STUDIES OF A SPERM ANTIGEN RECOGNIZED BY TWO MONOCLONAL ANTIBODIES

By

Patricia McChesney

B.A. The University of Colorado, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE FACULTY OF GRADUATE STUDIES

THE DEPARTMENT OF OBSTETRICS AND GYNAECOLOGY
REPRODUCTIVE AND DEVELOPMENTAL SCIENCES PROGRAMME

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 1994

© Patricia McChesney, 1994
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

The University of British Columbia
Vancouver, Canada

Date 9 Sept 94
Abstract

The monoclonal antibodies HSA-5 and HSA-6 were generated by others in Dr. Lee's lab against ionophore-treated acrosome reacted human sperm in 1989. They were submitted to the second WHO workshop for interlaboratory evaluations of their potential to be immunocontraceptive vaccine candidates, and were concluded to be "high priority candidates" in 1989.

The two antibodies recognize identical bands on a western blot of crude human sperm extract, which indicate that they recognize the same protein (HSAg-5/6). Indirect immunofluorescent assays performed in our lab have shown that the HSAg-5/6 is located in the equatorial region of methanol-fixed human sperm, as well as on the head and tail of mouse sperm. The antigen was immunolocalized on human sperm by Homyk et al (1993) who showed that HSAg-5/6 was located on both the inner and outer acrosomal membranes. The corresponding antigen from mouse sperm which binds to HSA-5 and HSA-6 was found to have characteristics identical to those of the human protein as analyzed by western blot assay.

The purpose of my project was to study the molecular nature of human HSAg-5/6, and isolate cDNA clones of the corresponding antigen from mouse. This characterization included studying the purified human antigen, and molecular cloning of the homologous protein from a mouse testis cDNA library.

From work performed by Dr. T. Yoshiki, it was established that purified HSAg-5/6 is a protein with a molecular mass of approximately 60kD. My studies established that the
isolated cDNA clone expresses the C-terminal region of the protein and accounts for approximately 13% of the total molecular weight, based upon the approximated molecular weight of the cloned fusion protein compared to the molecular weight of the native HSAg-5/6. However, antigenic sites for both monoclonal antibodies lie in this tail end, which suggests that the C-terminus of HSAg-5/6 is exposed to the environment such that an immune response occurs.

The studies of HSAg-5/6 described in this thesis not only supplement our current knowledge regarding sperm-specific proteins, but also aid in development of an immunocontraceptive vaccine.
**Table of Contents**

Abstract ................................................. ii

Table of Contents ........................................ iv

List of Abbreviations ..................................... vii

List of Figures ............................................ ix

Acknowledgment ........................................... x

Introduction ................................................ 1

Natural and Induced Immunoinfertility ................. 1

Anti-fertility Vaccines and Advantages ............... 2

Criteria for Candidates of Immunocontraceptive Vaccines .... 2

The Anti-Sperm vaccines .................................. 4

Potential Sperm Antigens for Antifertility Vaccine Development .... 6

HSAg-5/6 .................................................. 9

Objectives ............................................... 17

Materials and Methods .................................... 18

Chemicals .................................................. 18

Immunofluorescent Assays of Mouse Sperm and Mouse Testis .... 18

Immunoaffinity Chromatography .......................... 19

Peptide Sequencing of Purified Antigen .................. 20

Immunoscreening of Mouse Testis cDNA Library .......... 20

Fusion Protein Analysis .................................. 22

Western Blot .............................................. 23
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSA-5</td>
<td>Human Sperm monoclonal antibody, group A-5</td>
</tr>
<tr>
<td>HSA-6</td>
<td>Human Sperm monoclonal antibody, group A-6</td>
</tr>
<tr>
<td>HSAg-5/6</td>
<td>Cognate antigen recognized by HSA-5 and HSA-6</td>
</tr>
<tr>
<td>IgM</td>
<td>Immunoglobulin, M class</td>
</tr>
<tr>
<td>IgA</td>
<td>Immunoglobulin, A class</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin, G class</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecylsulfate-Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxynucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary Deoxynucleic acid</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein isothiocyanate</td>
</tr>
<tr>
<td>HCl</td>
<td>Hydrochloric acid</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>E. Coli</td>
<td>Escherichia Coli</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris Buffered saline</td>
</tr>
<tr>
<td>TBS-T</td>
<td>Tris buffered saline with 0.5% Tween-20</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecylsulfate</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>dH₂O</td>
<td>Deionized water</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>rpm</td>
<td>Rotations per minute</td>
</tr>
<tr>
<td>HBS-9505</td>
<td>Monoclonal antibody, class M, against Hepatitis B</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxy Adenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxy Cytosine triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxy Thymidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxy Guanidine triphosphate</td>
</tr>
<tr>
<td>λgt11</td>
<td>Species of phage used as a vector for cloning</td>
</tr>
<tr>
<td>Taq polymerase</td>
<td>DNA polymerase isolated from Thermus Aquaticus</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>kb</td>
<td>kilobases, or 1000 base pairs</td>
</tr>
<tr>
<td>UV light</td>
<td>Ultraviolet light</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction™, Perkin-Elmer Cetus</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Alignment Search Tool</td>
</tr>
<tr>
<td>kD</td>
<td>kilodalton</td>
</tr>
<tr>
<td>EcoR1</td>
<td>Restriction Enzyme, also site of cDNA insertion in λgt11 phage</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region following the termination codon in mRNA</td>
</tr>
<tr>
<td>ASA</td>
<td>Anti-sperm antibody</td>
</tr>
<tr>
<td>WHO</td>
<td>The World Health Organization</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene difluoride</td>
</tr>
</tbody>
</table>
List of Figures

Fig.

1 Immunofluorescent Assay of HSA-5 binding to the tail of mouse sperm. [11]

2 Indirect immunofluorescent assay of HSA-6 monoclonal antibody binding to frozen testicular sections of adult mouse. [14]

3 An SDS-PAGE gel (12% acrylamide) of HSAG-5/6 purified on an immunoaffinity chromatography column. [25]

4 Nitrocellulose membranes displaying plaques of E.Coli lysate containing β-galactosidase fusion protein which have positively bound to polyclonal HSA-5/6 antibody and to monoclonal HSA-5/HSA-6 antibody cocktail. [29]

5 Western blot analysis of β-galactosidase fusion protein. [32]

6 Nucleotide sequence of Clone #5 and deduced amino acid sequence. [35]

7 Diagram illustrating the β-galactosidase protein, and the predicted size of the resulting fusion protein. [39]

8 Hydrophilicity plot of the deduced amino acid sequence of Clone #5. [42]
Acknowledgment

While several people deserve credit for helping me accomplish this goal, there are two special people who have earned extra gratitude. First, I would like to thank my lab partner and dear friend Letticia. Without her, I would have not been able to continue my work. To my beloved husband Darrell, who has supported me through the best and worst of times, I am eternally indebted.

Additionally, I would like to thank Dr. T. Yoshiki for allowing me to display his data prior to journal publication (Fig. 1 and Fig. 2).

Thank you.
Introduction

Natural and Induced Immunoinfertility

It is well known that some sperm antigens are auto- or isoimmunogenic in humans.

It has been known for almost one hundred years that spermatozoa can cause an immune response when introduced into areas of the body which are not part of the reproductive tract (Landsteiner, 1899; Metchnikoff, 1899). Approximately 70% of vasectomized men form anti-sperm antibodies, and 30% of infertility is associated with anti-sperm antibodies in either the male or female partner (Naz, 1990). Furthermore, it has been shown that treatment of immunoinfertility with immunosuppressive agents results in lower antibody titres and often leads to successful conception (Hendry et al, 1979; Shulman, 1986).

Sperm antigens which lead to antibody formation are expressed after meiosis, when the sperm are protected from systemic circulation by the blood-testis barrier. As a result, these antigens are foreign to the immune system (Witkin, 1988). The sperm are also exposing "foreign" antigens to the female reproductive tract. However, it was observed that "autoantibodies to spermatozoa in males and isoantibodies in females are inhibited both by the physical isolation of spermatozoa from the systemic immune system and by active immunosuppression mechanisms" (Witkin, 1988)

Immune reactions are inhibited in men by T-suppressor cells in semen and in women by activation of T-suppressor cells following coitus, as well as physical barriers in both genders (Witkin, 1988).
However, introduction of sperm or sperm antigens into the systemic circulation leads to an immune response.

Anti-fertility Vaccines and Advantages

Experiments using spermatozoa as a vaccine to prevent fertility were conducted in the 1920’s and 1930’s. In 1932, Baskin published a report of 20 women who had received three intramuscular injections of fresh semen, seven days apart. In all cases but one the injections appeared harmless, and sperm cytotoxicity in serum was maintained for up to twelve months. There were no pregnancies reported under these conditions. However, the follow-up period was too short to assess subsequent fertility or health (Baskin, 1932).

Immunococontraception refers to utilizing the immune system to inhibit reproduction, and has a number of advantages over traditional methods of preventing pregnancy. Vaccines are not pharmacologically active and are specific to the immune system. Use and administration of vaccines can be accomplished by nurses or paramedical personnel with relative ease (Jones, 1982).

Criteria for Candidates of Immunococontrceptive Vaccines

Development of an immunococontrceptive vaccine requires consideration of a variety of factors. The most important is identifying an antigen which will not induce an immunopathological condition in the vaccine recipient. Therefore, the antigen must be specific to the tissues in the reproductive system. It has been suggested that there are a number of proteins found on sperm surface that are shared with somatic cell
plasma membranes (Naz et al, 1990). For example, a variety of laboratory induced ASA's have been shown to cross react with brain (Freund et al, 1955), kidney (Chaffee and Schachner, 1978) and erythrocytes (Kerek, 1974). There is also a risk that non-specific antibodies could react with soluble proteins normally found in body fluids such as milk, saliva and serum. It has been generally agreed that the utility of an antigen as a contraceptive vaccine is contingent upon its tissue specificity, its involvement in fertility and upon a sufficient immune response within the genital tracts such that fertility is intercepted (Naz et al, 1990).

The antigen chosen for immunocontraception should not normally be found in the recipient (Jones WR, 1982; Griffin D, 1990), or the antigen should be present transiently (as in the case of sperm presence in the female reproductive tract). Alternatively, a consistently present antigen may be shown to be in such low concentrations that a "low-grade immune response" would be insignificant to the vaccine recipient.

The capability of an antibody to intercept fertility requires that the antigen be accessible to the antibody and that the local immune response in the genital tract is sufficient to prevent conception. This requirement was demonstrated when a vaccine against acrosin was tested (De Ioannes et al, 1990). Acrosin is a sperm protein essential to fertilization. The resulting antisera was shown to prevent in vitro fertilization in mice (De Ioannes et al, 1990). However, when the antibody was used as a passive vaccine, it was ineffective when systemic immunization was induced in fertile female mice (De Ioannes et al, 1990). Despite a high serum antibody titre, the immune response in the reproductive tract was insufficient to prevent conception in vivo (Naz et al, 1990).
The final requirement of an contraceptive vaccine is that it must be produced in mass quantities. The antigen used for the vaccine must be produced under artificial conditions in order to accommodate the high demand. For example, it may be expressed as a fusion protein in a bacteria culture or it may be manufactured in a peptide synthesizing apparatus. Natural sources of antigen such as human tissues or fluids are impractical for large-scale vaccine production.

The World Health Organization’s (WHO) criteria for an immunocontraceptive vaccine suggested that a good vaccine candidate will (Griffin, 1990):

1. provide effective, safe fertility regulation when antigen is eliminated
2. be specific to its intended target
3. insure that the immune reaction is present in a controlled site and does not lead to immunopathology
4. include an antigen which is present transiently or at low levels compared to the immune response it elicits
5. will not elicit undesired responses
   eg. IgE production or an allergic reaction
6. can be chemically characterized and produced

The Anti-Sperm vaccines

One of the areas of active research in antifertility vaccines is antisperm agents. The World Health Organization (WHO) has recognized five steps involved in vaccine development (Griffin, 1990):

1. identification of events in reproduction which are accessible to immune intervention
For example, sperm-zona pellucida binding and entry of sperm into the female reproductive tract are events that are exposed to the immunesystem and are accessible to intervention.

2. identification of molecules which are necessary for reproduction whose elimination will result in safe, effective antifertility

3. development of vaccines utilizing those molecules

4. pre-clinical testing of the vaccine for safety and efficacy on lab animals

5. clinical testing of the vaccine for safety and efficacy in humans

One approach to identify sperm proteins which are involved in the fertilization process has been antifertility studies or in vitro fertilization experiments. The first step of the process can be accomplished by immunizing mice with human sperm. If the sperm are fresh then the antigens exposed to the murine immune system are those found on the surface of the sperm. The antigens and resulting antibodies have provided valuable information about the epididymal maturation of a spermatozoa (Myles et al, 1990), but more research is required to prove their efficacy (Mitchison, 1990). On the other hand, if the sperm have undergone the acrosome reaction prior to immunizing the mice, then the antigens exposed to the immune system may include those components which are essential for fertilization or sperm-egg plasma membrane associations.

The procedure of immunizing mice with target materials is the primary step of monoclonal antibody production. The resulting immune response provides valuable antibody producing B-cells which can be harvested from the murine spleen. These cells are fused with immortal mouse myeloma cells according to established
procedures (Davis et al, 1982). The new hybridoma cell line is cultured and tested for antibody production (Lee et al, 1984; Lee et al, 1984; Lee et al, 1984). Once antibody production has been established, then research into the nature of the target antigen commences.

The second step of the vaccine development process (Griffin, 1990) includes studying those antigens which illicit an immune response, and verification that the antibody production will prevent fertility. After experimenting with a variety of antibodies and cognate antigens, Lee and Wong (1986) concluded that monoclonal antibodies which bind to the acrosomal region of a sperm most effectively prevent fertilization in vitro and in vivo.

Potential Sperm Antigens for Antifertility Vaccine Development

The development of anti-sperm vaccine candidates has thus far progressed to the second step of WHO's program (Griffin, 1990), identifying antigens which illicit an immune response. Several candidates have been identified and studied.

Since 1983, Naz has been studying antigens found on the surface of intact spermatozoa. Antisera against Fertility Antigen-1 (FA-1) has been found to reduce human sperm fusion with zona-free hamster oocytes (Naz et al, 1984). It also inhibits fertility in female rabbits immunized with FA-1 (Naz et al, 1987).

Goldberg (1990) reported extensively on lactate dehydrogenase-C₄ (LDH-C₄). This enzyme is a sperm-specific (Wheat et al, 1983; Liang et al, 1986) isomer of lactate dehydrogenase and is found mainly on the tail of sperm (Goldberg, 1963; Blanco et al, 1963). This enzyme is not produced in females, and is isolated from the
male immune system by the blood-testis barrier. When used as a vaccine, LDH-C₄ provokes an immune response in both genders. However, active immunization against this protein in female mice, rabbits and baboons resulted in only partial fertility suppression (Goldberg, 1975; Wheat et al, 1983).

Two antigens found in guinea pig sperm, called PH-20 and PH-30 have been studied by Primakoff and Myles (Myles et al 1984, Primakoff et al 1985). PH-20 is a 64kD glycoprotein located on the postacrosome of guinea pig sperm. Nucleic acid/peptide sequence analysis revealed that it is a cell adhesion related protein (Blobel et al 1992). Active immunization against PH-20 resulted in complete, reversible sterilization of either male or female animals (Primakoff et al 1988). PH-30 consists of two subunits, and is essential for sperm-oocyte fusion (Blobel et al 1992). Immunization against PH-30 resulted in partial fertility suppression.

Leyton et al (1992) have studied a 95kD mouse sperm protein called P-95. This protein has been identified as a putative receptor with tyrosine kinase activity. It binds ZP-3 on mouse zonae pellucidae, and may participate in regulation of gamete interaction. P-95 is considered an excellent target for immunoregulation of fertility. However, a homologous human antigen has not been fully studied.

Rabbit Sperm Autoantigen (RSA) is a zona-pellucida binding protein isolated from rabbits by O’Rand (O’Rand et al, 1988). This lectin-like protein binds to the zona pellucida with high affinity (O’Rand et al, 1988) and has a molecular weight of 65-70kD. Western blot analyses using anti-RSA antibodies have shown that similar proteins are present in human and pig sperm (O’Rand et al, 1988). Interestingly, O’Rand has found a ten amino acid peptide sequence, called P10G, within the RSA family which represents an autoantigenic epitope that specifically binds to rabbit
autoantisera (O’Rand and Widgren, 1989). They have shown by ELISA that antibodies to RSA and P10G bind to human sperm lysate in a dose dependent manner (O’Rand et al, 1990). More studies are required to isolate the human RSA analog protein to determine its zona binding capabilities and structural similarity to RSA and the P10G peptide (O’Rand et al, 1990).

Herr et al have been studying a sperm specific protein called SP-10, recognized by the monoclonal antibody MHS-10 (Herr et al, 1990 a-d; Homyk et al, 1990; Write et al, 1990). This antigen is located in the intra-acrosomal region of human sperm. Western blot analysis demonstrated SP-10 to be a group of proteins ranging from 18kD to 34kD (Herr et al, 1990a). WHO-sponsored interlaboratory studies of the MHS-10 monoclonal antibody showed that it inhibited interactions between sperm and zona-free hamster oocytes (Anderson et al, 1987).

MSA-63 is an antigen found in the acrosome of mouse sperm and binds to the monoclonal antibody HS-63 (Liu et al, 1989). HS-63 was shown to inhibit in vitro fertilization of mouse (Lee et al, 1990). HS-63 was also shown to inhibit binding of human sperm to zona-free hamster oocytes. The purified MSA-63 is comprised of three major soluble glycoproteins with molecular weights of 50kD, 43kD and 42kD (Liu et al, 1989; Lee et al, 1990). MSA-63 has also been cloned from a mouse testis cDNA library. The full length cDNA is 1067 bases in length and includes a poly-A tail region and an open reading frame with 261 amino acids. The predicted protein size of MSA-63 is 27.9kD (Liu et al, 1992).

A sperm/placenta cross reacting antigen called STX-10 has been studied (Lee et al, 1993). This 75kD group of glycoproteins is found on the acrosome of methanol-fixed human sperm and is recognized by the monoclonal antibody HSA-10.
Human sperm penetration of zona-free hamster oocytes was reduced by nearly 75% in the presence of HSA-10 when compared to an unrelated antibody used as a negative control (Lee et al, 1993). HSA-10 was labeled as a "high priority" vaccine candidate by the second WHO workshop (Anderson, 1992).

The monoclonal antibodies HSA-5 and HSA-6 were developed in 1989. They were two of seven antibodies generated against ionophore-treated acrosome reacted human sperm. Both antibodies are IgM class and have been propagated in BALB/c mice as ascites fluid.

Following purification by 40% ammonium sulfate precipitation and passage through a Sephacyryl S-300 gel filtration column, the antibodies were conjugated to Affigel-10 agarose according to the protocol provided by BioRad Lab. This immunoaffinity column was used to purify the antigen from human semen extract. Rabbit antisera against the affinity-isolated cognate antigen were raised according to the standard protocol (Liu et al, 1989). The monoclonal antibody from ascites fluid and rabbit antisera were utilized to study the cognate antigen.

The HSA-5 and HSA-6 antibodies were submitted to the second WHO Workshop for interlaboratory evaluations of their potential to be immunocontraceptive vaccine candidates. Of the 69 submissions, only five were considered "high priority candidates," including HSA-5 and HSA-6 (Anderson, 1990). The interlaboratory evaluations determined that the antibodies do not react with adult somatic tissue of
human and mouse. It was also determined that HSA-5 and HSA-6 react with the sperm of human, mouse and chimpanzee. The antibodies were designated S71 and S72 by WHO Workshop (Anderson, 1992).

Furthermore, the monoclonal antibodies were shown to recognize identical bands on a western blot of crude human sperm extract, which indicated that they recognize the same protein (Anderson, 1992; Yoshiki et al, unpublished). However, sandwich enzyme immunoassays (SEIA) determined that each monoclonal antibody may recognize different antigenic regions (Yoshiki et al, unpublished).

Indirect immunofluorescent assays, performed by Dr. T. Yoshiki, of human and mouse sperm have shown that the antigen is located in the equatorial region of the acrosomal of methanol-fixed human sperm (data not shown). Similar assays demonstrated binding of the monoclonal antibodies to an antigen localized to the head and tail of mouse sperm. Fig. 1 shows HSA-5 binding to the tail of mouse sperm.
Fig. 1 Immunofluorescent Assay of HSA-5 binding to the tail of mouse sperm.

A: Sperm as seen under light microscopy
B: Sperm as seen under Ultraviolet light
   400X Magnification, photographic enlargement

Data courtesy of Dr. T. Yoshiki.
HSAg-5/6 protein has been localized in the acrosome of the human sperm and primarily on the tail of mouse sperm. The antigen was immunolocalized on human sperm by Homyk et al (1993) using light and transmission electron microscopy. Their research showed that HSAg-5/6 was located on both the inner and outer acrosomal membranes.

Frozen sections of mouse testis were examined by indirect immunofluorescent assay for antigen binding to HSA-5 and HSA-6. Immunohistological studies of frozen mouse testis sections were performed using the published protocol of Lee and Wong (1986) by Dr. T. Yoshiki (Fig. 2). These studies demonstrated that HSAg-5/6 is expressed post-meiotically in the mouse testis.
Fig. 2  Indirect immunofluorescent assay of HSA-6 monoclonal antibody binding to frozen testicular sections of adult mouse. Note the staining of the sperm tail is apparent in the center of the photo. 400X Magnification, photographic enlargement included

Data courtesy of Dr. T. Yoshiki.
Western blot assays were utilized to determine the molecular weight of the antigen recognized by the HSA-5 and HSA-6 monoclonal antibodies (Dr. T. Yoshiki, unpublished observations).

Western blot assay revealed that a 60 kilodalton (kD) protein band was immunoreactive with monoclonal antibodies HSA-5 and HSA-6 (not shown). Parallel western blot analysis of crude sperm extract confirmed an immunoreactive band at 60kD. A similar western blot was probed with rabbit antisera raised against the purified cognate antigen. The antisera recognized multiple bands in crude human sperm extract ranging from 11.5kD to 106kD. Both monoclonal and polyclonal antibodies recognized the 60kD band.

The corresponding antigen from mouse sperm which binds to HSA-5 and HSA-6 was also found to have characteristics identical to those of the human protein as analyzed by Western blot assay.

Following assessment of the molecular size and location of HSAg-5/6 on human and mouse sperm, it became important to study the protein on the molecular level. Such studies reveal the molecular nature, sequence, and structure of a protein.

In collaboration with our lab, Dr Mona Homyk, in the laboratory of Dr. J.C. Herr at the University of Virginia have cloned HSAg-5/6 from a human testis cDNA library (Homyk et al, unpublished observations). They have obtained a partial length sequence of the cDNA for this antigen. However, their cDNA fragment is only 600 base pairs (bp) in length and constitutes the C-terminal region of the protein. For our part of the collaboration, we set out to perform a similar study using a mouse testis cDNA library.
Objectives

The outcome of this project is to characterize the cognate antigen recognized by the monoclonal antibodies HSA-5 and HSA-6. Characterization of this protein will not only supplement current knowledge of sperm physiology, but it will also aid in developing an immunocontraceptive vaccine.

The purpose of my portion of the project was to study the molecular nature of human HSAg-5/6, and isolate cDNA clones of the corresponding antigen gene from mouse. This characterization included examining the antigen as isolated from human sperm extract, and molecular cloning the homologous protein from a mouse testis cDNA library.
Materials and Methods

Chemicals

The monoclonal antibodies HSA-5 and HSA-6 (in both ascites fluid and purified solution), as well as the polyclonal antisera were generously donated by Dr. C.Y.G. Lee. Triton X-100 was purchased from Sigma (St. Louis, MO). All analytic-grade reagents for gel electrophoresis, nitrocellulose and PVDF membranes for protein blot were obtained from BIO-RAD (Richmond, CA). Embedding medium for testicular sections was Tissue TekII, O.C.T. Compound No. 4583, from Lab-Tek. Bacterial culture media supplies were from Difco (Burlington, ON, Canada). Disposable petri dishes were purchased from Fisher Scientific (Vancouver, BC). Molecular biological supplies, fluorescein-isothiocyanate (FITC)-labeled goat anti-mouse/rabbit IgG+IgA+IgM, and horseradish peroxidase labeled goat anti-mouse/rabbit IgG+IgA+IgM were obtained from Gibco/BRL (Burlington, ON, Canada). The λgt11 phage mouse testis cDNA library was a gift from Dr. C. Lau of the University of California, San Francisco. Immunoscreening kits were from Clontech (Palo Alto, CA). Sequencing primers for λgt11 were purchased from New England Biolabs (New Hampshire).

Immunofluorescent Assays of Mouse Sperm and Mouse Testis

Fresh semen samples were obtained from the epididymus' of sacrificed mice. Sperm were washed three times in Ham's F-10, and allowed to swim up for 30 minutes at 37°C, and were diluted to 10⁴ sperm/ml. Drops of 5µl each were applied to 8-well slides and dried on a hot plate at 37°C for approximately five minutes.
Once dry, the sperm were fixed to the slides by immersion in methanol for ten minutes. Blocking solution consisting of 0.5% Bovine Serum Albumin and 0.1% Thimerosal in Phosphate buffered saline (PBS-BSA-TMS) was used to wash the slides three times prior to incubation in the solution for ten minutes. The sperm were probed with mouse monoclonal HSA-5 or HSA-6 ascites fluid in diluted in blocking buffer for one hour at 37°C. Following probing the slides were washed three times in the blocking solution. The primary antibody was detected by Goat Anti-mouse IgG + IgA + IgM conjugated to FITC for one hour at 37°C. After three final washes, the slides were examined under white and fluorescent light for antibody binding.

Similar assays were performed with mouse testicular sections. For these experiments, testes were removed from male mice 21 days after birth and frozen immediately at liquid nitrogen temperatures (approx. -196°C). They were then placed in an embedding medium at -60°C. Sections of 10µm thickness were cut with a cryotome and placed on slides. After air drying, the slides were rinsed with PBS-BSA-TMS and used for indirect immunofluorescence assay as described above.

This work was performed by Dr. T. Yoskiki and is displayed in Fig. 1 and Fig. 2.

Imunoaffinity Chromatography

An immunoaffinity column was prepared to purify HSAg-5/6 by Dr. C.Y.G. Lee. Monoclonal antibodies HSA-5 and HSA-6 were conjugated to Affigel-10 agarose according to the protocol provided by BioRad Lab.

In my experiments, the column was washed with 20ml of 0.1M glycine-HCl (pH=2.2) and equilibrated with 40ml PBS. Approximately 500ml of crude human
semen extract from frozen stocks, containing an average of 37 mg/ml protein, was passed through the column. After rinsing the column thoroughly with PBS, the bound proteins were eluted with 0.1M acetic acid, or glycine-HCl (pH=2.2). After neutralizing the pH with 1M Tris-Cl (pH=8.0) the fractions were quantitated spectrophotometrically at absorbance of 280 nm (O.D.). Samples containing the highest concentration of protein were pooled and further concentrated employing AMICON Centriprep concentrators. The protein samples were examined by SDS-PAGE using published methods (Laemmli, 1971).

Peptide Sequencing of Purified Antigen

Purified HSAg-5/6 was obtained and run on an SDS-PAGE (12 % acrylamide) gel. The protein bands were electrophoretically transferred to a PVDF membrane in a BIO-RAD transblot apparatus at 60 V for three hours. Following transfer, the membrane was rinsed with deionized water and stained with Coomassie Fast Blue (.025 %) for five minutes. The membrane was then de-stained overnight in a solution containing 50 % methanol and 50 % deionized water. The visible bands at 19 kD and 11 kD were cut out and allowed to dry at room temperature before being sent to the Biochemistry Department at the University of Victoria for peptide sequencing by the Edman degradation method.

Immunoscreening of Mouse Testis cDNA Library

A single colony of E. Coli strain Y1090* was isolated and grown overnight in LB broth containing 2% maltose at 37°C with constant agitation. Two hundred microlitres of the liquid culture was inoculated with approximately 7.5 x 10⁴
recombinant λgt11 phage. The mixture was incubated at 37°C for 20 minutes. LB soft top agar was added to a volume of 7 ml and poured onto LB agar plates containing 50μg/ml Ampicillin. The culture was then allowed to grow at 42°C for 2.5 - 3.5 hours. This procedure was repeated for 10 - 12 150mm LB agar plates. Following the initial incubation, 150mm nitrocellulose filters, previously saturated with 10mM IPTG, were overlayed onto the plates. The plates were then incubated in an incubator at 37°C for an additional 3.5 hours to allow growth of phage, expression of fusion protein, and lysis of the E. Coli. The filters were then removed from the plates and incubated in a blocking solution consisting of Tris-buffered saline containing 0.5% Tween-20 and 5% powdered skim milk for 1 hour at room temperature. Antibody probing of the filters began when both monoclonal antibodies HSA-5 and HSA-6 (ascites fluid) were added to the blocking solution to a concentration of 1 μl/ml. The filters were incubated in the antibody solution overnight at room temperature with constant agitation. The next morning, the filters were washed in Tris-buffered saline containing 0.5% Tween-20 three times for 5 minutes each and incubated in the blocking solution containing goat anti-mouse IgG+IgM+IgA conjugated with horseradish peroxidase (1 μl/ml) for 3 - 4 hours. After three further washes with TBS-T, and one with TBS the filters were incubated in a peroxidase substrate solution containing 3,3'-diaminobenzidine for 10 - 20 minutes to develop color.

This procedure was repeated with modifications for immunoscreening with the rabbit polyclonal antibody HSA-5/6. After the initial filters were removed, a second IPTG saturated filter was overlayed onto the same plate. The incubation times and conditions for the filters remained as above. However, antibody probing involved
rabbit antiserum diluted in blocking solution (1 μl/ml) followed by incubation with goat anti-rabbit IgG+IgM+IgA conjugated with horseradish peroxidase (1 μl/ml).

Positive plaques were identified and isolated for a second screening by the same method. Only those plaques which were positive for both antibody probes were chosen for subsequent screenings. All clones were screened at least four times or until plaques were 100% positive.

Following immunoscreening, positive clones were confirmed by an additional screening using antibodies which had been pre-absorbed in E. Coli lysate according to the published method (Sambrook et al, 1989).

**Fusion Protein Analysis**

Recombinant fusion protein consisting of β-galactosidase with an inserted peptide was expressed in the process of the immunoscreening protocol. Following growth, inoculation by positive λgt11 phage clones, and induction of β-galactosidase expression by IPTG, fusion protein was eluted from the LB top agarose for western blot analysis. Briefly, the top agarose was scraped into a 5ml culture tube and incubated in buffer containing sucrose, SDS, Tris, sodium-EDTA, and dH2O. After overnight incubation at room temperature, the mixture was homogenized and sonicated. The samples were then centrifuged at 10,000 rpm to remove the agarose and cellular debris. The protein laden supernatant (lysate) was removed and evaluated by spectrophotometry (O.D. 260) and western blot analysis (Towbin, 1979).
Western Blot

Lysate samples containing recombinant fusion protein were obtained and run on an SDS-PAGE gel (10% acrylamide). The samples were electrophoretically transferred to a nitrocellulose membrane in a BIO-RAD Trans-Blot apparatus at 60V for 3 hours. Following transfer, the membrane was incubated in a blocking solution of 5% powdered skim milk and 0.1% Tween-20 in Tris-buffered saline TBS-T) for 30 minutes at room temperature. The membrane was probed with a cocktail of HSA-5 and HSA-6 (ascites fluid) diluted to a concentration of 2 μl/ml in 5% milk-TBST overnight at room temperature with constant agitation. The mouse IgM monoclonal antibody HBS-9505 was used in a parallel experiment as a negative control. After washing the membrane three times in TBST, the antibodies were detected by incubation in Goat Anti-mouse IgG + IgA + IgM conjugated with horseradish peroxidase, 1 μl/ml in blocking buffer for 1.5 hours at room temperature with constant agitation. Color development was achieved by incubating the membrane in a horseradish peroxidase substrate solution containing 3,3'-diaminobenzidine.

DNA Sequence Analysis

The cDNA sequence and corresponding predicted peptide sequence of clone #5 were sent to the Basic local alignment search tool (BLAST) for a search of public sequence databases. Databases offered both nucleotide and peptide sequences for comparison.
Results

Purification of HSAg-5/6 from Human Semen Extract

HSAg-5/6 was purified from crude human semen extract by immunoaffinity chromatography. Protein fractions were eluted with 0.1M glycine HCl (pH=2.2) or a 0.1M acetic acid and were examined by SDS-PAGE (Laemelli, 1971)(Fig. 3). Three major protein bands of molecular weights 11kD, 19kD, and 60kD were observed, respectively. The results of this analysis are presented in Fig. 3.
Fig. 3 An SDS-PAGE (Laemmli, 1971) gel (12% acrylamide) of HSAg-5/6 purified on an immunoaffinity chromatography column. The gel was stained with a silver stain kit purchased from BioRad. Note the bands appearing at 60kD, 19kD and 12kD.

An immunoaffinity column (10ml) was prepared to purify HSAg-5/6 by Dr. C.Y.G. Lee, using the monoclonal antibodies HSA-5 and HSA-6. For the experiment shown, I washed the column twice with 20ml of 0.1M glycine-HCl (pH=2.2) and equilibrated it with 40ml PBS. Approximately 500ml of crude human semen extract, containing an average of 37mg/ml protein, was passed through the column. After rinsing the column thoroughly with 300 ml of PBS, the bound proteins were eluted with glycine-HCl (pH=2.2). Upon neutralizing the pH with 1M Tris-Cl (pH=8.0) the fractions were quantitated spectrophotometrically at absorbance of 280nm (O.D.280).

Samples containing the highest concentration of protein were pooled and lyophilized. The fractions were resuspended in 10mM NaOH to a concentration of 0.4 µg/µl. Approximately 8µg of protein was run in each lane of the gel.

Samples of purified HSA-5 and HSA-6 IgM immunoglobulin was donated by Dr. C.Y.G. Lee, and run in the gel to compare the purity of antigen and antibody. Approximately 100µg of IgM was run in each lane.

Lane 1     HSAg-5/6 Fractions #4 and #5
Lane 2     HSAg-5/6 Fractions #6 and #7
Lane 3     HSA-5 purified IgM Immunoglobulin
Lane 4     HSA-6 purified IgM Immunoglobulin
Lane 5     Standard Molecular Weight Marker,
Top to Bottom: 106kD
                        80kD
                        49.5kD
                        32.5kD
                        27.5kD
                        18.5kD
Peptide Sequencing

The lower molecular weight protein bands of purified HSAg-5/6, observed by SDS-PAGE were originally thought to be the sperm antigen which specifically binds to HSA-5 and HSA-6 monoclonal antibodies. As a result, these bands were transferred to a PVDF membrane and were subjected to peptide sequencing procedures. Despite several attempts, the Edman degradation sequencing revealed only seven amino acids on the N-terminal region of the 18kD protein.

\[
\begin{array}{ccccccc}
K & Q & E & G & R & D & D \\
\hline
\text{Lys Gln Glu Gly Arg Asp Asp}
\end{array}
\]

The short sequence was analyzed using a local database in the Biomedical Research Center. It was determined that the proteins are degradation products of a precursor to Semenogelin 1 and Semenogelin 2 (Lilja et al, 1987; Lilja et al, 1989).

Western Blot analysis of the 60kD protein was carried out by Dr. T. Yoshiki in Dr. Lee’s lab. His data confirmed the true size of the HSAg-5/6 protein.

Immunoscreening of a Mouse Testis cDNA Library

The cDNA segment which codes for expression of the antigen recognized by HSA-5 and HSA-6 has been cloned from a mouse testis cDNA library. The cDNA library was constructed in a λgt11 bacteriophage expression vector. Varying
quantities of bacteriophage were inoculated into *E. Coli* strain Y1090⁺, and expression of the β-galactosidase fusion protein was induced. The expressed fusion protein was blotted onto two identical nitrocellulose membranes. The membranes were subjected to an immunoassay using an HSA-5/6 monoclonal antibody cocktail (HSA-5/HSA-6) or rabbit antisera as primary probes for detecting the antigenic region of the β-galactosidase fusion protein. Sixteen independently derived clones have been isolated by this method.

Of the sixteen clones, eight were selected at random for further analysis of their cDNA inserts and characterization of the expressed fusion protein. The authenticity of these clones was verified by immunoassay using antibody probes which had been pre-adsorbed with *E. Coli* lysate according to published methods (Sambrook et al, 1989). Seven of the eight clones demonstrated 100% binding to both monoclonal and polyclonal antibody probes under these conditions. This procedure provided conclusive evidence that the clones express fusion protein with antigenic epitopes not otherwise found in *E. Coli*. Fig. 4 shows clone #5 as an example.
Fig. 4 Nitrocellulose membranes displaying plaques of E.Coli lysate containing β-galactosidase fusion protein which have positively bound to polyclonal HSA-5/6 antibody and to monoclonal HSA-5/HSA-6 antibody cocktail. Detection of antibody binding was achieved by incubation in goat anti-rabbit or goat anti-mouse IgG+IgA+IgM conjugated to horseradish peroxidase. Color development resulted from peroxidase substrate binding in the presence of 3,3'-diaminobenzidine.
Fig. 4

mHSA-5
mHSA-6

pHSA-5/6
Characterization of Isolated cDNA clones

**Fusion Protein Analysis**

The fusion protein generated during immunoscreening was examined by Western blot analysis to determine the molecular size. E.Coli were induced to express fusion protein with IPTG, and were lysed. The lysate containing fusion protein was run on an SDS gel and transferred to PVDF membrane which was probed with a cocktail of monoclonal HSA-5 and HSA-6 antibodies. The antibodies demonstrated specific binding to the fusion protein, a protein with molecular weight of approximately 115kD - 120kD (Fig. 5).
Fig. 5 Western blot analysis of β-galactosidase fusion protein.

Lane A: E. Coli lysate containing clone #5 fusion protein probed with HSA-5/HSA-6 antibody cocktail
Lane B: Negative control blot probed with HBS-9505 monoclonal antibody against hepatitis B virus.

The estimated molecular weight of the fusion protein is 115kD - 120kD.
DNA Sequencing

The cDNA insert segment for Clone #5 was amplified by PCR using sequencing primers designed for use with λgt11 vector. The resulting 900bp cDNA fragment was then submitted to the DNA sequencing facility operated by the Canadian Genetic Diseases Network. "Dye-Deoxy" cycle sequencing method provided a nucleotide sequence 860 base pairs in length (Fig. 6).
Fig. 6 Nucleotide sequence of Clone #5 and deduced amino acid sequence.
Note the RNA degradation site and poly-A tail signal sequence as denoted by ** and ***, respectively.
Fig. 6

**HSAg-5/6 Cloned Sequence #5**

**Deduced Amino Acid Sequence**

<table>
<thead>
<tr>
<th>Deduced Amino Acid Sequence</th>
<th>Deduced Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGAATTCCGQMTTCCGGAATTTCAGT</td>
<td>GGAATTCCGQMTTCCGGAATTTCAGT</td>
</tr>
<tr>
<td>Glu Phe Asn Ser Gly Ile Pro Phe Cys Thr Ile Lys Phe Asn His Asn Tyr Lys Asn Val Val Leu Gly Phe Thr Ile Asn</td>
<td>Glu Phe Asn Ser Gly Ile Pro Phe Cys Thr Ile Lys Phe Asn His Asn Tyr Lys Asn Val Val Leu Gly Phe Thr Ile Asn</td>
</tr>
<tr>
<td>ATG TAC AAT CCA GCC CAA CAA GAA GGG CCT GCC CCC CAC CAT CCT GGC CCA CTG ACG AGC CCT GAG ACA CCA GGG</td>
<td>ATG TAC AAT CCA GCC CAA CAA GAA GGG CCT GCC CCC CAC CAT CCT GGC CCA CTG ACG AGC CCT GAG ACA CCA GGG</td>
</tr>
<tr>
<td>Ile Met Tyr Asn Pro Ala Gin Gin Glu Gly Pro Ala Pro His His Pro Gly Pro Leu Ser Gin Ser Pro Glu Thr Pro Gly</td>
<td>Ile Met Tyr Asn Pro Ala Gin Gin Glu Gly Pro Ala Pro His His Pro Gly Pro Leu Ser Gin Ser Pro Glu Thr Pro Gly</td>
</tr>
<tr>
<td>TGG GAG GGT GAC CCA CAG GGT CCT GCA GGA AAT GAG GCA GGG TTC CAG GCA CAG AGT TCC CAG GAC TAG</td>
<td>TGG GAG GGT GAC CCA CAG GGT CCT GCA GGA AAT GAG GCA GGG TTC CAG GCA CAG AGT TCC CAG GAC TAG</td>
</tr>
<tr>
<td>Trp Glu Gly Asp Pro Gin Gly Leu Ala Gly Thr Asp Glu Ala Gly Phe Gin Ala Gin Ser Ser Gin Asp TER</td>
<td>Trp Glu Gly Asp Pro Gin Gly Leu Ala Gly Thr Asp Glu Ala Gly Phe Gin Ala Gin Ser Ser Gin Asp TER</td>
</tr>
<tr>
<td>GGGCCAAAGG CCCCCCAGGG GGGAGAACAT GGGCACTGCC CCAGAGACC CTAACGTGGG GCAGGGTGCA CATAGGCA</td>
<td>GGGCCAAAGG CCCCCCAGGG GGGAGAACAT GGGCACTGCC CCAGAGACC CTAACGTGGG GCAGGGTGCA CATAGGCA</td>
</tr>
<tr>
<td>AAGGGGAAGT CCCCATTAA CATTGAGCAA AAGGCTATGC AGCCCCAAG AATAGCATCTA TACAGGGCTGTT ACCCAGGTT</td>
<td>AAGGGGAAGT CCCCATTAA CATTGAGCAA AAGGCTATGC AGCCCCAAG AATAGCATCTA TACAGGGCTGTT ACCCAGGTT</td>
</tr>
<tr>
<td>TCCACAGGAA ACAGCTGCTG GCAGCTACAA AACATTAA<strong>CA GC</strong>TCTTGTG TGCAAGAAGAAA AAAAATTAAGA GGGGTTTTT</td>
<td>TCCACAGGAA ACAGCTGCTG GCAGCTACAA AACATTAA<strong>CA GC</strong>TCTTGTG TGCAAGAAGAAA AAAAATTAAGA GGGGTTTTT</td>
</tr>
<tr>
<td>GGGCATGGGG CATAGGGGAG GCAGCAGATG CCTCAAAAGA CTGAGQCTCC CTTGTGCGGC TCAGTCTGCC CTAACCCAGG</td>
<td>GGGCATGGGG CATAGGGGAG GCAGCAGATG CCTCAAAAGA CTGAGQCTCC CTTGTGCGGC TCAGTCTGCC CTAACCCAGG</td>
</tr>
<tr>
<td>TGAAGGTGG GGGCAGTGGG CCCTGACTGAC CCAATAAGGCT GGGGTGCTGG GAAGCTTGGG CACCAAAGGG GAATAAAA***TAG</td>
<td>TGAAGGTGG GGGCAGTGGG CCCTGACTGAC CCAATAAGGCT GGGGTGCTGG GAAGCTTGGG CACCAAAGGG GAATAAAA***TAG</td>
</tr>
<tr>
<td>GGGTGAAGGC ACCGGTCCTG CCCCACCTCA GTTGAGGGGAG CCAAGGCAA CTTCAGTGTC GGCACAGACA GCAGATTTGG TTGTTCTCT</td>
<td>GGGTGAAGGC ACCGGTCCTG CCCCACCTCA GTTGAGGGGAG CCAAGGCAA CTTCAGTGTC GGCACAGACA GCAGATTTGG TTGTTCTCT</td>
</tr>
<tr>
<td>AGTGGGAACGT TGTAGTCGTA GGTGATCTCC TCATCCACGC CGATGGGCTG CTTCAGAGTAG ATGACAATCT TTTGTGAGA</td>
<td>AGTGGGAACGT TGTAGTCGTA GGTGATCTCC TCATCCACGC CGATGGGCTG CTTCAGAGTAG ATGACAATCT TTTGTGAGA</td>
</tr>
<tr>
<td>CTCGGAAATTC CAGCTAGGCG CGGGTGGCTA CGTTAAGGAG TTGGTTCTGTGTT</td>
<td>CTCGGAAATTC CAGCTAGGCG CGGGTGGCTA CGTTAAGGAG TTGGTTCTGTGTT</td>
</tr>
</tbody>
</table>

**RNA Degradation Site**

**Poly-A Tail Signal Sequence**
**Deduced Amino Acid Sequence.**

Each of the clones was screened with monoclonal antibodies and polyclonal antisera, thus an expressed fusion protein was instrumental in locating the desired clones. Translation of the fusion protein from mRNA included codons of the cloned insert directly following the translated codons of the β-galactosidase in precise order. Logically, the amino acid sequence of the cloned insert was deduced by knowing the amino acid sequence of β-galactosidase and the cDNA sequence of the insert.

Clone #5 is 76 amino acids in length. It represents the C-terminal end of the HSAg-5/6 protein and includes a termination codon, an RNA degradation site and a poly-A tail signal sequence (see Fig. 6).

**Sequence Homology To Known Proteins.**

The cloned cDNA insert sequence and resulting amino acid sequence were subjected to comparison with known sequences listed in public databases. Computer analysis of both sequences revealed that HSAg-5/6 has no significant homology to any other known protein listed in public databases.

The results of the computer database search for the amino acid sequence of the cloned insert is shown in Appendix A.

**Deduced Molecular Size of Fusion Protein.**

When λgt11 phage containing the cDNA insert for HSAg-5/6 is inoculated into E.Coli and induced by IPTG, a fusion protein is expressed. This protein was detected
by both monoclonal antibodies and polyclonal antisera in the immunoscreening procedure.

The fusion protein consists of β-galactosidase with an inserted amino acid segment, in this case a fragment of the HSAg-5/6 protein. In its native state, β-galactosidase is a large protein with a molecular size of 112kD. The insert site has been specifically engineered within the DNA to be the six base sequence recognized by the EcoR1 restriction enzyme, located 1.7kb from the termination codon.

The deduced peptide sequence of this insert (clone #5) is 76 amino acids in length, and includes a termination codon, an RNA degradation site and a poly-A tail signal sequence (Fig. 5).

However, because the HSAg-5/6 insert segment includes a termination codon, approximately 1.7kD of the C-terminal region of β-galactosidase was not expressed. Rather, the HSAg-5/6 fragment insert occupied the C-terminal region of the fusion protein (as well as representing the C-terminal region of the HSAg-5/6 protein). Based upon these facts, the total fusion protein has an estimated molecular size of 119kD.

This hypothesis is represented in graphical form in Fig. 7.
Fig. 7 Diagram illustrating the β-galactosidase protein, the EcoRI insertion site for cDNA, the deduced amino acid sequence of clone #5, and the predicted size of the resulting fusion protein. The N-terminus is on the right side of the diagram, the C-terminus is on the left. The fusion protein was estimated to be 119kD based upon the deduced amino acid sequence and the published information on β-galactosidase.
Fig. 7

Cloned Sequence #5
76 Amino Acids
Includes termination
codon

Fusion Protein
Estimated MW 119 kD
*Predicted Antigenic Regions of Peptide Insert.*

Additionally, the cloned amino acid sequence was analyzed for predicted hydrophilic segments. As certain amino acid residues are hydrophilic and others are hydrophobic, segments of the total amino acid sequence can be evaluated for their ionic properties. Those regions which are hydrophilic would probably be exposed to solution, and would constitute possible antigenic regions where antibodies such as HSA-5 and HSA-6 are most likely to bind. The plot in Fig. 8 shows probable antigenic regions within hydrophilic areas of clone #5 insert site.
Fig. 8 Hydrophilicity plot of the deduced amino acid sequence of Clone #5. Areas shown above 0.00 are hydrophilic and possibly antigenic for HSA-5 and HSA-6. The antigenic region is estimated to be within the final 35 amino acids. (Graph courtesy of BioYang Lin, consultant from Dr. Michael Hayden’s laboratory)
Fig. 8

Translation of PCR5 Protein Toolbox Plot

Hydrophilicity Window Size = 7
Scale = Kyte-Doolittle
Discussion

Studies of the Monoclonal Antibodies - HSA-5 and HSA-6

The monoclonal antibodies HSA-5 and HSA-6 were generated against ionophore-treated acrosome reacted human sperm in 1989. Both antibodies are IgM class and were produced in BALB/c mice as ascites fluid.

The HSA-5 and HSA-6 antibodies were submitted to the second WHO workshop for interlaboratory evaluations of their potential to be immunocontraceptive vaccine candidates and were among five considered "high priority candidates," (Anderson, 1990). The evaluations determined that the antibodies do not cross-react with somatic tissue.

Furthermore, they were shown to recognize identical bands on a Western blot of crude human sperm extract, which indicated that they recognize the same protein (HSAg-5/6). It was suggested that these two antibodies were identified from cell lines which originated from the same hybridoma. However, sandwich enzyme immunoassays (SEIA) later determined that the antibodies recognize different antigenic epitopes of HSAg-5/6 (Yoshiki et al, unpublished).

Purification of the antibodies led to conjugation with Affigel-10 and the formation of an immunoaffinity column. This immunoaffinity column was used to purify the antigen from human semen extract, whereby rabbit antisera were raised according to the standard protocol (Liu et al, 1989). Both monoclonal antibody ascites fluid and rabbit antisera were utilized to study the cognate sperm antigen.

Localization of HSAg-5/6 on Human and Mouse Sperm
Indirect immunofluorescent assays of human sperm were utilized extensively to determine the location of expression of HSAg-5/6. Assays using methanol-fixed human sperm, as well as live human and mouse sperm were performed. It was clearly established that the antigen recognized by the HSA-5 and HSA-6 monoclonal antibodies is located in the equatorial region of acrosome-reacted sperm (data not shown). Live sperm assays were observed to show little binding to fresh sperm preparations, while the binding increased significantly (>38%) following a 24-hour incubation. Positive staining increased further subsequent to incubation of the sperm with calcium ionophore A23187 to induce the acrosome reaction (data not shown).

Indirect immunofluorescent assays with epididymal mouse sperm were also performed. In this case, the monoclonal antibodies HSA-5 and HSA-6 demonstrated binding to antigens localized on the principal piece of the tail and the head of the sperm, rather than the acrosomal region as seen in human sperm (Fig. 1). A few sperms showed positive staining on both the head and tail regions.

The immunofluorescent assays demonstrated that HSA-5 and HSA-6 bind to the acrosomal region of human sperm and the head and tail of mouse sperm. To date, we have little explanation for these observation. However, O’Rand et al (1984) found that their monoclonal antibodies bound to the head region of rabbit sperm and the mid-piece of the human sperm. It has been hypothesized that HSAg-5/6 in the mouse is expressed in late spermiogenesis as an intracellular protein which is shed among the cytoplasmic droplets and later attaches to the sperm tail. Goldberg (1963) noted that LDH-X expression in sperm follows a similar pattern of migration.

This hypothesis was supported by indirect immunofluorescent assays on frozen sections of mouse tissues and developing mouse testes. Using procedures developed
by Lee and Wong (1986), Dr. T. Yoshiki (unpublished) observed that the antigen recognized by HSA-5 and HSA-6 is expressed only in testicular and epididymal sperm of the adult mouse, but not in the immature testes, nor in somatic tissues. Furthermore, positive staining of the sperm occurred on the head and tail regions in the testis, but primarily on the tail region of epididymal sperm (unpublished results).

Transmission electron microscopy was performed by Homyk et al (1992) to precisely localize the HSAg-5/6 protein on human sperm. The epitopes for these antibodies were adversely affected by fixative chemicals. Therefore, the sperm were permeabilized and incubated in HSA-5 and HSA-6 monoclonal antibodies prior to embedding. In separate experiments, HSA-5 and HSA-6 were bound to human sperm and detected with a secondary antibody conjugated to gold particles. After counterstaining with a silver enhancer to monitor the bound gold particles, the sperm were observed under and electron microscope. The results showed that the HSAg-5/6 protein is located on both the inner and outer acrosomal membranes.

Purification and Studies of HSAg-5/6

HSA-5 and HSA-6 were conjugated with Affigel-10 to form an immunoaffinity column according to the protocol provided by BioRad Lab and loaded into a modified syringe. This immunoaffinity column was used to purify the HSAg-5/6 from human semen extract. The resulting protein fractions were examined by SDS-PAGE (Fig. 3). Three major protein bands of molecular weights 11kD, 19kD, and 60kD were observed.
The lower molecular weight protein bands of purified HSAg-5/6, observed by SDS-PAGE were originally thought to be the sperm antigen which specifically binds to HSA-5 and HSA-5 monoclonal antibodies. Consequently, these bands were transferred to a PVDF membrane and were subjected to Edman degradation sequencing at a service laboratory in the University of Victoria. Unfortunately, only seven amino acids on the N-terminal region of the protein were elucidated (KQEGRD). The short sequence was analyzed and it was determined that the proteins are degradation products of a precursor to Semenogelin 1 and Semenogelin 2. These proteins are secreted by the seminal vesicle and are structural contributors to the formation of a gelatinous matrix within human semen (Lilja et al. 1989). Prostate specific antigen (PSA) is an abundant prostatic protease also found in human semen. This protease has been shown to progressively fragment Semenogelin proteins during and after liquefaction of semen, following ejaculation (Lilja et al., 1987; Lee et al., 1989). Fragments of semenogelin proteins including antigenic epitopes have been localized to parts of the spermatozoa associated with locomotion. It is thought that these fragments may participate in activation of progressive sperm movement as the gelatinous matrix is disassembled by PSA following ejaculation.

We believe that semenogelin protein fragments in the preparation of the crude semen extract were co-purified along with HSAg-5/6 in the immunoaffinity isolation process. It is possible that semenogelin proteins bind to the Fc regions of antibodies in a manner similar to those described by Kamada (Kamada et al., 1991). They hypothesized that Fc binding proteins which coat spermatozoa, provide protection for the sperm from immuno-attack by the host. However, the 16kD-20kD proteins
described by Kamada bound only to IgG class antibodies, not to IgM class antibodies such as HSA-5 and HSA-6.

**Western Blot analysis of HSAg-5/6**

Western blot analysis was performed by Dr. T. Yoshiki with purified HSAg-5/6 and compared to similar Western blot analyses of crude sperm extract from human and mouse. The analyses revealed that only the 60kD band of purified HSAg-5/6 is immunoreactive with the HSA-5 and HSA-6 monoclonal antibodies (unpublished data).

The analyses which involved crude human and mouse sperm extracts demonstrated a range of immunoreactive proteins from 40kD to 60kD (unpublished data). When compared to the purified antigen, with a molecular weight of 60kD, he hypothesized that the 60kD band on the western blot of crude extract(s) is HSAg-5/6, while the other protein bands are degradation products. The presence of proteases in semen (Lee et al., 1989) suggests the possibility that HSAg-5/6 is digested into fragments prior to enzyme neutralization during the extract preparation. Dr. C.Y.G. Lee further speculated that the acidic conditions under which HSAg-5/6 is affinity purified eliminate the degradation products from the purified fractions; only the intact form of HSAg-5/6 binds to the immunoaffinity column.

**Cloning and Molecular Biology of HSAg-5/6**

Having verified the size and location of HSAg-5/6 on human and mouse sperm, it became important to study the protein on the molecular level. These procedures reveal the molecular nature, sequence, and structure of a protein.
In collaboration with our lab, Homyk and Herr (unpublished) have cloned HSAg-5/6 from a human testis cDNA library. They have obtained a 600bp sequence of the cDNA, which represents the C-terminal region of the protein. As a result of their limited success, we set out to perform a similar cloning study in a mouse testis cDNA library. It was thought that once the cDNA sequence of the mouse gene is elucidated, then the DNA fragment could be used to isolate a full length human cDNA clone.

**Immunoscreening**

A cDNA library cloned into λgt11 phage vectors (provided by Dr. C. Lau of the University of California, San Francisco) was subjected to screening for the HSAg-5/6 epitopes. The library contained approximately $1 \times 10^6$ independent cDNA inserts. Following adsorption of the phage in E.Coli, expression of a fusion protein consisting of β-galactosidase with a peptide insert was induced. The cells were then lysed and the proteins were blotted to nitrocellulose where they were accessible to binding by the HSA-5 and HSA-6 monoclonal antibodies, or polyclonal rabbit antisera.

Using these antibodies to detect expressed fusion protein, sixteen independent clones were isolated in the immunoscreening procedure. Eight of those clones were selected at random and subjected to further examination to determine the nature of the cDNA inserts.

**cDNA Sequence Analysis**

Sequencing of the cDNA insert demonstrated that the clones isolated in immunoscreening represented the C-terminus of the HSAg-5/6 protein plus part of the
untranslated region (UTR) of messenger RNA (mRNA). The DNA sequence was 860 base pairs in length. Included in the sequence was a termination codon - T A G, which was 235 bases from the origin (Fig 3).

Among the bases of the 3' UTR was a segment which may represent an RNA degradation site - A T T T A (Fig. 3) (Liu et al, 1992). Shaw and Kamen (1986) found that insertion of nucleotide AT sequences into the 3’ UTR of a human lymphokine gene caused the otherwise stable mRNA to become highly unstable in vivo. The sequence TTATTTTAT was identified (Caput et al, 1986) as being a conserved region in the 3’ UTR of mouse and human Tumor Necrosis Factor genes which may play a role in regulating expression of mRNA. The five base pair segment in the HSag-5/6 cDNA could play a regulatory role in the expression of HSag-5/6, and the stability of the mRNA, in the mouse testis during spermatogenesis.

Also included in the 3’ UTR of the HSag-5/6 cDNA insert was a signal for cleavage of the raw RNA transcript and addition of a poly-A tail. The sequence A A T A A A (Fig. 3) signals Poly-A Polymerase to add 100-200 adenosine residues to the 3’ end of the RNA chain. This is one of the first modifications which an RNA transcript undergoes in the nucleus of a cell (Alberts et al, 1989).

The cDNA sequence in its entirety and the deduced amino acid sequence were subjected to comparison of known sequences stored in the public databases accessed by the Basic Local Alignment Search Tool (BLAST) at the National Institutes of Health, Maryland (Altschul et al, 1990). There were no significant homologies found to any known DNA or peptide sequence found in the public databases. The results of the BLAST search for the deduced peptide sequence are shown in Appendix A.
The Deduced Amino Acid Sequence and Fusion Protein Analysis

The amino acid sequence of the cloned insert for HSag-5/6 was deduced based upon the fact that it was expressed as a peptide insert within the β-galactosidase protein in the λgt11 phage. Translation of the fusion protein from mRNA included codons of the cloned insert directly following the translated codons of the β-galactosidase in precise order. Therefore, the amino acid sequence of the cloned insert was deduced by examining the flanking DNA sequences of β-galactosidase (Glover, 1985) and the cDNA sequence of the insert.

The clone #5 is 76 amino acids in length. It represents the C-terminal end of the HSag-5/6 protein and includes a termination codon, an RNA degradation site and a poly-A tail signal sequence.

Human and mouse native HSag-5/6 was determined by Western blot to have a molecular weight of 60kD. Such a protein would contain roughly six hundred amino acids, and would be coded by an mRNA of approximately three kilobases in length. Using this reasoning, we estimate that the cDNA clone of HSag-5/6, which was isolated by immunoscreening represents 76 out of 600 amino acids, or 13% of the total protein.

As stated, the isolated clone is seventy six amino acids in length and includes a termination codon. It has a molecular weight of approximately 7.6kD. When this quantity is added to the known molecular weight of β-galactosidase (Glover, 1985), taking into account the termination codon in the insert and accordingly eliminating the tail segment of the carrier protein, the molecular weight of the fusion protein is estimated at 119kD (Fig. 7). This figure corresponds well to the molecular weight of the fusion protein demonstrated by Western blot in Fig. 5 at 115kD - 120kD.
**Structural Analysis of HSAg-5/6**

The deduced amino acid sequence of the cDNA insert was subjected to a hydrophilicity analysis. This process determined regions of the peptide which were likely exposed to antibody reaction, versus those regions which were hydrophobic and probably hidden from solution.

The hydrophilicity plot in Fig. 8, demonstrates that there are several areas of the C-terminal region of HSAg-5/6 which could be exposed to immune reaction in vivo. These areas elicited an immune response in a vaccinated host mouse, whereby the monoclonal antibodies HSA-5 and HSA-6 were originally raised. The hydrophilicity analysis suggests that the C-terminus of the HSAg-5/6 protein is exposed to the exterior environment in mouse sperm and similarly uncovered subsequent to the acrosome reaction in human sperm. This hypothesis is supported by the findings of Homyk (unpublished) who cloned the C-terminal region of human HSAg-5/6 from a human testis cDNA library by immunoscreening.

The clones which were isolated by immunoscreening were shown to be identical to each other (Fig. 10 and Fig. 12). The use of antibodies, particularly monoclonal antibodies, to screen fusion protein expression in the cloning process explains why the clones are identical. In order for the antibodies to hybridize to the fusion protein, the antigenic epitopes of HSAg-5/6 were required to be intact. This requirement greatly restricted the screening process to a limited number of cDNA inserts. This forced the clone selection to identical, or nearly-identical features.

**Future Studies for HSAg-5/6**
The antigen recognized by the monoclonal antibodies HSA-5 and HSA-6 requires further study to determine its true potential as an antifertility vaccine candidate.

The exact epitopes which bind the HSA-5 and HSA-6 monoclonal antibodies need to be elucidated. The hydrophilicity plot in Fig. 8 demonstrates which sequences of the peptide are likely epitopes. Short peptides based upon epitope sequences can be synthesized in an automated peptide synthesizing apparatus. These peptides could be tested for their antigenicity to the monoclonal antibodies by ELISA. Once the exact epitope is known, it alone could be a candidate for immunocontraceptive vaccine. The advantage of using a short peptide antigen as a vaccine (rather than an entire protein) lies in a lesser chance of other protein epitopes being common to host tissues.

Northern Blot assay of the mRNA from mouse testis must be achieved. Using the 860bp cDNA insert from HSAg-5/6 as a probe to visualize the native mRNA from mouse testis will establish the size of the full length cDNA. This assay could confirm or disconfirm the current belief that the native HSAg-5/6 protein has a molecular weight of 60kDa.

The full length cDNA needs to be isolated from the mouse testis cDNA library. Again, using the 860bp cDNA as a probe, the λgt11 phage library could be screened for complementary cDNA inserts. If the screening does not rely on an intact protein epitope, then the likelihood of encountering a larger segment of the cDNA insert increases substantially.

Once the full length cDNA insert is isolated, then the expressed fusion protein could be isolated. An immunoaffinity chromatography procedure involving either anti-λ-galactosidase antibodies or the HSA-5 and HSA-6 monoclonal antibodies could
be utilized to isolate the fusion protein from a crude E.Coli/phage lysate. Isolating the entire fusion protein lends the advantage of raising antisera to further confirm the validity of the cDNA insert. Once the fusion protein is isolated, it can be used to vaccinate an animal host, which will result in a protein specific antisera. The antisera will be subjected to ELISA, indirect immunofluorescent assays of human and/or mouse sperm and Western blot assays to confirm that it recognizes the same protein as the original HSA-5 and HSA-6 monoclonal antibodies.

The isolated fusion protein can also be subjected to partial peptide sequencing to confirm its homology to the original HSAG-5/6 protein.

Conclusions

Understanding the nature of molecules which have been chosen for antifertility vaccine candidates is essential. Subsequently, the outcome of this project is to characterize the cognate antigen recognized by the monoclonal antibodies HSA-5 and HSA-6, which are considered "high priority candidates" for immunocontraceptive vaccine.

From the results of my work on the molecular nature of HSAG-5/6, the following conclusions can be drawn:

1. Analysis of the isolated cDNA clones and expressed fusion protein suggest that the antigenic epitopes recognized by the monoclonal antibodies HSA-5 and HSA-6 are localized in the C-terminal region of the HSAG-5/6 protein.
2. Sequence analysis of the cDNA insert (including the 3' UTR), as well as the deduced peptide sequence demonstrated that the HSAg-5/6 protein has no significant homology to any other known protein or gene (Appendix A).
Bibliography


Anderson DJ (1992) WHO sperm antigen workshop, Task Force on Vaccines for Fertility Regulation

Baskin MJ (1932) Temporary sterilisation by the injection of human spermatozoa.


58


Yoshiki T, McChesney P, Homyk M, Herr JC, Lee CYG, Characterization of a sperm antigen recognized by two monoclonal antibodies, UNPUBLISHED

Appendix A

Computer search of peptide sequence databases. The deduced peptide sequence in Fig. 6 was the subject of the query.
Trying muncher... connected
National Center for Biotechnology Information (NCBI)

Experimental GENINFO(R) BLAST Network Service (Muncher)


To Obtain Documentation: send an e-mail message to 'blast@ncbi.nlm.nih.gov' with the word HELP in the body of the message. The documentation was last modified February 24th.

---

Tomorrow, April 14th 1994, for computer maintenance reasons, the BLAST servers will be shut down sequentially.

Nevertheless, we should be able to manage the regular number of requests but the service should be slower than usual.

We apologize for any inconvenience.

---

PEPTIDE SEQUENCE DATABASES
nr Non-redundant PDB+SwissProt+PIR+SPUpdate+GenPept+GUpdate, updated daily
    for efficient, complete searches of the five component databases:
pdb  Brookhaven Protein Data Bank, October 1993 Release
swissprot SWISS-PROT Release 28.0, March 1994
pir   PIR Release 39.0 (complete), December 31, 1993
spupdate SWISS-PROT cumulative weekly update to the major release
genpept CDS translations from GenBank(R) Release 82.0, April 9, 1994
gupdate cumulative daily updates to the major release of genpept
kabatpro Kabat Sequences of Proteins of Immunological Interest Release 5.0, August 1992	
tfd  TFD transcription factor (protein) database Release 7.0, June 1993
acr * Ancient Conserved Region subset of SWISS-PROT, Dec. 3, 1993
palu * six-frame translations of representative human Alu repeats

NUCLEOTIDE SEQUENCE DATABASES
nr Non-redundant PDB+GBJUpdate+GenBank+EmblUpdate+EMBL, updated daily
    for efficient, complete searches of the four component databases:
pdb  Brookhaven Protein Data Bank, October 1993 Release
genbank GenBank(R) Release 82.0 (no daily updates), April 9, 1994
gbupdate GenBank(R) cumulative daily updates to the major release
embl  EMBL Data Library, Release 38.0, April 1994
emblu EMBL Data Library cumulative daily updates to the major release
vector Vector subset of GenBank(R), LANL, April 23, 1992
rebase * Human and other primate Alu repeats, Dr. Jerzy Jurka, Sept. 1993
kabatnc Kabat Sequences of Nucleic Acid of Immunological Interest
Release 5.0, August 1992
epd Eukaryotic Promoter Database Release 35, June 1993
dbest ** Database of Expressed Sequence Tags Release 2.5, April 1st, 1994

* Databases that are not accessible through the NCBI Retrieve E-mail server.
** dbEST data are available from est_report@ncbi.nlm.nih.gov. Send a HELP
message to obtain instructions.

For a free subscription to "NCBI News", the NCBI newsletter, send a request
along with your name and postal mailing address to: info@ncbi.nlm.nih.gov

GenBank(R) Release 82.0 is available via anonymous ftp on ncbi.nlm.nih.gov

All direct submissions of sequences to the GenBank(R) database, including
those composed with AuthorIn, should be sent to the NCBI at any of the
following addresses. If data is submitted on diskette, please indicate
whether Mac or PC.

Postal mail: GenBank Submissions
National Center for Biotechnology Information
Building 38A, Room 8N-603
8600 Rockville Pike
Bethesda, MD 20894-0001
Voice: 301-496-2475

E-mail submissions of new sequences: gbsub@ncbi.nlm.nih.gov
E-mail submissions of updates: update@ncbi.nlm.nih.gov

GenBank is a registered trademark of the National Institutes of Health.

The help document for the BLAST E-mail server was last modified Feb. 24th.

BLASTP 1.3.11MP [29-Oct-93] [Build 14:35:03 Mar 3 1994]

Reference: Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers,

Query= hsag-5%6 frame +2
(78 letters)

Database: Non-redundant PDB+SwissProt+SPupdate+PIR+GenPept+GPeptupdate, 5:04 AM
EDT Apr 13, 1994
114,708 sequences; 32,528,034 total letters.
Searching............................................done

Highest-scoring Hit Extension vs. Count of Database Sequences

Histogram units: = 130 Sequences : less than 130 sequences

Highest
Score
1 Count-->
V
0 2540 | = = = = = = = = = = = = = = = = = = = = = = = = = =
Neighborhood word score threshold, T = 11
| 11 | 885 | ======= |
| 12 | 1200 | ======= |
| 13 | 1545 | ======= |
| 14 | 1702 | ======= |
| 15 | 2354 | ======= |
| 16 | 2855 | ======= |
| 17 | 3620 | ======= |
| 18 | 4852 | ======= |
| 19 | 6016 | ======= |
| 20 | 6544 | ======= |
| 21 | 7532 | ======= |
| 22 | 8139 | ======= |
| 23 | 7794 | ======= |
| 24 | 8947 | ======= |
| 25 | 7832 | ======= |
| 26 | 6909 | ======= |
| 27 | 6138 | ======= |
| 28 | 5152 | ======= |
| 29 | 4399 | ======= |
| 30 | 3397 | ======= |
| 31 | 2979 | ======= |
| 32 | 2321 | ======= |
| 33 | 1717 | ======= |
| 34 | 1570 | ======= |
| 35 | 1196 | ======= |
| 36 | 904  | ======= |
| 37 | 787  | ======= |
| 38 | 535  | ======= |

============= Expect = 8.2e+02, Observed = 2037 ==========

| 39 | 433 | == |
| 40 | 318 | == |
| 41 | 227 | = |
| 42 | 205 | = |
| 43 | 152 | = |
| 44 | 130 | = |
| 45 | 132 | = |

============= Expect = 77., Observed = 440 ==========

| 46 | 119 | : |
| 47 | 79  | : |
| 48 | 48  | : |
| 49 | 29  | : |
| 50 | 19  | : |
| 51 | 54  | : |
| 52 | 10  | : |

============= Expect = 7.3, Observed = 82 ==========

| 53 | 14  | : |
| 54 | 15  | : |

<<<<<<<<<<<<<< Expect = 3.7, Observed = 53 <<<<<<<<< Cutoff Score 55

| 55 | 20  | : |
| 56 | 5   | : |
| 57 | 3   | : |
| 58 | 3   | : |

============= Expect = 0.94, Observed = 22 ==========

| 59 | 10  | : |
| 60 | 5   | : |
| 61 | 4   | : |
| 62 | 2   | : |
| 63 | 1   | : |
| >63 | 0 | : |
**Smallest Poison**

**High Probability**

Sequences producing High-scoring Segment Pairs:

<table>
<thead>
<tr>
<th>Score</th>
<th>P(N)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>0.00095</td>
<td>2</td>
</tr>
<tr>
<td>55</td>
<td>0.0012</td>
<td>2</td>
</tr>
<tr>
<td>51</td>
<td>0.0015</td>
<td>3</td>
</tr>
<tr>
<td>49</td>
<td>0.0016</td>
<td>3</td>
</tr>
<tr>
<td>50</td>
<td>0.0029</td>
<td>2</td>
</tr>
<tr>
<td>52</td>
<td>0.0032</td>
<td>2</td>
</tr>
<tr>
<td>42</td>
<td>0.0060</td>
<td>3</td>
</tr>
<tr>
<td>42</td>
<td>0.0060</td>
<td>3</td>
</tr>
<tr>
<td>61</td>
<td>0.0062</td>
<td>2</td>
</tr>
<tr>
<td>45</td>
<td>0.019</td>
<td>2</td>
</tr>
<tr>
<td>44</td>
<td>0.019</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>0.031</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>0.031</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>0.031</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>0.031</td>
<td>3</td>
</tr>
<tr>
<td>52</td>
<td>0.031</td>
<td>3</td>
</tr>
<tr>
<td>51</td>
<td>0.031</td>
<td>3</td>
</tr>
<tr>
<td>45</td>
<td>0.037</td>
<td>2</td>
</tr>
<tr>
<td>48</td>
<td>0.040</td>
<td>2</td>
</tr>
<tr>
<td>47</td>
<td>0.046</td>
<td>4</td>
</tr>
<tr>
<td>53</td>
<td>0.066</td>
<td>2</td>
</tr>
<tr>
<td>55</td>
<td>0.069</td>
<td>2</td>
</tr>
<tr>
<td>48</td>
<td>0.078</td>
<td>2</td>
</tr>
<tr>
<td>51</td>
<td>0.080</td>
<td>3</td>
</tr>
<tr>
<td>45</td>
<td>0.089</td>
<td>2</td>
</tr>
<tr>
<td>42</td>
<td>0.090</td>
<td>3</td>
</tr>
<tr>
<td>41</td>
<td>0.10</td>
<td>2</td>
</tr>
<tr>
<td>48</td>
<td>0.15</td>
<td>3</td>
</tr>
<tr>
<td>55</td>
<td>0.19</td>
<td>2</td>
</tr>
<tr>
<td>45</td>
<td>0.22</td>
<td>2</td>
</tr>
<tr>
<td>38</td>
<td>0.22</td>
<td>2</td>
</tr>
<tr>
<td>44</td>
<td>0.23</td>
<td>2</td>
</tr>
<tr>
<td>48</td>
<td>0.25</td>
<td>3</td>
</tr>
<tr>
<td>55</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>47</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>47</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>47</td>
<td>0.26</td>
<td>2</td>
</tr>
<tr>
<td>47</td>
<td>0.26</td>
<td>2</td>
</tr>
<tr>
<td>59</td>
<td>0.26</td>
<td>2</td>
</tr>
<tr>
<td>52</td>
<td>0.27</td>
<td>2</td>
</tr>
<tr>
<td>50</td>
<td>0.29</td>
<td>3</td>
</tr>
<tr>
<td>54</td>
<td>0.32</td>
<td>2</td>
</tr>
<tr>
<td>42</td>
<td>0.33</td>
<td>3</td>
</tr>
<tr>
<td>49</td>
<td>0.33</td>
<td>3</td>
</tr>
<tr>
<td>44</td>
<td>0.33</td>
<td>2</td>
</tr>
<tr>
<td>46</td>
<td>0.35</td>
<td>2</td>
</tr>
<tr>
<td>46</td>
<td>0.35</td>
<td>2</td>
</tr>
</tbody>
</table>
WARNING: Descriptions of 108 database sequences were not reported due to the limiting value of parameter V = 100.
Length = 367

Score = 56 (26.0 bits), Expect = 4.7, P = 0.99
Identities = 12/35 (34%), Positives = 14/35 (40%)

Query: 28 NIMYNPAQEQGPAPHHPGPLSQSPGTQPGWEGDPQG 62
        N+ P  +GP  QPG P G PQG
Sbjct: 136 NVDQGPGAPGPGAPQGPGAPQGPGAPQGPGAPQG 170

Score = 52 (24.2 bits), Expect = 0.00095, Poisson P(2) = 0.00095
Identities = 12/36 (33%), Positives = 13/36 (36%)

Query: 33 PAQQEGPAPHHPGPLSQSPGTQPGWEGDPQGLAGTDE 68
        P  +GP  QPG P G PQG  E
Sbjct: 195 PGAPQGPGAPQGPGAPQGPGAPQGPGAPQGPGAPQ 230

(pirl | S22373 | S22373 proline-rich protein - mouse >gp | X63005 | MMPRP28_1
proline-rich protein [Mus musculus] >gp | X63004 | MMPRP4628_1
proline-rich protein [Mus musculus]
Length = 260

Score = 55 (25.6 bits), Expect = 6.3, P = 1.0
Identities = 10/38 (26%), Positives = 18/38 (47%)

Query: 24 GFTINIMYNPAQEQGPAPHHPGPLSQSPGTQPGWEGDPQ 61
        G  +++ +P+  PPHGP +P  +GP +
Sbjct: 33 GHSTTVSDPSPTQPQPQKHSQPQKPRQGSTQGPR 70

Score = 51 (23.7 bits), Expect = 0.0012, Poisson P(2) = 0.0012
Identities = 11/27 (40%), Positives = 13/27 (48%)

Query: 36 QEGPAPHHPGPLSQSPGTQPGWEGDPQG 62
        Q+GP P PP S P G+ QG
Sbjct: 74 QGGPPPPGPPQGSSQQRPQPQPGNQQQG 100

Score = 36 (16.7 bits), Expect = 0.12, Poisson P(3) = 0.11
Identities = 9/27 (33%), Positives = 10/27 (37%)

Query: 39 PAPHHHPGPLSQSPGTQPGWEGDPQGLAG 65
        P P G QP PG + P G
Sbjct: 80 PGPPPQGSSQQRPQPQPGNQQGPPPGQG 106

(pirl | S16907 | S16907 collagen alpha 1(IV) chain - bovine (fragments)
Length = 350

Score = 51 (23.7 bits), Expect = 26., P = 1.0
Identities = 10/21 (42%), Positives = 10/21 (47%)

Query: 52 GTPGWEGDPQGLAGTDEAGFQ 72
        GPGW GP  +GPQ
Sbjct: 251 GNPGWQTPGAPGPXEDPGFQ 271

Score = 42 (19.5 bits), Expect = 0.80, Poisson P(2) = 0.55
Identities = 9/21 (42%), Positives = 11/21 (52%)

Query: 51 PGTPGWEGDPQGLAGTDEAGF 71
        PG+PG G PG+ GF
Sbjct: 137 PGSPGPXGGPGVYPGLPGF 157

Score = 41 (19.1 bits), Expect = 0.0015, Poisson P(3) = 0.0015
Identities = 10/33 (30%), Positives = 12/33 (36%)

Query: 38 GPAPHHGPGLSQSPGTGPWEGDPQGLAGTDEAG 70
     G P + G P G D + A G
Sbjct: 65 GPGPXGVAGIPGPQGIPGLPDXGAXGEXGQAG 97

>pir S20500 | S20500 hydroxyproline-rich glycoprotein - r:ce >gp | X61280 | OSHPRG_P_1
    hydroxyproline-rich glycoprotein [Oryza sativa]
    Length = 369

Score = 49 (22.8 bits), Expect = 51, P = 1.0
Identities = 8/21 (38%), Positives = 11/21 (52%)

Query: 31 YNPAQQEGPAPHPHGAPLSQSP 51
       Y P A + P P + P + P
Sbjct: 198 YKPAPKPPTTPQYPAPPTYPK 218

Score = 45 (20.9 bits), Expect = 0.11, Poisson P(2) = 0.11
Identities = 8/17 (47%), Positives = 10/17 (58%)

Query: 30 MYNPAQQEGPAPHPGP 46
       M Y P + P A P + P P
Sbjct: 145 MYKPQPKPTAYPYQPAP 161

Score = 41 (19.1 bits), Expect = 0.0016, Poisson P(3) = 0.0016
Identities = 7/13 (50%), Positives = 9/13 (69%)

Query: 41 PHHPGLSQSQTPGTP 54
       P + H P S + P G P
Sbjct: 353 PYHKPPSYTPGPP 366

>sp | P17656 | CC02_CAEEL CUTICLE COLLAGEN 2. >sp | 17656 | CC02_CAEEL CUTICLE COLLAGEN 2. >pir | B31219 | B31219 collagen 2 - Caenorhabditis elegans
>gp | V00148 | CECOL2_1 Caenorhabditis elegans gene Col-2 coding for a collagen. [Caenorhabditis elegans] >gp | J01048 | CECOL2G_1 C.elegans (nematode) collagen 2 (col-2) gene, complete cds. [Caenorhabditis elegans]
    Length = 301

Score = 50 (23.2 bits), Expect = 0.0029, Poisson P(2) = 0.0029
Identities = 11/37 (29%), Positives = 17/37 (45%)

Query: 38 GPAPHHPGGLSQSPGTGPWEGDPQGLAGTDEAGFQAQ 74
       G P P + G P G + G P + G + G
Sbjct: 242 GPGPAGPKGPAPAAGPGADGPNPGPGTAGKPGGE 278

Score = 50 (23.2 bits), Expect = 0.0029, Poisson P(2) = 0.0029
Identities = 13/38 (34%), Positives = 16/38 (42%)

Query: 33 PAQQECPAPHHPGGLSQSPGTGPWEGDPQGLAGTDEAG 70
       P Q P + P G G G + G P + G G G G G
Sbjct: 146 PCQPCPGPGPPAGPAGPGPGPPDGNGGPNPSFGSPFG 183

Score = 44 (20.5 bits), Expect = 0.17, Poisson P(2) = 0.16

72
Identities = 11/31 (35%), Positives = 13/31 (41%)

Query:  32 NPAQQEGPAPHHPGPLSQSPGTPGWEGDTPQG 62  
+PA GPP F + G G G PG  
Sbjct:  175 SPAQPSGPAPAGPAGPAGPAGNGAPGAPG 205

Score = 52 (24.2 bits), Expect = 0.0032, Poisson P(2) = 0.0032
Identities = 11/27 (40%), Positives = 12/27 (44%)

Query:  39 PAPHHPGPLSQSPGTPGWEGDTPQGLAG 65  
P  P  P  GTPG G G +G  
Sbjct:  827 PPPGPSPPPPSQNGTPGTPGETGASG 853

Score = 52 (24.2 bits), Expect = 0.0032, Poisson P(2) = 0.0032
Identities = 10/29 (34%), Positives = 16/29 (55%)

Query:  50 SPGTPGWEGDTPQGLAGTDEAGFQAQSFQD 78  
+PGTPG GPGG + G +++ D  
Sbjct:  858 TPGTPGTPGTGGMMKAYKLGLVKRTYVD 886

Score = 45 (20.9 bits), Expect = 0.37, Poisson P(2) = 0.31
Identities = 11/28 (39%), Positives = 11/28 (39%)

Query:  33 PAQQEGPAPHHPGPLSQSPGTPGWEGDTPQG 60  
P Q G  G 5GTPG G P  
Sbjct:  835 PPNSQNGTPGTPGETGASGCGTPGTPG 862

>sp|P02454|CA11_RAT COLLAGEN ALPHA 1(I) CHAIN (FRAGMENTS).  
Length = 571

Score = 42 (19.5 bits), Expect = 5.7e+02, P = 1.0
Identities = 8/22 (36%), Positives = 10/22 (45%)

Query:  39 PAPHHPGPLSQSPGTPGWEGDTPQGLAG 65  
P  P  P  GPG +GP  
Sbjct:  211 PGGPPGAPAGAAGPAGNPGADGQP 232

Score = 41 (19.1 bits), Expect = 0.0060, Poisson P(3) = 0.0060
Identities = 10/25 (40%), Positives = 11/25 (44%)

Query:  48 SQSPGTPGWEGDTPQGLAGTDEAGFQ 72  
S PG PG +GD  AGQ  
Sbjct:  271 SGEPGAPGKDGDTGAKGEPGAGVQ 295

Score = 41 (19.1 bits), Expect = 0.0060, Poisson P(3) = 0.0060
Identities = 9/26 (34%), Positives = 12/26 (46%)

Query:  33 PAQQEGPAPHHPGPLSQSPGTPGWEGDTPQG 58  
P Q + P P  L + G PG +G  
Sbjct:  79 PGQRGPQPGQARGLPGTAGLPGMKG 104

Score = 40 (18.6 bits), Expect = 0.017, Poisson P(3) = 0.015
Identities = 11/30 (36%), Positives = 12/30 (40%)
Query:  41  PHHPGPLSQPSGTGWPQDPQGLAGTDEAG 70
    PP     SPG  G+GP  EAG
Sbjct:  342  PKGPAGERGSPPAGPKGSPGEAGRPGEAG 371

Score = 39 (18.1 bits), Expect = 13., Poisson P(2) = 1.0
Identities = 9/24 (37%), Positives = 12/24 (50%)

Query:  36  QEGPAPHHPGPLSQSPGTGWPQGD 59
       QPP  P  ++PG  G+GD
Sbjct:  526  QGPPGAPAAGPRNGNGAPNGDKG 549

Score = 35 (16.3 bits), Expect = 0.020, Poisson P(4) = 0.020
Identities = 7/23 (30%), Positives = 11/23 (47%)

Query:  52  GTPGWEGDPQGLAGTDEAGFQAQ 74
       GPG+GPLL++
Sbjct:  494  GKPGZZGVPDLGAPGSGARGE 516

>pir! A90559! CGRT1s collagen alpha 1(I) chain - rat (fragments)
   length = 571

Score = 42 (19.5 bits), Expect = 5.7e+02, P = 1.0
Identities = 8/22 (36%), Positives = 10/22 (45%)

Query:  39  PAPHHPGPLSQPSGTGWPQGD 60
       PPP  +PG  +G
Sbjct:  211  PGPPGAPAAGPAGNGGPDGQGQP 232

Score = 41 (19.1 bits), Expect = 0.0060, Poisson P(3) = 0.0060
Identities = 10/25 (40%), Positives = 11/25 (44%)

Query:  48  SQSPGTGWPQDPQGLAGTDEAGFQ 72
       SPG  PG++GD  AGQ
Sbjct:  271  SGEPGAPCNKTDGAKEPGPGVQ 295

Score = 41 (19.1 bits), Expect = 0.0060, Poisson P(3) = 0.0060
Identities = 9/25 (36%), Positives = 12/25 (48%)

Query:  33  PAQQGEQPAPHHPGPLSQPSGTGWPQGD 58
       Q+PP  P+L  +GP+G
Sbjct:  79  PGQRGPQGAPGLPGTALPGMKG 104

Score = 40 (18.6 bits), Expect = 0.017, Poisson P(3) = 0.016
Identities = 11/30 (36%), Positives = 12/30 (40%)

Query:  41  PHHPGPLSQPSGTGWPQDPQGLAGTDEAG 70
       PP  SPG  G+GP  EAG
Sbjct:  342  PKGPAGERGSPPAGPKGSPGEAGRPGEAG 371

Score = 37 (17.2 bits), Expect = 48., Poisson P(2) = 1.0
Identities = 9/24 (37%), Positives = 12/24 (50%)

Query:  36  QEGPAPHHPGPLSQPSGTGWPQGD 59
       QPP  P++PG  G+GD
Sbjct:  526  QGPPGAPAAGPRNGNGAPGBGAKG 549

Score = 35 (16.3 bits), Expect = 0.020, Poisson P(4) = 0.020
Identities = 7/23 (30%), Positives = 11/23 (47%)
Query:  52 GTPGWE GPQGLAGTDE AGFQAQ 74
    GP +GP L +G ++
Sbjct:  494 GKP GZGVPDLGAPGSPARGE 516

>sp\ P34545\ YNJ1_CAEEL HYPOTHETICAL 222.4 KD PROTEIN R10E11.1 IN CHROMOSOME III.
    >sp\ P34545\ YNJ1_CAEEL HYPOTHETICAL 222.4 KD PROTEIN R10E11.1 IN
    CHROMOSOME III. >gp\ Z29095\ CER10E11.1 R10E11.1, contains charged
    regions and a copy of the bromodomain [Caenorhabditis elegans]
    Length = 2015
Score = 61 (28.4 bits), Expect = 0.93, P = 0.61
Identities = 14/42 (33%), Positives = 18/42 (42%)

Query:  29 IMYNPAQQE GPAPHPGGLS QSPG TPGWE GPQGLAGTDE AG 70
    +M Q+ P H P +Q PG +GPQG G G
Sbjct:  184 MMMGAQGQQFPGMMHRYPYAQQGPPGAQGMFPGVS RGG 225
Score = 52 (24.2 bits), Expect = 0.0062, Poisson P(2) = 0.0062
Identities = 13/43 (30%), Positives = 17/43 (39%)

Query:  26 TINIMYNPAQQE GPAPHPGGLS QSPG TPGWE GPQGLAGTDE 68
    T N +M P+ PP GP PG P G+ D+
Sbjct:  138 TPNMMSPPSMGRVPGPSPGGPPGPQMRPGQPGMFQGDQ 180
>gp\ X16711\ HSCOL2A1R_1 COL2A1 gene product [Homo sapiens]
    Length = 1150
Score = 45 (20.9 bits), Expect = 2.1e+02, P = 1.0
Identities = 13/36 (36%), Positives = 14/36 (38%)

Query:  35 QQEGPAPHPGGLS QSPG TPGWE GPQGLAGTDE AG 70
    +Q PP L PG PG G E+G
Sbjct:  586 EQGAPGPSPQFQQLPQGGPQGGPQEGGKPGDQGVPGEAG 621
Score = 44 (20.5 bits), Expect = 0.78, Poisson P(2) = 0.54
Identities = 9/26 (34%), Positives = 14/26 (53%)

Query:  50 SPGTPGWE GPQGLAGTDE AGFQAQ 75
    +PGTPG GP T +G ++
Sbjct:  997 APGTPGPSPGPA GTG KQGDRG E 1022
Score = 41 (19.1 bits), Expect = 0.019, Poisson P(3) = 0.019
Identities = 11/36 (30%), Positives = 13/36 (36%)

Query:  39 PAPHHPGGLS QSPG TPGWE GPQGLAGTDEAGFQAQ 74
    PP G G PG G P + EG F +
Sbjct:  602 PGPPGEGGKPGDQGVPGEAGAPGLVGPRGERGFPG 637
>pir\ S18804\ S18804 collagen alpha 4(IV) chain - bovine (fragment)
    Length = 453
Score = 44 (20.5 bits), Expect = 2.8e+02, P = 1.0
Identities = 9/23 (39%), Positives = 10/23 (43%)
Query:  33 PAQQEGPAPHPGGLS QSPGTPG 55
PQQPP + PGPG
Sbjct: 23 PGDQGPDDPGPRGVPQPGPGPG 45

Score = 41 (19.1 bits), Expect = 2.1, Poisson P(2) = 0.88
Identities = 8/13 (61%), Positives = 9/13 (69%)

Query: 47 LSQSPGTPGWEGD 59
  LSPG +GD
Sbjct: 185 LGSPGPPGHKG 197

Score = 39 (18.1 bits), Expect = 0.019, Poisson P(3) = 0.019
Identities = 7/11 (63%), Positives = 8/11 (72%)

Query: 50 SPGTPGWEGDP 60
  +PGPGDP
Sbjct: 206 APGPPGMGDP 216

>pir| B40333| B40333 collagen alpha 1(II) chain precursor - African clawed frog
  Length = 1486

Score = 47 (21.8 bits), Expect = 1.1e+02, P = 1.0
Identities = 10/22 (45%), Positives = 12/22 (54%)

Query: 39 PAPHHPGPLSQPGTPGWEGDP 60
  LPG +GD
Sbjct: 1021 PGPVPPGLTGPGSGPGREGN 1042

Score = 45 (20.9 bits), Expect = 0.52, Poisson P(2) = 0.40
Identities = 9/26 (34%), Positives = 14/26 (53%)

Query: 50 SPGTPGWEGDPQGLAGTDEAGFQAQS 75
  +PGPG  +T+G+  S
Sbjct: 1068 APGAPGAPGAPGSGPGREGN 1093

Score = 41 (19.1 bits), Expect = 0.062, Poisson P(3) = 0.031
Identities = 9/27 (33%), Positives = 11/27 (40%)

Query: 33 PAQQEGPAHPGHLSQSPGTPGWEGD 59
  PPGPG +G
Sbjct: 352 PGPAGPAPGVPAGAPGPGSKGE 378

>gp| M63595| XELCOL2A1_1 alpha-1 type II collagen | Xenopus laevis|
  Length = 1486

Score = 47 (21.8 bits), Expect = 1.1e+02, P = 1.0
Identities = 10/22 (45%), Positives = 12/22 (54%)

Query: 39 PAPHHPGPLSQPGTPGWEGDP 60
  LPG +GD
Sbjct: 1021 PGPVPPGLTGPGSGPGREGN 1042

Score = 45 (20.9 bits), Expect = 0.52, Poisson P(2) = 0.40
Identities = 9/26 (34%), Positives = 14/26 (53%)

Query: 50 SPGTPGWEGDFQGLAGTDEAGFQAQS 75
  +PGPG  +T+G+  S
Sbjct: 1068 APGAPGAPGAPGSGPGREGN 1093

76
Score = 41 (19.1 bits), Expect = 0.032, Poisson P(3) = 0.031
Identities = 9/27 (33%), Positives = 11/27 (40%)

Query: 39 PAQEEGPAPHPGPGPLSQSPGTPGEGD 59
       P P P P P + G +
Sbjct: 352 PGAPGAPGPGAPGPGAPGSKGE 378

Score = 45 (20.9 bits), Expect = 2.1e+02, P = 1.0
Identities = 13/36 (36%), Positives = 15/36 (41%)

Query: 35 QQEGGPAPHPGPGPLSQSPGTPGEGD 70 P + Q F P L + G G P G E A G
Sbjct: 660 EQGPPGGSFGQGLPGGSPGEGGKPGDQGVPGEAG 695

Score = 43 (20.0 bits), Expect = 2.0, Poisson P(2) = 0.86
Identities = 9/26 (34%), Positives = 14/26 (53%)

Query: 50 SPCTPGWEGDPQGLAGTDEAGFQAQS 75 + P G G P G + T + G + + S
Sbjct: 1071 APGAPGSPGAPGPGPTGKQGDRGSES 1096

Score = 41 (19.1 bits), Expect = 0.032, Poisson P(3) = 0.031
Identities = 9/27 (33%), Positives = 11/27 (40%)

Query: 39 PAQEEGPAPHPGPGPLSQSPGTPGEGD 59
       P P P P P + G +
Sbjct: 355 PGAPGAPGPGAPGPGAPGSKGE 378

>pir|A40333|A40333 collagen alpha 1(II) chain precursor - African clawed frog
Length = 1432

Score = 45 (20.9 bits), Expect = 2.1e+02, P = 1.0
Identities = 13/36 (36%), Positives = 15/36 (41%)

Query: 35 QQEGGPAPHPGPGPLSQSPGTPGEGD 70 + Q F P L + G G P G E A G
Sbjct: 660 EQGPPGGSFGQGLPGGSPGEGGKPGDQGVPGEAG 695

Score = 43 (20.0 bits), Expect = 2.0, Poisson P(2) = 0.86
Identities = 9/26 (34%), Positives = 14/26 (53%)

Query: 50 SPCTPGWEGDPQGLAGTDEAGFQAQS 75 + P G G P G + T + G + + S
Sbjct: 1071 APGAPGSPGAPGPGPTGKQGDRGSES 1096

Score = 41 (19.1 bits), Expect = 0.032, Poisson P(3) = 0.031
Identities = 9/27 (33%), Positives = 11/27 (40%)

Query: 39 PAQEEGPAPHPGPGPLSQSPGTPGEGD 59
       P P P P P + G +
Sbjct: 355 PGAPGAPGPGAPGPGAPGSKGE 378
>pir|A31427|A31427 collagen alpha 2(V) chain precursor - human (fragment)

>gp|J04478|HUMC5A2B_1 COL5A2 gene product [Homo sapiens]

Length = 463

Score = 49 (22.8 bits), Expect = 52., P = 1.0
Identities = 9/22 (40%), Positives = 10/22 (45%)

Query: 39 PAPHPGPGSQPSPGTPGEGDP 60
PP + PG PG G P
Sbjct: 152 PGPRGPGIDGEPGVPQPGAP 173

Score = 47 (21.8 bits), Expect = 0.038, Poisson P(2) = 0.037
Identities = 9/21 (42%), Positives = 11/21 (52%)

Query: 51 PGTPGWEDPQPGLAGTDEAGF 71
PG PG +G+P EGF
Sbjct: 260 PGKPGEDGEPGRNGNPGEVGF 280

>gp|M90464|HUMCOL4A5X_1 alpha-5 type IV collagen [Homo sapiens]

Length = 911

Score = 48 (22.3 bits), Expect = 0.040, Poisson P(2) = 0.040
Identities = 9/20 (45%), Positives = 12/20 (60%)

Query: 51 PGTPGWEDPQPGLAGTDEAG 70
PG PG +G+P EG
Sbjct: 584 PGPGPKGPEGPPGKGERG 603

Score = 48 (22.3 bits), Expect = 0.040, Poisson P(2) = 0.040
Identities = 9/22 (40%), Positives = 11/22 (50%)

Query: 39 PAPHPGPGSQPSPGTPGEGDP 60
PP + PG PG G P
Sbjct: 268 PGPPGIRPGPPGPPGGEKGEKGEQG 292

>gp|L31345|CHKCOL3ALX_1 collagen type III [Gallus gallus]

Length = 666

Score = 47 (21.8 bits), Expect = 1.1e+02, P = 1.0
Identities = 9/17 (52%), Positives = 10/17 (58%)

Query: 44 PGPLSQSPGTPWEGDP 60
PP +S PG PG G P
Sbjct: 168 PQPISFGPPSGPSGPP 184

Score = 45 (20.9 bits), Expect = 0.30, Poisson P(2) = 0.25
Identities = 9/25 (36%), Positives = 11/25 (44%)

Query: 51 PGTPGWEDPQPGLAGTDEAGFQAQS 75
PG PG G+P G + +
Sbjct: 811 PGAPGQNGEPGKGERGPPGLRGEA 835
Score = 39 (18.1 bits), Expect = 0.061, Poisson P(3) = 0.078
Identities = 8/15 (53%), Positives = 10/15 (66%)

Query: 48 SQSPGTPGWEGDPQG 62
    $ SPG G +G+P G$
Sbjct: 733 SGSPGPKGDKGEPPG 747
Score = 39 (18.1 bits), Expect = 0.061, Poisson P(3) = 0.078
Identities = 8/15 (53%), Positives = 10/15 (66%)

Query: 41 PHHPGPLSQSPGTPGWEGD 59
    P P + PG+PG++G+
Sbjct: 333 PPGPPGTAGFPSPGFKGE 351
Score = 35 (16.3 bits), Expect = 0.047, Poisson P(4) = 0.046
Identities = 8/20 (40%), Positives = 12/20 (60%)

Query: 39 PAPHIPGPLSQSPGTPGWEG 58
    P P P PG PG G
Sbjct: 298 PGPGPTGERGRPGNPGPG 317

>gp | L02918 | MUSCOL5A2X_1 procollagen type V alpha 2 [Mus musculus]
    Length = 1437
Score = 53 (24.6 bits), Expect = 14., P = 1.0
Identities = 10/21 (47%), Positives = 12/21 (57%)

Query: 51 PGTPGWEGDPQGLAGTDEAGF 71
    PG PG +G+ T EGF
Sbjct: 258 PGKPDGEDEPGRNMTGEVG 278
Score = 48 (22.3 bits), Expect = 0.068, Poisson P(2) = 0.066
Identities = 9/22 (40%), Positives = 10/22 (45%)

Query: 39 PAPHIPGPLSQSPGTPGWEGD 60
    P P P PG PG G
Sbjct: 150 PGPRGPGDGEPGMGPQGAP 171

>pir | S08012 | S08012 collagen-like protein, placental - human (fragment)
    >gp | X15038 | HSCOLLHM_1 collagen-like protein [Homo sapiens]
    Length = 447
Score = 55 (25.6 bits), Expect = 6.7, P = 1.0
Identities = 13/34 (38%), Positives = 15/34 (44%)

Query: 39 PAPHHPGPLSQSPGTPGWEGDPQGLAGTDEAGFQ 72
    P PG PG +G+D G+ T GFQ
Sbjct: 83 PGKDPGPGKGPGVKGKGEKPNDCECGPPTLPEGFQ 116
Score = 46 (21.4 bits), Expect = 0.071, Poisson P(2) = 0.069
Identities = 13/32 (40%), Positives = 14/32 (45%)

Query: 34 AQQEGPAPAPHHPGPLSQSPGTPGWEGDPQGLAG 65
    A + GPG P PGTFG G PG G
Sbjct: 44 AGEPPGPGPGPGPGPGPGLPGDPDPPGPKG 75

79
>sp|P27393|CA24_ASCSU PROCOLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
>pir|S16366|S16366 collagen alpha 2(IV) chain precursor - pig roundworm >gp|M675071|NEMA2C4A_1 alpha-2 (IV) collagen gene product
[Ascaris suum]
Length = 1763

Score = 48 (22.3 bits), Expect = 0.081, Poisson P(2) = 0.078
Identities = 11/31 (35%), Positives = 13/31 (41%)

Query: 33 PAQQEGPAPHPGPGLSQPSGTGPWGEGDPQGL 63
    P  P  P  PG  +G+P  L
Sbjct: 580 PGAPGRPGPQPGDPQGKGEPTQL 610

Score = 48 (22.3 bits), Expect = 0.081, Poisson P(2) = 0.078
Identities = 12/32 (37%), Positives = 13/32 (40%)

Query: 47 LSQSPGTGPWGEGDPQGLAGTDEAGFQAQSFQD 78
    P  PG  +GP  +  +L
Sbjct: 1215 LKGEPGLPGLGQPGPRGEAGLPGAPGRD 1246

Score = 47 (21.8 bits), Expect = 0.16, Poisson P(2) = 0.15
Identities = 13/36 (36%), Positives = 13/36 (36%)

Query: 39 PAPHHHPGPLSQSPGTGPWGEGDPQGLAGTDEAGFQAQ 74
    P  P  L  PG  G  GP  EGF  Q
Sbjct: 1362 PGPQGPAGLPGPGGLGEPGLPGFGQKGETGFGQ 1397

>gp|M25984|CHKCOLA29_1 Chicken alpha-2 collagen gene type I gene, exon 52.
[Gallus gallus]
Length = 896

Score = 51 (23.7 bits), Expect = 27., P = 1.0
Identities = 11/27 (40%), Positives = 14/27 (51%)

Query: 51 PGTPGWEGDPQGLAGTDEAGFQAQSFQ 77
    PG  +GP  G  EGF  A+++L
Sbjct: 624 PGPPGPGPQPGPNGGYEVGFDAEYR 650

Score = 40 (18.6 bits), Expect = 8.8, Poisson P(2) = 1.0
Identities = 13/36 (36%), Positives = 13/36 (36%)

Query: 38 GPAPHHHPGPLSQSPGTGPWGEGDPQGLAGTDEAGFQ 72
    GPA  SQ  P  PG  +G++
Sbjct: 607 GPPVPVRGSHCSQPGAPPAGPPAGPGPQPGPQPGPQPGPGGGE 641

Score = 39 (18.1 bits), Expect = 0.083, Poisson P(3) = 0.080
Identities = 8/26 (30%), Positives = 11/26 (42%)

Query: 33 PAQQEQGPAPHPGPGLSQSPGTGPWG 58
    P++  P  P  +PG  G  G
Sbjct: 243 PGGERGEPPVPVPSGFAGPPGAAGQGP 268

>gp|L15194|MAUFEMRNA_1 Golden delicious apple fruit expressed mRNA, complete cds. [Malus domestica]
Length = 119
Score = 45 (20.9 bits), Expect = 1.6e-02, P = 1.0
Identities = 8/14 (57%), Positives = 9/14 (64%)

Query:  47 LSQSPGTPGWEGDP 60
       + SPGTPG G P
Sbjct:  48 MPMSPGTPGTPGTP 61

Score = 43 (20.0 bits), Expect = 0.093, Poisson P(2) = 0.089
Identities = 8/18 (44%), Positives = 9/18 (50%)

Query:  50 SPGTPGWEQDPQGLAGTD 67
       +PGTPG G P D
Sbjct:  54 TPQTPGTPGTPASARAKD 71

>gp|M20789|HUMC1A1_1 alpha-1 type I collagen [Homo sapiens]
Length = 589

Score = 42 (19.5 bits), Expect = 5.7e-02, P = 1.0
Identities = 8/22 (36%), Positives = 10/22 (45%)

Query:  39 PAPHHHPGPLSQSPGTPGWEGDP 60
       + P P + G PG +GP
Sbjct:  373 PGPPGPGAGAAGPAGNGAPDGQ 394

Score = 40 (18.6 bits), Expect = 5.6, Poisson P(2) = 1.0
Identities = 11/30 (36%), Positives = 12/30 (40%)

Query:  41 PHHPGPLSQSPGTPGWEGDPQGLAGT DEAG 70
       +P + SFG +G P EAG
Sbjct:  504 PKGPAGERGSPGPAGPKGSPGEAGRPGEAG 533

Score = 38 (17.7 bits), Expect = 0.084, Poisson P(3) = 0.090
Identities = 8/26 (30%), Positives = 12/26 (46%)

Query:  33 PAQQEGPAPHHPGPLSQSPGTPGWEG 58
       + P + P + L + GP G
Sbjct:  241 PGGERGPPGQGARGLPGTALPGMKG 266

Score = 35 (16.3 bits), Expect = 1.9, Poisson P(3) = 0.85
Identities = 8/21 (38%), Positives = 12/21 (58%)

Query:  51 PGTPGWEGDPQGLAGTDEAGF 71
       +PG G GP G GF
Sbjct:  475 PGPTGLPGPPGERGGPGRGSRGF 495

>pir|A33848|A33848 secretory protein 1 - midge (Chironomus tentans) (fragment)
>gp|M24276|CHISPA_1 C.tentans 140-kd secretory protein (sp140)
mRNA, partial cds, clone pCt140.1. [Chironomus tentans]
Length = 84

Score = 41 (19.1 bits), Expect = 0.10, Poisson P(2) = 0.10
Identities = 9/24 (37%), Positives = 12/24 (50%)

Query:  32 NPAQQEGPAPHHPGPLSQSPGTPG 55
       +N +E G A P +G PG G
Sbjct:  28 NGSKPEGKAPGNNGKSCSKPGKG 51
Score = 41 (19.1 bits), Expect = 0.10, Poisson P(2) = 0.10
Identities = 10/32 (31%), Positives = 15/32 (46%)

Query: 39 PAPHHPGPLSQSPGTPGWEGDPQQLAGTDEAG 70
  P  PG +S  PG +G  +G  +G  +G
Sbjct: 32 PEGKAPGNNNGKCSKPGKPNGKNGKSGSKPEG 63

>gp | M63473 | HUMA5CL18_1 type IV collagen [Homo sapiens]
  Length = 762

Score = 59 (27.4 bits), Expect = 1.8, P = 0.83
Identities = 12/32 (37%), Positives = 15/32 (46%)

Query: 39 PAPHHPGPLSQSPGTPGWEGDPQQLAGTDEAG 70
  P  P  +S  PG  PG  G+  +G  +G
Sbjct: 204 PGPPGPKGISPQPNPGLPQGPVGGGHPG 235

Score = 46 (21.4 bits), Expect = 0.13, Poisson P(2) = 0.12
Identities = 11/29 (37%), Positives = 12/29 (41%)

Query: 32 NPAQQEGPAPHHPGPLSQSPGTPGWEGDP 60
  NP  P  P  G  PG  +G
Sbjct: 218 NPGLPGEPGPVGGGHPGQQPGPGEKGP 246

Score = 46 (21.4 bits), Expect = 0.13, Poisson P(2) = 0.12
Identities = 12/33 (36%), Positives = 14/33 (42%)

Query: 26 TINIMYNPAQQEGPAPHHPGPLSQSPGTPGWEG 58
  TI  M  P  Q  P  +P+PG  G
Sbjct: 105 TIGDMGFPGPQVEGPPGSVPQGPQSPGLP 137

>pil!A37969!A37969 collagen alpha 5(IV) chain - human (fragment)
  Length = 763

Score = 59 (27.4 bits), Expect = 1.8, P = 0.83
Identities = 12/32 (37%), Positives = 15/32 (46%)

Query: 39 PAPHHPGPLSQSPGTPGWEGDPQQLAGTDEAG 70
  P  P  +S  PG  PG  G+  +G  +G
Sbjct: 205 PGPPGPKGISPQPNPGLPQGPVGGGHPG 236

Score = 46 (21.4 bits), Expect = 0.13, Poisson P(2) = 0.12
Identities = 11/29 (37%), Positives = 12/29 (41%)

Query: 32 NPAQQEGPAPHHPGPLSQSPGTPGWEGDP 60
  NP  P  P  G  PG  +G
Sbjct: 219 NPGLPGEPGPVGGGHPGQQPGPGEKGP 247

Score = 46 (21.4 bits), Expect = 0.13, Poisson P(2) = 0.12
Identities = 12/33 (36%), Positives = 14/33 (42%)

Query: 26 TINIMYNPAQQEGPAPHHPGPLSQSPGTPGWE 58
  TI  M  P  Q  P  +P+PG  G
Sbjct: 106 TIGDMGFPGPQVEGPPGSVPQGPQSPGLP 138
>gpIM311151 HUMCOL4A5_1 COL4A5 gene product [Homo sapiens]
Length = 772

Score = 59 (27.4 bits), Expect = 1.8, P = 0.83
Identities = 12/32 (37%), Positives = 15/32 (46%)

Query: 39 PAPHHGPGLSQSPGTPGWEGDPQGLAGTDEAG 70
PPP +S PG PG G +P +G G
Subject: 214 PGPPGPKGIGSPPGPNPGPGLGEPFPVPVGGGHPG 245

Score = 46 (21.4 bits), Expect = 0.13, Poisson P(2) = 0.12
Identities = 12/33 (36%), Positives = 14/33 (42%)

Query: 26 TINIMYNPAQQEGPAPHHGPGLSQSPGTPGWEG 58
MIPQ PG +PG +PGG
Subject: 115 TIGDMGPFPQVGPVGPQGPPGPGPLG 147

>spIQ014931 GP22_LITCA MAJOR MICROFILARIAL SHEATH PROTEIN PRECURSOR.
>pirJH07881 JH0788 sheath glycoprotein gp22 precursor - nematode
(Litomosoides carinii) >gpIM962321 LITGP22A_1 major microfilarial
sheath protein [Litomosoides carinii]
Length = 148

Score = 48 (22.3 bits), Expect = 61., P = 1.0
Identities = 10/31 (32%), Positives = 14/31 (45%)

Query: 35 QQEGPAPHHGPGLSQSPGTPGWEGDPQGLAG 65
QGP PPPL +P +P + +G
Subject: 57 QPMGPQPMEPFPQLPMPQPSQPMQVPDRSCSG 87

Score = 43 (20.0 bits), Expect = 0.13, Poisson P(2) = 0.12
Identities = 8/19 (42%), Positives = 9/19 (47%)

Query: 39 PAQQEGPAPHHGPGLSQSP 51
PGPP +PP +P
Subject: 50 PQPMGPQPMGPQPMEPQP 68

>spIPO59971 CA25_HUMAN PROCOLLAGEN ALPHA 2(V) CHAIN PRECURSOR.
Length = 1486

Score = 49 (22.8 bits), Expect = 55., P = 1.0
Identities = 9/22 (40%), Positives = 10/22 (45%)

Query: 39 PAPHHGPGLSQSPGTPGWEGDP 60
PPP +PG PG GP
Subject: 152 PGPRPQGIDGEPVGPGPGAP 173

Score = 47 (21.8 bits), Expect = 0.13, Poisson P(2) = 0.13
Identities = 9/21 (42%), Positives = 11/21 (52%)
Query: 51 PGPGGDPQGGLAGTDEAGF 71
   PG PG +G+ P   E G F
Shjct: 260 PGKPGEDEGPGRNGNPGGEVGF 280

>pir|S20833|S20833 Collagen alpha 1(XIV) chain - Chicken (fragments)
   Length = 56

Score = 62 (28.8 bits), Expect = 0.15, P = 0.14
Identities = 14/43 (42%), Positives = 16/33 (48%)

Query: 33 PAQQEGPAPHPGPGSQQSPGTPGWEGRDPQGLAG 65
   P +PG PHA H PSQP P DP AG
Shjct: 23 PGEKGPGAGHTGPGSQGPAGPPGYCDPSSCAG 55

>sp|P35085|CBP_DICDI CALCIUM BINDING PROTEIN. >gp|U03413|DDU03413_1 calcium binding protein [Dictyostelium discoideum]
   Length = 87

Score = 48 (22.3 bits), Expect = 73., P = 1.0
Identities = 11/25 (44%), Positives = 11/25 (44%)

Query: 41 PHHPGPGSQQSPGTPGWEGRDPQGLAG 65
   P PG QPG PG QG G
Shjct: 121 PGQPGYPPQQGPGAPGQYPQQQGQPG 145

Score = 39 (18.1 bits), Expect = 8.5, Poisson P(2) = 1.0
Identities = 9/23 (39%), Positives = 9/23 (39%)

Query: 33 PAQQEGPAPHPGPGSQQSPGTPG 55
   P Q P PG QPG G
Shjct: 92 PGQYPPQQGPGQPPGQPPGQPGQSG 114

Score = 37 (17.2 bits), Expect = 0.16, Poisson P(3) = 0.15
Identities = 8/18 (44%), Positives = 9/18 (50%)

Query: 49 QSPGTPGWEGRDPQGLAGT 66
   QPG PG QG+ T
Shjct: 180 QQPGQPGAYPPQGQGVQNT 197

>pir|PQ0612|PQ0612 collagen alpha 1(XVI) chain - human (fragment)
   >gp|SS7132|SS7132_1 type XVI collagen alpha 1 chain [Homo sapiens]
   Length = 1186

Score = 55 (25.6 bits), Expect = 7.1, P = 1.0
Identities = 13/34 (38%), Positives = 15/34 (44%)

Query: 39 PAPHPGPGSQQSPGTPGWEGRDPQGLAGTDEAGFQ 72
   P PG PG G +GD P + T GFQ
Shjct: 67 PGKEFGPGKPGKPGVKGEKGDPEVCPTLPEGFQ 100

Score = 46 (21.4 bits), Expect = 0.21, Poisson P(2) = 0.19
Identities = 13/32 (40%), Positives = 14/32 (43%)

Query: 34 AQEGPAPHPGPGSQQSPGTPGWEGRDPQGLAG 65
   A +GP P PGTFG GPG G
Shjct: 28 AGEPPGPPGPPGIGLPGTPGDPGPPGPKG 59
Score = 40 (18.6 bits), Expect = 1.1e+06, P = 1.0
Identities = 10/28 (35%), Positives = 11/28 (39%)

Query: 33 PAQQEGPAPHPGQLSPGTPGWEYDP 60
        P + P P P L P G G P
Sbjct: 806 PGLRGDPAGPPGGLMGPPFKEKLGHP 833
Score = 37 (17.2 bits), Expect = 1.1, Poisson P(3) = 0.67
Identities = 8/24 (33%), Positives = 11/24 (45%)

Query: 51 PGTPGWEYDPQGLAGTDEAGFQAQ 74
        PG PG +G P +G A A
Sbjct: 500 PGVPGLQGVPGNNGLPGQPLTHE 523

>gp | U00697 | U00697_1 orphan receptor COUP-TFII [Gallus gallus]
    Length = 410
Score = 45 (20.9 bits), Expect = 2.0e+02, P = 1.0
Identities = 12/36 (33%), Positives = 16/36 (44%)

Query: 39 PAPHPGQLSPGTPGWEYDPQGLAGTDEAGFQAQ 74
        P P +PT PG G P A A A Q Q
Sbjct: 28 PVQGPPAGTPHFTFQTPGPPGPPSTPAQSNASQAQQ 63
Score = 44 (20.5 bits), Expect = 0.25, Poisson P(2) = 0.22
Identities = 10/23 (43%), Positives = 12/23 (52%)

Query: 36 QEGPAPHPGQLSPGTPGWEY 58
        Q PAP GP +P TP G
Sbjct: 22 QPAPAPPQGPAGTPHTQTPG 44

>gp | M13027 | MUSC4A15_1 alpha-1 type IV collagen [Mus musculus]
    Length = 43
Score = 38 (17.7 bits), Expect = 0.25, Poisson P(2) = 0.22
Identities = 7/10 (70%), Positives = 7/10 (70%)

Query: 51 PGTPGWEYDP 60
        PG PG EG P
Sbjct: 18 PGLPGPEGPP 27
Score = 38 (17.7 bits), Expect = 0.25, Poisson P(2) = 0.22
Identities = 8/18 (44%), Positives = 8/18 (44%)

Query: 41 PHHPGQLSPGTPGWEY 58
        P PGP G PG G
Sbjct: 5 PGPGPGYDIKGEGLPG 22

>pir | S22215 | S22215 Collagen alpha 1(C) chain - Mouse (fragment)
    >gp | X69018 | MMA1XCOL_1 collagen-alpha-1 type X [Mus musculus]
        Length = 243
Score = 44 (20.5 bits), Expect = 2.7e+02, P = 1.0
Identities = 7/11 (63%), Positives = 9/11 (81%)

85
Query:  50 SPGTPGWEGDP 60
   +PG PG +GDP
Sbjct:  18 NPGLPGKKGDP 28

Score = 43 (20.0 bits), Expect = 0.26, Poisson P(2) = 0.23
Identities = 9/20 (45%), Positives = 10/20 (50%)

Query:  51 PGTPGWEGDPQGLAGTDEAG 70
   PG PG +GDP AG
Sbjct:  76 PGFPGBKDGPNPGAPGPAG 95

>sp!P12105!CA13_CHICK COLLAGEN ALPHA 1(III) CHAIN (FRAGMENTS).
   >pir!A05269!A05269 collagen alpha 1(III) chain - chicken
      (fragments)
      Length = 615

Score = 48 (22.3 bits), Expect = 75., P = 1.0
Identities = 12/39 (30%), Positives = 14/39 (35%)

Query:  32 NPAQQEGPAPHPGPLSQLSPGTPGWEGDPQGLAGTDEAG 70
      NP ++ PP GPG GPG EG
Sbjct:  100 NPGERGEPFPQGAGPQPQGPPGAGSPGKGEKEMGEPG 138

Score = 41 (19.1 bits), Expect = 3.0, Poisson P(2) = 0.95
Identities = 10/39 (25%), Positives = 17/39 (43%)

Query:  21 VVLGFTINMYNPAQQEGPAPHPGPLSQLSPGTPGWEGD 59
      +LG + QP P + PG+PG++G+
Sbjct:  50 IPLGECCPVCQTTQPTKGPGPGPGTAGFPGFSPGFKGE 88

Score = 37 (17.2 bits), Expect = 0.28, Poisson P(3) = 0.25
Identities = 9/22 (40%), Positives = 10/22 (45%)

Query:  51 PGTPGWEGDPQGLAGTDEAGFQ 72
      PG PG G+ GFQ
Sbjct:  254 PGHPGPAGNNGAPGKARGERFQ 275

>pir!S23810!S23810 collagen alpha 1(XVI) chain precursor - human
   >gp!M92542!HUMCOL16A_1 alpha-1 type XVI collagen [Homo sapiens]
   Length = 1603

Score = 55 (25.6 bits), Expect = 7.2, P = 1.0
Identities = 13/34 (38%), Positives = 15/34 (44%)

Query:  39 PAPHHPGPLSQLSPGTPGWEGDPQGLAGTDEAGFQ 72
      P PG PG +GDP + T GFQ
Sbjct:  484 PGKEGPPGKPGPKPGVKGKEKDPCVECPILPEGFQ 517

Score = 46 (21.4 bits), Expect = 0.28, Poisson P(2) = 0.25
Identities = 13/32 (40%), Positives = 14/32 (45%)

Query:  34 AQEQGPAPHPGPLSQLSPGTPGWEGDPQGLAG 65
      + + GP P PGPG GPG G
Sbjct:  445 AGEPGPPGLGPAGGLPGPGPDGPAGPKG 476

Score = 40 (18.6 bits), Expect = 1.1e+03, P = 1.0
Identities = 10/28 (35%), Positives = 11/28 (39%)
Query: 33 PAQQEGPAPHHPGPLSQSPGTPGWEGDP 60
  P + P P P L P G G GP
Sbjct: 1223 PGLRGDPGPAGPGPGLMGPGFPKFGKRTGHP 1250

Score = 37 (17.2 bits), Expect = 2.0, Poisson P(3) = 0.87
Identities = 8/24 (33%), Positives = 11/24 (45%)

Query: 51 PGTPGWEGDPQGLAGTDEAGFQAQ 74
  PG PG +G P +G A+ 
Sbjct: 917 PGVPGLQGVPGNNGLPGQGPLTAE 940

>gp !D14076! RATTESDYN_1 testicular dynamin [Rattus norvegicus]
  Length = 848

Score = 47 (21.8 bits), Expect = 1.1e+02, P = 1.0
Identities = 10/22 (45%), Positives = 10/22 (45%)

Query: 33 PAQQEGPAPHHPGPLSQSPGTP 54
  PA GPAP P P S P
Sbjct: 779 PASSRGPAAPSPGPGHSGAPP 800

Score = 45 (20.9 bits), Expect = 0.28, Poisson P(2) = 0.25
Identities = 9/22 (40%), Positives = 10/22 (45%)

Query: 39 PAPHPGPLSQSPGTPGWEGDP 69
  PP PGPL P + G P
Sbjct: 800 PVPFRPGPLPFPNSDSYGAP 821

>pir !S12898! S12898 Collagen alpha 2(VIII) chain - Bovine (fragment)
  Length = 469

Score = 45 (20.9 bits), Expect = 2.0e+02, P = 1.0
Identities = 10/30 (33%), Positives = 14/30 (46%)

Query: 33 PAQQEGPAPHPGPLSQSPGTPGWEGDPQG 62
  P + P H P + G-PG+G PG
Sbjct: 337 PGERPGLGAPAHPGPGPTPGXCEPGFTGRPGG 366

Score = 44 (20.5 bits), Expect = 0.29, Poisson P(2) = 0.25
Identities = 10/24 (41%), Positives = 11/24 (45%)

Query: 47 LSQSPGTGWEGDPQGLAGTDEAG 70
  L+ PG PG G P DE G
Sbjct: 446 LTGPPGPPGPGGAPGAFDETG 469

>gp !M58526! HUMCOLA5IV_1 alpha-5 type IV collagen [Homo sapiens]
  Length = 1544

Score = 59 (27.4 bits), Expect = 1.8, P = 0.84
Identities = 12/32 (37%), Positives = 15/32 (46%)

Query: 39 PAPHPGPLSQSPGTPGWEGDPQGLAGTDEAG 70
  PP P +5 PG PG G+P +G G
Sbjct: 1046 PGPPGPKGISGPGNPLPGPEGPGPGPGVGGHPG 1077
Score = 46 (21.4 bits), Expect = 0.29, Poisson P(2) = 0.25
Identities = 12/29 (41%), Positives = 14/33 (42%)

Query: 32 NPAQEGPAHPHPGGLSQQSPGTPGWEQGD 60
NP P P G PG + G
Sbjct: 1060 NPGLPGPVPVGGGHPGQPGPPGEKGP 1088

Score = 46 (21.4 bits), Expect = 0.29, Poisson P(2) = 0.25
Identities = 12/33 (36%), Positives = 14/33 (42%)

Query: 26 TINIMYPNAQEGPAHPHPGGLSQQSPGTPGWEQG 58
TI M P Q P + PG + PG G
Sbjct: 947 TIGDMGFPQPGVQGPPGPSPGPGQPGPGLPG 979

>sp!P02462|CA14_HUMAN PROCOLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
Length = 1569

Score = 47 (21.8 bits), Expect = 1.1e+02, P = 1.0
Identities = 8/15 (53%), Positives = 10/15 (66%)

Query: 51 PGTPGWEQDPQGLAG 65
PG PG + GDP + G
Sbjct: 150 PGLPGMKGDQPEILG 164

Score = 46 (21.4 bits), Expect = 0.30, Poisson P(2) = 0.26
Identities = 9/21 (42%), Positives = 10/21 (47%)

Query: 52 GTPGWEQDPQGLAGTGDEAFQ 72
GPW GP G + GFQ
Sbjct: 1269 GNPGWPGAPVPGPKGDQPFQ 1289

Score = 45 (20.9 bits), Expect = 0.58, Poisson P(2) = 0.44
Identities = 9/17 (52%), Positives = 10/17 (58%)

Query: 44 PGPSQGSPGTGWEQGD 60
PG + PG PG EG P
Sbjct: 638 PGKIVPLPGPPGAEGLP 654

Score = 44 (20.5 bits), Expect = 1.1, Poisson P(2) = 0.68
Identities = 10/21 (47%), Positives = 11/21 (52%)

Query: 39 PAPHHPGPLSQQSPGTGWEQGD 59
PP L SPG PG + GD
Sbjct: 645 PGPPGAEGLPGPGFQPGPD 665

>pir|S16876|CGHU4B collagen alpha 1(IV) chain precursor - human
>gp|M26576|HUMOCOL1A42_1 alpha-1 type IV collagen [Homo sapiens]
Length = 1669

Score = 47 (21.8 bits), Expect = 1.1e+02, P = 1.0
Identities = 8/15 (53%), Positives = 10/15 (66%)

Query: 51 PGTPGWEQDPQGLAG 65
PG PG + GDP + G
Sbjct: 150 PGLPGMKGDQPEILG 164

Score = 46 (21.4 bits), Expect = 0.30, Poisson P(2) = 0.26
Identities = 9/21 (42%), Positives = 10/21 (47%)

Query: 52 GTPGWGDPQGLAGTDEAGFQ 72
   GP + GFQ
Shjct: 1269 GNPGWPGAPGVPPKGDPGFQ 1289

Score = 45 (20.9 bits), Expect = 0.58, Poisson P(2) = 0.44
Identities = 9/17 (52%), Positives = 10/17 (58%)

Query: 44 PGPLSQSPGTGPGWEAGDP 60
   PG + PG EG P
Shjct: 538 PGKIVPLPGPAGELP 544

Score = 44 (20.5 bits), Expect = 1.1, Poisson P(2) = 0.68
Identities = 10/21 (47%), Positives = 11/21 (52%)

Query: 39 PAPHHPGPLSQSPGTGPGWED 59
   P P L SPG PG +GD
Shjct: 645 PGPPGAEGLPSPGPFPQGD 665

Query: 26 TINIMYNPAQQEGPAPHHPGPLSQSPGTGPGWED 58
   T I M Q P P +PG +G G
Shjct: 1028 TIGDMGFPGPQGVEGPPPGSGVPGPGPGLPGP 1060

>sp | P29400 | CA54_HUMAN COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR. >pir | S22917 | S22917
collagen alpha 5(IV) chain precursor - human
>gp | U04520 | HS4COL5A5 | COL4A5 gene product [Homo sapiens]
>gp | U04520 | HS4COL5A5 | COL4A5 gene product [Homo sapiens]
Length = 1685

Score = 59 (27.4 bits), Expect = 1.8, P = 0.84
Identities = 12/32 (37%), Positives = 15/32 (46%)

Query: 39 PAPHPPGPLSQSPGTGPGWED 59
   P P P + S PG PG +G + G
Shjct: 1127 PGPPGPKGISGPPGNPGPGLPGEPGVPVGGGHPG 1158

Score = 46 (21.4 bits), Expect = 0.30, Poisson P(2) = 0.26
Identities = 11/29 (37%), Positives = 12/29 (41%)

Query: 32 NPAQQEGPAPHHPGPLSQSPGTGPGWED 60
   NP P G G + P +G G
Shjct: 1141 NPGLPGEPFGPGPVGPGHGPQGPGEKP 1169

Score = 46 (21.4 bits), Expect = 0.30, Poisson P(2) = 0.26
Identities = 12/33 (36%), Positives = 14/33 (42%)

Query: 25 TINIMYNPAQQEGPAPHHPGPLSQSPGTGPGWED 58
   T I M Q P P +PG +G G
Shjct: 1028 TIGDMGFPGPQGVEGPPPGSGVPGPGPGLPGP 1060

>gp | Z22964 | CECOLA2IV_1 a2(IV) collagen [Caenorhabditis elegans]
   Length = 1735

Score = 52 (24.2 bits), Expect = 20., P = 1.0
Identities = 12/28 (42%), Positives = 13/28 (46%)

Query: 33 PAQQEGPAPHHPGPLSQSPGTGPGWED 60
   P Q P G L PGTPG +G P
Shjct: 725 PGQPFGPAKGDGGLPGLPGPGTPGPGMP 750
Score = 46 (21.4 bits), Expect = 0.32, Poisson P(2) = 0.27
Identities = 12/32 (37%), Positives = 16/32 (50%)

Query: 47 LSQSPGTPGWEQDPQGLAGTDEAGFQAQSFQD 78
  L SPG PG +G P ++GF Q Q+
Sbjct: 1250 LKGSPGYPQDGLPGIPGKLGDGSGFPGQPGQE 1281

Score = 36 (16.7 bits), Expect = 7.0, Poisson P(3) = 1.0
Identities = 10/30 (33%), Positives = 12/30 (40%)

Query: 41 PHHPGPLSQQSPGTPGWEQDPQGLAGTDEAG 70
  P P +Q + PG + GL GT G
Sbjct: 750 PGEPAPENQVNPAPPQGPLPGTQGEG 779

>sp|P981201|CA14_DROME PROCOLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
  >pir|A31893|A31893 collagen alpha 1(IV) chain precursor - fruit fly
    (Drosophila melanogaster) >gp|M23704|DROCA1A1A_1 alpha-1 type IV
    collagen [Drosophila melanogaster] >gp|M96575|DROCOL4G_1 type IV
    collagen [Drosophila melanogaster] >gp|J02727|DROCOLIV_1 type IV
    pro-collagen [Drosophila melanogaster]
  Length = 1775

Score = 50 (23.2 bits), Expect = 39., P = 1.0
Identities = 10/24 (41%), Positives = 13/24 (54%)

Query: 47 LSQSPGTPGWEQDPQGLAGTDEAG 70
  L+ PG PG +G+P T EG
Sbjct: 362 LNLPGNPGQKGEPRAGATQEPG 385

Score = 39 (18.1 bits), Expect = 0.34, Poisson P(3) = 0.29
Identities = 9/20 (45%), Positives = 9/20 (45%)

Query: 39 PAPHHPGPLSQQSPGTPGWEQ 58
  P P L G PG EG
Sbjct: 94 PGPLGPTGLKEMGGFPMEG 113

Score = 39 (18.1 bits), Expect = 0.34, Poisson P(3) = 0.29
Identities = 7/21 (33%), Positives = 9/21 (42%)

Query: 51 PGTPGWEQDPQGLAGTDEAGF 71
  PG PG +G P G+
Sbjct: 1149 PDAPGMGDLPGAAGAPGAVGY 1169

>pir|S17035|S17035 Collagen alpha 1(XIV) chain - Chicken
  Length = 156

Score = 54 (25.1 bits), Expect = 8.4, P = 1.0
Identities = 13/28 (46%), Positives = 13/28 (46%)

Query: 38 GPAPHHPGPLSQQSPGTPGWEQDPQGLAG 65
  GPAH P SQ P P DP AG
Sbjct: 168 GPAGHTGPGPSQGPAPGYPYCDPSCAG 195

Score = 42 (19.5 bits), Expect = 0.39, Poisson P(2) = 0.32
Identities = 9/29 (31%), Positives = 12/29 (41%)

Query: 32 NPAQQEGPAPHHPGPLSQQSPGTPGWEQDP 60
>pir|S34665|S34665 cuticular collagen - root-knot nematode (Meloidogyne
cognita) >gp|Z24734|MILEMIA_1 cuticular collagen [Meloidogyne
cognita]
Length = 286

Score = 42 (19.5 bits), Expect = 5.3e+02, P = 1.0
Identities = 9/20 (45%), Positives = 9/20 (45%)

Query: 39 PAPHHPGPLSQQSPGTPGWEG 58
  P P P S P G G
Sbjct: 226 PGPVPGAPGSPGPGAPGPQPG 245

Score = 37 (17.2 bits), Expect = 19., Poisson P(2) = 1.0
Identities = 8/20 (40%), Positives = 9/20 (45%)

Query: 41 PHHPGPLSQQSPGTGECWPWEGDP 60
  P P + P G G
Sbjct: 231 PAGPSGPAPGKPGAPGQPG 250

Score = 36 (16.7 bits), Expect = 4.0e+03, P = 1.0
Identities = 8/20 (40%), Positives = 9/20 (45%)

Query: 51 PGTGPWGDPQQLAGTDEAG 70
  P G P + G G
Sbjct: 143 PGPGPGPAPGKSSGAGPPG 162

Score = 35 (16.3 bits), Expect = 0.40, Poisson P(3) = 0.33
Identities = 7/14 (50%), Positives = 7/14 (50%)

Query: 33 PAQQEGPAHPHPGP 46
  P G P A H G P
Sbjct: 166 PGGPGPAPGHHGGP 179

>pir|A30296|A30296 collagen alpha (VII) chain - human (fragments)
Length = 62

Score = 49 (22.8 bits), Expect = 25., P = 1.0
Identities = 12/26 (46%), Positives = 12/26 (46%)

Query: 45 CPLSQSPGTGECWPWEGDPQQLAGTDEAG 70
  G L P G F G G E A G
Sbjct: 19 GELCIEPGBKGPLPAPGPPGPPKPEAG 44

Score = 39 (18.1 bits), Expect = 0.40, Poisson P(2) = 0.33
Identities = 10/29 (34%), Positives = 12/29 (41%)

Query: 30 MYNPAQQEGCPAPHPHPGPLSQQSPGTPGWEG 58
  +Y Q+G P G P G
Sbjct: 11 VYXEGQGGGELCIEPGBKGPLPAPGPPGPPG 39

Score = 35 (16.3 bits), Expect = 6.9, Poisson P(2) = 1.0
Identities = 10/33 (30%), Positives = 12/33 (36%)

Query: 33 PAQQEGPAHPHPGPLSQQSPGTGECWPWEGDPQQLAG 65
P + GP  G + GPG G G
Sbjct:  4 PRPEPGPVYXEKGQQGELCGEPKPGILPGPGP 36

>pir|A44309|A44309 type II collagen alpha 1 chain, COL2A1 - human (fragment)
length = 346

Score = 44 (20.5 bits), Expect = 2.6e+02, P = 1.0
Identities = 9/26 (34%), Positives = 14/26 (53%)

Query:  50 SPGTPGWEQPQGLAGTDEAGFQAQS 75
+PGTPG GP  T + G  ++
Sbjct:  315 APGTPGFGSFPAGPTGKQGDRGEA 340

Score = 43 (20.0 bits), Expect = 0.40, Poisson P(2) = 0.33
Identities = 9/22 (42%), Positives = 11/22 (50%)

Query:  39 PAPHPGPLSQSPGTPGWEGDP 60
+ P P P + GPG + GP
Sbjct:  268 PGVPQGGLTGPAGEPGRQGSP 289

WARNING: HSPs involving 158 database sequences were not reported due to the limiting value of parameter B = 50.

Parameters:
E = 10., S = 55 (25.6 bits), E2 = 0.14, S2 = 35
W = 3, T = 11 (5.1 bits), X = 22 (10.2 bits)
M = BLOSUM62
H = 1, V = 100, B = 50
-gapdecayrate 0.5 (the default)

Statistics:
Lambda = 0.322 nats/unit score, Lambda/in2 = 0.465 bits/unit score
K = 0.142, H = 0.667 bits/position
Expected/Observed high score = 58 (27.0 bits) / 63 (29.3 bits)
# of letters in query: 78
# of neighborhood words in query: 2172
# of exact words scoring below T: 0
Database: Non-redundant PDB+SwissProt+SPupdate+PIR+GenPept+GPupdate, 5:04 AM EDT Apr 13, 1994
# of letters in database: 32,528,034
# of word hits against database: 6,536,751
# of failed hit extensions: 4,809,644
# of excluded hits: 1,722,483
# of successful extensions: 4624
# of overlapping HSPs discarded: 4162
# of HSPs reportable: 462
# of sequences in database: 114,708
# of database sequences with at least one HSP: 208
No. of states in DFA: 462 (46 KB)
Total size of DFA: 70 KB (128 KB)
Time to generate neighborhood: 0.01u 0.00s 0.01t Real: 00:00:00
No. of processors used: 12
Time to search database: 33.78u 3.34s 37.12t Real: 00:00:06
Total cpu time: 33.87u 3.39s 37.26t Real: 00:00:06

WARNINGS ISSUED: 2

92
Date: Wed, 13 Apr 94 08:17:54
From: glee (gregory lee)
To: blast@ncbi.nlm.nih.gov

PROGRAM blastp
DATALIB nr
BEGIN
>HSAg-5/6 frame +2
EFRISGIPFCTIKFIHNYKINVVLGFTINIMYNPAQQEGRHPLSQGPGTWPQGEGDQPLATDEAGFQAQSFQD