GLYCOLYTIC GENES FROM ESCHERICHIA COLI AS EVOLUTIONARY PROBES

LIAN C. LYONS-LOOI

Glycolytic Genes from Escherichia Coli as Evolutionary Probes

By

© Lian C. Lyons-Looi B.Sc.(Hons.)

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry

Memorial University of Newfoundland

September 1986

St. John's Newfoundland

ISBN 0-315-45096-7

Abstract

Comparative studies based on the structure of ribosomes and their components suggested that there are three primary kingdoms: eubacteria, archaebacteria and urkaryotes. However, the validity of this hypothesis is being questioned. In order to resolve the evolutionary relationships among bacteria it is necessary to compare the primary structures of molecules other than those of the translational apparatus. The main objective of this project was to determine whether or not hybrid recombinant plasmids that contain sequences for glycolytic enzymes from *Escherichia coli* could be used to detect corresponding sequences in preparations of genomic DNA from distantly related organisms. Three plasmids from the Clarke-Carbon *E. coli* genomic library were chosen for this study. The plasmid pLC 16-4 contains the genes for triose-phosphate isomerase and phosphofructokinase, whereas it had been reported that pLC 10-47 and pLC 11-8 both contain the gene for enolase. Restriction endonuclease maps were constructed for these plasmids, and the location of the cloning vector was determined in each case by Southern blotting and probing with Col E1. Comparison of the restriction maps of pLC 10-47 and pLC 11-8 indicated that there were no sequences in common in the genomic inserts of these plasmids. Subsequent studies at the protein level showed that only pLC 10-47 contains a gene coding for enolase. Genomic DNA was prepared from representative species of the Enterobacteriaceae, other gram-negative bacteria, gram-positive bacteria, and archaebacteria. The genomic DNA was subjected to Southern blotting and probed with radioactively-labelled pLC10-47 or pLC 16-4. The results from these hybridization studies indicated that genes encoding glycolytic enzymes in *E. coli* were not able to cross-react with DNA from species that had diverged from *E. coli* more than 40 million years ago. Therefore, the genes for glycolytic enzymes will not be of use as long range evolutionary probes.

Acknowledgements

Any task could hardly be completed without the help and assistance of other people.

I would like to thank Dr. W.S. Davidson, my supervisor, for the help and supervison which he has so kindly provided throughout my research in his laboratory. His burst of enthusiasm, ideas and his willingness to discuss the project have made it possible for me to complete this thesis.

I would also like to thank the members of my supervisory committee, especially Dr. E. A. Barnsley for providing bacteria for this project and for allowing me to use the equipment in his laboratory, but most of all for his ability to read and get to the crux of my thesis so quickly and for providing valuable insights, suggestions and criticisms.

To people in Dr. Davidson's laboratory, I would like to express my appreciation for their friendship.

Last but not least to the School of Graduate Studies and NSERC, I would like to express my appreciation for providing funds for this project.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A_x	absorbance at x nm
cpm	counts per minute
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
eno	enolase gene
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
kbp	kilobase pairs
PAGE	polyacrylamide gel electrophoresis
PFK	phosphofructo-kinase protein
pfk	phosphofructo-kinase gene
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
TEMED	N,N,N',N'-tetramethylenediamine
TPI	triose-phosphate isomerase protein
tpi	triose-phosphate isomerase gene
UEP	unit evolutionary period
V	volts

Chapter 1

Introduction

1.1. The Origin of Life

The origin of life has long been a matter of speculation. It is unfortunate that one may never be able to ascertain the truth. Even so, continuous research using a variety of approaches may provide valuable information upon which intelligent speculation may be based. For instance, it has been well established that our solar system, including the Earth, originated as early as 4.6 billion years ago and that the history of the Earth may be delineated by an orderly sequence of fossils that is used to assign geological eras and periods (Figure 1-1; Futuyma, 1983).The first evidence of biochemical reactions and cellular activity occurs in the Precambrian era. The earliest Precambrian rocks are rich in unoxidized iron compounds that could not have formed in the presence of oxygen. This indicates that the atmosphere at that time lacked oxygen and ozone. Precambrian rocks dated 3.7 billion years old are somewhat different in that they contain peculiar layers of iron which almost certainly evince the presence of life (Futuyma, 1983). These rocks resemble deposits which are formed by iron-using bacteria. However, the oldest unquestionable traces of life are fossils of bacteria-like forms that were preserved in Swaziland (South Africa) sediments. Radioactive dating of these fossils indicates that they are between 3.2 to 3.5 billion years old (Dayhoff, 1978;

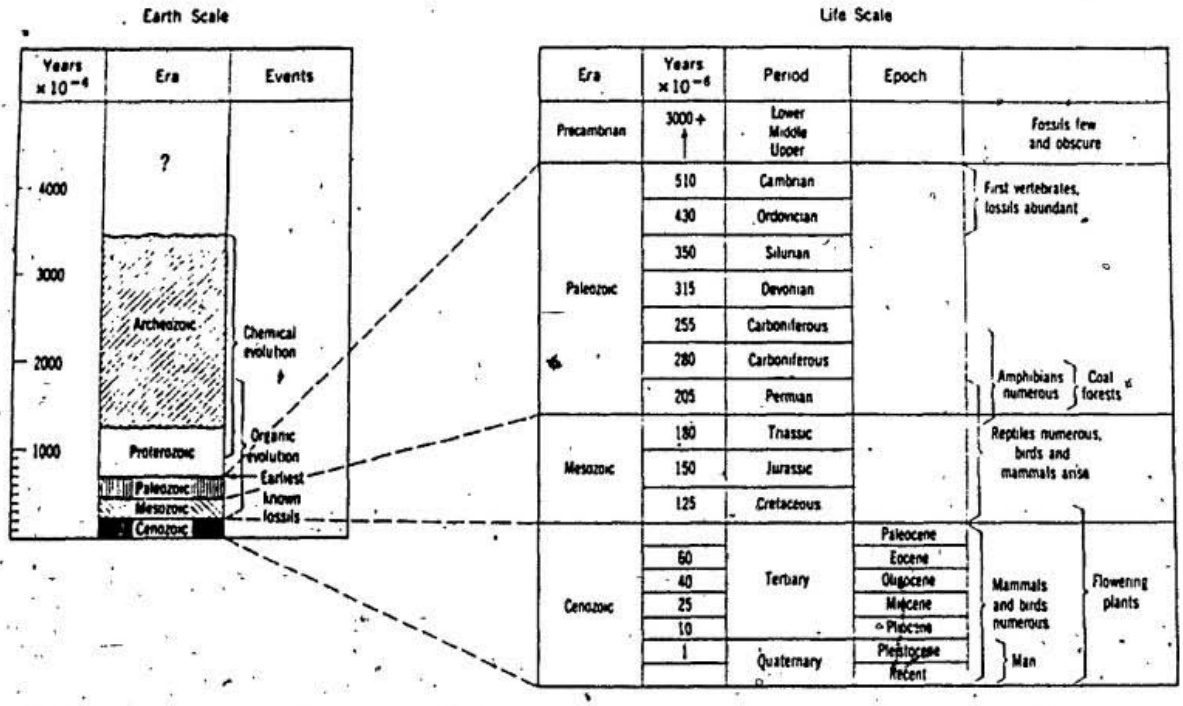


Figure 1-1: Geological Era and the Fossil Record

Source: Futuyma, 1983

Doolittle. 1980). These inhabitants of the earth were simple microscopic organisms. Many of them were comparable in size and complexity to modern bacteria although the conditions under which these organisms lived differred greatly from those prevailing today. Fossils formed one billion years later indicate the presence of blue-green algae (cyanobacteria) which were capable of photosynthesis and the production of oxygen. Rocks produced during this period and since then reveal an oxygen-rich atmosphere. Therefore, it appears that prokaryotic cells were the first life-forms on Earth and that they dominated most of the Precambrian era.

1.2. Implications of the Fossil Record

Defining the time of the appearance of the first eukaryotes from the fossil record is difficult. This is because the criteria by which the earliest eukaryotes differred from their prokaryotic counterparts are few and in many cases open to debate. Schopf (1978) has listed a series of criteria based on size, shape and morphological complexity that may be used to differentiate between eukaryotic and prokaryotic fossils. Structural characteristics of eukaryotes include branched filaments with internal cross walls, complex (e.g. flask-shaped) microfossils, large algal cysts, internal dense bodies resembling the residues of eukaryotic organelles, and tetrads of cells, possibly representing the products of meiosis. The presence of these traits have been used to mark the origin of eukaryotes. Unforturnately, some of these presumed eukaryotic structures are dubious since in experiments performed with cultures of living blue-green algae Knoll and Barghoorn (1975) showed that an organelle-like mass appeared in degenerating blue-green algal

cells. Oehler and co-workers (1976) have found that blue-green algae can form non-meiotic tetrads. These results cast doubt on the different dates that have been reported for the first appearance of the eukaryotic cells. Recently Vidal (1984) indicated that the origin of eukaryotes began 1.6 billion years ago in a form of green algae. It is reasonable then, to accept that a diversity of eukaryotic cells was present before 0.9 billion years ago, but animal fossils do not appear in profusion until the begining of the Cambrian period - 580 to 600 million years ago.

Based on comparative morphology, the fossil record seems to suggest that the prokaryotic genealogies are far more ancient than their eukaryotic counterparts. As a result it is generally accepted without other evidence, that eukaryotes evolved from prokaryotes. However, the questions of how and when this evolutionary event took place remain unanswered. The fossil evidence accumulated so far suggests that life is monophyletic: that is, all living organisms are descended from an ancient anaerobic heterotrophic prokaryote, similar to that of a modern *Clostridium* (Fox *et al.*, 1980). The first eukaryotic cell was probably a unicellular alga, derived from a prokaryotic blue-green alga, and it was from this simple eukaryotic cell that higher plants arose. Algae that lost their photosynthetic ability evolved to form organisms such as protists, fungi and animals (Figure 1-2; Doolittle, 1980; 1982). The blue-green bacteria are viewed as the intermediate between the prokaryotes and "lower" photosynthetic eukaryotes and thus represent a steady evolutionary progression.

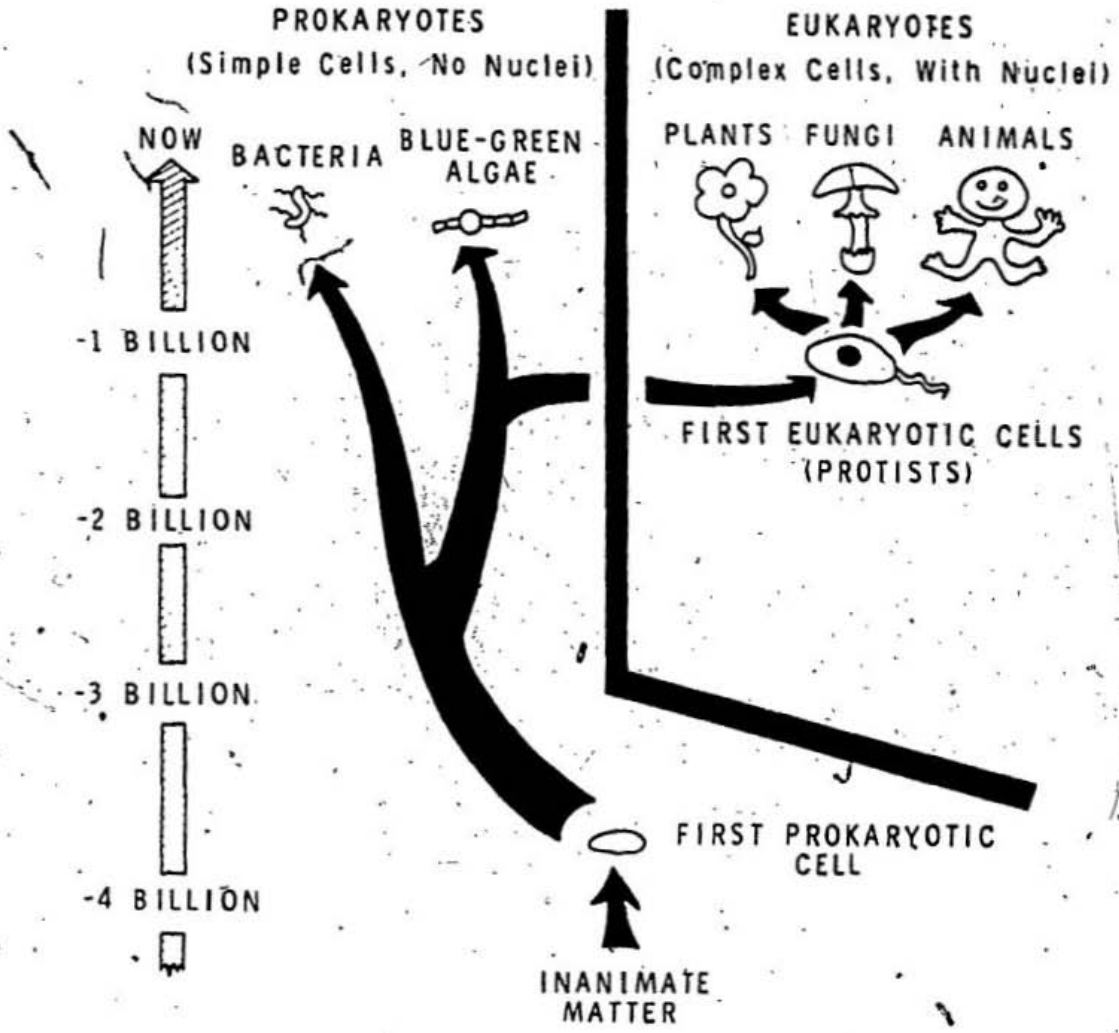


Figure 1-2: Evolutionary Progression of Eukaryotes and Prokaryotes

Source: Doolittle, 1980

1.3. Morphological Evolution

A tacit assumption of the fossil record is that morphological changes reflect genomic changes. It is generally accepted that there is a relationship between morphological and genomic evolution, but morphological complexity may not in fact be an accurate index of the extent of genomic evolution (Schopf *et al.*, 1975). This is supported by evidence (reviewed by Wilson *et al.*, 1977) that some vertebrates experience faster rates of phenotypic evolution than others. For example, frogs have remained essentially unchanged morphologically since the first frogs appeared whereas the present day mammals show a great variety of shapes compared with their common ancestor. Yet both frogs and mammals share the same rate of evolution at the protein level. It is difficult to assess the relative extent of genomic evolution of organisms from comparisons of morphological characteristics. One approach to investigate this conundrum has been to compare the sequences of specific gene products from different organisms.

1.4. Molecular Evolution

Many protein sequences have been determined and collected into a very useful, growing compendium: *The Atlas of Protein Sequences and Structures* (Dayhoff, 1978). Amino acid sequence comparisons of homologous proteins from different species revealed that the primary structures of proteins differ in a phylogenetically consistent manner. That is, the more closely related the species, the more similar are the amino acid sequences of their homologous proteins. Divergence times of organisms can be obtained from the geological record. Thus, the rates of amino acid substitution can be calculated for several proteins. It was

found that the rate of amino acid substitution is constant within a particular set of homologous proteins, but the rates are not necessarily the same for proteins having different functions (Wilson *et al.*, 1977). This is illustrated in Figure 1-3. These observations form the basis for the concept of the "Molecular Clock" (Dickerson and Geis, 1969).

The rate of protein evolution is measured in terms of unit evolutionary period (UEP), i.e. the time in million of years required for one per cent change in amino acid sequence to accumulate between two divergent lineages. The greater the UEP, the slower the rate of change. It has been observed that there is a good correlation between the rates of molecular evolution of proteins and their biological functions. For example, fibrinopeptides evolve faster than cytochrome c. Fibrinopeptides have little known function after they are cut out of fibrinogen when it is converted to fibrin in a blood clot. Virtually any amino acid change that permits the peptides to be removed may be acceptable. On the other hand, cytochrome c interacts with macromolecules such as cytochrome oxidase and cytochrome c reductase. Its function is more specific, and any change in the structure that alters the function is detrimental. Therefore, cytochrome c is more conserved and its UEP is greater than that of fibrinopeptides. Table 1-1 shows the UEPs for several proteins.

Cytochrome c is found in mitochondria of nucleated cells and this protein has been used to examine the evolutionary relationships among eukaryotes. The amino acid sequences of cytochrome c have been determined for more than 35 species and a phylogenetic tree based on comparisons of these sequences has been

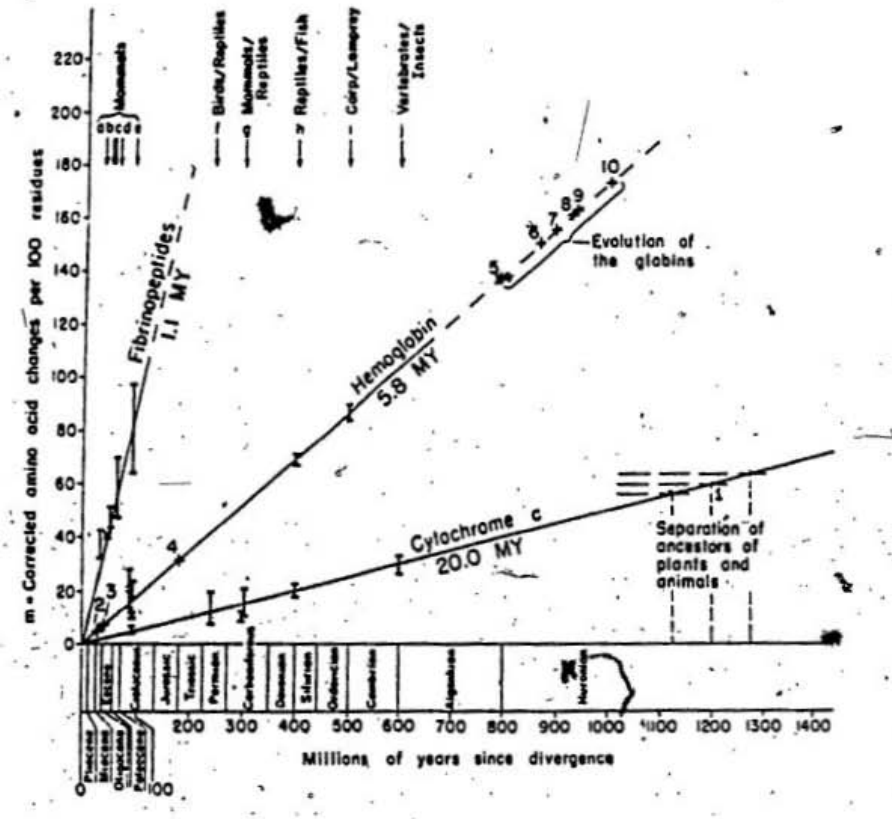


Figure 1-3: Rates of Protein Evolution

Source: Dickerson, 1971

Table 1-1: UEPs for Several Proteins

Protein	U.E.P. (milllion years)
Fibrinopeptide	1.1
Albumin	3
Myoglobin	6
Cyctochrome c	15
Triose-phosphate isomerase	19
GAPDH	20
Histone 4	400

GAPDH denotes Glyceraldehyde-3-phosphate dehydrogenase

constructed (Dickerson, 1971). This phylogenetic tree is consistent with conventional taxonomic relationships that have been derived from classical methods such as comparative anatomy, embryology and paleontology. Thus, evolutionary data at the molecular level provide an invaluable tool for determining relationships among organisms. In particular, it is possible to predict when organisms last shared a common ancestor even in the absence of a fossil record. Although some prokaryotes contain a protein with similar function to the cytochrome c from eukaryotic mitochondria, it is not known whether these proteins are homologous. Therefore, cytochrome c cannot be used to determine the evolutionary relationships between prokaryotes and eukaryotes.

1.5. The Evidence for Three Primary Kingdoms

The ability to translate messenger RNA into protein is a common feature of all living organisms. Hence, molecules associated with the translational apparatus have been used to investigate the evolutionary origin of prokaryotes and eukaryotes. Ribosomal RNA (rRNA) from the small ribosomal subunit was one of the first components to be investigated in detail (Woese and Fox, 1977). Comparative analyses of the structures of the 16S and 18S rRNAs from prokaryotes and eukaryotes involving an oligonucleotide cataloging approach were used to examine the phylogenetic relationships among these organisms. Each radioactively labelled rRNA was digested with ribonuclease T_1 and the resulting fragments were resolved by two-dimensional paper electrophoresis and then sequenced. Thus a catalog of oligonucleotide sequences characteristic of a specific organism was obtained. Association coefficients were calculated for all the pair-wise combinations of these catalogs. Although the relationship between these

association coeficients and the actual number of nucleotide sequence differences is unknown, these values were used to construct dendrograms. The phylogeny that resulted from this analysis indicated that living systems can be categorized into three primary lineages: the eubacteria, the urkaryotes and the archaebacteria (Woese and Fox, 1977; Fox *et al.*, 1980). It appears that these lineages are equidistant from one another, and that they diverged independently from a common ancestor (Figure 1-4).

The eubacterial kingdom contains all the typical bacteria. It can be subdivided into the cyanobacteria, the gram-positive, and the gram-negative bacteria. The urkaryotes, the predecessor of the eukaryotes, are defined by the similarities in the structures of their cytoplasmic 18S rRNAs. Eubacteria and urkaryotes correspond to prokaryote and eukaryote in the conventional sense. The third kingdom, the archaebacteria, comprise methanogens, thermoacidophiles and halobacteria (for review, see Doolittle, 1980). This hypothesis has received support from independent studies on the three dimensional structures of ribosomal small subunits (Lake *et al.*, 1982; Lake, 1983).

1.6. How Good is the Evidence for Three Primary Kingdoms?

Sequence comparisons of complete 16S or 16S-like rRNAs (18S) from organisms representative of the eubacteria, urkaryotes and archaebacteria revealed that the primary structure of the archaebacterial 16S rRNA is more similar to both its eubacterial and eukaryotic counterparts, than these two are to one another (Gupta *et al.*, 1983). These results suggest that the rate of evolution in the archaebacterial lineage is slower than in the other two lineages, and that

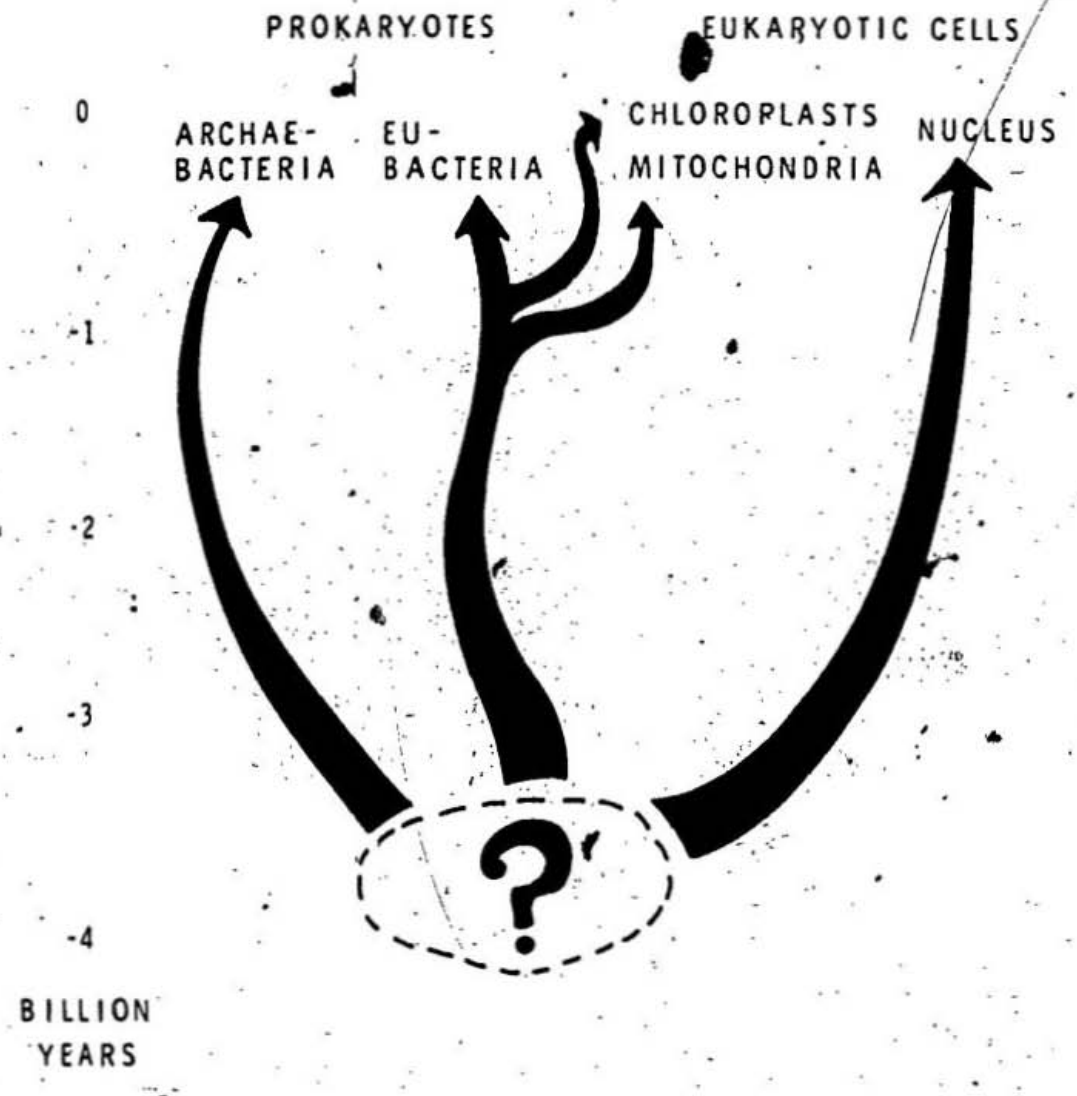


Figure 1-4: . Independent Evolution of Archaebacteria, Eubacteria and Urkaryotes

Source: Doolittle, 1980

the structure of the 16S rRNA in archaebacteria is closer to that of the common ancestor. One of the basic assumptions of phylogenetic studies at the molecular level is that the rate of evolution is constant along each lineage. Therefore, the comparison of the 16S rRNA sequences casts doubt on the dendrograms derived from the rRNA cataloging experiments.

Other studies on ribosomal components, namely comparisons of the amino terminal sequences of ribosomal proteins (Yaguchi *et al.*, 1982; Kimura and Langner, 1984), also question the relationships among the three primary kingdoms. These comparisons indicated that archaebacteria are more closely related to eukaryotes than to the eubacteria. Since the data accumulated from comparative studies on the molecules of the translational apparatus are insufficient to determine unequivocally the phylogenetic origins of prokaryotes and eukaryotes, molecules other those present in ribosomes should be examined. These macromolecules must satisfy the following criteria:

- the molecules must be ubiquitous in nature, that is be present in all organisms regardless of their taxonomic organization,
- they must be conserved molecules, with a slow but constant rate of evolution, and
- they should contain sufficient inherent information, derived from a common genetic locus, to allow a long range phylogenetic study.

1.7. Can Glycolytic Enzymes and Their Genes Serve as Evolutionary Probes?

Glycolytic enzymes seem to satisfy the above criteria. The UEPs of the glycolytic enzymes examined to date, are surpassed only by those of histones

(Wilson *et al.*, 1977; Fothergill-Gilmore, 1986; Table 1-1). This indicates that glycolytic enzymes are highly conserved. In addition, glycolytic enzymes are found in all organisms, and are easily isolated (Spring and Wold, 1971; Chin *et al.*, 1981a). Many of the enzymes involved in this pathway have been investigated in great detail (Fothergill-Gilmore, 1986).

For example, the amino acid sequences of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) have been determined for a variety of eukaryotes including, human, pig, rat, chicken, lobster and yeast, and three prokaryotes (*Thermus aquaticus*, *Bacillus stearothermophilus* and *Escherichia coli*) (Milner *et al.*, 1983; Branlant *et al.*, 1983; Holland *et al.*, 1983; Hocking and Harris, 1980;). These sequences are of similar length, approximately 333 amino acid residues, and their alignment clearly demonstrates that they are homologous proteins (Dayhoff *et al.*, 1978). Sequence comparisons of the primary structures of GAPDHs from the above organisms showed that there is approximately fifty percent homology between the enzymes of prokaryotes and eukaryotes. This indicates that this glycolytic enzyme contains sufficient information to resolve the relationship among prokaryotes and eukaryotes. Moreover, the similarity in length of glycolytic enzymes among organisms provides an additional advantage over rRNAs in which the alignment of their sequences requires insertions and/or deletions to give maximum matching (Brimacombe, 1984). Thus, glycolytic enzymes and their genes should provide an excellent means for studying long range evolution.

Chapter 2

Research Proposal

2.1. Purpose

The main purpose of this project was to investigate the possibility of using cloned *E. coli* genes encoding glycolytic enzymes for long range evolutionary studies. Based on the available protein data, it was anticipated that the study of these genes would provide some clues to the phylogenetic relationships within prokaryotes, and between prokaryotes and eukaryotes. This project was also designed to form the foundation for future studies concerning the regulation and expression of glycolytic genes in prokaryotes, and to determine how these prokaryotic genes differ from their eukaryotic counterparts.

2.2. Approach of Investigation

The source of the cloned *E. coli* glycolytic genes was the Clarke-Carbon clone bank. This clone bank was constructed by Clarke and Carbon in 1976. It is a genomic library which contains hybrid recombinant plasmids representing the entire *E. coli* genome. The plasmids were constructed by shearing total *E. coli* K-12 (strain CS520: *Hfr*C, *trp*A58,*met*B,*glyVsu*58) genomic DNA into fragments with an average size of 8.4×10^6 daltons. These fragments were treated with λ 5'-exonuclease and then inserted into the Eco R1 site of the Col E1 cloning vector

using poly dT-dA extensions. The resultant hybrid plasmids were transformed into another *E. coli* K-12 (strain JA200: F^+, *Δtrp*E5,*recA*,*thr*, *leu*, *lac*Y) (Clarke and Carbon, 1976). The hybrid plasmids pLC 16-4, pLC 10-47 and pLC 11-8 were selected for this project. These plasmids were first identified by Thomson *et al.* (1979) for their ability to complement *E. coli* mutations at the triose-phosphate isomerase (*tpi*) locus (pLC 16-4) or in the enolase gene (pLC 10-47 and pLC 11-8).

This project was divided into two sections. In the first part (Chapters 4 and 5) these three plasmids were characterized. This involved constructing restriction endonuclease maps, locating the position of the vector in each hybrid recombinant plasmid, and determining the approximate position of the glycolytic genes. The second part (Chapter 6) involved using the the cloned *E. coli* glycolytic genes to examine the extent of relatedness between organisms. Genomic DNA from representatives of the eubacteria, archaebacteria and eukaryotes was digested with restriction endonucleases and subjected to Southern blot analyses using the plasmids pLC 16-4 and pLC 10-47 as radioactively-labelled hybridization probes.

Chapter 3

Materials and Methods

3.1. Materials

Salmon sperm DNA, *Micrococcus lysodeikticus* DNA, *Clostridium perfringens* DNA, ribonuclease, proteinase K, and chemicals unless otherwise stated were purchased from Sigma Chemical Co., St Louis, MO., USA. Acrylamide, bisacrylamide, TEMED, and ammonium persulphate were BioRad products (BioRad Canada Ltd., Mississauga, Ont.). DEAE-Sephadex was bought from Pharmacia (Canada) Ltd., Dorval, Que. Restriction enzymes were obtained from Bethesda Research Laboratories (BRL) Inc., Gaithersburg, MD., USA. α^{32}P-dCTP and reagents for nick translation were bought from Amersham, Oakville, Ont.

3.2. Strains and Plasmids

Cultures of *Escherichia coli* (JA200) and *E. coli* containing either pLC 10-47, pLC 11-8 or pLC 16-4 were obtained from Dr. Barbara J. Bachmann at the *E. coli* Genetic Stock Center, New Haven, CT., USA. *E. coli* (WA802) harbouring Col E1 was a gift from Dr. L. Visentin, N.R.C, Ottawa, Ont. The names and sources of other organisms used in these studies are given in Table 3-1.

Table 3-1: Sources of Bacteria

Organism	Source	ATCC
Salmonella choleraesuis	Biology Dept., Memorial University	12011
Salmonella typhimurium	Biology Dept., Memorial University	N/A
Citrobacter freundii	Biology Dept., Memorial University	8090
Shigella sonnei	Biology Dept., Memorial University	25931
Proteus mirabilis	Biology Dept., Memorial University	9240
Enterobacter aerogenes	Dr. J. Wright, Memorial University	E13048
Erwinia carotovora	Dr. J. Wright, Memorial University	E 495
Serratia marcescens	Dr. J. Wright, Memorial University	E13880
Klebsiella pneumoniae	Biology Dept., Memorial University	E13883
Edwardsiella tarda	Dr. J. Wright, Memorial University	N/A
Yersinia enterocolitica	Dr. J. Wright, Memorial University	N/A
Pseudomonas putida	Dr. E. Barnsley, Memorial University	N/A
Bacillus subtilis	Dr. E. Barnsley, Memorial University	E23856
Lactobacillus casei	Biology Dept., Memorial University	7469
Halobacterium halobium	Dr. W.F. Doolittle, Dalhousie University	N/A
Halobacterium volcanii	Dr. W.F. Doolittle, Dalhousie University	N/A

ATCC denotes American Type Culture Collection

N/A denotes Not Available

3.3. Growth Media and Buffers

Yeast extract, bacto-tryptone and bacto-agar were purchased from British Drug House (Canada) Ltd., Halifax. The recipes for LB, M9 and other media are given in Appendix 1. The constituents of the commonly used buffers (e.g. 10 mM TE, SSC, TBE, Denhardt's solution) are described in Appendix 1.

3.4. Isolation of Plasmid DNA

A culture of *E. coli* containing the plasmid of interest was grown overnight at 37^{0}C without shaking in 10 ml LB. 1 ml of the culture was used to inoculate 250 ml LB and the *E. coli* were grown with vigourous shaking at 37^{0}C. When the A_{600} reached 0.5, chloramphenicol was added to a final concentration of 170 ug/ml and the culture was allowed to shake for another 12 to 15 hours. The cells were harvested by centrifugation and plasmid DNA was prepared according to the procedure of Maniatis *et al.* (1982), using the alkali/SDS lysis method. Covalently closed circular plasmid DNA was purified from crude extracts by equilibrium centrifugation in cesium chloride containing ethidium bromide. Plasmid DNA was extracted from the ultracentrifuge tube by puncturing the side with a 21 gauge needle, and drawing the solution into a syringe. Ethidium bromide was removed from the DNA by isoamyl alcohol extraction and the cesium chloride was removed by extensive dialysis against 10 mM TE. The purified plasmid DNA was recovered by ethanol precipitation at -20^{0}C followed by centrifugation. The plasmid DNA was dissolved in 10 mM TE. The integrity of DNA preparations was checked by agarose gel electrophoresis.

3.5. Restriction Endonuclease Digestions

Table 3-2 shows the recognition sequences of the restriction endonucleases used in this project. Samples of DNA were digested overnight at 37^{0}C in 30 ul reaction mixtures. Core buffer (supplied by the manufacturer) was used in all restriction endonuclease reactions except for Sma I when 20 mM KCl, 10 mM Tris/HCl, pH 8.0, 10 mM $MgCl_2$ and 1 mM dithiothreitol was used.

3.6. Agarose Gel Electrophoresis

DNA fragments generated by restriction endonuclease digestion were separated by electrophoresis in 0.8 or 1.0% agarose gels. The buffer system used was TBE. After electrophoresis the gels were stained in 0.5 ug/ml ethidium bromide for half an hour and the DNA was visualized by ultra-violet transillumination (Ultra-violet Products, Inc., San Gabriel, California, U.S.A.). The results were recorded by photographing the gels using a Polaroid MP-4 camera.

3.7. Growth of Prokaryotes

Bacterial cultures were started from single colonies from agar plates. Enteric bacteria and *Bacillus subtilis* were grown in LB at 37^{0}C. *Pseudomonas putida* was grown in LB at 30^{0}C. Species of Halobacteria were cultured at 37^{0}C in broth containing a high concentration of salt (See Appendix 1).

Table 3-2: Restriction Endonucleases and their Recognition Sequences

Restriction Endonuclease	Recognition Site
Bam HI	G`GATCC
Bgl I	GCCNNNN`NGGC
Bgl II	A`GATCT
Eco RI	G`AATTC
Hae III	GG`CC
Hind III	A`AGCTT
Hinf I	G`ANTC
Kpn I	GGTAC`C
Msp I	C`CGG
Pst I	CTCGA`G
Pvu II	CAG`CTG
Sac I	GAGCT`C
Sal I	G`TCGAC
Sma I	CCC`GGG
Taq I	T`CGA
Xba I	T`CTAGA

` denotes the site at which the endonuclease cleaves

3.8. Isolation of Bacterial DNA

Bacterial cultures from the organisms listed in Table 3-1 were allowed to grow until stationary phase. The cells were harvested, lyzed and genomic DNA was isolated by the procedure of Marmur (1961). These crude preparations of DNA and *Micrococcus lysodeikticus* DNA and *Clostridium perfringens* DNA purchased from Sigma were dissolved in 10 mM TE and RNA was removed by the addition of heat-treated ribonuclease at a final concentration of 100 ug/ml. After incubating for 1 hour at 37 ^{0}C, the solution was adjusted to 0.5% SDS and 100 ug/ml proteinase K was added. The mixture was further incubated at 50^{0}C for 2 hours and then deproteinized by extraction with an equal volume of phenol saturated with 50 mM Tris/HCl, 1 mM EDTA, 0.5% SDS (pH 7.5). The aqueous and phenol phases were separated by centrifugation at 4,000x g for 15 minutes at 4^{0}C. The aqueous layer was adjusted to 0.2 M sodium acetate (pH 5.1), and the nucleic acids were precipitated by the addition of two volumes of ethanol. The purified DNA was dissolved in 10 mM TE and dialyzed extensively against the same buffer before use. This procedure yields high molecular weight DNA with an A_{260}/A_{280} ratio of 1.8 to 2.0.

3.9. Southern Blotting

After agarose gel electrophoresis, DNA fragments were denatured *in situ* by soaking the entire gel in 300 ml of 0.5 M NaOH/1.5 M NaCl, with constant agitation at room temperature. The gels were incubated twice for 45 minutes. The gels were then neutralized in the presence of 0.5 M Tris/HCl, 3 M NaCl, pH 7.5 with agitation. This step was repeated twice for 30 minutes each time. After

neutralization, the DNA fragments from plasmids were transfered to Genescreen (New England Nuclear, Boston, MASS., USA) using 10X SSC, according to the manufacturer's instructions (Method I, Catalog No. NEF-972). For Southern blotting of genomic DNA, Biodyne A nylon membranes (Pall Ultrafine Filtration Corperation, Glen Cove, N.Y.) were used. DNA fragments were fixed on to the filters by baking the membranes for 2-4 hours at 80^0C under vacuum.

3.10. Nick Translation

Plasmid DNA was radioactively-labelled with α^{32}P-dCTP, using a nick translation kit purchased from Amersham. The reaction mixture was incubated at 10^0C for 90 minutes and the reaction was stopped by the addition of EDTA to a final concentration of 100 uM. Plasmid DNA was separated from unincorporated α^{32}P-dCTP by chromatography in Sephadex G-50 in 10 mM TE containing 0.1% SDS. The DNA labelled using this method routinely had a specific activity of 10^8 cpm per ug DNA.

3.11. Hybridization and Washing Conditions

The Genescreen filters were prehybridized in a solution containing 50% formamide, 5X SSC, 5X Denhardt's solution, 50 mM sodium phosphate (pH 6.8), 500 ug/ml salmon sperm DNA (sonicated, depurinated, single-stranded) and 1% glycine. Approximately 50 ul of the prehybridization solution per cm^2 of filter area was added to a sealable plastic pouch containing the GeneScreen. The pouch was heated, sealed and incubated at 37 ^{0}C with constant agitation for 24-36 hours. Hybridization was carried out in 25 ul of hybridization buffer per cm^2 of filter plus the radioactive probe (2500 cpm per unit cm area of the filter) at 42 ^{0}C

for 36 hours. The hybridization buffer consisted of 50% formamide, 5X SSC, 1X Denhart's solution, 20 mM sodium phosphate (pH 6.8) and 100 ug/ml salmon sperm DNA. After hybridization the filters were washed twice in 2X SSC, 0.1% SDS at room temperature for 30 minutes and then twice in 1X SSC, 0.1% SDS at 50 ^{0}C for the same period of time. The filters were dried at room temperature and subjected to autoradiography using Kodak X-Omat RP X-ray film (Kodak Canada, Inc., Toronto, Canada) for two to seven days, at -70 ^{0}C. A cassette with an intensifying screen (Dupont Cronex) was used to enhance the autoradiographic process.

3.12. Preparation of Extracts

E. coli was grown in M9 media, in the presence or absence of glucose. In the former case, glucose was added at a final concentration of 0.4%. Crude extracts were prepared from *E. coli* (JA200) alone and from *E. coli* (JA200) containing pLC 10-47, pLC 11-8 or pLC 16-4. Cell density was monitored by measuring the absorbance at 600 nm. The cells were harvested in the midlog phase of growth by centrifugation, and the cells were resuspended in 50 mM imidazole/HCl, pH 6.8 containing 10 mM $MgCl_2$. Toluene was added to the suspension to a final concentration of 1 % to aid in lysing the cells. Further lysis was completed by two minutes of sonication at 40 watts, using a Sonificer cell disruptor (Heat Systems, Ulltarsonics Inc., Plainview, N.Y). The cells were sonicated for 30 seconds, followed by cooling on ice water for one minute. The sonication and cooling steps were repeated four times. After sonication the cell extract was centrifuged for 10 minutes at 12,000x g at 4 ^{0}C, and the clear supernatant was used for enolase assays.

3.13. Enolase Assay

Enolase activity was measured by following the increase in the absorbance at 240 nm using a PYE UNICAM SP8-100 UV/Vis spectrophotometer. The assay mixture consisted of 50 mM imidazole/HCl, pH 6.8, 10 mM $MgCl_2$ and 2 mM of 2-phosphoglyceric acid., The protein concentration of an extract was estimated from its absorbance at 280 nm. Specific activities are given in terms of the change in absorbance at 240 nm per minute per ml of extract with an absorbance at 280 nm of 1.

3.14. SDS Polyacrylamide Gel Electrophoresis

Crude extracts of cultures, prepared as described in Section 3.12, were subjected to electrophoresis under denaturing conditions in 10% polyacrylamide gels using the procedure of Laemlli (1970). Bovine serum albumin, rabbit enolase, chicken lactate dehydrogenase and chymotrypsinogen were used as standards. Gels were stained with 0.1% Coomassie Brilliant Blue G250 in 1.75 % perchloric acid for two hours and then destained in 7% acetic acid.

3.15. DEAE-Sephadex Chromatography

Crude extracts prepared as described above were dialysed overnight at 4°C against 2000 volumes of 10 mM Tris/HCl, pH 8.5. The dialyzed extract was applied to a column (1.5 cm x 20 cm) of DEAE-Sephadex equilibrated with 10 mM Tris/HCl, pH 8.5. The column was washed with 50 ml of this buffer and then a linear gradient (0 to 500 mM) (5 mM per ml) of NaCl in 10 mM Tris/HCl, pH 8.5 was applied. After the gradient was completed, 1 M NaCl in Tris/HCl, pH 8.5 was applied to the column. Fractions were tested for enolase activity

using the standard assay and protein was monitored by following the absorbance at 280 nm.

Chapter 4

Characterization of pLC 16-4

4.1. Background

Triose-phosphate isomerase (TPI) [E.C. 5.3.1.1.] catalyzes the interconversion of dihydroxyacetone-phosphate and glyceraldehyde-3-phosphate. This protein is one of the best characterized glycolytic enzymes. The amino acid sequence of this protein has been determined from eukaryotes such as human, rabbit, chicken, coelacanth, and yeast, (Dayhoff, 1978; Lu *et al.*, 1984) and from a prokaryote, *Bacillus stearothermophilus* (Artavanis-Tsakonas and Harris, 1980; Kolb, 1980). Comparisons of the amino acid sequences showed that TPI is a highly conserved protein, suggesting that all TPI genes are descendents of a single primordial gene (Straus and Gilbert, 1985).

From genetic mapping of the *E. coli* chromosome, it is known that the genes encoding TPI and phosphofructo-kinase (PFK) are tightly linked to one another. The chromosomal location of the *tpi* locus is 87.8 min and the *pfk A* gene maps at the 87.7 min region of the chromosome (Bachmann, 1980). Thomson *et al.* (1979) have shown that the hybrid recombinant plasmid pLC 16-4 from the Clarke-Carbon clone bank (Clarke and Carbon, 1976) complements both TPI^- and PFK^- phenotypes. Therefore, it was decided to use pLC 16-4 as a source of the gene

encoding TPI and PFK from *E. coli* to examine the extent of sequence homology at this locus among prokaryotes and eukaryotes. It was hoped that the information derived from these studies would also indicate whether or not it would be feasible in the future to use hybrid plasmids containing *E. coli* glycolytic genes as probes to isolate the genes for glycolytic enzymes from archaebacteria. Before the recombinant plasmid pLC 16-4 could be used for this evolutionary study it was necessary to characterize the plasmid. A restriction endonuclease map was constructed using six base recognition enzymes. The location of the cloning vector Col E1 in pLC 16-4 was determined by comparing the restriction maps of pLC 16-4 and Col E1, and by probing the recombinant plasmid pLC 16-4 with radioactively-labelled Col E1.

4.2. Results

4.2.1. Restriction Map of pLC 16-4

The restriction endonuclease map for pLC 16-4 was constructed from an analysis of the fragment sizes produced by single, double, and triple digests of pLC 16-4. The number of fragments produced by a restriction enzyme depends on the number of restriction sites it recognizes. For a circular DNA, the number of fragments generated by a single enzyme is the same as the number of cleavage sites made by the enzyme whereas in a double digestion, the total number of fragments obtained equals the number of fragments generated by the first enzyme plus the number of fragments produced by the second enzyme.

DNA fragments travel through agarose gel matrices at rates that are

inversely proportional to the $\log_{10}$ of their sizes (Helling *et al.*, 1974). Thus, a larger fragment migrates at a slower rate than a smaller fragment. A plot of $\log_{10}$(size in base pairs) versus distance migrated by standard markers enables the approximate size of the unknown fragments to be estimated (Table 4-1; Figure 4-1). Knowing the distance travelled by the DNA fragment of interest, then its size can be determined from the standard plot.

The strategy used for constructing a restriction map was as follows (Refer to Figure 4-2 and Table 4-2): Bgl II (lane 2) cuts the plasmid pLC 16-4 at two sites to generate fragments that migrated 46 mm and 109 mm from the point of application. These distances correspond to sizes of 11.5 kbp and 2 kbp respectively. When a double digest using Eco RI and Bgl II was carried out (lane 3), three fragments with sizes 11.5, 1.4 and 0.6 kbp were produced (the 0.6 kbp fragment was too small to be seen on this gel, Figure 4-2). Table 4-2 summarizes the sizes of the restriction fragments of the digest in Figure 4-2.

This showed that Eco RI cleaves the 2 kbp (Bgl II - Bgl II) to give the two smaller fragments (Figure 4-3(b)]. A triple digest using Bgl II, Eco RI and Hind III (lane 4) generated four fragments, indicating that Hind III cuts the 11.5 kbp (Bgl II - Bgl II) into 7.5 and 4.0 kbp. The two possible permutations of this digest are as shown in Figure 4-3(c) and (d). The relative orientation of the restriction sites was achieved by examining the Hind III/Eco RI double digest. This produced two fragments of size 8.0 and 5.4 kbp (Figure 4-3(e)]. Therefore, as illustrated in Figure 4-3, the 5.4 kbp Eco RI - Hind III fragment must contain a Bgl II site : 1.4 kbp from the Eco RI and 4.0 kbp from the Hind III site i.e. Figure

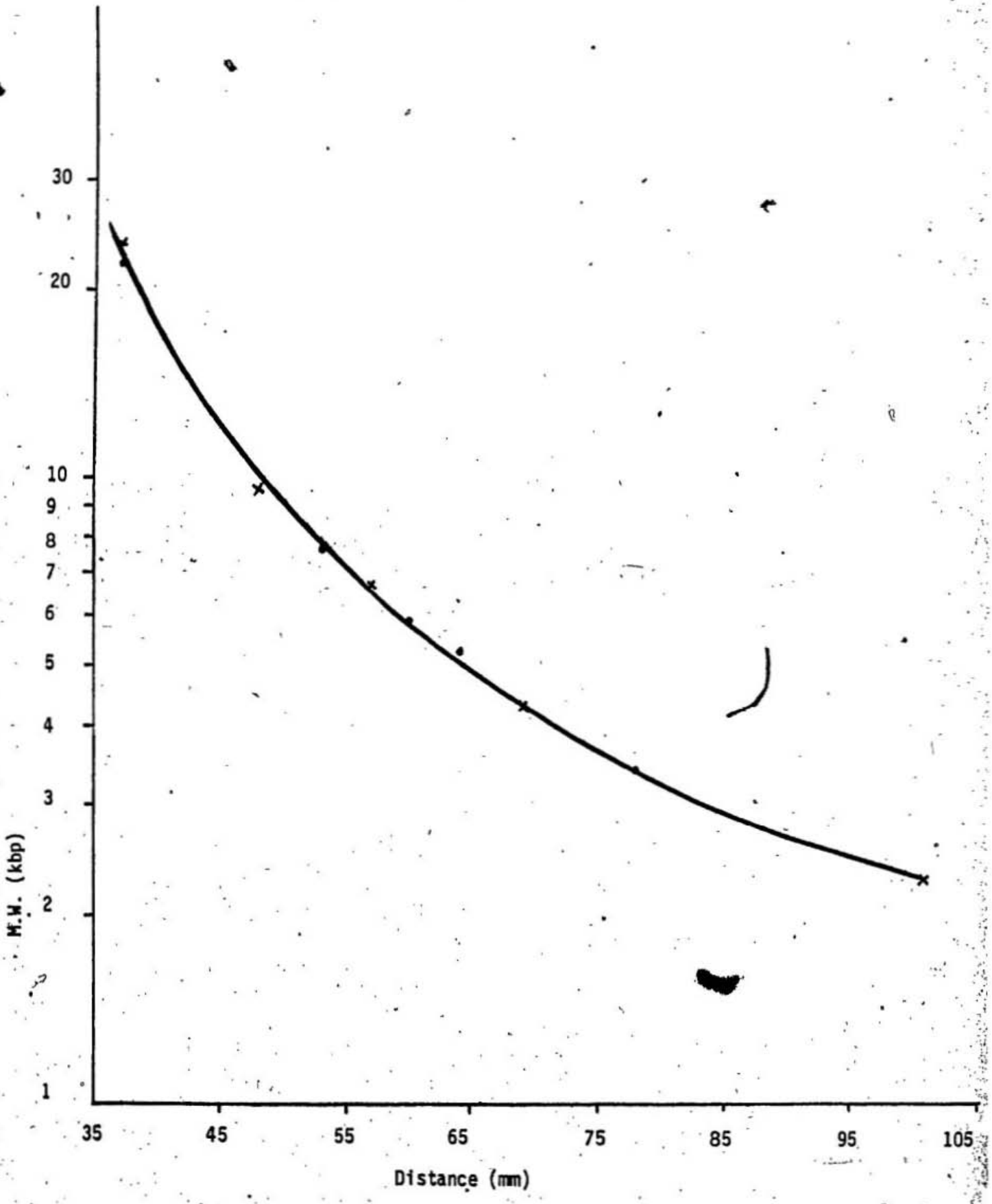
Table 4-1: Molecular Sizes and Distances Migrated by λ DNA Digested with Hind III or Eco RI

distance (mm)	size (kbp)
λ DNA/Hind III	
37	23.7
48	9.5
57	6.7
69	4.3
101	2.3
109	2.0
λ DNA/Eco RI	
37	21.8
53	7.6
60	5.9
64	5.2
78	3.4

The distance migrated by the DNA fragments was measured in millimeters (mm) from the point of origin i.e. where the DNAs were first loaded in the wells.

Figure 4-1: A Semi-logarithmic Plot of Molecular Weight Marker versus Distance Migrated on an Agarose Gel

The markers used were generated by digesting λ DNA with Hind III or Eco RI, and their sizes and the corresponding distances travelled are listed in Table 4-1.


30
20
10
9
8
7
6
5
4
3
2
1
M.W. (kbp)
35
45
55
65
75
85
95
105
Distance (mm)

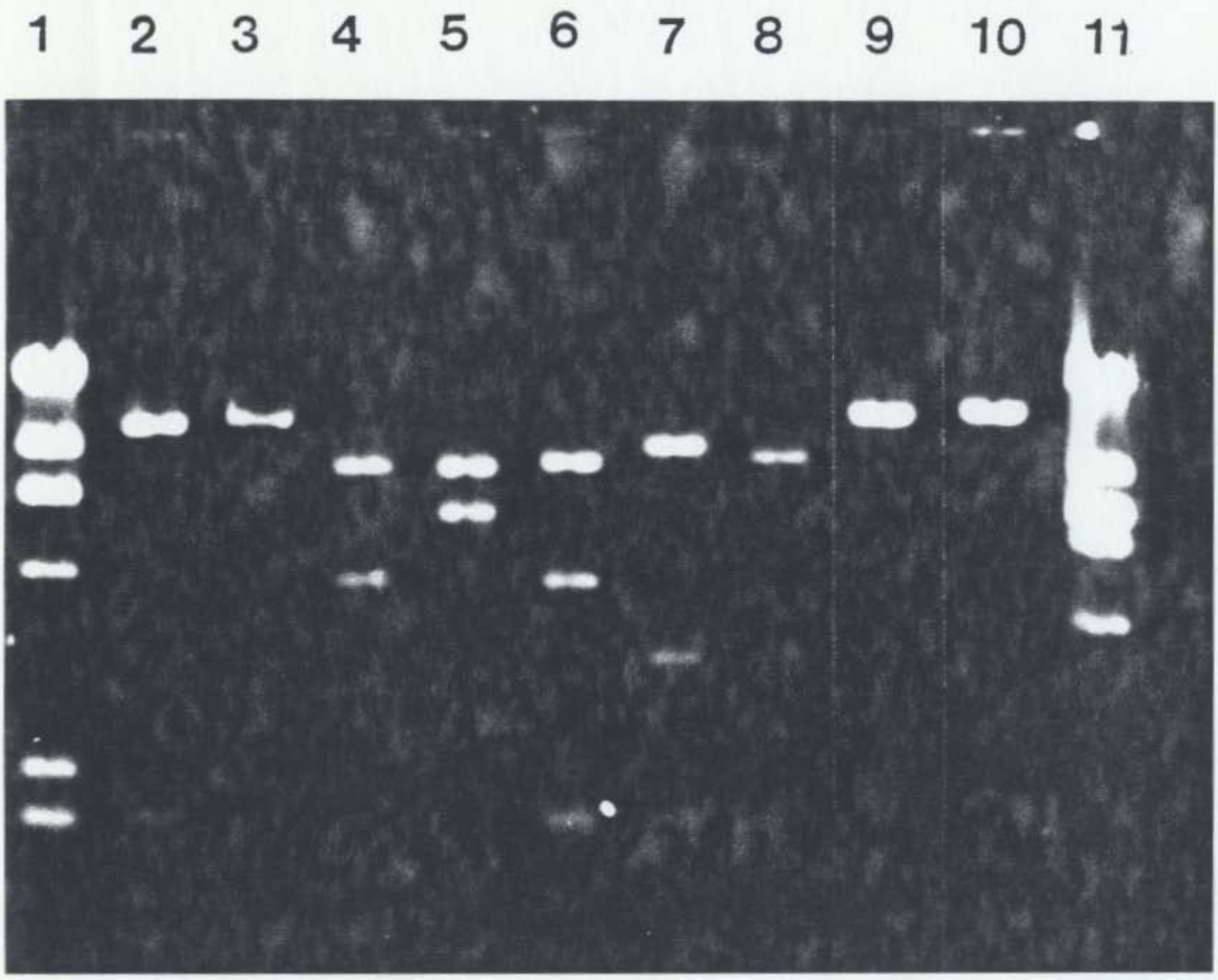


Figure 4-2: Restriction Endonuclease Analysis of pLC 16-4

The gel was electrophoresed at 100 V, in TBE buffer. Lanes 2-10 correspond to pLC 16-4 digested with : (2) Bgl II, (3) Bgl II/Eco RI, (4) Bgl II/Eco RI/Hind III, (5) Hind III/Eco RI, (6) Bgl II/Hind III, (7) Bgl II/Sac I, (8) Bgl II/Sac I/Hind III, (9) Sac I/Hind III, (10) Hind III. Lanes 1 and 11 correspond to λ DNA/Hind III and λ DNA/Eco RI markers respectively.

Table 4-2: A Summary of the Fragment Sizes Generated by Restriction Digests of pLC 16-4

Restriction Enzymes	Size (Distance) kbp mm
Bgl II	11.5 (46), 2.0 (109)
Bgl II/Eco RI	11.5 (46), 1.4 (128), 0.6
Bgl II/Eco RI/Hind II	7.5 (52), 4.0 (71), 1.4 (128), 0.6
Hind III/Eco RI	8.0 (73), 5.5 (60)
Bgl II/Hind III	7.5 (52), 4.0 (71), 2.0 (109)
Bgl II/Sac I	9.0 (48), 3.0 (83), 2.0 (109)
Bgl II/Sac I/Hind III	8.3 (51), 3.0 (83), 2.0 (109)
Sac I/Hind III	12.5 (44), 1.0
Hind III	13.5 (43)

See Figure 4-2 for the gel corresponding to these data

4-3(d) has the correct fragment size and orientation. In a similar manner, other endonucleases were mapped on pLC 16-4 (Figure 4-4).

4.2.2. Location of the Cloning Vector on pLC 16-4

The hybrid plasmid pLC 16-4 contains a piece of *E. coli* DNA insert, ligated into the cloning vector Col E1 by poly dA - dT tailing. The location of the cloning vector in pLC 16-4 was determined by a combination of restriction mapping and Southern blotting using radioactively-labelled Col E1 as the hybridization probe.

Dr. L. Visentin, N.R.C., Ottawa very kindly provided a restriction map for Col E1 (Figure 4-5). Col E1 has a unique site for Sma I. As pLC 16-4 was found to have only one Sma I site, this suggested that the unique Sma I restriction site is located in the region containing the vector, Col E1. Using this recognition site as a reference position, together with three Pst I and Pvu II recognition sites obtained from the Col E1 restriction map constructed by Dr. Visentin, the map for pLC 16-4 was established (Figure 4-4).

To locate the position of the cloning vector, Col E1, Southern transfer in conjunction with DNA-DNA hybridization was used. The Col E1 plasmid was used as a ^{32}P-labelled probe to identify the location of the cloning vector on the restriction map. The plasmid pLC 16-4 was first digested with various restriction enzymes, as seen in Figure 4-6(a). The fragments generated were then transfered to Genescreen and hybridized with radioactively-labelled K30 (cloning vector) probe. The probe being homologous to the cloning vector annealed to fragments

Figure 4-3: Four Site Map of pLC 16-4

B, E and H denote restriction sites of Bgl II, Eco RI and Hind III respectively

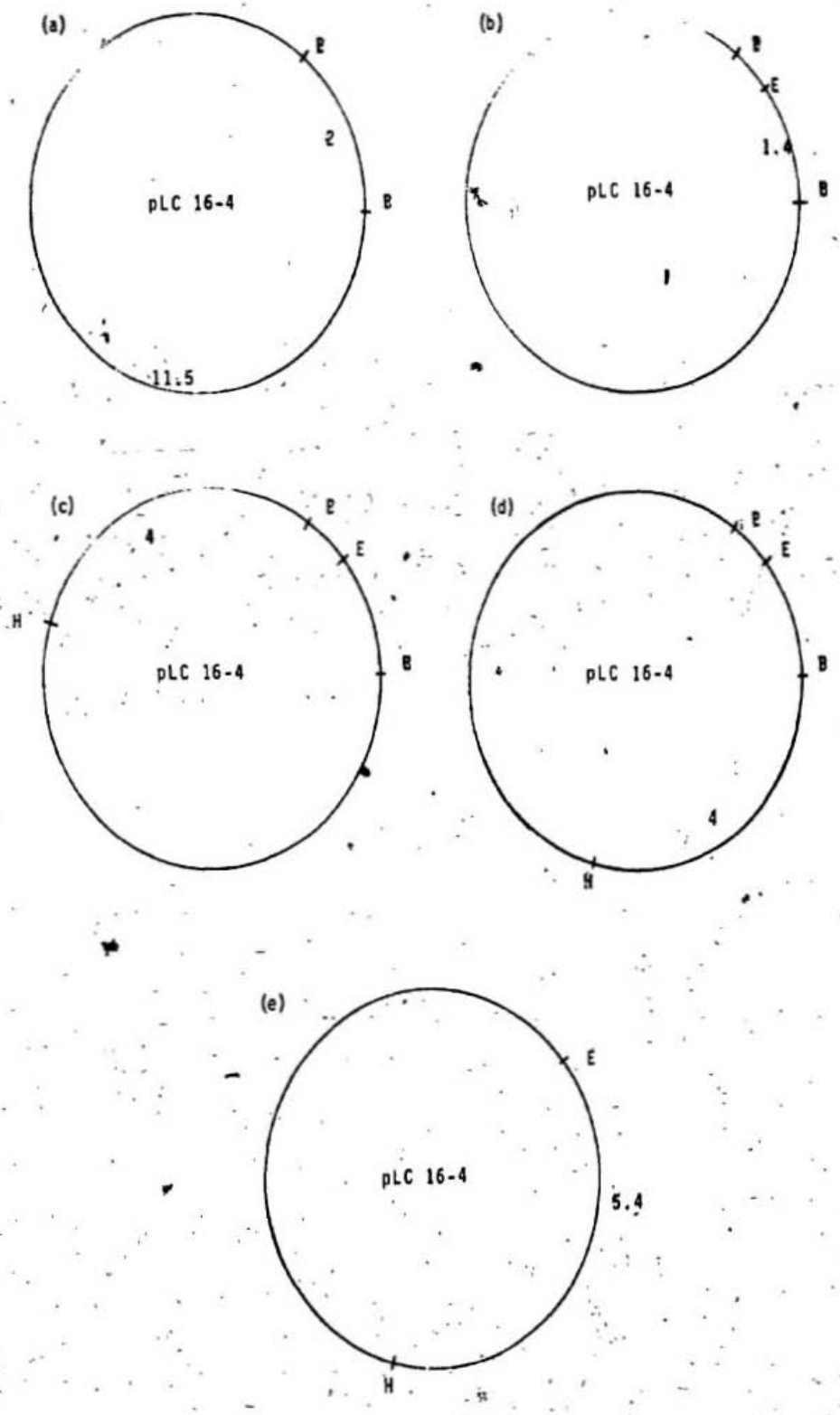
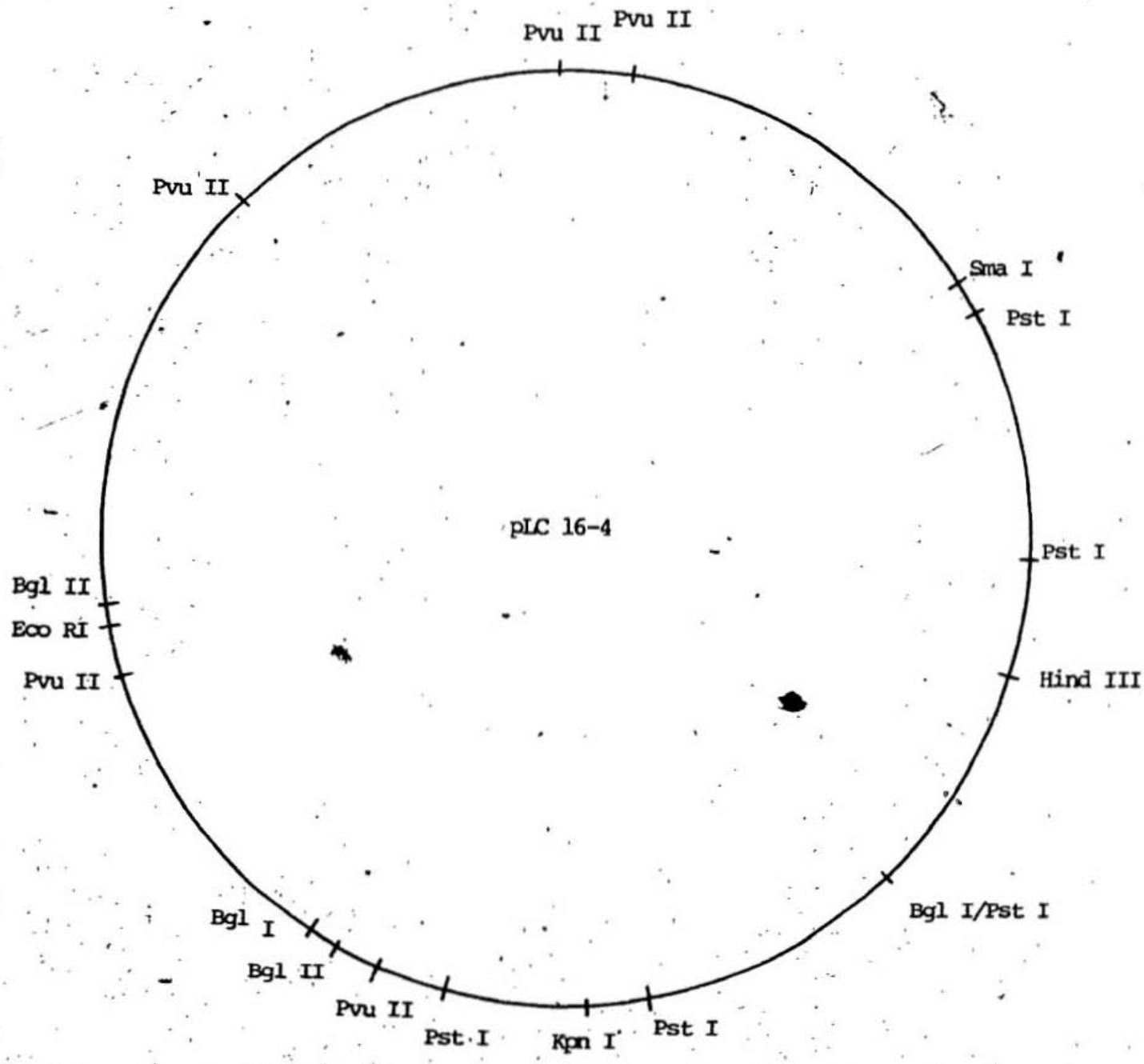

(a)
pLC 16-4
11.5
(b)
pLC 16-4
1.4
(c)
pLC 16-4
(d)
pLC 16-4
4
(e)
pLC 16-4
5.4

Figure 4-4: A Restriction Endonuclease Map of pLC 16-4


Pvu II
Pvu II
Pvu II
Sma I
Pst I
pLC 16-4
Pst I
Bgl II
Eco RI
Pvu II
Hind III
Bgl I/Pst I
Bgl I
Bgl II
Pvu II
Pst I
Kpn I
Pst I

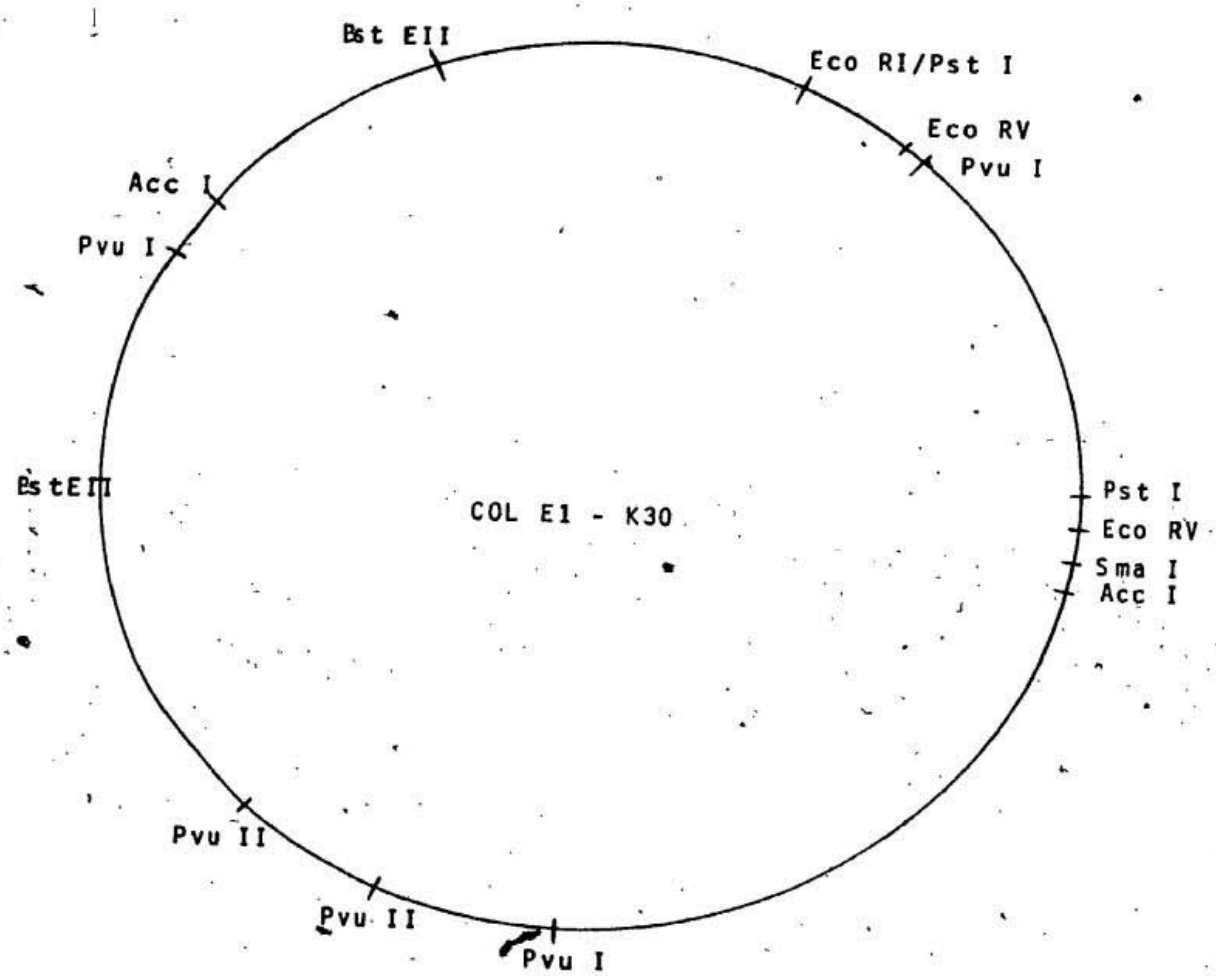


Figure 4-5: A Restriction Map of Col E1

Source: Dr. L. Visentin

of pLC 16-4 DNA containing the vector. These fragments were visualized by autoradiography (Figure 4-6(b)). For instance, the 1.4 kbp fragment of the pLC 16-4 digested with Pvu II (lane 4) or with Pvu II/Hind III (lane 13) is derived from the cloning vector. A fragment of the same size was seen when the cloning vector K30 was digested with Pvu II (lane 2).

4.3. Discussion

4.3.1. Restriction map of pLC 16-4

The size of pLC 16-4 is approximately 13.5 kbp, with the *E. coli* genomic DNA insert being 6.5 kbp. The restriction enzymes Hind III, Sac I, and Kpn I each has a unique site on the genomic fragment, and Sma I cuts the cloning vector once. During the course of this project three other laboratories independently worked on and published data on this plasmid. The restriction map for pLC 16-4 shown in Figure 4-4 is consistent with those generated by Shimosaka *et al.* (1982), Pichersky *et al.* (1984), and Hellinga and Evans (1985).

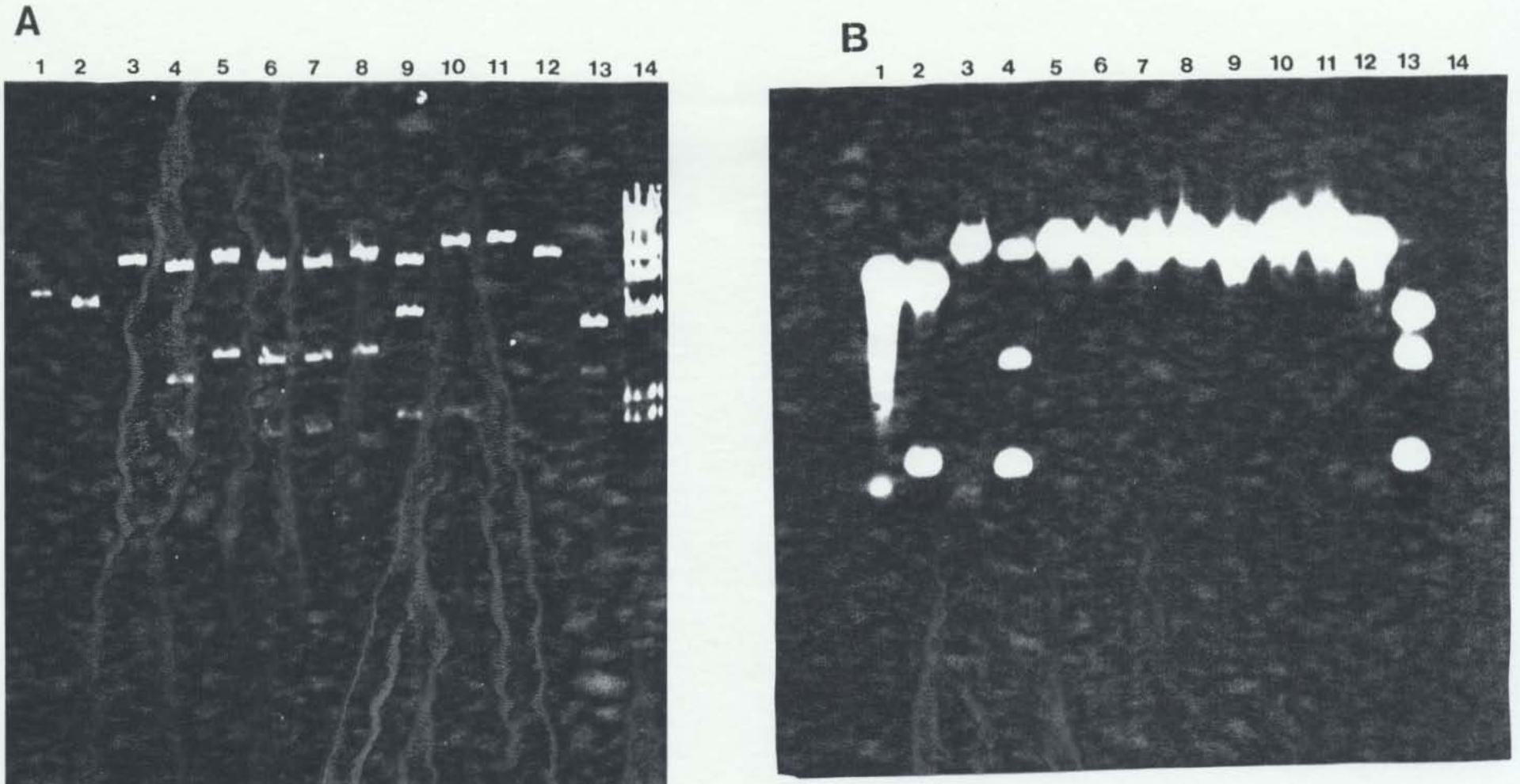
4.3.2. The *E. coli* Genes on pLC 16-4

The approximate positions of the *tpi* and *pfk A* genes on pLC 16-4 were determined by subcloning fragments into pBR322 (Shimosaka *et al.*, 1982). Shimosaka and co-workers (1982) showed that *tpi* and *pfk A* genes can be transcribed independently, in spite of the fact that they are closely linked to one another. Their results also indicated that an Eco RI site is located in the *pfk A* gene region since this endonuclease abolished the PFK activity. The *tpi* gene is located in the 0.9 kb Pst I fragment, and its nucleotide sequence has also been

Figure 4-6: Analysis of pLC 16-4 using Restriction Enzymes and Southern Blotting with the Cloning Vector K30 as Radioactively-Labelled Probe

(a) Lane 1 and 2 correspond to K30 digested with Pst I and Pvu II respectively. Lane 3-13 are restriction digest of pLC 16-4 with (3) Pst I, (4) Pvu II, (5) Bgl I, (6) Bgl I/Bgl II, (7) Bgl I/Bgl II/Hind III, (8) Bgl I/Hind III, (9) Bgl II/Hind III, (10) Bgl II, (11) Hind III, (12) Pst I/Hind III, (13) Pvu II/Hind III. Lane 14 corresponds to λ DNA-Hind III marker.

(b) Autoradiograph of Southern Blot of (a) Probed with ^{32}P-labelled Col E1.


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14

determined (Pichersky *et al.*, 1984). The entire region between the genes encoding TPI and PFK has now been sequenced (Hellinga and Evans, 1985). Two other genes have been identified between the glycolytic genes; they are (i) *sbp*, a gene for a periplasmic sulphate-binding protein, and (ii) *cdh*, a gene for the membrane-bound enzyme, CDP-diglyceride hydrolase.

Chapter 5

Characterization of pLC 10-47 and pLC 11-8

5.1. Background

5.1.1. Properties of enolase

2-Phosphoglycerate dehydrase, more commonly known as enolase, [E.C. 4.2.1.11], catalyzes the dehydration of D-2-phosphoglycerate to phosphoenolpyruvate. As one of the glycolytic enzymes acting on three carbon compounds, enolase is found in the entire spectrum of biological forms (Spring and Wold, 1971). This enzyme has been isolated from more than 20 different sources: mammals (Wold, 1971); fish (Pietkiewicz *et al.*, 1983); plants (Miernyk and Dennis, 1984); yeast (Chin *et al.*, 1981a) and bacteria (Spring and Wold, 1971) including an extreme thermophile (Stellwagen *et al.*, 1973). Because of its requirement for metal ions for proper catalytic function, enolase is often referred to as a metal-activated metalloprotein. Enolases isolated from different sources exhibit remarkably similar metal binding, metal activation and substrate specificity. In addition, it appears to have been preserved through evolution as a dimer (with the exception of certain thermophilic bacteria). Therefore, enolase should be ideal for studying evolutionary relationships between species.

5.1.2. Enolase in *E. coli*

The nucleotide sequences of the two yeast enolase genes have been determined (Holland *et al.*, 1981). The predicted amino acid sequences of the enolases encoded by these genes differ at only 20 of the 436 amino acid residues suggesting that these genes are the result of a recent duplication event. In addition, these genes are expressed differentially depending upon the carbon source in the media (McAlister and Holland, 1982). In contrast, only one form of enolase has been found in *E. coli* (Pfleiderer *et al.*, 1966; Spring and Wold, 1971). This is consistent with the *E. coli* chromosome containing a sin (*eno*) for enolase (Irani and Maitra, 1974; 1976), which appears to be constitutively expressed (Fraenkel and Vinopal, 1973; Thomson *et al.*, 1979).

Two hybrid recombinant plasmids (pLC 10-47 and pLC 11-8) isolated from the Clarke-Carbon *E. coli* genomic clone bank (Clarke and Carbon, 1976) have been found to complement a strain of *E. coli* with a point mutation in its enolase gene (Thomson *et al.*, 1979). As a start to comparing the structure and regulation of the *E. coli* enolase gene with its yeast counterparts, restriction endonuclease maps for pLC 10-47 and pLC 11-8 were constructed.

5.2. Results

5.2.1. Restriction Endonuclease Analysis of pLC 10-47 and pLC 11-8

Since only one locus has been identified for enolase on the *E. coli* chromosome, the working hypothesis was that the hybrid plasmids pLC 10-47 and pLC 11-8 both contain the *E. coli* enolase gene. If this were the case, these hybrid plasmids would represent overlapping clones of the *eno* locus. Therefore, there should be a region in the genomic DNA inserts which is common to both plasmids and this region should contain the enolase gene. In order to locate this region of homology, pLC 10-47 and pLC 11-8 were digested with restriction endonucleases that have six base pair recognition sequences and the fragment patterns were compared after separation by electrophoresis on agarose gels. Table 5-1 shows the results of these digests. The only fragments from pLC 10-47 and pLC 11-8 that comigrated were pieces of DNA that came from the cloning vector Col E1 (Table 5-1; Figure 5-1). For instance, the 1.2 kbp and 1.4 kbp fragments generated from Pst I and Pvu II digests respectively were seen in all the plasmids, pLC 10-47, pLC 11-8 and the K30 vector.

Since it was possible that pLC 10-47 and pLC 11-8 overlapped by only a very small region, a series of double and triple digests were carried out. These data were used to construct restriction endonuclease maps for both plasmids (see Section 4.2.1 for an explanation of the methodology employed). The maps are shown in Figure 5-2.

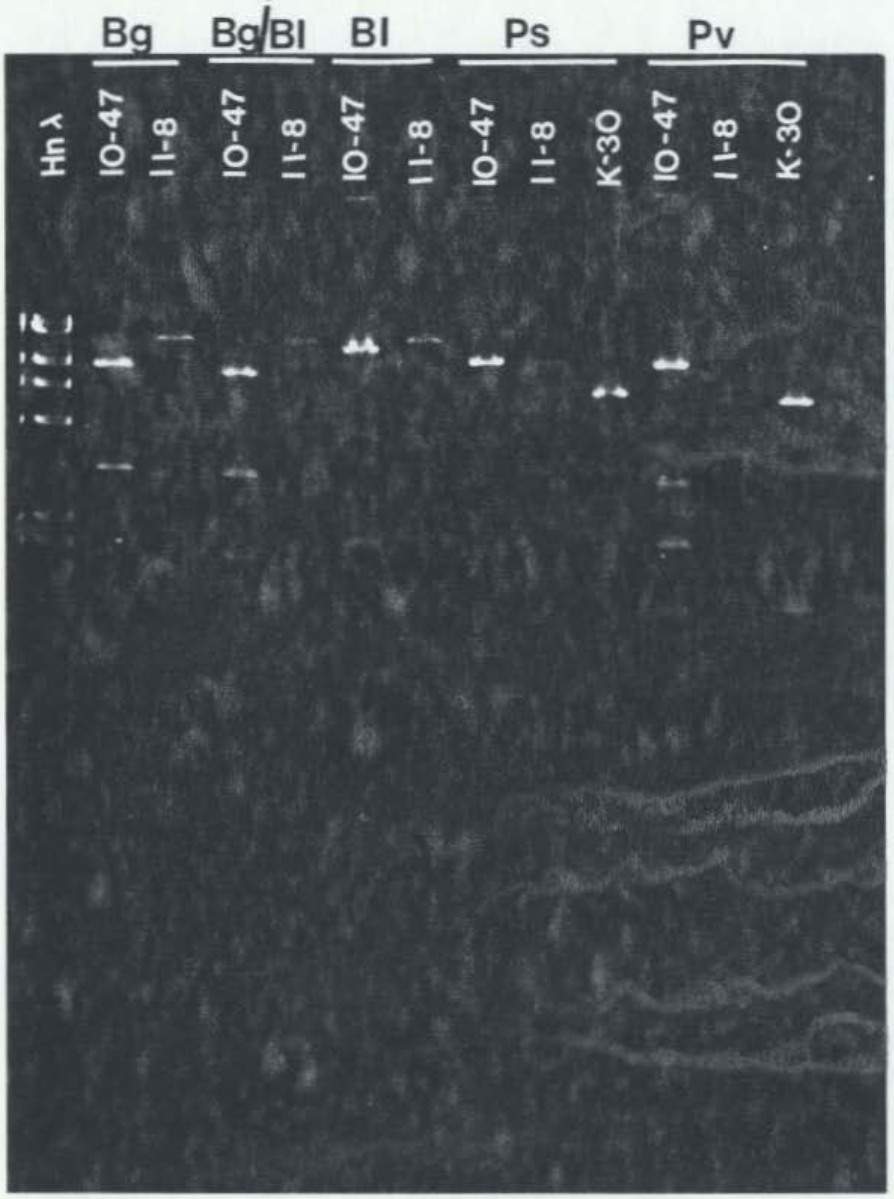


Figure 5-1: Restriction Endonuclease Analysis of pLC 11-8 and pLC 10-47 Bg, Bl, Ps and Pv denote Bgl I, Bgl II, Pst I and Pvu II respectively.

Table 5-1: Summary of Restiction Digests of pLC 11-8 and pLC 10-47

Restriction enzymes	pLC 10-47	pLC 11-8
Bam HI	19.3, 1.7	18
Bgl I	12.0, 5.6, 2.8, 0.8	18
Bgl II	21.0	15.0, 2.3, 1.35
Eco RI	21.0	NC
Hind III	21.0	NC
Kpn I	21.0	NC
Pst I	12.0, 4.8, 1.9, 1.2, 0.6, 0.4	8.1, 5.3, 2.6, 1.2, 0.95
Pvu II	6.7, 5.2, 2.6, 1.4, 1.1, 0.8	4.7, 3.8, 3.1, 2.0, 1.8, 1.4, 0.35
Sal I	14.5, 3.6, 1.8, 1.0	NC
Sma I	11.8, 9.7	18.5, 2.5
Xba I	21.0	NC

NC denotes no restriction site

The numbers denote the fragment sizes, measured in kilobase pairs (kbp)

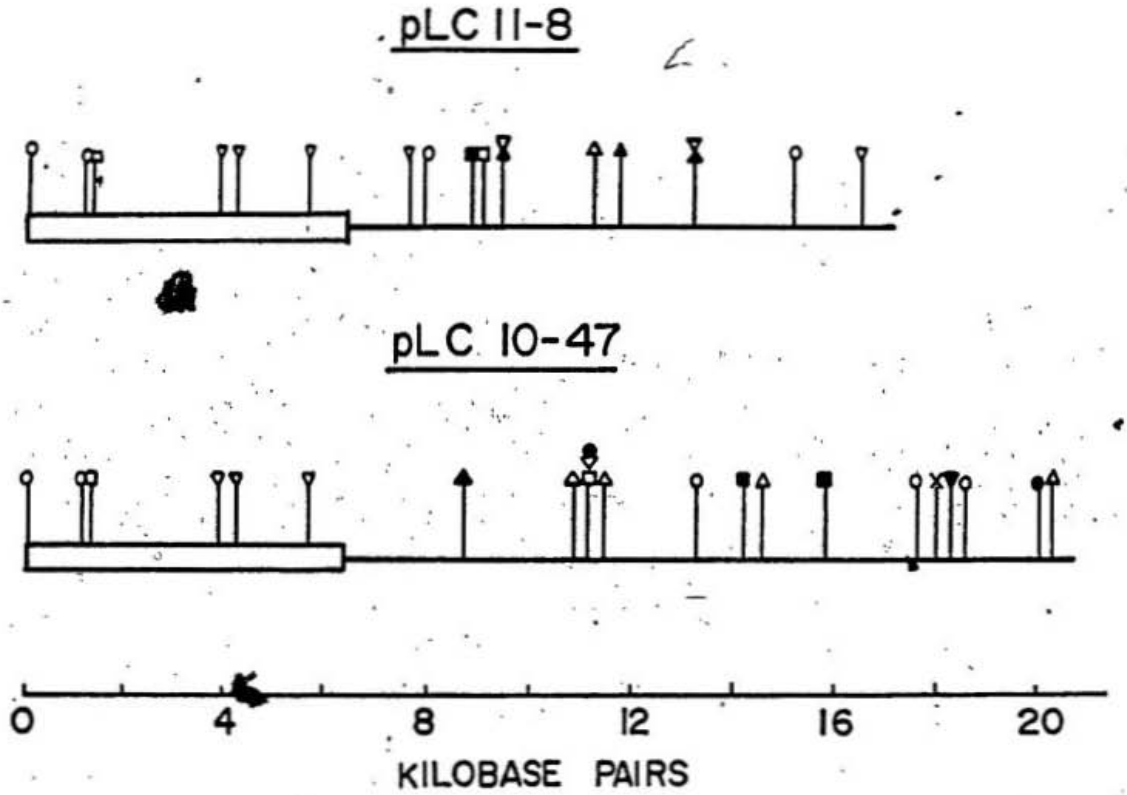


Figure 5-2: Restriction Endonuclease Maps for pLC 10-47 and pLC 11-8 The boxed regions correspond to the cloning vector, and the thin line represents the *E. coli* genomic DNA insert. The symbols used are : Bam HI (⍑), Bgl I (⍑), Bgl II (⍑), Eco RI (⍑), Hind III (⍑), Kpn I (⍑), Pst I (⍑), Pvu II (⍑), Sma I (⍑).

5.2.2. Locations of the Cloning Vector Col E1 on pLC 10-47 and pLC 11-8

The location of the cloning vector in each plasmid was determined by Southern blot analyses using Col E1 as the radioactively labelled probe. The Sma I, Pst I, and Pvu II sites on Col E1 were known (see Figure 4-5) and they were used in the final construction of the restriction maps of pLC 10-47 and pLC 11-8. Comparisons of the restriction endonuclease maps of pLC 10-47 and pLC 11-8 revealed no regions of homology apart from the cloning vector.

5.2.3. Further Searches for Homologous Sequences

The plasmid pLC 11-8 was digested with with several restriction enzymes and analyzed by Southern blotting using pLC 10-47 as the probe. The patterns (Figure 5-3) were identical to those produced by probing with Col E1 alone. In addition, when pLC 10-47, pLC 11-8, and Col E1 were digested with Hae III, Hinf I, Msp I, or Taq I (enzymes that recognize four base pairs) the only fragments in common to pLC 10-47 and pLC 11-8 were also present in the Col E1 digest (Figure 5-4). These results confirmed that, although it has been reported that pLC 10-47 and pLC 11-8 both complement the same enolase mutant, these plasmids do not share any sequences of *E. coli* DNA.

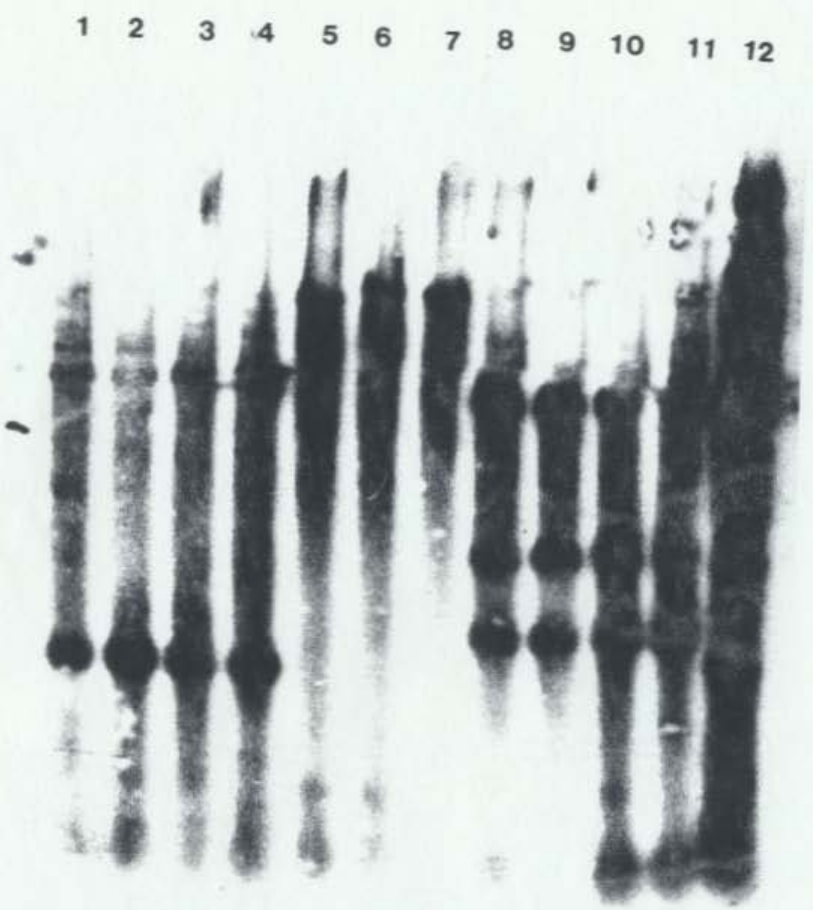


Figure 5-3: Search for Sequence Homology between pLC 10-47 and pLC 11-8

The restriction fragments of pLC 11-8 were separated on 0.8% agarose gel, transfered to Genescreen and then hybridized with radioactively-labelled pLC 10-47.

Lanes 1-11 correspond to restriction digest of pLC 11-8 using the following enzymes : (1) Pst I, (2) Pst I/Bgl I, (3) Pst I/Bgl I/Bam HI, (4) Pst I/Bam HI, (5) Bgl I/Bam HI, (6) Bgl I, (7) Bam HI, (8) Bam HI/Pvu II, (9) Bam HI/Pvu II/Bgl I, (10) Pvu II/Bgl I. (11) Pvu II. Lane 12 corresponds to pLC 10-47 digested with Pvu II.

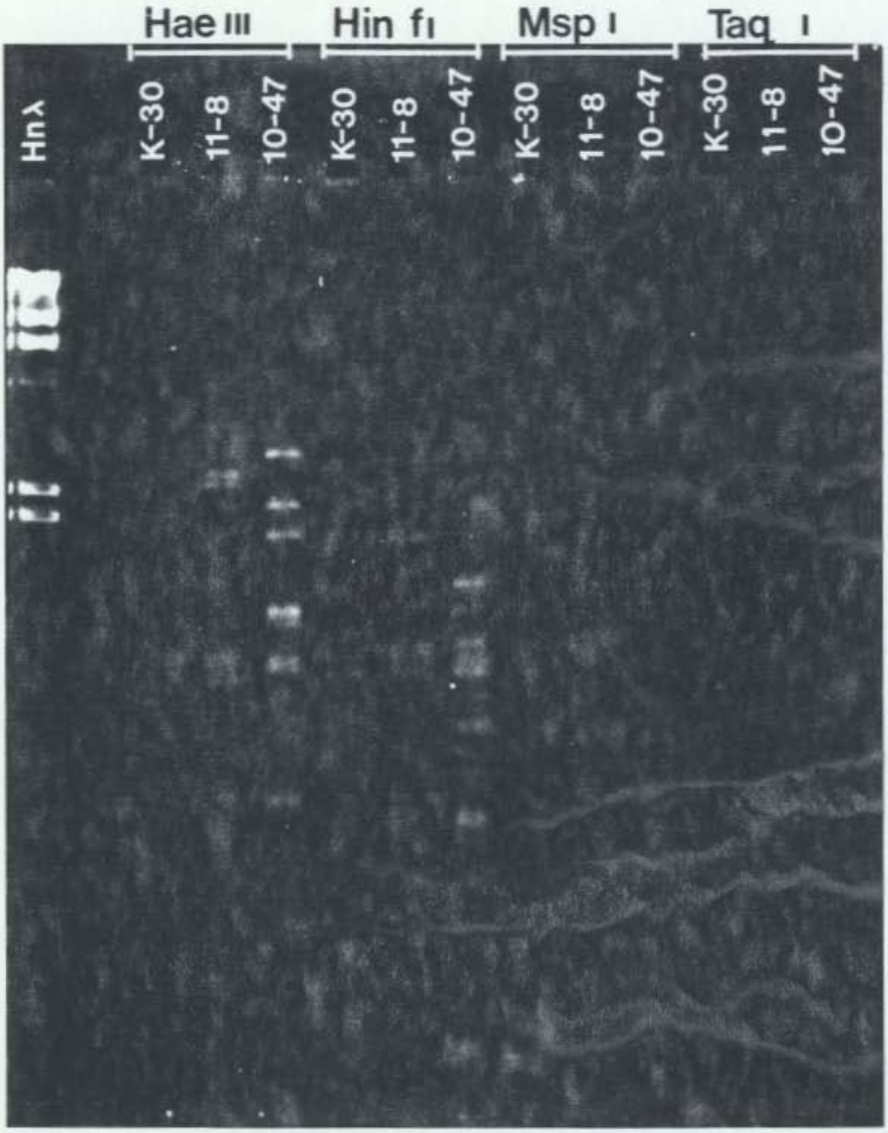


Figure 5-4: Restriction Digest Analysis of pLC 10-47, pLC 11-8 and K30 using Four Base Recognition Endonucleases

5.2.4. Possible Explanations for Results Obtained using Restriction Endonucleases

It was obvious from the above results that pLC 10-47 and pLC 11-8 do not share any common insert from the *E. coli* genomic DNA. This seemed to suggest two possibilities :

1. Like eukaryotes, *E. coli* harbors more than one enolase gene, and pLC 10-47 and pLC 11-8 plasmids carry different enolase genes,
2. There is only one enolase gene in *E. coli*. In this case, either the plasmid pLC 10-47 or pLC 11-8 contains this gene.

5.2.5. Search for Multiple Forms of Enolase in *E. coli*

In order to determine if there is more than one enolase gene in *E. coli*, the enolase gene product was examined. *E. coli* JA200 grown in the presence or absence of glucose were used to investigate if the enolase gene was differentially expressed as is observed in yeast. Sonicated crude extracts from exponentially growing cells were dialyzed and then chromatographed on DEAE-Sephadex. The enzymes were eluted using a linear salt gradient. Figure 5-4 shows that the elution profiles of enolase activity were identical for *E. coli* grown in the presence or absence of glucose. The presence of only one peak of enzyme activity in each case is consistent with there being only one enolase gene in *E. coli* and this gene not being differentially expressed.

The presence of a single peak of enzyme activity indicates that only one plasmid, either pLC 10-47 or pLC 11-8 contains the gene. The ability of one of the plasmids to complement the *eno* mutation could be via another genetic locus and there is a precedent for this phenomenon (Clarke and Carbon, 1978).

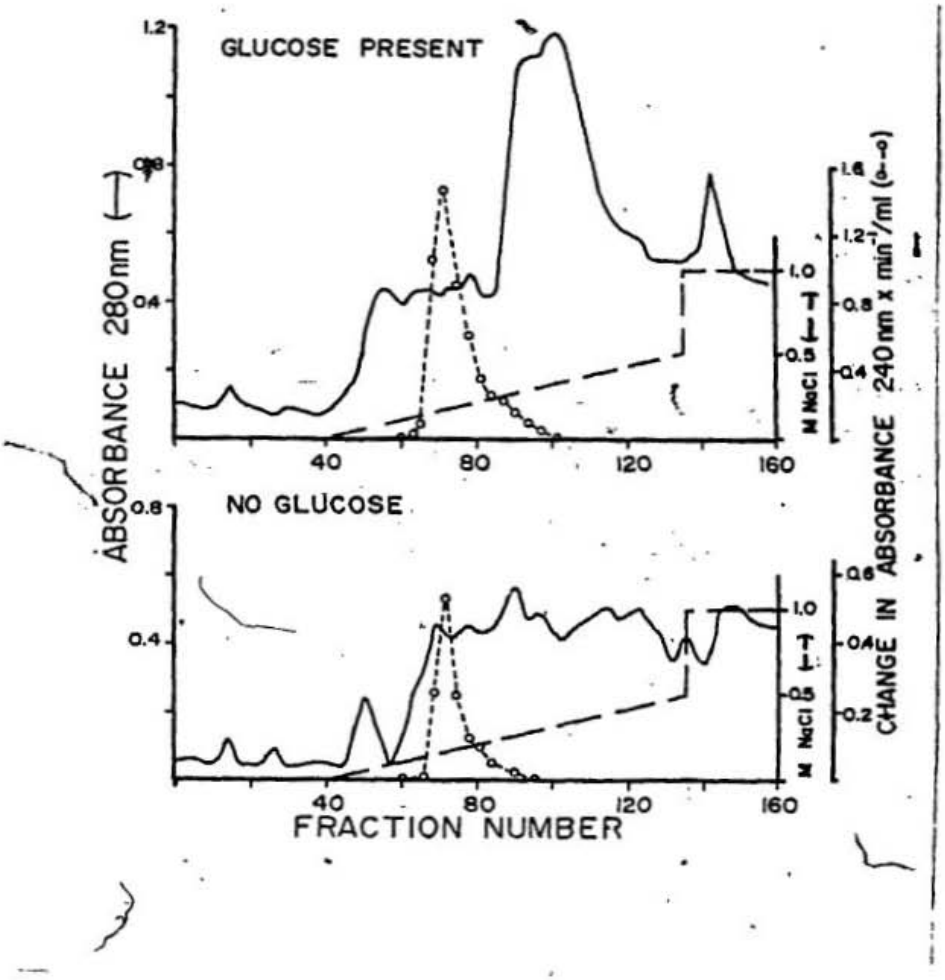


Figure 5-5: Chromatography of *E. coli* JA200 in DEAE-Sephadex

Crude extracts of *E. coli* were chromatographed on DEAE-Sphadex. The protein was eluted using a salt gradient.

5.2.6. Enolase Activity in Extracts of *E. coli* containing pLC 10-47, pLC 11-8 or pLC 16-4

In order to find out which plasmid carries the enolase gene, enolase activity in *E. coli* strains carrying the plasmids pLC 16-4, pLC 10-47 and pLC 11-8, and those without plasmid were assayed. The relative specific enolase activity for strains harboring pLC 16-4 or pLC 11-8 were similar to that of the host strain alone (Table 5-2). However, *E. coli* carrying pLC 10-47 contained 26 fold more enzyme activity than the rest of the strains. The increase in enzyme activity in cells containing pLC 10-47 was attributed to the presence of the enolase gene on the plasmid, and this is equivalent to a gene dosage effect. This result strongly suggests that the hybrid plasmid pLC 10-47 contains the enolase gene.

5.2.7. SDS Polyacrylamide Gel Electrophoresis (PAGE) of Extracts of *E. coli* containing pLC 10-47, pLC 11-8 or pLC 16-4

SDS PAGE analysis of crude protein extracts from *E. coli* JA200 cells carrying pLC 10-47, pLC 11-8, pLC 16-4 or without plasmid showed that the former cells but not the latter three, synthesized large amounts of a protein with an apparent molecular weight of approximately 45,000 daltons, which comigrated with the rabbit enolase (Figure 5-5). As enolase in *E. coli* consists of two identical subunits with a molecular weight of 45,000, this further supports the hypothesis that only pLC 10-47 contains the *E. coli* enolase gene.

Table 5-2: Enolase Activity of *E. coli* Containing Plasmids

E. coli	Plasmid	Glucose	Enolase Activity
JA200	none	-	0.48
JA200	none	+	1.0
JA200	pLC 16-4	+	1.2
JA200	pLC 11-8	+	0.9
JA200	pLC 10-47	+	26.0

The values of the enolase activity are relative to the activity of *E. coli* JA200 which did not carry any plasmid.

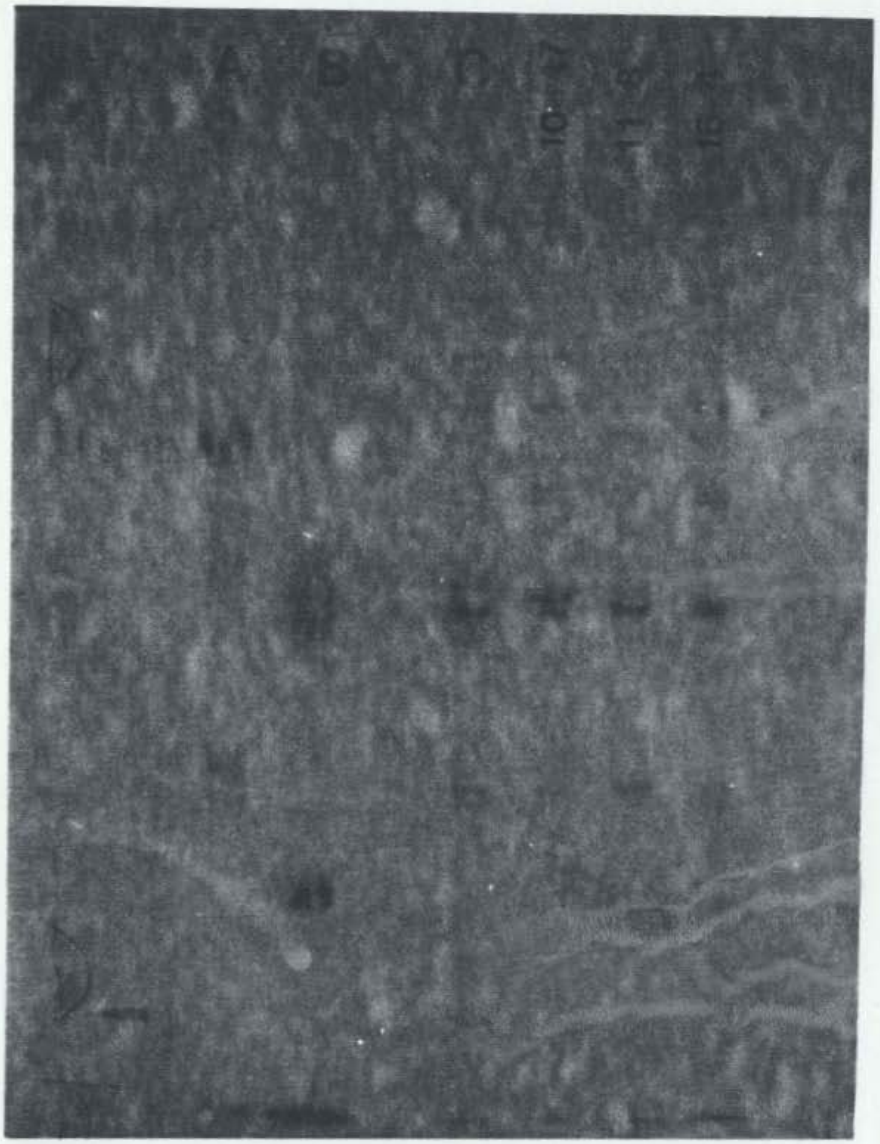


Figure 5-6: SDS PAGE Analysis of Extracts of *E. coli* Carrying pLC 10-47, pLC 11-8 or pLC 16-4

Extracts of *E. coli* containing pLC 10-47, pLC 11-8 and pLC 16-4, and those without plasmids lane C were subjected to electrophoresis in a 10% polyacrylamide-SDS gel and stained for protein. The molecular weight markers used were: lane A, bovine serum albumin (68,000) and chicken lactate dehydrogenase (35,000); lane B, rabbit enolase (45,000) and chymotrypsinogen (25,000).

5.3. Discussion

5.3.1. Enolase Genes

Eukaryotic organisms possess at least two genes for enolase. In mammals these genes are expressed in different tissues (Wold, 1971) and in plants the enolases may be compartmentalized in different organelles (Miernyk and Dennis, 1984). In yeast there are two enolase genes which arose by a recent gene duplication and they are known to be differentially expressed depending on the carbon source (McAlister and Holland, 1982). Since so much is now known about yeast enolase genes, comparisons of these genes with the corresponding gene in *E. coli* may yield insights into some of the differences in the regulation of "housekeeping" genes in prokaryotes and eukaryotes.

5.3.2. How Many Enolase Genes in *E. coli*?

Thomson *et al.* (1979) isolated several hybrid recombinant plasmids that complement *E. coli* glycolytic mutants. Among these are pLC 10-47 and pLC 11-8 which complement a point mutation at the *eno-2* locus. Classical genetic studies (Irani and Maitra, 1974; 1976) indicate that there is only one enolase locus in *E. coli* and there is no evidence for more than one active form of the enzyme (Pfleiderer *et al.*, 1966). These results indicate that both pLC 10-47 and pLC 11-8 should contain the single *E. coli* enolase gene. However, when the extent of DNA sequence homology in the *E. coli* genomic inserts of these two plasmids was compared, none was found.

Protein separation on DEAE-Sephadex using crude extracts from *E. coli*

and the subseqent enolase assay of each fraction showed one sharp peak of enzyme activity. The presence or absence of glucose in the growth media does not seem to effect the expression of enolase in *E. coli*. The enolase was eluted at approximately the same fraction, and the specific enzyme activity was the same for both cases. This suggests that enolase gene in *E. coli* is not differentially expressed as has been observed in yeast, and confirms the conclusion of Fraenkel and Vinopal (1973) that glycolytic genes in *E. coli* are not inducible. These results suggest that there is only one enolase gene and one enolase gene product in *E. coli*.

When enolase activity was measured in cells carrying pLC 10-47, pLC 11-8 and pLC 16-4, only pLC 10-47 exhibited an increase in enzyme activity over cells without any plasmids. This result suggests a gene dosage phemonenon, i.e. the difference in activity is due to the presence of an enolase gene on the plasmid which is translated to enzyme. This was further confirmed by SDS PAGE. A major band corresponding to the molecular weight of rabbit enolase was seen only in extracts of cells containing pLC 10-47. The results shown in this chapter are consistent with the hypothesis that there is only enolase gene in *E. coli*, and that pLC 10-47 but not pLC 11-8 carries this gene.

It is conceivable that pLC 11-8 complements the *eno*-2 mutation with a gene product other than enolase. A possibility would be that a suppressor tRNA is carried by pLC 11-8, and examples of this phenomenon have been reported (Ratzkin and Carbon, 1977; Clarke and Carbon, 1978). However, this would not explain the enhanced enolase activity reported by Thomson *et al.* (1979) in wild

type *E. coli* carrying pLC 11-8. It is probable the pLC 11-8 which was obtained from the *E. coli* Genetic Stock Center is not the same pLC 11-8 that was described by Thomson *et al.* (1979). The Clarke-Carbon library is carried on micro-titre plates, and contamination of wells or incorrect sampling is often a problem (unidentified reviewer for "Gene"). Attempts are now being made by Dr. W.S. Davidson to resolve this point. It was decided not to use pLC 11-8 for the evolutionary studies.

Chapter 6

Evolutionary Studies using pLC 16-4 and pLC 10-47 as Hybridization Probes

6.1. Background

Traditionally, organisms have been categorized into prokaryotes and eukaryotes. The question of the evolutionary origins of these two groups of organisms has been a matter of speculation ever since the fundamental distinction between prokayotic and eukaryotic cells became evident. Various methods have been used to study the phylogenetic relationships among these organisms. Quantitative immunological comparisons of proteins (Howe and Hershey, 1984), protein electrophoresis (Kersters and De Ley, 1975) and comparison of the sequences of conserved molecules are some of the commonest methods used (Wilson *et al.*, 1977). The macromolecules of interest should contain a "record" of their past evolutionary ancestry in order to allow construction of molecular phylogenetic trees. These trees provide a useful framework for analysis of rates of evolution and taxonomic relationships between organisms. Until recently, the study of molecular evolution was usually restricted to proteins. Now the emphasis has shifted to nucleic acids. This is because recombinant DNA techniques available today make sequencing of nucleic acid a relatively easy task.

Before the advent of recombinant DNA technology, evolutionary studies at the nucleic acid level were confined to the entire genome. For instance, information on nucleotide sequence relatedness among enteric bacteria has been derived from studies on hybridization of total bacterial DNAs (Brenner *et al.*, 1971). In this technique genomic DNA from a pair of different organisms was denatured, mixed and allowed to anneal. If there were sufficient conserved regions, the complementary bases will pair forming hybrid double-stranded DNAs. Thermal stabilities of the hybrid DNAs have been used as a measure of the extent of nucleotide sequence divergence among the bacteria, since mismatched bases lower the overall stability of the hybrid. The resistance of heteroduplexes (hybrid DNAs) to thermal denaturation was compared with homoduplexes, and the difference in thermostability between the two duplexes provided an approximate per cent base substitution. However, studies using the entire genome were found to vary widely, even for closely related species (Brenner *et al.*, 1972).

Later, major advances in evolutionary studies came with detailed analyses of repetitive DNA and coding regions, such as histone genes. However, the molecules under investigation are unique to eukaryotes. For instance, there are no histone genes and very little repetitive DNA in prokaryotes. This posed a problem for studying evolutionary relationships between prokaryotes and eukaryotes. Hence, evolutionary studies were extended to molecules that are common to both prokaryote and eukaryotes, e.g. ribosomal RNA (De Smedt and De Ley, 1977). It was the studies on partial sequence analysis of 16S or 16S-like rRNAs (eukaryotic 18S rRNA) which first led Woese and Fox (1977) to propose

the novel concept of three primary kingdoms, i.e. the eubacteria, urkaryotes and archaebacteria. This tripartite division of extant life is incompatible with the conventionally accepted view, and has revolutionalized the basic dichotomic classification of prokaryotes and eukaryotes.

However, one should be cautious because the observations which led to the proposal of the three primary kingdoms were derived only from studies on ribosomes and their components. Although Lake and his co-workers (1982; 1983) originally provided evidence in favor of the three-way split, they have recently challenged this scheme and proposed that eubacteria and halobacteria are more closely related to each other than they are to any other group of organisms (Lake *et al.*, 1985; Lake, 1986). Furthermore, it was suggested that eubacteria and halobacteria (which were originally grouped under archaebacterial kingdom) compose a monophyletic group known as the "photocytes" (Lake, 1986). The whole concept of photocytes and this novel classification has been opposed by other laboratories (Zillig, 1986; Lederer, 1986; Woese *et al.*, 1986). Since a study of the macromolecules associated with translation has not clarified unequivocally the phylogenetic origins of prokaryotes and eukaryotes, molecules other than those which are part of the ribosome should be examined in order to clarify the validity of the "three-way split". Glycolytic enzymes seem to be suitable for such an evolutionary study, as they are highly conserved molecules (Fothergill-Gilmore, 1986) and are universally distributed in nature (Van Valen and Maiorana, 1980). (Criteria to which molecules must conform for these evolutionary studies have been listed in Chapter 1.) Since glycolytic enzymes are ubiquitous and are highly

conserved, one could imagine that their genes would be sufficiently homologous to permit cross-hybridization between species that are distantly related. This Chapter describes the results of testing this hypothesis. The genes encoding *E. coli* glycolytic enzymes, specificly *eno*, *tpi* and *pfk A*, were used in this project to estimate how far in an evolutionary sense it is possible to use glycolytic genes as evolutionary probes.

6.2. Sequence Comparisons of Glycolytic Enzymes and Genes

In order to examine the potential of genes encoding *E. coli* glycolytic enzymes as long range evolutionary probes, the amino acid and nucleotide sequences of TPI, PFK A and enolase from various organisms were compared.

6.2.1. Comparison of TPI Sequences

The amino acid sequence for TPI has been determined from *E. coli* (Pichersky *et al.*, 1984); *B. stearothermophilus* (Artavanis-Tsakonas and Harris, 1980); yeast (Alber and Kawasaki, 1982); coelacanth, chicken, rabbit and human (Lu *et al.*, 1984) (see Figure 6-1).

The corresponding comparisons can also be made at the nucleic acid level for *tpi* from *E. coli* (Pichersky *et al.*, 1984) and yeast (Alber and Kawasaki, 1982), and the messenger RNA for the TPI from chicken muscle (Straus and Gilbert, 1985). These are shown in Figure 6-2.

Comparisons of these sequences (Figure 6-1; Figure 6-2) indicate that there is at least 40 to 50 per cent homology between the prokaryotic and eukaryotic TPIs (Table 6-1).

```
            1                                                  50
E. coli     --MRHPLVMGNWKLNGSRHMVHELVSNLRKELAGVAGCAVAIAPPEMYID
B. stearo.  --  K IIA    MHKTLAEAVQF EDVKGVPP- DEVDSVV   FLFRL
Yeast       -MA TFF G  F     LQSIK I ER NTASIP-ENVE V C  AT L
Coelacanth  AP- KFF G    M  DKKSLG  IQT NAAKVP-FTGEIVC    A L
Chicken     AP- KFF G    M  RKKSLG  IHT EGAKLS- DTE VCGA SI L
Rabbit      APS KFF G    M  RKKNLG  ITT NAAKVP- DTE VC   TA
Human       APS KFF G    M  RLLNLG  I T QGAKVP- DTE VCIG TA L

            51                                                 99
E. coli     MAKREAEGSHIMLGAQNVNLNLSGAFTGETSAAMLKDIGAQYIIIGHSE
B. stearo.  VQAADGT-DLQKI   TMHFABZ  T   V PV    L VT V L
Yeast       TSVSLVKKPQVTV    AY KA       N VDQI  C  KWV L
Coelacanth  F RLKVD-PKFGVA   CYKVSK       I P  I  C VTWV L
Chicken     F RLKLD-AK GVA   CYKVPK       I P  I     AWV L
Rabbit      F RQKLD-PK AVA   CYKVTN       I PG I  C  TWVVL
Human       F RQKLD-EK AA    CYKVTN       I PG I  C  TWVVL

            100                                                149
E. coli     RRTYHKESDELIAKKFAVLKEQGLTPVLCIGETEAENEAGKTEEVCARQ
B. stearo.    HMFA T  TVB  VLAAFTR  I II C  SLE RQ  E DA VSQV
Yeast         S FH      GQ VKHALG  VGVI     LE KK    LE VE
Coelacanth    HVFG      GQ VSHALSE  GV A    KLD R   I  G VFEV
Chicken       HVFG      GQ V HALAE  GVIA    KLD R   I  K VFQE
Rabbit        HVFG      GQ V HALSE  GVIA    KLD R   I  K VFE
Human         HVFG      GQ V HALNE  GVIA    KLD R   I  K VFE

            150                                                199
E. coli     IDAVLKTQGAAAFEGAVIAYEPVWAIGTGKSATPAQAQAVHKFIRDHIA
B. stearo.  EK LAGLTPQ--EVKIIL    L       S  ZB BS CGH  SVVS
Yeast       LN   EEVKD--WTNV V          LA   ED  DI AS  KFL
Coelacanth  TEVIADDVKD--WSKV L           T S Q S EL GKL KWLK
Chicken     TK IADNVKD--WSKV L           T   Q   E  EKL GWLK
Rabbit      TKVIADNVKD--WSKV L           T   Q   E  EKL GWLK
Human       TKVIADDVKD--WSKV L  N        T   Q  EE  EKL GWLK
```

Figure 6-1: The Comparison of Amino Acid Sequences of TPI from *E. coli*, *B. stearothermophilus*, Yeast, Coelacanth, Chicken, Rabbit and Human

Figure 6-1, continued

```
            200                                                    249
E. coli     -KVDANIAEQVIIQYGGSVNASNAAELFAQPDIDGALVGGASLKADAFA
B. stearo.  RLFGPEA AIR      KPD IRDFL   ZZ          EPAS L
Yeast       S LGDKA SELR L   A G  VTFKDKA V  F       PE- V
Coelacanth  EN SETV DS   I    TGATCK  ASE  V  F       P - V
Chicken     TH SDAV VQSR I    TGGDCK  AS H V  F       PE- V
Rabbit      SN SDAV QSTR I    TGATCK  AS   V  F       PE- V
Human       SN SDAV Q TR I    TGATCK   S   V  F       PE- V

            250
E. coli     VIVKAAEAAKQA
B. stearo   QL Q ---GRHE
Yeast       D INS----RN-
Coelacanth  EYKDV----R -
Chicken     D IN ----LH-
Rabbit      D IN ----  -
Human       DLIN ----  -
```

The *E. coli* TPI amino acid sequence is given in full in capital letters. Only differences in the sequences of other TPIs are indicated. Deletions introduced to maximize the homology of sequence comparisons are indicated by ---. B. stearo. is an abbreviation for *Bacillus stearothermophilus*. The reference for each TPI sequence is given in the text.

```
            1                                                   13
E. coli     ATG --- --- CGA CAT CCT TTA GTG ATG GGT AAC TTG AAA
Yeast           GCT --- A   AC  TTC   T   C GGT           T
Chicken     --- GCT CCC A G A G TTC   C     GGT   C       G   G

            14                                                  26
E. coli     CTG AAC GGC AGC CGC CAC ATG GGT CAC GAG CTG GTT TCT
Yeast       T A   . T TC  AAA   A TCC AT  A G   A A T     GAA
Chicken     A       GAG AAG A G  GC TTG GG        C A C CAC  CG

            27                                                  39
E. coli     AAC CTG CGT AAA GAG CTG GCA GGT GTT GCT GGC TGT GCG
Yeast         A T   AAC  CT  CT TCT ATC CCA ---  AA AAT GTC  AA
Chicken      CG     AA  GGC  CC AA  CTC TCG ---   C  A  ACC  A

            40                                                  52
E. coli     GGT GCA ATC GCA CCA CCG GAA ATG TAT ATC GAT ATG GCG
Yeast        T  TT     TGT   T   A  CT  CC   C T A   C TAC T T
Chicken      TG  TT TG  G  G C   T TC   C   C C T    T T   C

            53                                                  65
E. coli     AAG CGC GAA GCT GAA GGC AGC CAC ATC ATG CTG GGT GCG
Yeast       GTC TCT TTG  T  A G AAG CCA   A G    CT G C       T
Chicken     CGC  AG A G CT    T --- GCA A G   T GGA G T  CA   A

            66                                                  78
E. coli     CAA AAC GTG AAC CTG AAC CTG TCC GGC GCA TTC ACG GGT
Yeast            CC T   T     G GCT   T   T   T      C
Chicken      C     TGT T   AG  GTA  C  AAG   T   T       A   A

            79                                                  91
E. coli     GAA ACC TCT GCT GCT ATG CTG AAA GAC ATC GGC GCA CAG
Yeast             C  T   AC CAA A C   G   T G T   T   T A
Chicken       G  T  AGC C A   A     A C       T   T   A   T GCA

            92                                                 104
E. coli     TAC ATC ATC ATC GGT CAC TCT GAA CGT CGT ACT TAC CAC
Yeast        G  G T   T T             C     A A A A T       TT
Chicken      GG G G    C G   C       A   G   G A G CA  GTT TTT
```

Figure 6-2: Nucleotide Sequences Comparison of TPI from *E. coli*, Yeast and Chicken

Figure 6-2, continued

```
                105                                                 117
E. coli     AAA GAA TCT GAC GAA CTG ATC GCG AAA AAA TTC GCG GTG
Yeast       C C     GA      A G T C   T   T G C   G AC  AAG T C
Chicken     GG  G       T   G T       T     C G   G G G   T CAT

                118                                                 130
E. coli     CTG AAA GAG CAG GGC CTG ACT CCG GTT CTG TGC ATC GGT
Yeast       GCT TT   GT     T G C GG  GTC A C T     T
Chicken     GCT CTT  CT G A       C GG  GTC A C GCC   C   T   G

                131                                                 143
E. coli     GAA ACC GAA GCT GAA AAT GAA GCG GGC AAA ACT GAA GAA
Yeast            T TTG  AA       G A G   C   T   G     TTG   T
Chicken      G  AG CTG  A    G GA       T     T   G   G A G

                144                                                 156
E. coli     GTT TGC GCA CGT CAG ATC GAC GCG GTA CTG AAA ACT CAG
Yeast          GTT  AA A A   A T G A     T  C T   G   GAA GTT
Chicken      G  TT  A   AG ACC  C A A   T A T GCT G T  AC GT

                157                                                 169
E. coli     GGT GCT GCG GCA TTC GAA GGT GCG GTT ATC GCT TAC GAA
Yeast       AAG  AC --- ---  GG ACT AAC  TC     G
Chicken     AAG  AC --- ---  GG AGT AAG  T      C T   C   T   G

                170                                                 182
E. coli     CCT GTA TGG GCA ATC GGT ACT GGC AAA TCT GCA ACT CCG
Yeast         A   C       C   T       C.  T TTG G     T       A
Chicken       A   T       T       A       T     A     T       C

                183                                                 195
E. coli     GCT CAG GCA CAG GCT GTT CAC AAA TTC ATC CGT GAC CAC
Yeast        AA G T   T   A  A  A       GCT  C      A A A G TT
Chicken     CAA      T      AG       T G G AAG C G A A   G  TGG

                196                                                 208
E.coli      ATC GCT --- AAA GTT GAC GCT AAC ATC GCT GAA CAA GTG
Yeast       T G     TCC   G T G  GT  AC  G GCT   C AGC G   T
Chicken     C   AAA AGC C C   G TCT  A  GCT G T     C G TC  ACT
```

Figure 6-2, continued

```
            209                                                  221
E. coli     ATC ATT CAG TAC GGC GGC TCC GTA AAC GCG TCT AAC GCT
Yeast        GA   C TTA       T   T       CT       GT AGC       C
Chicken      GG   C ACT   T   A   T   A   C  CT   GT GGC      TG

            222                                                  234
E. coli     GCA GAA CTG TTT GCT CAG CCG GAT ATC GAC GGC GCG CTG
Yeast        TT ACC T C AAG  AC A   G T   G     T   T TTC T
Chicken     AAG         GCC T C      AT     G G   T     TTC   T

            235                                                  247
E. coli     GTT GGT GGT GCT TCT CTG AAA GCT GAC GCC TTC GCA GTA
Yeast         C                 T   G C A   A ---   T  GT  CT
Chicken             G           C   G C A   G ---   T  TG  AT

            248
E. coli     ATC GTT AAA GCT GCA GAA GCG GCT AAA CAG GCT
Yeast        A C   C T   --- --- --- --- G  A C ---
Chicken    T A C   T   A --- --- --- ---    T ---
```

The *E. coli* sequence for TPI is given in full. Only differences in the corresponding sequences of the TPIs of yeast and chicken are indicated. The numbers refer to the corresponding amino acid sequences. Deletions (---) have been introduced where necessary to maximize homology. The reference for each sequence is given in the text.

Table 6-1: Per cent Amino Acid and Nucleotide Sequence Differences of TPIs from *E. coli* (EC), *B. stearothermophilus* (BS), Yeast, Coelacanth (Coel), Chick, Rabbit and Human

	EC	BS	Yeast	Coel	Chick	Rabbit	Human
EC	-		62		65		
BS	63	-					
Yeast	56	66	-		59		
Coel	59	63	52	-			
Chick	57	64	49	24	-		
Rabbit	56	63	49	19	14	-	
Human	57	61	51	22	16	6	-

The values in the upper half of the matrix are per cent nucleotide sequence differences and those in the lower half are per cent amino acid sequence differences.

The above sequences were compiled using a Los Alamos Sequence Analysis computer program

The amino acids at position 96 to 102 are invariant in all the species examined to date. Other regions that show homology at the amino acid level include the sequences from positions 73 to 80, 167 to 179, 211 to 218 and 233 to 242. However, the corresponding regions at the nucleotide level are not quite as conserved. This is due to the degeneracy of the genetic code and in particular to base substitutions at the third position of codons that do not alter the amino acid sequence.

6.2.2. Comparison of PFK Sequences

Phosphofructo-kinase (PFK) [EC 2.7.1.11] is one of the key enzymes involved in the glycolytic pathway. It catalyzes the conversion of fructose-6-phosphate to fructose-1,6-diphosphate using ATP as the phosphorylation source. This enzyme is is subjected to allosteric activation and inhibition by various metabolites. In *E. coli*, PFK is encoded by two genes: *pfk A* and *pfk B*. These genetic loci map at 88 min and 38 min in the *E. coli* chromosome respectively. Eukaryotic PFKs are twice the size of the corresponding prokaryotic enzymes. It has been proposed that mammalian PFKs have evolved by duplication of a prokaryotic *pfk* (Poorman *et al.*, 1984). The amino acid sequences for PFK from *E. coli* (Daldal, 1984), *B. stearothermophilus* (Kolb *et al.*, 1980) and rabbit (Poorman *et al.*, 1984) have been determined and their alignment is shown in (Figure 6-3).

Comparison of the PFK amino acid sequences indicates approximately 30 to 40 per cent homology between the sequences, except for *E. coli* PFK B (Table 6-2). Certain regions of PFK are highly conserved: namely amino acid residues 4

to 11; 111 to 120; 151 to 171; and 281 to 297. However, PFK A and PFK B from *E. coli* showed less than 20 per cent homology. This is consistent with results that indicate that there is no immunological cross-reactivity between these two proteins (Daldal, 1984). There is no obvious homology between PFK B of *E. coli* and the PFKs of rabbit or *B. stearothermophilus*.

6.2.3. Comparison of Enolase Sequences

The amino acid sequence for enolase has been determined for yeast (Holland *et al.*, 1981; Chin *et al.*, 1981a; 1981b) and rat (neuronal and non-neuronal) (Sakimura *et al.*, 1985). Comparison of these sequences (Figure 6-4) indicates that there is at least 60 per cent homology between yeast and rat, and more than 80 per cent matches between neuronal and non-neuronal enolase sequences of rat (See Table 6-3). The corresponding comparisons were also made at the nucleotide level (Figure 6-5).

The amino acid sequences of enolase are more conserved than their corresponding nucleotide sequences. The amino acid sequence homology between the neuronal-specific and non-neuronal enolase of rat is greater than 80 per cent, whereas at the nucleotide level the match is approximately 75 per cent (Table 6-3). The enolase sequence comparisons between yeast and rat exhibit less homology. The amino acid and nucleotide sequences between the two species show approximately 60 per cent homology. It is of interest to note that the enolases within a species are more similar to one another than they are to the enolases in another species. This gives the impression that there has been independent duplications of the enolase genes in rat and yeast. This phenomenon

```
          1                                                  50
E. coli A IKKIGVLTSGGDAPGMNN-IRGVVRSALTEGLEVMGIYDGYLGLYEDRMV
B. stearo. - R      NS    AA  S   K IYH V  Y V H  A  IA-G--
Rab m-N   G A A     Q    AAV A   VGIFT AR FFVHE  Q  VDGGD-
Rab m-C   ----A MNV AP A   AAVRST  IG IQ NR LVVH  FE  AKG---
E. coli B -VR YT  L-APSLDS-AT TPQ----IYPEENCAVPHRCSNPGGGINVA

          51                                                 100
E. coli A QLDRYSVSDM---INRGG-TFLGCARCPEFRDENIRAVAIENLKN-GIDA
B. stearo. NIKKLEVGDVGDI H   - I YT      KT EGEKKG  Q  KH  QG
Rab m-N   HIREATWESVSMMLQL  - VI S   KD  EREG LR AH  VKR  TN
Rab m-C   QIEEAGWSYVGGWTGQ  -SK  SK TLPKKSFE----------------
E. coli B RAIAHLGGSATAIFPAGGA GEHLVSLLA--   VP-  TVEAKDWTRQN

          111                                                150
E. coli A LVVIGDDGSYMGAMRLTEMG----------------------FPCIGLP
B. stearo.     G    Q  KK   H ---------------------  V V
Rab m-N    C   G   LT  DTF SEWSDLLSDLQKAGKITAEEATRSSYLNIV  V
Rab m-C   ---------- E RKQFDELC----------------------I FVVI
E. coli B  H HVEASGEQYRFVMPGAA----------------------LNEDEFR

          151                                                200
E. coli A GTIDNDIKGTDYTIGFFTALSTVVEAIDRLRDTSSSHHAISVVEVMGRYC
B. stearo.      P   F    D   N ID   KI   AT  ERTY I     HA
Rab m-N    S    FC   M   TDS  HRITGIV AITT AQ  QRTF L     H
Rab m-C   A VS  VP S FSV AD  N ICTTC RIKQIAAGTKRVFII T  G
E. coli B QLEEQVLEIESGA LVISGSLPPGVKLEKLTQLI LRKNKGSAASSTVLG

          211                                                250
E. coli A GDLTLAAAIAGGCEFVVVPEVEFSREDLVNEIKAGIAKGKKH---AIVA
B. stearo.   IA WSGL   A TILI  ADYDMN VIARL R HER  ---S II
Rab m-N     Y A VTSLSC ADW FI  CPPDDNWE HLCRRLSETRTRGSRLN II
Rab m-C     Y ATM GL A ADAAYIF EP TIR  QANVEHLVQ M TTVKRGLVL
E. coli B  QG SA L  - NI L KPNQK L A--   RELTQPDDVRKA--- QEI
```

Figure 6-3: Amino acid sequence comparison of PFK from *E.coli*, *Bacillus stearothermophilus* and rabbit

Figure 6-3, continued

```
            251                                                  300
E. coli A   ITEH--------MCDVDELAHFIEKETG--RETRATVLGHIQRGGSPVPY
B. stearo.  VA GV-------GSG  -FGRQ QEA  --F   V      V      TAF
Rab m-N     VA GAIDSNGKPITSEGVKDLVYRRL --TD  V      V    T SAF
Rab m-C     RN KCNENYT---TDFIFNLYS EGK I-FDS KN GHMQ       T F
E. coli B   VNSG--------KAKRVVVSLGPQEAL VDSENCIQ VPPALKSQ T GA

            301                                                  350
E. coli A   DRILASRMGAYAIDLLLAGYGG-----------------RCVGIQ-NEQL
B. stearo.    V SA L   R VE    E GG-----------------        - N
Rab m-N       I G     VE VMA   E --------------TPDTPA  VSLSGN A
Rab m-C       NF TK    K-AMNWMA KIKESYRNGRIFANTPDSG-  LGMRKRA
E. coli B   GDR VGA            TLKLAENASLEEMVRFGVAAGSAATLNQGTR

            351                                                  400
E. coli A   VHHDIIDAIENMKRPFKGDWLDCAEKTVLMISEKGRFLYPETDDRSKNA
B. Stearo.   D  AE LA KHTIDQRMYALSK LSI----------------------
Rab m-N      RLPLMECVQVT DVTKAMDEKRFDEAMKLRGRSFMNNWEVYKL-LLAH
Rab m-C      FQPVTE-LQNQTDFEHRIPKEQWWLKLRPILKILAKYEIDLDTSEHAH
E. coli B   CSH DTQL YAYLSR-----------------------------------

            401
E. coli A   --------------
B. stearo.  --------------
Rab m-N     IRPPAPK-SGSYTV
Rab m-C     LEHISR R  EAT
E. coli B   --------------
```

The *E. coli* PFK A amino acid sequence is given in full in capital letters. Only differences in the sequences of other PKFs are indicated. Deletions introduced to maximize the homology of sequence comparisons are indicated by ---. B. stearo. is an abbreviation for *Bacillus stearothermophilus*, Rab m-N and Rab m-C correspond to N- and C-halves of PFK from rabbit muscle. The reference for each PFK sequence is given in the text.

Table 6-2: Per Cent Amino Acid Sequence Differences between *E. coli*, *Bacillus stearothermophilus* and Rabbit PFKs

	E. coli A	B. stearo.	Rabbit m-N	Rabbit m-C	E. coli B
E. coli A	-				
B. stearo.	59	-			
Rabbit m-N	62	56	-		
Rabbit m-C	68	63	72	-	
E. coli B	84	NH	NH	NH	-

NH denotes no obvious homology between the two sequences compared.

The abbreviations are as in Figure 6-3.

The above sequences were compiled using a Los Alamos Sequence Analysis computer program

```
          1                                                  50
YEAST p46 AVSKVYARSVYDSRGTPTVEVELTTEKGVFRSIVPSGASTGVHEALEMRD
YEAST p8
RAT NNE   SIL IH  EIF  N    D Y A L  AA      IY   L
RAT NSE   SIQ IW  EIL  N    D H A L  AA      IY   L

          51                                                 100
YEAST p46 GDKSKWMGKGVLHAVKNVNDVIAPAFVKANIDVKDQKAVDDFLISLDGTA
YEAST p8            MN  N   N   A     L          L
RAT NNE   N  TRF    SK  EHI KT    L SKKLN VE EKI QLM EM   E
RAT NSE    QRYL     K  DHI ST    LISSGLS VE EKL NLMLE    E

          101                                                150
YEAST p46 NKSKLGANAILGVSLAASRAAPAEKNVPLYKHLADLSKSKTSPYVLPVPF
YEAST p8              M  A          E
RAT NNE       F         VCK GAVG G   R I   A-GNPEV-T    A
RAT NSE       F         VCK GA   DL  R I Q A-GNSD-LI    A

          151                                                200
YEAST p46 LNVLNGGSHAGGALALQEFMIAPTGAKTFAEALRIGSEVYHNLKSLTKKR
YEAST p8
RAT NNE   F  I      NK  M    L V  SS R  M   A       NVI EK
RAT NSE   F  I      NK  M    L V  ES RD M L A   T   GVI DK

          201                                                250
YEAST p46 YGASAGNVGDEGGVAPNIQTAEEALDLIVDAIKAAGHDGKVKIGLDCASS
YEAST p8
RAT NNE     KD T     F   LENK   E LPS  AK  YTDQ V  M V A
RAT NSE     KD T     F   LENS   E VKE  DK  YTE MV  M V A

          251                                                300
YEAST p46 EFFKDGKYDLDFKNPNSDKSKWLTGPQLADLYHSLMKRYPIVSIEDPFAE
YEAST p8                E      VE  M
RAT NNE     YRA       S D- A RYI PD      K FI D          D
RAT NSE     YR        S  A P RCI  D  GA  QDFVRN          D
```

Figure 6-4: Amino Acid Sequence Comparison of Yeast and Rat Enolases

Figure 6-4, continued

```
            301                                                  350
YEAST p46   DDWEAWSHFFKTAGIQIVADDLTVTNPKRIATAIEKKAADALLLKVNQIG
YEAST p8                                    A
RAT NNE     Q   D  QK TA     V G           K AGE SCNC
RAT NSE     Q   A   K TANV     G          ER V E  CNC

            351                                                  400
YEAST p46   TLSESIKAAQDSFAAGWGVMVSHRSGETEDTFNADLVVGLRTGQIKTGAP
YEAST p8                  N
RAT NNE     SVT  LQ CKLAQSN          E      I       C
RAT NSE     SVT A Q CKLAQEN                 I       C

            401
YEAST p46   ARSERLAKLNQLLRIEEELGDNAVFAGENFHHGDKL
YEAST p8                          K  Y
RAT NNE     C       Y  I         SK K   RS RNPLAK
RAT NSE     C       Y  M         EE R   H  RNPSV
```

The amino acid sequence of the p46 yeast enolase is given in full. Differences in the sequences of the other enolases are indicated and identical amino acid residues are left as blanks. Deletions, introduced to maximize the homology of the sequence comparisons, are indicated by ---. RAT NSE and RAT NNE correspond to neuronal-specific and non-neuronal enolase sequence of rat respectively. The reference for each enolase sequences is given in the text.

```
            1                                                    13
YEAST p46   GCT GTC TCT AAA GTT TAC GCT AGA TCC GTC TAC GAC TCC
YEAST p8
RAT NNE     T C A T CTC   G A C C T   C     GAG A    TT
RAT NSE     T   A A CAG   G A C  GG   C C   GAG A    TG

            14                                                   26
YEAST p46   CGT GGT ACC CCA ACC GTC GAA GTC GAA TTA ACC ACC GAA
YEAST p8
RAT NNE       C   G  AT   C       T   G   G   T CTC TA       C
RAT NSE           G  AT   C       G   G   G   T CTC CAT   T  CC

            31                                                   39
YEAST p46   AAG GGT GTT TTC AGA TCC ATT GTC CCA TCT GGT GCT TCT
YEAST p8                                    T                C   C
RAT NNE       A     C C     C T G T GCG   G   C AGC      G   C
RAT NSE       A     C       C G G T GCA       C AG    G   C   C

            40                                                   52
YEAST p46   ACC GGT GTC CAC GAA GCT TTG GAA ATG AGA GAT GGT GAC
YEASR p8                                                    AA
RAT NNE       T   C A   T     G   C C A     C C C     C AA    T
RAT NSE       T   C A   T T   G   C C     G C A  G       G

            53                                                   65
YEAST p46   AAA TCC AAG TGG ATG GGT AAG GGT GTT TTG CAC GCT GTT
YEAST p8                                        A.  A        C
RAT NNE       G A   CGC  TC .     G           C  CA A G
RAT NSE         CAG CGT  AC T A   C   A       C C   A G       G

            66                                                   78
YEAST p46   AAG AAC GTC AAC GAT GTC ATT GCT CCA GCT TTC GTT AAG
YEAST p8      C           A C           G T           C
RAT NNE     G   C   A     T A A ACT       A   T     C G      GC
RAT NSE     G C C   A       AGC AC    C   A   G   C C   A C  GC
```

Figure 6-5: Nucleotide Sequences of Enolase from Yeast and Rat

Figure 6-5, continued

```
                79                                                  91
YEAST p46       GCT AAC ATT GAT GTT AAG GAC CAA AAG GCC GTC GAT GAC
YEAST p8           T A
RAT NNE         AAG   A C G A       GT   G   G G   AAG A T   C C G
RAT NSE         TCA GG  C C TC   G GT   G   G G   AAG C G   C A

                92                                                 104
YEAST p46       TTC TTG ATT TCT TTG GAC GGT ACT GCC AAC AAA TCC AAG
YEAST p8           T G          T      C          G
RAT NNE         C G A    C GAG A        C   A  AG  T       T
RAT NSE         C G A  C G GAG      T  G      AG

                105                                                117
YEAST p46       TTG GGT GCT AAC GCT ATC TTG GGT GTT TCT TTG GCT GCT
YEAST p8                                        C   C A    C
RAT NNE           T       A   T   C     C     A   G   C C       TC
RAT NSE           T   G   C   T   C     C         G   C C     C  TG

                118                                                130
YEAST p46       TCC AGA GCT GCC GCT GCT GAA AAG AAT GTC CCA TTA TAC
YEAST p8        G T       C   T                   C         G
RAT NNE          G   AG      GT   C  TG  GG     GGG   G   G C T
RAT NSE          G   AG      GG   A   C   G     G C T G   C C C   T

                131                                                143
YEAST p46       AAG CAC TTG GCT GAC TTG TCT AAG TCC AAG ACC TCT CCA
YEAST p8        C A
RAT NNE         CGT    A T   C        G C GGC AA  CCT GAG GTC A T
RAT NSE         CGC     A T     C A C   G   GG  AA  TCC GA  CTC A T

                144                                                158
YEAST p46       TAC GTT TTG CCA GTT CCA TTC TTG AAC GTT TTG AAC GGT
YEAST p8
RAT NNE         CTG CCG G C      C  TTC AAT G    T  ---  ---      C
RAT NSE         CTG CCC GT    G  CC TTT AAT G    T  ---  ---
```

Figure 6-5, continued

```
            157                                                   169
YEAST p46   GGT TCC CAC GCT GGT GGT GCT TTG GCT TTG CAA GAA TTT
YEAST p8                                                       C
RAT NNE        T   T       C AAC AAG       C A         G   C
RAT NSE     C  T   T       G AAC AAG       C A     G   G   C

            170                                                   182
YEAST p46   ATG ATT GCT CCA ACT GGT GCT AAG ACC TTC GCT GAA GCT
YEAST p8                                                       C
RAT NNE        C CTG   T GTG   G   A TCC T T     CGG        C
RAT NSE        C CTC     GTG       G    G    T CGG   T    C

            183                                                   195
YEAST p46   TTG AGA ATT GGT TCC GAA GTT TAC CAC AAC TTG AAG TCT
YEAST p8    A
RAT NNE     A   C C       A G A   G                 C       AAC
RAT NSE     A   C   C     G G     G   G         CA C C      GGG

            196                                                   208
YEAST p46   TTG ACC AAG AAG AGA TAC GGT GCT TCT GCC GGT AAC GTC
YEAST p8                          T
RAT NNE     G C  T    A G    AG       G AAA GAC   ACC   T   G
RAT NSE     G C  T      G C  AG       C AAG GA      AC       G

            209                                                   221
YEAST p46   GGT GAC GAA GGT GGT GTT GCT CCA AAC ATT CAA ACT GCT
YEAST p8                                                    C
RAT NNE        T  G       A T C   A   T       C  TG GAG AAC
RAT NSE     G  T       C   C T     C   C   T   C  TG GAG AAC

            222                                                   234
YEAST p46   GAA GAA GCT TTG GAC TTG ATT GTT GAC GCT ATC AAG GCC
YEAST p8                                                       T
RAT NNE     A       A C     G C   C A CCG TCT   C    T GCA AAG
RAT NSE     AGC             G C   G G AAG   A   C    T G C AAG
```

Figure 6-5, continued

```
                235                                                 247
YEAST p46       GCT GGT CAC GAC GGT AAG GTC AAG ATC GGT TTG GAC TGT
YEAST p8
RAT NNE           C   C T   ACT  AC C     T GTC       C A     T GTG
RAT NSE           C   C T   ACG  AA      A G GT    T     ATG   T GTG

                248                                                 260
YEAST p46       GCT TCC TCT GAA TTC TTC AAG GAC GGT AAG TAC GAC TTG
YEAST p8
RAT NNE           G   C   G       A   G   CT   C       T     C
RAT NSE           G       G   T   A  CGC   T   C   A

                261                                                 273
YEAST p46       GAC TTC AAG AAT CCA AAC TCT GAC AAA TCC AAG TGG TTG
YEAST p8                    C     G A
RAT NNE                   TC      G T ---   T GCC AG   CG    AC A C
RAT NSE           T       TC    T GCT ---     CCT      CGA   C A C

                274                                                 286
YEAST p46       ACT GGT CCT CAA TTG GCT GAC TTG TAC CAC TCC TTG ATG
YEAST p8              GTC G               A
RAT NNE           A CCC GAC   G C     C     C       A G       C   C
RAT NSE             GGG GAC   G C T  GG   CA C C       G GA    T G C

                287                                                 299
YEAST p46       AAG AGA TAC CCA ATT GTC TCC ATC GAA GAT CCA TTT GCT
YEAST p8
RAT NNE           GAC         G G   G       G           C      AC
RAT NSE         CC  AC   T   T G G           T       C       C  AC

                300                                                 313
YEAST p46       GAA GAT GAC TGG GAA GCT TGG TCT CAC TTC TTC AAG ACC
YEAST p8
RAT NNE         C G   C           T       CAG A G     ACA GCT   T
RAT NSE         C G               C         C A G     ACA GCC  AT
```

Figure 6-5, continued

```
            314                                                 326
YEAST p46   GCT GGT ATT CAA ATT GTT GCT GAT GAC TTG ACT GTC ACC
YEAST p8      C                                                 T
RAT NNE       A   C   C   G G G   G  GG         C C   A   G
RAT NSE     TC   C   C   G   A   G  G          C     G   G

            327                                                 339
YEAST p46   AAC CCA AAG AGA ATT GCT ACC GCT ATC GAA AAG AAG GCT
YEAST p8        GCT
RAT NNE       T     C G   C   C  AG     GCA  GC G A     T C
RAT NSE       C     C C   C  AG CGG   A G G   G G   C     C

            340                                                 352
YEAST p46   GCC GAC GCT TTG TTG TTG AAG GTC AAC CAA ATC GGT ACC
YEAST p8      T                           T
RAT NNE     TG  A   TGC C C C   C C   A   G       G   T   C T T
RAT NSE     TG  A   TG      C   C C               G       C T A

            353                                                 365
YEAST p46   TTG TCT GAA TCC ATC AAA GCT GCT CAA GAC TCT TTC GCT
YEAST p8                        G
RAT NNE     G   A C   G   T C G C G   G TG  A G CTG G C CAG T C
RAT NSE     G C A A     G       C     G TGC A G CTG G C CAG  AG

            366                                                 378
YEAST p46   GCC GGT TGG GGT GTT ATG GTT TCC CAC AGA TCT GGT GAA
YEAST p8      AAC           C
RAT NNE     AAT   C           C       G       T C         AG   G
RAT NSE     AA    C       G           G AGT   T C C        A

            379                                                 391
YEAST p46   ACT GAA GAC ACT TTC AAT GCT GAC TTG GTC GTC GGT TTG
YEAST p8                                        T
RAT NNE           G               T    C     C    G   G   G C C
RAT NSE       C   G       G       T    A     C C  A   G   A C
```

Figure 6-5, continued

```
            392                                                 404
YEAST p46   AGA ACT GGT CAA ATC AAG ACT GGT GCT CCA GCT AGA TCC
YEAST p8
RAT NNE     T C       G   G                   C   C TGC C     T
RAT NSE     T T   A       G                   C     TGC       T

            405                                                 417
YEAST p46   GAA AGA TTG GCT AAA TTG AAC CAA TTG TTG AGA ATC GAA
YEAST p8                    G
RAT NNE       G C C C     C   G  AC   T   G A C C T           G
RAT NSE         C T C     G   G  AC       G C C A     G   T

            418                                                 430
YEAST p46   GAA GAA TTG GGT GAC AAC GCT GTT TTC GCT GGT GAA AAC
YEAST p8                        G       C  A    C
RAT NNE       G     C     C AG    A   C AAG   T   C   C AGG TC
RAT NSE       G   G       G   G G G     CGC       G   A

            431
YEAST p46   TTC CAC CAC GGT GAC AAA TTA
YEAST p8                      G   G
RAT NNE         AGG A   CCC CTG GCC AAG
RAT NSE                       C G
```

The nucleotide sequence of the yeast p46 enolase gene is given in full. Only differences in the corresponding sequences of the yeast p8 enolase and rat neural-specific (NSE) and non-neuronal (NNE) are indicated. The numbers refer to the corresponding amino acid sequences. Deletions (---) have been introduced where necessary to maximize homology.

Table 6-3: Per Cent Amino Acid Sequence and Nucleotide Sequence Differences between Yeast and Rat Enolases

	Yeast p46	Yeast p8	Rat NNE	Rat NSE
Yeast p46	-	5	41	40
Yeast p8	4	-	42	40
Rat NNE	39	39	-	25
Rat NSE	39	39	17	-

The values in the upper half of the matrix are per cent nucleotide sequence differences and those in the lower half are per cent amino acid sequence difference.

The above sequences were compiled using a Los Alamos Sequence Analysis computer program

could also be explained by postulating concerted evolution of these genes (Zimmer *et al.*, 1980).

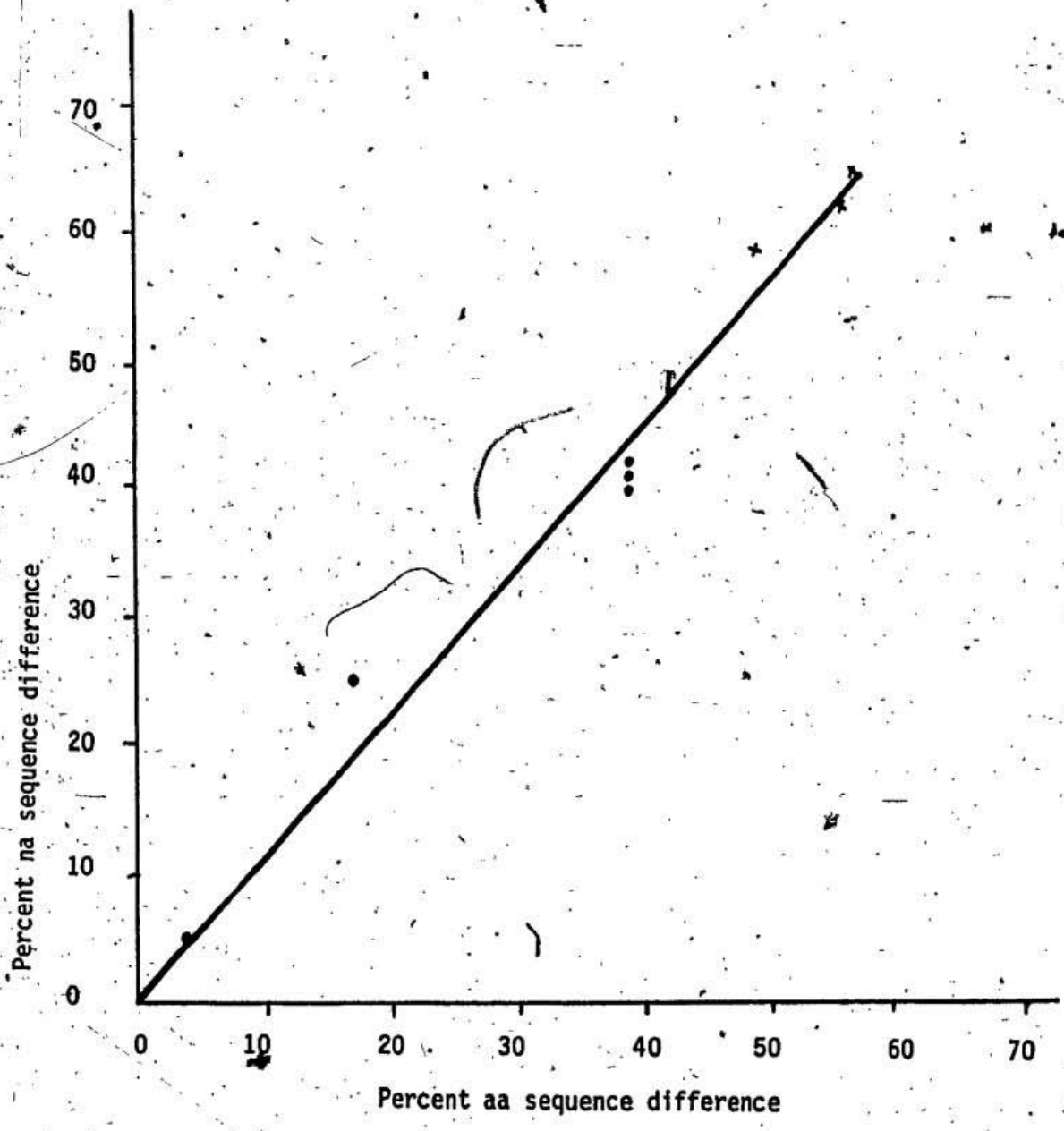
The per cent amino acid sequence difference of TPI between yeast and mammals is 49-51, whereas the yeast and rat enolases differ by approximately 40 per cent. It can be concluded from this comparison that enolase evolves even more slowly than TPI.

6.3. Potential of pLC 18-4 and pLC 10-47 as Evolutionary Probes

The amount of sequence divergence at the nucleic acid level parallels the per cent amino acid sequence difference very well (Figure 6-6). This is somewhat surprising as it is possible to have a situation where there are no amino acid sequence differences yet greater than 30 per cent difference in the nucleotide sequences because of the degeneracy of the genetic code. However, for hybridization purposes it is not the overall amount of sequence homology that is important but rather the existance of homology in sufficiently long stretch of DNA. As an example, the amino acid sequences of yeast and *E. coli* TPIs are identical at residues 234 to 241 but in this region at the nucleic acid level there is a stretch of 22 nucleotides in which there are 2 substitutions. The theoretical temperature at which these sequences would hybridize in 6X SSC is 45^0C. (See Meinkoth and Wahl (1984) for a review of the parameters that influence hybridization of nucleic acid probes and their detection of sequences of DNA fixed to solid supports.)

Figure 6-6: Comparison of the Amount of Sequence Differences at the Amino Acid and Nucleic Acid Level

closed circles represent data obtained from enolase sequence comparison and crosses represent those from TPI


70
60
50
40
30
20
10
0
Percent na sequence difference
0
10
20
30
40
50
60
70
Percent aa sequence difference

It is most unusual to be able to predict the optimum conditions for hybridization and washing when using a piece of DNA as a heterologous probe and these parameters are commonly determined by trial and error. Cloned genes for a variety of proteins have been used as heterologous probes to identify and isolate the corresponding genes from different organisms: e.g. a *Drosophila* actin gene for sea urchin actin genes (Schuler and Keller, 1981; Durica *et al.*, 1980); a yeast cytochrome c gene for rat cytochrome c genes (Scarpulla *et al.*, 1981); a sea urchin histone gene for chicken histone genes (Engel and Dodgson, 1981); the chicken glyceraldehyde-3-phosphate dehydrogenase gene for the corresponding yeast genes (Musti *et al.*, 1983); human globin cDNA clones for old world monkey globin genes (Martin *et al.*, 1983); the calmodulin gene from the electric eel for a chicken calmodulin gene (Putkey *et al.*, 1983); a rat cDNA for preproinsulin for the chicken insulin gene (Perler *et al.*, 1980); the *Klebsiella nif* genes for *nif* genes from a blue-green alga (Mazur *et al.*, 1980); and a *Drosophila* gene for the major heat-shock protein was used to detect related sequences in mouse and yeast genomic DNA (Moran *et al.*, 1983). Based on the observations of Moran *et al.* (1983) it was decided to use two sets of hybridization conditions in this study: the first was considered "high" stringency (40^0C, 50 per cent formamide, 5X SSC) and the other was "low" stringency (37^0C, 30 per cent formamide, 5X SSC). Under less stringent conditions than the latter Moran *et al.* (1983) found that there was too much non-specific hybridization.

6.4. Results

6.4.1. Characterization of Genomic DNAs

Genomic DNAs were isolated from organisms representative of eubacteria, archaebacteria and eukaryotes. Electrophoresis of these DNA preparations in agarose gels indicated that the average sizes of the genomic DNAs were greater than 25 kilobasepairs. The DNA preparations from *Shigella sonnei* and *Halobacterium volcanii* gave bands on the gel that had the appearance of plasmids (Figure 6-7). There was the suggestion that the DNA prepared from *Serratia marcescens* and *Klebsiella pneumoniae* also contained plasmids.

6.4.2. Hybridization Studies: Enterobacteriaceae DNAs

The genomic DNAs obtained from members of the Enterobacteriaceae were digested with Eco RI or Hind III. The fragments that were generated were separated by gel electrophoresis (Figure 6-8), denatured *in situ*, transfered to Biodyne A nylon membrane and hybridized with radioactively-labelled K30, pLC 10-47 or pLC 16-4 under two different conditions ("high" or "low" stringency).

6.4.2.1. Hybridization using K30 cloning vector as probe

K30 is the Col E1 plasmid that was used as the cloning vector for the *E. coli* genomic library (Clarke and Carbon, 1976) from which pLC 10-47 and pLC 16-4 were derived. The reason for using K30 as a probe was to determine if there were sequences in the genomic DNA preparations that would cross-react with the cloning vector in addition to the *E. coli* inserts of pLC 10-47 and pLC 16-4.

K30 hybridized with the DNAs prepared from *Shigella*, *Klebsiella* and

Figure 6-7: Analysis of Prokaryotic Genomic DNAs by Gel Electrophoresis

2 ug of each genomic DNA was loaded on a 0.8% agarose gel and electrophoresis was carried out at 100 V.

(1) *Escherichia coli*; (2) *Shigella sonnei*; (3) *Serratia marcesteens*; (4) *Klebsiella pneumoniae*; (5) *Pseudomonas putida*; (6) *Bacillus subtilis*; (7) *Clostridium perfringens*; (8) *Micrococus lysodeikticus*; (9) *Halobacterium volcanii*; (10) *Halobacterium halobium*; (11) *Halobacterium marismortui*; (12) *Halobacterium trapanicum*.; (13) λ cut with Hind III; (14 and 15) *Escherichia*; (16) *Shigella*; (17) *Salmonella typhimurium*; (18) *Citrobacter freundii*; (19) *Serratia marcescens*; (20) *Enterobacter aerogenes*; (21) *Klebsiella pneumoniae*; (22) *Edwardsiella tarda*; (23) *Proteus mirabilis*; (24) *Erwinia carotovora*.

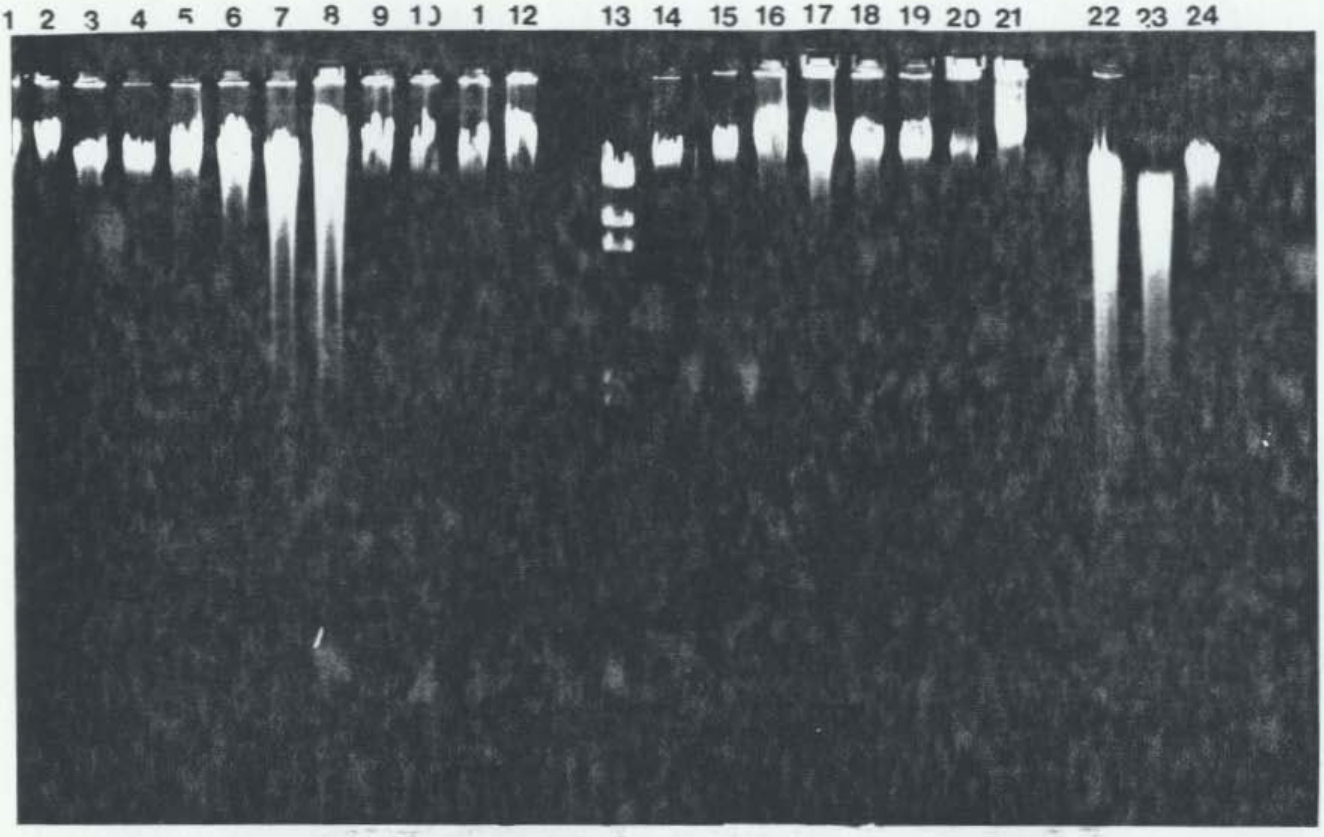
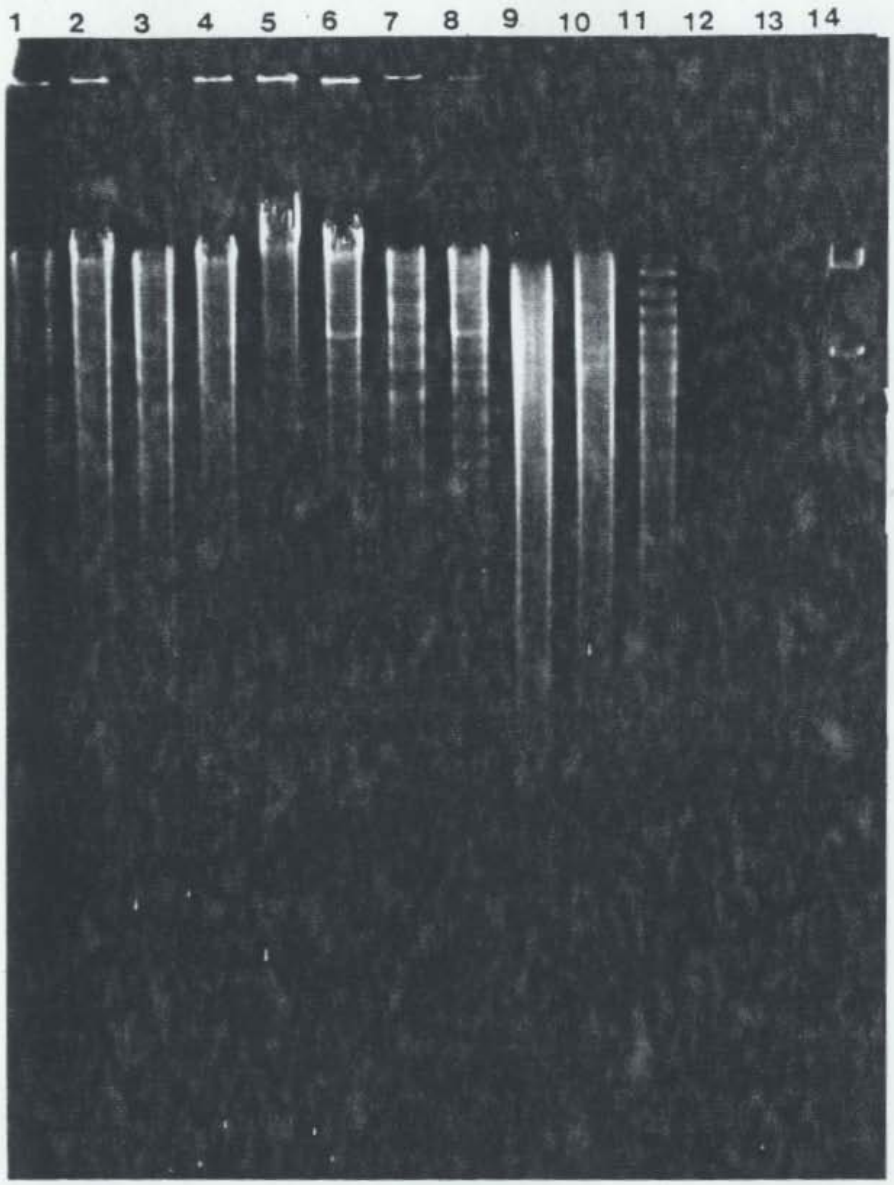

1 2 3 4 5 6 7 8 9 10 1 12 13 14 15 16 17 18 19 20 21 22 23 24

Figure 6-8: Electrophoretic Pattern of Genomic DNAs from Enterobacteriaceae Digested with Hind III

Restriction fragments were separated on a 0.8% agarose gel at 100 V. (1) *Escherichia coli*, (2) *Shigella sonnei*, (3) *Salmonella choleraesuis*, (4) *Salmonella typhimurium*, (5) *Citrobacter freundii*, (6) *Serratia marcescens*, (7) *Enterobacter aerogenes*, (8) *Klebsiella pneumoniae*, (9) *Edwardsiella tarda*, (10) *Yersinia enterocolitica*, (11) *Proteus mirabilis*, (12) *Erwinia carótovora*. Lane 13 contained no DNA and lane 14 corresponds to λ DNA cut with Hind III.


1 2 3 4 5 6 7 8 9 10 11 12 13 14

Serratia (Figure 6-9 and Figure 6-10). Similar patterns were seen irrespective of hybridization stringency except that the bands were more intense at low stringency. The interaction with the *Shigella* DNA was strongest whereas the signals with *Serratia* and *Klebsiella* were less, but of equal intensity. The hybridization of K30 with *Shigella*, *Klebsiella* and *Serratia* genomic preparations corresponds to interactions with plasmids in these strains. Plasmids have been isolated and partially characterized from these strains (W.S. Davidson, unpublished observations).

6.4.2.2. Hybridization with pLC 16-4 as Probe

Hybridization of the genomic DNA from Enterobacteriaceae at low stringency with the radioactively-labelled plasmid pLC 16-4 showed a wide range of cross-reactions (Figure 6-11 B and C). The following order of hybridization was observed: *Shigella* > *Escherichia* > or = *Salmonella* = *Citrobacter* > *Serratia* = *Klebsiella*. Some hybridization was occassionally observed with *Yersinia*, *Proteus* and *Enterobacter*. However, no hybridization (or occassionally non-specific hybridization) was seen with *Edwardsiella* or *Erwinia*.

The extent of cross-reaction seen with *Shigella* was probably due to the presence of a plasmid with sequences related to Col E1. Some of the cross-hybridization seen with *Serratia* and *Klebsiella* was also attributed to the presence of Col E1 sequence-related plasmids.

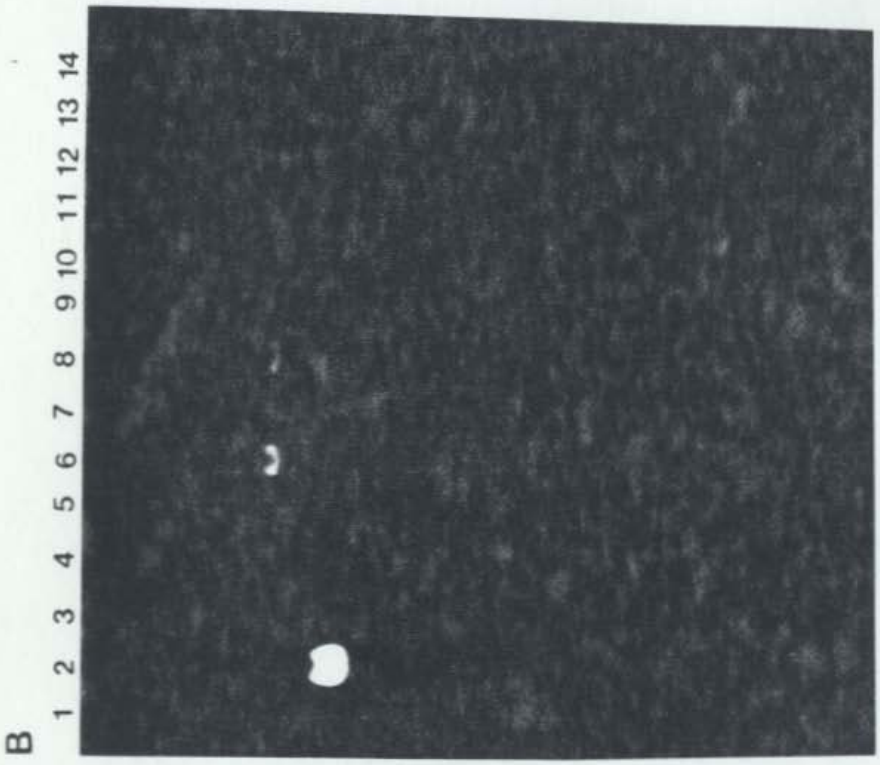
Hybridization at high stringency showed an overall reduction in cross-reactivity of most DNAs (Figure 6-11 A). The bands in the *Shigella* lane were still prominent and hybridization was evident with *Salmonella sp.*, *Citrobacter*,

Figure 6-9: Southern Blot of Enterobacteriaceae DNA probed with K30 at High Stringency

Restriction fragments generated by Eco RI (A) or Hind III (B) were separated on a 0.8% agarose gel at 100 V.

(1) *Escherichia coli*, (2) *Shigella sonnei*, (3) *Salmonella choleraesuis*, (4) *Salmonella typhimurium*, (5) *Citrobacter freundii*, (6) *Serratia marcescens*, (7) *Enterobacter aerogenes*, (8) *Klebsiella pneumoniae*, (9) *Edwardsiella tarda*, (10) *Yersinia enterocolitica*, (11) *Proteus mirabilis*, (12) *Erwinia carotovora*.

Lane 13 contained no DNA and lane 14 corresponds to λ DNA cut with Hind III.


B
1 2 3 4 5 6 7 8 9 10 11 12 13 14

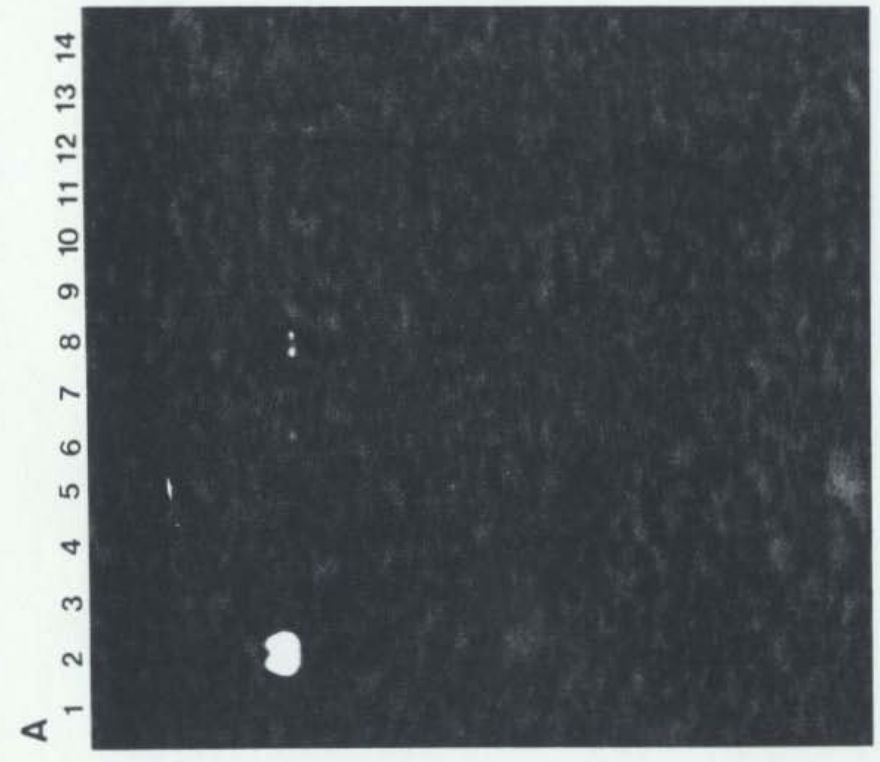
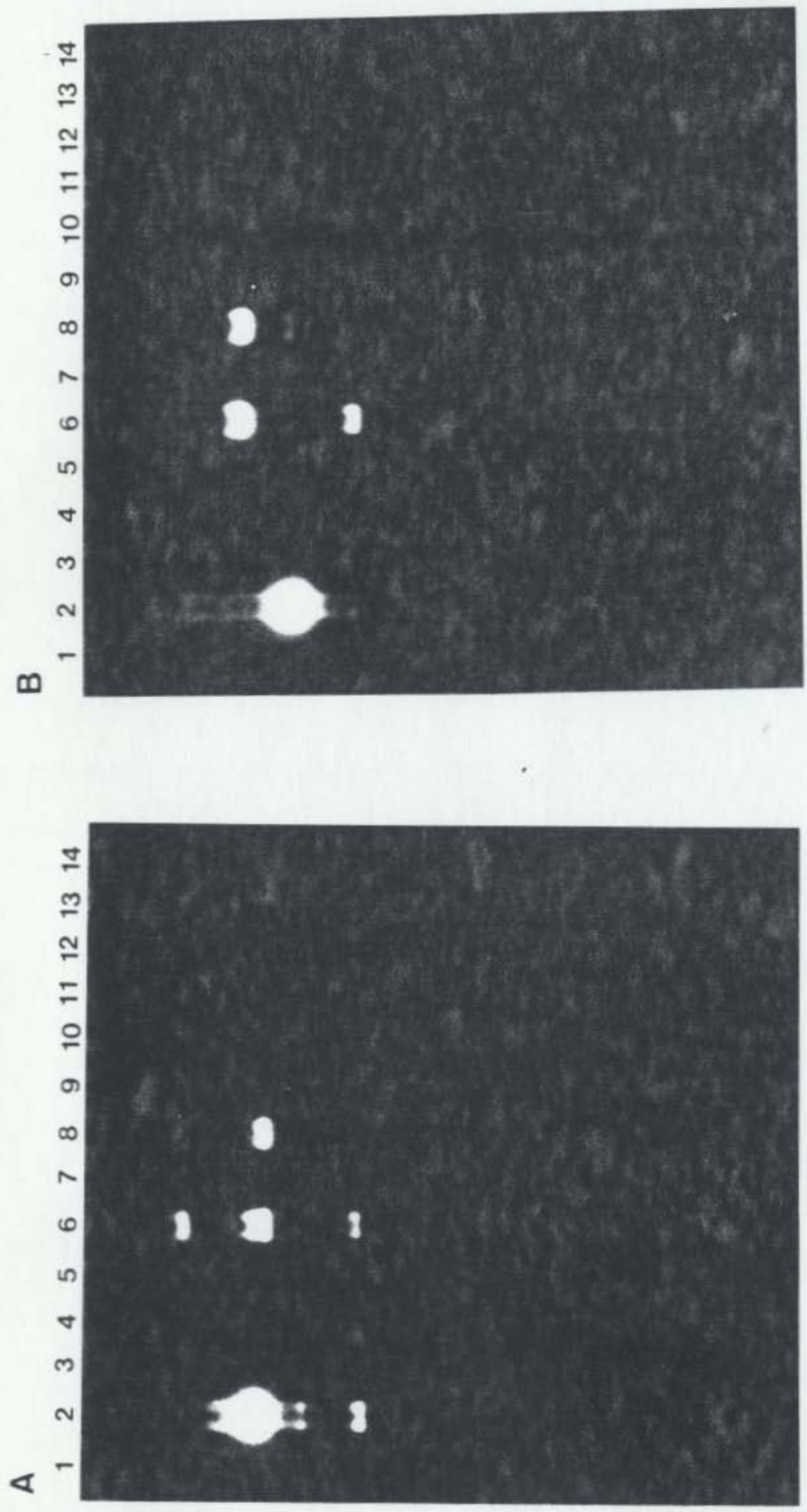

A
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 6-10: Southern Blot of Enterobacteriaceae DNA probed with K30 at Low Stringency

Restriction fragments generated by Eco RI (A) or Hind III (B) were separated on a 0.8% agarose gel at 100 V.

(1) *Escherichia*, (2) *Shigella*, (3) *Salmonella choleraesuis*, (4) *Salmonella typhimurium*, (5) *Citrobacter*, (6) *Serratia*, (7) *Enterobacter*, (8) *Klebsiella*, (9) *Edwardsiella*, (10) *Yersinia*, (11) *Proteus*, (12) *Erwinia*.

Lane 13 contained no DNA and 14 corresponds to λ DNA cut with Hind III.


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Serratia and *Klebsiella* but these bands were less intense than those obtained when hybridization was carried out at low stringency. There was very faint or no hybridization seen with *Yersinia*, *Proteus*, *Enterobacter*, *Edwardsiella* or *Erwinia*. Hybridization at high stringency with pLC 16-4 proved unsuccessful in detecting related sequences in all members of the Enterobacteriaceae. Therefore, it was decided that only low stringency hybridization conditions would be used in the following studies.

6.4.2.3. Hybridization with pLC 10-47 as Probe

Low stringency hybridization conditions were used to attempt to detect sequences related to pLC 10-47 in members of the Enterobacteriaceae. The results were similar to those seen when pLC 16-4 was used as the probe. Various amounts of hybridization were seen with DNA from *Escherichia*, *Shigella*, *Salmonella*, *Citrobacter*, *Serratia* and *Klebsiella* (Figure 6-12) but there was little or no cross-reaction with *Yersinia*, *Proteus*, *Enterobacter*, *Edwardsiella* or *Erwinia* DNA.

6.4.3. Hybridization Studies: Prokaryotic and Eukaryotic DNAs

Genomic DNA from representative members of the Enterobacteriaceae, *Bacillus subtilis*, *Pseudomonas putida*, *Clostridium perfringens*, *Micrococcus lysodeikticus*, *Halobacterium sp.*, chicken and human were used in a comparative hybridization study. These DNAs were cut with the restriction enzymes Eco RI or Hind III and then subjected to Southern blot analysis with K30, pLC 10-47 or pLC 16-4 as probes. Hybridization was carried out at low stringency. There was no hybridization of any of these genomic DNA preparations with K30.

Figure 6-11: Southern Blot of Enterobacteriaceae DNA Probed with pLC 16-4

Restriction fragments generated by Eco RI (A and B) or Hind III (C) were separated on a 0.8% agarose gel at 100 V.

(1) *Escherichia*, (2) *Shigella*, (3) *Salmonella choleraesuis*, (4) *Salmonella typhimurium*, (5) *Citrobacter*, (6) *Serratia*, (7) *Enterobacter*, (8) *Klebsiella*, (9) *Edwardsiella*, (10) *Yersinia*, (11) *Proteus*, (12) *Erwinia*.

Lane 13 contained no DNA and lane 14 corresponds to λ DNA cut with Hind III.

Blot A was probed at high stringency whereas B and C were probed at low stringency.

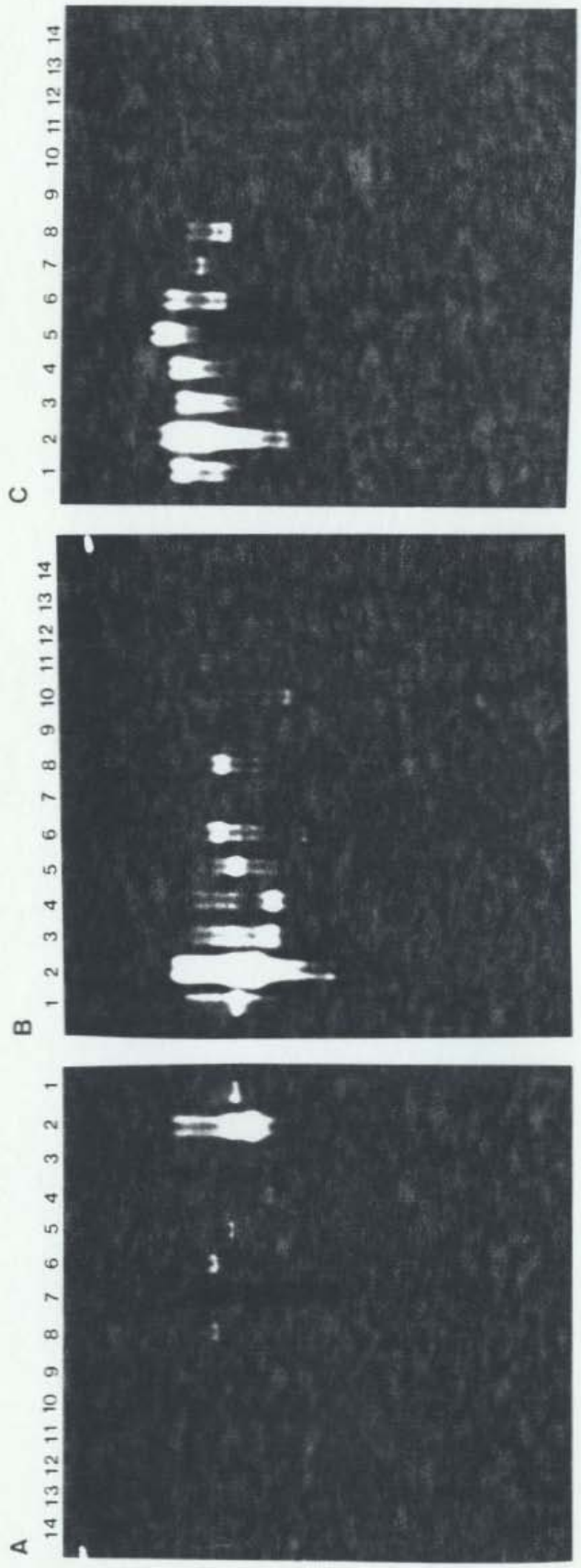
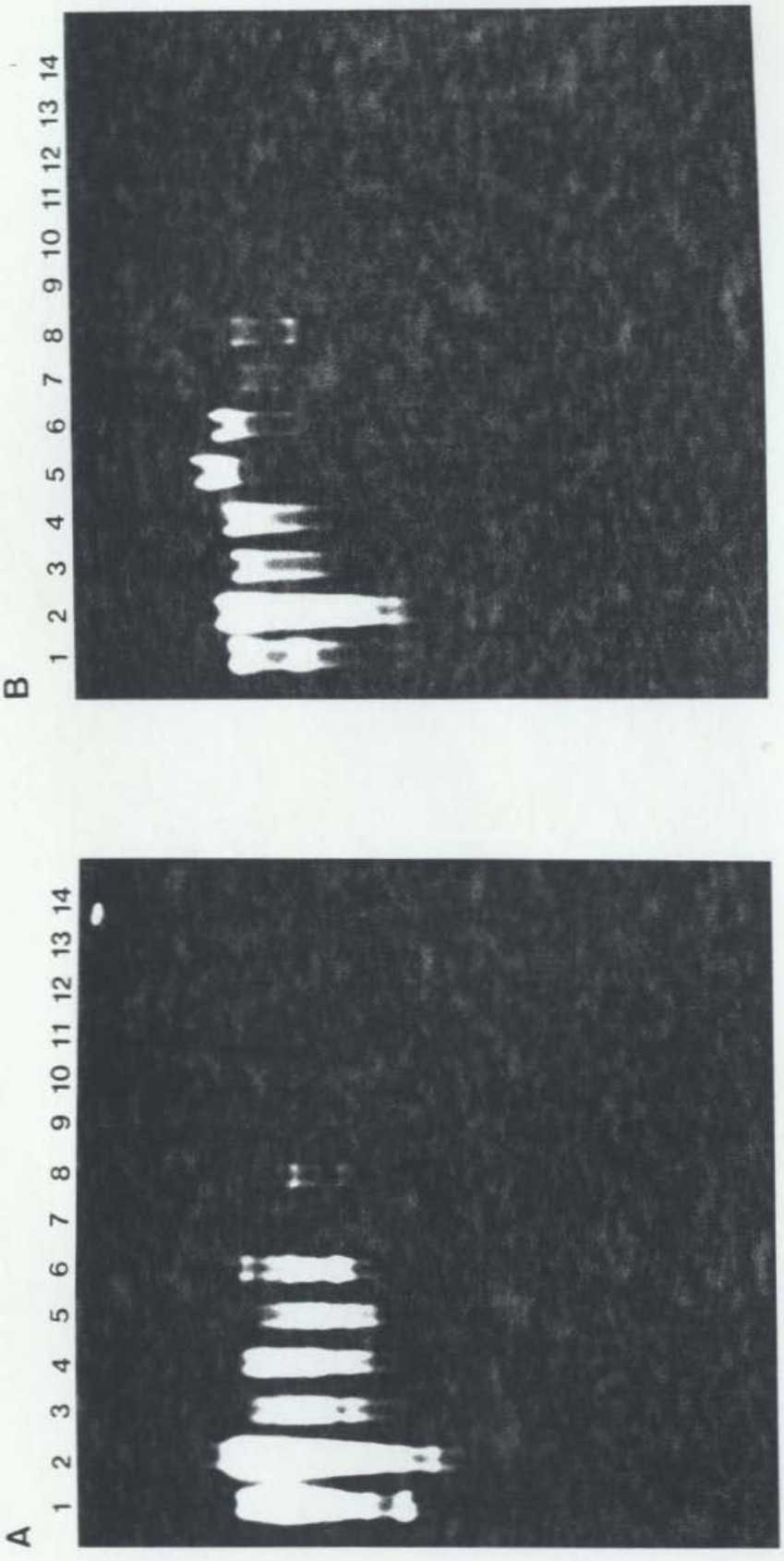

A
14 13 12 11 10 9 8 7 6 5 4 3 2 1
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14
C
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 6-12: Southern Blot of Enterobacteriaceae DNA Probed with pLC 10-47

Restriction fragments generated by Eco RI (A) or Hind III (B) were separated on a 0.8% agarose gel at 100 V.

(1) *Escherichia*, (2) *Shigella*, (3) *Salmonella choleraesuis*, (4) *Salmonella typhimurium*, (5) *Citrobacter*, (6) *Serratia*, (7) *Enterobacter*, (8) *Klebsiella*, (9) *Edwardsiella*, (10) *Yersinia*, (11) *Proteus*, (12) *Erwinia*. Lane 13 contained no DNA and lane 14 corresponds to λ DNA cut with Hind III.


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14

6.4.3.1. Hybridization with pLC 16-4 as Probe

The recombinant plasmid pLC 16-4 was used in an attempt to detect sequences in genomic DNAs from bacteria and eukaryotes that correspond to the *tpi* or *pfk* genes from *E. coli*. There was no obvious cross-reaction with eukaryotic DNA. Hybridization was strong with *Escherichia*, *Salmonella* and *Citrobacter* DNA but rather weak with *Yersinia* and *Pseudomonas* (Figure 6-13).

6.4.3.2. Hybridization with pLC 10-47 as Probe

Only DNA from *Escherichia*, *Salmonella*, *Citrobacter*, *Yersinia* and *Pseudomonas* showed any hybridization with pLC 10-47 (Figure 6-14). The bands appearing on the *Yersinia* and *Pseudomonas* samples were rather faint compared with those of *Escherichia*, *Salmonella* and *Citrobacter*. There was no cross-reactivity with the other DNAs, except for the non-specific smear seen with chicken DNA.

6.5. Discussion

Bacterial genomes are presumed to have descended from a common ancestor, diverging from one another as each genome underwent a succession of changes such as base substitutions, duplications, and rearrangements (Riley and Anilionis 1978; 1980; Anilionis and Riley, 1980). However, the dates at which bacteria diverged from their common ancestors are at present open to debate. It has been suggested that prokaryotes can be divided into two primary kingdoms based on comparisons of the three-dimensional structures of their ribosomes and partial nucleotide sequences of rRNA. As only molecules involved in the translational apparatus were used to make this prediction, it is important to

Figure 6-13: Southern Blot of Genomic DNAs Probed with pLC 16-4

Restriction fragments generated by Eco RI (A) or Hind III (B)

were separated on a 0.8% agarose gel at 100 V.

(1) *Escherichia coli*; (2) *Salmonella choleraesuis*;

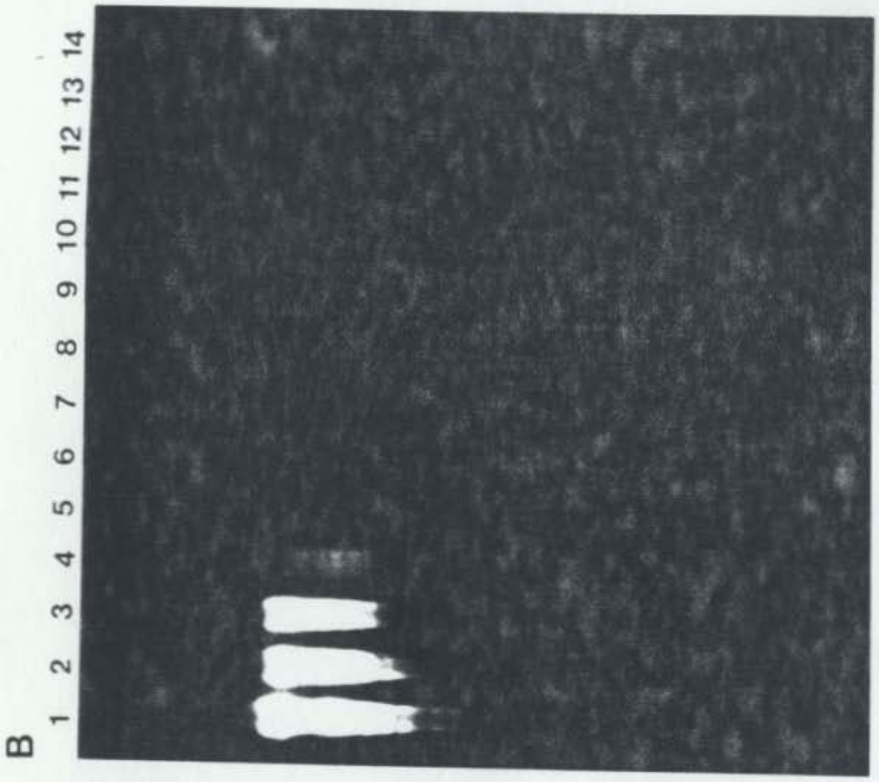
(3) *Citrobacter*; (4) *Yersinia*; (5) *Pseudomonas*;

(6) *Bacillus*; (7) *Clostridium*; (8) *Micrococcus*;

(9) *Lactobacillus*; (10) *Halobacterium volcanii*; (11)

Halobacterium halobium; (12) Human; (13) Chicken. Lane

14 contained λ DNA cut with Hind III.


B
1 2 3 4 5 6 7 8 9 10 11 12 13 14

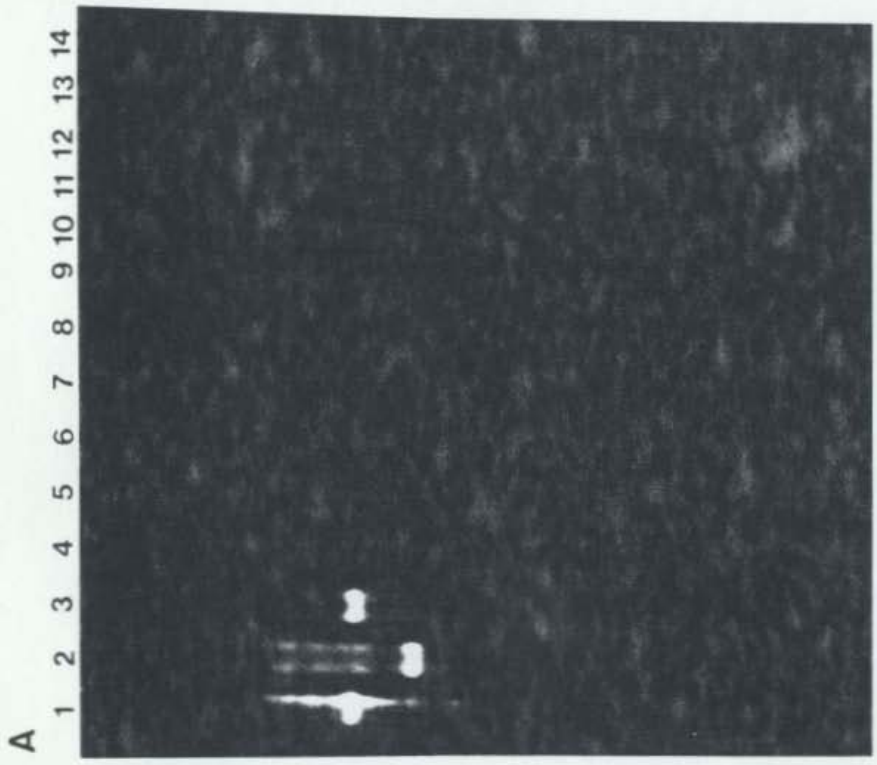
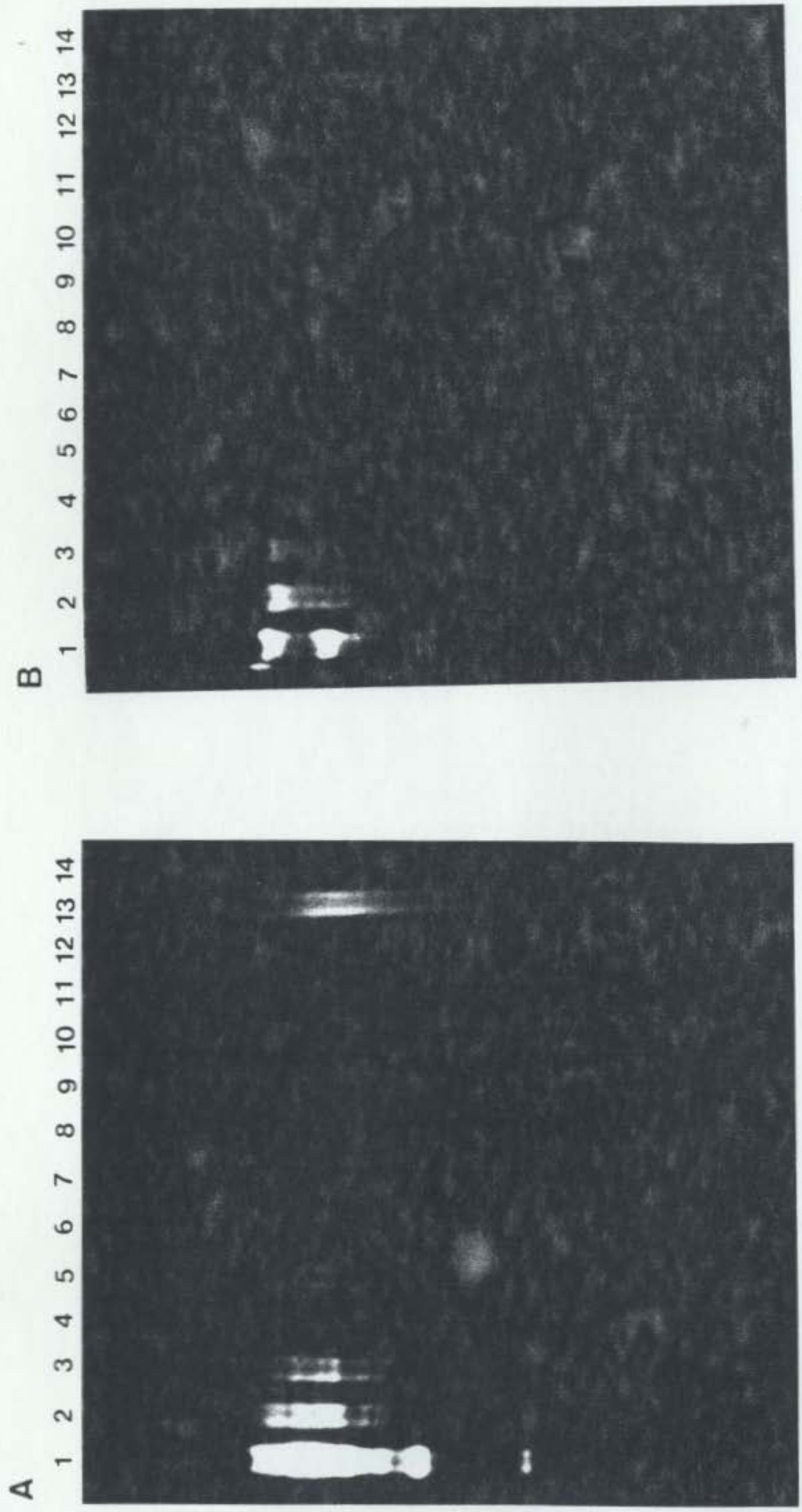

A
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 6-14: Southern Blot of Genomic DNAs Probed with pLC 10-47

Restriction fragments generated by Eco RI (A) or Hind III (B) were separated on a 0.8% agarose gel at 100 V.

(1) *Escherichia coli*; (2) *Salmonella choleraeguis*; (3) *Citrobacter*; (4) *Yersinia*; (5) *Pseudomonas*; (6) *Bacillus*; (7) *Clostridium*; (8) *Micrococcus*; (9) *Lactobacillus*; (10) *Halobacterium volcanii*; (11) *Halobacterium halobium*; (12) Human; (13) Chicken. Lane 14 contained λ DNA cut with Hind III.


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14

examine other molecules to determine if they comply with the current hypothesis. In an attempt to gain a greater understanding of prokaryotic evolution the hybrid plasmids pLC 10-47 and pLC 16-4 were used as radioactively-labelled probes to search for nucleotide sequences corresponding to the glycolytic genes *tpi*, *pfk A* and *eno* among organisms representative of the three primary kingdoms.

The results from Southern blot studies indicated that there was significant hybridization between the plasmids pLC 16-4 and pLC 10-47 and *Escherichia* DNA. As these hybrid recombinant plasmids contain inserts of the *Escherichia* genome the intensity of the signal in such a hybridization served as a positive control and indicated 100 per cent homology. The interpretation of results was made difficult when it was found that the cloning vector for pLC 16-4 and pLC 10-47 (K30) gave positive results with the preparations of DNA from *Shigella*, *Serratia*, and *Klebsiella*. It has since been established that this cross-reaction was due to these strains of bacteria containing plasmids that contain sequences related to Col E1 (K30) (W.S. Davidson, personal communication).

The only members of the family Enterobacteriaceae, other than *Escherichia*, to show strong hybridization signals with pLC 16-4 and pLC 10-47 were *Shigella*, *Salmonella sp.*, *Serratia*, *Citrobacter* and *Klebsiella*. In the case of *Yersinia*, *Proteus* and *Enterobacter* the bands were much less intense and sometimes undetectable. No cross-hybridization was seen with DNA from *Edwardsiella* or *Erwinia*. *Pseudomonas* was the only non-member of the Enterobacteriaceae to show any cross-hybridization with pLC 10-47 or pLC 16-4.

Several laboratories have used Southern blot analysis and/or hybridization studies in order to examine the evolutionary relationships among bacteria. For example the genes from *Klebsiella pneumoniae* that encode nitrogenase hybridized to the nitrogenase genes from a diverse array of other organisms that are able to fix nitrogen (Ruvkun and Ausubel, 1980). These organisms included gram-negative bacteria, gram-positive bacteria and cyanobacteria. However, the *nif* genes did not detect homologous sequences in closely related species that do not fix nitrogen. Another bacterial gene that has been used successfully to detect similar sequences in phylogenetically distant organisms is the *tuf* gene from *E. coli*. This gene, which encodes the elongation factor (Tu) that functions in protein synthesis, is able to detect homologous sequences in *Chromatium vinosum*, a purple sulphur photosynthetic bacterium, (Filer and Furano, 1980) and in the chloroplast genome of *Chlamydomonas reinhardtii* (Watson and Surzycki, 1982). The following order of extent sequence homology between the *E. coli tuf* gene and DNA from other members of the Enterobacteriaceae was obtained: *Salmonella* = *Shigella* > *Enterobacter* = *Serratia* > *Proteus* > *Erwinia* > *Yersinia*. The extent of cross-reaction between *E. coli tuf* and *Pseudomonas* DNA was greater than that between *E. coli tuf* and *Yersinia* DNA (Filer *et al.*, 1981).

Riley and Anilionis (1980) assessed the extent of relatedness of several portions of the *E. coli* genome with those of other enteric bacteria. Their results with *tna*, *trp*, and *thy* indicated that *Escherichia* is most closely related to *Shigella* followed by *Salmonella*, *Erwinia*, *Citrobacter*, *Enterobacter*, *Klebsiella*

and then *Serratia*. Similar results were obtained when *lac* was used as the probe except that no homologous sequences were found in *Salmonella typhimurium*. This is consistent with a lack of an inducible β-galactosidase in this species. The *speC* gene of *E. coli* only detected similar sequences in *Salmonella*, *Citrobacter*, *Klebsiella* and *Enterobacter* although biosynthetic ornithine decarboxylase activity was detected in *Edwardsiella*, *Proteus*, *Yersinia* and *Serratia* (Wright and Boyle, 1984). (*Shigella* was not included in this study.) The results from Southern blot analysis and hybridization studies using specific probes are in agreement with reassociation studies that have used entire genomes (Brenner and Falkow, 1971). These earlier studies concluded that *Escherichia* and *Shigella* are very closely related. *Citrobacter* and *Salmonella* are thought to be the next nearest, with *Klebsiella* and *Enterobacter* more distantly related. *Serratia*, *Erwinia* and *Edwardsiella* appear to be even more distantly related and the *Proteus* group bears little sequence similarity.

Direct nucleotide sequencing of the *trpG* regions of *Escherichia*, *Shigella*, *Salmonella* and *Serratia* revealed that *Escherichia* is most closely related to *Shigella*. There are about three times as many nucleotide substitutions between *Salmonella* and *Escherichia* or *Shigella* than there are between *Escherichia* and *Shigella*. *Serratia* is a very distant relative of the other three and it appears as if a fusion of the *trpG* and *trpD* genes occurred in the common ancestor of *Escherichia*, *Shigella* and *Salmonella* after it branched off from the line leading to *Serratia* (Nichols *et al.*, 1980).

Bearing in mind that some of the cross-reactions seen with *Shigella*,

Serratia and *Klebsiella* were due to the presence of Col E1 plasmid-related sequences, the results that were obtained using pLC 16-4 and pLC 10-47 as probes are consistent with previous studies and hypotheses concerning the phylogenetic relationships among the Enterobacteriaceae. The limited homology among members of the Enterobacteriaceae that was observed using plasmids harboring glycolytic genes as evolutionary probes suggests that the selective pressures on the DNA sequences of *eno*, *tpi* and *pfk A* have been less stringent than for other genes such as *tuf*, *nif*, *tna*, *trp* and *thy*. It has been assumed throughout this work that any hybridization observed between pLC 10-47 or pLC16-4 and bacterial genomic DNAs would likely be due to the genes for glycolytic enzymes. This of course need not be the case. The lack of hybridization indicates that the glycolytic genes are not participating, but positive results could be due to non-glycolytic genes carried on these plasmids.

6.6. Conclusion

The point of this thesis was to test the possibility that genes for glycolytic enzymes could be used to detect similar sequences in the genomes of organisms distantly related to *Escherichia*. The results that were obtained with two independent plasmids carrying *E. coli* glycolytic genes indicate that this is not possible under the hybridization conditions tested. It has been estimated that *Escherichia* and *Erwinia* last shared a common ancestor about 40 million years ago whereas the divergence times for *Escherichia* and *Enterobacter* and *Escherichia* and *Proteus* are 20 and 30 million years ago respectively (Filer *et al.*, 1981). If these estimates are reliable, then it is probable that the genes for

glycolytic enzymes will only be useful for detecting the corresponding genes in species that have been separated for less than 30 million years. This does not imply that comparison of the amino acid sequences of glycolytic enzymes would not be useful for long range evolutionary studies, but rather that this information must be obtained by more conventional methods.

Chapter 7

References

1. Alber, T. and Kawasaki, G. 1982. Nucleotide sequence of the triose phosphate isomerase gene of *Saccharomyces cerevisiae*. J. Mol. App. Genet. 1: 419-434.

2. Anilionis, A. and Riley, M. 1980. Conservation and variation of nucleotide sequences within related bacterial genomes: Enterobacteria. J. Bacteriol. ,143: 366-376.

3. Artavanis-Tsakonas, S. and Harris, J.I. 1980. Primary structure of triosephosphate isomerase from *Bacillus stearothermophilus*. Eur. J. Biochem. 108: 599-611.

4. Bachmann, B.J. 1980. Linkage map of *Escherichia coli* K-12, Edition 7. Microbiol. Reviews 47: 180-230.

5. Branlant, G., Flesch, G. and Branlant, C. 1983. Molecular cloning of the glyceraldehyde-3-phosphate dehydrogenase of *Bacillus stearothermophilus* and *Escherichia coli*, and their expression in *Escherichia coli*. Gene 25: 1-7.

6. Brenner, D.J. and Falkow, S. 1971. Molecular relationships among members of the enterobacteriaceae. Adv. Genet. 16: 81-118.

7. Brimacombe, R. 1984. Conservation of structure in ribosomal RNA. TIBS 9: 273-277.

8. Chin, C.C., Brewer, J.M., Eckard, E. and Wold, F. 1981a. The amino acid sequence of yeast enolase: Preparation and characterization of peptides produced by chemical and enzymatic fragmentations. J. Biol. Chem. 256: 1370-1376.

9. Chin, C.C., Brewer, J.M. and Wold, F. 1981b. The amino acid sequence of yeast enolase. J. Biol. Chem. 256: 1377-1384.

10. Clarke, L. and Carbon, J. 1976. A colony bank containing synthetic Col E1 hybrid plasmid representative of the entire *E. coli* genome. Cell 9: 91-99.

11. Clarke, L. and Carbon, J. 1978. Functional expression of cloned yeast DNA in *Escherichia coli*: Specific complementation of argininosuccinate lyase (*arg H*) mutations. J. Mol. Biol. 120: 517-532.

12. Daldal, F. 1984. Nucleotide sequence of gene *pfkB* encoding the minor phosphofructokinase of *Escherichia coli* K-12. Gene 28: 337-342.

13. Dayhoff, M.O. (Ed) 1978. In "Atlas of Protein Sequence and Structure", vol. 5, suppl. 3, Natl. Biomed. Res. Found., Washington, D.C.

14. De Smelt, J. and De Ley. J. 1977. Intra- and intergeneric similarities of *Agrobacterium* ribosomal ribonucleic acid cistrons. Int. J. Syst. Bacteriol. 27: 222-239.

15. Dickerson, R.E. and Geis, I. 1969. *The Structure and Action of Proteins*, W.A.Benjamin, Inc., Menlo Park, California.

16. Dickerson, R.E. 1971. The structure of cytochrome c and the rates of molecular evolution. J. Molec. Evolution 1: 26-45.

17. Doolittle, W.F. 1980. Revolutionary concepts in evolutionary cell biology. TIBS 5: 146-149.

18. Doolittle, W.F. 1982. Evolutionary molecular biology: where is it going? Can. J. Biochem. 60: 83-90.

19. Durica, K.S., Schloss, J.A. and Crain, Jr. W.R. 1980. Organization of actin gene sequences in the sea urchin: Molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin. Proc. Natl. Acad. Sci. USA 77: 5683-5687.

20. Engel, J.D. and Dogson, J.B. 1981. Histone genes are clustered but not tandemly repeated in the chicken genome. Proc. Natl. Acad. Sci. USA 78: 2856-2860.

21. Fraenkel, D.G. and Vinopal, R.T. 1973. Carbohydrate metabolism in bacteria. Ann. Rev. Microbiol. 27: 69-100.

22. Filer, D. and Furano, V. 1980. Portions of the gene encoding factor Tu are highly conserved in prokaryotes. J. Biol. Chem. 255: 728-734.

23. Filer, D., Dhar, R. and Furano, A.V. 1981. The conservation of DNA sequences over very long periods of evolutionary time. Eur. J. Biochem. 69-76.

24. Fothergill-Gilmore, L.A. , 1986. The evolution of the glycolytic pathway. TIBS 11: 47-51.

25. Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Leuhrsen, K.R., Chen, K.N. and Woese, C.R. 1980. The phylogeny of prokaryotes. Science 209: 457-463.

26. Futuyma, D.J. 1983. *Science on Trial*, p.73., Pantheon Books, New York.

27. Gupta, R., Lanter, J.M. and Woese, C.R. 1983. Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an Archaebacterium. Science 221: 656-659.

28. Helling, R.B., Goodman, H.M. and Boyer H.W. 1974. Analysis of Eco R1 fragments of DNA from lambdoid bacteriophages and other viruses by agarose gel electrophoresis. J. Virol. 14: 1235.

29. Hellinga, H.W. and Evans, P.R. 1985. Nucleotide sequence and high level expression the major *Escherichia coli* phosphofructokinase. Eur. J. Biochem. 149: 363-373.

30. Hocking, J.D. and Harris, J.I. 1980. D-glyceraldehyde-3-phosphate dehydrogenase : Amino acid sequence of the enzyme of the extreme thermophile *Thermus aquaticus*. Eur. J. Biochem. 108: 571-573.

31. Holland, M.J., Holland, J.P., Thill, G.P. and Jackson, K.A. 1981. The primary structures of two yeast enolase genes. J. Biol. Chem. 256: 1385-1395.

32. Holland, J.P., Labieniec, L., Swimmer, C. and Holland, M.J. 1983. Homologous nucleotide sequences at the 5' termini of messenger RNAs synthesized from the yeast enolase and glyceraldehyde-3-phosphate dehydrogenase gene families. J. Biol. Chem. 258: 5291-5299.

33. Howe, J. G. and Hershey, J.W.B. 1984. The rate of evolutionary divergence of initiation factors IF2 and IF3 in various bacterial species determined quantitatively by immunoblotting. Arch. Microbiol. 140: 187-192.

34. Irani, M.H. and Maitra, P.K. 1974. Isolation and characterization of *Escherichia coli* mutants defective in enzymes of glycolysis. Biochem. Biophys. Res. Commun. 56: 127-133.

35. Irani, M.H. and Maitra, P.K. 1976. Glyceraldehyde 3-P Dehydrogenase, glycerate 3-P kinase and enolase mutants of *Escherichia coli* : Genetic studies. Molec. Gen. Genet. 145: 65-71.

36. Kersters, K. and De Ley, J. 1975. Identification and grouping of bacteria by numerical analysis of their electrophoretic protein patterns. J. Gen. Microbiol. 87: 333-342.

37. Kimura, M. and Langner, G. 1984. The complete amino acid sequence of the ribosomal protein HS3 from *Halobacterium marismortui*, an archaebacterium. FEBS Letters 175: 213-218.

38. King, M. and Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. Science 188: 107-116.

39. Knoll, A.H. and Barghoorn, E.S. 1975. Precambrian eukaryotic organisms: A reassessment of evidence. Science 190: 52-54.

40. Kolb, E., Hudson, P.J. and Harris, J.I. 1980. Phosphofructokinase: Complete amino-acid sequence of the enzyme from *Bacillus stearothermophilus*. Eur. J. Biochem. 108: 587-597.

41. Lake, J.A., Henderson. E., Clark, M.W. and Matheson, A.T. 1982. Mapping evolution with ribosome structure: Intralineage constancy and interlineage variation. Proc. Natl. Acad. Sci. USA 19: 5948-5952.

42. Lake, J.A. 1983. Evolving ribosome structure: Domains in archaebacteria, eubacteria and eucaryotes. Cell 33: 318-319.

43. Lake, J.A., Clark, M.W., Henderson, E., Fay, S.P., Oakes, M., Scheinman, A., Thornber, J.P. and Mah, R.A. 1985. Eubacteria, halobacteria, and the origin of photosynthesis: The photocytes. Proc. Natl. Acad. Sci. USA 82: 3716-3720.

44. Lake, J.A. 1986. In defence of bacterial phylogeny. Nature 321: 657-658.

45. Laemmli, U.K. 1970. Cleavage of structural proteins during assemble of bacterial phage. Nature 277: 680-685.

46. Lederer, H. 1986. Archaebacterial status quo is defended. Nature 320: 220.

47. Lu, H.S., Yuan, P.M. and Gracy, R.W. 1984. Primary structure of human triosephosphate isomerase. J. Biol. Chem. 259: 11958-11968.

48. Maleszka, R., Wang, P.Y. and Schneider, H. 1982. A Col E1 hybrid plasmid containing *Escherichia coli* genes complementing D-xylose negative mutants of *Escherichia coli* and *Salmonella typhimurium*. Can. J. Biochem. 60: 144-151.

49. Maniatis, T., Fritsch, E.F. and Sambrook, J. 1982. *In* Molecular Cloning laboratory manual. Cold Spring Harbor Laboratory Publication.

50. Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. 3: 208-218.

51. Martin, S.L., Vincent, K.A. and Wilson, A.C. 1983. Rise and fall of the delta globin gene. J. Mol. Biol. 164: 513-528.

52. Mazur, B.J., Rice, D. and Haselkorn, R. 1980. Indentification of blue-green algal nitrogen fixation genes by using heterologous DNA hybridization probes. Prc. Natl. Acad. Sci. USA 77: 186-190.

53. McAlister, L. and Holland, M.J. 1982. Targeted deletion of a yeast enolase structural gene. J. Biol. Chem. 257: 7181-7188.

54. Meinkoth, J. and Wahl, G. 1984. Hybridization of nucleic acids immobilized on solid supports. Analytic Biochem. 138: 267-284.

55. Miernyk, J.A. and Dennis, D.T. 1984. Enolase isozymes from *Ricinus communis*: Partial purification and characterization of the isozymes. Arch. Biochem. Biophys. 233: 643-651.

56. Milner, R.J., Brow, M.D., Cleveland, D.W., Shinnick, T.M. and Sutcliffe, J.G. 1983. Glyceraldehyde-3-phosphate dehydrogenase protein and mRNA are both differentially expressed in adult chickens but not chick embryos. Nucleic Acids Research 11: 3301-3315.

57. Moran, L.A., Chauvin, M., Kennedy, M.E., Korri, M., Lowe, D.G., Nicholson, R.C. and Perry, M.D. 1983. The major heat-shock protein (hsp 70) gene family: related sequences in mouse, *Drosophilia* and yeast. Can. J. Biochem. 61: 489-499.

58. Musti, A.M., Zehner, Z., Bostian, K.A., Paterson, B.M. and Kramer, R.A. 1983. Transcriptional mapping of two yeast genes coding for glyceraldehyde-3-phosphate dehydrogenase isolated by sequence homology with the chicken gene. Gene 25: 133-143.

59. Nichols, B.P., Miozzari, G.F., van Cleemput, M., Bennett, G.N. and Yanofsky, C. 1980. Nucleotide sequences of the *trpG* regions of *Escherichia coli*, Shigella dysentariae, Salmonella typhimurium] and *Serratia marcescens*. J. Mol. Biol. 142: 503-517.

60. Oehler, J.H., Oehler, D.Z. and Muir, M.D. 1976. On the significance of tetrahedrad tetrads of Precambrian algal cells. Orig. Life 7: 259-267.

61. Pace, N. R., Stahl, D.A., Lane, D.J. and Olsen, G.J. 1985. Analyzing natural microbial populations by rRNA sequences. ASM News 51: 4-12.

62. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Doggson, J. 1980. The evolution of genes: the chicken preproinsulin gene. Cell 20: 555-566.

63. Pfleiderer, G., Neufahrt-Kreiling, A., Kaplan, R.W. and Fortnagel, P. 1966. Biochemical, immunological, and genetic investigations of the multiple forms of yeast enolase. Biochemische Zeitschrift 346: 269-281.

64. Pichersky, E. et al. 1984. Nucleotide sequence of the triose phosphate isomerase gene of *Escherichia coli*. Mol. Gen. Genet. 195: 314-320.

65. Pietkiewicz, J., Kustrzeba-Wojcicka, I. and Wolna E. 1983. Purification and properties of enolase from carp *Cyprinus carpio* comparison with enolases from mammals' muscles and yeast. Comp. Biochem. Physiol. 75B: 693-698.

66. Poorman, R.A, Randolph, A., Kemp, R.G. and Heinrikson, R.L. 1984. Evolution of phosphofructo-kinase gene duplication and creation of new effector sites. Nature 309: 467-469,

67. Putkey, J.A., Ts'ui, K.F., Tanaka, T., Lagace, L. Stein, J.P., Lai, E.C. and Means, A.R. 1983. Chicken calmodulin genes. J. Biol. Chem. 258: 11864-11870.

68. Ratzkin, R. and Carbon, J. 1977. Functional expression of cloned yeast DNA in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 74: 487-491.

60. Riley, M. and Anilionis, A. 1978. Evolution of the bacterial genome. Ann. Rev. Microbiol. 32: 519-560.

70. Riley, M. and Anilionis, A. 1980. Conservation and variation of nucleotide sequences within related bacterial genomes: Enterobacteria. J. Bacteriol. 143: 366-376.

71. Ruvkun, G.B. and Ausubel, F.M. 1980. Interspecies homology of nitrogenase genes. Proc. Natl. Acad. Sci. USA 77: 191-195.

72. Sakimura, K., Kushiya, E., Obinata, M. and Takahashi, Y. 1985. Molecular cloning and the nucleotide sequence of cDNA to mRNA for non-neuronal enolase ($\alpha\alpha$ enolase) of rat brain and liver. Nucleic Acids Research 13: 4365-4378.

73. Scarpulla, R.C., Agne, K.M., Borlthel, J. and Wu, R. 1981. Structures of a gene and pseudogene for rat cytochrome c. DNA 1: 96.

74. Schopf, T.J.M., Raup, D.M., Gould S.J. and Simberloff, D.S. 1975. Genomic versus morphologic rates of evolution: Influence of morphologic complexity. Paleobiology 1: 63-70.

75. Schopf, J.W. 1978. The evolution of the earliest cells. Sci. Am. 239: 110-138.

76. Schuler, M.A. and Keller, E.B. 1981. The chromosomal arrangement of two linked actin genes in the sea urchin *S. purpuratus*. Nucleic Acid Research 9(3): 591-604.

77. Shimosaka, M., Fukuda, Y., Murata, K. and Kimura, A. 1982. Application of hydrid plasmids carrying glycolysis genes to ATP production by *Escherichia coli*. J. Bacteriol. 152: 98-103.

78. Spring, T. G. and Wold, F. 1971. The purification and characterization of *Escherichia coli* enolase. J. Biol. Chem. 246: 6797-6802.

79. Stellwagen, E., Cronlund, M.M. and Barnes, L.D. 1973. A thermostable enolase from the extreme thermophile *Thermus acquaticus* YT-1. Biochemistry 12: 1552-1559.

80. Straus, D. and Gilbert, W. 1985. Chicken triosephosphate isomerase complements an *Escherichia coli* deficiency. Proc. Natl. Acad. Sci. USA 82: 2014-2018.

81. Thomson, J. 1977. *E. coli* phosphofructokinase synthesized *in vitro* from a Col E1 hybrid plasmid. Gene 1: 347-356.

82. Thomson, J., Gerstenberfer, P.D., Goldberg, D.E., Gociar, E., de Silva, A.O. and Fraenkel, D.G. 1979. Col E1 hybrid plasmids for *Escherichia coli* genes of glycolysis and the hexose monophosphate shunt. J. Bacteriol. 137: 502-506.

83. Upholt, W.B. 1977. Estimation of DNA sequence divergence from comparison of restriction endonuclease digests. Nucleic Acid Res. 4(5): 1257-1265.

84. Van Valen, L.M. and Maiorana, V.C. 1980. The archaebacteria and eukaryotic origins. Nature 287: 248-250.

85. Vidal, G. 1984. The oldest eukaryotic cells. Sci. Am. 250: 48-57.

86. Watson, J.C. and Surzycki, S.J. 1982. Extensive sequence homology in the DNA coding for elongation factor Tu from *Escheichia coli* and the *Chlamydomonas reinhardtii* chloroplast. Proc. Natl. Acad. Sci. USA 79: 2264-2267.

87. Wilson, A.C., Carlson, S.S., and White, T.J. 1977. Biochemical evolution. Ann. Rev. Biochem. 46: 573-639.

88. Woese, C.R. and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. U.S.A. 74: 5088-5090.

89. Wold, F. 1971. Enolase. In *The Enzymes* (Edited by Boyer, P.D.) 3rd edition, vol 5, pp499-538. Academic Press, New York.

90. Wright, J.M. and Boyle, S.M. 1984. Intergeneric homology of the *speC* gene encoding biosynthetic ornithine decarboxylase in *Escherichia coli*. J. Bacteriol. 159: 1074-1076.

91. Yaguchi, M., Visentin, P., Zuker, M., Matheson, A.T., Roy, C. and Strom, A.R. 1982. Amino-terminal sequences of ribosomal proteins from the 30S subunit of archaebacterium *Halobacterium cutirubrum*. Zbl. Bakt. Hyg. I. Abt. Orig. C3: 200-208.

92. Zillig, W. 1986. Archaebacterial status quo is defended. Nature 320: 220.

93. Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan. Y. W. and Wilson, A.C. 1980. Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc. Natl. Acad. Sci. USA 77: 2158-2162.

Chapter 8

Appendix 1

1. LB (Luria-Bertani) Medium Per litre:

Bacto-tryptone	10 g
Bacto-yeast extract	5 g
NaCl	10 g

Adjust pH to 7.5 with sodium hydroxide.

2. M9 Medium Per litre:

Na_2HPO_4	6 g
KH_2PO_4	3 g
NaCl	0.5 g
NH_4Cl	1 g

Adjust pH to 7.4, autoclave, cool, and then add:

1 M $MgSO_4$	2 ml
20% glucose	10 ml
1 M $CaCl_2$	0.1 ml

The above solutions should be sterilized separately by filtration (glucose) or autoclaving.

3. Growth Medium for Halobacteria Per litre:

Tryptone	2.5 g
Yeast extract	5 g
NaCl	25 g
$MgSO_4.7H_2O$	15 g
KCl	5 g
$CaCl_2.6H_2O$	0.2 g

4. 50 X Denhardt's solution Per litre:

Ficol	5 g
Polyvinyl-pyrrolidone	5 g
Bovine serum albumin (BSA)	5 g

5. 20X SSC Per litre:

NaCl	175.3 g
Trisodium citrate	88.2 g
Adjust pH to 7.0 using NaOH	

6. 50X TAE (Tris-Acetate) Per litre:

Tris base	242 g
Glacial acetic acid	57.1 ml
0.5 M EDTA (pH 8.0)	100 ml

7. 5X TBE (Tris-Borate) Per litre:

Tris base	54 g
Boric acid	27.5 ml
0.005 M EDTA (pH 8.0)	20 ml

8. TE

10 mM Tris/HCl (pH 8.0)
1 mM EDTA (pH 8.0)