The Cloning and Sequencing of Canine Granulocyte/Macrophage Growth Factor

by

GERALD F. FARNELL

A THESIS

Presented to the Robert D. Clark
Honors College of the University of Oregon
in partial fulfillment of the requirments
for the degree of
Bachelor of Arts, Honors College

June 1990

APPROVED: [redacted]

Dr. Richard A. Nash M.D., Adviser
Fred Hutchinson Cancer Research Center
Seattle, Washington

ABSTRACT

Granulocyte/macrophage colony-stimulating factor (GM-CSF) is an essential hematopoietic growth factor for the differentiation and proliferation of granulocytes and macrophages. In the work reported here, a human cDNA clone was used to isolate canine GM-CSF clones from a cDNA library derived from canine peripheral blood mononuclear cells and spleen cells. A genomic library was necessary to obtain the complete sequence. A full-length cDNA of the coding region was produced by polymerase chain reaction (PCR) leading to a clone of approximately 450 base pairs long. Canine GM-CSF showed a significant homology to the human GM-CSF both at the nucleotide (80%) and amino acid (70%) levels throughout the coding regions. It will now be possible to characterize further the activities of this factor in a pre-clinical animal model after expression of the protein.

ACKNOWLEDGEMENTS

I wish to extend my deepest thanks to Dr. Richard Nash for his support throughout the preparation of this manuscript and my time at the Fred Hutchinson Cancer Research Center in Seattle. Also, I would like to thank Dr. Ulla Sarmiento and Dr. Rainer Storb for their valuable help and guidance during my stay at the "Hutch." I also wish to thank Dr. Aaron Novick who originally helped me get in contact with the researchers in Seattle and who also stepped in late in the project as my third committee member. In addition, the help of Professor Dennis Todd in the actual writing of this manuscript was greatly appreciated. And finally, I want to thank my parents and my brother Mike, for allowing me to live and work in Seattle, for without whose help none of this would have been possible.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	BACKGROUND	2
III.	MATERIALS AND METHODS	8
	RNA Preparation	10
	cDNA Library	11
	Genomic Library	16
	Screening	16
	Preparation of DNA for Sequencing	19
	Sequencing	19
	Polymerase Chain Reaction (PCR)	20
IV.	RESULTS	23
V.	DISCUSSION	26
VI.	REFERENCES	29

LIST OF FIGURES

1. Compartmentalization of Hematopoiesis..... 3

2. mRNA Preparation from Canine Lymph Nodes..... 9

3. cDNA Synthesis..... 12

4. Ligation of cDNA Inserts into Lambda Zap II Arms..... 14

5. Recombinant Phage DNA Packaging and Amplification..... 15

6. Screening Procedure for both Libraries..... 17

7. Polymerase Chain Reaction (PCR)..... 21

8. Splicing of Genomic Exon I to cDNA clone by using PCR..... 24

9. Nucleotide Sequence of Canine and Human GM-CSF..... 25

10. Amino Acid Sequences of Dog and Human GM-CSF..... 25

Introduction

The research discussed here is concerned with acquiring the DNA sequence of canine granulocyte/macrophage-colony stimulating factor (GM-CSF) for the purpose of producing the protein in large quantities. The protein will then be used in conjunction with bone marrow transplantion in a preclinical model, the dog, to test the factor's effects. The importance of this research stems from its possible applications in cancer treatment, as well as its use in *in vivo* experiments concerned with broadening our knowledge of how the blood regulates its production.

Bone marrow transplantation is used in treating aplastic anemia and leukemia and is currently being tested to determine if it would improve results in other malignant diseases. A major complication of this procedure is the period of time when the patient is nearly defenseless against infection, leading to serious illness and death. It is thought that this period of vulnerability could be shortened by the use GM-CSF, thereby increasing the likelihood of a successful recovery. For these new clinical protocols to be used, they must first be tested in an appropriate animal model.

Background

Hematopoietic growth factors (Gr. *haima*, blood + *poiesis*, formation) variously characterized as colony-stimulating factors (CSF's) and interleukins, compose a family of proteins with biological specificity defined by their ability to support proliferation and differentiation of hematopoietic cells, more commonly known as blood cells. Lymphocytes, and other blood cells in the marrow, release these factors when activated and these factors seem to play a major role in the regulation of hematopoiesis.

Hematopoiesis is the process by which blood cells proliferate and differentiate. The key cells in sustaining hematopoiesis are the primitive hematopoietic stem cells. The arrangement of blood-forming tissues requires that stem cells be able both to self-generate and to generate differentiating progeny. These cells constitute an exceedingly small fraction (approximately 0.2 percent) of the total hematopoietic population even during active repopulation of hematopoietic cells.[1]

Hematopoietic tissues can be figuratively represented as a series of functional compartments of increasing size, with individual cells in each compartment being able to generate variable numbers of progeny entering the next developmental compartment.[2](see figure 1) Thus stem cells generate progenitor cells, which in turn generate morphologically identifiable progeny that progressively

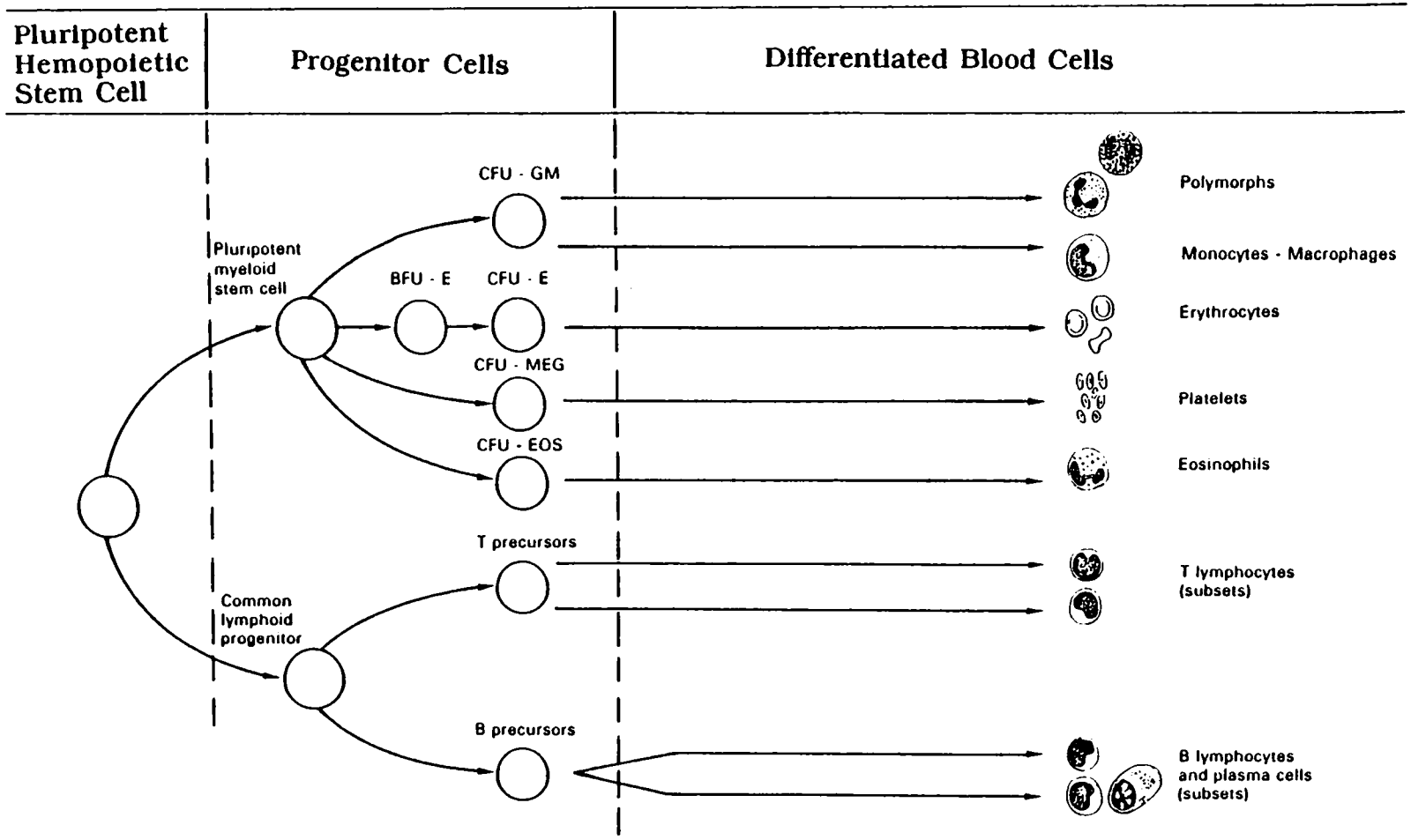

Figure 1: Compartmentalization of Hematopoiesis. From: Atlas of Blood Cells: Function and Pathology, by D. Zuckner-Franklin et al., Lea and Febiger, Philadelphia, PA, 1981, p. 16.

mature and ultimately lose their capacity for further proliferation, becoming the mature cells entering the blood.

Progenitor cells are committed to differentiate along a specific pathway of hematopoiesis and go on to become the many distinct lineages of mature cells found in the peripheral blood and other tissues: granulocytic (polymorphic), monocytic, erythroid (red cells), megakaryocytic (platelets), eosonophilic, T-lymphoid and B-lymphoid.(see figure 1)

In the early 1970's, researchers were attempting to identify the factors specific for the granulocyte-macrophage lineage. As different types of tissues were used for the purification of these regulators different forms of CSF were characterized.[3] By the 1980's, the improvement in protein chemistry technology had allowed the identification of four distinct types of CSF: M-CSF, G-CSF, GM-CSF and Multi-CSF or IL-3.[4] The various CSF's were identified by a prefix indicating the major target cell originally thought to be stimulated by low concentrations of the CSF. Therefore, the first two were originally shown to stimulate macrophage and granulocyte populations respectively, while GM-CSF stimulated both. Multi-CSF, on the other hand, had been shown to stimulate the proliferation not only of granulocytic and macrophage cells, but also of eosoniphils, erythroid, megakaryocytic and stem cells.[5] However, with more recent studies, the activity of these factors seems not

to be as confined to the certain cell types as was previously thought.

The multipotential hematopoietin GM-CSF has been shown to be necessary for GM-colony formation in vitro by the fact that the removal of intrinsic sources of the factor in culture blocks GM-colony formation.[6] Transfer of developing colonies in cultures with GM-CSF to cultures lacking GM-CSF also causes immediate cessation of colony growth followed rapidly by the death of colony cells.[7] It can then be concluded that GM-CSF is not simply a trigger for the start of GM-colony formation but is required continuously throughout colony growth. In addition, there was evidence that in the presence of high concentration of GM-CSF, most GM-progenitors became granulocytes and conversely, at low concentrations of GM-CSF, the progenitors became macrophages. This correlation between GM-CSF concentration and colony morphology led to the conclusion that the concentration of GM-CSF determines which of the two alternative pathways of differentiation colony cells enter as the colony increases in size. GM-CSF was seen as operating a switch mechanism affecting cellular differentiation according to the factor concentration.[8] It has also been shown that RNA synthesis in mouse bone marrow cells is increased with application of the factor. Autoradiographic analysis of RNA synthesis in cells preincubated with GM-CSF revealed the greatest increase in RNA synthesis was found in already differentiating GM-

cells.[9] This finding, in addition to others, suggests the possiblity that GM-CSF may be able to induce heightened functional activity in preformed granulocytes and macrophages.

GM-CSF has possible applications in bone marrow transplantion because of the factor's ability to cause GM-cells to proliferate, and thus increase the body's defense against infection. Bone marrow transplants are performed to cure benign hematopoietic diseases, such as aplastic anemia, as well as to treat malignant diseases such as leukemia. Leukemia treatment involves high doses of chemotherapy and radiation, but when these doses are high enough to maximize the chances for a cure, the normal marrow is destroyed. Immediately after the marrow transplantation, a patient lacks granulocytes, one of the types of white blood cells that engulf invading bacteria, and is therefore very susceptible to infection. This condition persists until the new bone marrow engrafts and slowly replenishes the pool of anti-infection cells, the granulocytes, monocytes and lymphocytes.

The animal model used in this research, the dog, has been an important model in the advancement of bone marrow transplantation. Many of the clinical protocols used today are based on studies previously done in this animal model. Contributions include the development of pretransplant preparative regimens, the importance of histocompatibility matching, the phenomenon of graft resistance, the effect of

pretransplant transfusions, the prophylaxis and treatment of graft-versus-host disease (GVHD), and studies on the nature of graft-host tolerance and immunologic reconstitution.[10]

For more than two decades, the dog has served as a model for studying the principles and techniques of marrow grafting applicable to man. There many reasons for using this model. First, dogs have a major histocompatibility complex, called DLA, similar to man's HLA. Large canine families are readily available for genetic studies of histocompatibility antigens, and they provide matched littermate pairs simulating the HLA-identical human sibling pairs. An added attraction of the dog is the availability of animals with spontaneous malignant and non-malignant hematologic diseases resembling those encountered in man. Finally, the dogs are easier to deal with both logistically and financially in comparison with other biologically suitable donors.[11]

Due to the species specificity of some of these factors, including GM-CSF, it was important to have the canine product for use in the dog rather than the human or mouse products, since they may have shown little or no activity. The use of growth factors will allow important questions with respect to transplantation, reconstitution, GVHD and hematopoiesis to be reinvestigated in greater detail. In addition, the physiological effects of the different hematopoietic growth factors on white blood cells and the immune system can also be studied more closely.

Materials and Methods

To understand the following methods a general knowledge of DNA, cDNA and messenger RNA is necessary. DNA is the master blueprint of the cell and contains that inherited information needed for the cell to survive. Messenger RNA, or mRNA, is the molecule that carries information from the DNA to the protein factories of the cell, the ribosomes. It is simply a small copy of a portion of the DNA. It differs from DNA in that it is single rather than double-stranded. Messenger RNA is smaller than the gene from which it is derived because it has gone through an editing process that splices out introns, segments of the sequence that are not involved in the coding of the particular protein, leaving those segments, called exons, that correlate directly to the final amino acid sequence of the protein. Also characteristic of mRNA is a segment of adenine bases, referred to as a poly-A tail.

Complementary DNA, or cDNA, on the other hand, is simply a copy of the mRNA. Because of its different support structures it can be sequenced.

The process by which the DNA sequence for GM-CSF was learned started with the isolation of mRNA from lymphocytes, contained in the lymph nodes, and spleen cells. Complementary DNA was then produced from this mRNA and the cDNA corresponding to the GM-CSF gene was isolated using a cDNA library and DNA probes. A genomic library was used as

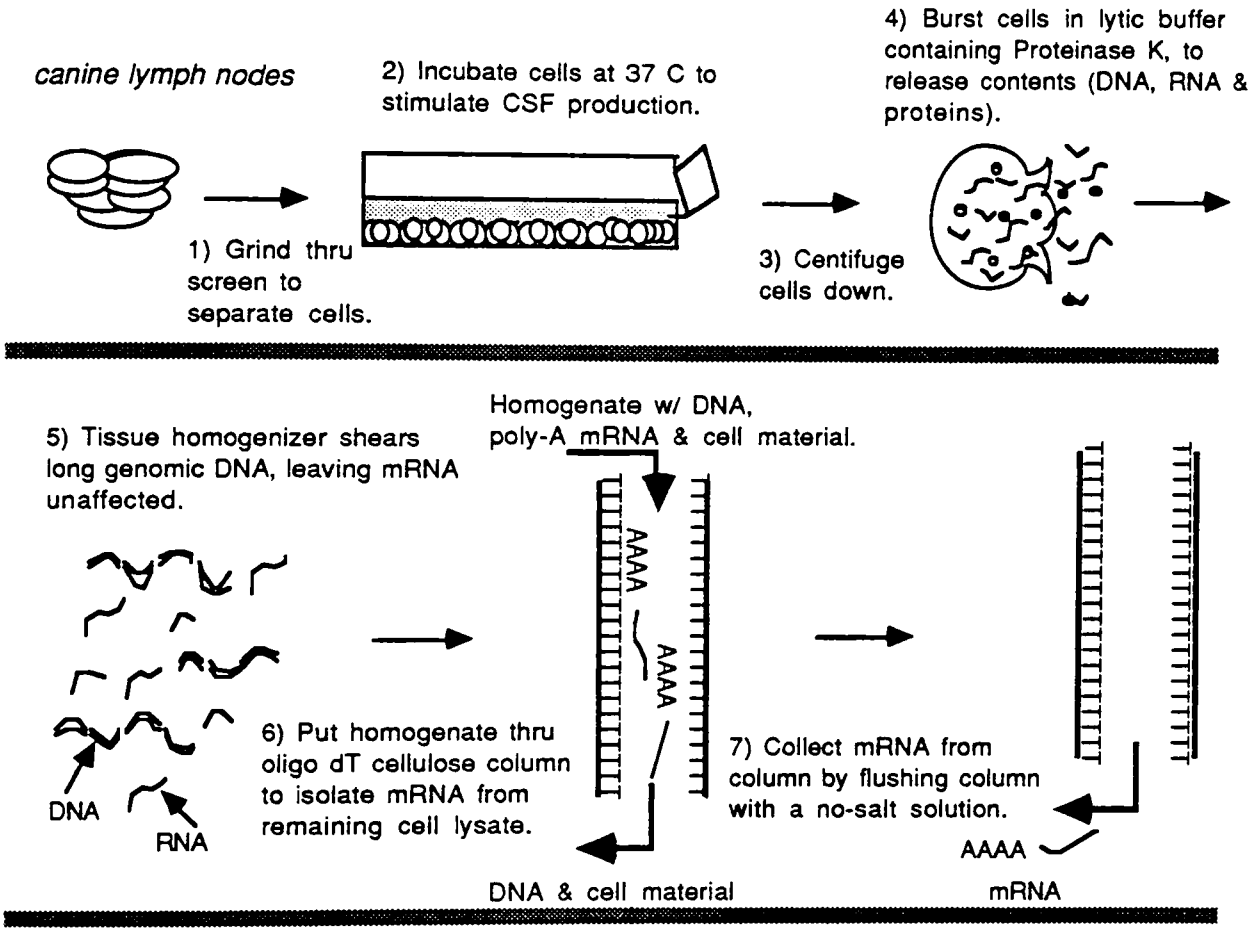

Figure 2:mRNA preparation from canine lymph nodes.

well to provide that portion of the sequence not contained in the cDNA clones. The screening with DNA probes allowed us to identify certain clones that were likely to contain the GM-CSF sequence and when used repeatedly, the screening procedure purified the clone as well. From this point, the cDNA clones were readied for sequencing by an *in vivo* excision and sequenced. The genomic clones were cut up into smaller fragments and reisolated. With this more manageable size the genomic clones were then sequenced. With the

entire sequence known, a full length copy was spliced together.

RNA Preparation

RNA was isolated from stimulated lymphocytes obtained from both mesenteric and non-mesenteric canine lymph nodes.

Nodes were harvested from a sacrificed animal from as many areas as possible and handled separately according to type of node.(see figure 2) In a petri dish with a small amount of medium the fat was trimmed from the nodes, the nodes were then cut into smaller pieces and finally ground through mesh screens to separate the cells. The cells were incubated for 3-4 days at 37 degrees centigrade in medium containing the following:

Waymouth's media
10% dog serum
Penicillin/Streptomycin
Essential amino acids
1% PHA\5% PWM

The PHA/PWM acts as a stimulator for RNA synthesis in T-cells, thus increasing the amount of mRNA for GM-CSF and other proteins. These PHA/PWM stimulated cells were then centrifuged and counted. The non-mesenteric lymph nodes and mesenteric nodes each yielded approximately $1\text{-}5\times10^{9}$ cells. (From this point on only RNAse-free glassware was used to reduce the possibility of RNA degradation.) These cells were then pelletted and re-suspended in a lytic buffer (150mM NaCl, 10mM Tris, 1% SDS) containing proteinase K.

This caused the cells to burst, emptying their contents into the surrounding medium.(The proteinase K reduced the possibility of RNAse activity in the culture, as well as eliminating other contaminating proteins in suspension.) The long genomic DNA strands were then sheared by the use of a tissue homogenizer, leaving the shorter messenger RNA strands unaffected. The homogenized tissue was incubated at 50 degrees centigrade for 1-2 hours and then frozen in the lytic buffer at -70 degrees centigrade.

This homogenate was then repeatedly put through a oligo-T cellulose column which selectively binds the poly-A mRNA in the presence of salt. The poly-A mRNA was then washed off the column with a no salt elution buffer. Sodium acetate and ethanol were added to the enriched mRNA to precipitate it out of solution and clean it up. The mRNA was then resuspended in water and stored at -70 degrees centigrade.[12]

cDNA Library

The cDNA library used in this research was produced commercially by Stratagene from the mRNA isolated on the oligo-dT column. The protocol for its production is outlined below with a brief description of the major steps.

The conversion of mRNA into cDNA for the purpose of cloning is a complex, interrelated series of enzyme-catalyzed reactions which lead to the production of a cDNA

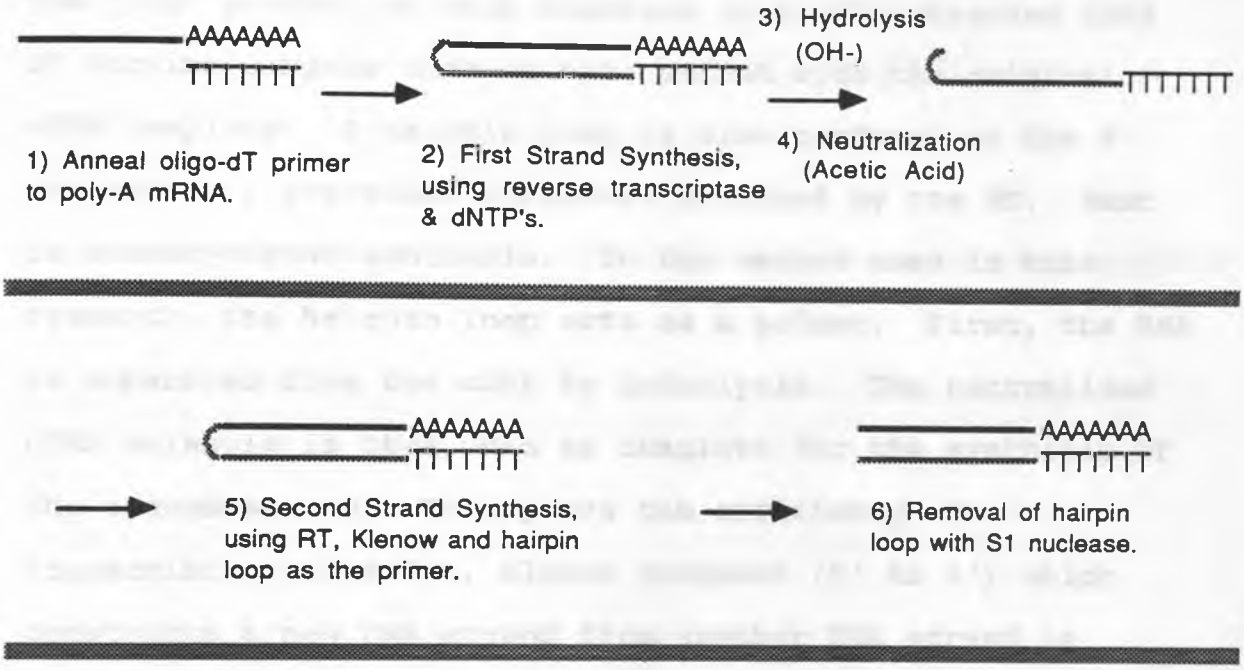

Figure 3: cDNA synthesis.

library. In this context, a library is defined as a group of recombinant DNA segments contained within virus for the purpose of replication. They are used so that a particular segment of DNA, or gene, can be duplicated and therefore more easily isolated for further study. There are many routes to a cDNA library. The first step in cDNA cloning is the synthesis of a DNA strand complementary to the mRNA sequence.(see figure 3) It begins with first-strand synthesis. This reaction requires template RNA, a complementary primer, reverse transcriptase (RT), and the nucleotides making up DNA (dNTP's). The RT is the enzyme that builds a new strand of DNA (5' to 3') from the information contained in the template strand of RNA. Oligo-dT is the most frequent choice as a primer because it can hybridize (bind) to the poly-A tails of template mRNA.

The final product of this reaction is single-stranded cDNA of varying lengths that is base paired with the original mRNA template. A hairpin loop is also produced at the 5' end and is a transient structure produced by the RT. Next is second-strand synthesis. In the method used in this research, the hairpin loop acts as a primer. First, the RNA is separated from the cDNA by hydrolysis. The neutralized cDNA molecule is then used as template for the synthesis of the second-strand. To improve the efficiency of transcribing naked DNA, Klenow fragment (5' to 3') which constructs a new DNA strand from another DNA strand is supplemented with reverse transcriptase. Apparently, the combination of enzymes is more effective for transcribing regions with secondary structure. With completion of second-strand synthesis, the hairpin loop is removed by the action of S1 nuclease which cleaves single stranded DNA. With the completed cDNA molecule, a method was selected for coupling the insert (the double-stranded cDNA) to the vector (phage DNA).(see figure 4) This was done by first blunting of the cDNA ends with Klenow fragment, methylating the Eco RI sites within the cDNA and subsequently, the covalent attachment of symmetric, double-stranded oligomers (linkers) to the ends of the cDNA, which contain a recognition site for the Eco RI restriction endonuclease. Linkers are used in high molar excess relative to inserts to help drive the ligation reaction of insert with linker. The high concentration of linkers competitively inhibits self-

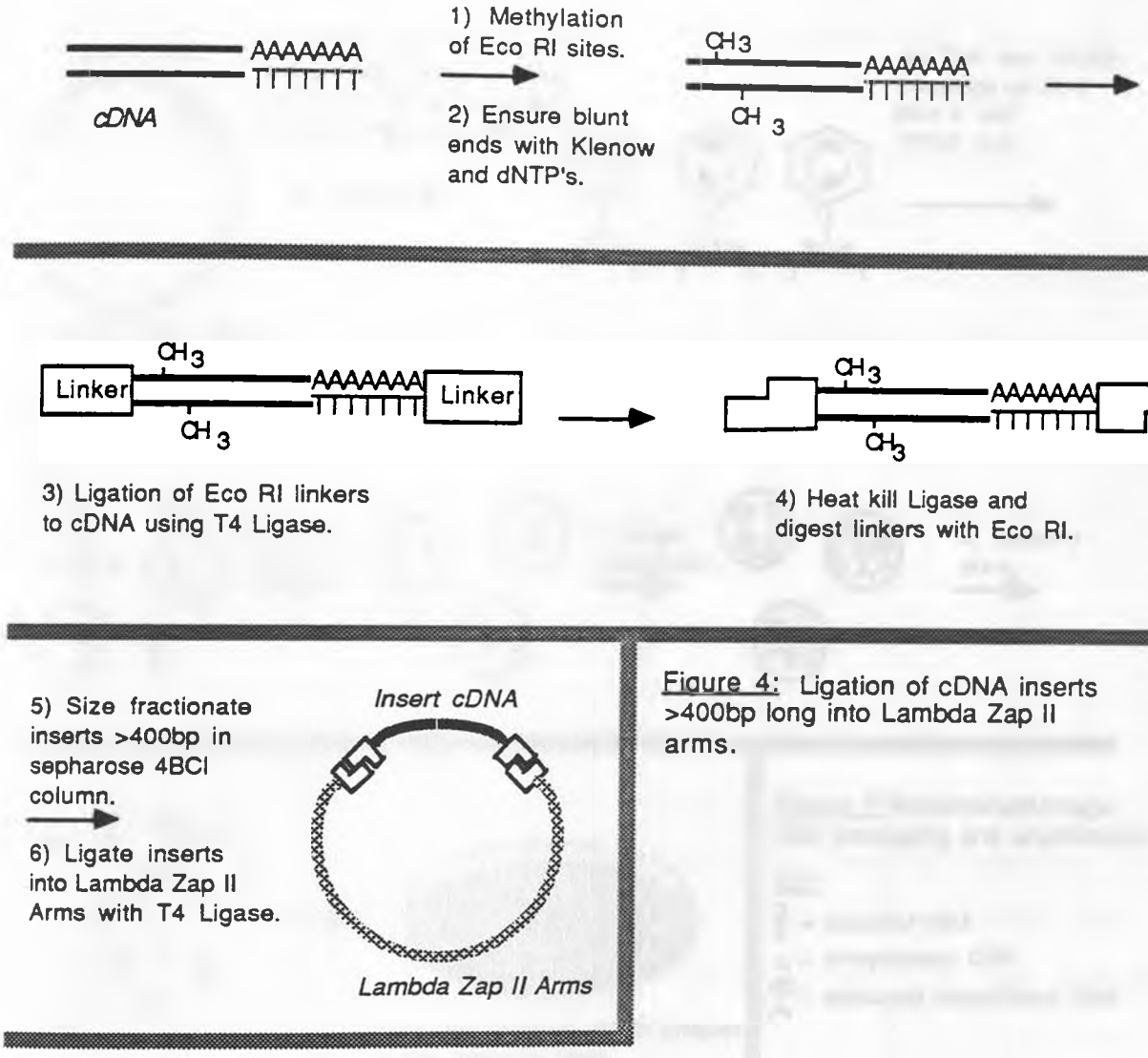

Figure 4: Ligation of cDNA inserts >400bp long into Lambda Zap II arms.

ligation of inserts. Restriction endonucleases are enzymes that make double strand cuts at specific DNA sequences. The specific enzyme used in the production of this library was EcoRI, which recognizes the following sequence:

G(A A T T C
C T T A A)G

(The asymetric cut made by this enzyme is shown by the brackets and line.) Eco RI is used to cut the linkers attached to the cDNA to generate the staggered termini

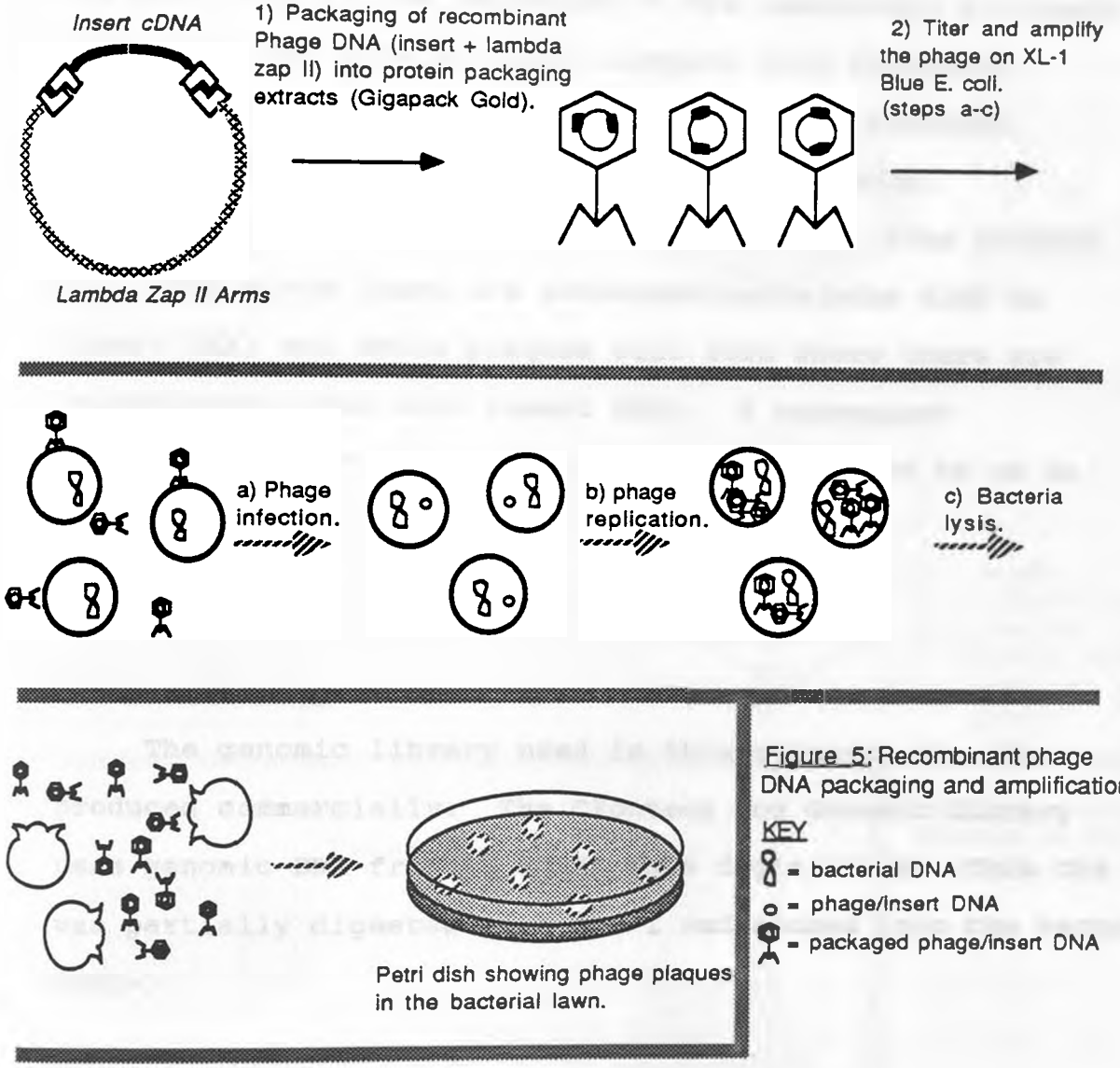

Figure 5: Recombinant phage DNA packaging and amplification.

capable of being ligated to a compatible vector. To protect the insert DNA from the activity of this enzyme, the double-stranded cDNA is methylated. Methylation prevents specific sequences within the insert from being cleaved as well. The Lambda-Zap II arms (Stratagene) are then digested with Eco-RI and the insert and arms (vector) are ligated with T4 DNA ligase.

The next step is the packaging of the Lambda-Zap II/insert DNA. This was done by using Gigapack Gold Packaging extracts (Stratagene) and then titering the packaged recombinant DNA on XL-1 Blue (E. coli) bacteria. (see figure 5) Upon plating these bacteria, blue plaques will form where there are nonrecombinants(arms with no insert DNA) and white plaques will form where there are recombinants(arms with insert DNA). A subsequent amplification of this library was done and sent to us in Seattle.

Genomic Library

The genomic library used in this research was also produced commercially. The Clontech Dog Genomic Library uses genomic DNA from an adult male dog's liver. This DNA was partially digested with Mbo I and cloned into the vector EMBL-3.

Screening

We used nucleic acid probes to screen first the cDNA library and later the genomic library. Originally a probe produced by PCR of the human GM-CSF cDNA was used. Once the first clones were sequenced, new canine cDNA probes were used. To locate the desired clone from the library the following steps were carried out: (see figure 6) 1) Phage

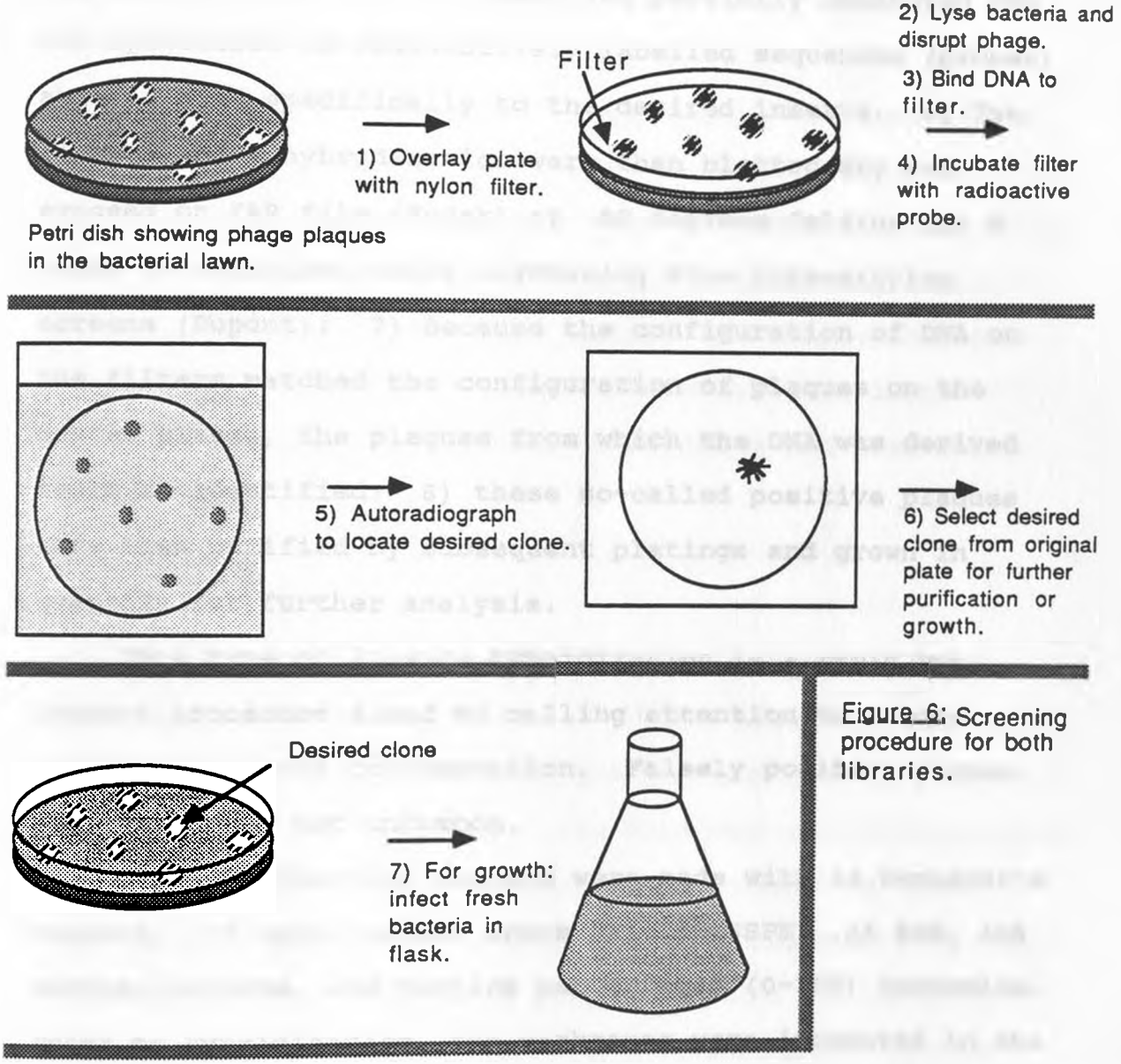

Figure 6: Screening procedure for both libraries.

infected bacteria were grown on plates of NZY agar (C600 bacteria with the cDNA clones and LE392 bacteria with the genomic clones); 2) Nylon membranes, called filters, (Hybond by Amersham) were then laid over the plates and further processed; 3) Phage and bacteria were disrupted and lysed *in situ* on the filters; 4) the DNA was bound to the

filter while the RNA was hydrolyzed, in .5M NaOH/1.5M NaCl and neutralized; 5) the resulting partially denatured DNA was hybridized to radioactively labelled sequences (probes) able to bind specifically to the desired inserts. 6) The filters after hybridization were then blotted dry and exposed on XAR film (Kodak) at -80 degrees Celsius for 6 hours to overnight using Lightening Plus Intensifying screens (Dupont); 7) Because the configuration of DNA on the filters matched the configuration of plaques on the master plates, the plaques from which the DNA was derived could be identified; 8) these so-called positive plaques were then purified by subsequent platings and grown in quantity for further analysis.

This type of *in situ* hybridization is a rapid but inexact procedure aimed at calling attention to clones worthy of serious consideration. Falsely positive clones were therefore not uncommon.

Our hybridization buffers were made with 1X Denhardt's reagent, 100 ug/ml salmon sperm DNA, 5X SSPE, .1% SDS, 10% dextran sulfate, and varying percentages (0-50%) formamide. Prior to hybridization, the membranes were incubated in the hybridization buffer for at least an hour. Hybridization took place at varied temperatures (room temp. to 65 degrees Celsius) and at varied lengths of time (2hrs. to overnight). Hybridization time, temperature and formamide concentration were varied depending on the stringency of conditions wanted.

Preparation of DNA for Sequencing

cDNA Library

The Lambda Zap II vector was designed to allow in vivo excision and recircularization of the cloned insert contained within the vector to form a plasmid containing the cloned insert. This procedure was done according to the protocol sent to us from Stratagene. The double stranded Bluescript/insert plasmid DNA was then grown in large quantities, using the protocol of Maniatis and then ready for sequencing.[13]

Genomic Library

The EMBL3 vector DNA was isolated using the procedure of Maniatis with the omission of second centrifugation after PEG precipitation and in its place the addition of SDS and Proteinase K. A 23 kb insert was then cut out of the vector using Sal I restriction nuclease and separated on a low melting point gel. This insert was then digested with Bam HI and Sal I and a 4.5 kb fragment positive for the cGM-CSF gene was subcloned into Bluescript. Once again this was grown in large quantities and ready for sequencing.[14]

Sequencing

The double stranded plasmid DNA was sequenced by using the S-35 dideoxy chain termination reaction kit, Sequenase-

version 2.0 (US Biochemical Corp.) and the protocols suggested by USBC. T3 and T7 primers were used so that sequencing could occur on both the sense and nonsense strands of DNA. The sequencing reactions were then run on 40% acrylamide gels, one with a 37.5 acrylamide:bis ratio (stretch gel-for sequence away from primer), the other gel with a 19:1 ratio (wedge gel-for sequence closer to primer). These gels were run at approximately 25W for 16 hours and 60W for 4-6 hours, respectively. The gels were then dried and exposed on XAR film for about 2-3 days.

Polymerase Chain Reaction - (PCR)

For the purposes of both DNA splicing and DNA amplification a process known as PCR was used. The method is based on the repetition of a set of three steps, denaturation, annealing of primers, and primer extension. All these steps are conducted in succession under somewhat different and controlled temperature conditions. In the PCR technique, this typical set of three steps is referred to as a cycle.(see figure 7)

Step 1- The double-stranded DNA sample referred to as the template is denatured by incubation at high temperature. The two strands, now dissociated, will remain free in solution until the temperature is lowered sufficiently to allow annealing.

Step 2- The extension primers are a pair of short pieces of DNA (oligonucleotide) that anneal to sites flanking the

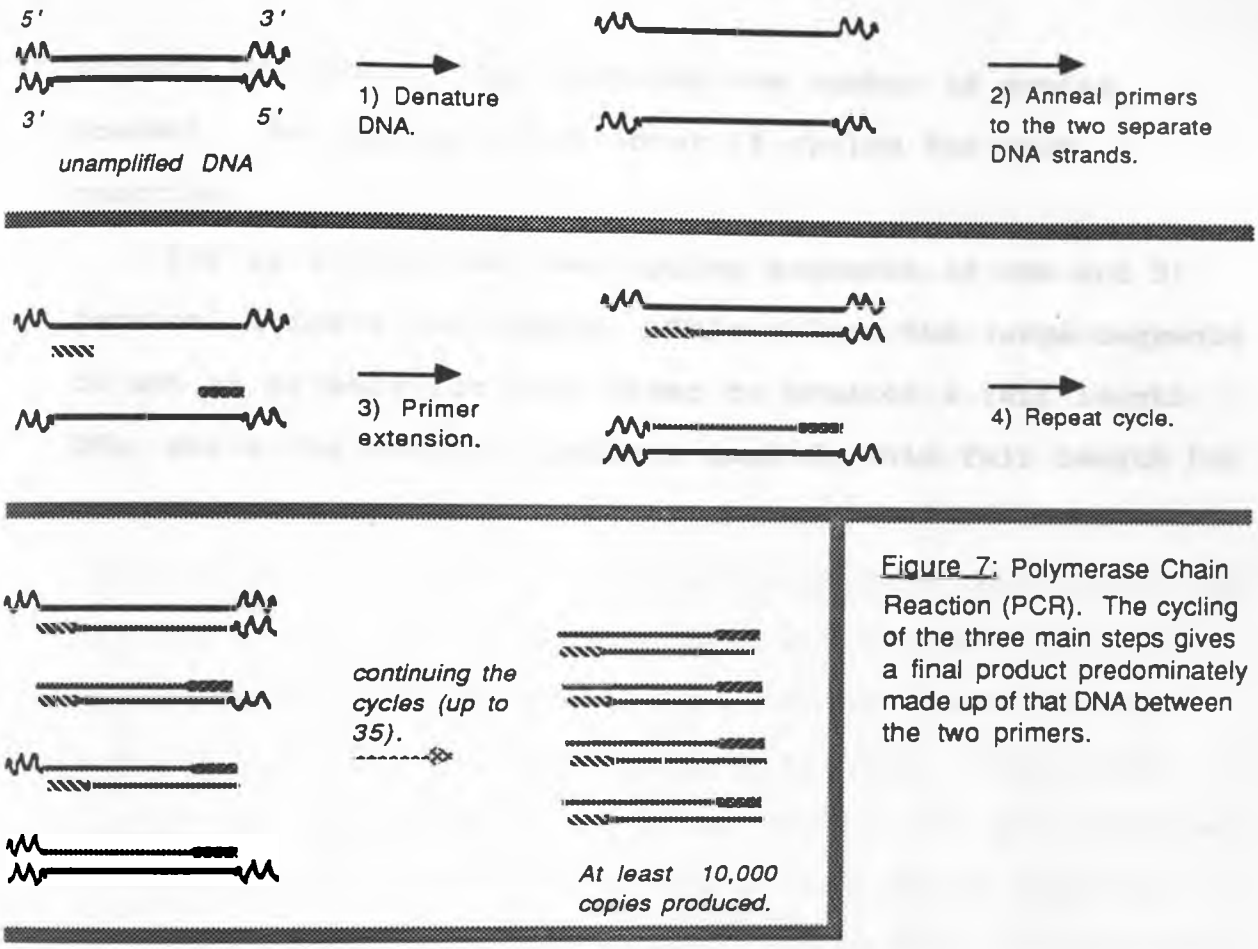

Figure 7: Polymerase Chain Reaction (PCR). The cycling of the three main steps gives a final product predominately made up of that DNA between the two primers.

region to be amplified. Each primer in the pair will anneal to only one of the strands of DNA. Because the primers are present in large excess over the DNA template, the formation of the primer-template complex will be favored over the reassociation of the two DNA strands when the temperature is lowered.

Step 3- The DNA Polymerase mediated (5' to 3') extension of the primer-template complex is the final step in the cycle. Due to the high temperatures involved in the denaturing step, the temperature-sensitive Klenow fragment is replaced with a more stable DNA Polymerase, *Taq* DNA Polymerase. With each cycle the DNA contained between the

primers is copied, thus doubling the number of copies present. We typically ran about 35 cycles for each reaction.

For splicing, two overlapping segments of DNA and 5' terminal primers are needed. This allows the large segments to act as primers for each other to produce a full length DNA, while the small 5' primers amplify this full length DNA in subsequent cycles.

Results

The purpose of this research was to isolate and determine the DNA sequence of canine GM-CSF. It was accomplished by the use of both a canine genomic library and a cDNA library constructed from polyadenylated mRNA isolated from canine lymph nodes. Originally, a human GM-CSF DNA probe was used in the screening process. From this a 727bp (base pair) clone was isolated and sequenced which contained the poly A tail, but was missing around 100 base pairs of the expected 5'end. A 250bp probe was produced from this clone by PCR with sequence specific primers. Other cDNA clones were isolated with this new probe, yet none contained any additional information on the 5'-end of the sequence.

Therefore, we began to search the genomic library for a positive clone. Using the same 250bp probe, a clone containing a 23kb (kilo-base) genomic fragment, G-21, was isolated. This fragment was then digested with numerous restriction enzymes, run on an agarose gel, and analyzed by the method of Southern.[15] A positive 4.5kb (Bam HI + Sal I) fragment was then isolated and subcloned into pBluescript. The portion of this fragment containing exon I of the GM-CSF gene was sequenced, and since the original cDNA clone extended 80bp into exon I, this was all that was needed.

Exon I was then amplified by PCR and was found to be 160bp long. The original cDNA clone and the exon I fragment were then mixed together with two flanking 20bp

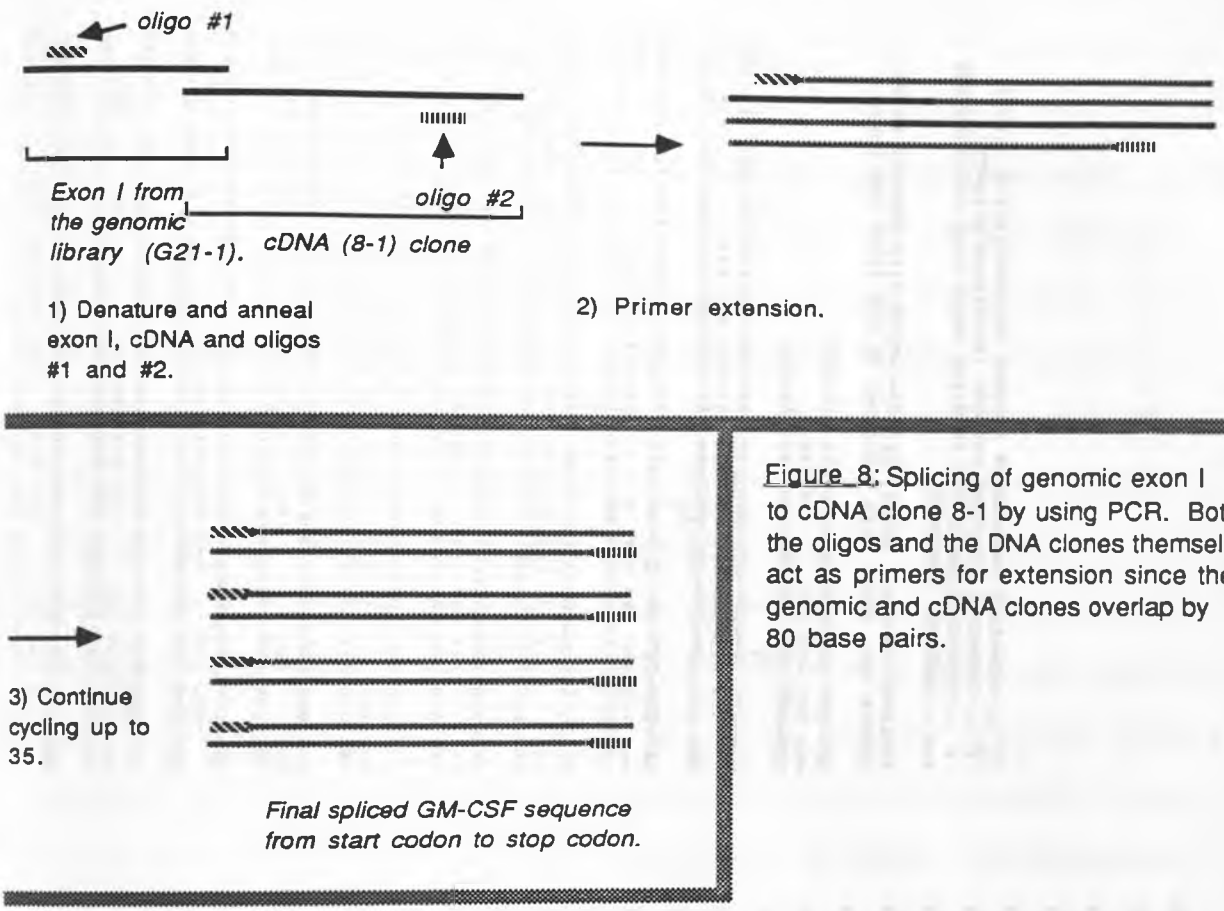

Figure 8: Splicing of genomic exon I to cDNA clone 8-1 by using PCR. Both the oligos and the DNA clones themselves act as primers for extension since the genomic and cDNA clones overlap by 80 base pairs.

oligonucleotides and put through PCR cycles to produce a full length cDNA clone for canine GM-CSF.(see figure 8) This full length clone was then cloned into pBluescript(SK-) II and sequenced.

The complete cDNA sequence for canine GM-CSF is 809bp's long and is shown in figure 9, with the complete amino acid sequence, as predicted from the cDNA sequence, shown in figure 10. The 5' end of the cDNA encodes a region of 17 amino acids with a potential initiator methionine and many of the characteristics expected of a hydrophobic signal peptide for a secreted protein.[16]

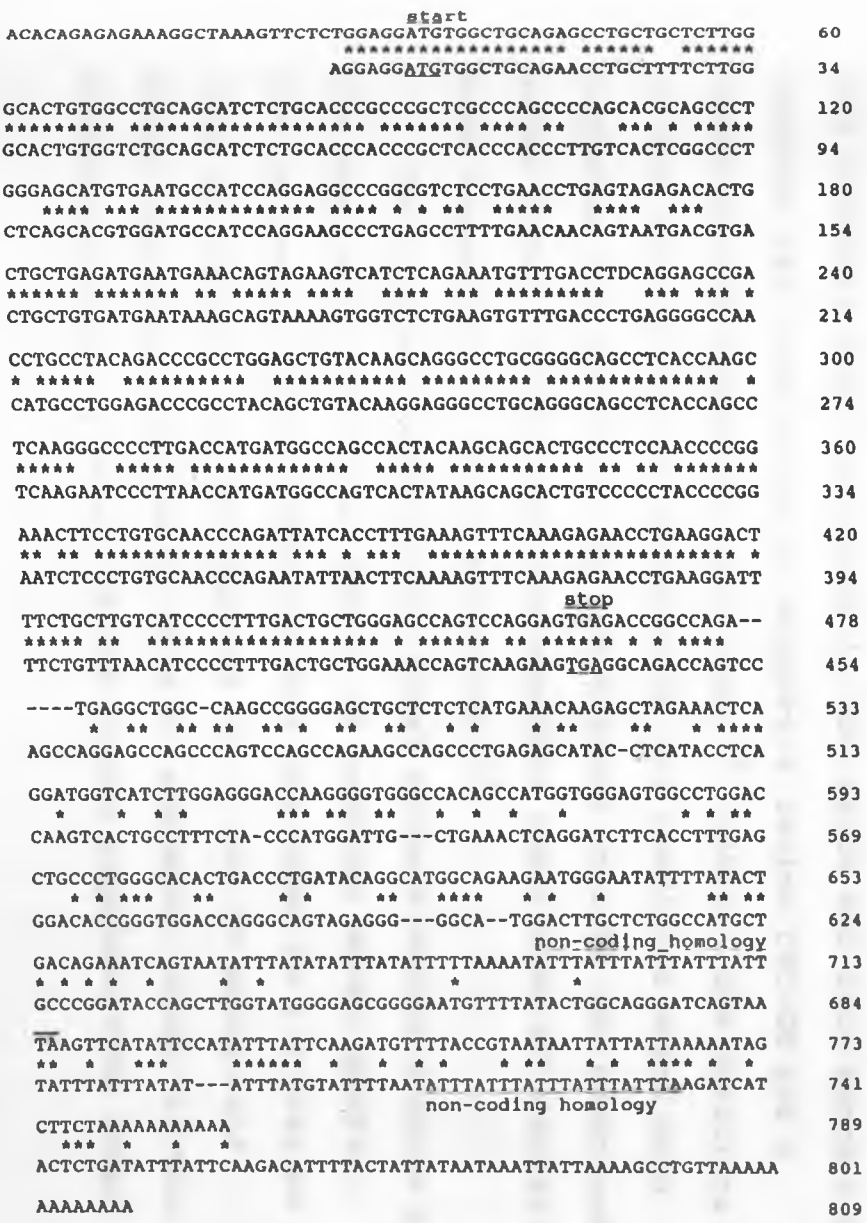

Figure 9: Nucleotide sequence of canine (bottom) and human (top) GM-CSF. Highlights of the sequence are noted.

Signal Peptide(up to Ser-17)

```
Met-Trp-Leu-Gln-Asn-Leu-Leu-Phe-Leu-Gly-Thr-Val-Val-Cys-Ser      15
*** *** *** ***     *** ***     *** *** *** ***     *** ***
Met-Trp-Leu-Gln-Ser-Leu-Leu-Leu-Leu-Gly-Thr-Val-Ala-Cys-Ser      15

Ile-Ser)Ala-Pro-Thr-Arg-Ser-Pro-Thr-Leu-Val-Thr-Arg-Pro-Ser      30
*** ***)*** ***     *** *** ***  *          ***     ***
Ile-Ser)Ala-Pro-Ala-Arg-Ser-Pro-Ser-Pro-Ser-Thr-Gln-Pro-Trp      30

Gln-His-Val-Asp-Ala-Ile-Gln-Glu-Ala-Leu-Ser-Leu-Leu-Asn-Asn      45
    *** ***     *** *** *** *** ***          *** *** ***
Glu-His-Val-Asn-Ala-Ile-Gln-Glu-Ala-Arg-Arg-Leu-Leu-Asn-Leu      45

Ser-Asn-Asp-Val-Thr-Ala-Val-Met-Asn-Lys-Ala-Val-Lys-Val-Val      60
***     ***     *** *** ***          ***     *** *
Ser-Arg-Asp-Thr-Ala-Ala-Glu-Met-Asn-Glu-Thr-Val-Glu-Val-Ile      60

Ser-Glu-Val-Phe-Asp-Pro-Glu-Gly-Pro-Thr-Cys-Leu-Glu-Thr-Arg      75
*** ***     *** ***             *** *** *** ***     *** ***
Ser-Glu-Met-Phe-Asp-Leu-Gln-Glu-Pro-Thr-Cys-Leu-Gln-Thr-Arg      75

Leu-Gln-Leu-Tyr-Lys-Glu-Gly-Leu-Gln-Gly-Ser-Leu-Thr-Ser-Leu      90
***     *** *** ***     *** ***     *** *** *** ***     ***
Leu-Glu-Leu-Tyr-Lys-Gln-Gly-Leu-Arg-Gly-Ser-Leu-Thr-Lys-Leu      90

Lys-Asn-Pro-Leu-Thr-Met-Met-Ala-Asn-His-Tyr-Lys-Gln-His-Cys     105
***     *** *** *** *** *** ***     *** *** *** *** *** ***
Lys-Gly-Pro-Leu-Thr-Met-Met-Ala-Ser-His-Tyr-Lys-Gln-His-Cys     105

Pro-Pro-Thr-Pro-Glu-Ser-Pro-Cys-Ala-Thr-Gln-Asn-Ile-Asn-Phe     120
*** *** *** *** ***  *      *** *** *** ***     ***     ***
Pro-Pro-Thr-Pro-Glu-Thr-Ser-Cys-Ala-Thr-Gln-Ile-Ile-Thr-Phe     120

Lys-Ser-Phe-Lys-Glu-Asn-Leu-Lys-Asp-Phe-Leu-Phe-Asn-Ile-Pro     135
    *** *** *** *** *** *** *** *** *** ***          *** ***
Glu-Ser-Phe-Lys-Glu-Asn-Leu-Lys-Asp-Phe-Leu-Leu-Val-Ile-Pro     135

Phe-Asp-Cys-Trp-Lys-Pro-Val-Lys-Lys                             144
*** *** *** ***     *** ***
Phe-Asp-Cys-Trp-Glu-Pro-Val-Gln-Glu                             144
```

Figure 10: Amino acid sequences of dog (top) and human (bottom) GM-CSF. HIghlights are noted with symbols below.

*** = complete homology
* = conservative change (similar amino acid replacement)
ⓒ = conserved cysteine residue (disulfide bridges)
ⓝ = conserved asparagine residue (glycosylation sites)

Discussion

In this study both cDNA and genomic libraries were used to determine the complete cDNA sequence for canine GM-CSF. There is evidence to support the assumption that this sequence corresponds to the canine GM-CSF gene. First, there is a high degree of homology between the known human sequence and the sequence we have isolated, both at the nucleotide level and the amino acid level. The cDNA sequence for GM-CSF is compared to the human sequence and shown in figure 9. Overall, there is a 61% homology between the two species, with greater homology in the protein coding region. The amino acid sequences of the two species, shown in figure 10, reveal a 70% homology. Second, the sequence has recently been expressed in *cos* cells and this expressed protein has been shown to support a canine CFU-GM and a human GM-CSF dependent cell line.[17] Lastly, the positions of the four cysteine residues in the predicted amino acid sequence are identical to those found in both the human and mouse forms, indicating the possible importance of disulfide bond formation in these molecules.[18]

Two regions outside the final protein coding regions of the mRNA also show some degree of homology among species. One is the 5' signal sequence and the other is a 3' region rich in thymine (T) and adenine (A). The 17 amino acid signal peptide, encoded by this 5' sequence, shows all the characteristics of other GM-CSF signal peptides from other

species and is involved in the secretion of the protein that produces it. The existence of a T\A rich sequence in the 3' non-coding region of our isolated cDNA agrees with that found in both the human and mouse forms. These long stretches of 40 to 60 nucleotides consisting of T's and A's, beginning roughly 690 bp after the start codon in the dog and 630 bp after the start codon in the human, are found in other secreted factors, notably the human and mouse forms of interferon (alpha)2.[19] This sequence is important in destabilizing the GM-CSF mRNA so that the protein will not be continuously produced in the cell.

With the sequence of canine GM-CSF known, the protein is now being produced by Amgen Inc., using a bacterial expression vector. This recombinant form of the protein will not be identical to that found naturally in the dog. Rather it will be lacking the substantial carbohydrates (glycosylation sites) usually associated with asparagine residues (one of the amino acids). This is caused by the inability of the bacteria to attach carbohydrates to the proteins they produce. The specific role of glycosylation in proteins is unclear; sometimes it is involved in the activity of the protein or its binding, while at other times its purpose is unknown. Studies in which the carbohydrate was enzymatically stripped from the CSF show that it is neither involved in the molecule's *in vivo* activity nor in its receptor binding.[20] These studies allow the assumption to be made that the non-glycosylated form of GM-CSF, which

will be used in the later *in vivo* studies, will exhibit levels of activity similar to those found in the native protein.

Presently, GM-CSF is being administered to many patients as an experimental post-transplant regimen to test whether it is able to shorten the period of time needed for hematopoietic reconstitution. The hope is that GM-CSF will indeed shorten this period and thereby allow more intensive treatments while lowering the patient's risk of infection. In addition, the use of the factors in the model systems will aid in our understanding of the stem cell's role in hematopoietic reconstitution.

The need for information about the primitive stem cell is long overdue, yet the only way to study this particular cell is through bone marrow transplant. There are no *in vitro* assays for the stem cell, and this makes the canine model that much more important. Therefore, the recombinant form of GM-CSF in conjunction with other drugs will be used to maximize the stem cell's ability to reconstitute the hematopoietic populations after bone marrow transplant in the dog, in the hopes that those findings will improve the current management of clinical bone marrow transplantation.

References

1. Metcalf, Donald, The Molecular Control of Blood Cells, Cambridge, MA: Harvard Press, 1988. p.13.

2. ibid., p.7.

3. ibid., p.34

4. Moore, Malcolm A.S., "The use of hematopoietic growth and differentiation factors for bone marrow stimulation," Important Advances in Oncology '88, Vincent T. DeVita Jr.(ed.), Philadelphia, PA: J.P. Lippincott Co., 1988. p. 31.

5. ibid., p.32.

6. Moore, M.A.S., Williams, N. 1972. J. Cell. Physiol. 80: 195-206.
   Moore, M.A.S., Williams N., Metcalf, D. 1973. J. Nat. Cancer Inst. 50: 591-602.
   Messner, H.A., Till, J.E., McCulloch, E.A. 1973. Blood 42: 701-710.

7. Metcalf, D., Foster, R. 1967. Proc. Soc. Exp. Biol (NY) 126: 758-762.
   Poron, M., Sachs, L. 1968. J Cell Physiol. 72: 247-250.

8. Metcalf, D., "Regulation by colony-stimulating factor of granulocyte and macrophage colony formation *in vitro* by normal and leukemic cells," Control of Proliferation in Animal Cells, R. Baserga (ed.), New York: Cold Spring Harbor Laboratory, 1974. pp.887-905.
   Moore, M.A.S., Spitzer, G. Metcalf, D., Penington, D.G. 1974. Brit. J. Haemat. 27: 47-55.

9. Burgess, A.W., Metcalf, D. 1977. J. Cell. Physiol. 90: 471-484.

10. Storb, Rainer, Deeg, H. Joachim. 1985. Plasma Ther. Transfus. Technol. 6: 303-312.

11. ibid., pp. 303-312.

12. Aviv, H., Leder, P. 1972. Proc. Nat. Acad. Sci. USA 69: 1408-1412.

13. Maniatis, T., Fritsch, E.F., Sombrook, J., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1982. pp. 371-372.

14. ibid., pp. 371-372.

15. Southern, E. 1975. J. Mol. Biol 98: 503.

16. Watson, M.E.E. 1984. Nucleic Acids Res. 12: 5145-5164.

17. Nash, Richard. Personal communication 4-20-90.

18. Cantrell, Michael A. et al. 1985. Proc. Natl. Acad. Sci. USA 82: 6250-6254.

19. Cantrell, Michael A. et al. 1985.
Lawn, R.M. et al. 1981. Proc. Natl. Acad Sci.USA 78: 5435.
Shaw, G.D. et al. 1983. Nucleic Acids Res. 11: 555.

20. Metcalf, D.(1988), p. 35.