophila* proteins remains to be revealed. These *Drosophila* sequences were chosen as outgroups to LMPX and 7, and were used to root the trees in Fig. 3c and 3d.

Phylogenetic analysis of DNA sequence from the C-terminal region

Phylogenetic analysis was performed with 184 codons of the C-terminal portion of the PSBM5/8 genes (N=20) in the attempt to better understand the evolution of these paralogous genes (Fig.4a & b). Both models (TrNef+I+G and GTR+I+G) of DNA substitution were applied to maximum likelihood (ML) bootstrap analysis using the PAUP 4.0* program as described above. Because of the sample size, only 100 bootstrap repetitions were performed. The unrooted ML trees that resulted (Fig. 4a & 4b) suggest that the more complex GTR model may fit this data more appropriately since it appears to better resolve the invertebrate LMP X clade (Fig. 4a, invertebrate LMP X produces an unresolved polytomy with 89% support). Interestingly, the finding that the tunicate LMPX clusters, for example, with coral LMPX than with amphioxus LMP X may suggest that the LMP X proteins being compared in the invertebrates are not true orthologous copies (which may be case in animals with two LMP X genes). Alternatively, this may be a simple lack of resolution at this area of the tree.

The GTR model suggests that there is rate heterogeneity (following a gamma distribution) in among-site substitution rates and unequal nucleotide frequencies. The major difference in the way the paralogous genes are evolving appears to lie with LMPX

(see Fig 4c as example), since in trees from both models the topology of LMP7 is essentially the same. Given the among-site rate heterogeneity, is it constant among all lineages of the tree?

Preliminary tests of the molecular clock hypothesis

The molecular clock hypothesis (H_0) was tested in the paralogous genes, LMP X and 7 using Felsenstein's F84 model (modified F81 by allowing unequal rates and transition/transversion ratio biases) of the PHYLIP package (Felsenstein, 1981; Felsenstein, 1995). This model assumes one rate of substitution with unequal base frequencies and three free parameters for base frequencies. The LRT can be performed to test the clock as long as the only variation between the two ML results is the presence or absence of clock-like behavior (Felsenstein, 1995; Huelsenbeck & Crandell, 1997).

The exact same data used above was tested under this model so that the likelihood scores were compared between the two resulting trees representing clock and no-clock behavior. Using N=20 taxa of both paralogous genes (X and 7), the clock hypothesis is rejected. The resulting trees, with and without clock assumption are significantly different (see table 2 for tree scores; trees not shown). The data set was separated, so that each set consisted of orthologous members of that gene, and tested. The LRTs (see table 2) indicate that LMP7 appears to follow clock-like behavior (null hypothesis cannot be rejected), while LMPX does not. It is of particular interest that if the agnathan LMP X genes are added to the ML analysis of LMP7, the clock-like behavior is unchanged. This does not hold if any other LMP X sequence (for example, adding the amphioxus LMPX) is added to the data set. These preliminary tests appear to suggest rate heterogeneity

among the LMPX orthologs. Because of this possibility, both data sets were tested in the ModelTest program to determine which model best fits each set of genes.

Results from ModelTest suggest that LMP7 best fits the TrNef+I+G (logL= -3470.46) and SYM+I+G (logL= -3469.32) models (SYM = symmetrical model, Zharkikh & Li, 1993). Both models are under equal base frequencies and similar I and G values, with each assuming a unique substitution rate matrix. LMPX best fits the TrNef+I+G (logL= -3619.49; with equal base frequencies) and the GTR +I+G (logL= -3597.15; with unequal base frequencies) models of nucleotide substitution.

The molecular clock was retested using the SYM+I+G model for LMP7 and the GTR+I+G model for LMPX, along with the previously described GTR+I+G model for both LMPX and 7 together. As with the F84 model test above, the LRT was performed on the results of ML analysis with and without the enforced clock assumption (in PAUP* 4.0b8 program). The results of this analysis (table 2b) supports those found using the F84 model (table 2a) where clock-like behavior only appears in the LMP 7 lineages.

Hydrophobicity profiling

Hydropathy profiling (Kyte & Doolittle, 1982) was performed on the N-terminal propeptides of the LMP X and 7 proteins. The results indicate that human LMP7e1 propeptide shares hydropathy characteristics with LMPX (see Fig. 5), and this may influence its inability to properly incorporate into the immunoproteasome. Hydropathy profiles of pairwise aligned human LMPX, LMP7 e1, and e2 propeptides suggest that a critical region exists between aligned position (residue) 16 and 33 (see Fig 5, lower panel). Biochemical and functional analyses indicate that this is a major region

associated with proper LMP7e2 assembly into the immunoproteasome (Kingsbury *et al.*, 2000).

The possibility that an LMPX paralog may have served LMP7-like functions prior to its divergence as LMP7 in later vertebrates warrants further investigation. Hydrophobicity profiling of the N-terminal propeptide was performed with agnathan LMPX and human LMPX and both human LMP7e1 and e2 (see Fig. 6). The propeptide of lamprey LMPX (panels a-c) shares hydropathy profiles (position 16 - 33) similar to human LMP7e2 but not LMP7e1 or LMPX. This implies that the lamprey LMPX may share the functional requirements for incorporation into an immunoproteasome since its propeptide appears to share equivalent properties with LMP7e2. This incorporation, though, is unrelated to the subunit's catalytic capability since LMP7 contains 20-25 unique amino acids residues spanning its catalytic core (see Fig. 2). A similar pattern was seen with hagfish LMPX (panels d-f).

This technique was further applied to invertebrate LMPX propeptides to address the same question, and attempt to uncover a phylogenetic divergence period for acquiring LMP7-like functionality. It appears that *Botryllus* (tunicate) LMPX (Fig. 7) shares overlapping hydrophobicity in the corresponding region of human LMP7e2 (position 16-33), which is a property not shared with human LMPX. The coral LMPX described in this report instead shares hydropathy characteristics with human LMPX and LMP7e1 in the corresponding region (Fig. 8). Unexpectedly, though, it was found that sponge LMPX (Fig. 9) shares a similar overall profile to agnathan LMPX and human LMP7e2. Having the entire genome sequence available (GenBank database; http://www.ncbi.nlm.nih.gov/), we see than in *Drosophila*, which also has two PSMB5

(LMPX-like) and PSMB2 (LMPY-like, not discussed here) molecules, a similar pattern is evident. One PSMB5 shares hydrophobicity with human LMP7e2 propeptide while the other appears divergent from both X and 7 (data not shown). This data may suggest that two types or copies of PSMB5 (LMPX-like) genes have existed since early diploblastic animals, with each acquiring separate N-terminal propeptide characteristics. These findings also suggest that the LMP7 propeptide has ancient functional characteristics.

Discussion

LMP X and 7 (PSMB 5 and 8) are paralogous genes which diverged from the common ancestral beta subunit (20S-like) of the ubiquitous housekeeping proteasome complex. Orthologous copies of X and 7 have remained quite conserved, primarily over the length of the 20S core/catalytic region. The two paralogous genes differ via several amino acid positions evenly distributed over the length of the 20S region (see Fig. 2) and extensively at the first 60-80 aa of the N-terminal (propeptide) region. The propeptide region appears to have been gained at some point after the LMPX divergence from the ancestral beta subunit. This acquisition occurred before the divergence of metazoans because the yeast and plant beta proteasome subunits are PSMB5-like and contain an N-terminal peptide of similar length (Hughes, 1997).

This report describes an LMPX-like homologue (SeLMPX) from a Cnidarian, the endosymbiont-free gorgonian coral, *Swiftia exserta*. Alignment of SeLMPX with other PSMB5 and PSMB8 proteins (LMP X and 7) indicates the presence of conserved residues along the length of the C-terminal catalytic core (20S region) which are unique to LMPX. The correspondingly different residues, along with the unique properties of its

N-terminal propeptide, are what make LMP7 functionally restricted to immunoproteasomes and more effective at processing antigen for display on MHC class I proteins (Fehling *et al.*, 1994; Griffin *et al.*, 1998; Kingsbury *et al.*, 2000).

Phylogenetic analysis places SeLMPX among other LMPX genes (PSMB5). Previous studies did not include invertebrate LMPX sequences in their analyses because they were not available (Hughes, 1997; Kandil *et al.*, 1996) or they were omitted (Takezaki *et al.*, 2002) because they produced "incorrect topologies". This is the first attempt to produce PSMB5/8 phylogenies in the presence of invertebrate LMPX subunit proteins (See table 1; Sponge, Coral, Tunicate, Amphioxus, and Fruit Fly). Invertebrate PSMB5 (LMPX) proteins mostly produce unresolved polytomies (Fig. 3d & 3e), while the correct position of the agnathan orthologs has typically been difficult to determine (Kandil *et al.*, 1996; Takezaki *et al.*, 2002). Maximum likelihood analysis at the DNA level (Fig. 4a-c) supports monophyletic origins for LMP X with difficult to resolve topologies for the invertebrate LMP X genes (Fig. 4a & 4b). The polytomy clustering of the invertebrate LMP X genes may suggest the comparison of paralogous, rather than orthologous, copies of LMP X (resulting in soft polytomies).

Based on the phylogenetic analysis at both the protein and nucleotide level, one may predict the existence of more than one paralogous copy of LMPX (PSMB5) in some invertebrates. This is supported by the finding of two paralogous copies in *Drosophila*, while partial gene sequence data suggests a similar senario in corals (Dishaw, unpublished observations). A valid concern arises, since producing phylogenies using true orthologous copies of a gene is now more difficult. The hypothesis that more than one LMPX gene exists is further supported by the finding that some invertebrate PSMB5

propeptides share hydrophobicity profiles with human LMP7e2 propeptides while others (like SeLMPX described here) more closely resemble human LMPX. At least one gene duplication event, therefore, appears to have occurred prior to the divergence of LMP7. More data from various phyla are now necessary to determine if duplication has been lineage-specific.

The point at which LMP7 diverged in phylogeny has been very unclear (Hughes, 1997). Although the divergence event has been estimated to be at about 600mya (Hughes, 1997), no LMP7-like sequences have been found in agnathans or in deuterostome invertebrates. It has been suggested that the shark LMP7b sequence is a precursor of bonafide LMP7 (functionally similar to mammalian LMP7) as it contains several amino acid residues unique to LMPX (Hughes, 1997; Kandil *et al.*, 1996). Phylogenetic analysis suggests that the shark LMP7b-like paralog was established much earlier than other vertebrate LMP7 genes and that agnathan LMPX may be tightly linked to this event (Fig. 3c as example). Therefore, a second paralogous gene may exist in agnathans but not appear as LMP7 because the functional constraints associated with adaptive immunity and/or immunoproteasomes had not evolved.

Our phylogenetic results, and the work of others (Hughes, 1997; Kandil *et al.*, 1996; Takezaki *et al.*, 2002), has suggested that the agnathan PSMB5 position in phylogeny is roughly unresolved. The unique properties of shark LMP7b further suggests that this gene may have existed before the PSMB8 (LMP7) divergence in jawed vertebrates. It has been proposed (Hughes, 1994; Jensen, 1976; Orgel, 1977) that prior to gene duplication, proteins may be serving more than one function. After a duplication event, one of the paralogs will rapidly evolve and may become specialized and fixed to

fill that second function. The proteasome family of proteins appears to be a good representative of this scenario (Hughes, 1997)

The N-terminal propeptide of LMPX and 7 determine the protein's incorporation into, and the functional assembly of, the proper proteasome. Specifically, the propeptide of LMP7 is required for LMP7's incorporation into the immunoproteasome (Kingsbury *et al.*, 2000). This is supported by data from LMP7 knockout mice which do not have properly assembled immunoproteasomes because LMPX can not incorporate in place of (substitute) LMP 7 (Fehling *et al.*, 1994; Griffin *et al.*, 1998). The propeptide of LMP7 has become specialized for this function (Griffin *et al.*, 1998; Kingsbury *et al.*, 2000). Substitution of LMP7 propeptide onto LMPX (with propeptide deletion) results in the incorporation of LMPX into LMP2-positive immunoproteasomes (Kingsbury *et al.*, 2000) but does not produce functionally equivalent structures. In humans, alternative splicing of two copies of the first exon produce a catalytically active LMP7e2 and one that is not incorporated into the immunoproteasome (LMP7e1) (Fruh *et al.*, 1992; Kingsbury *et al.*, 2000).

Hydropathy profiling indicates that some invertebrates contain PSMB5 proteins with N-terminal propeptides that specifically share hydrophobicity profiles or patterns with the propeptide of LMP7e2 (functional version of human LMP7). Invertebrates that contain two copies of PSMB5 appear to have one that shares hydropathy characteristics with LMP7e2 propeptide. These results suggest a need to reinvestigate invertebrates and agnathans for second copies of LMPX (which may not be detectable if functionally down-regulated). This data proposes that because of the divergent but unique characteristics of the propeptide, LMPX-like subunits (particularly in agnathans) may

share some functional characteristics with vertebrate LMP7, and that in invertebrates, a second copy of LMPX (PSMB5) may exist. Functional studies of LMPX in these animals, as well as shark LMP7b, should provide intriguing results.

From an immunological perspective, the proteasome evolved independently of adaptive immunity. The abrupt appearance of adaptive immunity (Agrawal *et al.*, 1998; Marchalonis & Schluter, 1998) via rearranging antigen receptors and the MHC was facilitated by the recruitment of pre-existing and independently evolving proteasomal subunits that when assembled into what would become the "immunoproteasome" were more efficient at generating peptide antigens for display on MHC (Tanaka & Kasahara, 1998).

PSMB5 (LMPX) gene duplication appears to have occurred early in phylogeny. This event(s) occurred much earlier than the PSMB8 (LMP7) divergence and recruitment into vertebrate immunoproteasomal pathways. Hence, prior to the period in which a copy was recruited into the immunoproteasome, the PSMB5-paralogs may have been fulfilling PSMB8-like functions (i.e., processing peptides for immunity). Therefore, until the point in phylogeny where the functional constraints of adaptive immune pathways modified the newly recruited paralog, attempting to characterize it as a bonafide LMP7 may be difficult. The shark LMP7b is a molecular "fossil" since it may have preserved characteristics of the ancestral form. This mode of gene/protein sharing, also seen in other beta subunit paralogs, has been proposed to have been common in the evolution of the proteasome (see Hughes, 1994; Hughes, 1997).

Furthermore, in studies to determine the minimal requirements for MHC antigen processing and display pathways, some elegant experiments have been performed in

insect cell lines (*Drosophila*) (Deng *et al.*, 1998; Schoenhals *et al.*, 1999). It was originally shown that, following transfection, insect cell-expressed recombinant human MHC class I molecules are displayed empty (without bound peptide) (Jackson *et al.*, 1992). Recent work now indicates that this occurs because of the absence of a tapasin-like gene in arthropods (Schoenhals *et al.*, 1999). Recombinant expression of tapasin in these cell lines retains MHC molecules until they are loaded with antigen. Cell-surface expression of MHC molecules in these cell lines are bound with intracellular peptide (Schoenhals *et al.*, 1999).

In these studies, MHC molecules can be precipitated with bound peptide of intracellular proteasomal origin even though insects lack the "immunoproteasome"-specific subunit genes seen in vertebrates. This further suggests that the constitutive house-keeping proteasomal subunits are minimally sufficient for generation of peptides that can be "displayed" to the immune system. The evolutionary recruitment of the INFγ inducible forms (such as LMP7, LMP2, and MECL) into the antigen processing pathway appears to increase efficiency of generating a more diverse repertoire of peptides (Griffin *et al.*, 1998). LMP7 knockout mice demonstrate decreased efficiency of class I presentation, but not complete abolishment (Fehling *et al.*, 1994). These examples further support that hypothesis that constitutive proteasomal subunits can process peptides for immune display and that prior to PSMB8 divergence, the PSMB5 paralogs were fulfilling PSMB8-like functions. Our preliminary observations suggest that invertebrates and agnathans should be further investigated for other paralogous copies of PSMB5-like genes.

References

Agrawal A., Eastman Q. M., and Schatz D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394:** 744-751.

Belich M. P., Glynne R. J., Senger G., Sheer D., and Trowsdale J. (1994). Proteasome components with reciprocal expression to that of the MHC-encoded LMP proteins. *Current Opinion in Biology* **4:** 769-776.

Deng Y., Gibbs J., Bacik I., Porgador A., Copeman J., Lehner P., Ortmann B., Cresswll P., Bennink J. R., and Yewdell J. W. (1998). Assembly of MHC Class I molecules with biosynthesized endoplasmic reticulum-targeted peptides is ineffecient in insect cells and can be enhanced by protease inhibitors. *Journal of Immunology* **161:** 1677-1685.

Driscoll J., and Finley D. (1992). A controlled breakdown: antigen processing and turnover of viral proteins. *Cell* **68:** 823-825.

Fehling H. J., Swat W., Laplace C., Kuhn R., Rajewsky K., Muller U., and von Boehmer H. (1994). MHC class I expression in mice lacking the proteasome subunit LMP7. *Science* **265:** 1234-1237.

Felsenstein J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17:** 368-376.

Felsenstein J. (1985). Confidence limits on phylogenetics: an approach using the bootstrap. *Evolution* **39:** 783-791.

Felsenstein J. (1995). PHYLIP: Phylogeny Inference Package, version 3.57c.

Fitch W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree. *Systematic Zoology* **20:** 406-416.

Fruh K., Yang Y., Arnold D., Chambers J., Wu L., Waters J. B., Spies T., and Peterson P. A. (1992). Alternative exon usage and processing of the major histocompatibility

complex-encoded proteasome subunits. *Journal of Biological Chemistry* **267:** 22131-22140.

Griffin T. A., Nandi D., Cruz M., Fehling H. J., Kaer L. V., Monaco J. J., and Colbert R. A. (1998). Immunoproteasome assembly: Cooperative incorporation of interferon gamma -inducible subunits. *Journal of Experimental Medicine* **187:** 97-104.

Hall T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95-98.

Hehling H. J., Swat W., Laplace C., Kuhn R., Rajewsky K., Muller U., and von Boehmer H. (1994). MHC class I expression in mice lacking the proteasome subunit LMP-7. *Science* **265:** 1234-1237.

Hoop T. P., and Woods K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences* **78:** 3824-3829.

Huelsenbeck J. P., and Crandell K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* **28:** 437-466.

Huelsenbeck J. P., and Rannala B. (1997). Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276:** 227-232.

Hughes A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London B* **256:** 119-124.

Hughes A. L. (1997). Evolution of the proteasome components. *Immunogenetics* **46:** 82-92.

Jackson M. R., Song E. S., Yang Y., and Peterson P. A. (1992). Empty and peptide-containing conformers of class I major histocompatibility complex molecules expressed in *Drosophila melanogaster* cells. *Proceedings of the National Academy of Sciences* **89:** 12117-12121.

Jensen R. A. (1976). Enzyme recruitment in the evolution of new function. *Annual Review of Microbiology* **30:** 409-425.

Kandil E., Namikawa C., Nonaka M., Greenberg A. S., Flajnik M. F., Ishibashi T., and Kasahara M. (1996). Isolation of low molecular mass polypeptide cDNA clones: implications for the origin of MHC class I-restricted antigen presentation. *Journal of Immunology* **156:** 4245-4253.

Kingsbury D. J., Griffin T. A., and Colbert R. A. (2000). Novel propeptide function in 20S proteasome assembly influences beta subunit composition. *Journal of Biological Chemistry* **275:** 24156-24162.

Kloetzel P.-M. (2001). Antigen processing by the proteasome. *Nature Reviews* **2:** 179-187.

Kumar S. (1996). A stepwise algorithm for finding minimum evolution trees. *Molecular Biology and Evolution* **13:** 584-593.

Kumar S., Tamura K., Jakobsen I. B., and Nei M. (2001). Mega2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics (submitted).*

Kyte J., and Doolittle R. F. (1982). A Simple Method for Displaying the Hydrophobic Character of a Protein. *Journal of Molecular Biology* **157:** 105-142.

Manufacturer's protocol M. (1997). TriReagent -RNA, DNA, Protein Isolation Reagent, Molecular Research Center, Inc., Cincinnati, OH.

Marchalonis J. J., and Schluter S. F. (1998). A Stochastic Model for the Rapid Emergence of Specific Vertebrate Immunity Incorporating Horizontal Transfer of Systems Enabling Duplication and Combinatorial Diversification. *Journal Of Theoretical Biology* **193:** 429-444.

Monaco J. J. (1992). A molecular model of MHC class I-restricted antigen processing. *Immunology Today* **13:** 173-178.

Nei M., and Kumar S. (2000). "Molecular Evolution and Phylogenetics," Oxford University Press, Oxford.

Nicholas K. B., and Nicholas Jr. H. B. (1997). GeneDoc: a tool for annotating and editing multiple sequence alignments. *Distributed by author*.

Orgel L. E. (1977). Gene-duplication and the origins of proteins with novel functions. *Journal of Theoretical Biology* **67:** 773.

Posada D., and Crandall K. A. (1998). ModelTest: testing the model of DNA substitution. *Bioinformatics* **14:** 817-818.

Preston G. M. (1996). Polymerase chain reaction with degenerate oligonucleotide primers to clone gene family members. *In* "Basic DNA and RNA Protocols" (A. J. Harwood, Ed.), pp. 514, Humana Press, Totowa, NJ.

Rivett A. J. (1993). Proteasomes: multicatalytic proteinase complexes. *Biochemical Journal* **291:** 1-10.

Rodriguez F., Oliver J. L., Marin A., and Medina J. R. (1990). The general stochastic model of nucleotide substitution nucleotide substitution. *Journal of Theoretical Biology* **142:** 485-501.

Rzhetsky A., and Nei M. (1993). Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution* **10:** 1073-1095.

Salter-Cid L., and Bigger C. H. (1991). Alloimmunity in the Gorgonian Coral *Swiftia exserta*. *Biological Bulletin* **181:** 127-134.

Schoenhals G. J., Krishna R. M., Grandea III A. G., Spies T., Peterson P. A., Yang Y., and Fruh K. (1999). Retention of empty MHC class I molecules by tapasin is essential to reconstitute antigen presentation in invertebrate cells. *The EMBO Journal* **18:** 743-753.

Shastri N., Schwab S., and Serwold T. (2002). Producing Nature's Gene-Chips: The Generation of Peptides for Display by MHC Class I Molecules. *Annual Review of Immunology* **20:** 463-493.

Stothard P. (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28:** 1102-1104.

Swofford D. L. (1998). PAUP*: Phylogenetic analysis using parsimony (and other methods), Sinauer Associates, Sunderland, MA.

Takezaki N., Zaleska-Rutczynska Z., and Figueroa F. (2002). Sequencing of amphioxus *PSMB5/8* gene and phylogenetic position of agnathan sequences. *Gene* **282:** 179-187.

Tamura K., and Nei M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10:** 512-526.

Tanaka K., and Kasahara M. (1998). The MHC class I ligand-generating system: roles of immunoproteasomes and the interferon-gamma-inducible proteasome activator PA28. *Immunological Reviews* **163:** 161-176.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., and Higgins D. G. (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24:** 4876-4882.

Thomson S., and Rivett A. J. (1993). Processing of N3, a mammalian proteasome beta-type subunit. *Biochemistry Journal* **315:** 733-738.

Zhang Y., and Frohman M. A. (1997). Using Rapid Amplification of cDNA Ends (RACE) to Obtain Full-Length cDNAs. *In* "cDNA Library Protocols" (I. G. Cowell, and C. A. Austin, Eds.), pp. 61-88, Humana Press Inc, Totowa, NJ.

Zharkikh A., and Li W.-H. (1993). Inconsistency of the maximum-parsimony method: The case of five taxa with a molecular clock. *Systematic Biology* **42:** 113-125.

Zwickl P., Grziwa A., Puhler G., Dahlmann B., Lottspeich F., and Barmeister W. (1992). Primary structure of the Thermoplasma proteasome and its implications for the structure, function, and evolution of the multicatalytic proteinase. *Biochemistry* **31:** 964-972.

Swiftia –LMPX homologue

```
       1   N  M  A  A  L  G  L  A  A  E  Y  E  K  K  S  F  Y  E  I  L
       1 CCAACATGGCCGCTCTTGGCTTAGCTGCAGAATACGAAAAGAAAAGTTTTTATGAaATTT
      21   N  P  K  S  A  L  C  I  D  F  N  Y  E  N  G  F  S  F  P  P
      61 TAAATCCAAAATCTGCACTGTGTATCGATTTTAACTATGAAAACGGTTTCTCCTTCCCAC
      41   V  A  N  P  A  E  F  L  S  Q  C  T  S  G  D  D  N  I  K  I
     121 CAGTAGCAAATCCAGCAGAGTTTCTAAGCCAATGCACAAGCGGTGACGACAACATAAAgA
      61   Q  F  A  H  G  T  T  T  L  A  F  K  F  Q  H  G  V  I  V  A
     181 TTCAATTTGCTCATGGcACAACAACaTTAGCATTTAAATTTCAACATGGTGTCATTGTGG
      81   V  D  S  R  A  T  A  G  S  Y  I  A  S  Q  T  V  K  K  V  I
     241 CAGTAGATTCCAGAGCTACTGCTGGATCATATATAGCATCTCAGACTGTAAAGAAAGTGA
     101   E  I  N  P  Y  L  L  G  T  M  A  G  G  A  A  D  C  S  Y  W
     301 TAGAAATAAATCCTTACCTTCTGGGGACAATGGCAGGTGGAGCAGCTGATTGTTCATACT
     121   E  R  V  L  A  K  Q  C  R  I  Y  E  L  R  N  K  E  R  I  S
     361 GGGAACGCGTTCTAGCCAAGCAATGCAGGATATACGAGCTACGTAACAAGGAGAGAATAT
     141   V  A  A  A  S  K  L  L  A  N  M  V  Y  Y  Y  R  G  M  G  L
     421 CTGTTGCAGCTGCATCCAAGTTACTGGCAAACATGGTGTATTATTACAGGGGAATGGGAC
     161   S  M  G  T  M  I  C  G  W  D  K  R  G  P  G  L  Y  Y  V  D
     481 TTTcAATGGGAACGATGATTTGTGGTTGGGATAAACGGGGTCCTGGACTTTACTATGTTG
     181   S  D  G  S  R  L  S  N  N  I  F  S  V  G  S  G  S  T  F  A
     541 ACAGTGATGGAAGTCGATTATCCAACAACATTTTCTCGGTGGGATCTGGGTCAACATTCG
     201   Y  G  V  L  D  S  G  Y  R  P  D  L  S  V  Q  E  A  Y  D  L
     601 CATATGGAGTTCTTGACAGCGGATATCGTCCTGATCTGTCTGTCCAAGAGGCATATGACC
     221   G  E  R  A  I  Y  H  A  T  H  R  D  A  Y  S  G  G  V  V  N
     661 TGGGAGAGAGGGCAATTTATCATGCAACACACAGAGATGCATACAGTGGTGGAGTTGTGA
     241   M  Y  H  M  Q  E  T  G  W  V  K  V  S  Q  N  D  V  G  Q  L
     721 ACATGTACCATATGCAAGAGACAGGCTGGGTCAAAGTATCTCAAAATGATGTTGGTCAAC
     261   H  Y  K  Y  Q  D  E  K  R  *
     781 TGCATTACAAATATCAGGATGAGAAGAGATAAGAACTTTGCAAAGACATATTTCTATCGT
     841 AGTGTTGTTAGAGGTTTGTGTCACATTTTATTAAACTTTTCTGAGAACATCTATGCTTAC
     901 AGTAGGTACAGTGCTGTCAGTGAAAAGATTCAATCATATTAAATCCAACTTGATAAAAAA
     961 AAAAAAAAAAA
```

```
Primers used:
Degenerate primers as follows:
S–GCIGCIGAYTGYCARTWYTGG (Kandil et al. 1996)
As–TTRTCCCAICCRCADATCAT

Race primers as such:
LMPX/7-3'S1–GAACGCGTCCTAGCCAAGCA
LMPX/7-3'SN–CTGTTGCAGCTGCATCCAAG
LMPX/7–5'AS1-GTAACTTGGATGCAGCTGCA
LMPX/7-5'ASN-GGATGCAGCTGCAACAGATA
LMPX/7-5'ASN2-CTGCCACAATGACACCATGT
LMPX/7-5'AS3-GATCTCCTTGCCCAGCTTCG
LMPX/7-5'AS4-GCCACAATGACACCATGTTG
```

Figure 1. SeLMPX full-length sequence and deduced amino acid translation. Primers used for the production of the full-length gene, by generating over-lapping RACE-PCR products.

Animal	Gene	Synonym, allele, or paralog	Accession No.
Human, *Homo sapiens*	LMPX	PSMB5, MB1 (syn)	NM_002797
Mouse, *Mus musculus*	LMPX	"	AF060091
Chicken, *Gallus gallus*	LMPX	"	AB001935
Shark, *Ginglymostoma cirratum*	LMPX	"	D64058
Zebrafish, Danio rerio	LMPX	"	AF032391
Hagfish, Myxine glutinosa	LMPX	"	D64054
Lamprey, Petromyzon merinus	LMPX	"	D64055
Amphioxus, *Branchiostoma lanceolatum*	LMPX	"	AF449497
Tunicate, *Botryllus schlosseri*	LMPX	"	X97729
Sponge, *Geodia cydonium*	LMPX	"	X97728
Coral, *Swiftia exserta*	LMPX	LMPX1	??
Human, *Homo sapiens*	LMP7	PSMB8, Ring10; LMP7c (syn)	U17497
Human	LMP7	LMP7e1 (par.)	Z14982
Human	LMP7	LMP7e2 (par.)	Z14982
Mouse, *Mus musculus*	LMP7	LMP7b (allele)	U22032
Mouse	LMP7	LMP7d (allele)	U22031
Mouse	LMP7	LMP7s,k,f (allele)	U22033
Rat, *Rattus sp.*	LMP7		D10727
Pig, *Sus scrofa*	LMP7		AF059493
African clawed frog, *Xenopus laevis*	LMP7	LMP7a (allele)	D44540
Xenopus laevis	LMP7	LMP7b (allele)	D44549
Medaka fish, *Oryzias latipes*	LMP7		D89725
Zebrafish, *Danio rerio*	LMP7		AF032390
Shark, *Ginglymostoma cirratum*	LMP7	LMP7a (par.)	D64057
Shark, *Ginglymostoma cirratum*	LMP7	LMP7b (par.)	D64056
Trypanosma, *Trypanosoma brucei*	PSMB	20S PSM (syn.)	CAC08538
Fruit fly, *Drosophila melanogaster*	CG9868	PSMB5-like (syn.)	AAF46978
Drosophila	PRSMB5	PSMB5-like (syn.)	AAF58748
Fungi, Yeast, *Saccharomyces cerevisiae*	PSMB	PRG1, Ring10, 20S (syn.)	M96667
Bacteria, *Methanosarcina thermophila*	PSMB	20S (syn.)	MTU22157
Plant, *Arabidopsis thaliana*	PSMB	20S (syn.)	NP_172765
Plant, Chick Pea, *Cicer arietinum*	PSMB	20S (syn.)	CAA0903
Human	LMP2	PSMB9, Ring 12 (syn.)	2118154
Mouse	LMP2	"	2467365
Frog, *Xenopus*	LMP2	"	2055297
Medaka	LMP2	"	2055311
Zebrafish	LMP2	"	2654064
Human	LMPY	PSMB6, Delta (syn.)	1362909
Mouse	LMPY	"	984940
Rat	LMPY	"	286248
Frog, *Xenopus laevis*	LMPY	"	2055299
Zebrafish, *Danio rerio*	LMPY	"	2654062
Lamprey, *Lampetra japonica*	LMPY	"	2055301
Tabacco, *Nicotiana tabacum*	LMPY-like	Delta (syn.)	1743356
Yeast, *Saccharomyces cerevisiae*	LMPY-like	Delta (syn.)	2832891
Drosophila, *Drosophila melanogaster*	CG8392	PRSMB6-like (syn.)	AAF58077
Drosophila	Prosbeta2	PRSMB6-like (syn.)	AAF49685
Human	LMP3	N3, PSMB4 (syn.)	D26600
Mouse	LMP3	N3 (syn.)	P99026
Rat	LMP3	N3 (syn.)	L17127
Xenopus	LMP3	N3 (syn.)	X62709
Drosophila	LMP3	PSMB4, N3-like (syn.)	XP_082336
Coral, *Swiftia*	LMP3	N3-like (syn.)	??

Table 1. Database accession numbers of sequences used throughout this study. Synonyms for gene names are labeled (syn.), as are paralogous genes (par.), and alleles (allele).

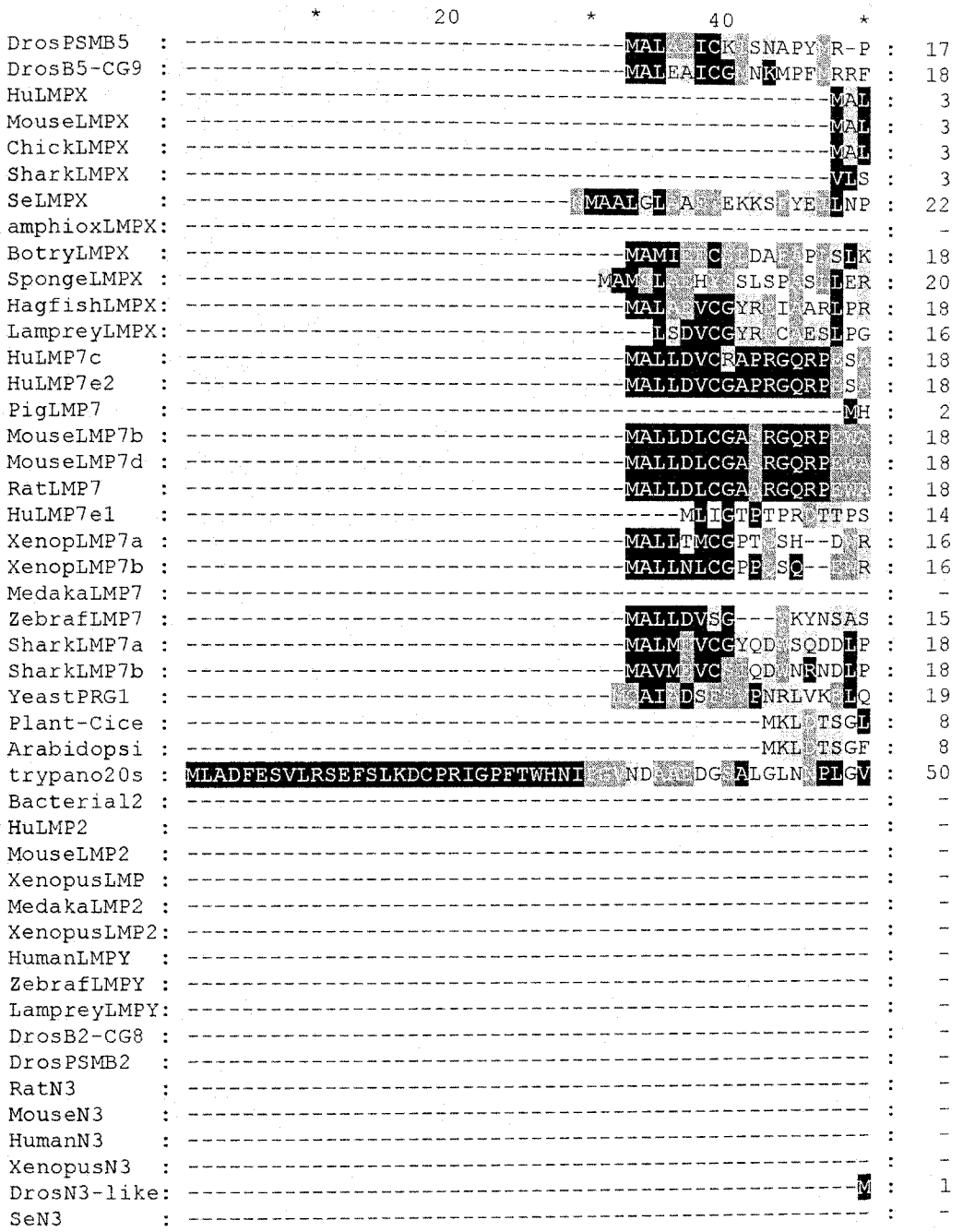
Gene tree	Log *L* (no clock)	Log *L* (clock)	df = (n-2)	Significance
LMP X + 7	-7175.55	-7204.26	18	P<0.001, sig.
LMP X	-3905.25	-3922.30	7	P<0.001, sig.
LMP 7	-3675.51	-3679.71	9	P>0.25, not sig.

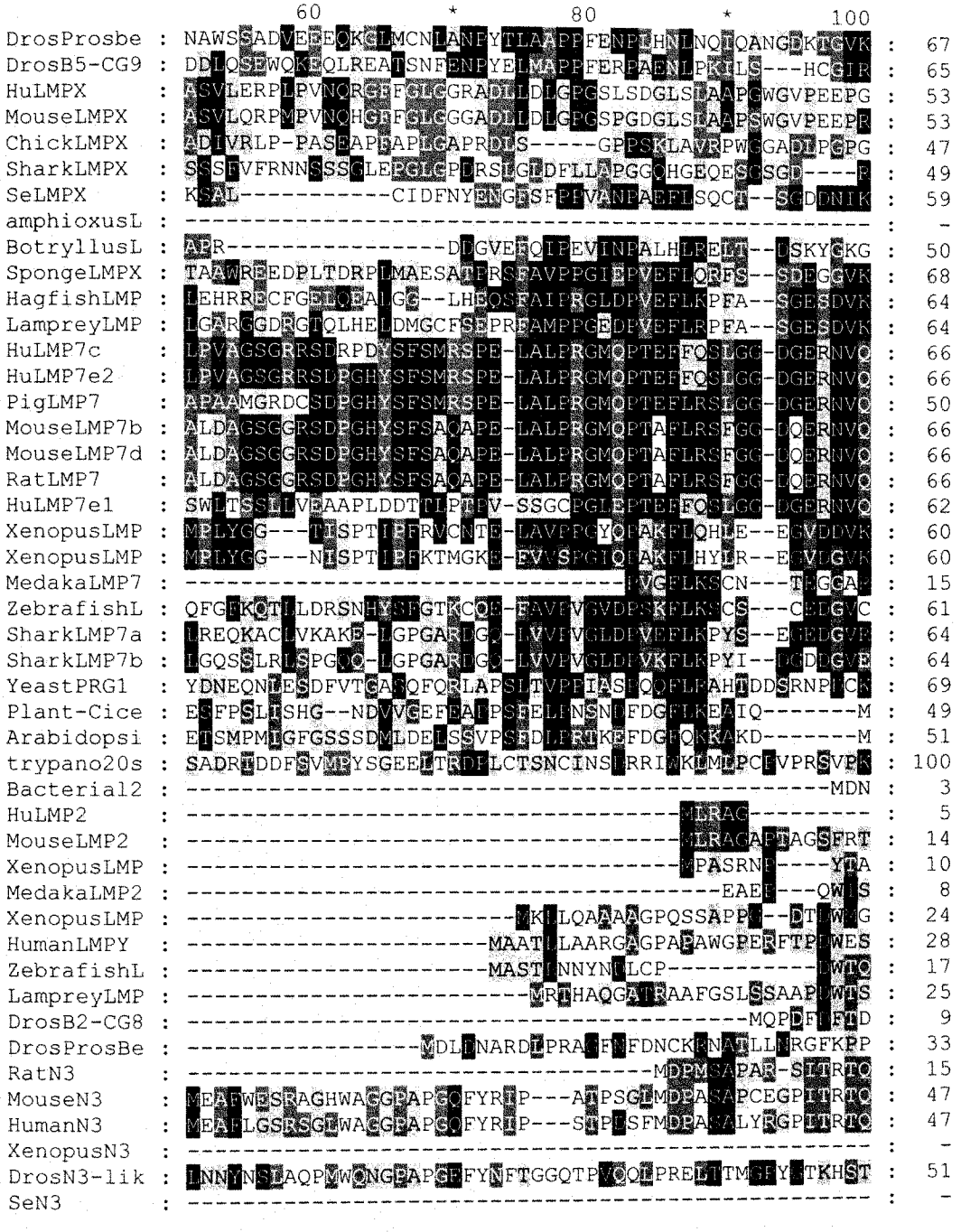
Table 2a. Maximum Likelihood test of the molecular clock hypothesis (H_0). ML trees were estimated using the Felsenstein F84 model (1993) under global rearrangements and random addition of sequences (jumbled 10x). Equal substitution rates and a transition/ transversion ratio of 2.0 was assumed. The molecular clock is tested with the Likelihood Ratio Test (because both trees have the same unrooted topology) following a Chi square distribution with (n-2) degrees of freedom as such: $\Delta = 2$(diff in logL scores). The significance of likelihood ratio statistic (Δ) is determined using df=n-2 where n= number of taxa examined. Here, the likelihood ratio test does not reject the molecular clock hypothesis (H_0) for the LMP 7 phylogeny (P>0.25). LMPX and LMPx+7 phylogenies (with and without clock assumption) produces significant (P<0.001) likelihood differences and rejects the molecular clock hypothesis. There appears to be significant rate heterogeneity in the evolution of LMP X genes (see text).

Gene tree	Log *L* (no clock)	Log *L* (clock)	df = (n-2)	Significance
LMP X + 7	-6790.07	-6817.78	18	P<0.001, sig.
LMP X	-3597.15	-3608.37	7	P<0.005, sig.
LMP 7	-3465.66	-3471.49	9	P>0.15, not sig.

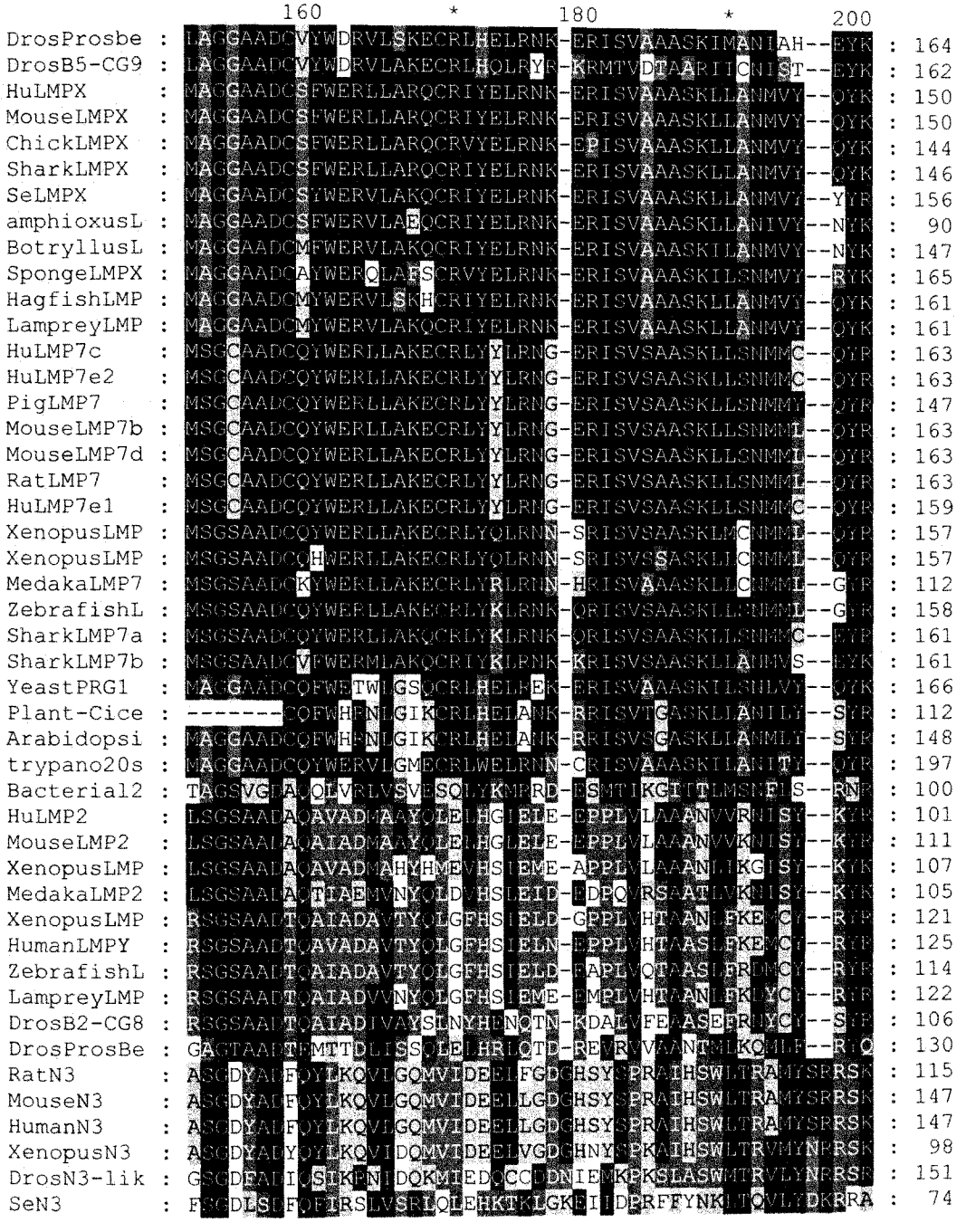
Table 2b. Maximum Likelihood test of the molecular clock hypothesis (H_0) under the appropriate models of DNA substitutions (see text). LRTs were calculated as in table 2a. Again, the likelihood ratio test does not reject the molecular clock hypothesis (H_0) for the LMP 7 phylogeny (P>0.15). LMPX and LMPx+7 phylogenies (with and without clock assumption) produces significant (P<0.001 and P<0.005) likelihood differences and rejects the molecular clock hypothesis.

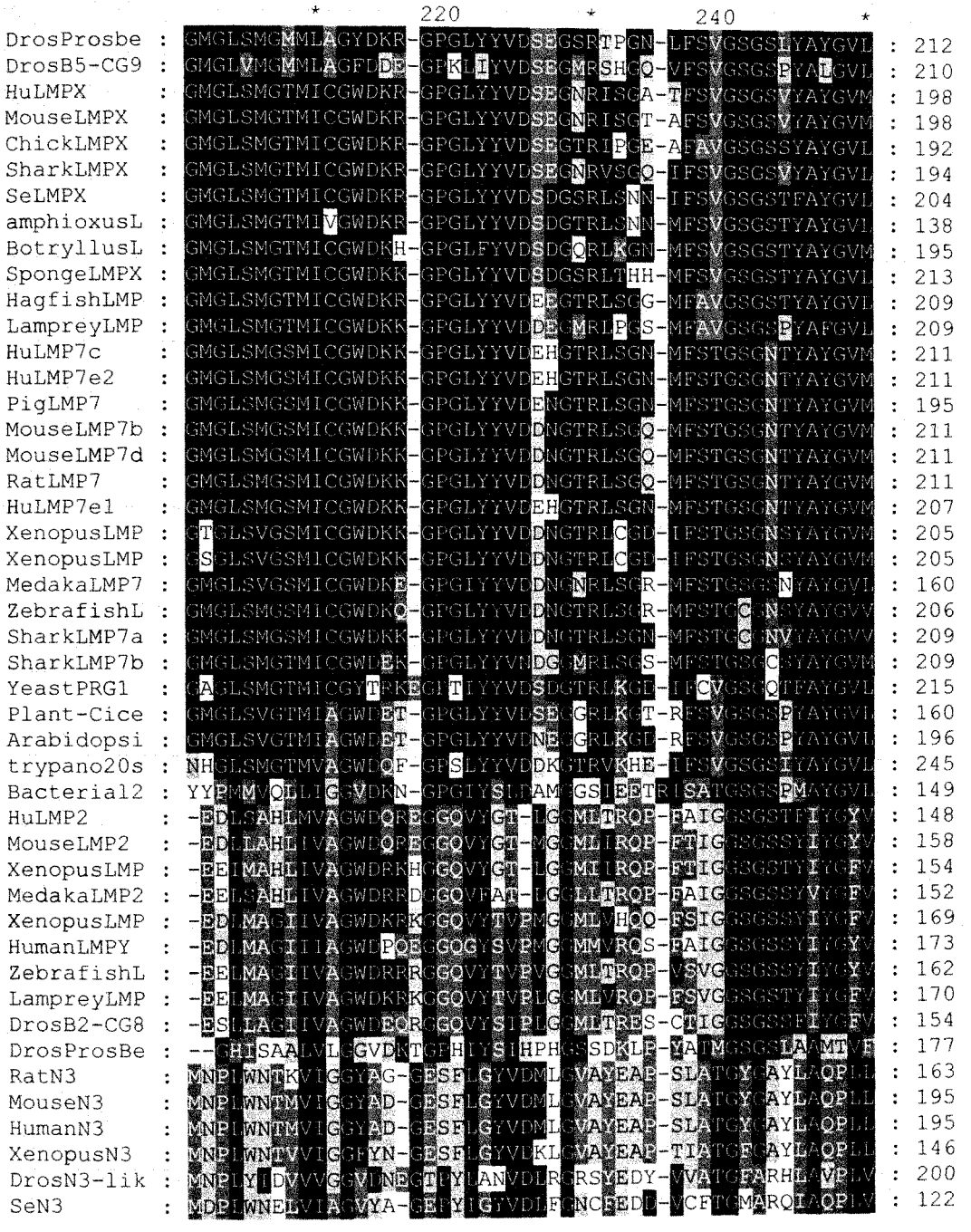
Figure 2. Clustal X alignment of the full-length polypeptide sequence of selected beta proteasome subunits.

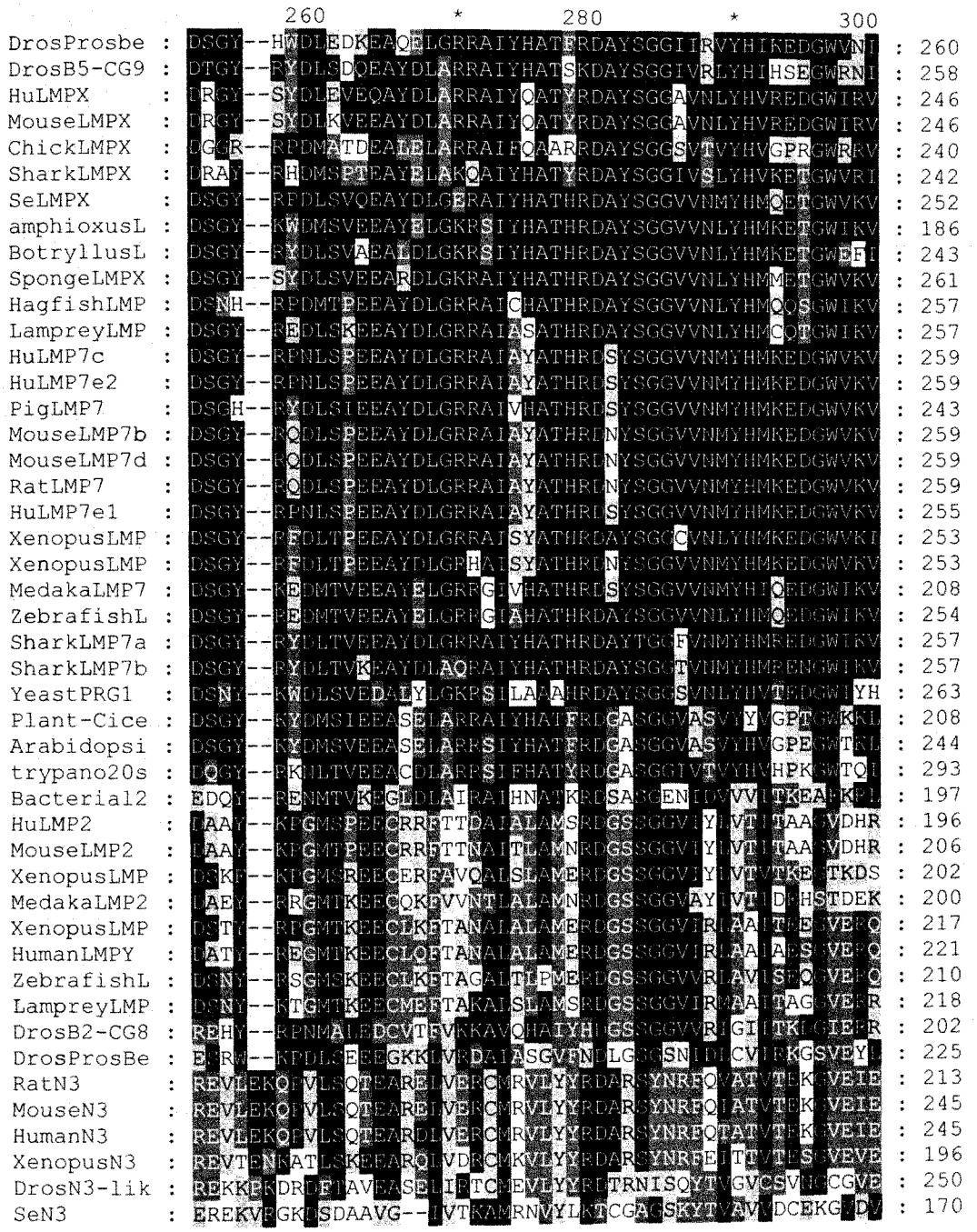


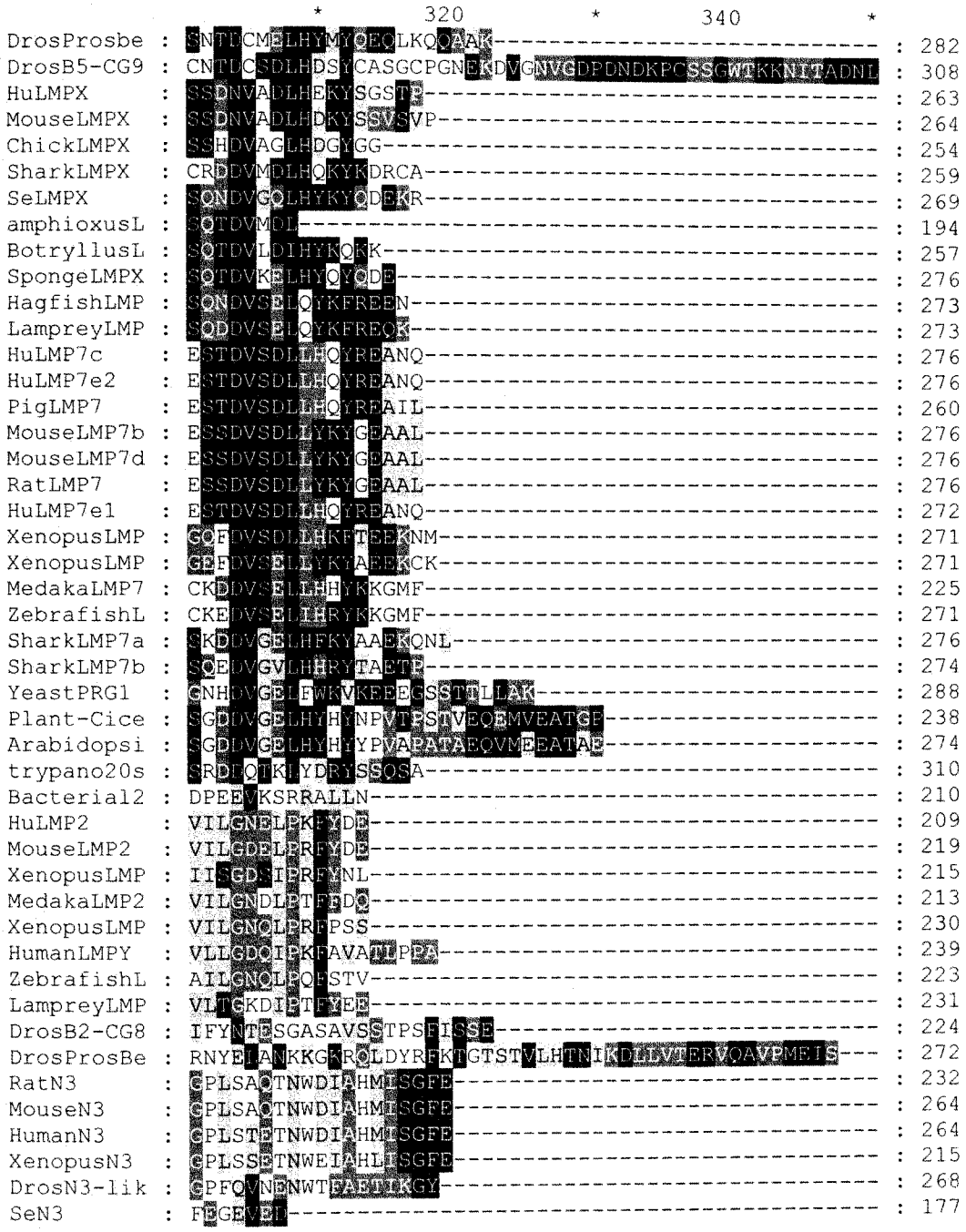


Propeptide cleavage site for the beta subunits


160
*
180
*
200
DrosProsbe : 164
DrosB5-CG9 : 162
HuLMPX : 150
MouseLMPX : 150
ChickLMPX : 144
SharkLMPX : 146
SeLMPX : 156
amphioxusL : 90
BotryllusL : 147
SpongeLMPX : 165
HagfishLMP : 161
LampreyLMP : 161
HuLMP7c : 163
HuLMP7e2 : 163
PigLMP7 : 147
MouseLMP7b : 163
MouseLMP7d : 163
RatLMP7 : 163
HuLMP7e1 : 159
XenopusLMP : 157
XenopusLMP : 157
MedakaLMP7 : 112
ZebrafishL : 158
SharkLMP7a : 161
SharkLMP7b : 161
YeastPRG1 : 166
Plant-Cice : 112
Arabidopsi : 148
trypano20s : 197
Bacterial2 : 100
HuLMP2 : 101
MouseLMP2 : 111
XenopusLMP : 107
MedakaLMP2 : 105
XenopusLMP : 121
HumanLMPY : 125
ZebrafishL : 114
LampreyLMP : 122
DrosB2-CG8 : 106
DrosProsBe : 130
RatN3 : 115
MouseN3 : 147
HumanN3 : 147
XenopusN3 : 98
DrosN3-lik : 151
SeN3 : 74


220
240
DrosProsbe : 212
DrosB5-CG9 : 210
HuLMPX : 198
MouseLMPX : 198
ChickLMPX : 192
SharkLMPX : 194
SeLMPX : 204
amphioxusL : 138
BotryllusL : 195
SpongeLMPX : 213
HagfishLMP : 209
LampreyLMP : 209
HuLMP7c : 211
HuLMP7e2 : 211
PigLMP7 : 195
MouseLMP7b : 211
MouseLMP7d : 211
RatLMP7 : 211
HuLMP7e1 : 207
XenopusLMP : 205
XenopusLMP : 205
MedakaLMP7 : 160
ZebrafishL : 206
SharkLMP7a : 209
SharkLMP7b : 209
YeastPRG1 : 215
Plant-Cice : 160
Arabidopsi : 196
trypano20s : 245
Bacterial2 : 149
HuLMP2 : 148
MouseLMP2 : 158
XenopusLMP : 154
MedakaLMP2 : 152
XenopusLMP : 169
HumanLMPY : 173
ZebrafishL : 162
LampreyLMP : 170
DrosB2-CG8 : 154
DrosProsBe : 177
RatN3 : 163
MouseN3 : 195
HumanN3 : 195
XenopusN3 : 146
DrosN3-lik : 200
SeN3 : 122


260
*
280
*
300
DrosProsbe : DSGY--HWDLEDKEAQELGRRAIYHATFRDAYSGGIIRVYHIKEDGWVNI : 260
DrosB5-CG9 : DTGY--RYDLSDQEAYDLARRAIYHATSKDAYSGGIVRLYHIHSEGWRNI : 258
HuLMPX : DRGY--SYDLEVEQAYDLARRAIYQATYRDAYSGGAVNLYHVREDGWIRV : 246
MouseLMPX : DRGY--SYDLKVEEAYDLARRAIYQATYRDAYSGGAVNLYHVREDGWIRV : 246
ChickLMPX : DGGR--RPDMATDEALELARRAIFQAARRDAYSGGSVTVYHVGPRGWRRV : 240
SharkLMPX : DRAY--RHDMSPTEAYELAKQAIYHATYRDAYSGGIVSLYHVKETGWVRI : 242
SeLMPX : DSGY--RFDLSVQEAYDLGERAIYHATHRDAYSGGVVNMYHMQETGWVKV : 252
amphioxusL : DSGY--KWDMSVEEAYELGKRSIYHATHRDAYSGGVVNLYHMKETGWIKV : 186
BotryllusL : DSGY--RYDLSVAEALDLGKRSIYHATHRDAYSGGVVNLYHMKETGWEFI : 243
SpongeLMPX : DSGY--SYDLSVEEARDLGKRAIYHATHRDAYSGGVVNLYHMMETGWVKV : 261
HagfishLMP : DSNH--RPDMTPEEAYDLGRRAICHATHRDAYSGGVVNLYHMQQSGWIKV : 257
LampreyLMP : DSGY--REDLSKEEAYDLGRRAIASATHRDAYSGGVVNLYHMCQTGWIKV : 257
HuLMP7c : DSGY--RPNLSPEEAYDLGRRAIAYATHRDSYSGGVVNMYHMKEDGWVKV : 259
HuLMP7e2 : DSGY--RPNLSPEEAYDLGRRAIAYATHRDSYSGGVVNMYHMKEDGWVKV : 259
PigLMP7 : DSGH--RYDLSIEEAYDLGRRAIVHATHRDSYSGGVVNMYHMKEDGWVKV : 243
MouseLMP7b : DSGY--RQDLSPEEAYDLGRRAIAYATHRDNYSGGVVNMYHMKEDGWVKV : 259
MouseLMP7d : DSGY--RQDLSPEEAYDLGRRAIAYATHRDNYSGGVVNMYHMKEDGWVKV : 259
RatLMP7 : DSGY--RQDLSPEEAYDLGRRAIAYATHRDNYSGGVVNMYHMKEDGWVKV : 259
HuLMP7e1 : DSGY--RPNLSPEEAYDLGRRAIAYATHRDSYSGGVVNMYHMKEDGWVKV : 255
XenopusLMP : DSGY--RFDLTPEEAYDLGRRAISYATHRDAYSGGCVNLYHMKEDGWVKI : 253
XenopusLMP : DSGY--RFDLTPEEAYDLGRHAISYATHRDNYSGGVVNMYHMKEDGWVKV : 253
MedakaLMP7 : DSGY--KEDMTVEEAYELGRRGIVHATHRDSYSGGVVNMYHIQRDGWIKV : 208
ZebrafishL : DSGY--REDMTVEEAYELGRRGIAHATHRDAYSGGVVNLYHMQRDGWIKV : 254
SharkLMP7a : DSGY--RYDLTVEEAYDLGRRAIYHATHRDAYTGGFVNMYHMREDGWIKV : 257
SharkLMP7b : DSGY--RYDLTVKEAYDLAQRAIYHATHRDAYSGGTVNMYHMRENGWIKV : 257
YeastPRG1 : DSNY--KWDLSVEDALYLGKRSILAAAHRDAYSGGSVNLYHVTEDGWIYH : 263
Plant-Cice : DSGY--KYDMSIEEASELARRAIYHATFRDCASGGVASVYYVGPTGWKKL : 208
Arabidopsi : DSGY--KYDMSVEEASELARRSIYHATFRDCASGGVASVYHVGPEGWTKL : 244
trypano20s : LQGY--PKNLTVEEACDLARRSIFHATYRDCASGGIVTVYHVHPKGWTQI : 293
Bacterial2 : EDQY--RENMTVKEGLDLAIRAIHNATKRDSASGENIDVVVITKEAKKFL : 197
HuLMP2 : LAAY--KFGMSPEECRRFTTDAIALAMSRDGSSGGVIYLVTLTAAAVDHR : 196
MouseLMP2 : LAAY--KFGMTPEECRRFTTNAITLAMNRDGSSGGVIYLVTLTAAAVDHR : 206
XenopusLMP : DSKF--KFGMSREECERFAVQALSLAMERDGSSGGVIYLVTVTKEETKDS : 202
MedakaLMP2 : LAEY--RRGMTKEECQKFVVNTLALAMNRLGSSGGVAYLVTLDPHSTDEK : 200
XenopusLMP : DSTY--RFGMTKEECLKFTANALALAMERDGSSGGVIRLAALTHEGVEPQ : 217
HumanLMPY : LATY--REGMTKEECLQFTANALALAMERDGSSGGVIRLAALAPSGVEPQ : 221
ZebrafishL : DSNY--RSGMSKEECLKFTAGALTLPMERDGSSGGVVRLAVLSQQGVEPQ : 210
LampreyLMP : DSNY--KTGMTKEECMEFTAKALSLAMSRDGSSGGVIRMAALTAGGVEPR : 218
DrosB2-CG8 : REHY--RPNMALEDCVTFVKKAVQHAIYHIGSSGGVVRIGITKIGIEPR : 202
DrosProsBe : ESRW--KPDLSEEEGKKLVEDAIASGVFNLLGGGSNIDICVIPKGSVEYL : 225
RatN3 : REVLEKQFVLSQTEAREIVERCMRVLYYRDARSYNRFQVATVTLKGVEIE : 213
MouseN3 : REVLEKQFVLSQTEAREIVERCMRVLYYRDARSYNRFQIATVTLKGVEIE : 245
HumanN3 : REVLEKQFVLSQTEARDIVERCMRVLYYRDARSYNRFQTATVTLKGVEIE : 245
XenopusN3 : REVTENKATLSKEEARQIVDRCMKVLYYRDARSYNRFEITTVTLSGVEVE : 196
DrosN3-lik : REKKPKDRDFTAVEASELIFTCMEVLYYRETRNISQYTVGVCSVNCGVE : 250
SeN3 : EREKVPGKDSDAAVG--IVTKAMRNVYLKTCGAGSKYTVAVVDCEKGIDV : 170


320
340
DrosProsbe : 282
DrosB5-CG9 : 308
HuLMPX : 263
MouseLMPX : 264
ChickLMPX : 254
SharkLMPX : 259
SeLMPX : 269
amphioxusL : 194
BotryllusL : 257
SpongeLMPX : 276
HagfishLMP : 273
LampreyLMP : 273
HuLMP7c : 276
HuLMP7e2 : 276
PigLMP7 : 260
MouseLMP7b : 276
MouseLMP7d : 276
RatLMP7 : 276
HuLMP7e1 : 272
XenopusLMP : 271
XenopusLMP : 271
MedakaLMP7 : 225
ZebrafishL : 271
SharkLMP7a : 276
SharkLMP7b : 274
YeastPRG1 : 288
Plant-Cice : 238
Arabidopsi : 274
trypano20s : 310
Bacterial2 : 210
HuLMP2 : 209
MouseLMP2 : 219
XenopusLMP : 215
MedakaLMP2 : 213
XenopusLMP : 230
HumanLMPY : 239
ZebrafishL : 223
LampreyLMP : 231
DrosB2-CG8 : 224
DrosProsBe : 272
RatN3 : 232
MouseN3 : 264
HumanN3 : 264
XenopusN3 : 215
DrosN3-lik : 268
SeN3 : 177

```
DrosProsbe : ------- :   -
DrosB5-CG9 : QTKLATV : 315
HuLMPX     : ------- :   -
MouseLMPX  : ------- :   -
ChickLMPX  : ------- :   -
SharkLMPX  : ------- :   -
SeLMPX     : ------- :   -
amphioxusL : ------- :   -
BotryllusL : ------- :   -
SpongeLMPX : ------- :   -
HagfishLMP : ------- :   -
LampreyLMP : ------- :   -
HuLMP7c    : ------- :   -
HuLMP7e2   : ------- :   -
PigLMP7    : ------- :   -
MouseLMP7b : ------- :   -
MouseLMP7d : ------- :   -
RatLMP7    : ------- :   -
HuLMP7e1   : ------- :   -
XenopusLMP : ------- :   -
XenopusLMP : ------- :   -
MedakaLMP7 : ------- :   -
ZebrafishL : ------- :   -
SharkLMP7a : ------- :   -
SharkLMP7b : ------- :   -
YeastPRG1  : ------- :   -
Plant-Cice : ------- :   -
Arabidopsi : ------- :   -
trypano20s : ------- :   -
Bacterial2 : ------- :   -
HuLMP2     : ------- :   -
MouseLMP2  : ------- :   -
XenopusLMP : ------- :   -
MedakaLMP2 : ------- :   -
XenopusLMP : ------- :   -
HumanLMPY  : ------- :   -
ZebrafishL : ------- :   -
LampreyLMP : ------- :   -
DrosB2-CG8 : ------- :   -
DrosProsBe : ------- :   -
RatN3      : ------- :   -
MouseN3    : ------- :   -
HumanN3    : ------- :   -
XenopusN3  : ------- :   -
DrosN3-lik : ------- :   -
SeN3       : ------- :   -
```

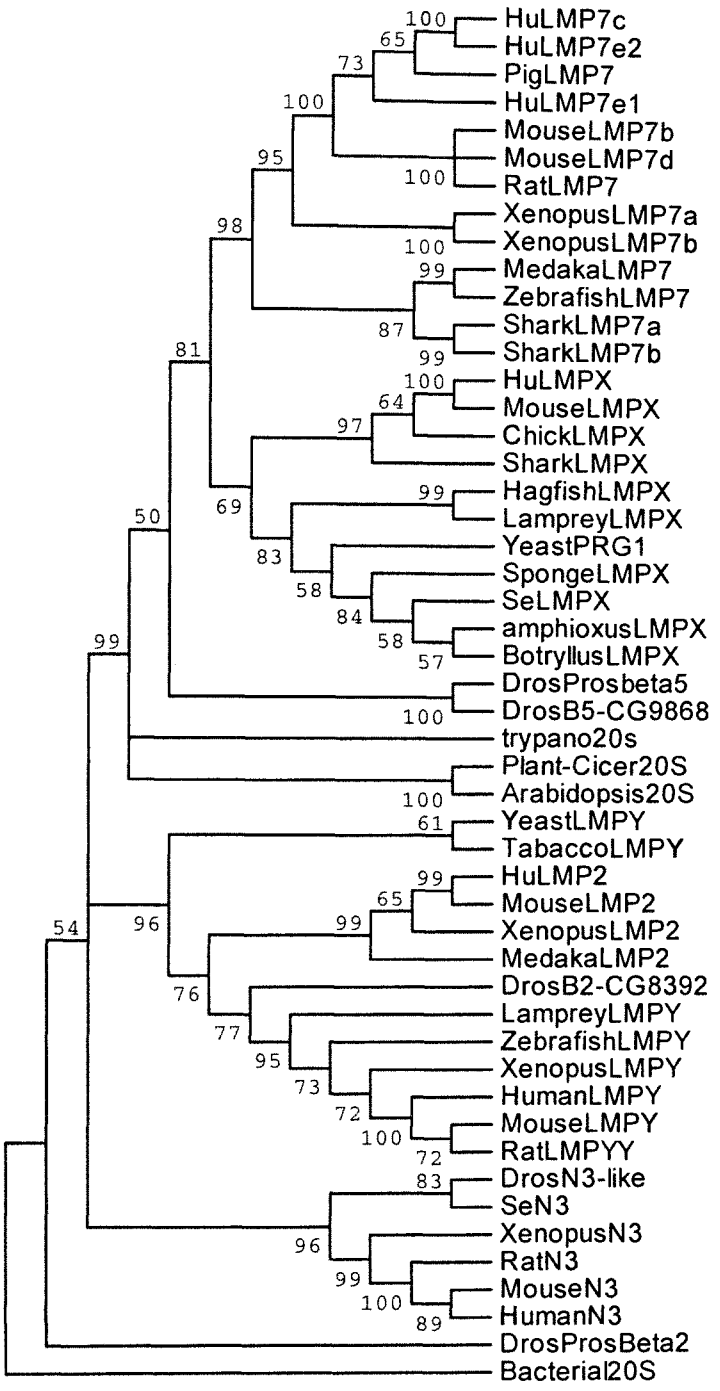


Figure 3a. Minimum evolution bootstrap tree (condensed at 50% support) using Poisson corrected distance of full length proteasome beta subunit protein sequences, N= 50, 357 sites analyzed. Tree rooted with bacterial 20s beta subunit protein. 10,000 bootstrap replicates with pairwise deletion of gaps. Se- Coral sequence; SeLMPX and SeN3. Accession numbers can be found in table 1.

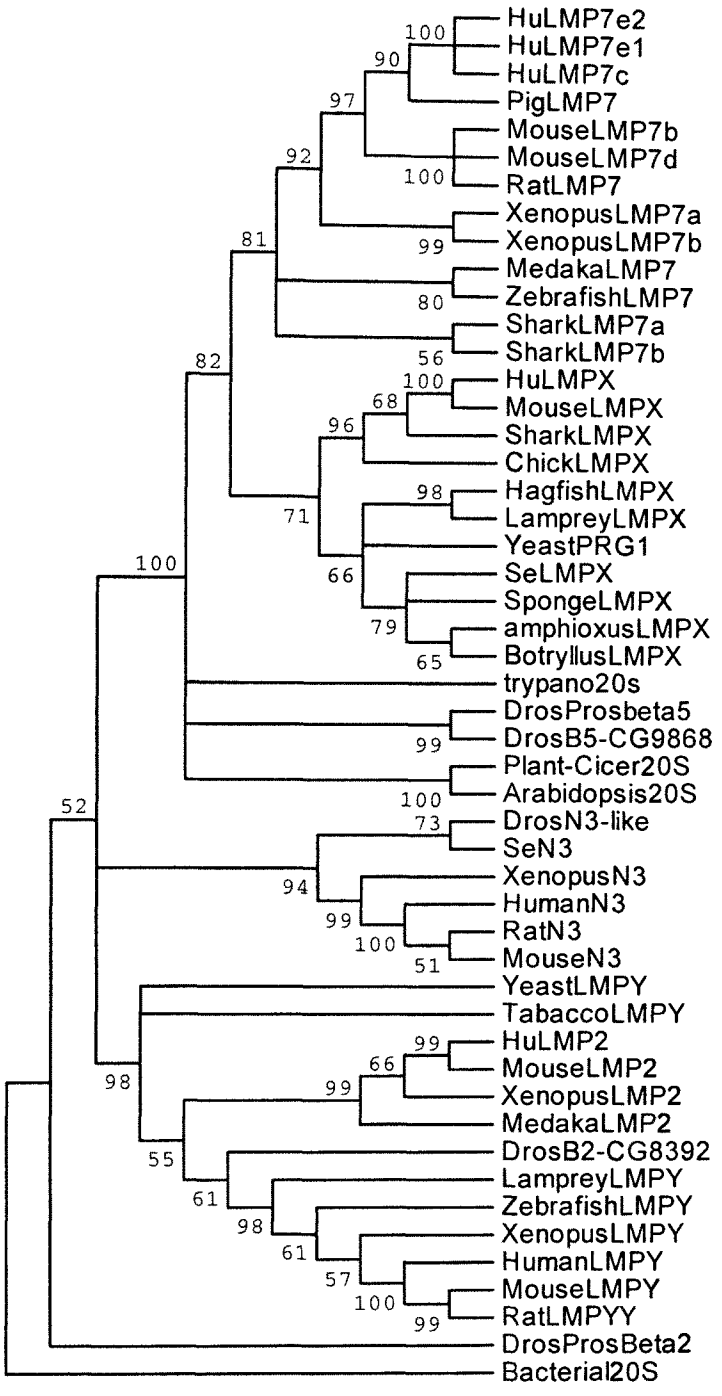


Figure 3b. Minimum evolution bootstrap tree (condensed at 50% support) using Poisson corrected distance of proteasome beta subunit protein sequences, in C-terminal conserved 20S core; N= 50, 252 sites analyzed. Tree rooted with bacterial 20s beta subunit protein. 10,000 bootstrap replicates with pairwise deletion of gaps. Se- coral sequence; SeLMPX and SeN3. Accession numbers can be found in table 1.

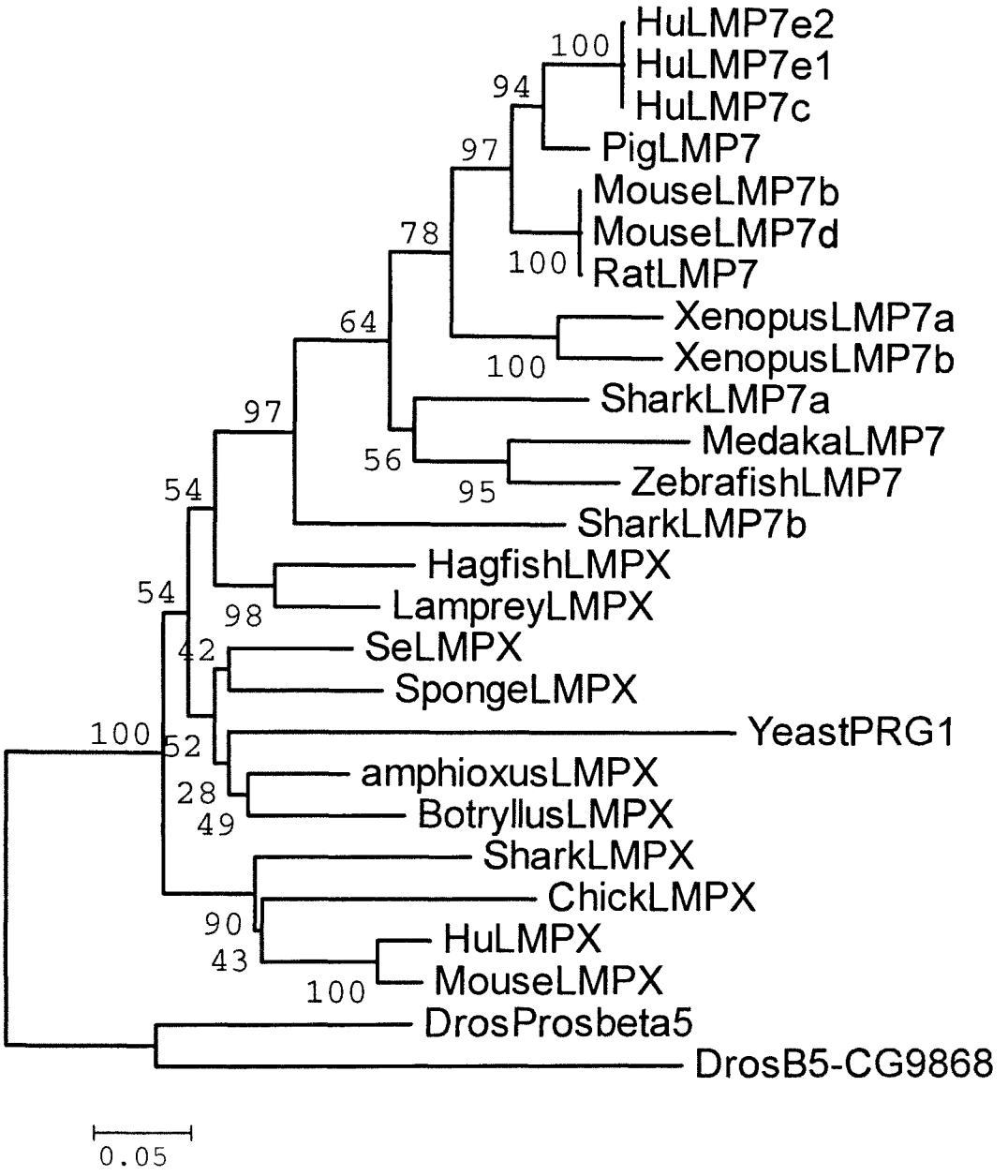


Figure 3c. Minimum evolution bootstrap tree (10,000 replicates) under Poisson-corrected distance model of the C-terminal portion (252 sites) of the proteasome beta genes, X and 7. Tree was rooted with the *Drosophila* PSMB5 sequences, DrosProsbeta5 and DrosB5-CG9868. The yeast PRG1 sequence (PSMB5-like, see Fig. 3a and 3b) has the longest branch length, greater than 2x the length of the longest branch from the sponge or coral (SeLMPX).

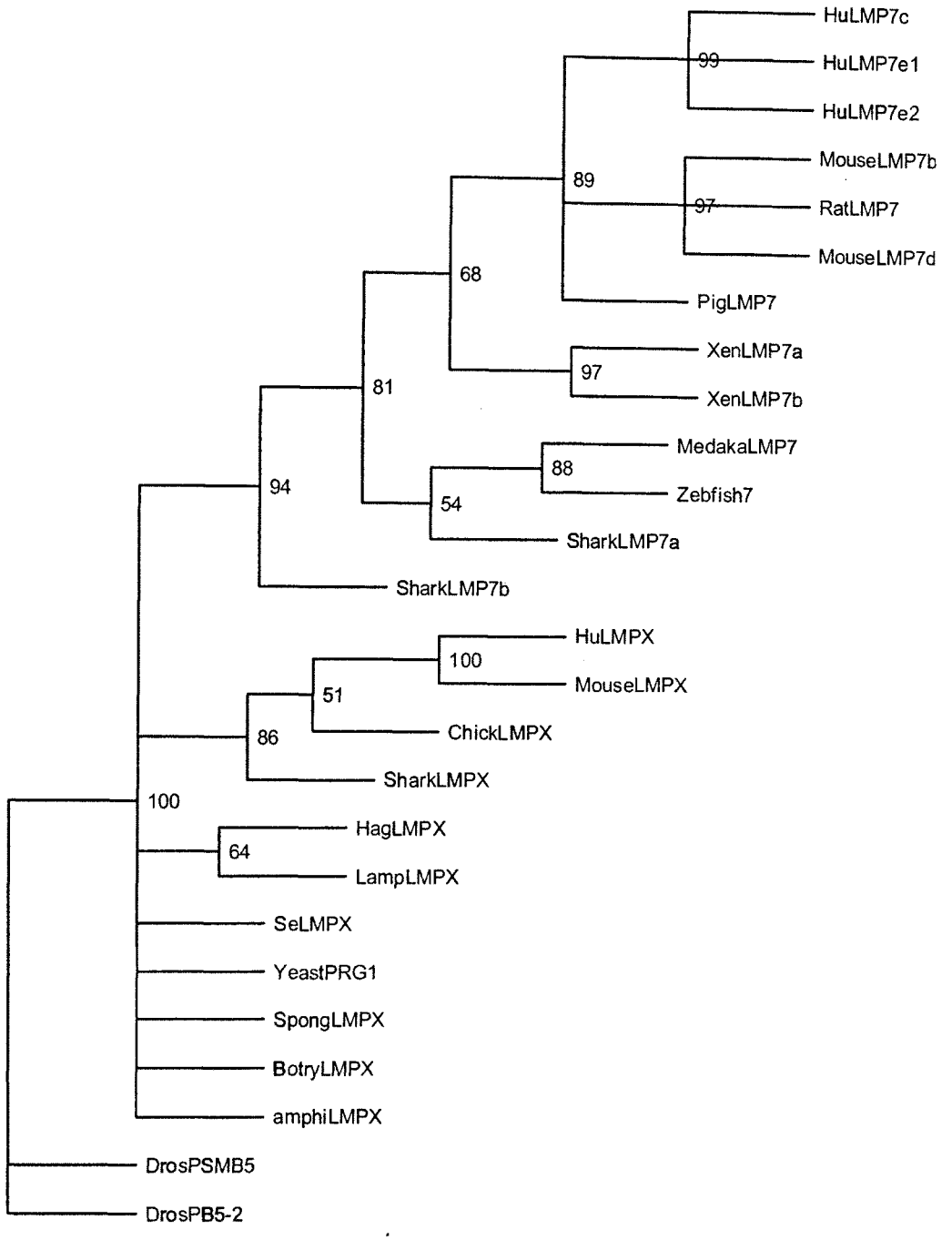


Figure 3d. Maximum parsimony bootstrap tree (500 replicates), produced by heuristic search criteria using 203 sites (104 informative) of the C-terminal region (protein level), and the TBR branch swapping algorithm with random addition of sequences and 10 repetitions (random) of the proteasome beta genes, X and 7. Tree rooted with the *Drosophila* sequences, DrosPSMB5 & PB5-2

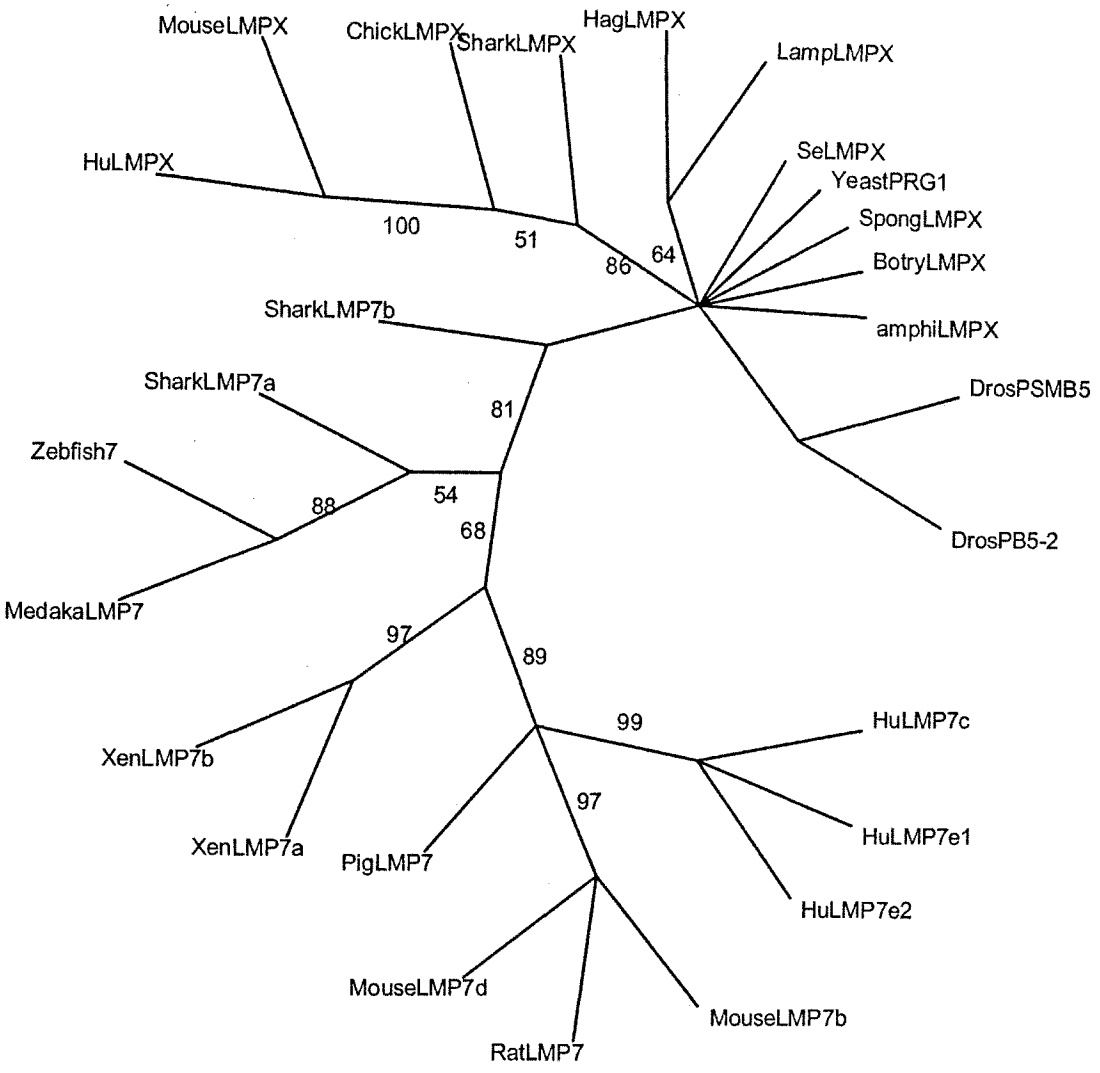


Figure 3e. Unrooted maximum parsimony bootstrap tree (500 replicates) from Figure 3d, produced by heuristic search criteria using 203 sites (104 informative) of the C-terminal region (protein level), and the TBR branch swapping algorithm with random addition of sequences and 10 repetitions (random) of the proteasome beta genes, X and 7.

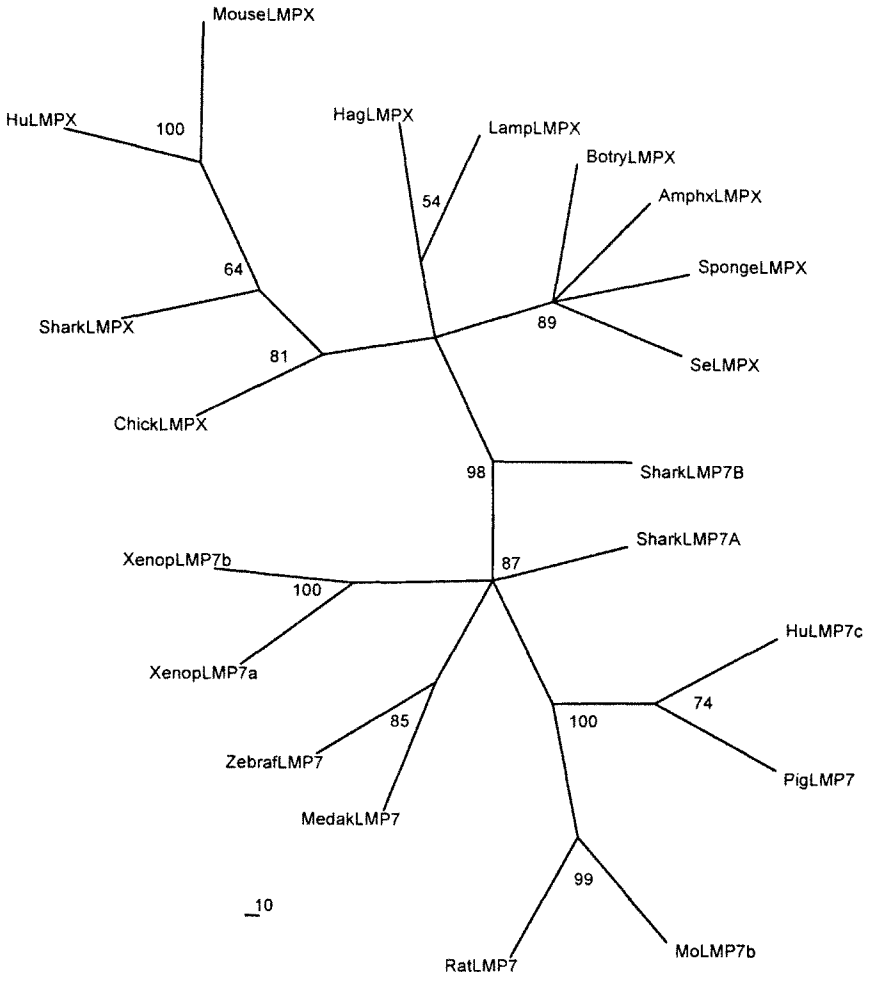


Figure 4a. Unrooted maximum likelihood bootstrap tree under the TrNef +I+G model of nucleotide substitution (logL = -6832.92). Coral sequence is, SeLMPX. Note unresolved polytomy for invertebrate LMPX. See text.

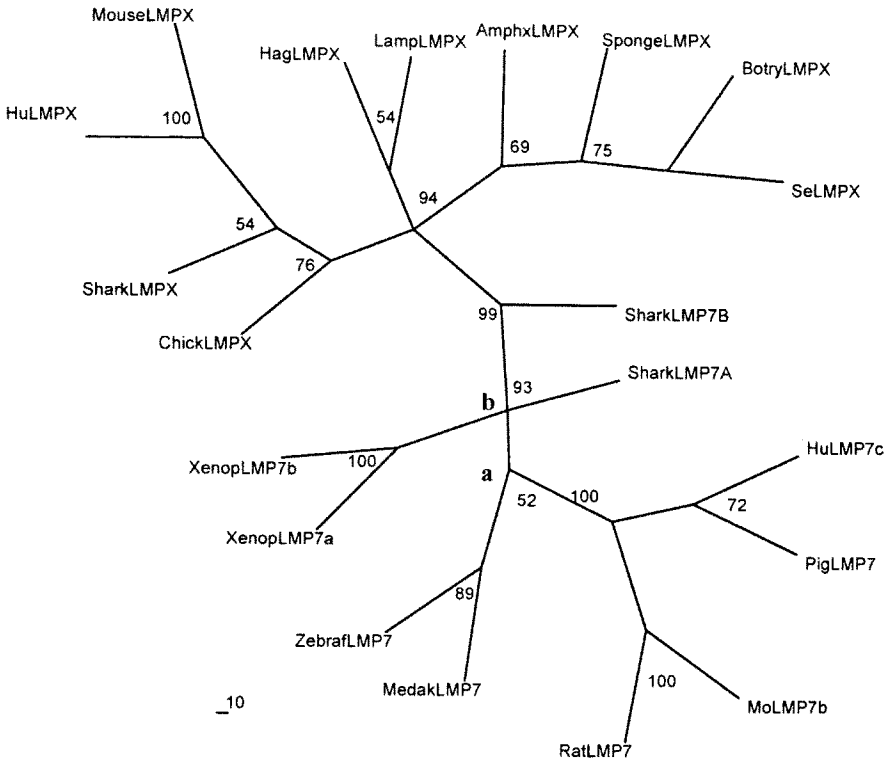


Figure 4b. Unrooted maximum likelihood bootstrap tree under the GTR +I+G model of nucleotide substitution (logL = -6794.76). If node "a", with a low bootstrap support, is collapsed into node "b", the LMP7 phylogeny produces a monophyletic node where Xenopus, Fish, and Shark LMP 7A share a common node with the mammalian LMP7 (as in Fig.4a).

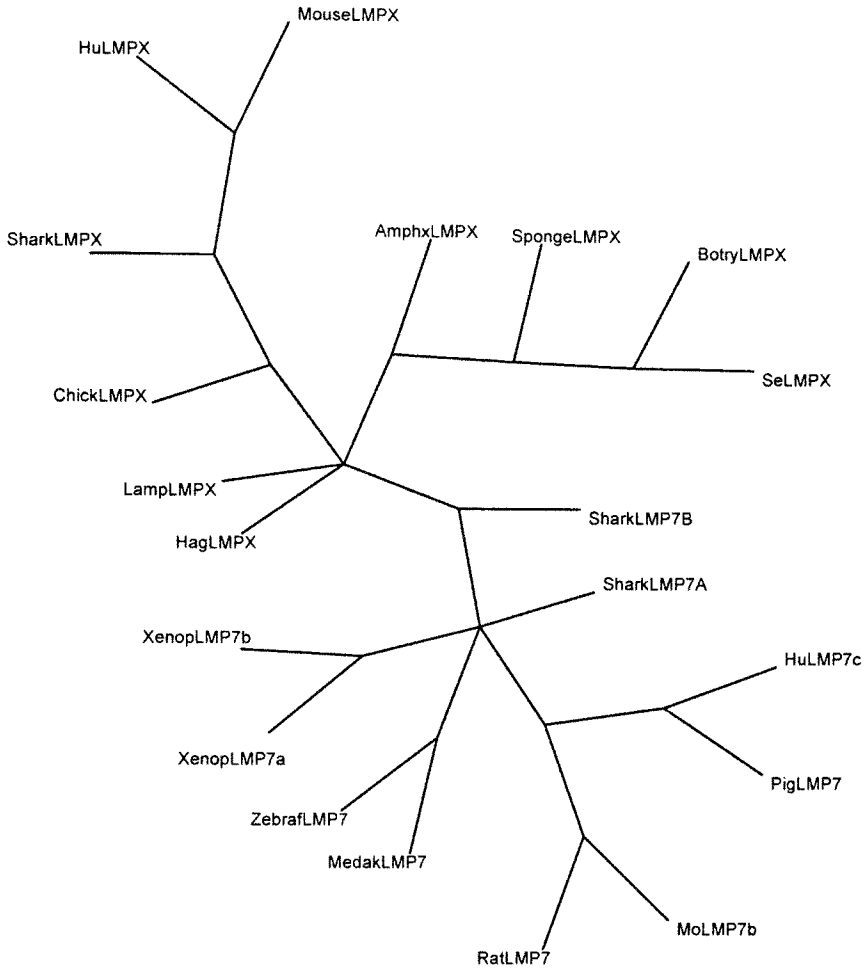
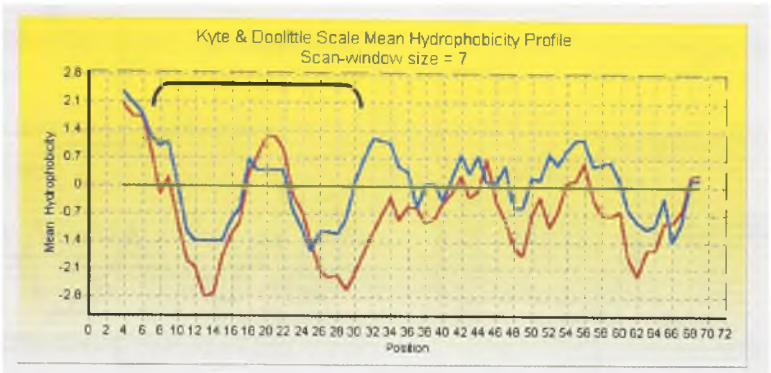


Figure 4c. Unrooted maximum likelihood bootstrap tree under the GTR +I+G model of nucleotide substitution (logL = -6794.76) from Fig. 4b. Node "a" was collapsed into node "b" the LMP7 phylogeny produces a monophyletic node where Xenopus, Fish, and Shark LMP 7A share a common node with the mammalian LMP7 (as in Fig.4a).

A.



B.

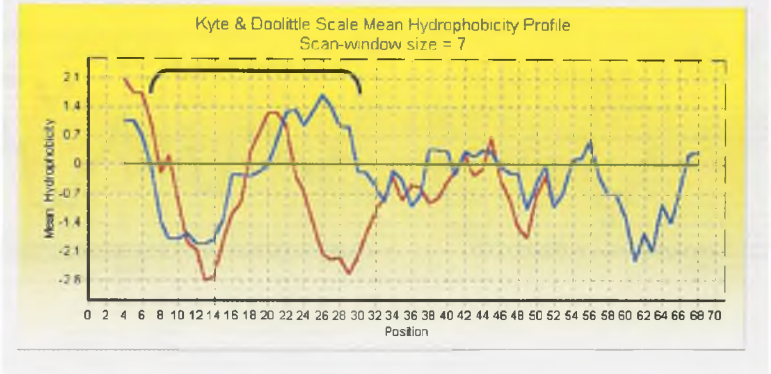


Figure 5. Hydrophobicity profiling of the N-terminal propeptide of Human LMP X and LMP7 (e1 and e2) paralogs. Human LMP7e2 propeptide is red in all cases. A. Profile of HuLMPe2 and HuLMPX. B. Profile of HuLMP7e2 and e1 isotypes. Major region of difference lies in the region of residue number 15-33. **Note** region under black horizontal bracket is area of interest.

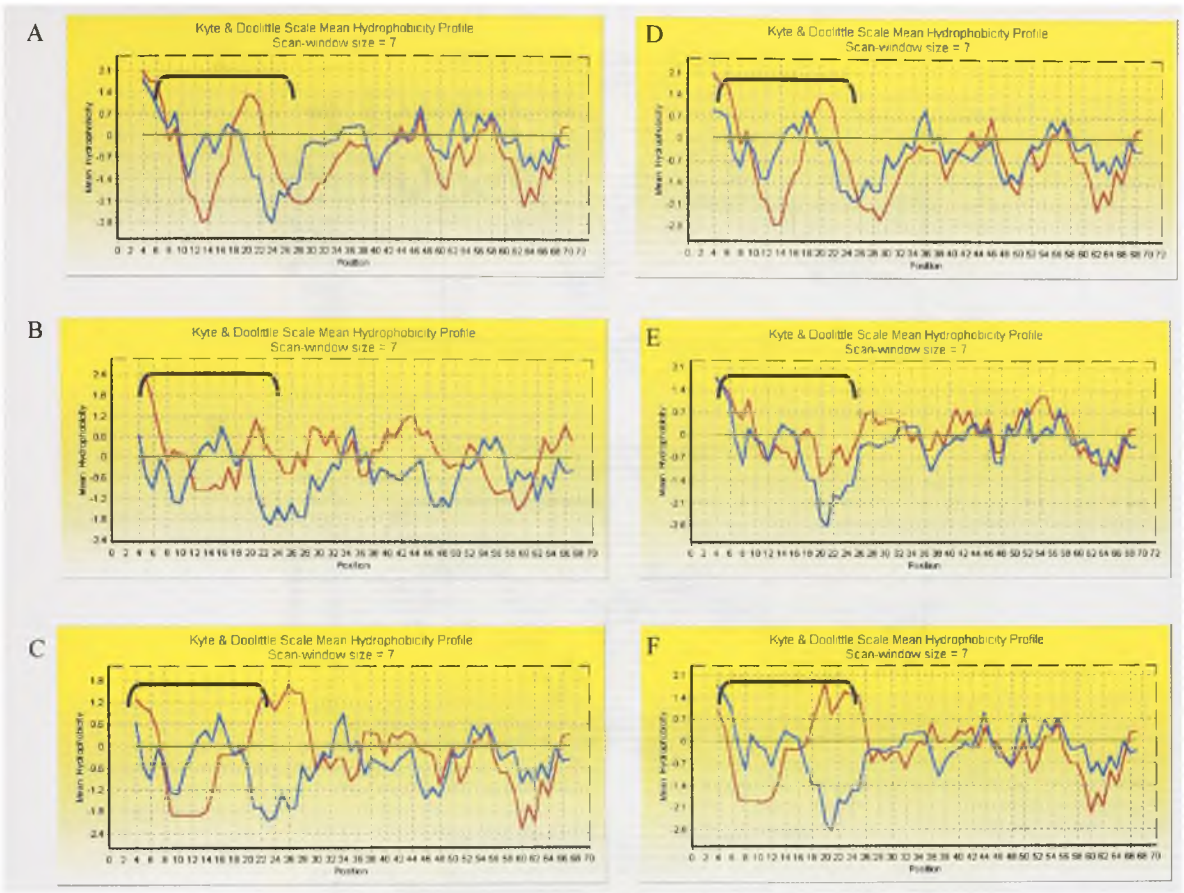


Figure 6. Hydrophobicity profiling of the N-terminal propeptide of aganthan LMPX and human LMP X and LMP7 (e1 and e2). Human LMP7e2 is red in all cases. A. Lamprey LMPX and HuLMP7e2. B. Lamprey LMPX and HuLMPX. C. Lamprey LMPX and HuLMP7e1. D. Hagfish LMPX and HuLMP7e2. E. Hagfish LMPX and HuLMPX. F. Hagfish LMPX and HuLMP7e1. Agnathan LMP X propeptide most resembles the propeptide of the active form of human LMP7. **Note** region under black horizontal bracket is area of interest.

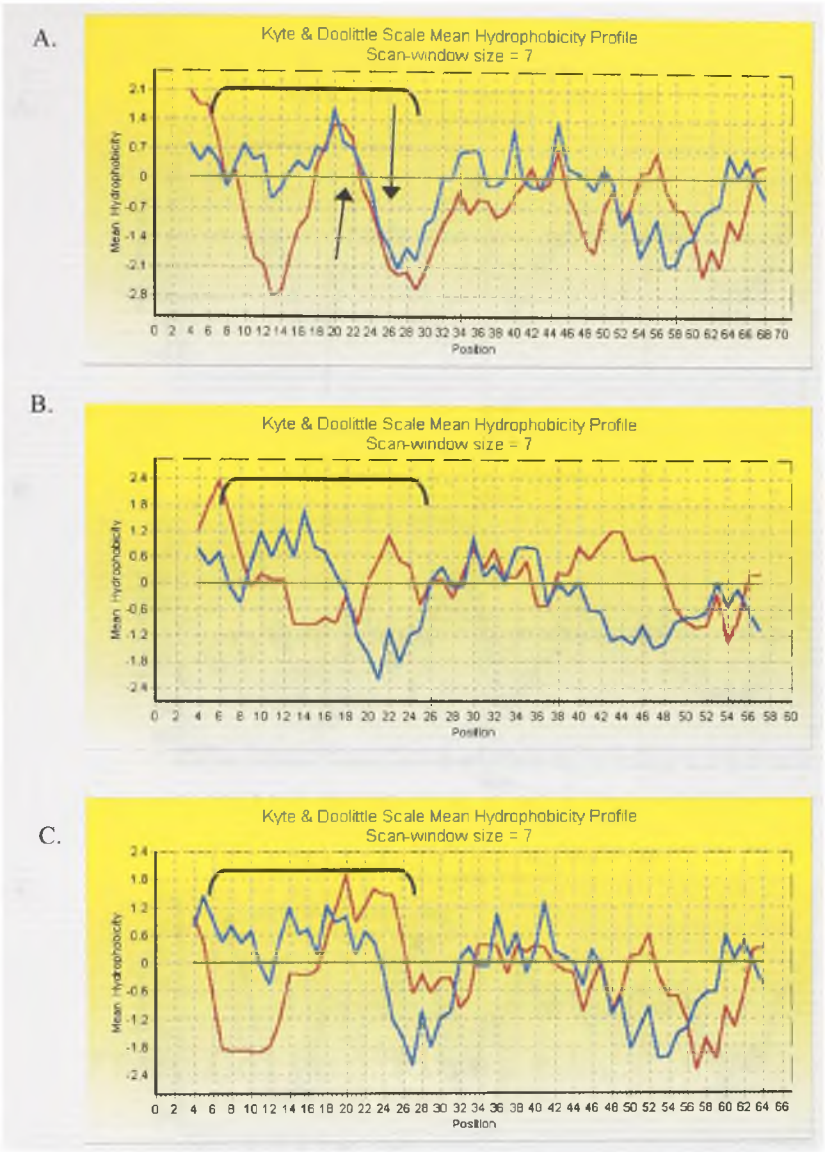


Figure 7. Hydrophobicity profiling of the N-terminal propeptide of urochordate, *Botryllus* LMPX and human LMP X and LMP7 (e1 and e2). Human propeptide is red in all cases. A. *Botryllus* LMPX and HuLMP7e2. B. *Botryllus* LMPX and HuLMPX. C. *Botryllus* LMPX and HuLMP7e1. **Note** region under black horizontal bracket is area of interest.

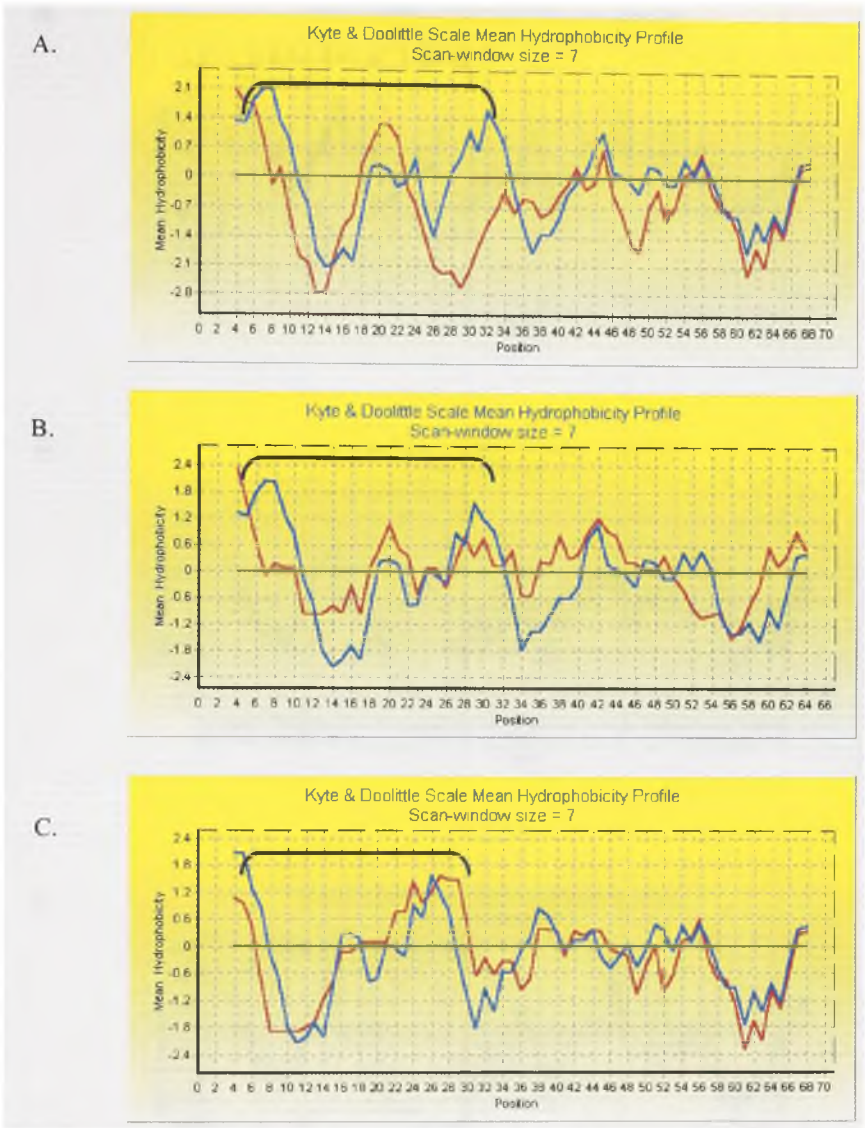


Figure 8. Hydrophobicity profiling of the N-terminal region of coral LMPX and human LMP X and LMP7 (e1 and e2). Human propeptide is red in all cases. A. Coral LMPX and HuLMP7e2. B. Coral LMPX and HuLMPX. C. Coral LMPX and HuLMP7e1. **Note** region under black horizontal bracket is area of interest.

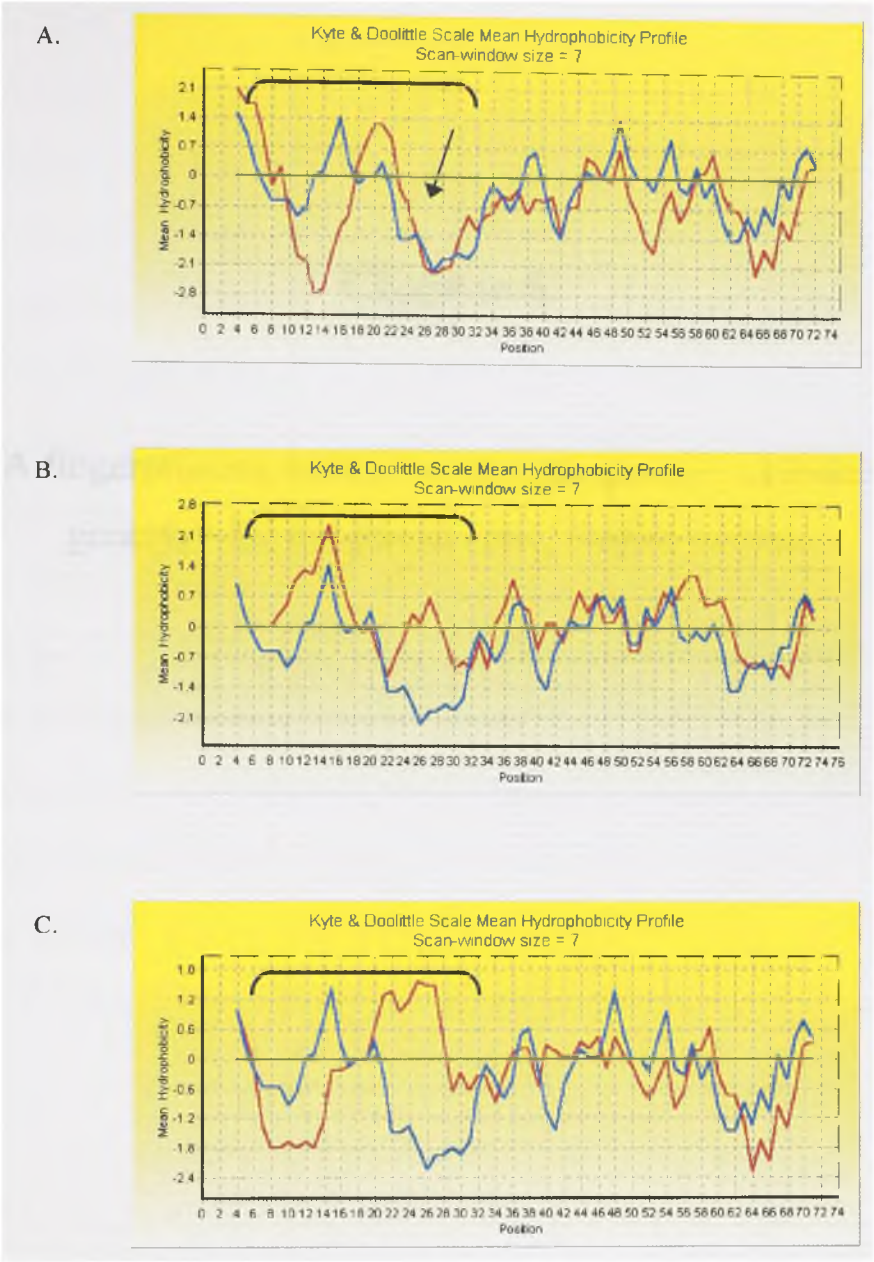


Figure 9. Hydrophobicity profiling of the N-terminal region of Sponge LMPX and Human LMP X and LMP7 (e1 and e2). Human propeptide is red in all cases. A. Sponge LMPX and HuLMP7e2. B. Sponge LMPX and HuLMPX. C. Sponge LMPX and HuLMP7e1. **Note** region under black horizontal bracket is area of interest.

Chapter 6

A DNA fingerprinting method to estimate genetic relatedness and genotype the gorgonian coral, *Swiftia exserta.*

Abstract

Studies of histocompatibility have demonstrated that the gorgonian coral *Swiftia exserta* (Cnidaria, Anthozoa) fulfills the three minimal criteria (Hildemann *et al.*, 1979) (cytotoxicity, specificity, and altered secondary response) characterizing adaptive-type immunocompetence (Salter-Cid & Bigger, 1991). Though primary allograft recognition and rejection responses occur within a narrow range (7-9 days, Salter-Cid & Bigger, 1991), deviations from this range are sometimes encountered (Olano, C. and C.H. Bigger, unpublished observations). *Swiftia*, a deep water soft coral, is a convenient animal model not only because it maintains well in laboratory conditions, but it lacks endosymbiotic zooxanthellae. This lack of symbionts (external source of nucleic acids) is critical for molecular applications involving non-specific polymerase chain reaction (PCR) primers. Our lab has adopted a simple and reliable DNA fingerprinting method (Zietkiewicz *et al.*, 1994) for genotyping intraspecific corals for our immunogenetic studies of histo(in)compatibility. This method permits the calculation of similarity indices (Lynch, 1990) for the estimation of pairwise genetic relatedness.

Introduction

DNA fingerprinting is based upon the idea that polymorphic genetic markers will provide sufficient information on individualization as well as calculation of genetic relatedness or distance (Jeffreys *et al.*, 1985; Smouse & Chevillon, 1998). Microsatellites are tandemly repeating nucleotide units (1-5bp) that exhibit high mutation rates and are distributed throughout eukaryotic genomes (e.g., (Hancock, 1996). Analysis of sets of single loci has been used extensively for classification of familial relatedness (e.g., parentage) since some loci and their inheritance are highly informative (Blouin *et al.*, 1996; Bowcock *et al.*, 1994; Hancock, 1996; Rosenbaum & Deinard, 1998; Schlotterer & Pemberton, 1998; Weber, 1990). Multilocus analysis, though, can provide information on closely related organisms without prior knowledge of their genome's contents (Chakraborty & Jin, 1993; Jeffreys *et al.*, 1985; Nagaoka & Ogihara, 1997; Zietkiewicz *et al.*, 1994).

DNA fingerprinting is useful in studying individual relatedness, population substructure and species relatedness because the investigator can examine a number of bands (or loci) and determine the proportion of fragments shared between individuals or between offspring and parents. Genetic distance can generally be estimated in a small sample size as long as a large number of loci are examined (Chakraborty & Jin, 1993; Danforth & Freeman-Gallant, 1996; Jeffreys *et al.*, 1985; Lynch, 1990; Lynch, 1991; Nei, 1978 ; Nei & Roychoudhury, 1974). Determining genetic relatedness within a population, though, is quite complex involving multivariable considerations (Danforth & Freeman-Gallant, 1996; Smouse & Chevillon, 1998; Zhivotovsky & Feldman, 1995).

Because we are interested in the effects of intra-specific genotypic relatedness on allograft recognition and rejection reactions, a multilocus fingerprinting method was desired that would provide high resolution for genetically undefined individuals. In this study, microsatellite loci were targeted because they have been shown to be very informative in both relatedness and ecological analyses of closely related individuals (e.g., Hearne *et al.*, 1992; Rosenbaum & Deinard, 1998; Schlotterer & Pemberton, 1998; Zhivotovsky & Feldman, 1995). Primers are designed with a (CA)n repeat and two (or more) 3' anchoring nucleotides [$(CA)_8RG$] so they bind and extend into the inter-repeat region (for detail see Zietkiewicz et al., 1994; and Figure 1). Because microsatellites tend to cluster in certain regions of eukaryotic genomes (e.g., Dib *et al.*, 1996), this technique is likely to produce an abundance of informative bands in the PCR-range of 100-2000 bp.

In comparison to other fingerprinting methods, the described technique is convenient because it requires little DNA (PCR is utilized), primers are easy to design, and results can be obtained in 48 hours. Complex, reproducible patterns of bands are produced which may be used to estimate genetic relatedness. Because microsatellites evolve at considerably high rates (reviewed in Hancock, 1996), they appear to provide informative loci for multilocus fingerprinting of a diverse array of organisms (Detter *et al.*, 1998; Morgante & Olivieri, 1993; Nagaoka & Ogihara, 1997; Weber, 1990; Zhivotovsky & Feldman, 1995; Zietkiewicz *et al.*, 1994).

Materials and Methods

Animals

Nine gorgonian soft corals, *Swiftia exserta* (Phylum Cnidaria, Class Anthozoa), were collected off the southeast coast of Florida (USA) and maintained in the laboratory as previously described (Salter-Cid & Bigger, 1991). Animals were maintained alive until needed for experimental use. Pieces of tissue were removed from the animal and either directly homogenized in DNA extraction buffer (see below) or pulverized in liquid nitrogen and then homogenized. Random collected tissues recovered from the field were dehydrated in an ethanol gradient and stored in 70% ethanol at –20 °C until ready for use.

Preparation of Genomic DNA

Genomic DNA was extracted directly from homogenized tissue utilizing DNAzol (MRC, Cleveland, OH, USA) as previously described (see chapter 2). Integrity of genomic DNA was determined by visualization in ethidium-bromide stained TAE-agarose gel and quantitated with the Spect3000 instrument (BioRad). Purified RNA-free genomic DNA was stored in water at 4°C or in 70% ethanol at -20°C until ready for use.

Microsatellite- anchored polymerase chain reaction for multilocus fingerprinting

Genome-based multilocus microsatellite fingerprinting was performed as previously described (Zietkiewicz *et al.*, 1994). Essentially, primers were designed so that they contained the microsatellite repeat of choice within the sequence [i.e., $(CA)_{10}$] and include a pair of anchoring nucleotides to help minimize "slippage" [i.e., $(CA)_8RG$;

see figure 1]. Primers were used alone or in pairs (against more than one type of repeat). Primers that anchor on the 3' end amplify the intervening sequence of two microsatellite regions and provide information on the intervening sequence rather than the microsatellites themselves. Primers with 5' end anchoring nucleotides anchor at the 5' end of the microsatellites and amplify both the microsatellites and the intervening sequence (see figure 1). This type of amplification generally produces more polymorphism because it includes microsatellite loci which could be expanding or shrinking (Blouin *et al.*, 1996; Hearne *et al.*, 1992). Additionally, 5' anchored primers have the potential of producing banding patterns that are not reproducible because of primer slippage due to imperfect binding between primer and template.

Based on modifications of the previously described procedure (Zietkiewicz *et al.*, 1994), 50-100ng of genomic DNA was amplified in 50ul of total reaction volume containing 50pmol of primer [$(CA)_8RG$], 200uM of dNTP, 2% formamide or 1-2% of DMSO and 1.5U of Taq polymerase (Qiagen). The target DNA was amplified for 35 cycles [95°C for 5min followed by 30 cycles of 95°C for 1min, 52-55°C for 1min, and 72°C for 2min, and finalized with a 10min extension at 72°C]. Primers were 5'-labeled with 6-FAM (PE-Biosystems) fluorescent tags.

PCR products were cleaned up by ethanol precipitation using ammonium acetate as the precipitating salt. The products were resuspended and an aliquot analyzed for fingerprint patterns. Typically, the products were resuspended in 20ul of water and 1-2 ul were combined with formamide containing loading buffer and loaded onto a sequencing gel, containing Rox-labeled internal lane standards (PE-Biosystems).

Analysis of microsatellite banding patterns

Fluorescently-labeled products were separated on 6% polyacrylamide sequencing gels and analyzed on an ABI –Prism 377 automated sequencer utilizing the GeneScan (ver 2.0 and 3.0) software. Fingerprints were analyzed and scored using the Genotyper software (ver 1.0 and 2.0) package. All reactions were performed in triplicate to help resolve ambiguities. In some cases, bands are very close together. These may represent stutter artifacts (Biosystems, 1997) rather than heterozygosity. Typically, stutters are not reproducible while true heterozygotes are (see figure 3). In many cases of ambiguity, a 'bin' is created in a 2-3bp span where bands within this region are counted as one band (Biosystems, 1997). Since many bands result from these multilocus fingerprints, areas of ambiguity can also be skipped without severe compromise to the data.

Once the reproducible bands (shared and unique) are chosen, a matrix is created to calculate similarity indices. In this case, each column represents the individual characters (bands) and each row is a unique individual being fingerprinted. The matrix allows the calculation of similarity indices (Lynch, 1990), probability of a genotypic match, and parsimony-based production of trees showing individual relatedness. First, frequencies of shared bands, x, are calculated and averaged across all individuals (Jeffreys *et al.*, 1985). Assuming Hardy-Weinberg Law, the probability of two individuals sharing any one band is then calculated as $x = 2q - q^2$ (Jeffreys *et al.*, 1985), where q is the frequency of a specific band in the population. By calculating the mean number of bands per individual, m, the probability of two randomly chosen individuals sharing the same complete fingerprint can be estimated by x^m (Jeffreys *et al.*, 1985).

Pairwise comparisons of the individual characters in the matrix allows for the calculation of similarity indices (Lynch, 1990). The similarity index [S = 2Nab / (Na + Nb), where Nab is the number of shared bands and Na and Nb are the number of bands in each pairwise comparison] is computed for each pair of animals compared and provides a relative measure of genetic distance based on the fingerprinting method used (Lynch, 1990). Since each band is a character, maximum parsimony (Fitch, 1971; Swofford & Maddison, 1987; Swofford *et al.*, 1996) analysis of the matrix is performed to construct a tree of pairwise relatedness values. This pattern is useful for testing and/or supporting relationships that are suggested by the similarity indices. The relatedness values are imported into PAUP (Swofford, 1998) to produce maximum parsimony genotypic trees. Because the genetic basis of the bands or fragments is unknown, it is impossible to assign specific bands to a locus. This issue of non-independence of characters can have confounding effects on familial relatedness studies and in estimating population divergence (Dowling *et al.*, 1996). As long as a sufficient number of bands are analyzed (which will have to be determined experimentally) the proper resolution can be attained in the estimation of relatedness (Blouin *et al.*, 1996; Bowcock *et al.*, 1994; Lynch, 1990; Morgante & Olivieri, 1993; Nagaoka & Ogihara, 1997; Schlotterer & Pemberton, 1998; Zhivotovsky & Feldman, 1995; Zietkiewicz *et al.*, 1994).

Results

PCR with fluorescently labeled primer

Fingerprints were generated using fluorescently labeled primers which appear in reproducible manner in triplicate runs (see figure 2 for example data). From the overall fingerprints, bands were chosen that were consistently present among the sampled individuals for the $(CA)_8RG$ primers (figure 1). Different variations of primer sequence, based on the repeat chosen and the position of the anchoring nucleotides, affect the conditions chosen for PCR. With the advent of gradient thermal cycling, in one run the optimal melting temperature can be determined for a given primer. 5'- prime anchoring primers, because they can slip into imperfect matches and wobble at lower annealing temperatures, should have the annealing temperatures approach the Tm as much as possible to force perfect matches. These types of modifications help alleviate stutter artifact which can appear to be double-banded phenotypes (see figure 3).

There are a few considerations that we found are essential in producing consistent patterns. The use of a denaturant that relaxes secondary structure, such as DMSO or formamide, is very important for the production of consistent banding patterns. The consistent integrity of the genomic DNA is also important as highly degraded DNA can lead to the loss of specific bands or the production of very light bands due to low levels of amplification. The purity of the DNA, free from protein contamination, is essential as well because protein bound to the template will inhibit or compromise the efficiency of the amplification. Under these criteria, we were able to produce consistent fingerprinting

patterns. The presence or absence of a band, and the nature of the band, is easy to predict from triplicate results (see figure 3).

Data matrix and the similarity index

Fifty-six positions were chosen between 94 and 340bp, appearing as bands among the differently sampled individuals (sexually outbreed population, N=9). This data was put into a matrix as described above. After scoring the matrix for the presence (1) or absence (0) of a band, it was determined that the mean number of bands per individual, m, was 21.7. The average frequency of shared bands, x, across the matrix was determined to be 0.39. The probability of two randomly chosen individuals sharing one locus can be estimated as such: $x = 2q - q^2$ where q was estimated to be 0.22 by solving the quadratic term $q^2 - 2q + 0.39$. Therefore, assuming this to be a purely outbred population of animals based on egg and sperm dispersal among corals, the probability of two randomly chosen individuals sharing an identical fingerprint pattern can be estimated as x^m, which was determined to be 1.3×10^{-9} for this data set. The fingerprint patterns from this data set support the depiction of *Swiftia* as a sexually outbred population.

Pairwise comparisons of the data matrix were produced in the computation of similarity indices (Lynch, 1990) as a measure of relative genetic distance (similarity). The relationships derived from this genetic distance information can be tested by performing character-based analysis (i.e., Parsimony) directly from the matrix data. Can microsatellite-based fingerprints and the associated genetic distance information be related to recognition of self vs. nonself in immunological phenomena? Or do

microsatellites diverge in ways independent of the divergence of antigenic nonself constituents?

Data matrix and genotypic trees

Because the bands are individual characters, the data matrix can be analyzed (independent of similarity index calculation) directly by maximum parsimony methods to produce genotypic trees of pairwise relatedness. The data matrix produced in this study was imported into PAUP (as described above) and a maximum parsimony tree was produced by the branch and bound method (see figure 4). Based on band-sharing data, the resulting character-based tree of the individuals is supported by the calculated similarity indices (genetic distance), see figure 4. Therefore, a tree produced before the similarity indices are calculated can provide preliminary information on relatedness.

Discussion

Utilizing fluorescently-labeled microsatellite-anchoring primers, we generated consistent genome-wide fingerprint patterns from the gorgonian coral, *Swiftia exserta.* We were able to localize the distribution of 56 bands among the tested animals using the $(CA)_8RG$ primer. Fingerprints were generated in triplicate and consistently produced bands were scored as present or absent in comparison with the 56 positions previously chosen. Band-sharing data produced an estimate (value between 0 – 1, where two individuals with a similarity index of 1 have an identical pattern of bands) of genetic relatedness among all animals tested. Overall, the data can be used for heterozygosity

testing, calculation of similarity indices, and for the production of pairwise genotypic trees.

This fingerprinting method will be used to estimate genetic relatedness in the gorgonian corals used in our allografting experiments. We hypothesize that multilocus microsatellite fingerprinting will provide sufficient genetic resolution so that recorded differences in tissue-recognition and reaction times can be correlated to pairwise genetic distance. Increased resolution may become necessary and achievable by the addition of new microsatellite anchoring primers. This is based on the fact that additional markers increase genotypic-specific resolution (Chakraborty & Jin, 1993; Danforth & Freeman-Gallant, 1996; Jeffreys *et al.*, 1985; Lynch, 1990; Lynch, 1991; Nei, 1978; Nei & Roychoudhury, 1974; Slatkin, 1995; Smouse & Chevillon, 1998; Zhivotovsky & Feldman, 1995). At this time, though, we have no data to indicate that genetic distance estimation based on shared microsatellite loci will relate to the degree of genetic distance required for allograft recognition in histoincompatibility reactions.

References

Biosystems P. A. (1997). "GeneScan Reference Guide," Perkin Elmer Corp., Foster City, CA.

Blouin M. S., Parsons M., Lacaille V., and Lotz S. (1996). Use of Microsatellite Loci to Classify Individuals by Relatedness. *Molecular Ecology* **5:** 393-401.

Bowcock A. M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J. R., and Cavalli-Sforza L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368:** 455-457.

Chakraborty R., and Jin L. (1993). Determination of relatedness between individuals using DNA fingerprinting. *Human Biology* **65:** 875-895.

Danforth B. N., and Freeman-Gallant C. R. (1996). DNA fingerprinting data and the problem of non-independence among pairwise comparisons. *Molecular Ecology* **5:** 221-227.

Detter J. C., Nguyen Q. A., and Kingsmore S. F. (1998). Identification of novel simple sequence length polymorphisms (SSLPs) in mouse by interspersed repetitive element (IRE)-PCR. *Nucleic Acids Research* **26:** 4091-4092.

Dib C., Faure S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Hazan J., Seboun E., Lathrop M., Gyapay G., Morissette J., and Weissenbach J. (1996). A Comprehensive Genetic Map of the Human Genome Based on 5,264 Microsatellites. *Nature* **380:** 152-154.

Dowling T. E., Moritz C., Palmer J. D., and Rieseberg L. H. (1996). Nucleic Acids III: Analysis of fragments and restriction sites. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), pp. 249-320, Sinauer Associates, Inc., Sunderland, MA.

Fitch W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree. *Systematic Zoology* **20:** 406-416.

Hancock J. M. (1996). Microsatellites and other simple sequences in the evolution of the human genome. *In* "Human Genome Evolution" (M. Jackson, T. Strachan, and G. Dover, Eds.), pp. 191-211, BIOS Scientific Publishers, Oxford.

Hearne C. M., Ghosh S., and Todd J. A. (1992). Microsatellites for Linkage Analysis of Genetic Traits. *Trends in Genetics* **8:** 288-294.

Hildemann W. H., Bigger C. H., and Johnston I. S. (1979). Histoincompatibility Reactions and Allogeneic Polymorphism Among Invertebrates. *Transplantation Proceedings* **11:** 1136-1142.

Jeffreys A. J., Wilson V., and Thein S. L. (1985). Individual-Specific 'fingerprints' of Human DNA. *Nature* **316:** 76-78.

Lynch M. (1990). The Similarity Index and DNA Fingerprint. *Molecular Biology and Evolution* **5:** 584-599.

Lynch M. (1991). Analysis of population genetic structure by DNA fingerprinting. *In* "DNA fingerprinting: approaches and applications" (T. Burke, G. Dolf, A. J. Jeffreys, and R. Wolff, Eds.), pp. 217-229, Birkhauser Verlag, Basel.

Morgante M., and Olivieri A. M. (1993). PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* **3:** 175-182.

Nagaoka T., and Ogihara Y. (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theoretical Applied Genetics* **94:** 597-602.

Nei M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583-590.

Nei M., and Roychoudhury A. K. (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* **76:** 379-390.

Rosenbaum H. C., and Deinard A. S. (1998). Caution before claim: an overview of microsatellite analysis in ecology and evolutionary biology. *In* "Molecular Approaches to Ecology and Evolution" (R. DeSalle, and B. Schierwater, Eds.), Birkhauser Verlag, Basel.

Salter-Cid L., and Bigger C. H. (1991). Alloimmunity in the Gorgonian Coral *Swiftia exserta*. *Biological Bulletin* **181:** 127-134.

Schlotterer C., and Pemberton J. (1998). The use of microsatellites for genetic analysis of natural populations -- a critical review. *In* "Molecular Approaches to Ecology and Evolution" (R. DeSalle, and B. Schierwater, Eds.), Birkhauser Verlag, Basel.

Slatkin M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139:** 457-462.

Smouse P. E., and Chevillon C. (1998). Analytical aspects of population-specific DNA fingerprinting for individuals. *Journal of Heredity* **89:** 413-150.

Swofford D. L. (1998). PAUP*: Phylogenetic analysis using parsimony (and other methods), Sinauer Associates, Sunderland, MA.

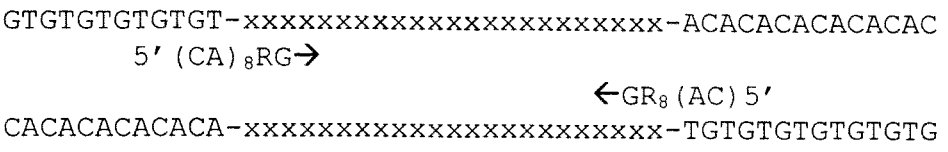
Swofford D. L., and Maddison W. P. (1987). Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* **87:** 199-229.

Swofford D. L., Olsen G. J., Waddell P. J., and Hillis D. M. (1996). Phylogenetic Inference. *In* "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), pp. 407-514, Sinauer Associates, Inc., Sunderland, MA, USA.

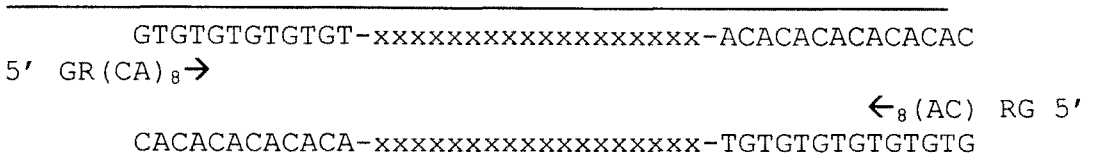
Weber J. L. (1990). Informativeness of Human (dC-dA)n-(dG-dT)n polymorphisms. *Genomics* **7:** 524-530.

Zhivotovsky L. A., and Feldman M. W. (1995). Microsatellite variability and genetic distance. *Proceedings of the National Academy of Science* **92:** 11549-11552.

Zietkiewicz E., Rafalski A., and Labuda D. (1994). Genome Fingerprinting by Simple Sequence Repeat (SSR)- Anchored Polymerase Chain Reaction Amplification. *Genomics* **20:** 176-183.



3'anchored primers bind to their corresponding dinucleotide repeat in opposite orientations and amplify the intervening region.



5' anchored primers bind to their corresponding dinucleotide repeat in opposite orientations and amplify the repeat and the intervening region.

Figure 1. Schematic representation of primers and primer binding sites for 3'-anchored microsatellite primers for the (CA)n repeat (top panel). Lower panel illustrates primer and primer binding sites for 5' anchored microsatellite primers. Intervening sequences in both are indicated by "x".

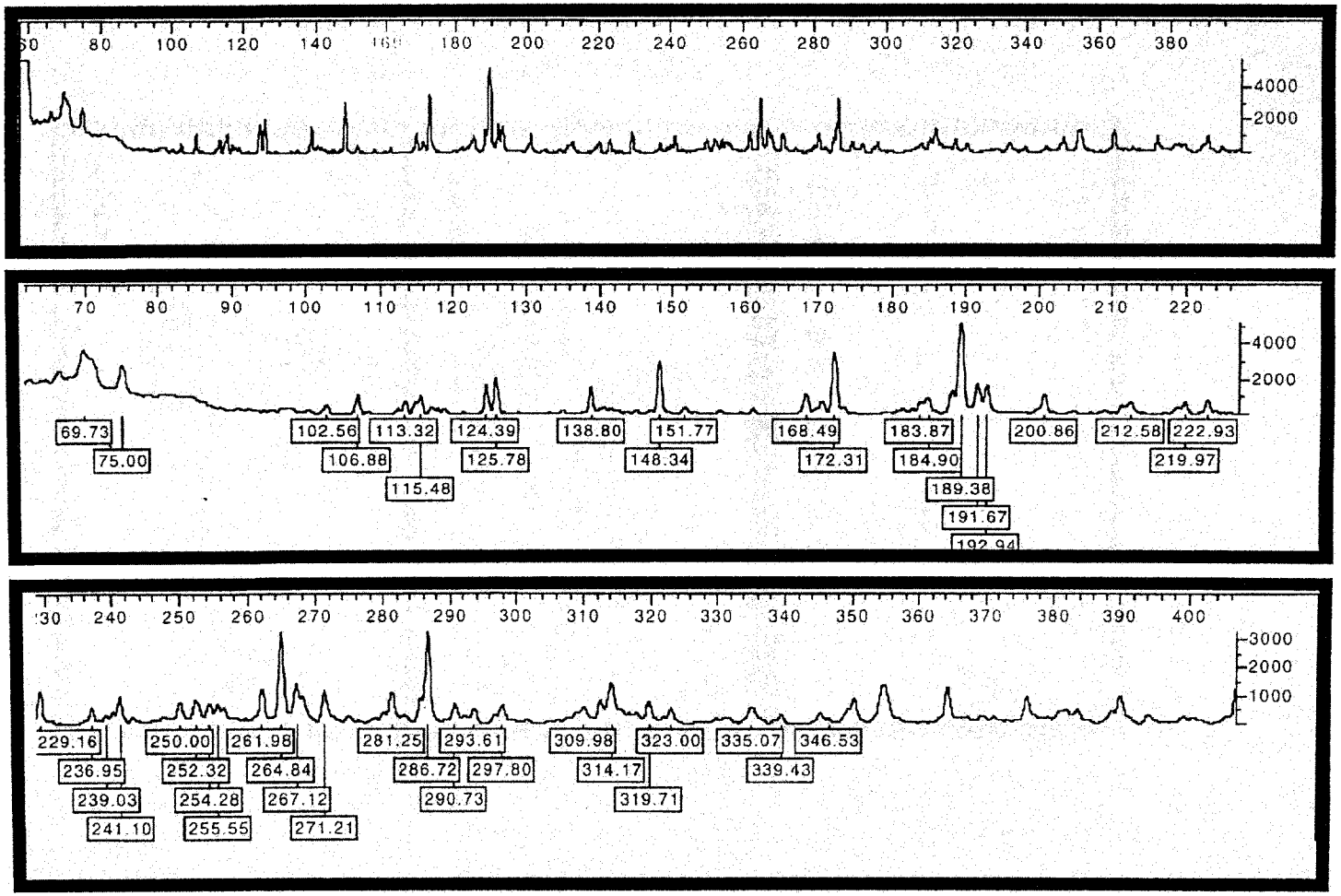


Figure 2. Genotyper output figure illustrating an example of a multilocus microsatellite fingerprint for one animal (top panel). The top panel was split in half and the enlarged regions are represented by the middle and lower panel. All potentially relevant bands or loci are size labeled. To increase resolution of close or overlapping peaks, one can zoom in at smaller window increments (e.g., at 50bp invervals)

Figure 3.

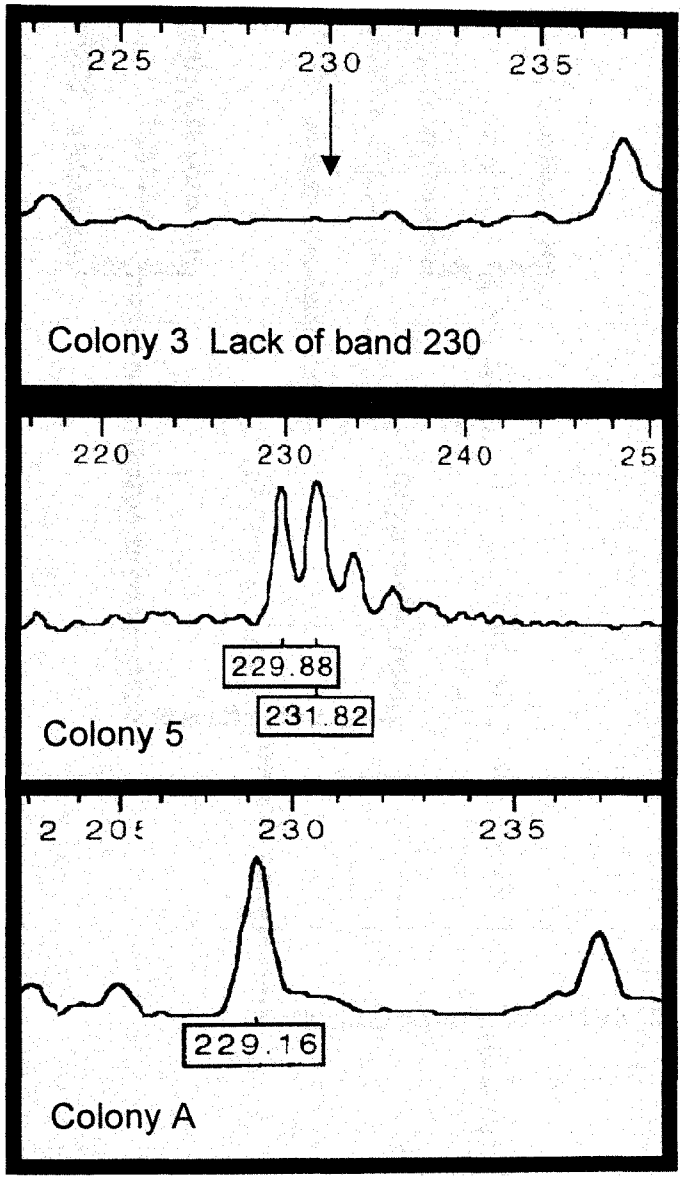


Figure 3. Top panel illustrates an example of a missing locus at position 230, which in the middle and lower panels are positive. Middle panel illustrates a true heterozygote because this pattern was reproducible, and in the lower panel a true homozygote at that position. Note: there is no sequence evidence proving that two bands of the same size from two individuals represent the same locus.

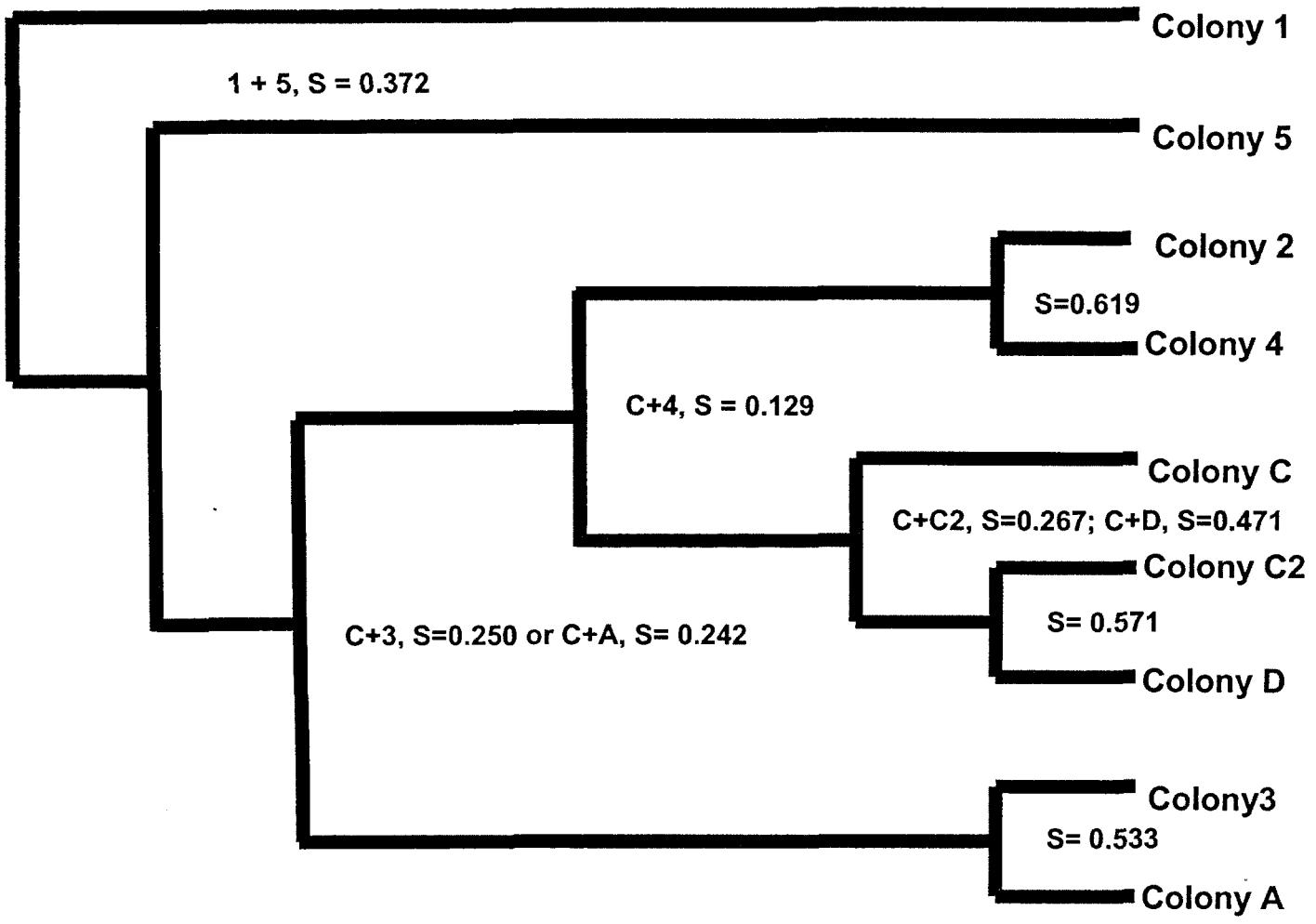


Figure 4. Maximum parsimony-derived genotypic tree illustrating pairwise relationships from band-sharing data. Similarity indices are in support of the branching patterns. Branch lengths not related to genetic distance.

Chapter 7

Final Discussion

In comparative biology, controversy frequently arises when one attempts to make evolutionary references, comparisons, and implications of homology (e.g., Klein, 1989; Klein, 1997). Comparative biology studies are biomedically essential because, in some cases, looking at simpler, diverse representative organisms often elucidates more complex pathways found in humans (e.g., pathways associated with signaling, apoptosis, and some cancers). These results can also contribute to our general understanding of intricate biological pathways and networks. Hence, studies from invertebrates with presumably simpler body and tissue plans may provide a plethora of new discoveries. Unfortunately, much of biology (i.e., comparative immunology) is plagued by biases associated with 'homochauvinisms' (Klein, 1995), which may in turn blind us to revealing the basics of biological phenomena. Despite some criticism about some aspects of invertebrate immunity (Klein, 1989; Marchalonis & Schluter, 1990), studies uninhibited by human biases are warranted (Klein, 1995; Klein, 1997). Comparative immunology, though, continues to suffer from predisposed beliefs and mistaken assumptions that evolution is a linear phenomena with humans at the pinnacle of that process (e.g., Klein, 1997; Liu & Shaw, 2001). Even as 'comparative' immunologists, many of the associated disciplines that are utilized (i.e., molecular biology and molecular evolution) are not completely understood and often misrepresented.

The molecular systematics literature has recently strengthened the argument against intermediate taxa and proposes that, based on molecular and protein data, metazoan phylogeny experienced one major split (Adoutte *et al.*, 2000) into protostomes and deuterostomes (P-D). This is a critical argument because if the recently calculated

time of divergence (Doolittle *et al.*, 1996) is correct, then for at least 670 million years these two lineages have evolved independently. This provides reasoning for studying extant organisms that predate this divergence point in search of common genes, clusters, and domains that have diverged for over a half billion years (see chapter one). In comparative immunology, a significant amount is known about very few animal models. While a few protostome representatives have been studied in depth (i.e., earthworms, some arthropods, and a few molluscs), most work has focused on deuterostomes (echinoderms, tunicates, and vertebrates). A marginal diversity of immunological phenomena has been explored in protostome and deuterostome invertebrates (Arala-Chaves & Sequeira, 2000; Cooper *et al.*, 2002; Gillespie *et al.*, 1997; Gross *et al.*, 1999; Medzhitov & Janeway Jr., 2000; Raftos, 1994; Rinkevich, 1996), yet very little information is known about immune defense strategies from lineages predating the P-D divergence.

We have been interested in the immunological defense strategies of the gorgonian coral, *Swiftia exserta*. Functional and cellular phenomena have been described in this animal (Olano, 1993; Olano & Bigger, 2000; Salter-Cid & Bigger, 1991), but molecular data has been lacking. The purpose of this work was to establish methods to purify and manipulate nucleic acids from this animal and to apply those techniques to the analysis of genes associated with immune defenses. In the past several chapters, the successful application of now routine nucleic acid procedures for the isolation and characterization of genes from *S. exserta*, have been described. The techniques can be routinely and reproducibly applied to Cnidarians and a diverse number of animal models and tissue types.

In chapter 2, 'simple and reproducible nucleic acid extraction procedures' are described. A collection of previously characterized nucleic acid isolation and molecular biology techniques were evaluated and assembled into a series of methods applicable to difficult tissues rich in proteoglycans, polysaccharides, nucleases, and other noxious materials. The methods described to isolate genomic DNA and RNA have been successfully and routinely used in this work for: restriction digestion, membrane blotting (Southerns and Northerns), RT-PCR and cloning, RACE-PCR, DNA sequencing, cDNA library construction, and PCR of genomic DNA (see chapters 2 and 3).

Genomic DNA isolated with the techniques described can also be used for genotyping. This is essential to the development of a fingerprinting technique (chapter 6) that will help to efficiently and reliably estimate genetic relatedness. Described in chapter 6 is a fingerprinting approach, and the associated statistical analyses, based on microsatellite-anchored PCR for multilocus investigation. This approach generates complex banding patterns which are unique to each individual. Assembling of this data into a matrix allows for the calculation of similarity indices, where genetic distance and relatedness can be estimated. The resulting technique is technically simple and reproducible, and with proper validation, will be experimentally applied. We are interested in genetic fingerprinting as a means of determining the effects of genotypic relatedness on allograft recognition and rejection. This will provide clues into the nature of immune specificity in this animal.

Further confirmation of the applicability of the described molecular techniques was the isolation and characterization of components (genes) from two gene families that appear to be conserved throughout phylogeny. Further efforts have been applied to the

isolation of several other immunologically relevant and important gene members, but will not be discussed here.

As the non-complement alpha-2 macroglobulin-like (A2M) paralogs are the only TEPs which have been isolated from protostome invertebrates, it was very unlikely that a complement-like protein would be found outside the deuterostome lineage. Thus, the isolation of a complement C3-like gene (SeC3) from a coral (chapter 4) is a very significant finding since Cnidarians predate the protostome-deuterostome split and complement-like genes (and/or precursors), therefore, appear to exist outside the deuterostomes.

To further understand the origins and evolution of thiolester-containing proteins (TEPs), a coral TEP homologue was pursued (see chapter 4). RT-PCR with degenerate primers produced a product with sequence similarity to vertebrate C3 and A2M proteins. Northern blot analysis showed the gene to be approximately 6kb. The entire open reading frame of 5.8kb was attained with 3' and 5' rapid amplification of cDNA ends (RACE). The deduced polypeptide has been determined to contain a thiolester site, the C3-specific catalytic histidine, an anaphylatoxin region, and two arginine-rich cleavage sites. Hydrophobicity profiling has predicted the conservation of a major receptor binding site which in vertebrates is C-terminal to the anaphylatoxin region. The binding site in mammals is recognized by three complement receptors and two factors (B and H) which are involved in complement regulation. Southern blotting suggests a complex genomic nature for SeC3, a trait conserved in mammalian TEPs. While functional studies are currently underway, physiochemical and structural properties of the deduced polypeptide, along with phylogenetic analysis, indicate it to be a coral C3-like homologue.

Phylogenetic analyses indicate that the ancestor of the paralogous C3, C4, and C5 genes was C3-like. The invertebrate complement-related genes that have been characterized as C3-like are not orthologous to vertebrate C3 but instead represent extant forms of the C3/C4/C5 ancestor, which appear to have C3-like characteristics. Unlike vertebrate A2M and insect TEPs, the coral protein shows significant similarity to the C-terminal region of C3/C4/C5 (a characteristic of these complement forms). A complement-related protein from this level of phylogeny demands a re-evaluation of TEP evolution.

Findings from this work have led us to propose that the ancestral protein to the TEP-family may not have been A2M-like, but instead C3-like and that A2M, alternatively, represents a divergent paralogous gene retained in protostome invertebrates (manuscript in preparation). The protostome ancestor, therefore, may have lost the C3-like paralog through a chromosomal deletion event. One potentially relevant argument in support of the physiological importance of a C3-like gene present in metazoans is the independent assembly of the prophenoloxidase (proPO) activating pathway in the protostome lineage (a lineage which lacks the C3-like paralog). The proPO pathway appears to have evolved independently in protostomes and is an enzymatically controlled pathway which shares some functional similarities with the complement pathways of vertebrates (Cerenius & Soderhall, 1995; Sritunyalucksana & Soderhall, 2000). Of further interest is the fact that the central component of this pathway, prophenoloxidase, is a paralog of hemocyanin and is a thiolester-containing protein that, like C3, can bind covalently to its target (Sritunyalucksana *et al.*, 1999). These paralogs, though, are unrelated to the TEP family of A2M, C3, C4, and C5 described in chapter 4. Functional characterization (Dishaw et al., work in progress) of the C3-like protein in a coral (an

animal that lacks a vascular system or coelomic cavity) should provide intriguing insight into the primordial functionality of the C3/C4/C5 ancestor.

The current work also describes the isolation and characterization of beta proteasome subunit proteins from *Swiftia* (see chapter 5). At least three have been isolated to date, but for the purposes of this manuscript, only one is described. The proteasome, as described earlier (chapter 5), is a multimeric complex of protein subunits that are directly involved in the proteolytic digestion of proteins into small peptides (Kloetzel, 2001; Rivett, 1993; Shastri *et al.*, 2002). Of all the associated subunits, the beta type are on the inner ring of the complex and come into direct contact with the proteins to be digested. This is a normal intra-cellular housekeeping phenomenon in all prokaryotic and eukaryotic organisms (Monaco & Nandi, 1995; Rivett, 1993).

At some point in evolution, a few of these inner beta subunits underwent gene duplication events. The resulting paralogous proteins would assemble into a 'new' type of proteasome, one that would be more efficient at generating antigen for the adaptive immune system via MHC class I display. This is how the 'immunoproteasome' got its name, because it is specialized for efficient antigen production targeted for MHC class I molecules (Belich *et al.*, 1994; Driscoll & Finley, 1992; Kingsbury *et al.*, 2000; Tanaka & Kasahara, 1998). Some of these inner components are known as LMP (low molecular weight polypeptides) molecules, and an example of two are the LMPX and LMP7 paralogs. The former is a housekeeping, constitutive form and the latter is IFN-gamma inducible for integration into the immunoproteasome (Griffin *et al.*, 1998).

The time of divergence of X and 7 has been a controversial issue since molecular evidence appears to indicate the divergence event occurred 600mya (Hughes, 1997), a

time which predates jawed fish divergence by about 150my. This has been difficult to confirm because LMP7 has not been found in agnathans or deuterostome invertebrates. In our interest to understand the phylogeny of these beta subunit proteins, we pursued a homologue of LMPX/7 in *Swiftia* (see chapter 5).

A partial LMPX-like gene fragment was cloned out of the coral using degenerate PCR. The entire gene sequence was then completed with RACE-PCR (see chapter 3 and 5). Phylogenetic analysis confirms it as an LMPX homologue, clustering with and/or as an outgroup to other invertebrate and vertebrate LMPX orthologs. In vertebrates, the major difference in X and 7 lies in the N-terminal propeptide region and in 20-25 residues evenly spaced along the catalytic core. The propeptide is responsible for ensuring proper integration of the subunits into the appropriate proteasome, and the residue difference (within the catalytic core) gives LMP7 its catalytic specificity (Griffin *et al.*, 1998; Kingsbury *et al.*, 2000).

Phylogenetic analyses of LMP X and LMP 7, using maximum likelihood (character based, at DNA level) and distance based methods (on the protein level), were performed to help resolve the phylogenetic relationships of these paralogous genes. Molecular clock tests, using the likelihood ratio test (LRT) approach, suggests that although these two paralogous genes appear to be evolving very slowly they do not appear to be following a molecular clock (constant rates of evolution in all lineages) if analyzed together. If the analysis is performed separately for the two paralogous sets of genes, the molecular clock appears to be preserved in the LMP 7 lineage (demonstrating an important functional constraint from sharks to humans), whereas it is not in the LMP X lineage(s) (see chapter 5). LMP X appears to be evolving at heterogeneous rates

among the various invertebrate and vertebrate lineages, and suggests that there may be less functional constraints than that associated with LMP 7 or that LMP X may be serving more than one function in different lineages of organisms or in different branches of phylogeny.

Hydrophobicity profiling was performed on the N-terminal propeptide (see chapter 5). Analysis showed that human LMPX and LMP7 have specific hydrophobicity differences in the corresponding region that had been shown previously by biochemical manipulation to be catalytically critical (Kingsbury *et al.*, 2000). Hydropathy profiles of agnathan propeptide sequences indicate a shared pattern with human LMP7e2 and not LMPX. The results also indicate that in the invertebrates proteins, some share hydropathy profiles with LMP X propeptides (like coral LMPX) and others share with LMP7e2. These findings make important suggestions about the evolution of LMP X and 7 (see chapter 5). A more complete comprehension of LMP7's origins will come from a re-investigation of agnathans and invertebrates for the existence of paralogous copies of LMPX.

The findings described in this work establish the coral (and Cnidarians) as essential animal models in the study of immune system origins. Studies from these animals, which predate the P-D split, may help unravel the complex nature of the origins and diversification of some immune response genes. The established protocols discussed in this manuscript and the results obtained with them prove that the coral is an appropriate animal model for gene expression analysis and immunophylogenetic studies.

References

Adoutte A., Balavoine G., Lartillot N., Lespinet O., Prud'homme B., and de Rosa R. (2000). Special Feature: The new animal phylogeny: Reliability and implications. *PNAS* **97:** 4453-4456.

Arala-Chaves M., and Sequeira T. (2000). Is there any kind of adaptive immunity in invertebrates? *Aquaculture* **191:** 247-258.

Belich M. P., Glynne R. J., Senger G., Sheer D., and Trowsdale J. (1994). Proteasome components with reciprocal expression to that of the MHC-encoded LMP proteins. *Current Opinion in Biology* **4:** 769-776.

Cerenius L., and Soderhall K. (1995). Crustacean Immunity and Complement; a Premature Comparison? *American Zoologist* **35:** 60-67.

Cooper E. L., Kauschke E., and Cossarizza A. (2002). Digging for innate immunity since Darwin and Metchnikoff. *BioEssays* **24:** 319-333.

Doolittle R. F., Feng D.-F., Tsang S., Cho G., and Little E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271:** 470-477.

Driscoll J., and Finley D. (1992). A controlled breakdown: antigen processing and turnover of viral proteins. *Cell* **68:** 823-825.

Gillespie J. P., Kanost M. R., and Trenczek T. (1997). Biological Mediators of Insect Immunity. *Annual Review of Entomology* **42:** 611-643.

Griffin T. A., Nandi D., Cruz M., Fehling H. J., Kaer L. V., Monaco J. J., and Colbert R. A. (1998). Immunoproteasome assembly: Cooperative incorporation of interferon gamma -inducible subunits. *Journal of Experimental Medicine* **187:** 97-104.

Gross P. A., Al-Sharif W. Z., Clow L. A., and Smith L. A. (1999). Echinoderm immunity and the evolution of the complement system. *Developmental and Comparative Immunology* **23:** 429-442.

Hughes A. L. (1997). Evolution of the proteasome components. *Immunogenetics* **46:** 82-92.

Kingsbury D. J., Griffin T. A., and Colbert R. A. (2000). Novel propeptide function in 20S proteasome assembly influences beta subunit composition. *Journal of Biological Chemistry* **275:** 24156-24162.

Klein J. (1989). Are Invertebrates Capable of Anticipatory Immune Responses? *Scandinavian Journal of Immunology* **29:** 499-505.

Klein J. (1995). Ehrlich and Darwin: Homochauvinism in immunology. *Immunology and Cell Biology* **73:** 103-108.

Klein J. (1997). Homology Between Immune Respones in Vertebrates and Invertebrates: Does it Exist? *Scandanavian Journal of Immunology* **46:** 558-564.

Kloetzel P.-M. (2001). Antigen processing by the proteasome. *Nature Reviews* **2:** 179-187.

Liu Y., and Shaw S. (2001). The human genome: an immuno-centric view of evolutionary strategies. *Trends in Immunology* **22:** 227-229.

Marchalonis J. J., and Schluter S. F. (1990). On the Relevance of Invertebrate Recognition and Defence Mechanisms to the Emergence of the Immune Response of Vertebrates. *Scandanavian Journal of Immunology* **32:** 13-20.

Medzhitov R., and Janeway Jr. C. (2000). Fly immunity: great expectations. *Genome Biology* **1:** 106.1-106.4.

Monaco J. J., and Nandi D. (1995). The genetics of proteasomes and antigen processing. *Annual Review of Genetics* **29:** 729-754.

Olano C. T. (1993). Cellular Aspects of Alloimmunity and Other Responses in the Gorgonian *Swiftia exserta*, Master's Thesis, Florida International University, Miami.

Olano C. T., and Bigger C. H. (2000). Phagocytic activities of the gorgonian coral, Swiftia exserta. *Journal of Invertebrate Pathology* **76:** 176-184.

Raftos D. A. (1994). Allorecognition and Humoral Immunity in Tunicates. *In* "Primordial Immunity: Foundations for the Vertebrate Immune System" (G. Beck, G. S. Habicht, E. L. Cooper, and J. J. Marchalonis, Eds.), New York Academy of Science, New York.

Rinkevich B. (1996). Immune responsiveness in marine invertebrates revisited: the concourse of puzzles. *In* "New Directions in Invertebrate Immunology" (K. Soderhall, S. Iwanaga, and G. R. Vasta, Eds.), pp. 55-90, SOS Publications, Fair Haven, NJ.

Rivett A. J. (1993). Proteasomes: multicatalytic proteinase complexes. *Biochemical Journal* **291:** 1-10.

Salter-Cid L., and Bigger C. H. (1991). Alloimmunity in the Gorgonian Coral *Swiftia exserta. Biological Bulletin* **181:** 127-134.

Shastri N., Schwab S., and Serwold T. (2002). Producing Nature's Gene-Chips: The Generation of Peptides for Display by MHC Class I Molecules. *Annual Review of Immunology* **20:** 463-493.

Sritunyalucksana K., Cerenius L., and Soderhall K. (1999). Molecular cloning and characterization of prophenoloxidase in the black tiger shrimp, *Penaeus monodon. Developmental and Comparative Immunology* **23:** 179-186.

Sritunyalucksana K., and Soderhall K. (2000). The proPO and clotting system in crustaceans. *Aquaculture* **191:** 53-69.

Tanaka K., and Kasahara M. (1998). The MHC class I ligand-generating system: roles of immunoproteasomes and the interferon-gamma-inducible proteasome activator PA28. *Immunological Reviews* **163:** 161-176.

Appendices

Appendix to Chapter 1

Part I:

Biology of the animal

Part II:

Maintenance of the animal

Part I: Biology of Animal

Biology of *Swiftia exserta*

S. exserta is a gorgonian octocoral of the sub-order holaxonia (Phylum: Cnidaria, Class: Anthozoa, Order: Gorgonia). It forms branching colonies composed of a rigid, mostly proteinacious gorgonin, central axis surrounded or enveloped by coenchyme (colonial tissue) and richly studded with polyps (Hyman, 1940). The coenchymal tissue contains gastrodermal tubes (solenia) that connect the gastrovascular cavities of the individual polyps. These solenia also connect to the larger longitudinal canals that run parallel to the axis itself. This creates a network by which nutrients can be shuttled to all cells of the colony. The coenchyme is also rich in calcareous spicules of various sizes and shapes (Goldberg, 2001; West, 1998). Directly on the axial skeleton grows a thin layer of epidermis, that secretes the protein-rich central rod during growth. The outer surface of the coenchymal tissue of the entire colony is also covered by epithelium.

This coral exhibits the typical Cnidarian three layered body plan (epidermis, mesoglea, and gastrodermis). The coenchyme's outer epithelium covers a thick, cell-rich gelatinous mesoglea in which are embedded spicules and through which runs the solenia. This animal is diploblastic (two tissue layer) because the mesoglea is not a true tissue layer. The coenchymal cells are loosely arranged throughout the mesoglea and around the spicules. It is through the mesoglea that amoeboid cells mostly travel. There are several cell types found in the coenchyme in general: epithelial cell, globular granular cell, granular amoebocyte, globular gland cell, cnidocytes, sclerocytes, mesogleal cells, and axial epithelial cells (Olano, 1993).

Phenomenological data on histo(in)compatibility studies

Some of the cells listed above appear to be directly involved in immune responsiveness and/or wound healing, i.e. granular amoebocytes (Bigger & Olano, MS in preparation). Our lab has also provided evidence of specific alloimmunity involving what appears to be a type of immunological "memory" (Salter-Cid & Bigger, 1991). The elaborate cellular components of these responses are currently being studied (Bigger & Olano, 1994; Olano, 1993; Olano & Bigger, 2000; Salter-Cid & Bigger, 1991). Although much information now exists on the nature of allograft rejection and wound healing responses in this animal, the genes and proteins associated with the reactions are unknown.

In arranging grafts (see Part II) to collect RNA for gene-expression studies, it was important to be able to predict, with some confidence, the microscopic phenomena within the involved tissues. This provided a timeline for estimating when the tissue-collection should occur. Studies in the past (Olano, 1993) have established average timepoints for these histological events. These observations indicate that the recognition and rejection of foreign tissue (primary reactions) can be divided into three stages. Stage 1 (recognition) occurs approximately 3-5 days post graft, and although macroscopically appearing normal, at the microscopic level cells begin to cross the graft interface, amoebocytes accumulate, and the epithelium is noticeably disrupted. Stage 2 (cytotoxicity-early events) can also occur within 3-5 days, and in this case spicules are extruded at the interface, mucus is produced, and/or swelling becomes visible at the macroscopic level. Microscopically, the coenchyme begins to decrease in diameter while the cell density at the interface increases with some granular amoebocytes crossing the interface. Stage 3

(cytotoxicity-late events/tissue death) typically occurs between 4-7 days, and includes necrosis at the graft interface (either bi- or unilateral), with the sloughing off of dead (and sometimes *apoptotic*) tissues, previously produced mucus, and spicules. These events can occur quicker or take significantly longer, and appear to be dependent on genotype.

Studies of wound healing (Olano, 1993) have classified the events into 8 stages with complete healing of wounds within an average of 18 days (wound size: 0.5cm). Broadly speaking, healing takes place in three major stages: sealing (1hr) of exposed internal tissue (done in 24hrs), regeneration of coenchyme (up to 5 days), and fusion of regenerating tissues (up to 18 days). In as little as one hour of wounding, exposed tissues are sealed by surrounding epithelial cells, while spicules are extruded. The release of spicules is either serving a defensive purpose or they are being removed to increase cellular mobility and tissue reorganization.

Between 12-24hrs, spicules are no longer evident and granular amoebocytes migrate into the wounded area and seal off the exposed gastrodermal canals. Within 48hrs of the wound, a macroscopically visible layer of tissue is evident over the axial skeleton, with the tip of the healing tissue composed of mostly granular amoebocytes. By 72 hours, a cone-shaped moving front is apparent, with coenchymal cells filling in rapidly behind granular amoebocytes. Then, within 4-5 days, fronts meet and cells begin to mix. Between 5-18 days, the number of spicules again increase, gastrodermal canals fuse, mesoglea becomes continuous, and finally fusion, or healing of the wound, is complete.

Part II: Maintenance of the animal and grafting procedures

Maintenance of animal

Swiftia exserta is a relatively easy animal to work with. *Swiftia* is a deep water coral (taken from 25-30m of water off S.E. Florida) that does not have endosymbiotic zooxanthellae. Therefore, it has evolved to live in deep, cool, and relatively dark waters. Because *Swiftia* is an active feeder of planktonic organisms (carnivorous), it is fed freshly hatched *Artemia* (brine shrimp) at 24-48 hr intervals.

In the attempt to duplicate their environment, salinity (33-35ppt), temperature (19-21°C), and (14/10 hr) light/dark cycles were maintained. Every attempt was made to prevent undo stress on the animals, as stress has been shown to compromise the health of the animals, as well as the experimental outcomes. The healthiest animals generally provide the most consistent results in experiments and hence most experimental manipulations were conducted on fresh coral tissue. New animals were allowed to acclimate for two weeks prior to onset of experiment. If any change in the tank conditions occurred, the animals were allowed to reacclimate before any experimentation.

Grafting methods and considerations

Grafting methods used by our lab have been well described previously (Olano, 1993; Salter-Cid & Bigger, 1991). Though a hardy animal, *Swiftia's* coenchymal tissue is quite delicate and improper handling may cause undo stress or physical damage, which may lead to complete sloughing of the tissue (rapid death). Depending on the applications for which grafts were produced, tissue samples harvested for nucleic acid extraction and molecular applications were never handled bare-handed. This caution

prevented cross contamination of human cellular material onto tissues from which nucleic acids were extracted.

Most of the grafting procedures employed by the lab have involved pair-wise contact of the tissues. This was performed by inserting the tissues into special holders, where opposite pieces contact each other in an "X" fashion producing a small contact/sensitization area. To increase the "contact zone" in this study, parallel pairwise grafts were produced instead.

The process of parallel grafting involved real silk suture (#2-6, with 4 being optimum size) which is inert to the animal and does not contain nucleic acids (as some cotton thread may). Silk suture knots were also easy to untie with forceps. Approximately one inch branch pieces were placed in gentle contact under seawater, in a shallow glass bowl. Gently, with fingers and/or forceps, the small branch pieces were manipulated into optimal orientation so that when tied together they each received maximum tissue contact.

The branches were lightly knotted together in 2-3 locations along the length of the pair. One extra-long piece of suture was used (out of the three) as the suspending line. This allowed the tied branches to be suspended from a floatation device in the aquaria. The pieces suspended in the aquaria during the duration of the experiments, in the same conditions as other non-experimental animals (which included normal feeding routines). Collection of tissues was usually performed at least 12 hrs after the animals were scheduled to eat to reduce the possibility of cross-contamination with *Artemia* nucleic acids (in the coral gastrovascular cavity).

After the experimental time period, the entire paired piece was harvested and either directly embedded in paraffin for sectioning (untying was not necessary) or directly pulverized in liquid nitrogen for RNA extraction. Untying was only performed in instances where RNA was to be harvested from the tissues separately (differential gene expression analysis). For all gene sequences recovered, gene-specific primers were designed and tested against aquarium seawater and *Artemia* cDNA and genomic DNA to confirm lack of contamination.

References:

Bigger C. H., and Olano C. T. (1994). Allo-immune Responses of the Gorgonian Coral, *Swiftia exserta*. *Developmental and Comparative Immunology* **18:** S111.

Goldberg W. M. (2001). The schlerites and geographic distribution of the gorgonian Swiftia exserta (Coelenterata: Octocorallia: Holaxonia). *Bullentin of the Biological Society of Washington* **10:** 100-109.

Hyman L. H. (1940). "The Invertebrates: Protozoa through Ctenophora," McGraw-Hill Book Co., New York.

Olano C. T. (1993). Cellular Aspects of Alloimmunity and Other Responses in the Gorgonian *Swiftia exserta*, Master's Thesis, Florida International University, Miami.

Olano C. T., and Bigger C. H. (2000). Phagocytic activities of the gorgonian coral, Swiftia exserta. *Journal of Invertebrate Pathology* **76:** 176-184.

Salter-Cid L., and Bigger C. H. (1991). Alloimmunity in the Gorgonian Coral *Swiftia exserta*. *Biological Bulletin* **181:** 127-134.

West J. M. (1998). The dual role of sclerites in a gorgonian coral: conflicting functions of support and defence. *Evolutionary Ecology* **12:** 803-821.

Appendix to Chapter 3

Detailed protocols described

RNA Extraction using TriReagent

Materials

1.5ml RNase-free tubes with fitting pestles
or ceramic mortar/pestle and liquid nitrogen
TriReagent (Molecular Research Center)
Chloroform or BCP (Bromochloroporpane)
Isopropanol
1.2M NaCl / 0.8M NaCitrate
70% Ethanol
DEPC-treated ultra pure water
Fresh tissue or tissue fixed in RNAlater (Ambion).

Methods

1) Homogenize about a 1cm piece of fresh tissue (or tissue fixed in RNAlater [Ambion]) from *Swiftia* in 1ml of TriReagent. This can be scaled up for larger sample sizes ground in ceramic mortal and pestle under liquid nitrogen. After homogenization and addition of larger volumes of TriReagent, the solution can be distributed into 1ml aliquots and the procedure followed as normal (see below).
2) Allow homogenate to sit at room temperature for 10 min to allow complete dissociation of proteins from the nucleic acids.
3) Spin down debris for 10min at 16000xg and transfer liquid phase (homogenate) by decanting.
4) Add 100ul of BCP or 200µl of Chloroform per ml of homogenate. I prefer BCP because we get cleaner phase separation and less DNA carry-over. Mix vigorously for 15seconds and allow to sit at room temperature for 15min.
5) Spin at full speed (12-16000xg) for 15min for optimal phase separation.
6) Transfer aqueous phase (by pipetting) into a new tube. All tubes and subsequent handling of homogenate and RNA should be with special attention to maintaining a sterile work area. See Molecular Cloning appendix (Sambrook, 2001) for instructions on how to maintain an RNase-free work environment.
7) Precipitate RNA from homogenate under high salt conditions to minimize polysaccharide co-precipitation. This is done by adding 200µl of room temperature isopropanol and 200µl of salt solution (0.8M NaCitrate/1.2M NaCl made in DEPC-water). Mix by inversion and allow RNA to precipitate at room temperature for 10-15min. Do not put on ice or in freezer because cold temperatures encourage the co-precipitation of contaminants.
8) Pellet RNA by spinning at 12000xg for 10min at 4°C or room temperature. Spinning at 4°C is a good idea because it keeps the tubes from warming up too much during the spin.

9) Decant supernatant and wash RNA pellet with ice-cold 70% ethanol (made with DEPC-treated water). Wash several times until the ethanol is no longer pink. Spin for a few minutes after each wash if the pellet dislodges from wall of tube.
10) Completely remove ethanol with pipette tip and quick spin to recover residual ethanol and remove as well. Allow to dry at room temperature, with lids open but covered with KimWipes (to keep dust out), for about 15min.
11) Resuspend pellet in 20µl of ultra-pure DEPC-treated water. Warm at 65°C for 15min to assure complete solubilization of RNA. Remove an aliquot for quantification.

Genomic DNA extraction

Materials

1.5ml tubes with fitting pestles.
or ceramic mortar/pestle and liquid nitrogen
DNAzol (Guanidine based extraction buffer, MRC)
Polyvinylpyrrolidone (PVP)
2-mercaptoethanol (2-ME)
Phenol:Chloroform:Isoamyl Alchohol (25:24:1)
Chloroform
Absolute ethanol (RT)
70% ethanol (cold)
Ultra pure water

Methods

The method described here is much longer than the standard method recommended for DNAzol extractions (manufacture's instructions, 1hr procedure). The standard procedure works for *Swiftia*, but the resulting DNA is not very clean. Much contamination co-precipitates, so the following method has been developed which produces very consistently clean DNA for Southern blotting, restriction digestion, and PCR. On some occasions, as all persons working with *Swiftia* will learn, if the animal is producing extra mucus or other noxious substances the extraction may not be as clean as desired. Very little can be done about that, except repeating the procedure with fresh tissue.

1) Prepare about 10ml of DNAzol with 2% PVP and heat for 10min at 65°C to bring into solution. Add 200µl of 2-ME and 150ul of Proteinase K (20mg/ml stock) after DNAzol mixture cools down. This will be the DNAzol homogenizing solution. Good for 1week at 4°C.
2) Homogenize approximately a 1cm piece of *Swiftia* tissue in 1ml of DNAzol with a plastic pestle in a 1.5ml tube. This can be scaled up for more tissue and ground (to powderize) in ceramic mortar and pestle under liquid nitrogen. I prefer the latter scaled-up version because more DNA results from one extraction, and all of it comes from the same homogenization reaction. The powerized tissue also dissolves nicely in the DNAzol.
3) Allow the homogenized tissue to rock at room temperature for 20-30min.
4) Extract the 1ml fractions with 500µl of phenol:chloroform:isoamyl (25:24:1). Mix well and allow to sit for 10min at RT. Spin at full speed for 10min to separate phases. If the interface is not tight and the phases not distinct, continue to spin for an extra 5-10min.
5) Transfer the aqueous phase to new tube. Depending on how thick the interface is, I sometimes repeat step 4 a second time. The presence of even a slight interface after the second extraction is a good indication that the extraction was necessary. Avoid any contamination with the interface, especially with the lipids and fat that dance around into the aqueous. If necessary, sometimes the second extraction is what pulls that material out and avoiding it is difficult unless the pipette tip is put in through the aqueous and into the organic. Pulling out most of the organic phase makes it much easier to pipette off the top, aqueous phase. A respin will be necessary to re-tighten the interface.

6) Transfer the aqueous to a new tube and extract with chloroform. After spinning, the interface should be clear. Transfer out the aqueous into a new tube which contains 500ul of RT ethanol.
7) Mix well by inversion (never vortex genomic DNA, which will shear), and store at RT for 10min to assure precipitation.
8) Spin to pellet the DNA at 5000-8000xg for 5min. Do not spin longer or at higher speeds. Higher speeds will spin down many more contaminants and pack the DNA so tight that it will not go into solution very well.
9) Remove supernatant by decanting. DNA pellet may not be visible until washed with 70% ethanol.
10) Add cold 70% ethanol and wash by vigorous inversion. Repeat 2-3x, leaving the last wash overnight if desired. This is a good time to stop, and it allows the DNA to clean overnight in 70%. Store at 4°C if doing this.
11) Spin to assure that pellet is bound to tube, and remove ethanol. Respin and remove all the ethanol with a pipette tip.
12) Allow to dry at RT, up to 30min if necessary.
13) Resuspend in 50µl of water. Add 5µl of 10mg/ml RNase A solution and heat at 50-65°C for 20min. This step assures that the DNA goes into solution while the contaminating RNA is destroyed.
14) After step 13, extract with Ph:ch (50µl). This extraction is much cleaner and only needs to be performed once. Allow to sit at room temperature for 10-15min and then spin at full speed for 2-5min. Transfer aqueous phase to new tube and extract the same way with 50µl of chloroform.
15) Combine all tubes of DNA (from same animal only), or precipitate separately.
16) Add 1/2vol of 7.5M Ammonium Acetate and 2.5vol of cold absolute ethanol. Ammonium produces much cleaner precipitations that NaAcetate. Mix by inversion. If DNA strands do not become visible, freeze for 20min at -80° or 1-2hrs at –20 °C.
17) Spin down the DNA. Dry. Resuspend. Warm for 15-20min at 65° to assure solubilization. Quantitate.

→ For previously resuspended DNA, which one suspects is still contaminated with polysaccharides, follow CTAB purification procedure described in Chapter 2. Note, this does not always work. There are some substances that once they co-purify, appear to bind to the DNA in an irreversible manner. This is rare, but does appear to happen to difficult tissues such as pancreas, liver, muscle, and plant material (see literature).

Protocol for generation of cDNA

Most reactions in this lab are performed using Superscript II or its thermostable derivative, Thermoscript (Invitrogen). It is suggested that the protocols that come with the enzyme be followed, where the user can adjust the amount of RNA used, the type (total vs. mRNA), how it should be treated, and what primers to use to prime the reverse transcription reaction. In this reaction, the mRNA is reverse transcribed into cDNA with the MMLV-RNase H minus recombinant enzyme. This enzyme has been modified so that it lacks RNase H activity, resulting in longer cDNA transcripts. The overall procedure is as follows:

1) An RNase-free environment is absolutely essential. Wash gloves frequently in 95% ethanol and use RnaseZap (Ambion) on gloves, pipetter, and work area. See Molecular Cloning (Sambrook, 2001) manual on how to maintain such a work place. Run gel of RNA and verify integrity. Good cDNA reactions are completely dependent on the integrity and purity of starting RNA.
2) Use up to 5ug of total RNA or less of mRNA, 1ul of 100uM degenerate antisense primer, or 1ul of 20μM gene specific primer, or 1μl of 5-10μM Oligo-dT primer (or 100-200ng of Random Hexamers).
3) Add 1μl of 10mM dNTPs (a mixture of all four).
4) Heat at 80°C for 5min. Quench on ice-water bath immediately, and do not move tube out for 2min. Cold quench can best be done by stabbing the tube directly into a block of dry ice.
5) Then add the rest of the materials, so that the final volume is 20μl.
6) Add 4μl of 5x buffer (comes with enzyme). Sometimes this buffer has the DTT (dithiotheritol) in it. Otherwise, add 1ul of 0.1M DTT. Then add 1ul of RNase inhibitor enzyme (various merchants), and 1ul of RT enzyme (DEPC-treated water to 20ul final volume). The enzyme is added last, after the reaction mixture is mixed.
7) Either do the RT reaction in a thermocycler with a heated lid, or overlay with mineral oil, place in a water bath and avoid any change in volume due to evaporation. The instructions for Superscript II suggest 42°C for 1hr. I like 42°C for 1hr, 50C° for 15min, and 60°C for 15min (or some derivative of this). The enzyme is denatured at 85°C for 5min.
8) For 5' RACE using gene-specific internal primers, or for difficult templates, the RT reaction is performed with Thermoscript at 65°C. DMSO, or other ingredients commonly used in PCR for difficult templates, cannot be used in RT reactions because they inhibit the RT enzyme. As does any residual ethanol not removed from the RNA pellet.
9) After the RT reaction, 1μl of RNase H is added and incubated for 20min at 37°C to nick the RNA. This makes second strand synthesis more efficient, or the initial extension in PCR, especially for long templates. Dilute the reaction to 50μl (optional if the gene you seek is expressed at low levels). Use 1-2μl directly as template in PCR (this amount is also dependent on an empirical knowledge of how common the gene's expression is under the appropriate conditions; how much mRNA do you expect to be present?)

Standard PCR protocols

All new users of PCR should read an introductory chapter now usually found in any molecular methods manual (see Molecular Cloning, Sambrook and Russell, 2001) for important background information on requirements of the reaction (i.e., $MgCl_2$), binding kinetics of primers, types of enzymes to use, and the importance of using hot start technology. Many of these will come with experience, but a user with some basic knowledge of PCR will become much better at trouble-shooting reactions.

Materials

0.2ml thin walled PCR tubes
10x PCR buffer (comes with enzyme)
25mM $MgCl_2$
10mM dNTPs (mixture of the four)
10-20μM Primers
PCR quality ultra pure water
Taq Polymerase (I have used various Taq enzymes with equal success)

I have used Qiagen, Promega, AmpliTaq (PE), Pfx derivatives, and Biolase (BioLine, Midwest Scientific).All work very well. My preferred Taq is Qiagen, albeit, it is expensive but very robust and easily gives consistently long amplifications. Pfx derivatives are great for long templates and has proofreading capabilities, so artifacts are virtually eliminated. Each enzyme requires its own buffers because some have special pH requirements or work best in the presence of a mild detergent (like Triton X-100).

Methods

1) Produce a master mix for 20 reactions. This will give enough mix for about 19 reactions (because of pipetting error). Determine how much of each will be needed for one PCR tube and multiply by twenty. The master mix will usually contain the buffer, $MgCl_2$, dNTPs, Taq and water. I usually do PCR reactions in a 50μl total volume and calculate so that I am distributing 40μl of the master mix into each tube. The added template, primers, and PCR water to 50μl total.
2) Each tube will contain a final concentration of 1x buffer, 200μM dNTPs, 10-20 pmol of each primer (1μl of 10-20μM primer), 1.5-2.0 mM $MgCl_2$, and about 1-1.5U of Taq.
3) The standard reaction conditions are an initial melting stage (for about 5min), then a cycling program consisting of about 30 cycles (melt, anneal, and extend). The standard melting temperature is 94-95°C, 55-65°C for the annealing temperature, and 72°C for the extension. After the cycling program ends, there is always a final 5-15min extension step required for Taq to run through and make sure all double stranded products are complete and not staggered. This also assures the addition of the non-template 3' dATP to the products, which is a critical element required for TA-cloning technology.
4) The typical times are: initial melting for 2-5min, and then in the cycling program each step is set at 30s-1min. The extension step of the cycling program can be set at 1-3min for longer templates because in PCR Taq extends about 1kb per min and takes longer as the dNTPs and primers run out.
5) The annealing temperature is dependent on the melting temperature of the primer (Tm) and the best results are attained by using an annealing temperature that is 5-10°C below the Tm of the primers. Sometimes imperfect matches of the primers

will incorporate and amplify non-specific templates and one may have to adjust the annealing temperature to a few degrees below the Tm to eliminate such artifacts. The invention of thermocyclers that have a gradient temperature function are great for determining the optimal annealing temperature as twelve different temperatures can be tested in one reaction.

6) All reactions should be set up on ice. The programmed thermocycler is run and allowed to reach the initial melting temperature and paused. The tubes are put from ice directly into the 95°C and the machine is then unpaused. This is called modified hot start (the alternative is to use commercial hotstart enzymes which do not function until they have been thermally activated). It removes all non-specifically bound primers from the template, melts the template into single stranded products, and then allows the primers to bind in their appropriate template-specific locations.

Degenerate PCR

--optimized from personal experience

Materials

100μM degenerate primers
Master mix made with 400uM dNTPs
Ultra pure PCR water
Clean hood, laminar flow hood

Methods

1) Degenerate PCR, because it is done at such low annealing temperatures, should be prepared in a sterile environment to prevent the amplification of non-specifc contaminating products.
2) Set up a standard PCR reaction but add 1-2ul of 100μM each primer. This is a high concentration, but necessary because the target primer is actually at a very low concentration in highly degenerate primer mixtures.
3) The template should be cDNA that was primed with the antisense degenerate primer. This limits the possible number of non-specific products that can be amplified compare to cDNAs primed with oligo-dT.
4) The PCR cycling program should run about 40-45 cycles because the degenerate PCR produced product is usually a light band, and may be hidden behind smeary background.
5) If no product arises from the degenerate PCR, try different cDNAs, newly made cDNAs from new RNA, cDNAs primed from oligo-dT, and RNA from induced tissues (the gene sought may be down regulated). As a last resort, try amplifying under higher Mg concentrations which encourage imperfect matches.
6) Well designed degenerate PCR primes, along with performing the PCR reaction under a gradient annealing program (37-55°C) will aid in producing a clean strong product. If a homologue exists (or any sequence sharing significant sequence similarity), it will be picked up, even if considerable divergence has occurred.
7) The cycling parameters should be set at 1min for each step.

Rapid Amplification of cDNA Ends (RACE)-PCR

--"Classic RACE" protocol, derived from the original Frohman technique (Zhang and Frohman, 1997).

Materials

Standard Master Mix, described above
Race Primers: Qt (at 20μM and 5μM), Q1 and Q0 each at 20μM
Qt— 5' CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC(T17)-3
Q0—5' CCAGTGAGCAGAGTGACG
Q1—5' GAGGACTCGAGCTCAAGC
Gene-specific primers (sense and antisense, each at 20uM)
Freshly prepared cDNAs
7.5M Ammonium Acetate
100% and 70% Ethanol
TdT enzyme, 5x buffer, and 1mM dATP
DMSO
Ultra pure PCR water

Methods

3'RACE

1) Make cDNAs with Qt primers (1μl of 20μM Qt), using 5μg of total RNA or 2μg of mRNA if available. Make sure to RNase H the reaction.
2) Depending on prevalence of transcript of interest, dilute cDNAs to 50ul or leave at 20μl. Use 1-2μl as template for PCR.
3) Since the cDNAs were primed with Qt primers, the PCR reaction can be performed using a gene-specific sense primer and Q0 (Q1 can be used for a nested reaction). The first step of RACE is second strand extension to produce double stranded cDNAs as template for PCR.
4) The program for PCR will include a second strand extension before cycling begins. This will be: 95° for 5min (initial melting), 50-60°C for 2min (annealing of sense primer to cDNAs), and extend at 72°C for 15-30min. Then cycling begins (note that Q0 primer is present during first strand extension, but is not supposed to doing anything). With 3' RACE on *Swiftia*, I have gotten non-specific amplification of other cDNAs. This apparently happens because there are cDNAs that are recognized by the Q0 and Q1 primers during 2nd strand extension. This produces templates for PCR because we get double stranded products with Q0 priming sites at both ends. This can be reduced by leaving Q0 out of the tubes until cycling begins (adding after 5th cycle) so that second strand synthesis only occurs on the correct template by the gene-specific sense primers.
5) The PCR cycling program is then performed at a high temperature because the Tm of the Q0/Q1 primers is above 70°C. As long as the gene-specific primer's Tm is similarly high, the annealing can be set to 65-68°C. This first round of PCR is called RACE1. During PCR, one may use DMSO (0.5-3% reaction volume; I have found that 1% works in most cases) to relax secondary structure and allow long difficult targets to amplify.
6) After RACE1, a second round of PCR is performed. This is nested PCR and is critical to perform because non-specific products are not uncommon in RACE. So by performing a second round of PCR with a nested (internal to the last

primer) gene-specific primer and Q1 as the other nested primer, only true RACE products will result because only they have the internal gene-specific sequence to bind the nested primer. To do this reaction, take 1µl of RACE1 and dilute 1:50 and use 1µl as template in RACE2 with nested primers. RACE2 does not include the 2nd strand synthesis, but is a regular PCR cycle. Many of the rare gene products become visible in the second round of RACE, after receiving two rounds of exponential amplification.

7) Separate products on a 2-2.5 % TAE gel with a 1kb ladder. This assures clean separation of bands, if more than one is produced.

5'RACE

1) 5'RACE is very similar to 3' RACE and also uses two nested reactions, RACE1 and 2. There are some important differences to consider. In 5'RACE, we are extending cDNAs into unknown territory and have no information regarding the 5' end of our gene of interest. Therefore, there is no priming site to use during PCR so we must add some type of binding region to the 3' end of extended cDNAs.
2) For difficult templates and long 5' gene-specific regions that one may be interested in, the RT reaction is performed with Thermoscript (Invitrogen) to relax secondary structure and allow the enzyme to extend to the end of the gene. If secondary structure issues are a concern and Thermoscript was used for this purpose, then it is important to address the same concerns during PCR because the amplification may suffer the same consequences otherwise. During PCR, one may use DMSO (0.5-3% reaction volume; I have found that 1% works in most cases) to relax secondary structure and allow long difficult targets to amplify.
3) The cDNA RT reaction is performed with 1µl of 20uM gene-specific antisense primer, 1µl of Thermoscript and extended for 1hr at 65°C, increasing the temperature gradually to 72° over the last 15-20min of the reaction.
4) RNase H digest the reaction at 37°C for 20min. Precipitate the cDNAs with ammonium acetate and ethanol. Freeze for 30min at -80° and spin down pellet. Wash well with 70% ethanol. Dry pellet.
5) Resuspend the cDNAs in 11µl of ultra pure water. Add 4µl of 5x TdT buffer, 4µl of 1mM dATP, and 1µl of TdT enzyme. Polyadenylate the 3'ends of the cDNA at 37°C for 10min. Stop the reaction by incubating at 75°C for 5min. Add water to 50µl and this is your polyadenylated 5'RACE cDNAs ready for PCR.
6) Use 1-2µl of the cDNA in RACE1 PCR reaction.
   Set up as follows: 1µl of gene-specific antisense primers, 1µl of Q0 and 1µl of Qt(5µM). The Qt will be the primer to bind initially to the polyadenylated cDNAs and extend in the 2nd strand synthesis reaction. It is 4x less concentrated so that it does not interfere in subsequent cycling (runs out faster).

7) Perform 1st strand synthesis in thermocyler as before but this time the annealing temperature is at 52°C (because Qt has lower Tm)→ 95° for 5min, 52° for 2min, and 72° for up to 30min. Go directly into cycling program of 95° for 30s, 60-68° for 1min, and 72° for 2-3min (most 5'race products expected are large). Use DMSO at 1% in all PCR reactions expected to contain long products or difficult templates. After RACE1-PCR, dilute 1µl 1:50 as before and use 1µl in RACE2-PCR with nested primers.

Northern transfer of RNA to a positively charged nylon membrane

--- from Molecular Cloning, 3rd Ed, Sambrook and Russell, 2001

Prior to setting up gel for transfer, total (or mRNA) RNA is prepared by quantitating and loading into 1% MOPS-buffer-based formaldehyde gel. This assures that the RNA is run under denaturing conditions. Protocols for casting a denaturing formaldehyde gel can also be found in Molecular Cloning, 3rd edition.

Materials

Methylene blue soln (for staining membrane), 0.02% in 0.3M NaAcetate pH 5.5
Soaking soln (0.01M NaOH/3M NaCl)
0.2x SSC with 1% SDS
20x SSC
Transfer buffer (0.01M NaCl/ 3M NaCl)

Methods

1) Hydrolyze the RNA in the formaldehyde gel as such: rinse the gel in DEPC-H20. Soak for 20min in 5 gel volumes of 0.01M NaOH/ 3M NaCl.
2) Transfer gel into 10 gel volumes of 20x SSC for 40min
3) Trim the gel, cut the left hand corner of the gel.
4) Set up apparatus, see MolCloning book (similar to Southern transfer).
5) Fill dish with transfer buffer.
6) Immerse nylon membrane in DI water and then soak for 5min in 10xSSC. Cut the corner of the membrane to match the corner of the gel.
7) Set up gel, upside down, and apply the membrane. Stack the papers 5-8cm tall.
8) Allow the transfer to occur for 2 hours. Do not allow it to exceed 2hrs, as will lead to high background staining especially if RNA probes are used.
9) Transfer the membrane to 300ml of 6xSSC and agitate for 5min.
10) Drain the membrane and allow to sit on a dry sheet of blotting paper for a few minutes.
11) Stain membrane if desired in methylene blue soln until rRNAs can be seen (3-5min). Mark at the edge of the membrane the positions of the rRNA with a ball point pen.
12) Destain the membrane in 0.2xSSC/1% SDS for 15min.
13) With positively charged nylon, the RNA does not need to be fixed with UV—only if neutral transfer used, and we do not do this here).
14) Proceed with prehybridization. If not, dry membrane, wrap in aluminum foil and store in zip-lock bag in the fridge (4°C).

Southern Transfer of Genomic DNA to positively charged nylon membranes

---from Molecular Cloning, 3rd ed. Sambrook and Russell, 2001.

Prior to Southern transfer, 2.5-5μg of RNA-free Genomic DNA is digested to completion by digesting in the presence of an appropriate restriction enzyme (under manufacture's recommendations) for 48hrs. Only digestion above 24hrs can assure complete digestion, especially of genomic DNA. After digestion, the genomic DNA is separated (at 5v/cm) on a 0.7% TAE agarose gel.

Materials

Alkaline Transfer buffer
- 0.4N NaOH
- 1M NaCl

Depurination of DNA
- 0.2N HCl

Neutralization buffer (alkaline transfer to nylon)
- 0.5M Tris-Cl (pH 7.2)
- 1M NaCl

Methods

1) After electrophoresis in 0.7% TAE agarose gel, depurinate by submerging gel into the 0.2N HCl for several minutes till the bromophenol blue turns yellow. Immediately rinse gel in DI water.
2) Soak gel for 15min at RT in several volumes of alkaline transfer buffer with gentle agitation. Change solution and continue for another 20 min.
3) Cut bottom right side of the gel (corner) off for orientation and remove area above the wells, area where a DNA ladder was run and any extra gel areas with no DNA.
4) Float the nylon membrane in DI water till it saturates and then place in transfer buffer for a few minutes.
5) Prepare transfer apparatus as usually for capillary transfer (see Molecular Cloning 3rd edition).
6) After transfer for 8-24hrs, soak membrane in neutralization buffer (15min) and proceed to prehybridization or cross link in the Stratalinker (optional).

Generating probes

The simplest and most popular method to produce radioactively-labeled probe (at high activity) is the Random Priming approach. This method works wonderfully, produces very little background, and is a very quick labeling reaction (less than 30min total). I use the MegaPrime Labeling System (Amersham BioSciences, cat # RPN1604). This protocol utilizes the Klenow fragment which extends the randomly bound primers from a few dozen to a few hundred bases.

1) Generate a PCR product to use a probe. Run the PCR reaction out in a 2% gel and Gel purify (Qiagen gel extraction kit) the PCR product. Then follow manufactures instructions for random priming the template.
2) In short, 5ng of PCR product is used, along with each dNTP-- except the one corresponding to the radioactive one of choice. Add enzyme and allow primers to bind at RT for 10min. Then, add 5µl of α-^{32}P-dCTP and incubate for 10min at 37°C.
3) After the reaction, remove the unincorporated nucleotides by running product through a G50 spin column (ProbeQuant G-50, Amersham Biosciences, cat # 27-5335-01). Only takes 2min total.
4) Within about 30min, one can go from PCR product to purified radioactively labeled probe.

The other method that I have used to generate probes consists of making RNA probes by run-off transcription. This method can be used to make radioactive RNA probes by using α-^{32}-P -rNTP or Dig-labeled probes by using Dig-labeled rNTPs to incorporate into the RNA. Both probes are made the same way. It consists of a cloned PCR product into a vector that contains T7 and SP6 (or T3) promoter sites. After sequencing and determining which strand is antisense, the vector is digested with the appropriate restriction enzyme to cut immediately after the PCR product opposite the side of the polymerase binding site which produces the antisene strand.

1) Use appropriate polymerase and the cut vector (RE digested) as template. Follow manufacture's instructions on performing the transcription reaction using the correct polymerase. Essentially, use template, 10x transcription buffer, rNTPs (either radioactive or Dig-labeled), polymerase and incubate for 2hrs at 37°C.
2) Stop reaction at 65°C 5min and destroy template with RNase-free DNase for 20min at 37°C.
3) Dig-labeled probes can be cleaned with phenol:chlorofrom or directly precipitated with 4M LiCl and ethanol. I prefer to use ammonium acetate and ethanol. Precipitate, spin, and wash pellet with 70% ethanol. Resuspend in DEPC water and use immediately or keep at -80°C.
4) Radioactively labeled probes are run through the G-50 columns described above (much safer and quicker than precipitating).
5) Run 1-2µl of probe through agarose gel (only for non-radioactive probes) and determine integrity of probe and relative abundance (make sure to run a standard, of known amount of DNA).
6) RNA probe is ready to be used in blot hybridizations or *in situ* hybridization reactions.

Method for Southern Hybridization

Using random prime labeled double stranded DNA as probe
---from Molecular Cloning 3rd ed.; Sambrook and Russell, 2001

Materials

Phosphate-SDS buffer (used as prehybridization and hybridization solution)
- 0.5M NaPO4 (pH 7.2)
- 1mM EDTA (pH 8)
- 7% SDS (w/v)
- 1% (w/v) BSA

Phosphate –SDS solution 1
- 40mM NaPO4 Buffer (pH7.2)
- 1mM EDTA (pH 8)
- 5% SDS and 0.5% Fraction 5 grade BSA.

Phosphate –SDS solution 2
- 40mM NaPO4 buffer (pH 7.2)
- 1mM EDTA (pH8) and 1% SDS

Methods

1) Soak the nylon membrane (with bound DNA) in 6xSSC for 2-5min.
2) Prehybridize membrane in roller bottle in preheated oven at 65°C for 1-2hrs.
3) If the radiolabeled probe is double-stranded DNA, denature it by heating for 5min at 100°C and quickly chill on ice water bath (ice water slushy, not ice).
4) Pour out prehyb buffer and to the hyb buffer, add the probe, mix and add to bottle.
5) Hybridize overnight at 65°C.
6) Remove membrane from bottle, place it in a tray with several hundred mls of PO4/SDS soln 1 at 65°C. Agitate the tray and repeat once more.
7) After 5min, pour off and rinse in PO4-SDS soln 2 for 5min each time, **8x**.
8) Blot membrane on paper towels and wrap in Saran Wrap and expose to X-ray film for 16-24hrs at -70°C

Stipping probes from the membrane- first consult the membrane manufacturer's suggestions or do the following:
0.4M NaOH for 30min at 42°C and then wash in 0.1xSSC/0.1%SDS/0.2M Tris-Cl (pH7.6) for 30 min at same temperature. Check membrane with hand-held counter, and re-expose if necessary to verify the membranes have been stripped.

Methods for Northern Hybridization

Using random-prime labeled double stranded DNA as probe
---from Molecular Cloning 3rd ed.; Sambrook and Russell, 2001

Materials

Prehyb buffer

0.5M NaPO4 (pH 7.2)
7% SDS
1mM EDTA

Methods

1) Incubate the membrane for 2hrs at 68°C in 10-20ml of prehyb buffer.
2) Denature the probe at 100°C for 5min and chill on ice-water bath.
3) Add the denatured probe to the prehyb buffer and continue incubation for 12-16hrs.
4) After hybridization, remove membrane and place in box containing 100-200ml of 1xSSC/0.1% SDS at room temperature. Place on a platform shaker and agitate for 10min.
5) Transfer the membrane to another container with 100-200ml of 0.5xSSC/0.1%SDS prewarmed to 68°C. Agitate gently for 10min at 68°C preferably back in the oven.
6) Repeat the washing from step 5, two more times.
7) Blot the membrane dry and wrap in Saran Wrap and expose to X-ray.

Stripping the membrane-- for 1-2hrs do the following: large volume of 10mM Tris-Cl (pH7.4)/ 0.2% SDS preheated to 70-75°C. Alternatively, use 50% formamide with 0.1xSSC/0.1%SDS preheated to 68°C.
Extreme (if necessary): wash filter in boiling 0.1xSSC/0.1% SDS for 15min; repeat until membrane is clean.

Northern Hybridization using RNA-labeled probes

--from Krumlauf, 1996.

Materials

50x Denhardt's Solution

0.05% (w/v) BSA, 0.05% (w/v) polyvinyl pyrolidone, and 0.05% (w/v) Ficoll 400.

Prehybridization buffer

60% formamide, 5x SSC, 5x Denhardt's, 50mM NaP04 buffer (pH 6.8), 250μg/ml of sheared denatured salmon sperm DNA, 100ug/ml of yeast tRNA, 1%SDS. Make with DEPC-treated water.

Hybridization buffer

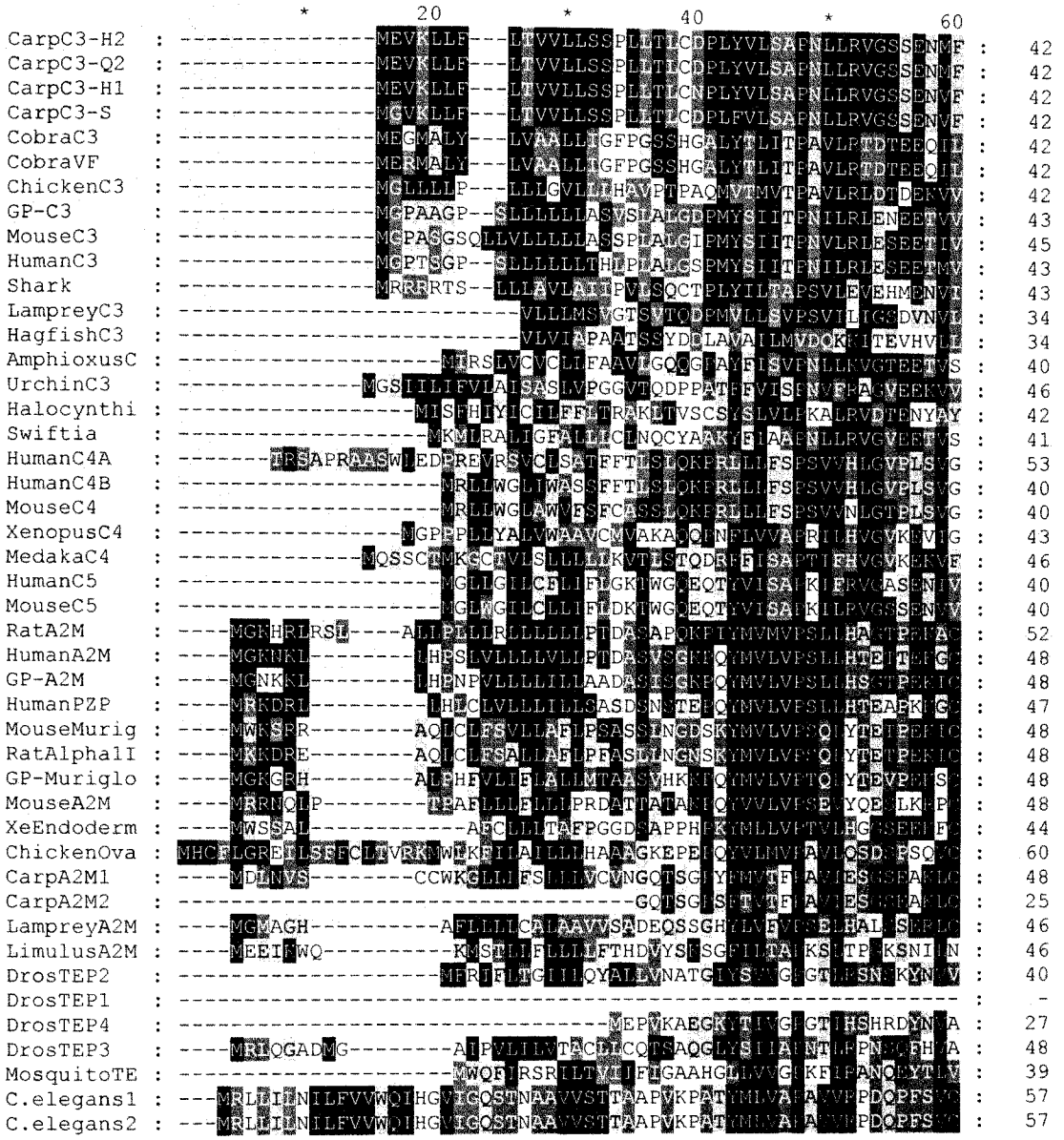
60% formamide, 5x SSC, 5x Denhardt's, 50mM NaP04 buffer (pH 6.8), 250μg/ml of sheared denatured salmon sperm DNA, 100μg/ml of yeast tRNA, 1%SDS (v/v), 10% dextran sulfate (w/v), and the appropriately labeled probe. Make with DEPC-treated water.

Methods

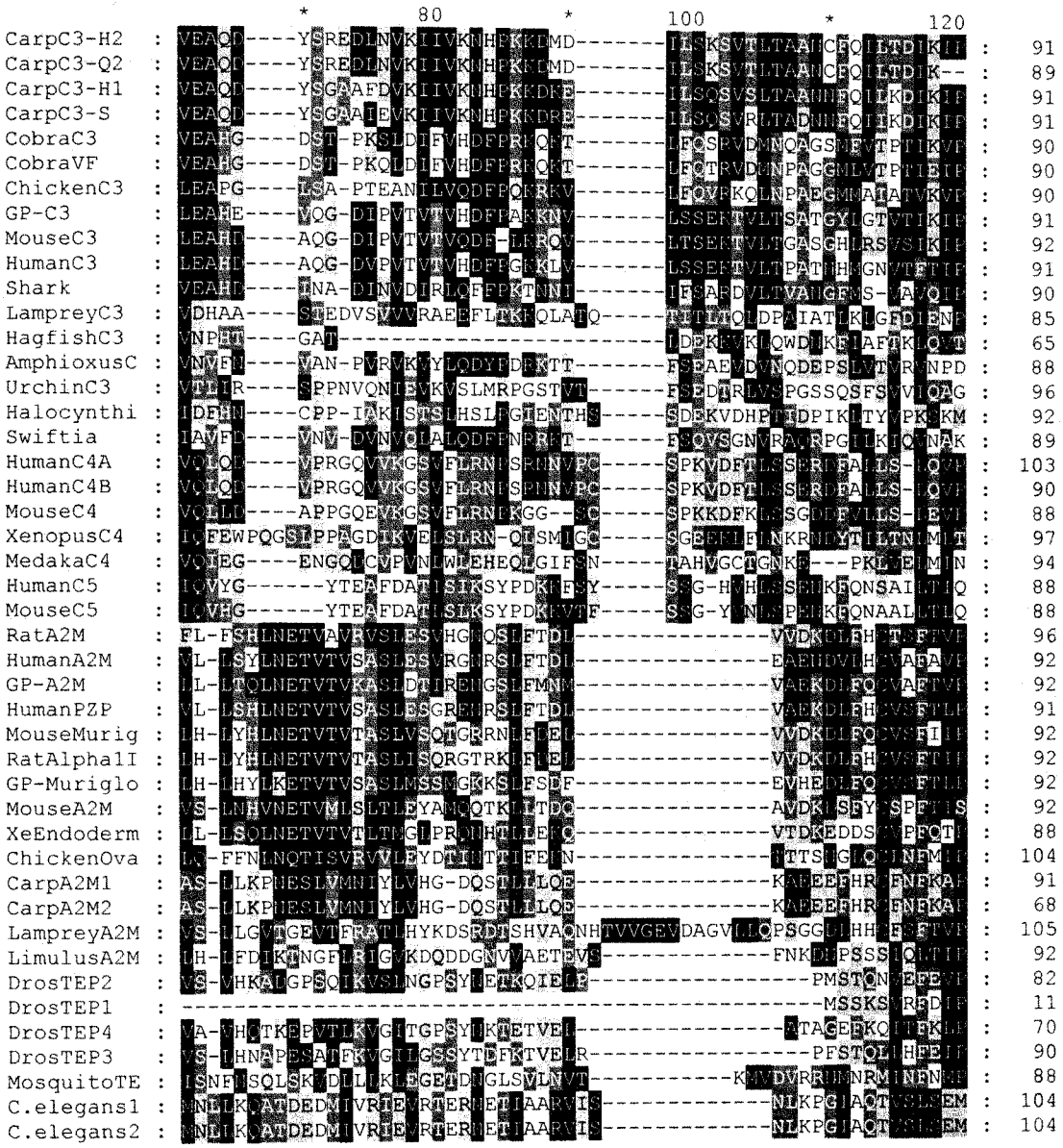
1) Place membrane in hybridization chamber and prehybridize for 2-4hrs at 60-65°C.
2) Place probe in 10-15ml of hybridization buffer and prewarm at hybridization temperature (65°C) for 20min.
3) Hybridize overnight (12-24hrs).
4) Rinse membrane in several hundred milliliters of 2xSSC at room temperature.
5) Wash filter in several hundred mls of 0.1x SSC/0.5% SDS at 75-80°C for 1hr.
6) Wash a second time with new buffer.
7) Check for background with hand-held monitor; a third wash may be necessary.
8) Remove filter from wash buffer, and wrap in Saran wrap and expose to film.
9) Membranes may be stripped by washing in 70% formamide at 90°C for 20min. Expose membrane to verify stripped. RNA probes are difficult to remove.

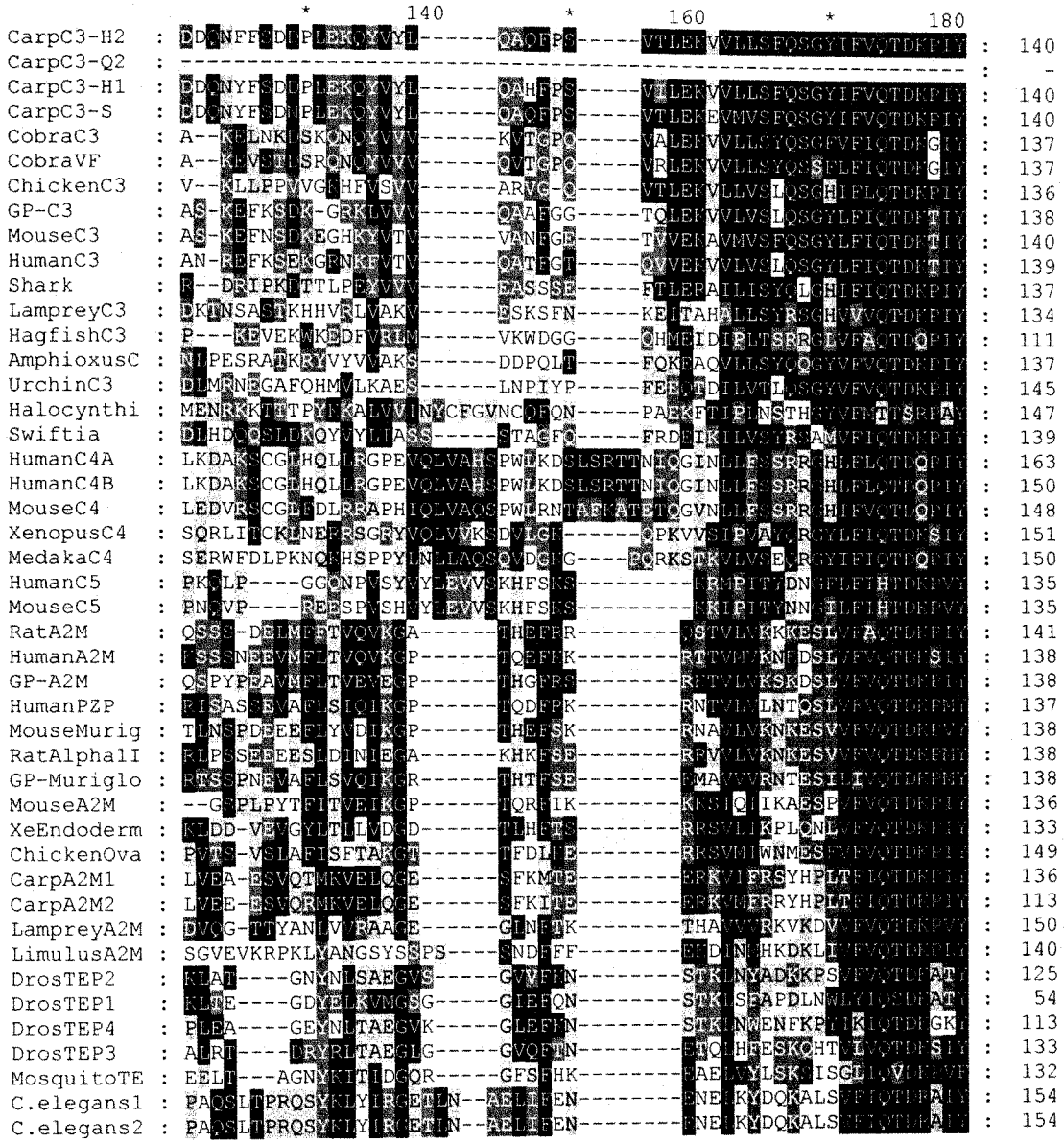
Appendix to Chapter 4

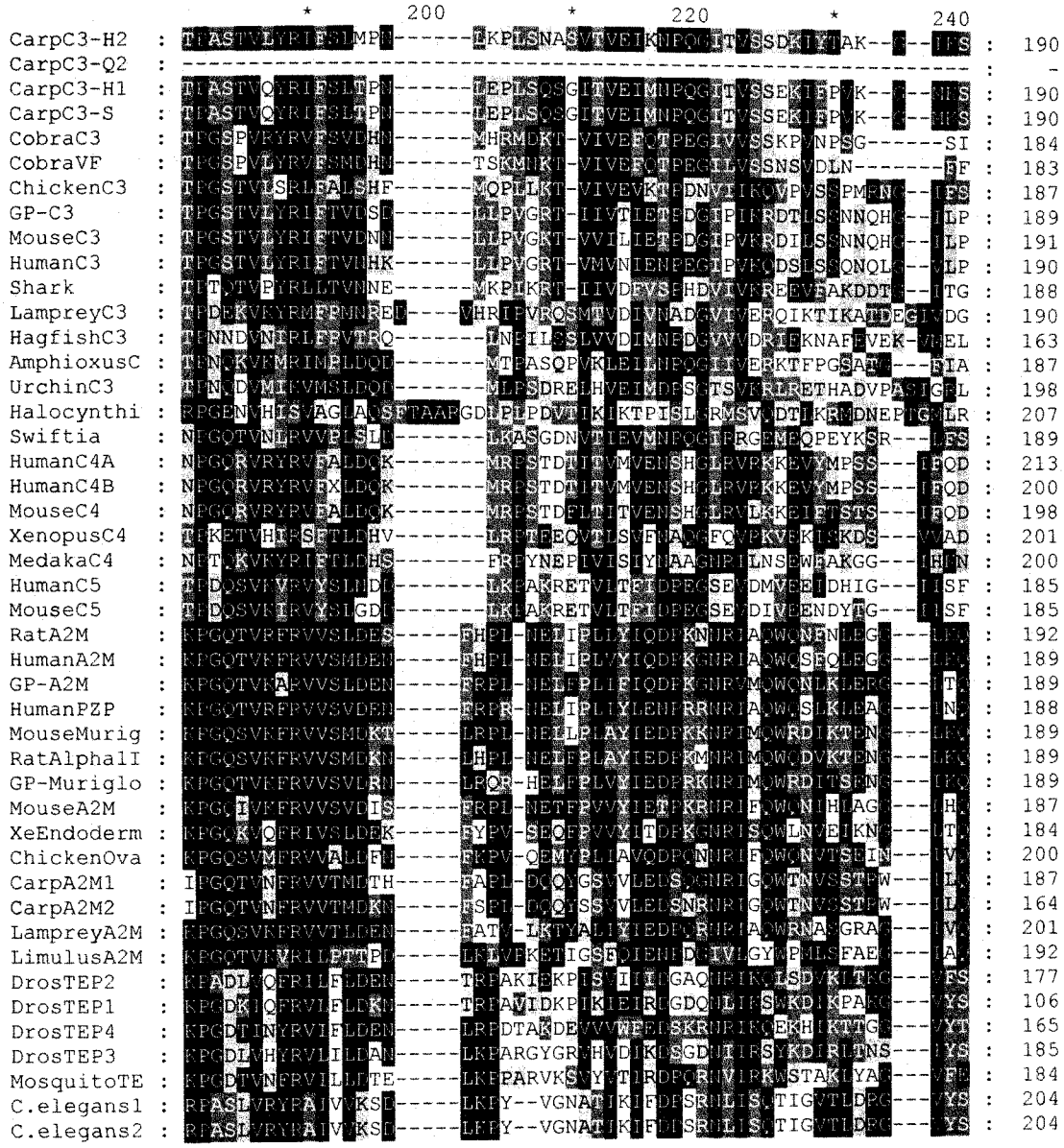
Full-length protein sequence alignment of multiple members of the TEP family, N=45.

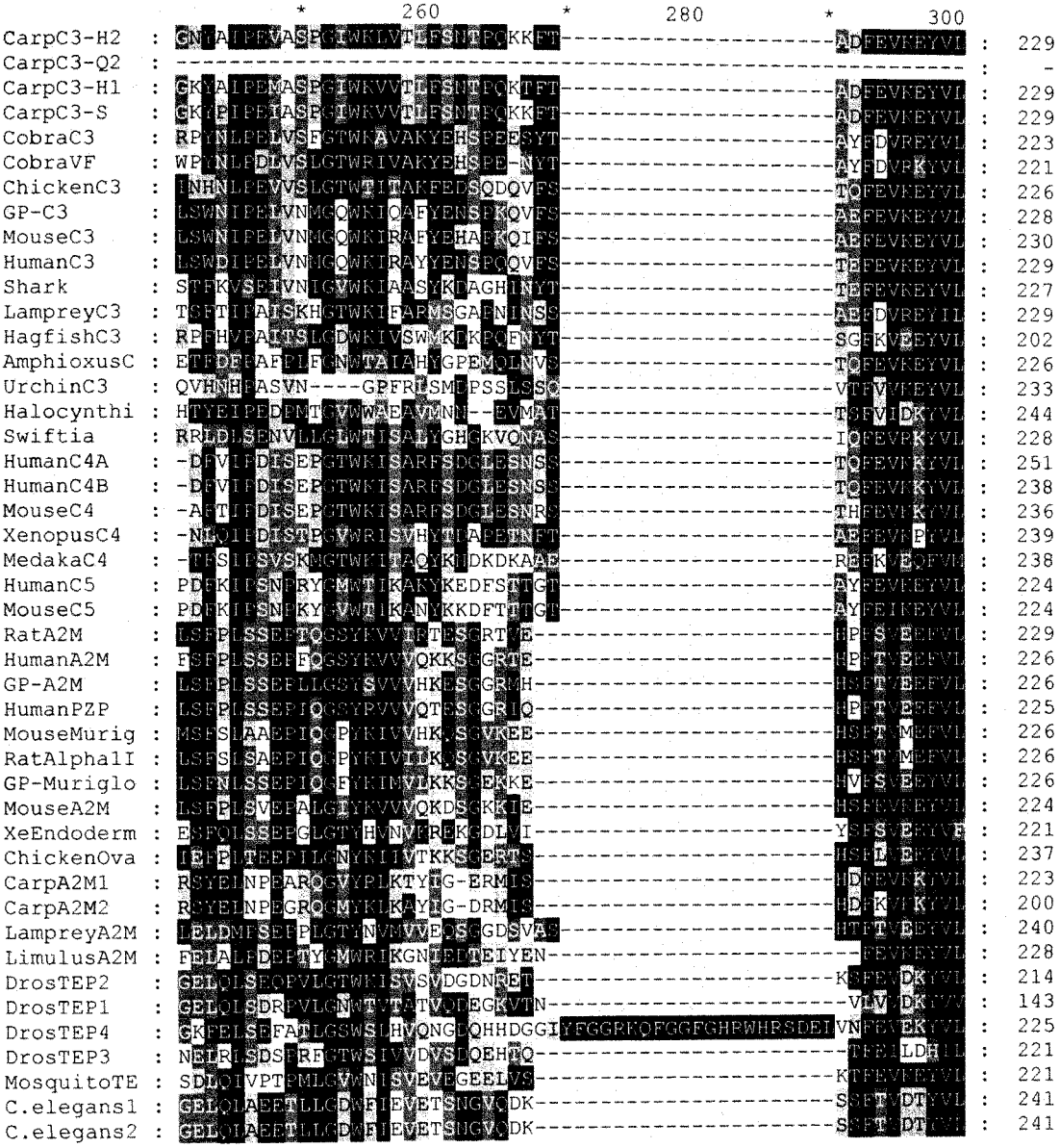


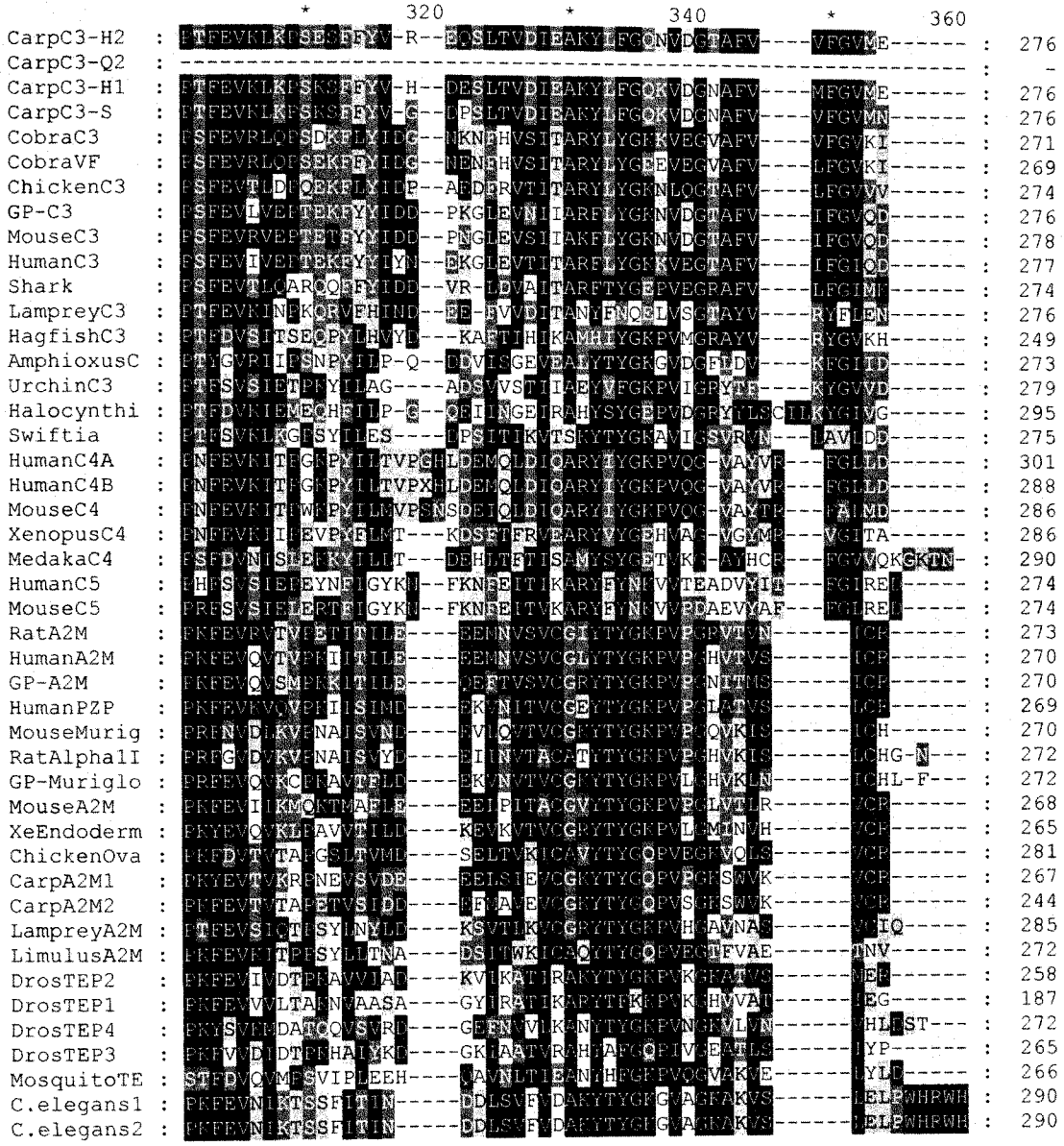
Appendix to Chapter 4. Full-length polypeptide sequence alignment of the TEP family, or the so-called alpha-2 macroglobulin family. Alignment continues for the next 36 pages. Sequences, from top to bottom, include C3, C4, C5 and then A2M-like protein sequences.

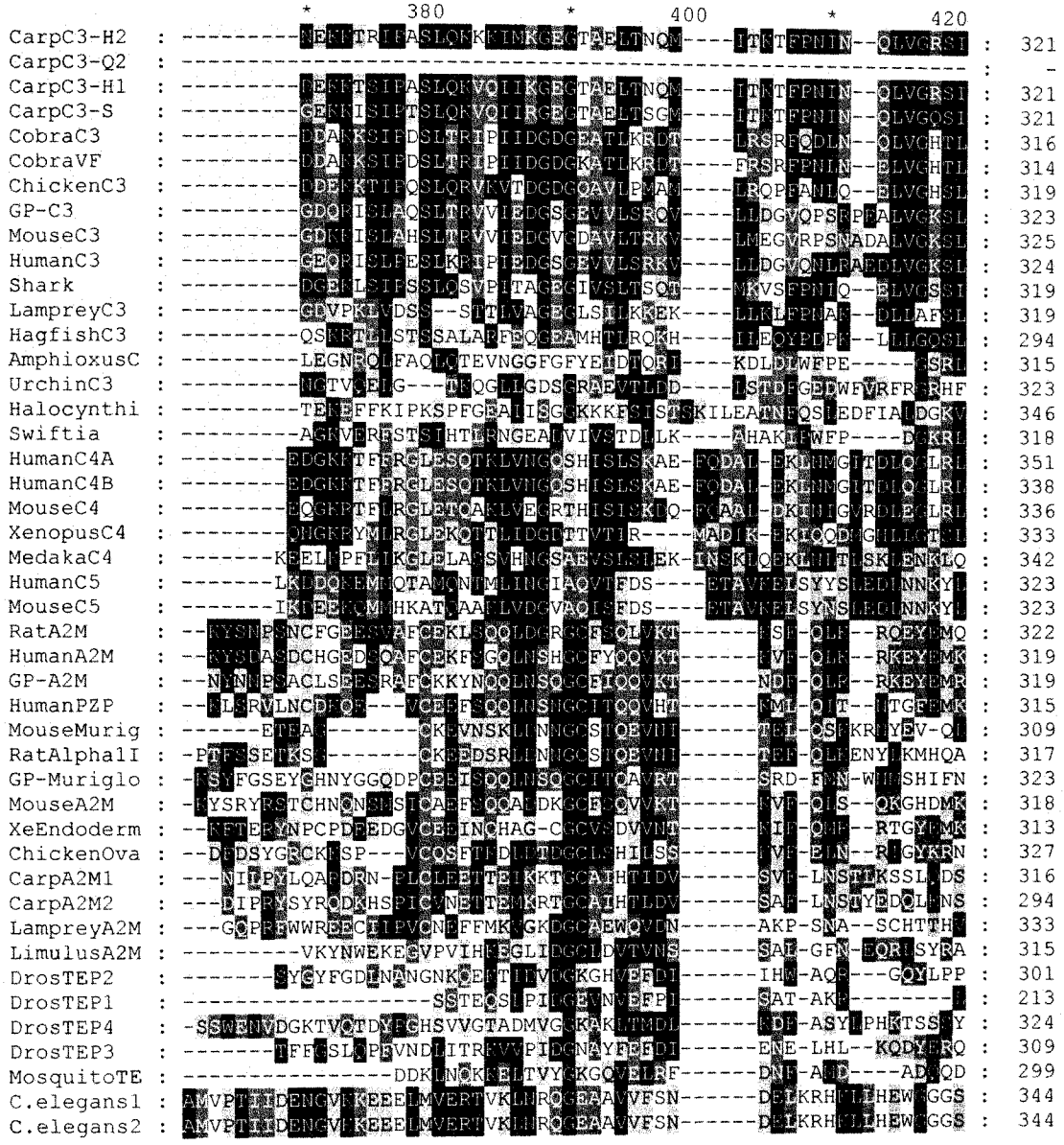


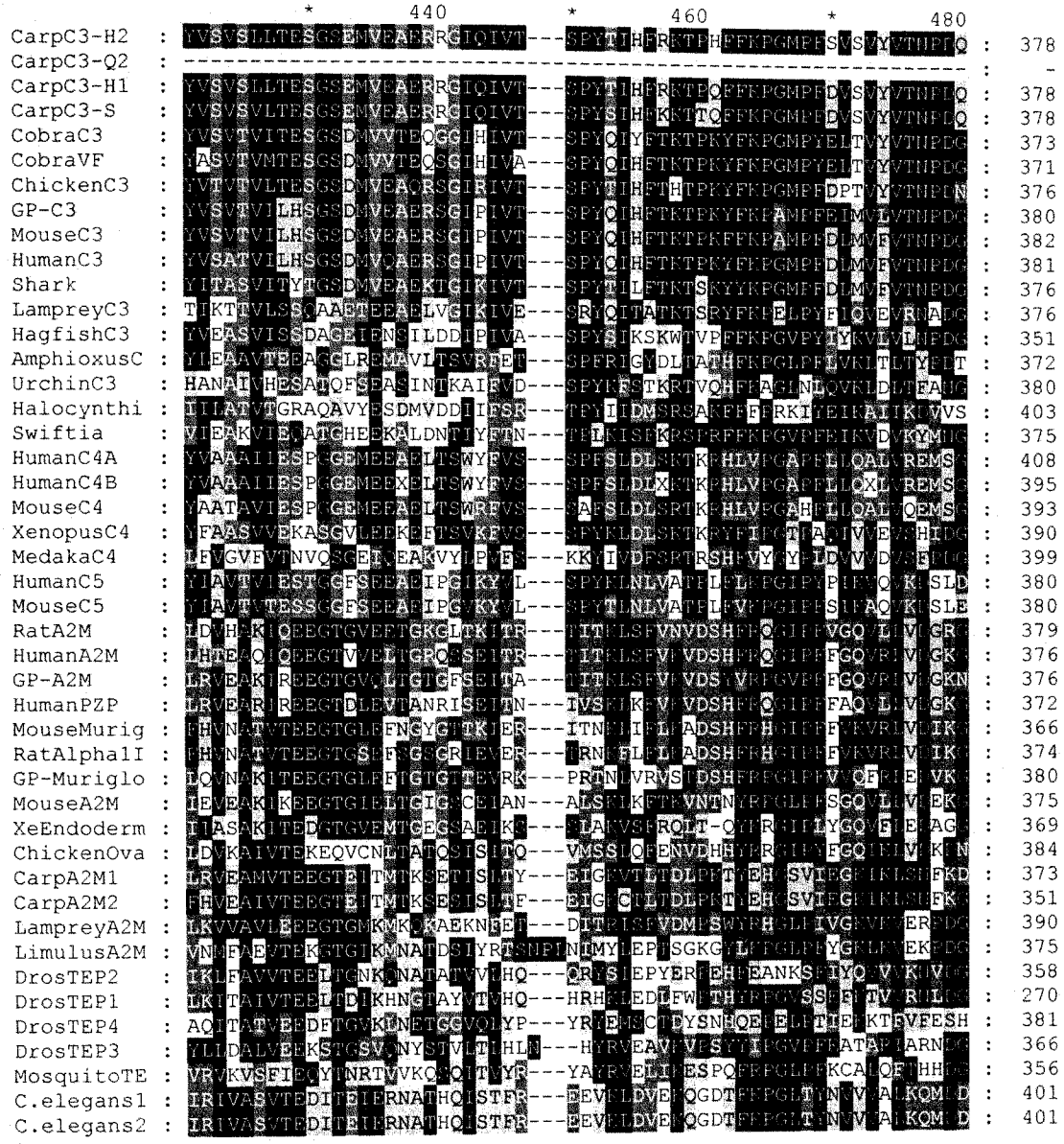


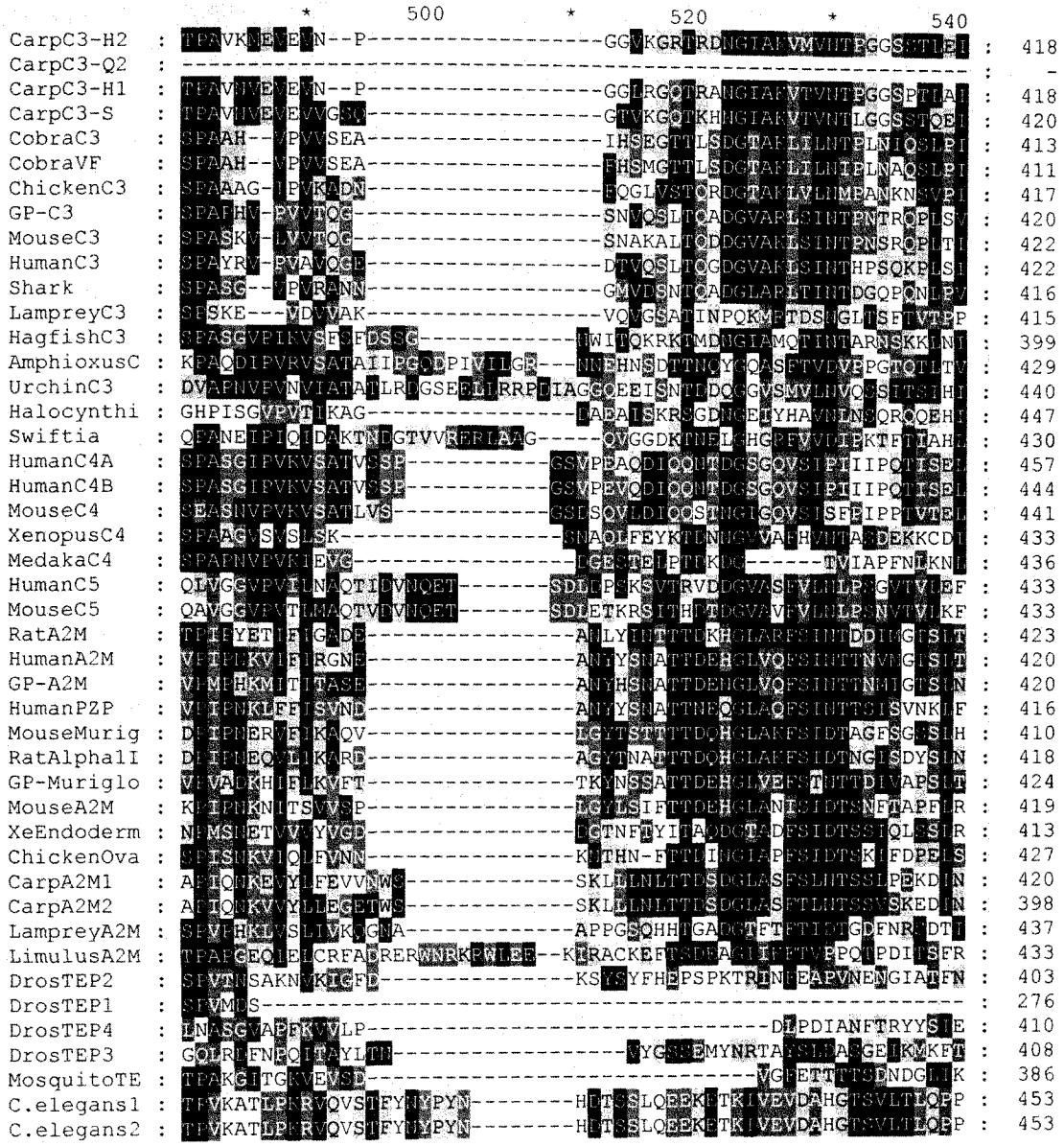

200
220
240
CarpC3-H2
CarpC3-Q2
CarpC3-H1
CarpC3-S
CobraC3
CobraVF
ChickenC3
GP-C3
MouseC3
HumanC3
Shark
LampreyC3
HagfishC3
AmphioxusC
UrchinC3
Halocynthi
Swiftia
HumanC4A
HumanC4B
MouseC4
XenopusC4
MedakaC4
HumanC5
MouseC5
RatA2M
HumanA2M
GP-A2M
HumanPZP
MouseMurig
RatAlphaII
GP-Muriglo
MouseA2M
XeEndoderm
ChickenOva
CarpA2M1
CarpA2M2
LampreyA2M
LimulusA2M
DrosTEP2
DrosTEP1
DrosTEP4
DrosTEP3
MosquitoTE
C.elegans1
C.elegans2

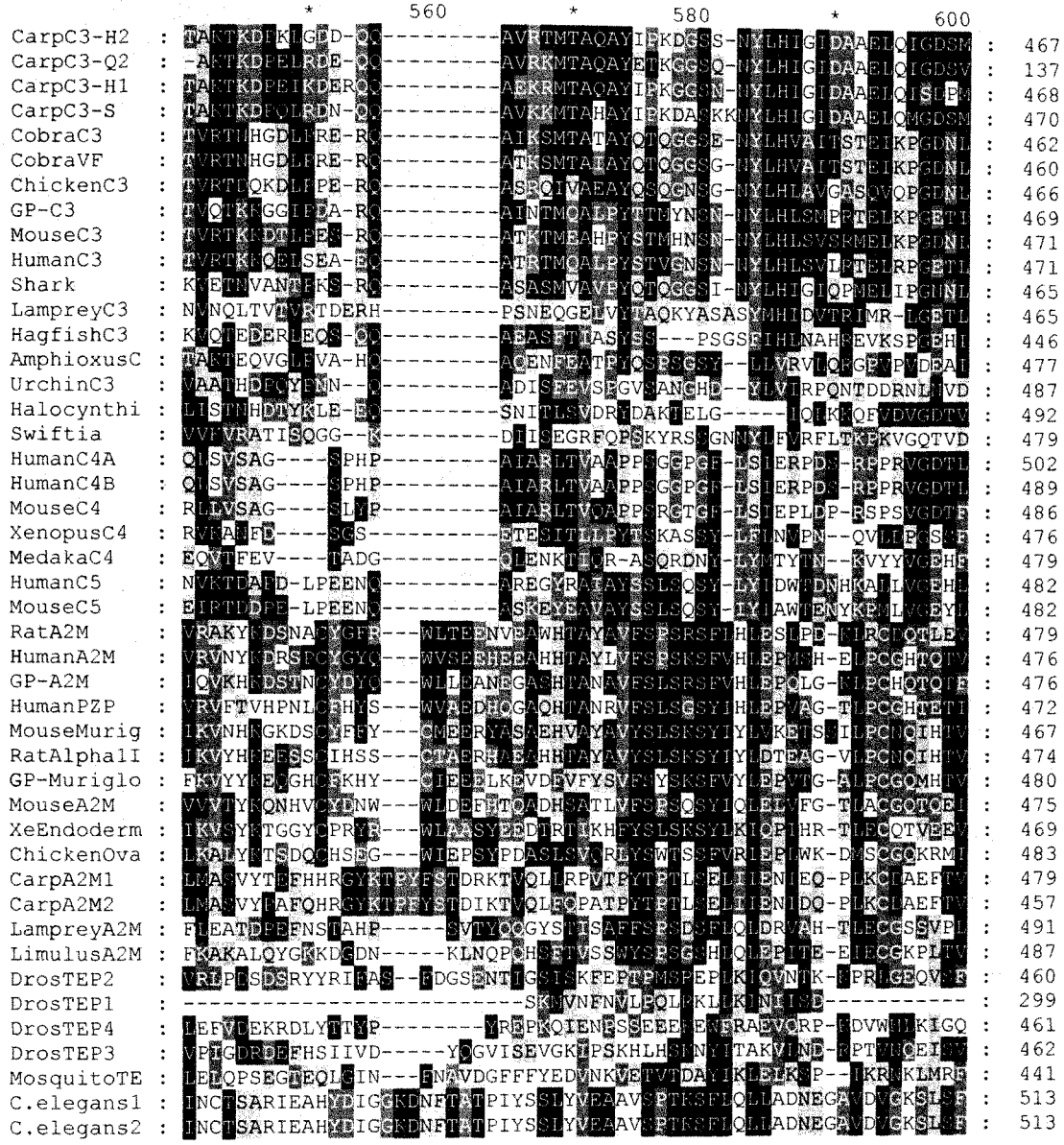

260
280
300
CarpC3-H2 : 229
CarpC3-Q2 : -
CarpC3-H1 : 229
CarpC3-S : 229
CobraC3 : 223
CobraVF : 221
ChickenC3 : 226
GP-C3 : 228
MouseC3 : 230
HumanC3 : 229
Shark : 227
LampreyC3 : 229
HagfishC3 : 202
AmphioxusC : 226
UrchinC3 : 233
Halocynthi : 244
Swiftia : 228
HumanC4A : 251
HumanC4B : 238
MouseC4 : 236
XenopusC4 : 239
MedakaC4 : 238
HumanC5 : 224
MouseC5 : 224
RatA2M : 229
HumanA2M : 226
GP-A2M : 226
HumanPZP : 225
MouseMurig : 226
RatAlpha1I : 226
GP-Muriglo : 226
MouseA2M : 224
XeEndoderm : 221
ChickenOva : 237
CarpA2M1 : 223
CarpA2M2 : 200
LampreyA2M : 240
LimulusA2M : 228
DrosTEP2 : 214
DrosTEP1 : 143
DrosTEP4 : 225
DrosTEP3 : 221
MosquitoTE : 221
C.elegans1 : 241
C.elegans2 : 241

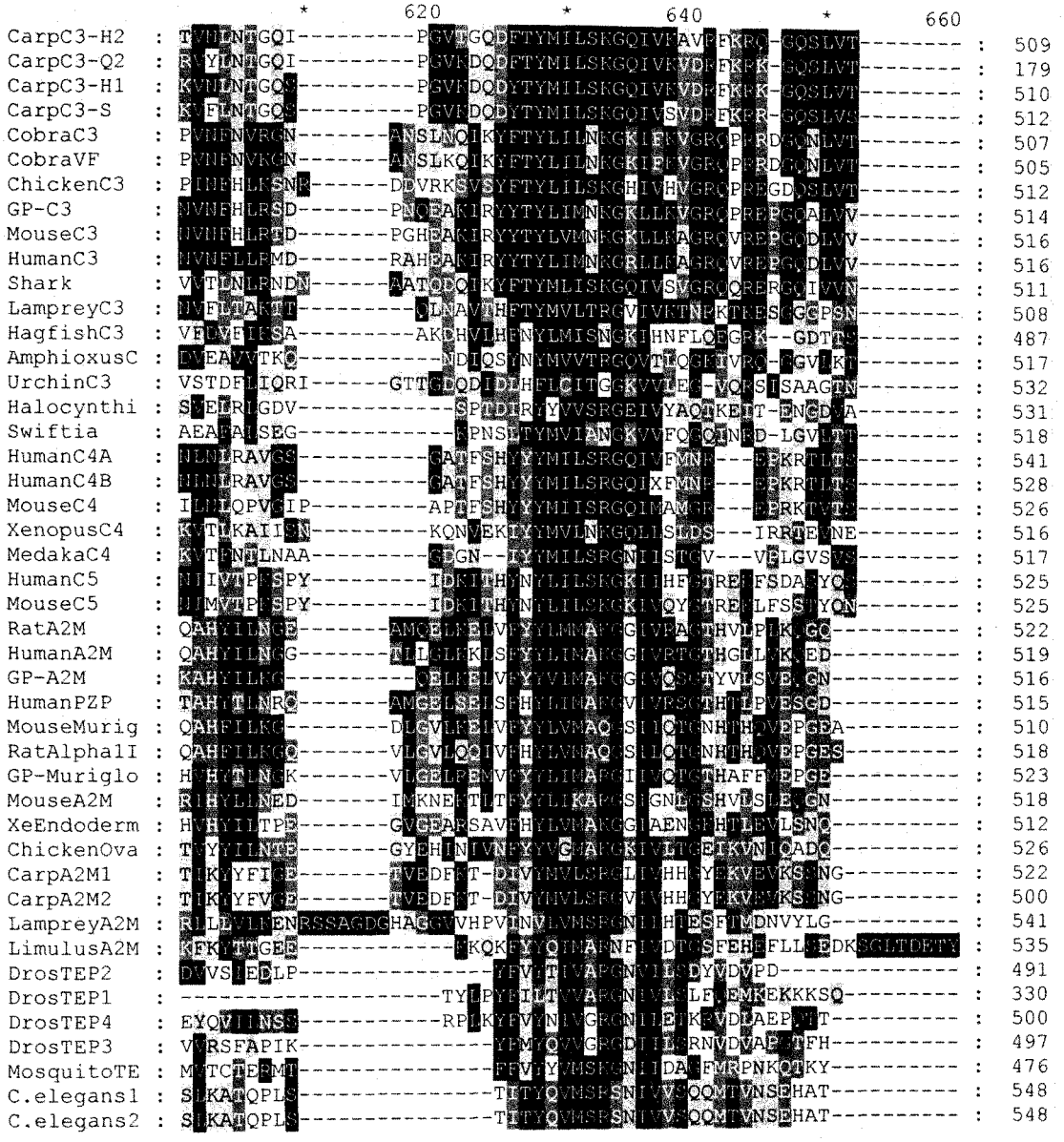

320
340
360
CarpC3-H2
CarpC3-Q2
CarpC3-H1
CarpC3-S
CobraC3
CobraVF
ChickenC3
GP-C3
MouseC3
HumanC3
Shark
LampreyC3
HagfishC3
AmphioxusC
UrchinC3
Halocynthi
Swiftia
HumanC4A
HumanC4B
MouseC4
XenopusC4
MedakaC4
HumanC5
MouseC5
RatA2M
HumanA2M
GP-A2M
HumanPZP
MouseMurig
RatAlpha1I
GP-Muriglo
MouseA2M
XeEndoderm
ChickenOva
CarpA2M1
CarpA2M2
LampreyA2M
LimulusA2M
DrosTEP2
DrosTEP1
DrosTEP4
DrosTEP3
MosquitoTE
C.elegans1
C.elegans2

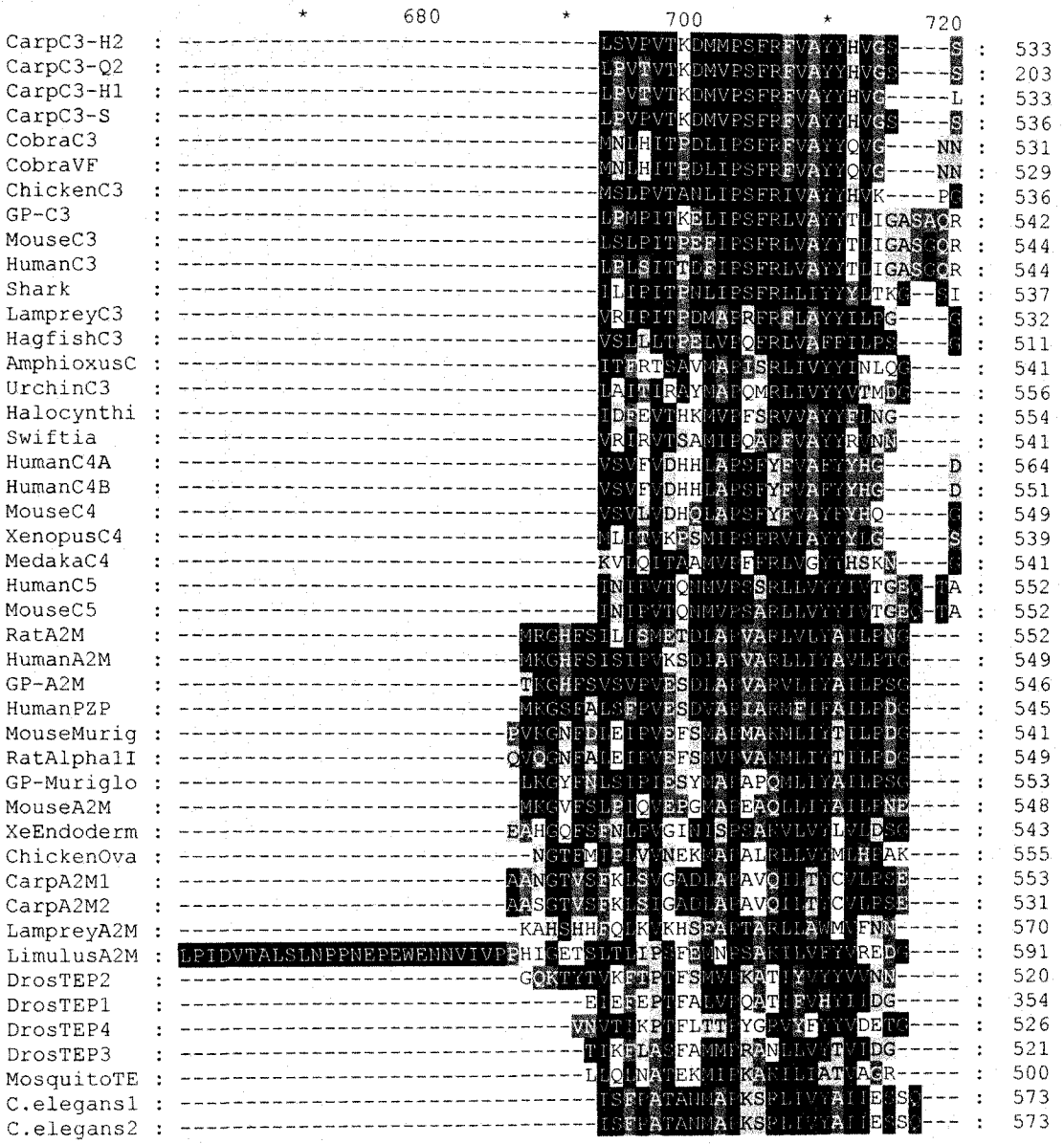

* 380 * 400 * 420
CarpC3-H2 : 321
CarpC3-Q2 : -
CarpC3-H1 : 321
CarpC3-S : 321
CobraC3 : 316
CobraVF : 314
ChickenC3 : 319
GP-C3 : 323
MouseC3 : 325
HumanC3 : 324
Shark : 319
LampreyC3 : 319
HagfishC3 : 294
AmphioxusC : 315
UrchinC3 : 323
Halocynthi : 346
Swiftia : 318
HumanC4A : 351
HumanC4B : 338
MouseC4 : 336
XenopusC4 : 333
MedakaC4 : 342
HumanC5 : 323
MouseC5 : 323
RatA2M : 322
HumanA2M : 319
GP-A2M : 319
HumanPZP : 315
MouseMurig : 309
RatAlpha1I : 317
GP-Muriglo : 323
MouseA2M : 318
XeEndoderm : 313
ChickenOva : 327
CarpA2M1 : 316
CarpA2M2 : 294
LampreyA2M : 333
LimulusA2M : 315
DrosTEP2 : 301
DrosTEP1 : 213
DrosTEP4 : 324
DrosTEP3 : 309
MosquitoTE : 299
C.elegans1 : 344
C.elegans2 : 344

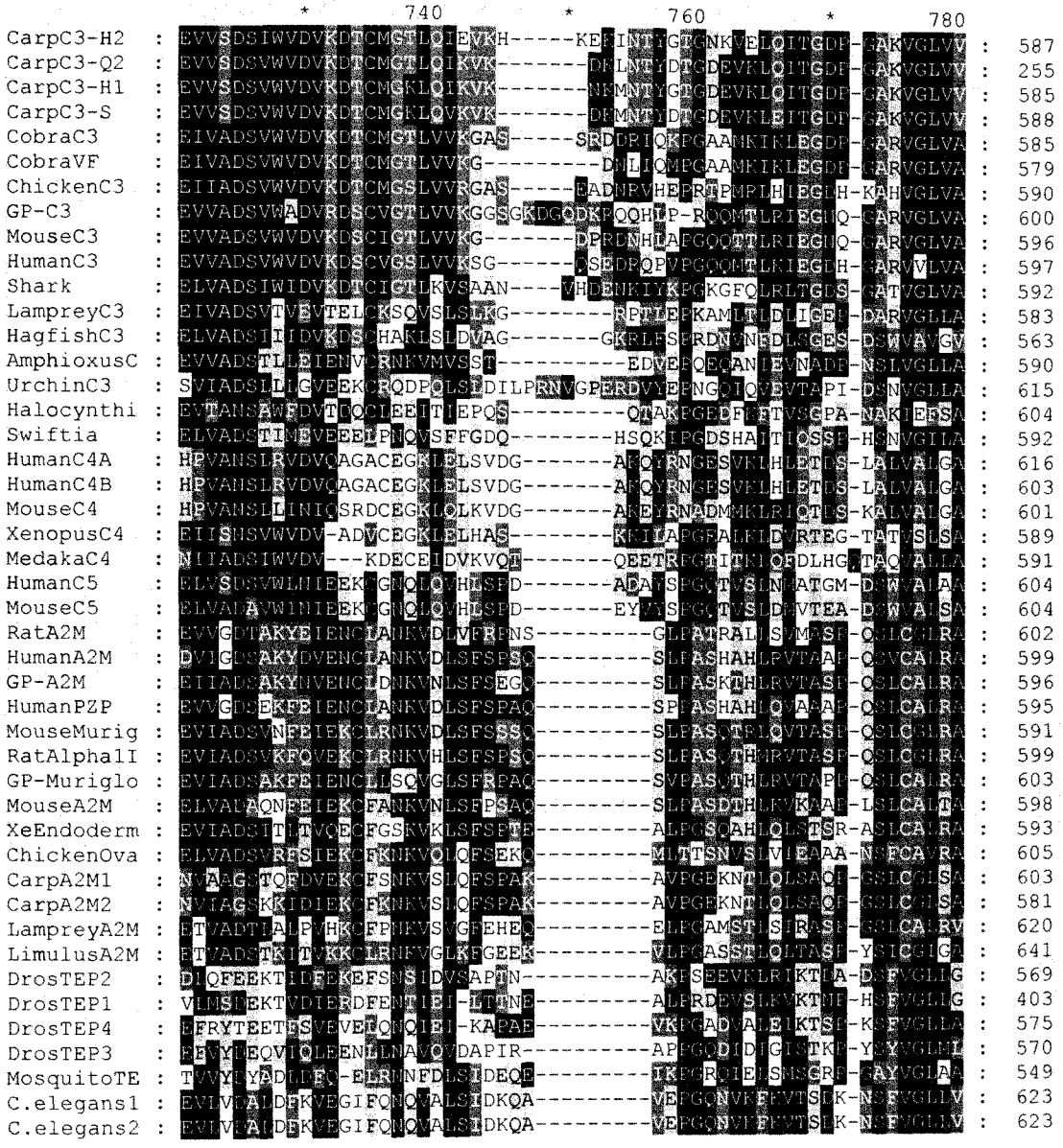

440
460
480
CarpC3-H2
CarpC3-Q2
CarpC3-H1
CarpC3-S
CobraC3
CobraVF
ChickenC3
GP-C3
MouseC3
HumanC3
Shark
LampreyC3
HagfishC3
AmphioxusC
UrchinC3
Halocynthi
Swiftia
HumanC4A
HumanC4B
MouseC4
XenopusC4
MedakaC4
HumanC5
MouseC5
RatA2M
HumanA2M
GP-A2M
HumanPZP
MouseMurig
RatAlpha1I
GP-Muriglo
MouseA2M
XeEndoderm
ChickenOva
CarpA2M1
CarpA2M2
LampreyA2M
LimulusA2M
DrosTEP2
DrosTEP1
DrosTEP4
DrosTEP3
MosquitoTE
C.elegans1
C.elegans2

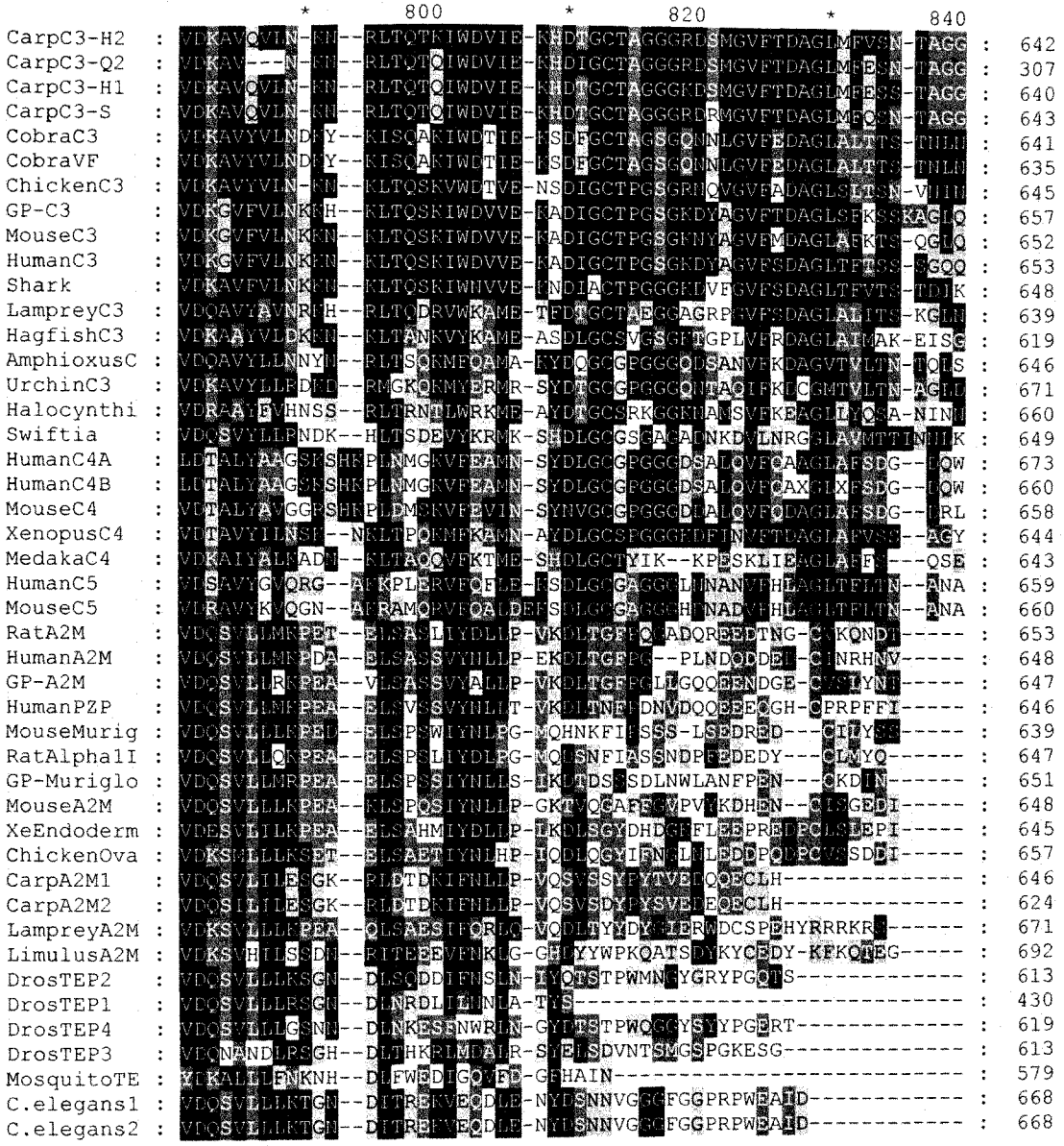

500
520
540
CarpC3-H2
CarpC3-Q2
CarpC3-H1
CarpC3-S
CobraC3
CobraVF
ChickenC3
GP-C3
MouseC3
HumanC3
Shark
LampreyC3
HagfishC3
AmphioxusC
UrchinC3
Halocynthi
Swiftia
HumanC4A
HumanC4B
MouseC4
XenopusC4
MedakaC4
HumanC5
MouseC5
RatA2M
HumanA2M
GP-A2M
HumanPZP
MouseMurig
RatAlpha1I
GP-Muriglo
MouseA2M
XeEndoderm
ChickenOva
CarpA2M1
CarpA2M2
LampreyA2M
LimulusA2M
DrosTEP2
DrosTEP1
DrosTEP4
DrosTEP3
MosquitoTE
C.elegans1
C.elegans2

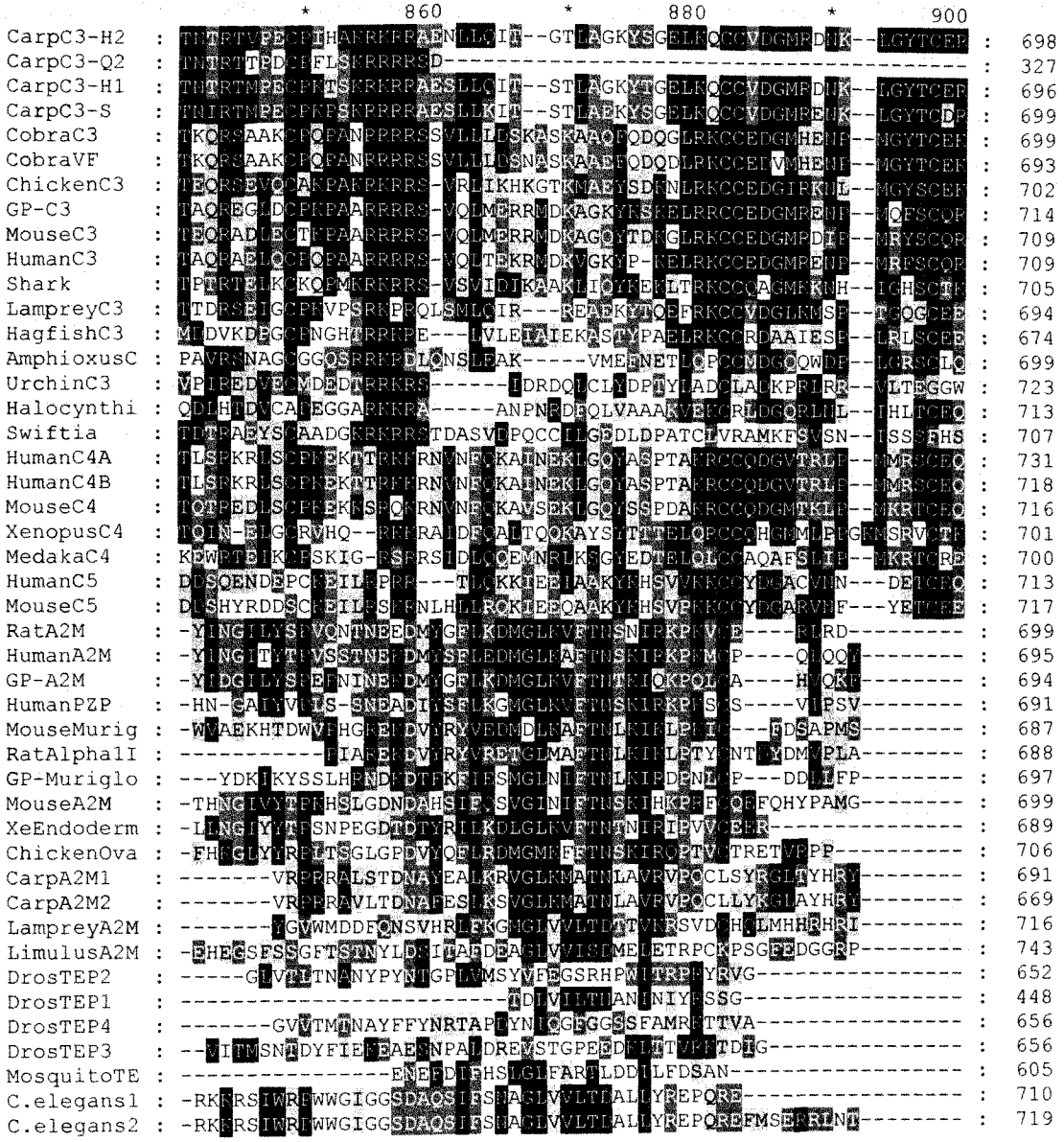

560
580
600
CarpC3-H2 : 467
CarpC3-Q2 : 137
CarpC3-H1 : 468
CarpC3-S : 470
CobraC3 : 462
CobraVF : 460
ChickenC3 : 466
GP-C3 : 469
MouseC3 : 471
HumanC3 : 471
Shark : 465
LampreyC3 : 465
HagfishC3 : 446
AmphioxusC : 477
UrchinC3 : 487
Halocynthi : 492
Swiftia : 479
HumanC4A : 502
HumanC4B : 489
MouseC4 : 486
XenopusC4 : 476
MedakaC4 : 479
HumanC5 : 482
MouseC5 : 482
RatA2M : 479
HumanA2M : 476
GP-A2M : 476
HumanPZP : 472
MouseMurig : 467
RatAlpha1I : 474
GP-Muriglo : 480
MouseA2M : 475
XeEndoderm : 469
ChickenOva : 483
CarpA2M1 : 479
CarpA2M2 : 457
LampreyA2M : 491
LimulusA2M : 487
DrosTEP2 : 460
DrosTEP1 : 299
DrosTEP4 : 461
DrosTEP3 : 462
MosquitoTE : 441
C.elegans1 : 513
C.elegans2 : 513

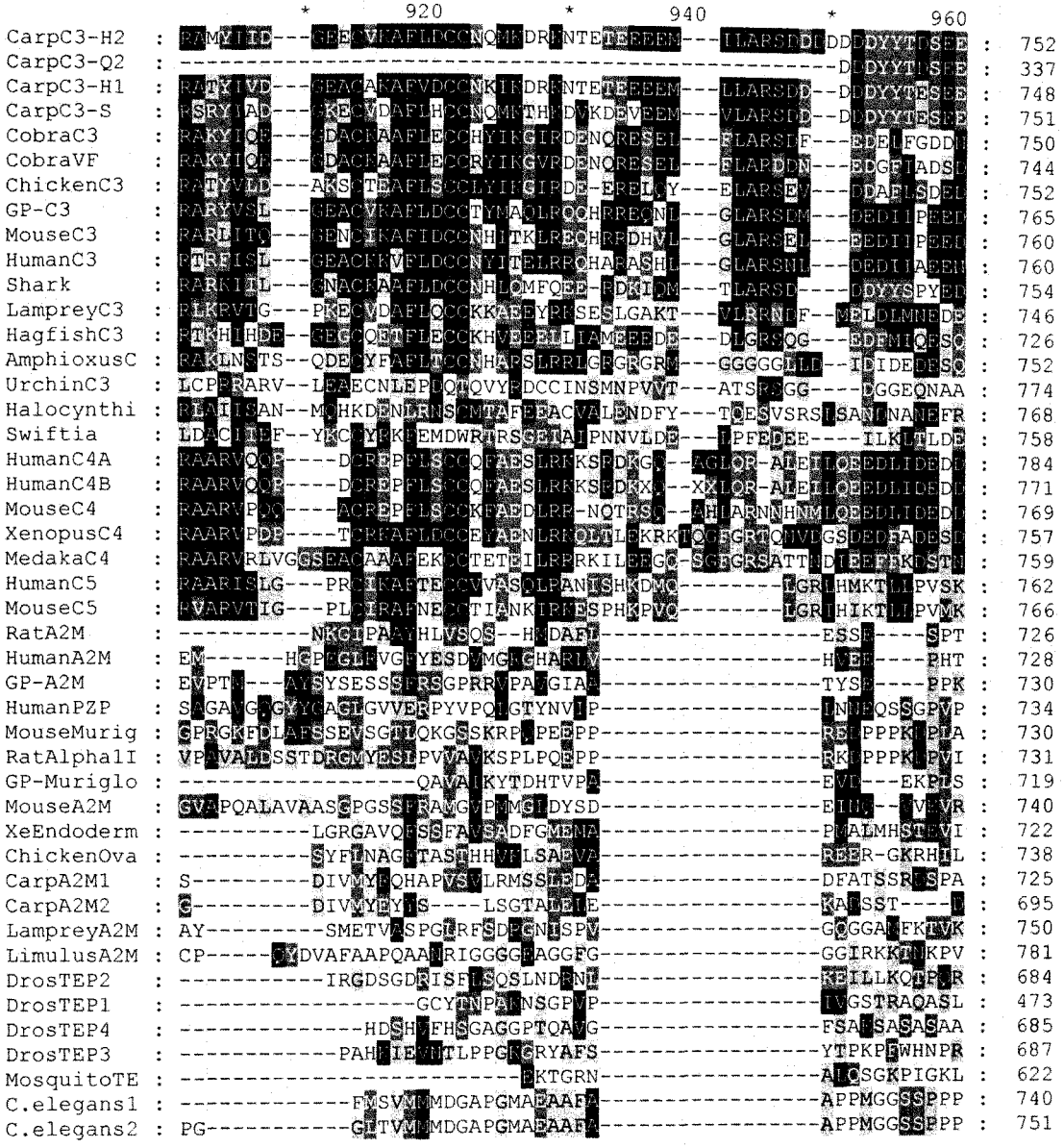

620
640
660
CarpC3-H2 : 509
CarpC3-Q2 : 179
CarpC3-H1 : 510
CarpC3-S : 512
CobraC3 : 507
CobraVF : 505
ChickenC3 : 512
GP-C3 : 514
MouseC3 : 516
HumanC3 : 516
Shark : 511
LampreyC3 : 508
HagfishC3 : 487
AmphioxusC : 517
UrchinC3 : 532
Halocynthi : 531
Swiftia : 518
HumanC4A : 541
HumanC4B : 528
MouseC4 : 526
XenopusC4 : 516
MedakaC4 : 517
HumanC5 : 525
MouseC5 : 525
RatA2M : 522
HumanA2M : 519
GP-A2M : 516
HumanPZP : 515
MouseMurig : 510
RatAlpha1I : 518
GP-Muriglo : 523
MouseA2M : 518
XeEndoderm : 512
ChickenOva : 526
CarpA2M1 : 522
CarpA2M2 : 500
LampreyA2M : 541
LimulusA2M : 535
DrosTEP2 : 491
DrosTEP1 : 330
DrosTEP4 : 500
DrosTEP3 : 497
MosquitoTE : 476
C.elegans1 : 548
C.elegans2 : 548

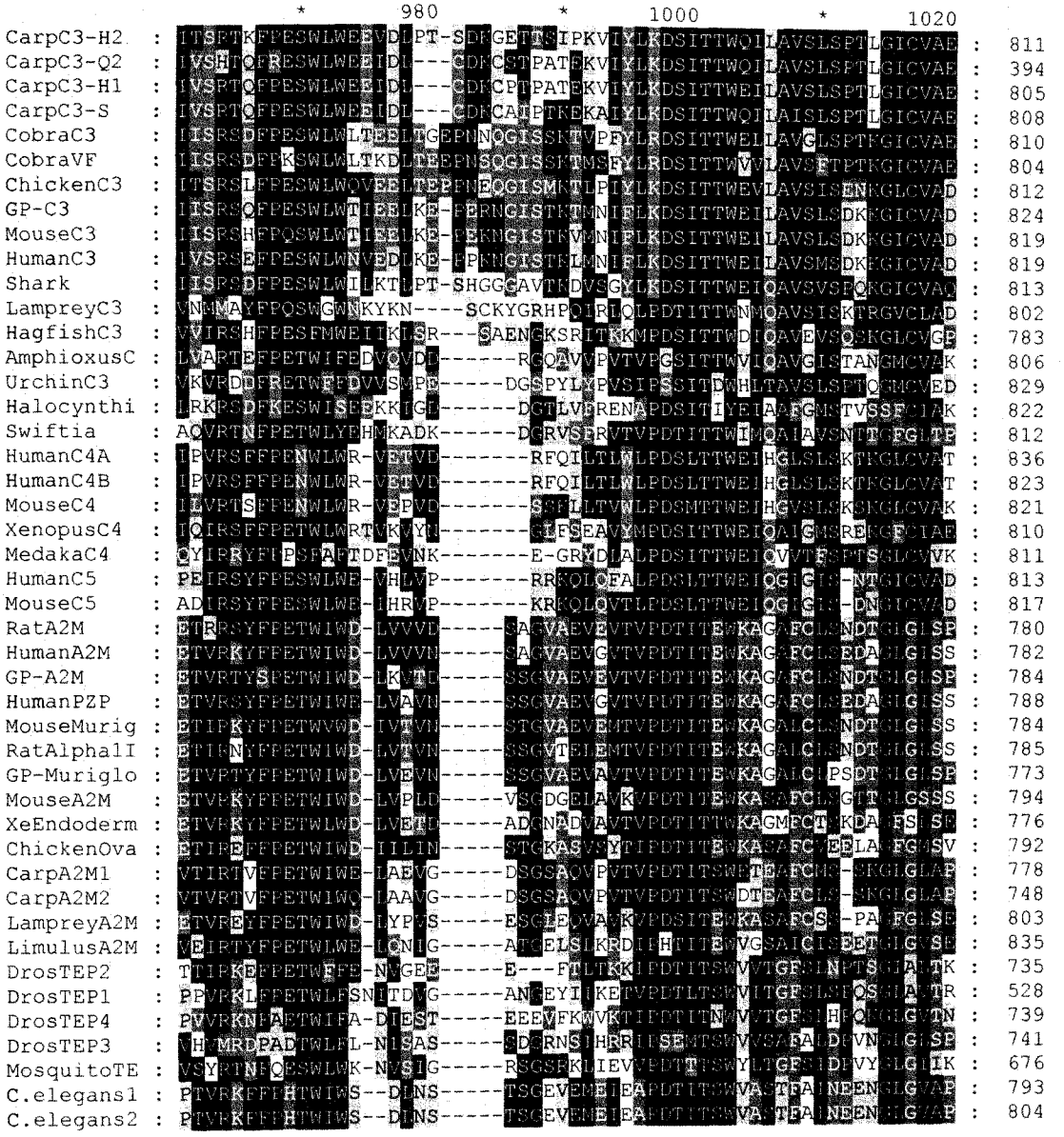

680
700
720
CarpC3-H2 : 533
CarpC3-Q2 : 203
CarpC3-H1 : 533
CarpC3-S : 536
CobraC3 : 531
CobraVF : 529
ChickenC3 : 536
GP-C3 : 542
MouseC3 : 544
HumanC3 : 544
Shark : 537
LampreyC3 : 532
HagfishC3 : 511
AmphioxusC : 541
UrchinC3 : 556
Halocynthi : 554
Swiftia : 541
HumanC4A : 564
HumanC4B : 551
MouseC4 : 549
XenopusC4 : 539
MedakaC4 : 541
HumanC5 : 552
MouseC5 : 552
RatA2M : 552
HumanA2M : 549
GP-A2M : 546
HumanPZP : 545
MouseMurig : 541
RatAlpha1I : 549
GP-Muriglo : 553
MouseA2M : 548
XeEndoderm : 543
ChickenOva : 555
CarpA2M1 : 553
CarpA2M2 : 531
LampreyA2M : 570
LimulusA2M : 591
DrosTEP2 : 520
DrosTEP1 : 354
DrosTEP4 : 526
DrosTEP3 : 521
MosquitoTE : 500
C.elegans1 : 573
C.elegans2 : 573

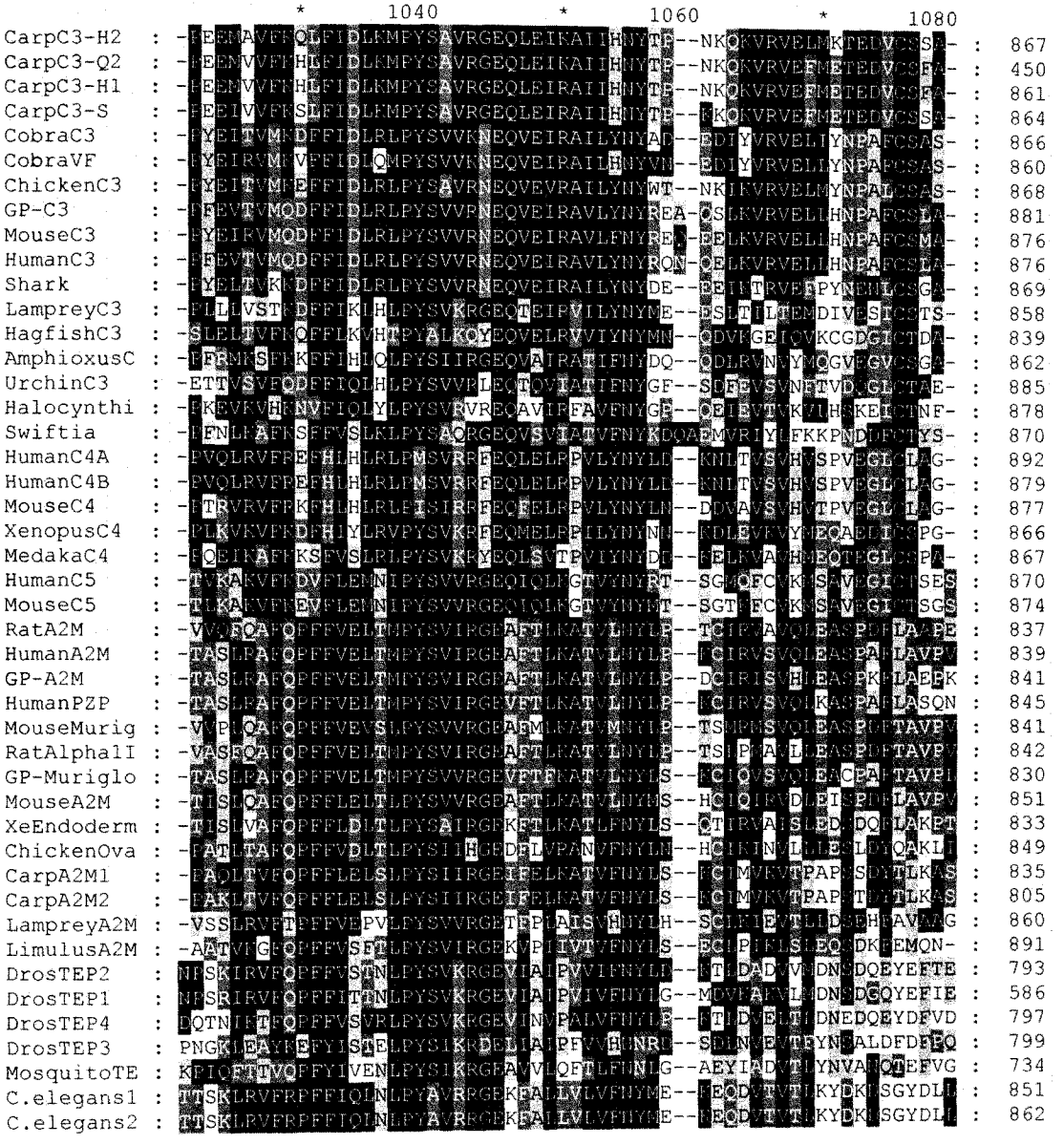

740
760
780
CarpC3-H2 : 587
CarpC3-Q2 : 255
CarpC3-H1 : 585
CarpC3-S : 588
CobraC3 : 585
CobraVF : 579
ChickenC3 : 590
GP-C3 : 600
MouseC3 : 596
HumanC3 : 597
Shark : 592
LampreyC3 : 583
HagfishC3 : 563
AmphioxusC : 590
UrchinC3 : 615
Halocynthi : 604
Swiftia : 592
HumanC4A : 616
HumanC4B : 603
MouseC4 : 601
XenopusC4 : 589
MedakaC4 : 591
HumanC5 : 604
MouseC5 : 604
RatA2M : 602
HumanA2M : 599
GP-A2M : 596
HumanPZP : 595
MouseMurig : 591
RatAlpha1I : 599
GP-Muriglo : 603
MouseA2M : 598
XeEndoderm : 593
ChickenOva : 605
CarpA2M1 : 603
CarpA2M2 : 581
LampreyA2M : 620
LimulusA2M : 641
DrosTEP2 : 569
DrosTEP1 : 403
DrosTEP4 : 575
DrosTEP3 : 570
MosquitoTE : 549
C.elegans1 : 623
C.elegans2 : 623

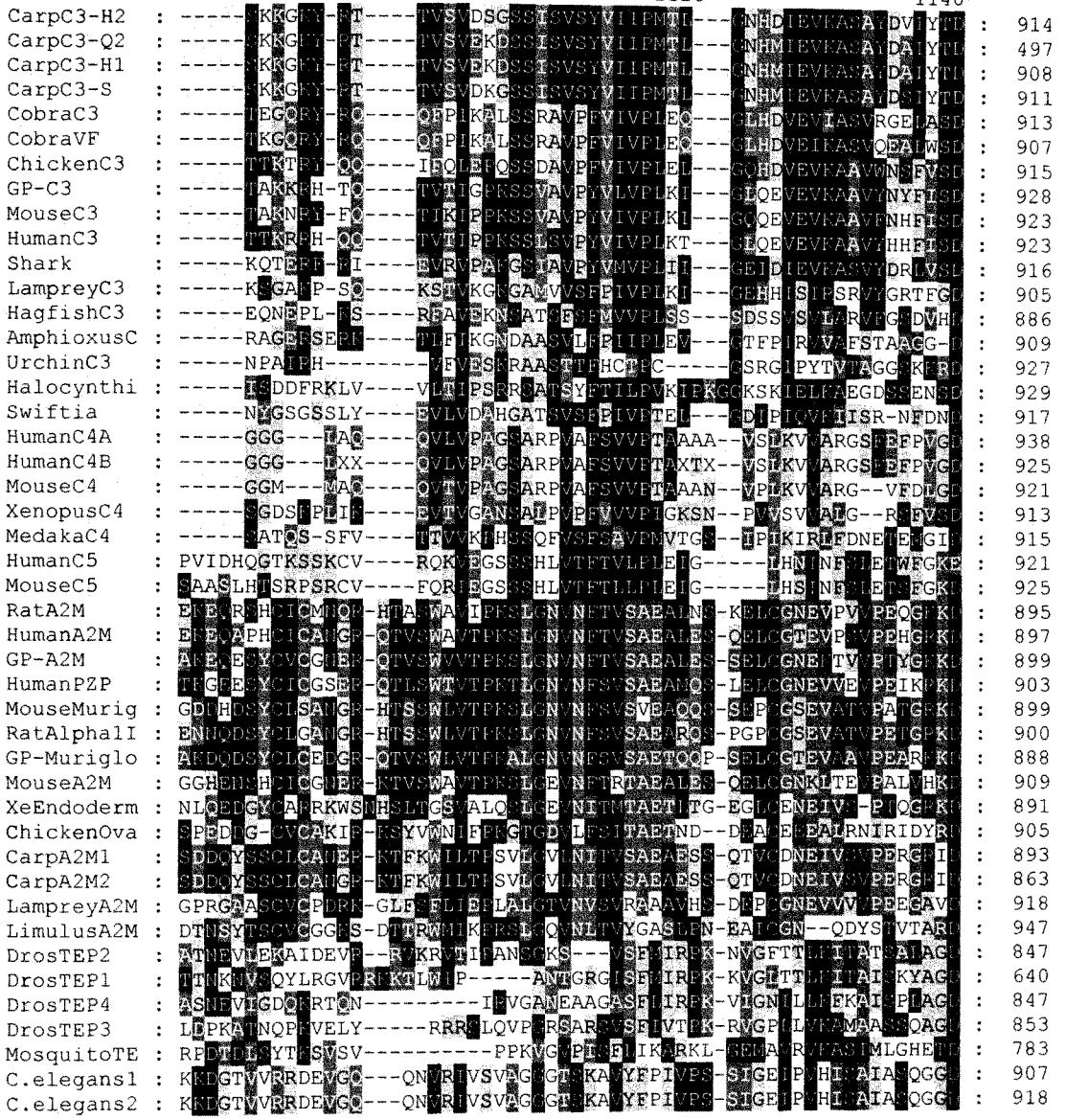

* 800 * 820 * 840
CarpC3-H2 : 642
CarpC3-Q2 : 307
CarpC3-H1 : 640
CarpC3-S : 643
CobraC3 : 641
CobraVF : 635
ChickenC3 : 645
GP-C3 : 657
MouseC3 : 652
HumanC3 : 653
Shark : 648
LampreyC3 : 639
HagfishC3 : 619
AmphioxusC : 646
UrchinC3 : 671
Halocynthi : 660
Swiftia : 649
HumanC4A : 673
HumanC4B : 660
MouseC4 : 658
XenopusC4 : 644
MedakaC4 : 643
HumanC5 : 659
MouseC5 : 660
RatA2M : 653
HumanA2M : 648
GP-A2M : 647
HumanPZP : 646
MouseMurig : 639
RatAlpha1I : 647
GP-Muriglo : 651
MouseA2M : 648
XeEndoderm : 645
ChickenOva : 657
CarpA2M1 : 646
CarpA2M2 : 624
LampreyA2M : 671
LimulusA2M : 692
DrosTEP2 : 613
DrosTEP1 : 430
DrosTEP4 : 619
DrosTEP3 : 613
MosquitoTE : 579
C.elegans1 : 668
C.elegans2 : 668

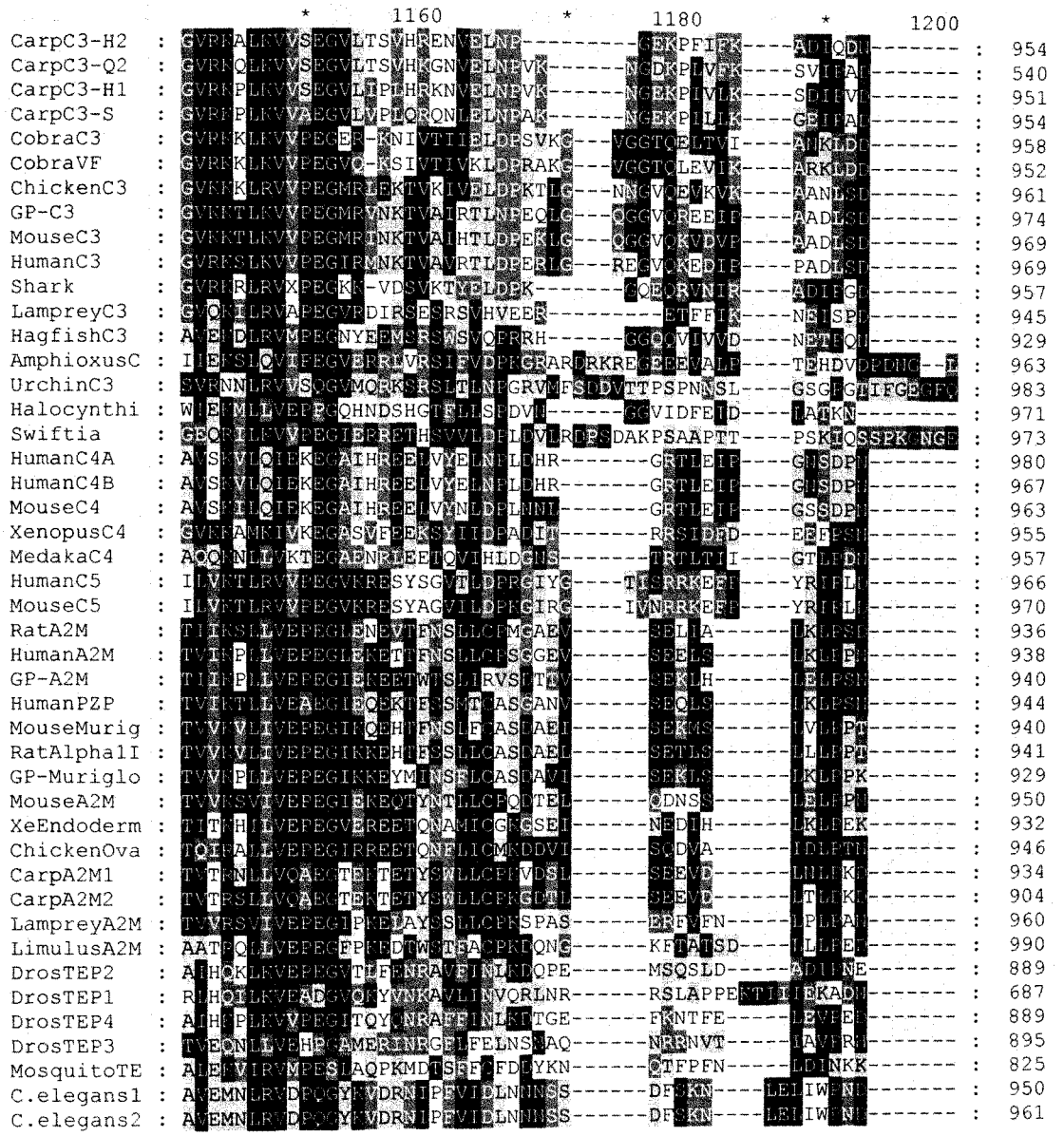

860
880
900
CarpC3-H2
CarpC3-Q2
CarpC3-H1
CarpC3-S
CobraC3
CobraVF
ChickenC3
GP-C3
MouseC3
HumanC3
Shark
LampreyC3
HagfishC3
AmphioxusC
UrchinC3
Halocynthi
Swiftia
HumanC4A
HumanC4B
MouseC4
XenopusC4
MedakaC4
HumanC5
MouseC5
RatA2M
HumanA2M
GP-A2M
HumanPZP
MouseMurig
RatAlpha1I
GP-Muriglo
MouseA2M
XeEndoderm
ChickenOva
CarpA2M1
CarpA2M2
LampreyA2M
LimulusA2M
DrosTEP2
DrosTEP1
DrosTEP4
DrosTEP3
MosquitoTE
C.elegans1
C.elegans2

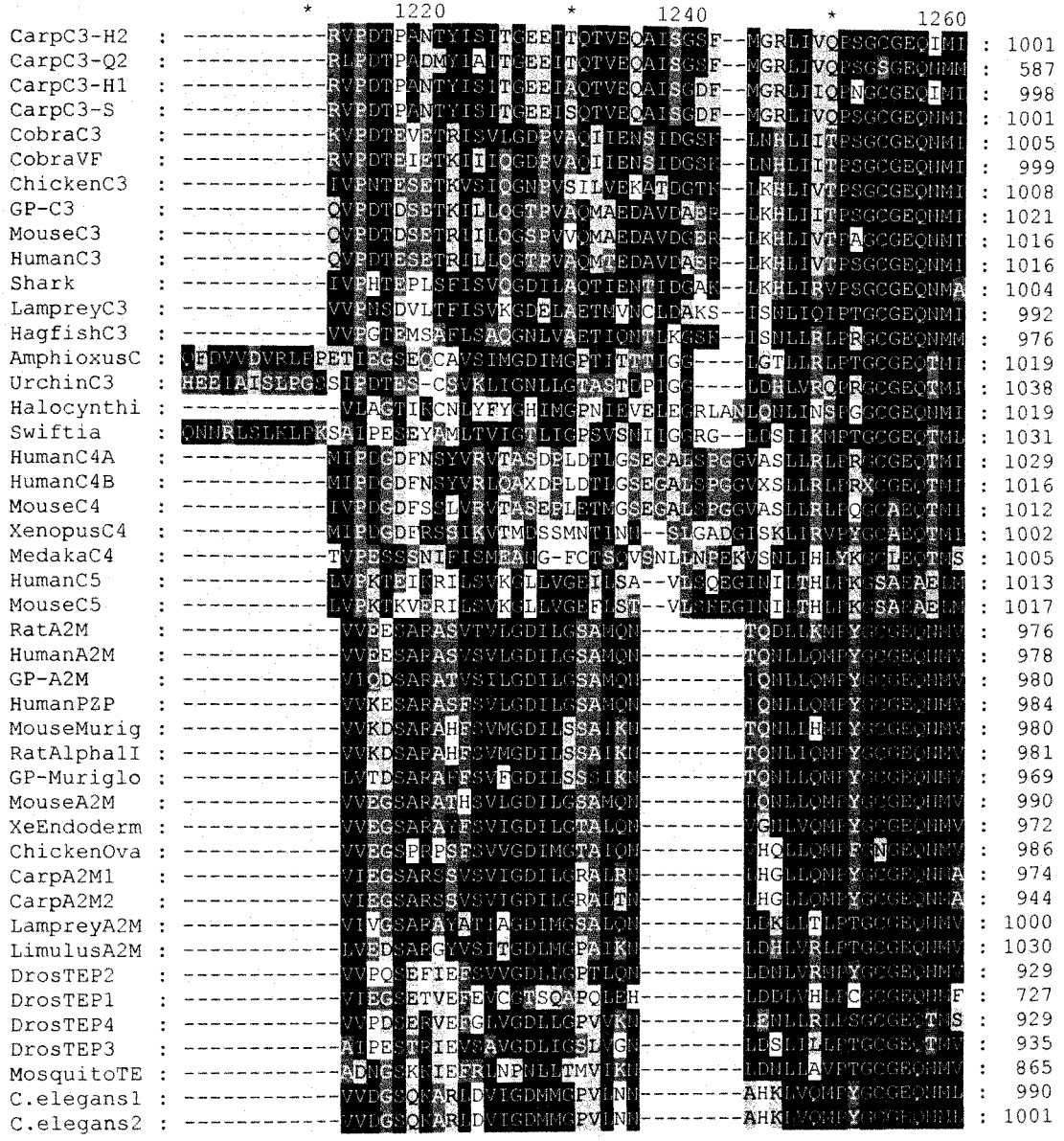


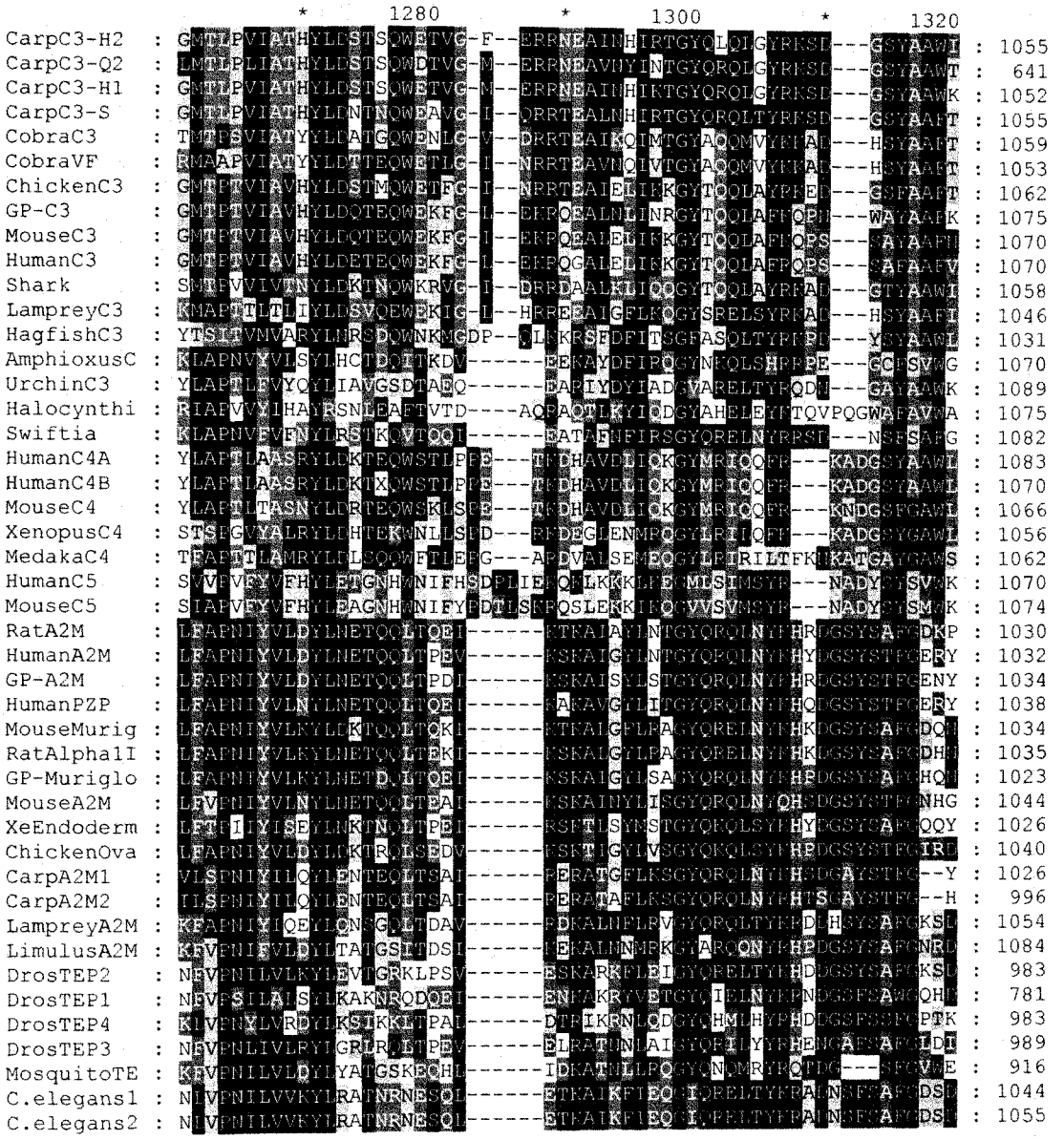

980
1000
1020
CarpC3-H2 : 811
CarpC3-Q2 : 394
CarpC3-H1 : 805
CarpC3-S : 808
CobraC3 : 810
CobraVF : 804
ChickenC3 : 812
GP-C3 : 824
MouseC3 : 819
HumanC3 : 819
Shark : 813
LampreyC3 : 802
HagfishC3 : 783
AmphioxusC : 806
UrchinC3 : 829
Halocynthi : 822
Swiftia : 812
HumanC4A : 836
HumanC4B : 823
MouseC4 : 821
XenopusC4 : 810
MedakaC4 : 811
HumanC5 : 813
MouseC5 : 817
RatA2M : 780
HumanA2M : 782
GP-A2M : 784
HumanPZP : 788
MouseMurig : 784
RatAlpha1I : 785
GP-Muriglo : 773
MouseA2M : 794
XeEndoderm : 776
ChickenOva : 792
CarpA2M1 : 778
CarpA2M2 : 748
LampreyA2M : 803
LimulusA2M : 835
DrosTEP2 : 735
DrosTEP1 : 528
DrosTEP4 : 739
DrosTEP3 : 741
MosquitoTE : 676
C.elegans1 : 793
C.elegans2 : 804

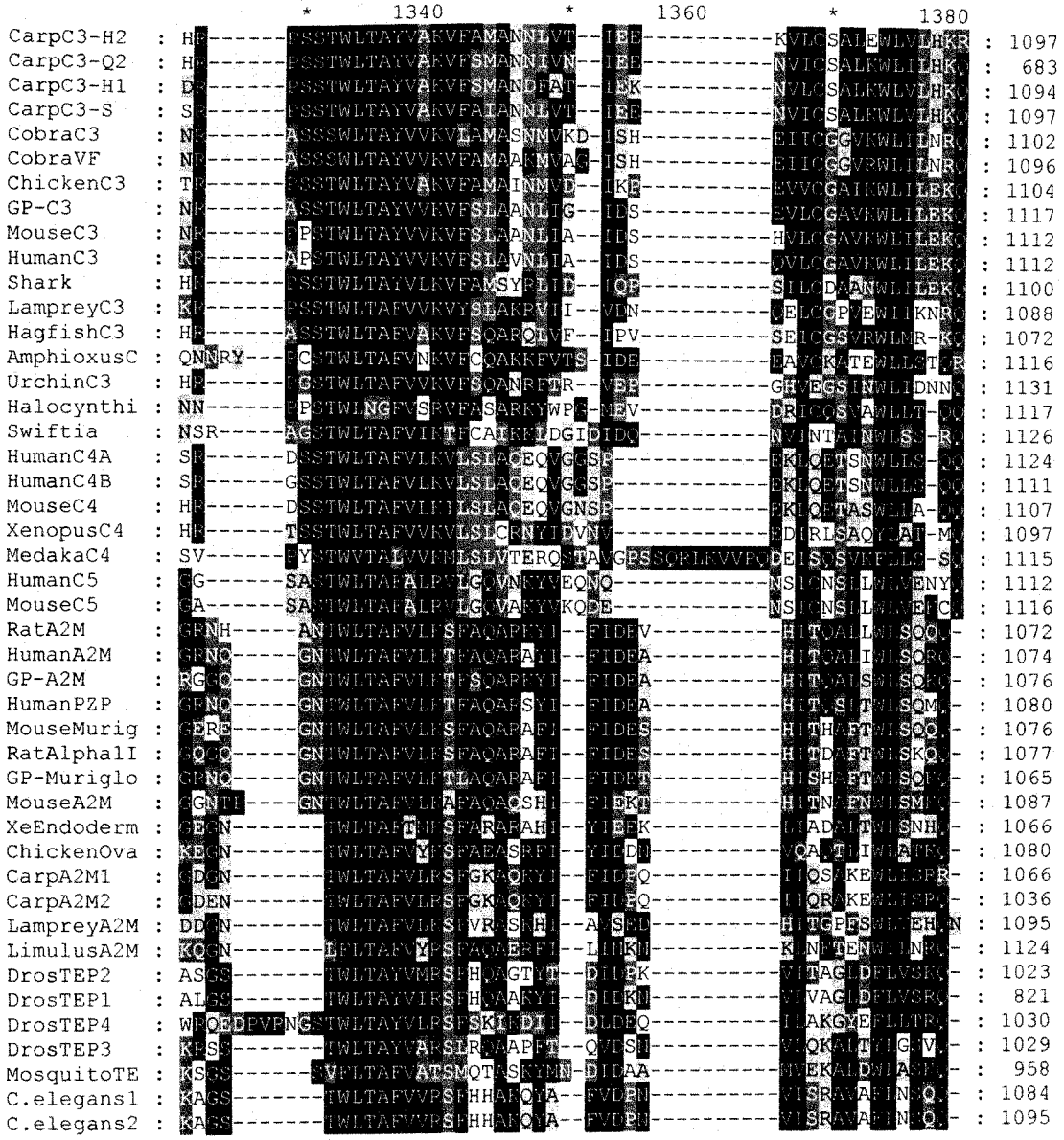

1040
1060
1080
CarpC3-H2 : 867
CarpC3-Q2 : 450
CarpC3-H1 : 861
CarpC3-S : 864
CobraC3 : 866
CobraVF : 860
ChickenC3 : 868
GP-C3 : 881
MouseC3 : 876
HumanC3 : 876
Shark : 869
LampreyC3 : 858
HagfishC3 : 839
AmphioxusC : 862
UrchinC3 : 885
Halocynthi : 878
Swiftia : 870
HumanC4A : 892
HumanC4B : 879
MouseC4 : 877
XenopusC4 : 866
MedakaC4 : 867
HumanC5 : 870
MouseC5 : 874
RatA2M : 837
HumanA2M : 839
GP-A2M : 841
HumanPZP : 845
MouseMurig : 841
RatAlpha1I : 842
GP-Muriglo : 830
MouseA2M : 851
XeEndoderm : 833
ChickenOva : 849
CarpA2M1 : 835
CarpA2M2 : 805
LampreyA2M : 860
LimulusA2M : 891
DrosTEP2 : 793
DrosTEP1 : 586
DrosTEP4 : 797
DrosTEP3 : 799
MosquitoTE : 734
C.elegans1 : 851
C.elegans2 : 862

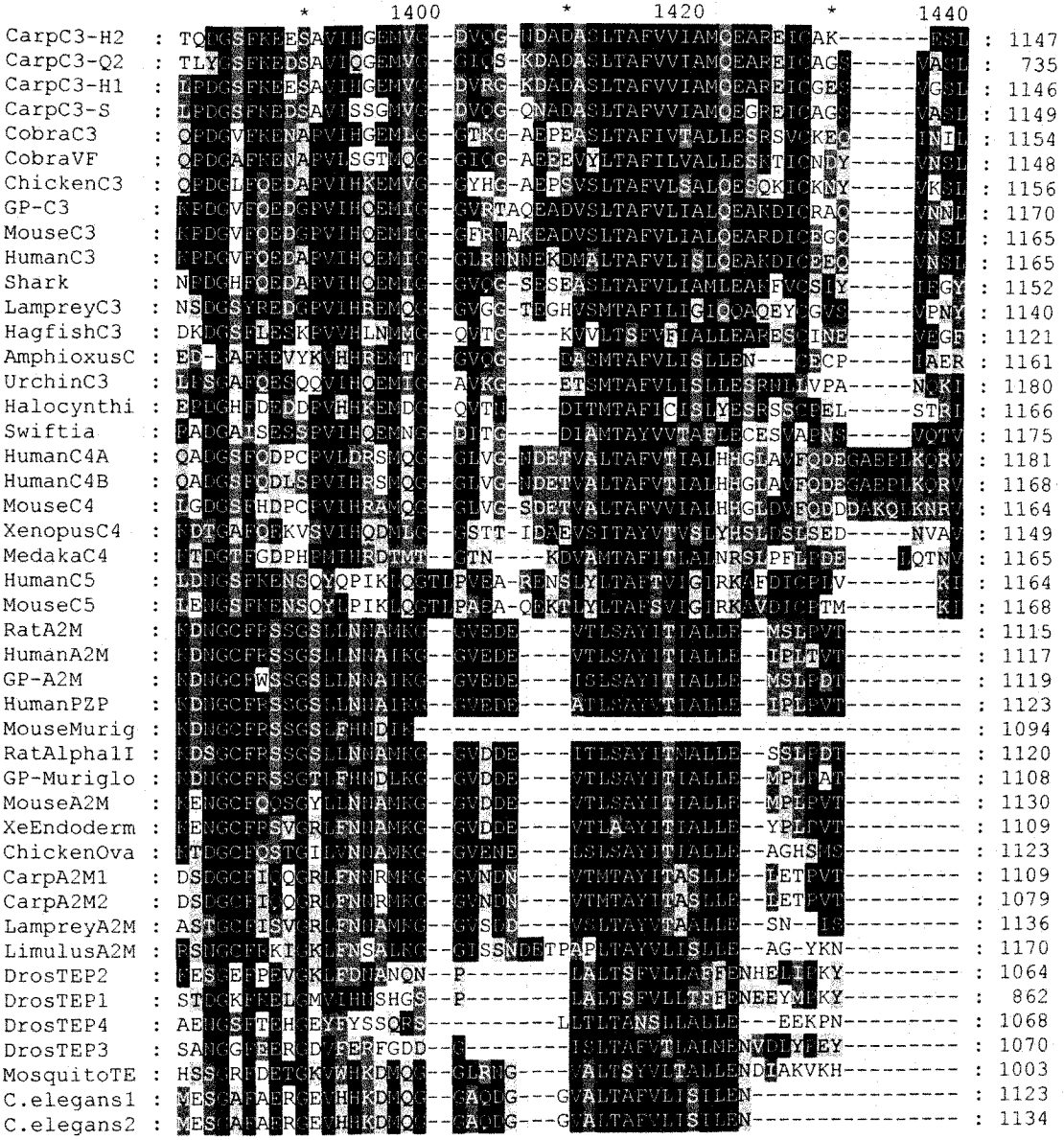

1100
1120
1140
CarpC3-H2
CarpC3-Q2
CarpC3-H1
CarpC3-S
CobraC3
CobraVF
ChickenC3
GP-C3
MouseC3
HumanC3
Shark
LampreyC3
HagfishC3
AmphioxusC
UrchinC3
Halocynthi
Swiftia
HumanC4A
HumanC4B
MouseC4
XenopusC4
MedakaC4
HumanC5
MouseC5
RatA2M
HumanA2M
GP-A2M
HumanPZP
MouseMurig
RatAlpha1I
GP-Muriglo
MouseA2M
XeEndoderm
ChickenOva
CarpA2M1
CarpA2M2
LampreyA2M
LimulusA2M
DrosTEP2
DrosTEP1
DrosTEP4
DrosTEP3
MosquitoTE
C.elegans1
C.elegans2

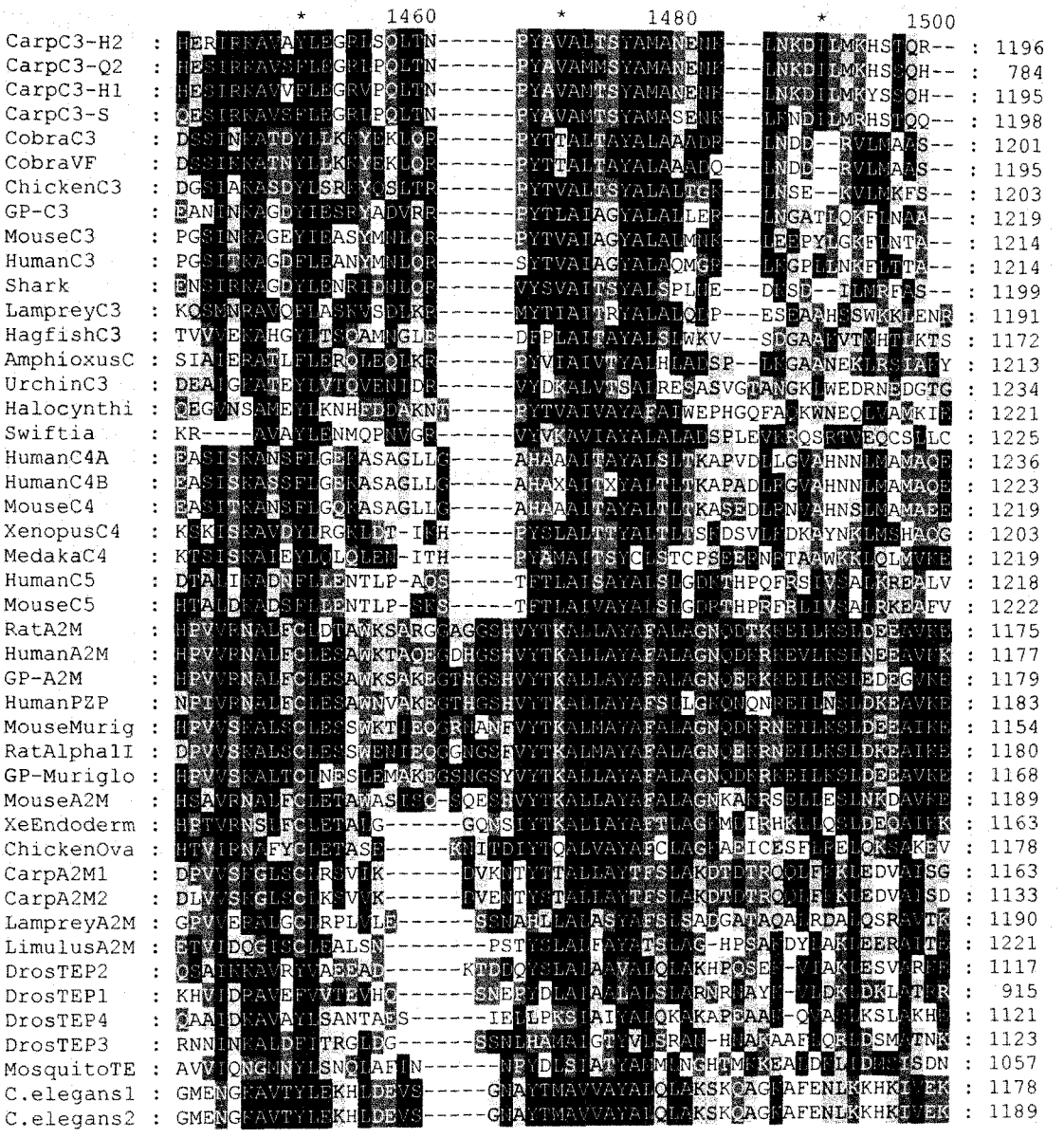

1160
1180
1200
CarpC3-H2 : 954
CarpC3-Q2 : 540
CarpC3-H1 : 951
CarpC3-S : 954
CobraC3 : 958
CobraVF : 952
ChickenC3 : 961
GP-C3 : 974
MouseC3 : 969
HumanC3 : 969
Shark : 957
LampreyC3 : 945
HagfishC3 : 929
AmphioxusC : 963
UrchinC3 : 983
Halocynthi : 971
Swiftia : 973
HumanC4A : 980
HumanC4B : 967
MouseC4 : 963
XenopusC4 : 955
MedakaC4 : 957
HumanC5 : 966
MouseC5 : 970
RatA2M : 936
HumanA2M : 938
GP-A2M : 940
HumanPZP : 944
MouseMurig : 940
RatAlpha1I : 941
GP-Muriglo : 929
MouseA2M : 950
XeEndoderm : 932
ChickenOva : 946
CarpA2M1 : 934
CarpA2M2 : 904
LampreyA2M : 960
LimulusA2M : 990
DrosTEP2 : 889
DrosTEP1 : 687
DrosTEP4 : 889
DrosTEP3 : 895
MosquitoTE : 825
C.elegans1 : 950
C.elegans2 : 961

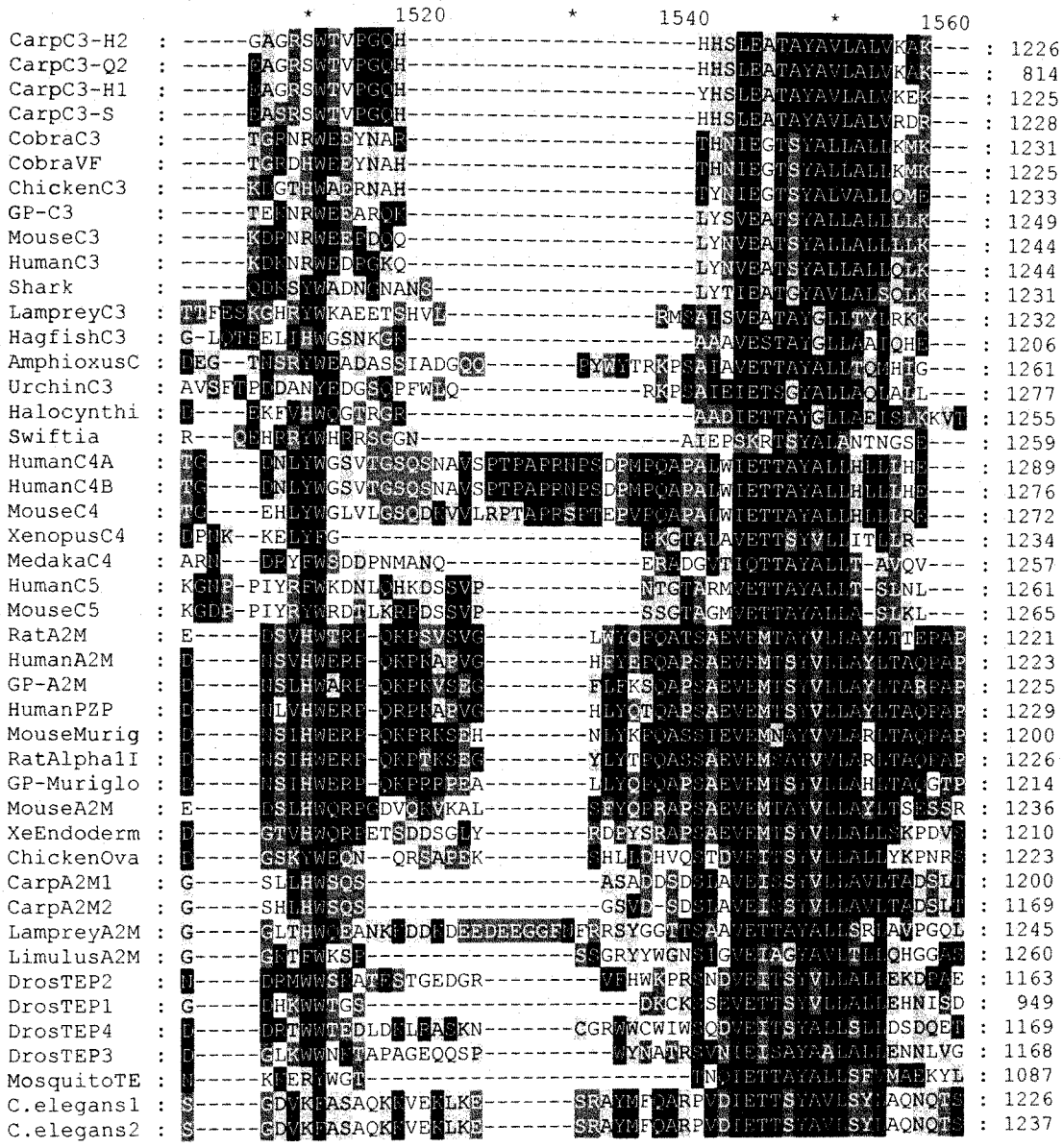

1220
1240
1260
CarpC3-H2 : 1001
CarpC3-Q2 : 587
CarpC3-H1 : 998
CarpC3-S : 1001
CobraC3 : 1005
CobraVF : 999
ChickenC3 : 1008
GP-C3 : 1021
MouseC3 : 1016
HumanC3 : 1016
Shark : 1004
LampreyC3 : 992
HagfishC3 : 976
AmphioxusC : 1019
UrchinC3 : 1038
Halocynthi : 1019
Swiftia : 1031
HumanC4A : 1029
HumanC4B : 1016
MouseC4 : 1012
XenopusC4 : 1002
MedakaC4 : 1005
HumanC5 : 1013
MouseC5 : 1017
RatA2M : 976
HumanA2M : 978
GP-A2M : 980
HumanPZP : 984
MouseMurig : 980
RatAlpha1I : 981
GP-Muriglo : 969
MouseA2M : 990
XeEndoderm : 972
ChickenOva : 986
CarpA2M1 : 974
CarpA2M2 : 944
LampreyA2M : 1000
LimulusA2M : 1030
DrosTEP2 : 929
DrosTEP1 : 727
DrosTEP4 : 929
DrosTEP3 : 935
MosquitoTE : 865
C.elegans1 : 990
C.elegans2 : 1001

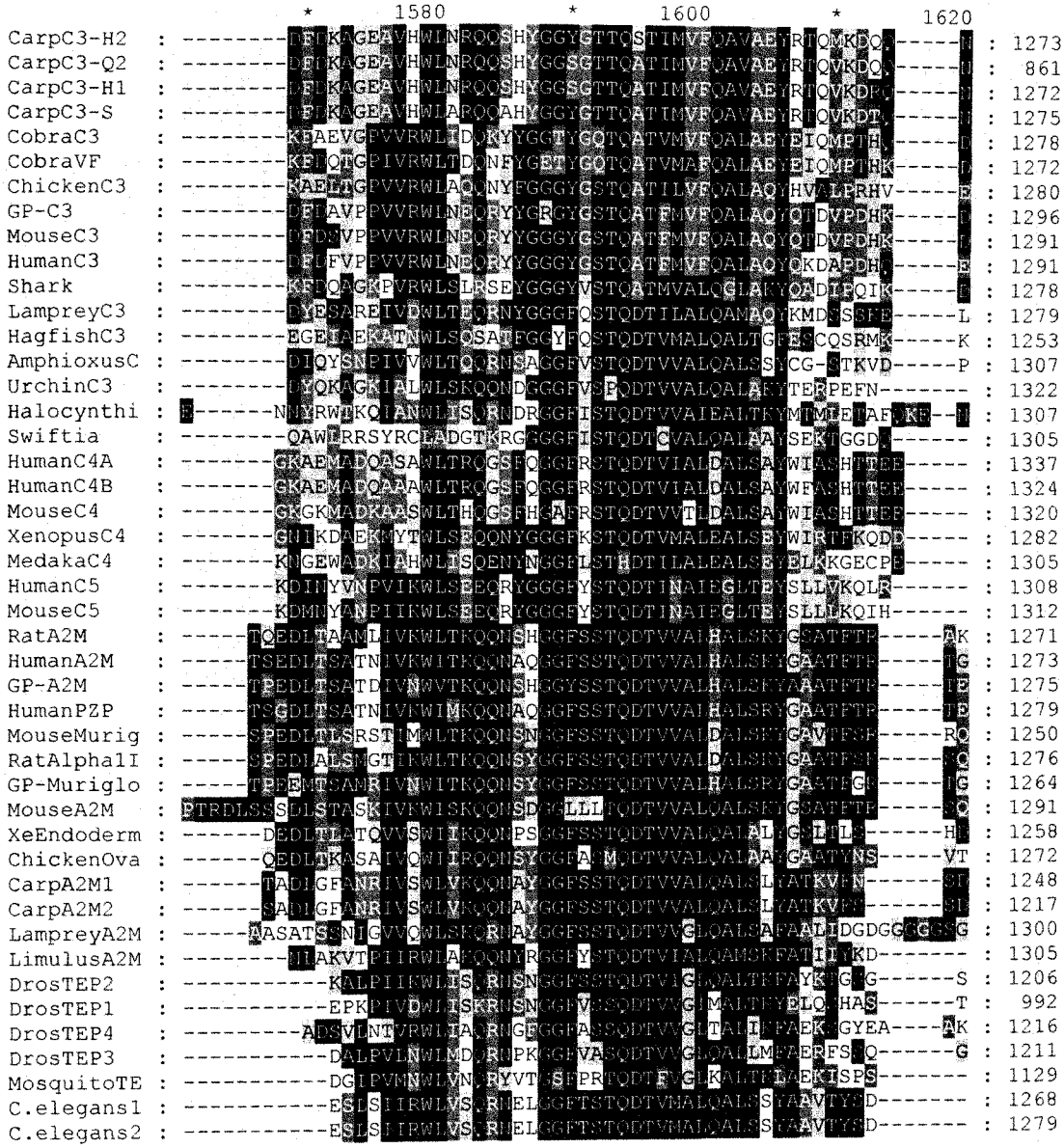

1280
1300
1320
CarpC3-H2 : 1055
CarpC3-Q2 : 641
CarpC3-H1 : 1052
CarpC3-S : 1055
CobraC3 : 1059
CobraVF : 1053
ChickenC3 : 1062
GP-C3 : 1075
MouseC3 : 1070
HumanC3 : 1070
Shark : 1058
LampreyC3 : 1046
HagfishC3 : 1031
AmphioxusC : 1070
UrchinC3 : 1089
Halocynthi : 1075
Swiftia : 1082
HumanC4A : 1083
HumanC4B : 1070
MouseC4 : 1066
XenopusC4 : 1056
MedakaC4 : 1062
HumanC5 : 1070
MouseC5 : 1074
RatA2M : 1030
HumanA2M : 1032
GP-A2M : 1034
HumanPZP : 1038
MouseMurig : 1034
RatAlpha1I : 1035
GP-Muriglo : 1023
MouseA2M : 1044
XeEndoderm : 1026
ChickenOva : 1040
CarpA2M1 : 1026
CarpA2M2 : 996
LampreyA2M : 1054
LimulusA2M : 1084
DrosTEP2 : 983
DrosTEP1 : 781
DrosTEP4 : 983
DrosTEP3 : 989
MosquitoTE : 916
C.elegans1 : 1044
C.elegans2 : 1055

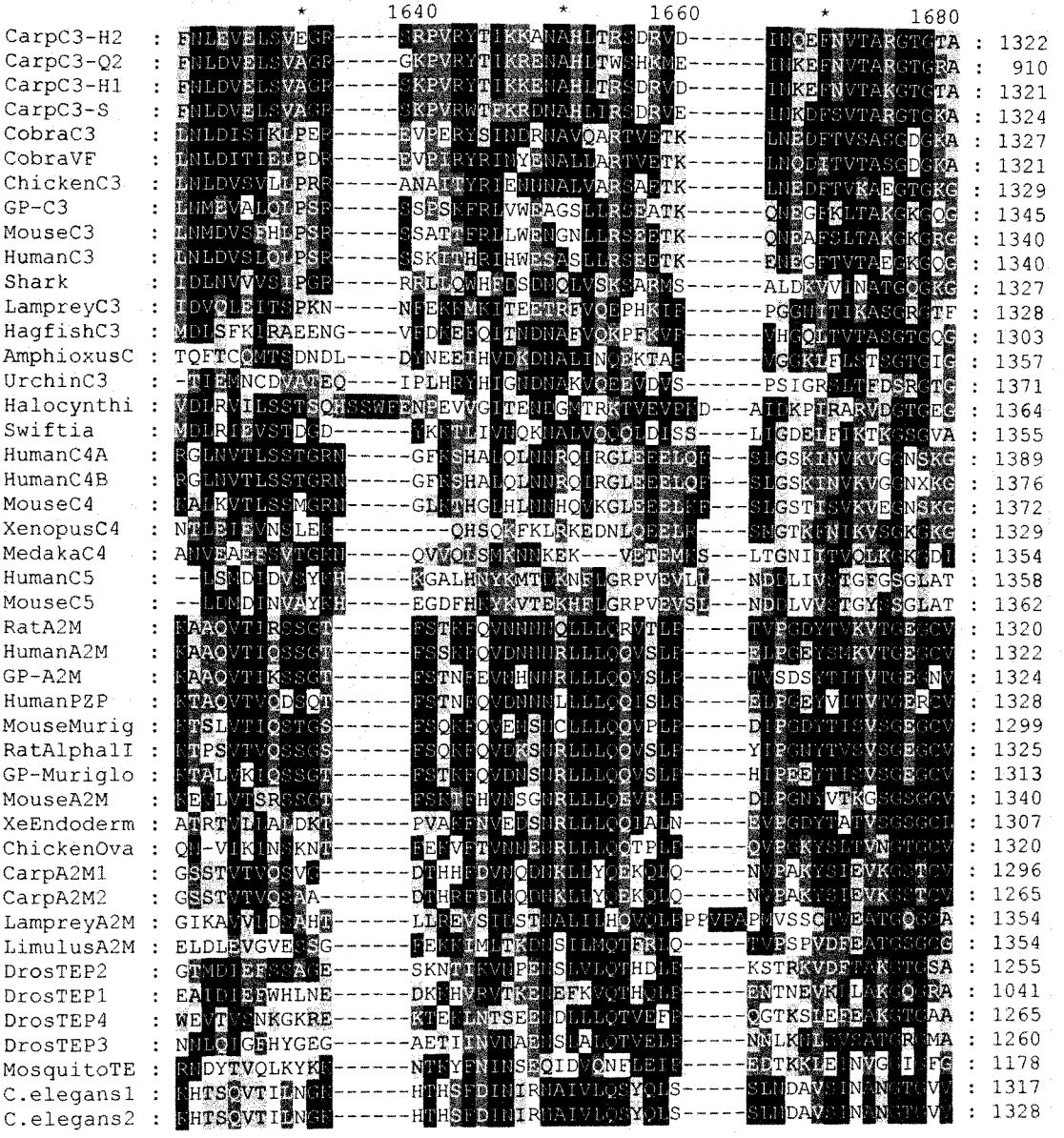

1340
1360
1380
CarpC3-H2 : 1097
CarpC3-Q2 : 683
CarpC3-H1 : 1094
CarpC3-S : 1097
CobraC3 : 1102
CobraVF : 1096
ChickenC3 : 1104
GP-C3 : 1117
MouseC3 : 1112
HumanC3 : 1112
Shark : 1100
LampreyC3 : 1088
HagfishC3 : 1072
AmphioxusC : 1116
UrchinC3 : 1131
Halocynthi : 1117
Swiftia : 1126
HumanC4A : 1124
HumanC4B : 1111
MouseC4 : 1107
XenopusC4 : 1097
MedakaC4 : 1115
HumanC5 : 1112
MouseC5 : 1116
RatA2M : 1072
HumanA2M : 1074
GP-A2M : 1076
HumanPZP : 1080
MouseMurig : 1076
RatAlpha1I : 1077
GP-Muriglo : 1065
MouseA2M : 1087
XeEndoderm : 1066
ChickenOva : 1080
CarpA2M1 : 1066
CarpA2M2 : 1036
LampreyA2M : 1095
LimulusA2M : 1124
DrosTEP2 : 1023
DrosTEP1 : 821
DrosTEP4 : 1030
DrosTEP3 : 1029
MosquitoTE : 958
C.elegans1 : 1084
C.elegans2 : 1095

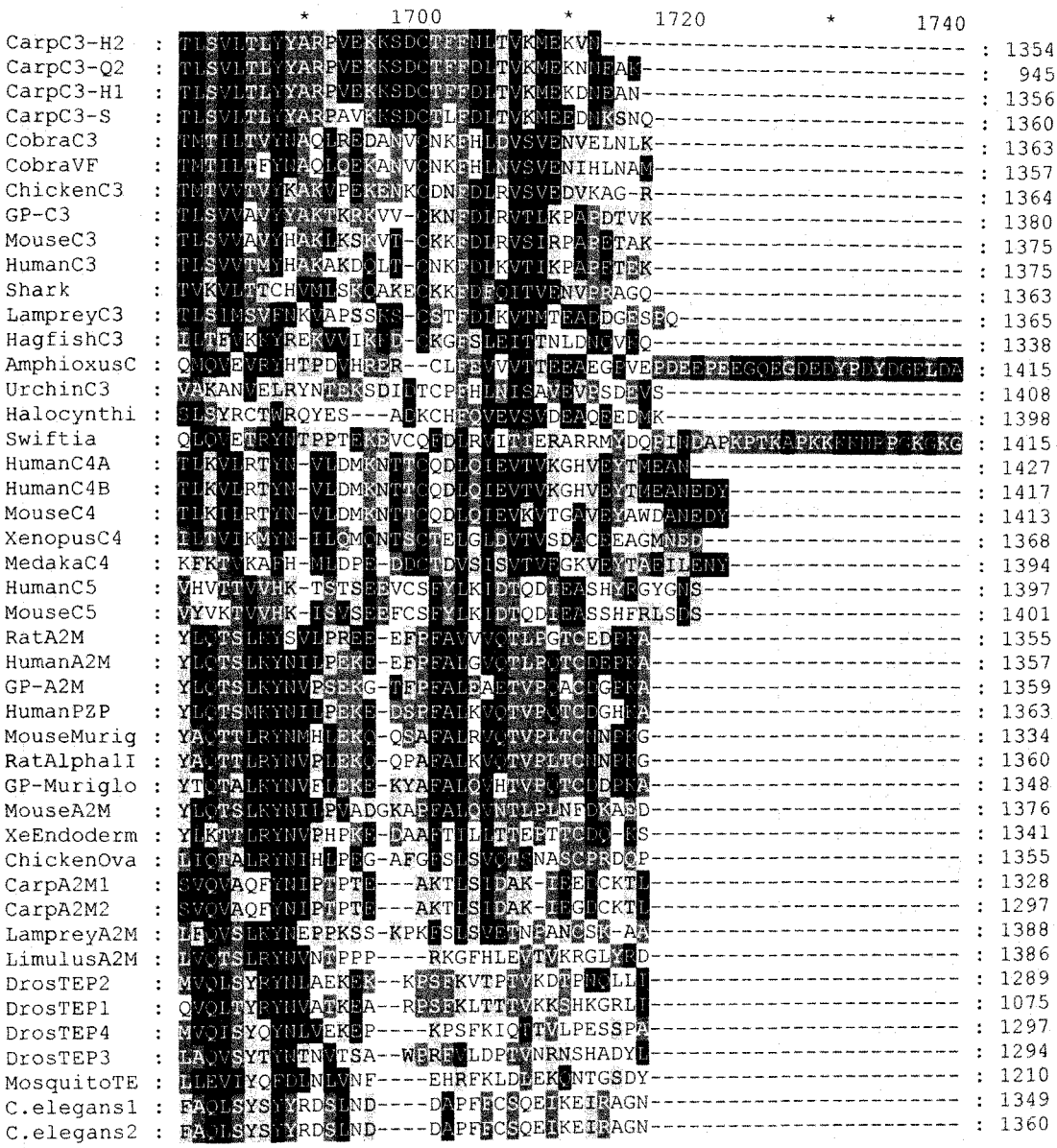

1400
1420
1440
CarpC3-H2 : 1147
CarpC3-Q2 : 735
CarpC3-H1 : 1146
CarpC3-S : 1149
CobraC3 : 1154
CobraVF : 1148
ChickenC3 : 1156
GP-C3 : 1170
MouseC3 : 1165
HumanC3 : 1165
Shark : 1152
LampreyC3 : 1140
HagfishC3 : 1121
AmphioxusC : 1161
UrchinC3 : 1180
Halocynthi : 1166
Swiftia : 1175
HumanC4A : 1181
HumanC4B : 1168
MouseC4 : 1164
XenopusC4 : 1149
MedakaC4 : 1165
HumanC5 : 1164
MouseC5 : 1168
RatA2M : 1115
HumanA2M : 1117
GP-A2M : 1119
HumanPZP : 1123
MouseMurig : 1094
RatAlpha1I : 1120
GP-Muriglo : 1108
MouseA2M : 1130
XeEndoderm : 1109
ChickenOva : 1123
CarpA2M1 : 1109
CarpA2M2 : 1079
LampreyA2M : 1136
LimulusA2M : 1170
DrosTEP2 : 1064
DrosTEP1 : 862
DrosTEP4 : 1068
DrosTEP3 : 1070
MosquitoTE : 1003
C.elegans1 : 1123
C.elegans2 : 1134

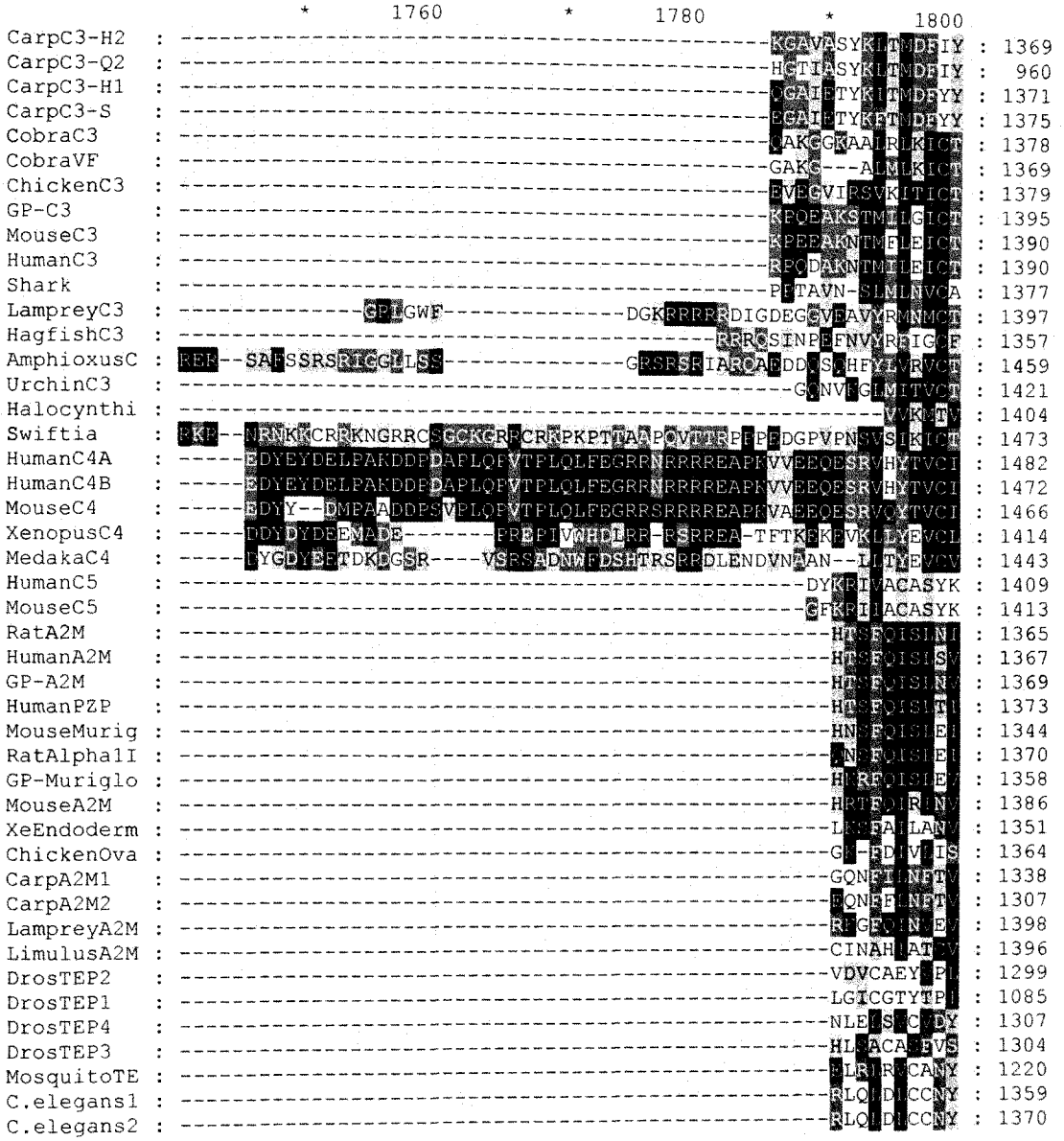

1460
1480
1500
CarpC3-H2 : 1196
CarpC3-Q2 : 784
CarpC3-H1 : 1195
CarpC3-S : 1198
CobraC3 : 1201
CobraVF : 1195
ChickenC3 : 1203
GP-C3 : 1219
MouseC3 : 1214
HumanC3 : 1214
Shark : 1199
LampreyC3 : 1191
HagfishC3 : 1172
AmphioxusC : 1213
UrchinC3 : 1234
Halocynthi : 1221
Swiftia : 1225
HumanC4A : 1236
HumanC4B : 1223
MouseC4 : 1219
XenopusC4 : 1203
MedakaC4 : 1219
HumanC5 : 1218
MouseC5 : 1222
RatA2M : 1175
HumanA2M : 1177
GP-A2M : 1179
HumanPZP : 1183
MouseMurig : 1154
RatAlpha1I : 1180
GP-Muriglo : 1168
MouseA2M : 1189
XeEndoderm : 1163
ChickenOva : 1178
CarpA2M1 : 1163
CarpA2M2 : 1133
LampreyA2M : 1190
LimulusA2M : 1221
DrosTEP2 : 1117
DrosTEP1 : 915
DrosTEP4 : 1121
DrosTEP3 : 1123
MosquitoTE : 1057
C.elegans1 : 1178
C.elegans2 : 1189

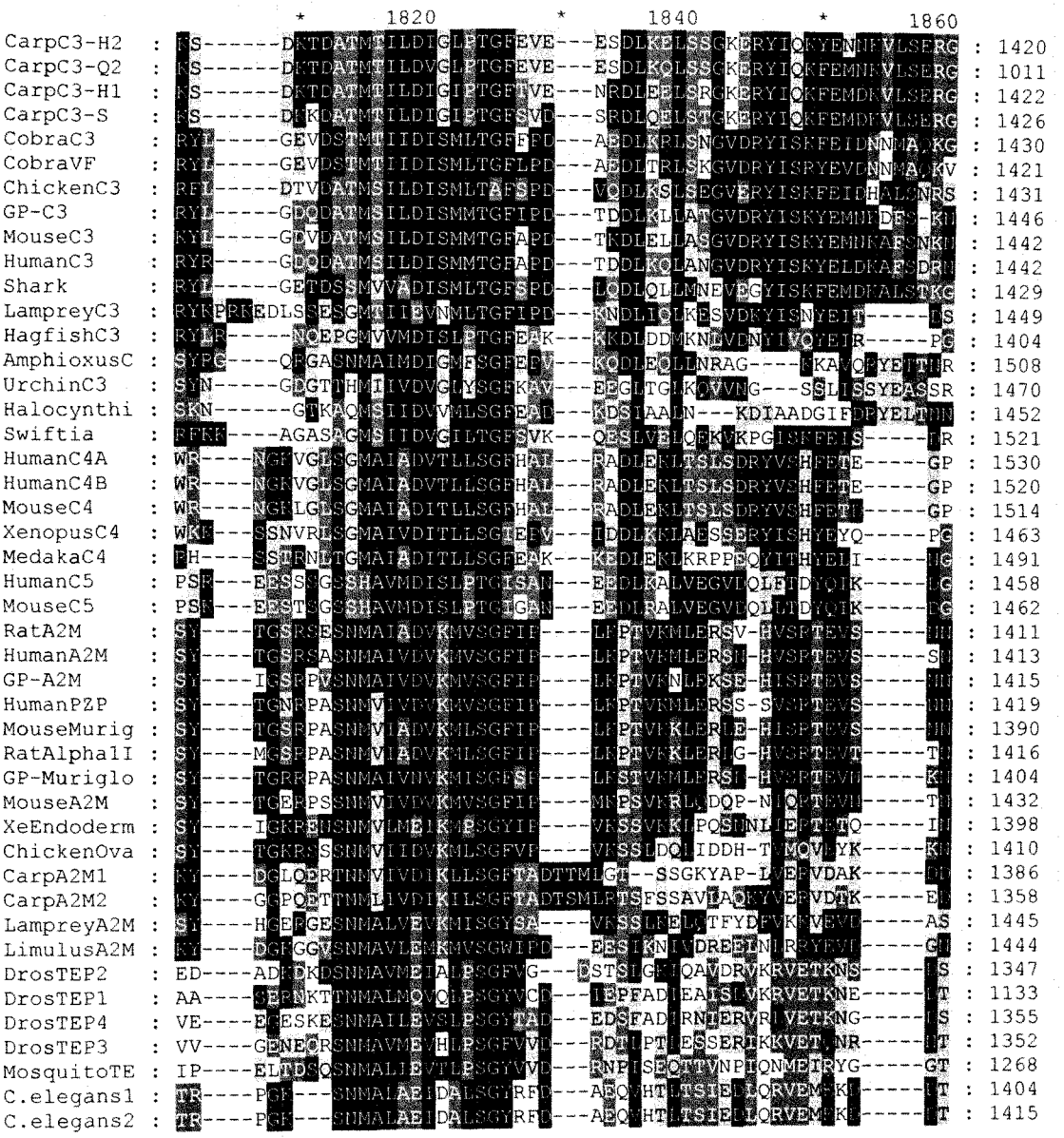

1520
1540
1560
CarpC3-H2 : 1226
CarpC3-Q2 : 814
CarpC3-H1 : 1225
CarpC3-S : 1228
CobraC3 : 1231
CobraVF : 1225
ChickenC3 : 1233
GP-C3 : 1249
MouseC3 : 1244
HumanC3 : 1244
Shark : 1231
LampreyC3 : 1232
HagfishC3 : 1206
AmphioxusC : 1261
UrchinC3 : 1277
Halocynthi : 1255
Swiftia : 1259
HumanC4A : 1289
HumanC4B : 1276
MouseC4 : 1272
XenopusC4 : 1234
MedakaC4 : 1257
HumanC5 : 1261
MouseC5 : 1265
RatA2M : 1221
HumanA2M : 1223
GP-A2M : 1225
HumanPZP : 1229
MouseMurig : 1200
RatAlpha1I : 1226
GP-Muriglo : 1214
MouseA2M : 1236
XeEndoderm : 1210
ChickenOva : 1223
CarpA2M1 : 1200
CarpA2M2 : 1169
LampreyA2M : 1245
LimulusA2M : 1260
DrosTEP2 : 1163
DrosTEP1 : 949
DrosTEP4 : 1169
DrosTEP3 : 1168
MosquitoTE : 1087
C.elegans1 : 1226
C.elegans2 : 1237


1580
1600
1620
CarpC3-H2 : 1273
CarpC3-Q2 : 861
CarpC3-H1 : 1272
CarpC3-S : 1275
CobraC3 : 1278
CobraVF : 1272
ChickenC3 : 1280
GP-C3 : 1296
MouseC3 : 1291
HumanC3 : 1291
Shark : 1278
LampreyC3 : 1279
HagfishC3 : 1253
AmphioxusC : 1307
UrchinC3 : 1322
Halocynthi : 1307
Swiftia : 1305
HumanC4A : 1337
HumanC4B : 1324
MouseC4 : 1320
XenopusC4 : 1282
MedakaC4 : 1305
HumanC5 : 1308
MouseC5 : 1312
RatA2M : 1271
HumanA2M : 1273
GP-A2M : 1275
HumanPZP : 1279
MouseMurig : 1250
RatAlpha1I : 1276
GP-Muriglo : 1264
MouseA2M : 1291
XeEndoderm : 1258
ChickenOva : 1272
CarpA2M1 : 1248
CarpA2M2 : 1217
LampreyA2M : 1300
LimulusA2M : 1305
DrosTEP2 : 1206
DrosTEP1 : 992
DrosTEP4 : 1216
DrosTEP3 : 1211
MosquitoTE : 1129
C.elegans1 : 1268
C.elegans2 : 1279

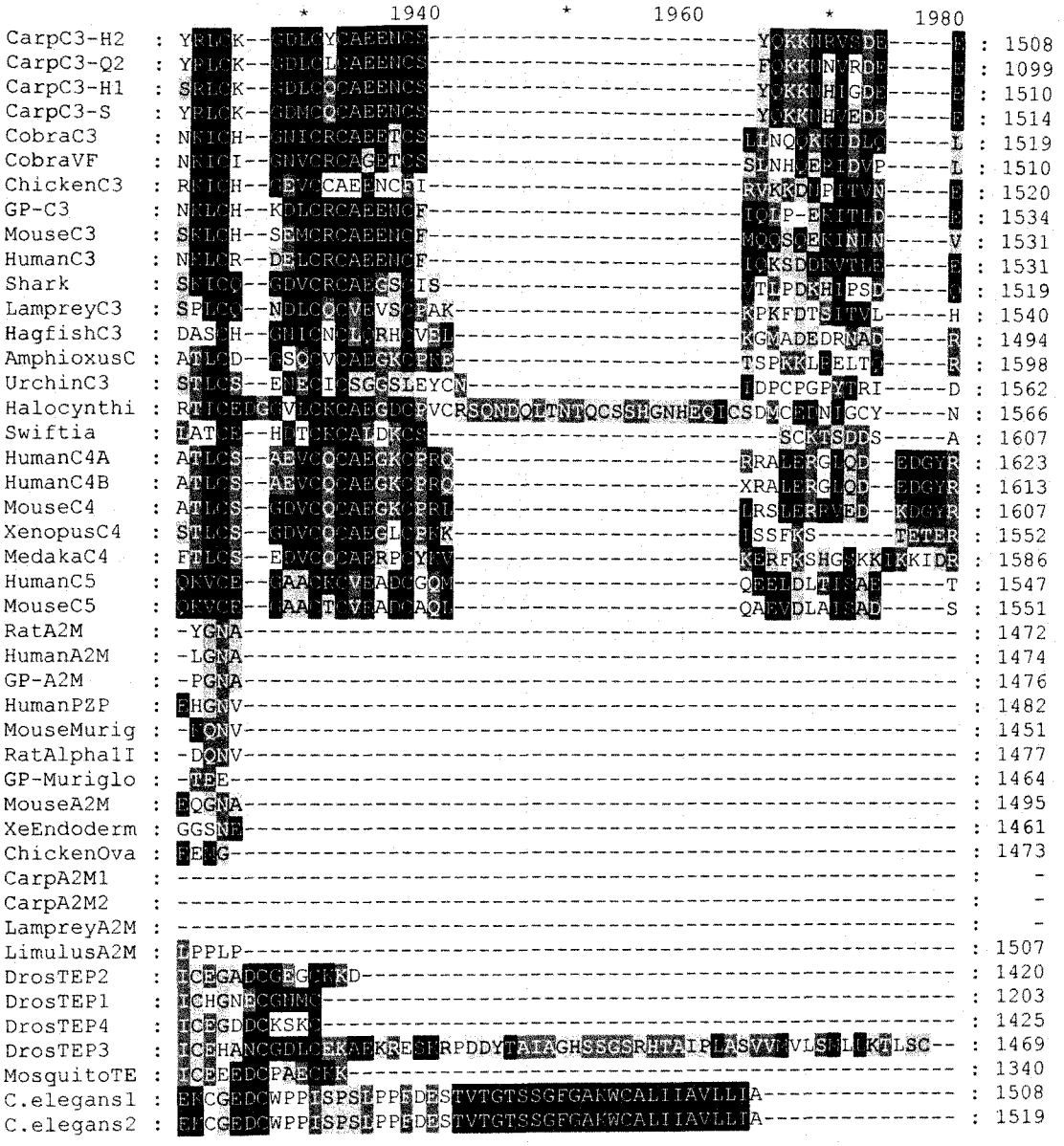

1640
1660
1680
CarpC3-H2 : 1322
CarpC3-Q2 : 910
CarpC3-H1 : 1321
CarpC3-S : 1324
CobraC3 : 1327
CobraVF : 1321
ChickenC3 : 1329
GP-C3 : 1345
MouseC3 : 1340
HumanC3 : 1340
Shark : 1327
LampreyC3 : 1328
HagfishC3 : 1303
AmphioxusC : 1357
UrchinC3 : 1371
Halocynthi : 1364
Swiftia : 1355
HumanC4A : 1389
HumanC4B : 1376
MouseC4 : 1372
XenopusC4 : 1329
MedakaC4 : 1354
HumanC5 : 1358
MouseC5 : 1362
RatA2M : 1320
HumanA2M : 1322
GP-A2M : 1324
HumanPZP : 1328
MouseMurig : 1299
RatAlpha1I : 1325
GP-Muriglo : 1313
MouseA2M : 1340
XeEndoderm : 1307
ChickenOva : 1320
CarpA2M1 : 1296
CarpA2M2 : 1265
LampreyA2M : 1354
LimulusA2M : 1354
DrosTEP2 : 1255
DrosTEP1 : 1041
DrosTEP4 : 1265
DrosTEP3 : 1260
MosquitoTE : 1178
C.elegans1 : 1317
C.elegans2 : 1328

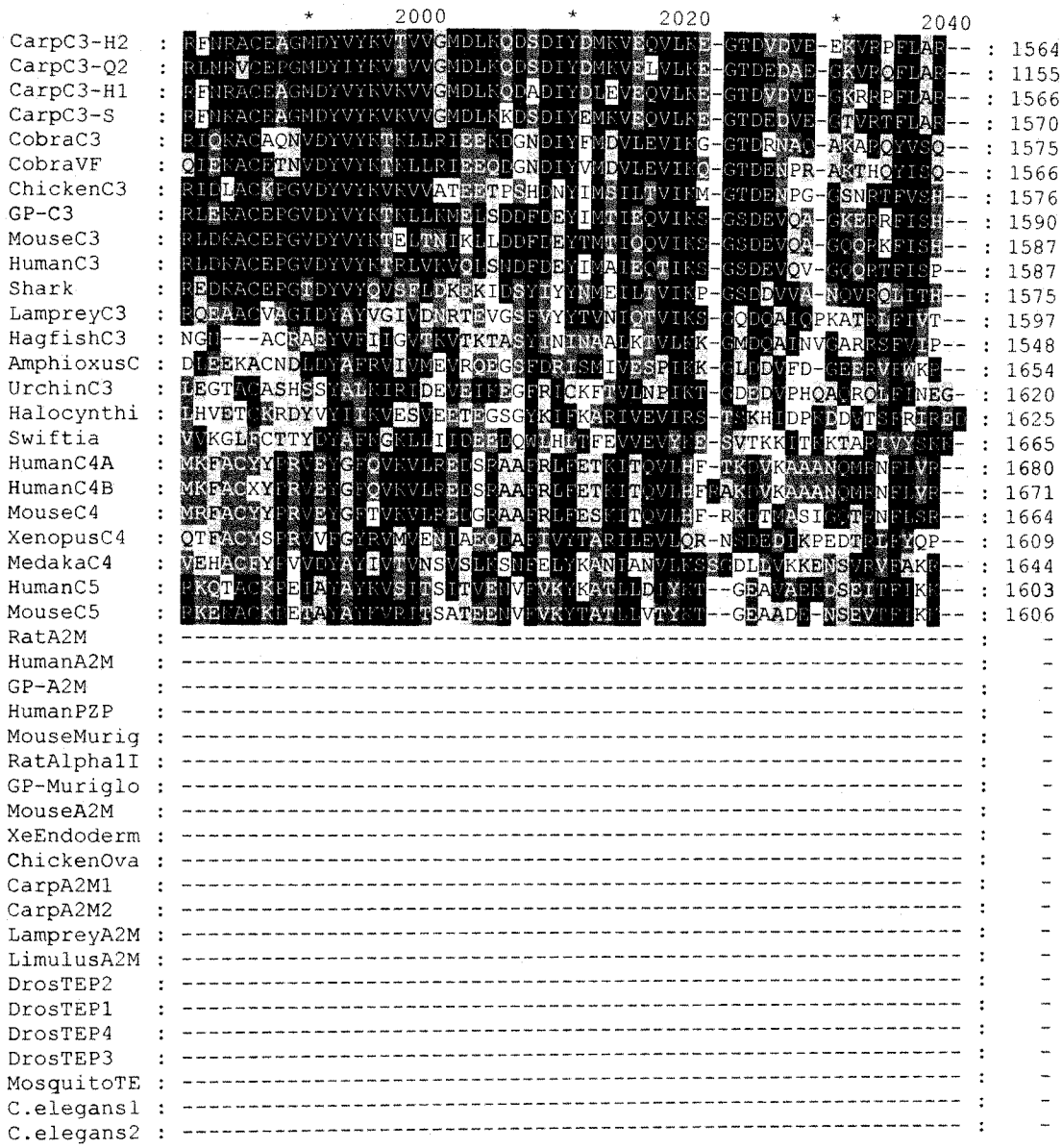

1700
1720
1740
CarpC3-H2 : 1354
CarpC3-Q2 : 945
CarpC3-H1 : 1356
CarpC3-S : 1360
CobraC3 : 1363
CobraVF : 1357
ChickenC3 : 1364
GP-C3 : 1380
MouseC3 : 1375
HumanC3 : 1375
Shark : 1363
LampreyC3 : 1365
HagfishC3 : 1338
AmphioxusC : 1415
UrchinC3 : 1408
Halocynthi : 1398
Swiftia : 1415
HumanC4A : 1427
HumanC4B : 1417
MouseC4 : 1413
XenopusC4 : 1368
MedakaC4 : 1394
HumanC5 : 1397
MouseC5 : 1401
RatA2M : 1355
HumanA2M : 1357
GP-A2M : 1359
HumanPZP : 1363
MouseMurig : 1334
RatAlpha1I : 1360
GP-Muriglo : 1348
MouseA2M : 1376
XeEndoderm : 1341
ChickenOva : 1355
CarpA2M1 : 1328
CarpA2M2 : 1297
LampreyA2M : 1388
LimulusA2M : 1386
DrosTEP2 : 1289
DrosTEP1 : 1075
DrosTEP4 : 1297
DrosTEP3 : 1294
MosquitoTE : 1210
C.elegans1 : 1349
C.elegans2 : 1360

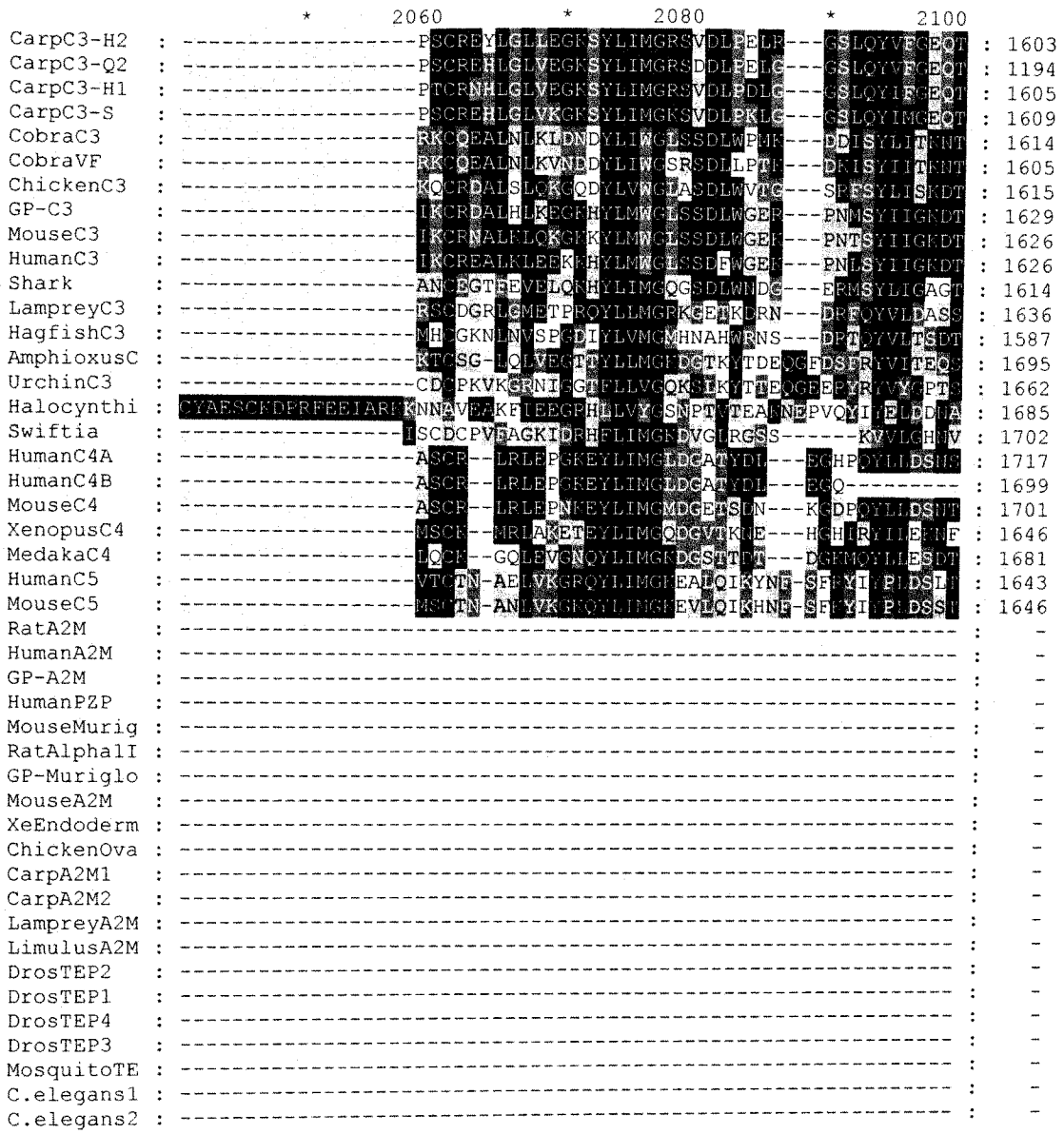

1760
1780
1800
CarpC3-H2 : 1369
CarpC3-Q2 : 960
CarpC3-H1 : 1371
CarpC3-S : 1375
CobraC3 : 1378
CobraVF : 1369
ChickenC3 : 1379
GP-C3 : 1395
MouseC3 : 1390
HumanC3 : 1390
Shark : 1377
LampreyC3 : 1397
HagfishC3 : 1357
AmphioxusC : 1459
UrchinC3 : 1421
Halocynthi : 1404
Swiftia : 1473
HumanC4A : 1482
HumanC4B : 1472
MouseC4 : 1466
XenopusC4 : 1414
MedakaC4 : 1443
HumanC5 : 1409
MouseC5 : 1413
RatA2M : 1365
HumanA2M : 1367
GP-A2M : 1369
HumanPZP : 1373
MouseMurig : 1344
RatAlpha1I : 1370
GP-Muriglo : 1358
MouseA2M : 1386
XeEndoderm : 1351
ChickenOva : 1364
CarpA2M1 : 1338
CarpA2M2 : 1307
LampreyA2M : 1398
LimulusA2M : 1396
DrosTEP2 : 1299
DrosTEP1 : 1085
DrosTEP4 : 1307
DrosTEP3 : 1304
MosquitoTE : 1220
C.elegans1 : 1359
C.elegans2 : 1370


1820
1840
1860
CarpC3-H2 : 1420
CarpC3-Q2 : 1011
CarpC3-H1 : 1422
CarpC3-S : 1426
CobraC3 : 1430
CobraVF : 1421
ChickenC3 : 1431
GP-C3 : 1446
MouseC3 : 1442
HumanC3 : 1442
Shark : 1429
LampreyC3 : 1449
HagfishC3 : 1404
AmphioxusC : 1508
UrchinC3 : 1470
Halocynthi : 1452
Swiftia : 1521
HumanC4A : 1530
HumanC4B : 1520
MouseC4 : 1514
XenopusC4 : 1463
MedakaC4 : 1491
HumanC5 : 1458
MouseC5 : 1462
RatA2M : 1411
HumanA2M : 1413
GP-A2M : 1415
HumanPZP : 1419
MouseMurig : 1390
RatAlpha1I : 1416
GP-Muriglo : 1404
MouseA2M : 1432
XeEndoderm : 1398
ChickenOva : 1410
CarpA2M1 : 1386
CarpA2M2 : 1358
LampreyA2M : 1445
LimulusA2M : 1444
DrosTEP2 : 1347
DrosTEP1 : 1133
DrosTEP4 : 1355
DrosTEP3 : 1352
MosquitoTE : 1268
C.elegans1 : 1404
C.elegans2 : 1415


1880
1900
1920
CarpC3-H2 : 1480
CarpC3-Q2 : 1071
CarpC3-H1 : 1482
CarpC3-S : 1486
CobraC3 : 1490
CobraVF : 1481
ChickenC3 : 1491
GP-C3 : 1506
MouseC3 : 1502
HumanC3 : 1502
Shark : 1489
LampreyC3 : 1509
HagfishC3 : 1463
AmphioxusC : 1567
UrchinC3 : 1530
Halocynthi : 1511
Swiftia : 1581
HumanC4A : 1589
HumanC4B : 1579
MouseC4 : 1573
XenopusC4 : 1522
MedakaC4 : 1550
HumanC5 : 1516
MouseC5 : 1520
RatA2M : 1468
HumanA2M : 1470
GP-A2M : 1472
HumanPZP : 1477
MouseMurig : 1447
RatAlpha1I : 1473
GP-Muriglo : 1461
MouseA2M : 1490
XeEndoderm : 1456
ChickenOva : 1469
CarpA2M1 : 1442
CarpA2M2 : 1406
LampreyA2M : 1503
LimulusA2M : 1502
DrosTEP2 : 1406
DrosTEP1 : 1192
DrosTEP4 : 1414
DrosTEP3 : 1411
MosquitoTE : 1327
C.elegans1 : 1463
C.elegans2 : 1474


1940
1960
1980
CarpC3-H2 : 1508
CarpC3-Q2 : 1099
CarpC3-H1 : 1510
CarpC3-S : 1514
CobraC3 : 1519
CobraVF : 1510
ChickenC3 : 1520
GP-C3 : 1534
MouseC3 : 1531
HumanC3 : 1531
Shark : 1519
LampreyC3 : 1540
HagfishC3 : 1494
AmphioxusC : 1598
UrchinC3 : 1562
Halocynthi : 1566
Swiftia : 1607
HumanC4A : 1623
HumanC4B : 1613
MouseC4 : 1607
XenopusC4 : 1552
MedakaC4 : 1586
HumanC5 : 1547
MouseC5 : 1551
RatA2M : 1472
HumanA2M : 1474
GP-A2M : 1476
HumanPZP : 1482
MouseMurig : 1451
RatAlpha1I : 1477
GP-Muriglo : 1464
MouseA2M : 1495
XeEndoderm : 1461
ChickenOva : 1473
CarpA2M1 : -
CarpA2M2 : -
LampreyA2M : -
LimulusA2M : 1507
DrosTEP2 : 1420
DrosTEP1 : 1203
DrosTEP4 : 1425
DrosTEP3 : 1469
MosquitoTE : 1340
C.elegans1 : 1508
C.elegans2 : 1519


2000
2020
2040
CarpC3-H2 : 1564
CarpC3-Q2 : 1155
CarpC3-H1 : 1566
CarpC3-S : 1570
CobraC3 : 1575
CobraVF : 1566
ChickenC3 : 1576
GP-C3 : 1590
MouseC3 : 1587
HumanC3 : 1587
Shark : 1575
LampreyC3 : 1597
HagfishC3 : 1548
AmphioxusC : 1654
UrchinC3 : 1620
Halocynthi : 1625
Swiftia : 1665
HumanC4A : 1680
HumanC4B : 1671
MouseC4 : 1664
XenopusC4 : 1609
MedakaC4 : 1644
HumanC5 : 1603
MouseC5 : 1606
RatA2M :
HumanA2M :
GP-A2M :
HumanPZP :
MouseMurig :
RatAlpha1I :
GP-Muriglo :
MouseA2M :
XeEndoderm :
ChickenOva :
CarpA2M1 :
CarpA2M2 :
LampreyA2M :
LimulusA2M :
DrosTEP2 :
DrosTEP1 :
DrosTEP4 :
DrosTEP3 :
MosquitoTE :
C.elegans1 :
C.elegans2 :


2060
2080
2100
CarpC3-H2 : 1603
CarpC3-Q2 : 1194
CarpC3-H1 : 1605
CarpC3-S : 1609
CobraC3 : 1614
CobraVF : 1605
ChickenC3 : 1615
GP-C3 : 1629
MouseC3 : 1626
HumanC3 : 1626
Shark : 1614
LampreyC3 : 1636
HagfishC3 : 1587
AmphioxusC : 1695
UrchinC3 : 1662
Halocynthi : 1685
Swiftia : 1702
HumanC4A : 1717
HumanC4B : 1699
MouseC4 : 1701
XenopusC4 : 1646
MedakaC4 : 1681
HumanC5 : 1643
MouseC5 : 1646
RatA2M : -
HumanA2M : -
GP-A2M : -
HumanPZP : -
MouseMurig : -
RatAlpha1I : -
GP-Muriglo : -
MouseA2M : -
XeEndoderm : -
ChickenOva : -
CarpA2M1 : -
CarpA2M2 : -
LampreyA2M : -
LimulusA2M : -
DrosTEP2 : -
DrosTEP1 : -
DrosTEP4 : -
DrosTEP3 : -
MosquitoTE : -
C.elegans1 : -
C.elegans2 : -


2120
2140
CarpC3-H2 : 1640
CarpC3-Q2 : 1230
CarpC3-H1 : 1642
CarpC3-S : 1646
CobraC3 : 1651
CobraVF : 1642
ChickenC3 : 1652
GP-C3 : 1666
MouseC3 : 1663
HumanC3 : 1663
Shark : 1651
LampreyC3 : 1673
HagfishC3 : 1620
AmphioxusC : 1732
UrchinC3 : 1698
Halocynthi : 1740
Swiftia : 1728
HumanC4A : 1754
HumanC4B : -
MouseC4 : 1738
XenopusC4 : 1683
MedakaC4 : 1718
HumanC5 : 1676
MouseC5 : 1680
RatA2M : -
HumanA2M : -
GP-A2M : -
HumanPZP : -
MouseMurig : -
RatAlpha1I : -
GP-Muriglo : -
MouseA2M : -
XeEndoderm : -
ChickenOva : -
CarpA2M1 : -
CarpA2M2 : -
LampreyA2M : -
LimulusA2M : -
DrosTEP2 : -
DrosTEP1 : -
DrosTEP4 : -
DrosTEP3 : -
MosquitoTE : -
C.elegans1 : -
C.elegans2 : -

VITA

LARRY J. DISHAW

DATE OF BIRTH 17 March 1975

PLACE OF BIRTH Hialeah, Fl

CIVIL STATUS Married

EDUCATION B.S. Biology, Florida International University, 1997

PROFESSIONAL EXPERIENCE

6/92 - 8/97	Lab Research Technician, Marine Laboratory, Dept. of Biology, FIU.
6/98 - 8/98	Tropical Biology Summer Fellowship ($2000), FIU.
6/99 - 8/99	Tropical Biology Summer Fellowship ($2000), FIU.
8/97 -12/99	Teaching Assistant, Immunology Lab, Genetics Lab, and Marine Biology.
6/00 - 8/00	Tropical Biology Summer Fellowship ($2000), FIU.
6/00	Travel Award ($1000), American Association of Immunologists.
6/01 - 8/01	Tropical Biology Summer Fellowship ($2000), FIU.
6/01 - 8/01	Comparative Immunology Institute Fellowship ($2000), FIU.
01/00- 09/02	Funded by the NIH MBRS Rise program.
02/02	Travel Award ($1000), American Society of Biochemistry and Molecular Biology.
5/97 – 12/02	Graduate Student, Biology, Florida International University.

RESEARCH INTERESTS

Molecular phylogeny of immune response genes.
Invertebrate immunology & comparative immunology.

DISSERTATION TOPIC

Immunophylogenetic aspects of a coral.

PUBLICATIONS AND PRESENTATIONS

Dishaw, L. J. and C.H. Bigger. 1999. Coral Microsatellite DNA. First Annual Biology Research Symposium. Oral Presentation.

Dishaw, L. J. and C.H. Bigger. 1999. Microsatellite Anchored DNA-fingerprinting in a Gorgonian Coral, *Swiftia exserta.* Mid Atlantic Society

of Developmental and Comparative Immunology. Washington, D.C. (Oral Presentation).

Bigger, C.H., C. Olano, and L.J. Dishaw. 1999. Mechanisms of allograft rejection in the gorgonian coral Swiftia exserta. Mid Atlantic Society of Developmental and Comparative Immunology. Washington, D.C.

Dishaw, L. J. and C.H. Bigger. 2000. Microsatellite Anchored DNA-fingerprinting in a Gorgonian Coral, *Swiftia exserta.* Second Annual Tropical Biology Research Symposium. (Oral Presentation).

Dishaw, L.J., S.L. Smith, and C.H. Bigger. 2000. Sequence analysis of partial cDNA clones from a coral, encoding a thiolester-containing protein. 8th Congress of the International Society of Developmental and Comparative Immunology. Cairns, Australia. (Oral Presentation).

Dishaw, L.J., S.L. Smith, and C.H. Bigger. 2001. Sequence and phylogenetic analysis of partial cDNA clones from a coral, encoding a thiolester-containing A2M/C3-like protein. FASEB. 15(4): 531.2.

Piatigorsky, J., B. Norman, L.J. Dishaw, L. Kos, J. Horwitz, P.J. Steinbach, and Z. Kozmik. 2001. J3-crystallin of the jellyfish lens: Similarity to saposins. Proceedings of the National Academy of Sciences. 98(22): 12362-67.

Dishaw, L.J., M.L. Herrera, S.L. Smith, and C.H. Bigger. 2002. Evidence for an ancestral C3-like protein in a gorgonian coral, *Swiftia exserta.* Fourth Annual Tropical Biology Research Symposium. (Oral Presentation).

Dishaw, L.J., M.L. Herrera, S.L. Smith, and C.H. Bigger. 2002. Evidence for an ancestral C3-like protein in a gorgonian coral, *Swiftia exserta. FASEB.* 16(4): 510.6.

Dishaw, L.J., C.H. Bigger, and S.L. Smith. 2002. Ancient origins for the C3-like ancestor of complement components C3/C4/C5. *International Immunopharmacology.* 2(3): 1293.

Dishaw, L. J., S.L. Smith, and C.H. Bigger. 2002. Precambrian origins for the C3-like ancestor of vertebrate C3/C4/C5. Submitted.