

AN ABSTRACT OF THE THESIS OF

Sarla Purohit for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on April 26, 1984.

Title: Bacteriophage T4-Coded Dihydrofolate Reductase: Cloning and Structural Analysis of its Gene and Organization of Flanking Regions.

Abstract Approved: Redacted for Privacy

The dihydrofolate reductase gene of bacteriophage T4 has been investigated at the structural level to elucidate the unusual regulation of expression of this gene and possible evolutionary relationships between the phage and drug resistance factors-encoded dihydrofolate reductases (DHFR).

A 1.1-kilobase-pair restriction fragment of the T4 multiple mutant dec8 containing frd, the structural gene for DHFR, has been cloned in E. coli by using pBR322 as the vector. It has been demonstrated, both by restriction and biochemical analysis, that the recombinant plasmid pSP19 carrying this fragment contains the entire frd gene, and the enzyme synthesized by these transformants is indistinguishable from T4 DHFR. Overproduction of this enzyme has been achieved by inserting this fragment into the hyper-expression vector pGW7.

The length of the cloned fragment, as determined by the base modification technique of Maxam and Gilbert, is 1073 base pairs. Two of the four open reading frames identified are the structural genes for DHFR and part of

thymidylate synthase (TS). The frd gene is of 579 base pairs and is translated into 193 amino acid residues. frd and td genes show a four-base-pair overlap with each other. Positions of these genes were identified by correlating the predicted and experimentally determined sequences of 20 N-terminal amino acid residues of DHFR (T. Bailey) and TS (162).

Two additional open reading frames are located upstream to frd. The first open reading frame (ORF), from the 5' end, when translated, shows sequence homology to the conserved active site region of all known DHFRs. The second reading frame, a small leader sequence, flanks the 5' end of frd. At least four complementary inverted repeats have been identified in the segment that includes the leader sequence and 20 nucleotides from the 5' region of frd. Also, the termination codons of the ORF and the leader sequence overlap with the Shine and Dalgarno sequence of the leader and frd, respectively.

This fragment also contains at least five sequences which show homology to the consensus sequence of E. coli promoters; however, no conventional transcriptional termination sequence is present.

Comparison of the deduced amino acid sequence of both T4 and R-plasmid enzymes does not show homology more than what is observed among other DHFRs, suggesting that these enzymes are not unusually closely related.

Bacteriophage T4-Coded Dihydrofolate Reductase:  
Cloning and Structural Analysis of its Gene  
and Organization of Flanking Regions

by

Sarla Purohit

A THESIS

submitted to

Oregon State University

in partial fulfillment of the requirements  
for the degree of

Doctor of Philosophy

Completed April 26, 1984

Commencement June 1984

APPROVED:

Redacted for Privacy

---

Professor of Biochemistry and Biophysics in charge of  
Major

Redacted for Privacy

---

Head of Department of Biochemistry and Biophysics

Redacted for Privacy

---

Dean of Graduate School

Date thesis is presented April 26, 1984

Typed by Esther Kissling for Sarla Purohit

## ACKNOWLEDGMENTS

I thank Dr. Christopher K. Mathews for his guidance and support. His patience, stimulating discussions and incisive criticisms were most valuable.

I acknowledge the help of Dr. Richard K. Bestwick during the early phases of this research, and wish to thank Dr. Henry W. Schaup for his helpful insights.

I also thank my children, Amit, Anita and Sumit, and my husband, Kris, without whose encouragement, cooperation and support, I could not have completed this work.

## TABLE OF CONTENTS

	<u>Page</u>
I. INTRODUCTION.....	1
A. Background.....	1
B. Compartmentation of DNA Synthesis and Multienzyme Complex for DNA Precursor Biosynthesis.....	1
C. Thymidylate Synthase as the Phage Induced Enzyme.....	3
D. Dihydrofolate Reductase as a Viral Gene Product.....	3
E. Isolation of Dihydrofolate Reduc- tase Mutants.....	5
F. Functions of Dihydrofolate Reduc- tase.....	6
G. Mechanisms of Drug Resistance.....	9
H. Subunit Structure and Properties of Dihydrofolate Reductase.....	13
I. Primary Structure of Dihydro- folate Reductases.....	17
J. Importance and Interest in the Study of Bacteriophage T4 Dihydrofolate Reductase.....	19
K. Mutants of Bacteriophage T4 that Synthesize Cytosine Containing DNA....	33
L. Present Work.....	36
II. MATERIALS.....	38
A. Reagents.....	38
B. Bacterial Strains.....	39
C. Bacteriophage T4 Strains.....	40
D. Plasmid Vectors.....	40
III. PROCEDURES.....	41
A. Preparation of High Titer Phage Stocks.....	41
B. Purification of Phage Particles.....	41
C. Isolation of DNA from Phage Particles.....	42
D. Isolation and Purification of Plasmid DNA.....	43
E. Large Scale Purification of Plasmid Vector pGW7.....	46

## TABLE OF CONTENTS (continued)

	<u>Page</u>
F. Large Scale Purification of Recombinant pSP19 DNA.....	47
G. Small Scale Isolation of Plasmid DNA.....	47
H. Digestion of DNA with Restriction Enzymes.....	49
I. Digestion of DNA Isolated from T4 Deletion Mutants with EcoR V.....	50
J. Isolation of Restriction Fragments...	51
K. Transformation of Cells by Calcium Chloride Procedure.....	52
L. Molecular Cloning of T4 <u>frd</u> Gene.....	53
M. Molecular Cloning of 1.1-kb Hind III Fragment Isolated from pSP19 into Hyperexpression Vector pGW7.....	55
N. Molecular Cloning of 0.9-kb Mbo II-Hind III Fragment.....	57
O. Growth of Transformed Bacterial Cells for DHFR Assay.....	58
P. Preparation of Phage-Infected Cultures.....	59
Q. Culturing Cells Containing pSP104 for Overproduction of DHFR.....	59
R. Preparation of Crude Extract.....	59
S. Fractionation of <u>E. coli</u> and T4 DHFRs.....	60
T. Dihydrofolate Reductase Assay.....	60
U. Nucleotide Sequence Determination....	61
V. Sequencing Gels.....	69
W. Nick Translation.....	71
X. Hybridization.....	73
IV. RESULTS.....	75
A. Molecular Cloning of T4 <u>frd</u> Gene.....	75
B. Digestion of pSP19 with Restriction Enzymes.....	77
C. Heat Stability of Dihydrofolate Reductases from pSP19.....	82
D. Fractionation of Enzymes by Ammonium Sulfate.....	82
E. Sensitivity of Dihydrofolate Reductases to Trimethoprim.....	87
F. Sensitivity of Dihydrofolate Reductases to Antiserum.....	87

## TABLE OF CONTENTS (continued)

	<u>Page</u>
G. Recloning of 1.1-kb Fragment in the Hyperexpression Vector pGW7.....	91
H. Nucleotide Sequence Determination of the 1.1-kb Hind III Fragment from T4.....	101
I. Study of Deletion Mutants in the 63-32 Region.....	132
V. DISCUSSION.....	136
A. Molecular Cloning of 1.1-kb Hind III Fragment.....	136
B. Nucleotide Sequence of the Cloned 1.1-kb Fragment.....	140
VI. REFERENCES.....	151
VII. APPENDIX.....	162

## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1 Reactions of DNA precursor biosynthesis in T4 phage-infected <u>E. coli</u>	2
2 Structures of folic acid and several folate antagonists	8
3 Map positions of <u>frd</u> and <u>td</u> on partial map of T4 showing early genes controlling reactions of DNA precursor biosynthesis.	23
4 Early transcription unit of the rII region of bacteriophage T4 showing multiple promoters	28
5 Map of 63-32 region of T4 showing deletions in <u>del1</u> , <u>del7</u> and <u>del9</u> mutants	32
6 Generation of viable bacteriophage T4 containing dC-DNA	35
7 Preparative scale separation of 16-kb fragment of T4 dC-DNA from multiple mutant, <u>dec8</u> digested by Sma I, on 0.4% agarose gel	76
8 Restriction map of pSP19	78
9 Partial restriction map of T4 16-kb Sma I fragment	79
10 Determination of fragment sizes produced by digestion of pSP19 by Hind III, EcoR I and Sal I	80
11 Heat stability of dihydrofolate reductase activity in crude extracts	84
12 Ammonium sulfate fractionation of dihydrofolate reductase activity in crude extracts	85
13 Trimethoprim sensitivity of dihydrofolate reductase	89
14 Cloning of the T4 <u>frd</u> gene into a high-expression vector	95

<u>Figure</u>	<u>Page</u>
15 Separation of Sal I restriction fragments of pGW7-1.1-kb recombinant plasmids on agarose gel	97
16 SDS polyacrylamide gel stained by Coomassie brilliant blue R, showing overproduced T4 dihydrofolate reductase	103
17 Strategy of sequencing the T4 <u>frd</u> gene	106
18 Separation of fragments that are labelled with $^{32}\text{p}$ on one end	108
19 Polyacrylamide-urea gel showing part of the sequence of the 360-base pairs Sal I-Hind III fragments	111
20 Nucleotide sequence of the non-coding strand of the Hind III fragment encompassing the <u>frd</u> gene	113
21 Alternative secondary structures of the segment between 259 to 358 nucleotides from the 5' end of the fragment	118
22 Determinations of sizes of fragments produced by digestion of 1.1-kb fragment by Mbo II	120
23 List of putative promoter sequences for <u>frd</u> and <u>td</u> genes of T4	124
24 Sequence alignment of six dihydrofolate reductases according to structural equivalence	127
25 Sequence homologies among dihydrofolate reductases	130
26 Comparison of EcoR V restriction patterns of T4 mutants deleted in 63-32 region	134
27 Hybridization of DNA from T4 mutants deleted in 63-32 region	135

## LIST OF TABLES

<u>Table</u>		<u>Page</u>
I	Inhibition of dihydrofolate reductases by folate antagonists	10
II	Properties of dihydrofolate Reductases	14
III	Conditions for DNA modification reactions	65
IV	Protocol for preparation of urea-polyacrylamide gels	70
V	Electrophoresis conditions for sequencing gels	72
VI	Fragment sizes resulting from restriction digestion of pSP19	81
VII	Heat stability of dihydrofolate reductases at 43°C	83
VIII	Ammonium sulfate fractionation of dihydrofolate reductases	86
IX	Sensitivity of ammonium sulfate fractions to trimethoprim	90
X	Antiserum sensitivity of dihydrofolate reductases	92
XI	Fragment sizes of pGW7-1.1-kb recombinant plasmids digested with Sal I	98
XII	Dihydrofolate reductase activity in <u>E. coli</u> RRI carrying recombinant plasmids	100
XIII	Amino acid composition of T4 dihydrofolate reductase	104

Bacteriophage T4-Coded Dihydrofolate Reductase: Cloning  
and Structural Analysis of its Gene and Organization of  
Flanking Regions

I. INTRODUCTION

A. Background

Infection of Escherichia coli by T-even bacteriophage results in the synthesis of a large number of enzymes specified by the bacteriophage genome, reviewed by Wood and Revel (1) and Mosig (2). A number of these enzymes duplicate pre-existing host functions in DNA precursor biosynthesis to support the enhanced rate of DNA synthesis known to occur after T4 infection (Fig. 1). Thymidylate synthase (TS:EC.2.1.1.45) and dihydrofolate reductase (DHFR:EC.1.5.1.3) are representatives of this class of enzymes. Another class of T4-coded enzymes are unique to phage infection. The enzyme deoxycytidylate hydroxymethylase (HMase; EC. 2.1.2.b) belongs to this second category and functions in the synthesis of hydroxymethyl dCMP (Hm dCMP). Information about the enzymes in DNA precursor synthesis has been reviewed (3-6).

B. Compartmentation of DNA Synthesis and Multienzyme  
Complex for DNA Precursor Biosynthesis

As mentioned above, infection of bacteria by T-even phages results in increased phage specific DNA synthesis, whereas host macromolecular synthesis is almost completely arrested and host DNA is degraded by phage specific

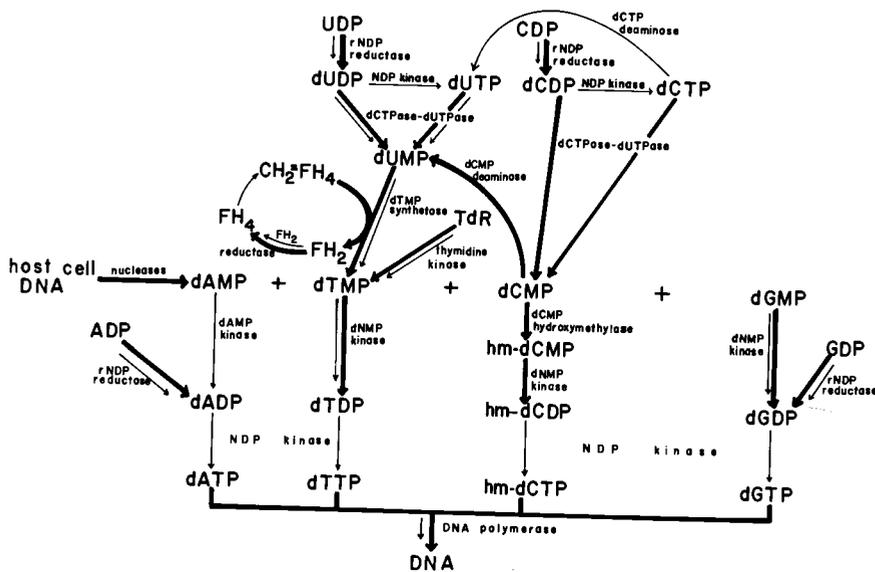


Figure 1. Reactions of DNA precursor biosynthesis in T4 phage-infected *E. coli*. Reactions catalyzed by virus-coded and preexisting host cell enzymes are identified by heavy and light arrows respectively (6).

nucleases. Phage DNA is protected against the action of these and the host nucleases by the presence of a modified base, hydroxymethyl cytosine (HMC), which is glucosylated. Compartmentation of DNA synthesis in the form of a DNA precursor-synthesizing multienzyme complex has been proposed (7, 8). This complex consists of a number of bacteriophage-specific early enzymes and two of the host enzymes, deoxyadenylate kinase (dAMP kinase: EC.2.7.4.11) and nucleoside diphosphokinase (NDP kinase: EC.2.7.4.6). Almost all the enzymes, including TS, that are involved in DNA precursor biosynthesis have been shown to be the components of this complex. However, DHFR, until recently, was not included in this list. This may be due to very weak associations of its physical location, with DHFR being stripped out during isolation. Allen et al. (9) have demonstrated that under milder lysis conditions DHFR is indeed present in this multienzyme complex.

#### C. Thymidylate Synthase as the Phage Induced Enzyme

Thymidylate synthase as the virus-induced enzyme was first demonstrated by Flaks and Cohen (10, 11). They showed a several fold increase in enzyme activity after infection of E. coli B by T-even phage. The evidence for thymidylate synthase as a viral gene product will be discussed in a later section.

#### D. Dihydrofolate Reductase as a Viral Gene Produce

Evidence that T6 phage induces DHFR was first pub-

lished by Mathews and Cohen (12). These authors identified the phage-specific enzyme by its difference in ammonium sulfate precipitation pattern from that of the host E. coli enzyme and the ability of the bacteriophage T6-induced enzyme to utilize NADH in place of NADPH as the hydride donor. However, the efficiency of NADH utilization was only one fourth of NADPH utilization. Dihydrofolate reductase in T6-infected cells was approximately 20-fold higher than that of uninfected E. coli. Induction of both DHFR and TS was also observed in other T-even bacteriophage infected cells. When the properties of partially purified DHFRs from T-even and T5 phages were compared (13,14), they all showed similar pH optima, sedimentation coefficients, pyridine nucleotide specificity and inhibition by trimethoprim (TMP), but showed considerable differences in inactivation by urea and heat. Although all bacteriophage DHFRs are highly sensitive to inhibition by aminopterin, T4 enzyme is the most and T2 enzyme is the least sensitive to this inhibitor.

Another pronounced difference between T-even phage DHFRs is in their sensitivity to heat. T4 enzyme was the least and T6 enzyme the most sensitive to heat inactivation.

From these studies it was confirmed that these bacteriophages do indeed induce phage specific DHFRs and that it is not the bacterial enzyme that is derepressed

on phage infection as was suggested by Warner and Lewis (15). The possibility of derepression of host enzyme was suggested because of the kinetics of formation of T4 DHFR, which differ from those of other early enzymes (15,16).

The conclusion that bacteriophage T4 codes for its DHFR was further supported by isolation of a number of bacteriophage T4 frd (wh in old terminology) mutants (17, 18, 19).

#### E. Isolation of Dihydrofolate Reductase Mutants

To isolate T4 frd mutants, the mutagenized phage particles were plated on uracil-requiring host mutants lacking deoxycytidine deaminase in the presence of cytidine as the only source of pyrimidines. The mutants were selected by a distinctive white halo surrounding the plaque (17). The authors suggested that the mutant phage with a white halo phenotype is defective in its ability to induce DHFR. In the absence of this enzyme, pools of reduced folate decrease; this in turn inhibits TS activity and reduces conversion of dUMP to dTMP. As a result dUMP accumulates and is excreted in the surrounding medium, where it supports the growth of uninfected cells in the vicinity giving the plaque a white halo phenotype. Later, almost simultaneously it was shown that these mutants with the abnormal plaque morphology are indeed deficient in phage-specific DHFR (18,19).

Most of the mutants are missense but one of these frd mutants, frd 11, is an amber mutant and synthesizes DHFR when grown on amber suppressor; this enzyme shows slightly altered properties from those of DHFR from wild type phage (19). This suggested that gene frd is in fact the structural gene for T4 DHFR.

#### F. Functions of Dihydrofolate Reductase

Dihydrofolate reductase has a crucial role in biosynthesis of thymidine nucleotides (Figure 1). In the reaction catalyzed by thymidylate synthase, methylenetetrahydrofolate ( $\text{CH}_2=\text{FH}_4$ ) is the one carbon donor to produce deoxythymidylate (dTMP) and dihydrofolate ( $\text{FH}_2$ ). Adequate levels of tetrahydrofolate ( $\text{FH}_4$ ) for the TS-catalyzed reaction and also for the biosynthesis of purines and amino acids are provided by the reaction catalyzed by DHFR. Because of this unique role, DHFR is a target enzyme for a number of antineoplastic, anti-parasitic, immunosuppressant and antibacterial agents (20,21). The folate antagonists, such as methotrexate (MTX), trimethoprim (TMP) and aminopterin (Figure 2) are competitive inhibitors of DHFR with respect to the substrate  $\text{FH}_2$ . Inhibition of the reduction of  $\text{FH}_2$  results in decreased pools of deoxyribonucleoside triphosphates (dNTPs), thus reducing the rate of DNA synthesis. However, dTMP can be synthesized by phosphory-

Figure 2. Structures of folic acid and several folate antagonists:

- A. folic acid: N-(para[(2-amino-4-hydroxy-6-pteridinyl)-methyl) amino] benzoyl) glutamic acid
- B. aminopterin: 4-amino-4 deoxyfolic acid
- C. methotrexate: 4-amino-4 deoxy-N10 methyl folic acid
- D. trimethoprim: 2,4-diamino-5-(3',4',5'-trimethoxy-benzoyl)-pyrimidine
- E. Chlorophenyl triazine = 2,4-diamino-1-(3'-chlorophenyl)-6,6-dimethyl-1,6-dihydro-1,3,5,-s-triazine;
- F. butylphenyl triazine: 1-(p-butylphenyl)-1,2-dihydro-2,2-dimethyl-4,6-diamino-s-triazine

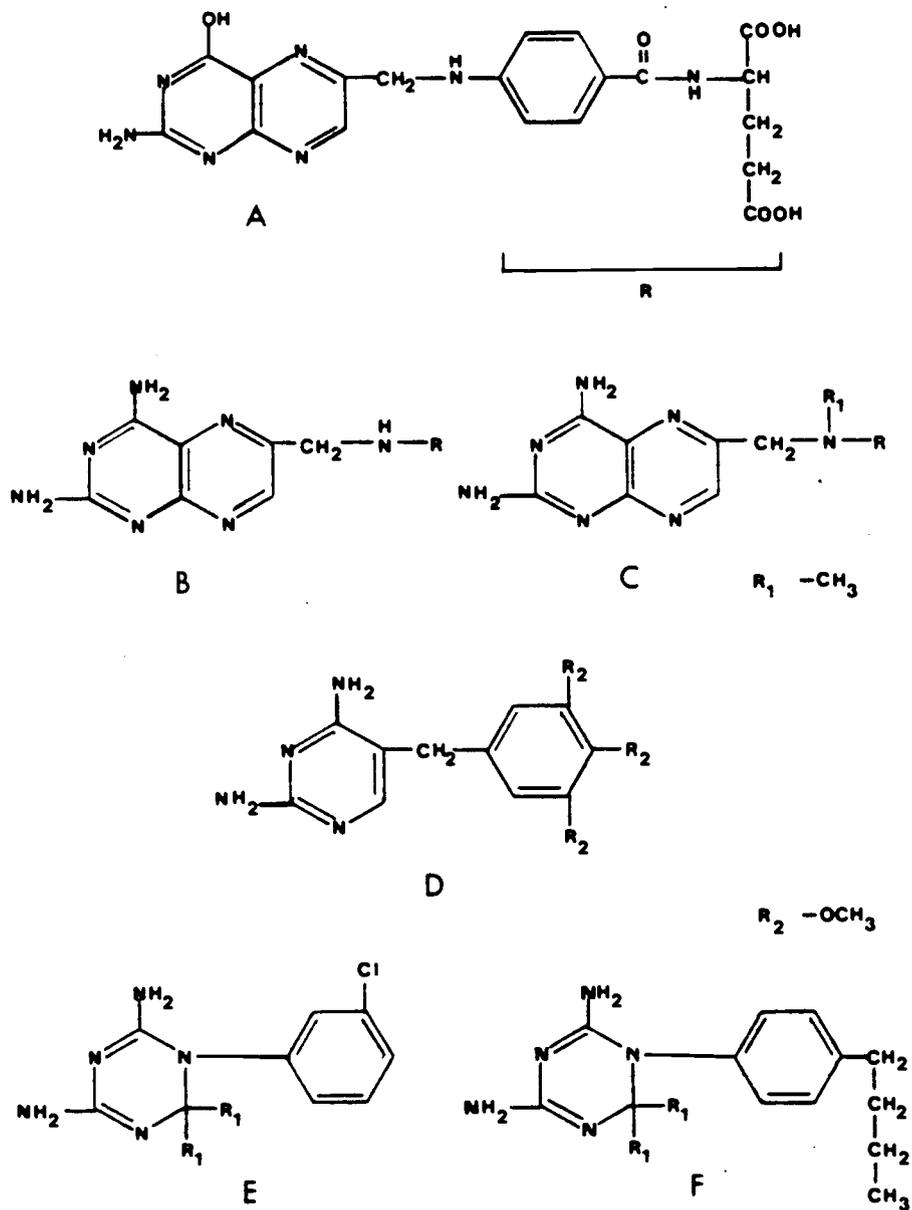


Figure 2. Structures of folic acid and several folate antagonists.

lation of thymidine (TdR) when available, catalyzed by thymidine kinase and all dNTPs can be replenished by salvage pathways. However, the levels of DNA precursor pools are not high enough to sustain the DNA synthesis of fast proliferating cells like neoplasms.

The concentrations of MTX for 50% inhibition ( $I_{50}$ ) for bacterial and vertebrate DHFRs are between 1 and 10 nM (Table I), but  $I_{50}$  for TMP varies by several orders of magnitude (22). The clinical use of TMP as an antibacterial agent is a consequence of its ability to inhibit the DHFR of an invading procaryotic organism at concentrations that do not affect the host enzyme.

A drawback with the continued use of folate analogues is that cells acquire resistance to these drugs and this limits their use (27). Resistant mutants have been used in the study of mechanisms of drug resistance. Based on these studies, a number of different mechanisms have been proposed for the development of resistance to folate analogues in animal cells in culture and also for bacteria.

## G. Mechanisms of Drug Resistance

### 1. Overproduction of DHFR

Foremost on this list is the resistance to drugs by overproduction of DHFR. Overproduction of an enzyme is possible either by gene amplification or by change in

Table I. INHIBITION OF DIHYDROFOLATE REDUCTASES  
BY FOLATE ANTAGONISTS

Enzyme Source	I <sub>50</sub> ( μM )				Refer- ence
	Metho- trexate (MTX)	Trimetho- prim (TMP)	Tria- zines*	TMP/ Tria- zines	
Human Liver	0.009	300	55	5.5x10 <sup>2</sup>	(54)
<u>E. coli</u>	0.006	0.005	650	7.7x10 <sup>16</sup>	(60)
<u>Crithidia fascicu- lata</u>	0.001	15	-	-	(58)
Bacterio- phage T4	0.004	3.3	1.5 <sup>a</sup>	2.2	(14, 95)
Plasmid Type I (R-483)	4.4	57	700	8x10 <sup>-2</sup>	(64)
Plasmid Type II (R-67)	1,100	70,000	800	87.5	(60)
<u>Plasmo- dium berghei</u>	-	0.0085	0.0009	9.4	(54)

\* 1-(p-butylphenyl)-s-triazine (Figure 2)

<sup>a</sup> 1-(3'-chlorophenyl)-s-triazine (Figure 2)

regulation of the expression of the gene.

(a) Overproduction of DHFR by Gene Amplification

Hakala et al. (28) in 1961, reported that cells in culture become resistant to MTX when grown in stepwise increasing concentrations of the drug. These MTX-resistant cells show elevated levels of DHFR. This increase in DHFR has been linked to gene amplification (29, 30, 31). In some murine sublines the amplified genes have been stably incorporated into a single abnormal chromosome in tandem (32, 33). In other MTX-resistant cell lines the amplified genes are associated with small chromosomal elements, the "double minute chromosomes." These genes are progressively lost when the selection pressure is removed (34,35).

(b) Overproduction of DHFR due to alteration in Regulatory Mechanisms

Some of the TMP-resistant mutants of E. coli isolated by Sheldon and Brenner (36) demonstrated increased levels of DHFR. One of these mutants, fol-60, showed both TMP-resistant (TMP<sup>R</sup>) and temperature-sensitive (ts) phenotypes. Smith and Calvo (37) demonstrated that increased levels of DHFR in this mutant were the result of elevated levels of fol mRNA produced due to increased rate of transcription. But the possibility of gene amplification was not addressed. By nucleotide sequence determinations of the fol gene from a number of TMP<sup>R</sup>

mutants which were isolated by Sheldon and Brenner (36) including fol-60 (38) the authors attributed the increased transcription of fol gene as a consequence of mutation in the -35 region of the fol promoter (39). The ts phenotype of fol-60 was explained by a second mutation outside the fol coding region.

Earlier work of Sirotnak and McCuen (40) with aminopterin-resistant mutants of Diplococcus pneumoniae also suggests that the hyperproduction of DHFR may be due to transcriptional control

## 2. Synthesis of Altered Enzyme and Autoregulation

In some cases resistance to folate analogues is acquired due to synthesis of an altered enzyme that shows decreased affinity for a particular drug. Low affinity of DHFR for MTS has been reported in murine sublines of L5178 Y (41,42), 3T6 mouse embryofibroblasts (43), and D. pneumoniae (44), and for pyrimethamine in Streptococcus faecium (45) and also in the protozoan Plasmodium berghei (46). These altered enzymes with low affinities to folate analogues also show changes in  $K_m$ 's and pH optima (22). D. pneumoniae cells with mutations in the DHFR structural gene also show increased synthesis of DHFR (47). It seems that the enzyme formation is autoregulated (48). Gronenborn and Davies (49) have demonstrated that DHFR from Lactobacillus casei binds to DNA in general, and with increased affinity to DNA carrying

the gene for DHFR. These results suggest that it may also regulate its own synthesis (49).

### 3. Altered Transport System

Methotrexate is actively transported and enters cells by the carrier system of reduced folates (50, 51, 52). In some cases resistance to MTX is developed by modifications of the folate transport system (53, 54, 55). In other drug resistant bacteria, the resistance is acquired by production of drug-resistant enzymes different from those produced by the host. These enzymes are coded by drug resistance factors (56, 57) and will be discussed in detail in the next section.

## H. Subunit Structure and Properties of Dihydrofolate Reductase

### 1. Monomeric Enzymes

Dihydrofolate reductases from bacterial and vertebrate sources are monomers of 18-22,000 daltons and contain 159-189 amino acids residues. These enzymes have similar  $K_m$ s for NADPH but differ in most respects (22). The vertebrate enzymes show two pH optima, one at pH 4-5.5 and another between pH 7-8.3, whereas the pH optima for bacterial enzymes are between pH 6-7 (Table II). Vertebrate enzymes are activated by mercurials that have no effect on bacterial enzymes. Differences in inhibition of bacterial and phage enzymes

Table II. PROPERTIES OF DIHYDROFOLATE REDUCTASE

Enzyme Source	Molecular Weight	Number of Subunits	pH Optima	Reference
Human liver	22,000	1	5.5, 7.8	23
<u>E. coli</u>	18,500	1	7	24, 58
<u>C. fasciculata</u>	107,000	2	-	25
Bacteriophage T4	44,500	2	7	59
R-Plasmid Type I (R-483)	35,000	2	-	24
R-Plasmid Type II (R-67)	37,000	4	-	24
R-Plasmid Type III	16,900	1	-	60

by folate antagonists have been discussed earlier.

In contrast to bacterial and vertebrate DHFRs, the enzymes coded by protozoa, drug resistance factors and by bacteriophage T4 are multimeric and have higher molecular weights. These DHFRs differ from monomeric enzymes in drug resistance (Table I). Also there are pronounced differences in the inhibition pattern of bacterial and vertebrate enzymes by TMP and triazines. Bacterial enzymes are much more sensitive to TMP as compared to vertebrate DHFRs, whereas the opposite response is observed for triazines (Table I and ref. 23).

## 2. Multimeric Enzymes

### (a) Dihydrofolate Reductases from Protozoa

The molecular weights of enzymes from certain parasites, such as trypanosomes and crithidia (61), plasmodia (62), coccidia (63), and Leishmania (31), range between 100,000-240,000. Ferone and Roland (25) have demonstrated that the purified enzyme from Crithidia fasciculata occurs as a bifunctional protein. The DHFR protein also possesses TS activity. This protein is a dimer of MW 107,000 and is composed of two identical subunits each of approximately 56,700. Dihydrofolate reductase and TS from P. berghei also coelute from a Sephadex G-200 column (25) and an apparent molecular weight of 150,000 has been calculated.

(b) Dihydrofolate Reductase Encoded by Drug  
Resistance Factors

In 1972, Fleming et al. (64) reported that resistance to TMP in E. coli and Klebsiella aerogenes is conferred by drug resistance factors, the "R-plasmids." It was later on demonstrated (56, 57) that these R-plasmids encode DHFRs that are different from the chromosomal enzymes. Investigators from Burroughs Wellcome Laboratories (24, 60) classified the R-plasmid coded enzymes into three classes on the basis of their subunit structures and drug resistance characteristics. Type I enzymes are dimers of about 35,000 daltons (65) and are moderately resistant to TMP, whereas Type II enzymes, although they have molecular weights similar to those of Type I, are tetramers of identical subunits (66). However, Type III enzyme (60) is a monomer with molecular weight similar to that of bacterial enzymes but is slightly more resistant to TMP than the host enzyme.

Type II dihydrofolate reductases coded by drug resistance plasmids R-67 and R-388 show greater than 70% sequence homology with each other but no homology has been observed with bacterial, vertebrate or the Type I enzyme coded by another R-plasmid, R-483 (67, 68, 69, 70). However, Type I enzyme coded by R-483 shows 30-35% sequence homology with E. coli DHFR (70).

(c) Dihydrofolate Reductase from Bacteriophage T4

Bacteriophage T4 coded DHFR is a dimer of 44,500 daltons with a subunit molecular weight of 23,000 (59). The  $I_{50}$  for TMP of this phage enzyme is also similar to those of R-plasmid Type I enzymes. However, the bacteriophage T4 enzyme does not cross-react immunologically with the enzyme coded by plasmid R-483 (71). DHFRs coded by T4 and R-483 also differ in MTX binding.

I. Primary Structure of Dihydrofolate Reductases

By comparing the amino acid sequences of E. coli and S. faecium DHFRs, it was concluded that the N-terminal region is conserved (72). Since then, amino acid and nucleotide sequences for several of the bacterial, vertebrate and some of the R-plasmid-induced DHFRs have been determined (reviewed in 22, 70). Dihydrofolate reductases from several sources show greater than 70% homology in their amino acid sequence but only less than 30% with bacterial DHFRs. Also, sequence homology among different bacterial DHFRs is less than 30%. Amino acid sequences of R-plasmid Type II enzymes from two different plasmids, R-388 and R-67, although highly homologous, do not show any homology to bacterial or vertebrate enzymes (68, 67, 69). However, like bacterial DHFRs, R-plasmid Type I enzyme (R-483) shows about 30% sequence homology to bacterial as well as to vertebrate enzymes (70).

In the last few years, three-dimensional structures of E. coli DHFR-MTX binary complex (73), L. casei DHFR-MTX-NADPH ternary complex (74), as well as one of the vertebrate enzymes, the chicken liver DHFR-phenyl triazine ternary complex (75) have been determined by X-Ray crystallography. Interestingly, irrespective of the size, amino acid sequence and the nature of the complex (binary or ternary), overall backbone foldings of bacterial and vertebrate dihydrofolate reductases are similar (73, 74, 75). The main structural feature of these DHFRs is an eight-stranded  $\beta$ -sheet, beginning at the amino terminus and ending with a single antiparallel strand at the carboxy terminus. They all contain four  $\alpha$ -helices and the remaining residues are involved in the formation of the connecting loops.

Volz et al. (75) aligned amino acid sequence of seven DHFRs, three bacterial and four vertebrate enzymes. This alignment was done according to the X-Ray structure equivalence of three of these enzymes, from E. coli, L. casei and chicken liver. Most of the insertions and deletions are in the loop regions connecting different secondary structures. At this point, it is worth mentioning that the residues involved in substrate and coenzyme binding are conserved and do not fall in this connecting region.

About 21 residues of L. casei DHFR have been implicated in coenzyme interaction and 17 residues in MTX or substrate binding (22). Two residues, ala 6 and leu 19 (met 1 in E. coli), are highly conserved and are common to both coenzyme and MTX interactions. Additional highly conserved residues that are involved in contact with coenzyme are ile 13, gly 14, trp 21, gly 42, thr 45, leu 62, ala 97, gly 98, gly 99 and phe 103 (tyr in E. coli and chicken liver) and those that interact with the inhibitor are asp 26 (glu in vertebrate and R-plasmid Type I enzymes), phe 30, ser 48, leu or lys 54 except in avian where the corresponding residue is glutamine, arg 57, thr 116 and thr 146. More than 50% of these conserved residues are present in the first 50 N-terminal residues. The rest are distributed throughout the molecule. Only three of a total of 19 highly conserved residues are present in the C-terminal third of the molecule.

## J. Importance and Interest in the Study of Bacteriophage

### T4 Dihydrofolate Reductase

#### 1. General

For decades, folate analogues have been used as clinical reagents for the treatment of a variety of parasitic infections, but the mechanisms of their actions were not known. Large quantities of purified protein are required for studies of interactions of these folate

analogues with the target enzyme, DHFR. About twenty years ago, these studies were not feasible because of the low level of the enzyme in target tissues. The discovery that bacteriophage T4 codes for DHFR and that a 20-fold excess of protein is synthesized upon infection (12) was of great advantage.

Erickson and Mathews (76) purified the bacteriophage T4-induced DHFR to homogeneity by affinity chromatography. The affinity ligand was the folate analog N<sup>10</sup>-formylaminopterin. This principle was later used for the purification of DHFR from a variety of sources (22).

Electrophoretically homogeneous T4 DHFR was used as a model system for studying the binding of different ligands (77). The data suggested that the highly basic N<sup>1</sup> of the pteridine moiety of MTX interacts with an acidic residue present in the active site of DHFR. This acidic residue was later identified by X-Ray crystallography as asp-27 of E. coli dihydrofolate reductase (73).

## 2. Dihydrofolate Reductase as the Structural Protein

Interest in the study of bacteriophage T4 dihydrofolate reductase was enhanced by the finding that T4 DHFR is a structural component of the T4 tailfiber baseplate (78, 79). These authors demonstrated that urea-

treated tail plate substructures of bacteriophage T4 contain the total DHFR activity that was present in the phage ghost (78).

For some time the identity and origin of this structural DHFR were much debated. The reasons for this uncertainty were a number of contradictory properties of structural and soluble DHFRs (80).

The missense mutants, such as frd 1, in the gene for DHFR, synthesize inactive enzyme. The isolated tail structures from this mutant also lack the DHFR activity. Furthermore, phage infectivity is inhibited by incubation of phage particles with NADPH (78) which is a substrate for DHFRs (13). These data suggest that both structural and soluble enzymes are the same protein. However, reversal of inactivation by  $\text{NADP}^+$ , inactivation of frd I mutants lacking the soluble DHFR by  $\text{NADP}^+$ , inactivation of phage particles by incubating with  $\text{NADPH}^+$  although the soluble enzyme has no folate reductase activity (14) and lastly, the inactivation of phage particles more rapidly at pH 5.0, than at pH 7.3, even when the soluble enzyme has no activity at low pH, suggests that the structural protein is different from the soluble enzyme.

Based on the genetic data, it has been conclusively shown that both of these enzymes, soluble and structural DHFRs, are the products of the same gene (80). Mathews (80) constructed a hybrid phage T4 frd<sup>T6</sup> that

carries the frd gene from T6 in T4 genetic background. It was accomplished by crossing T4 missense frd mutant, frd 2 (18) with wild type bacteriophage T6, and a number of back crossings with T4 frd 2, each time selecting for the T6-specific DHFR plus phenotype. As mentioned earlier, DHFRs from these phages differ in temperature sensitivity. The soluble enzyme from the hybrid phage, T4 frd<sup>T6</sup>, shows the properties of T6 enzyme. Furthermore, these hybrid particles were heat inactivated with kinetics similar to those of wild type bacteriophage T6. Interestingly, although the T6 enzyme is more sensitive to heat, both wild type T6 and the hybrid T4 frd<sup>T6</sup> particles are heat inactivated at a slower rate than that of the wild type T4. This result implies that the properties of the protein change when it is present in a rigid structure like the tail plate (80). This interpretation also explains the anomalous properties of the structural DHFR observed by Kozloff et al. (78).

Further evidence in support of the two proteins being the same molecular entities comes from studies of inactivation of phage particles by antiserum raised against homogeneous DHFR (61). Similar results were also obtained by Dawes and Goldberg (82). These authors also showed that the structural DHFR functions in the early stages of infection, and for this role active enzyme is not required. This interpretation was sup-

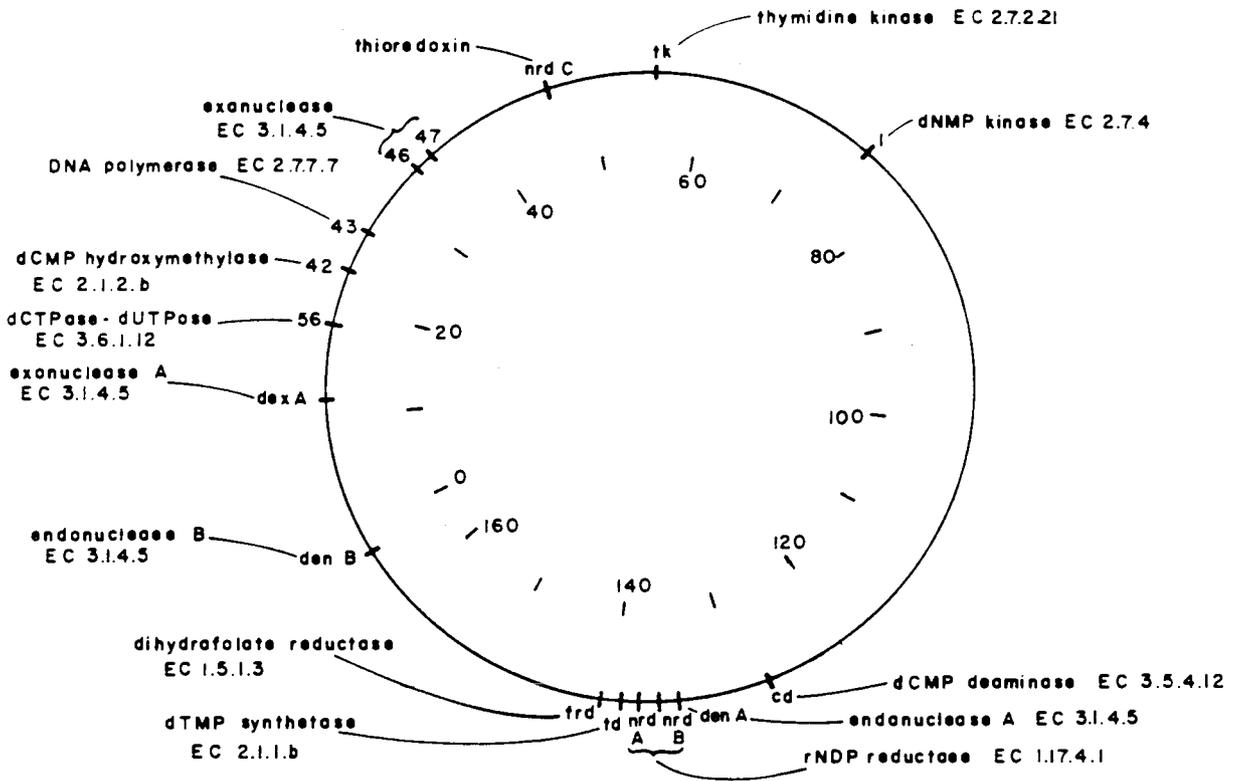


Figure 3. Map positions on frd and td on partial map of T4 showing early genes controlling reactions of DNA precursor biosynthesis. Numbers in the interior represent distances in kilobases from the rIIA/rIIB cistron divide (6).

ported by the production of phage particles containing bound aminopterin, a strong competitive inhibitor of DHFR, in place of the normal pteroylhexaglutamate. Pteroylhexaglutamate has been shown as the component of bacteriophage T4 tail baseplate (83, 84). Mosher and Mathews (85), using labelled anti DHFR antibodies, demonstrated that DHFR is a component of outer wedges of the baseplate.

### 3. Thymidylate Synthase as a Structural Protein

Thymidylate synthase, like DHFR, is a phage-coded early enzyme and duplicates the preexisting host functions. It is functionally associated with DHFR in thymidine nucleotide biosynthesis and the gene td, that codes for thymidylate synthase, maps very near to frd (Figure 3) on the T4 genetic map (17).

Thymidylate synthetase from bacteriophage T4, like other known thymidylate synthases, is a dimer of 58,000 with a subunit  $M_r$  of 29,000. The antiserum against purified T4 thymidylate synthase (86, 87) neutralizes phage T4 infectivity. No neutralization of bacteriophage T2 and T6 infectivity has been observed although this antiserum cross reacts with the soluble TS from these phages on double diffusion plates.

These observations are interesting for two reasons. First they suggest that TS, like DHFR, is a structural protein. Second, the soluble and structural thymidylate

synthases from bacteriophage T2 and T6 show differences in reactivity towards antiserum raised against purified bacteriophage T4 TS. This supports the hypothesis (80) that the incorporation of a protein in a rigid structure results in a conformational change in the protein structure. It is further supported by the inactivation studies of hybrid phage T4 $\underline{td}^{T6}$  constructed by Capco and Mathews (87) by a technique similar to that used for T4 $\underline{frd}^{T6}$ , using in this case the T4 $\underline{td}8$  mutant. This hybrid phage is inactivated by antiserum produced against purified DHFR and TS. Inactivation by antibodies against DHFR is expected because of its T4 genetic background. However, inactivation by TS antiserum is interesting because phage T6 is not inactivated by anti TS antibodies.

Altered temperature sensitivity of the hybrid T4 $\underline{td}^{T6}$  (87) and  $\underline{ts}$  mutants in  $\underline{td}$  (88) supports the above inference of TS as a structural component and that more likely it is present in the baseplate. It has been demonstrated that TS is in fact located in the hub of the tail baseplate (89, 90, 79).

#### 4. Regulation of Bacteriophage T4 Dihydrofolate Reductase

Mathews (16) and Warner and Lewis (15) observed that DO amber mutants of bacteriophage T4, defective in DNA synthesis, overproduce some of the early enzymes but

show no effect on the activity of DHFR.

The genes of DHFR and a number of other early enzymes map in a cluster on the T4 linkage map (Figure 3). Restriction mapping of the T4 genome (91, 92) and cloning of T4 DNA (93) located the frd gene in a 1.1 Kb Hind III fragment, about 143 Kb clockwise from the rIIA/rIIB cistron divide. It has been demonstrated (92, 93) that this fragment also carries a part of the gene for TS.

Bacteriophage T4 genes are divided into three main groups based upon time of expression: early, middle and late. Regulation of expression of these genes has been extensively reviewed (94, 95, 96). For the expression of early genes, T4-specific DNA synthesis is not required. Genes belonging to this group are mainly those required for DNA precursor biosynthesis and DNA replication. The early genes are further subdivided into immediate early (IE), and delayed early (DE). Immediate early genes are expressed within 30 seconds after infection and their expression is chloramphenicol (CAM) insensitive. Both unmodified and modified forms of the host RNA polymerase holoenzymes transcribe these genes both in vivo and in vitro. Delayed early genes, on the other hand, are expressed approximately two minutes after infection and are CAM sensitive. Expression of some of the immediate early genes is required for the expression of the delayed early

genes. Chloramphenicol sensitivity is attributed to the termination of transcription at rho sites when translation is not taking place or proceeding at a very slow rate. By the above mentioned criteria, frd is an IE gene and td a delayed early (26).

The regulation of early genes is very complex. The early transcription units which constitute both IE and DE genes are in some way similar to bacterial operons. These are transcribed from a single promoter and are polycistronic. These transcription units are approximately 1,500 to 2,000 nucleotides long (97). However, contrary to the operon model, these genes may also be transcribed from additional promoters located upstream of a particular gene in the transcription unit. An example of such a unit is the rII region between gene 39 and rIIB (96). This transcription unit consists of genes 39, 60, rIIA and rIIB, and is transcribed from the early promoter (PE) located upstream of gene 39 (Figure 4). Both rIIA and rIIB genes are also transcribed from another promoter, the PrIIA as a single transcript. However, a third promoter, the middle promoter (PM) is utilized for the transcription of rIIB. All of these transcripts terminate at the same rho independent termination site present at the end of the transcription unit; however, there is a rho dependent termination site at the IE/DE junction.

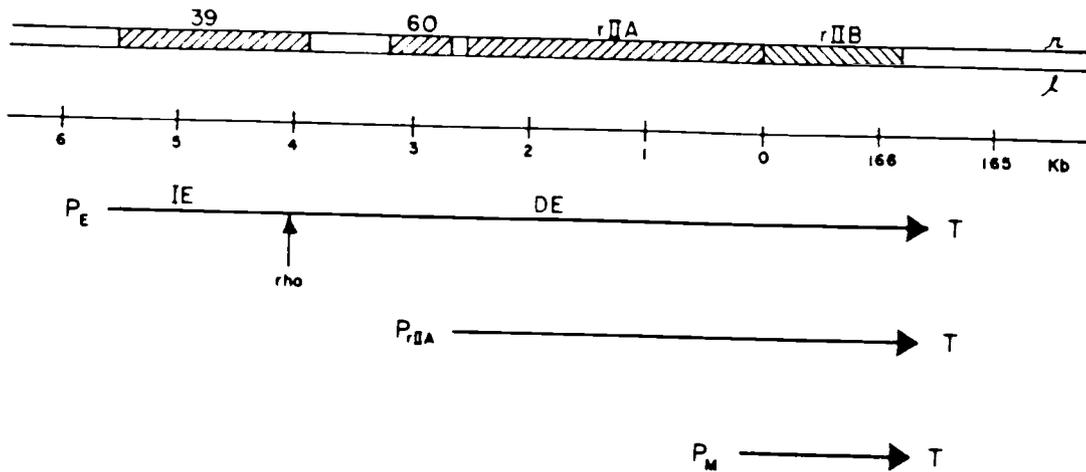


Figure 4. Early transcription unit of the rII region of bacteriophage T4 showing multiple promoters. P<sub>E</sub>, early promoter; P<sub>rIIA</sub>, presumed promoter for the transcription of rII genes; P<sub>M</sub>, middle promoter. Horizontal arrow indicates the direction of 1 strand specific transcription; the vertical arrow shows the site of proposed rho dependent termination site and marks the IE (immediate early) - DE (delayed early) division of this early transcription unit (96).

Both for the readthrough at IE/DE junctions and for the recognition of middle promoters, gene product mot is required (98, 99). mot itself is an immediate early gene and codes for a protein of approximately 24,500 daltons (96). A number of mutants in the mot gene have been isolated in different laboratories. These mutants exhibit altered expression of several genes. The folate analog-resistant mutants far P14 and far P85 (100, 101) are regulatory mutants. These mutants overproduce DHFR and a number of early enzymes, such as dCMP deaminase, thymidine kinase, dCTPase, and gene products of rIIA and rIIB (101, 102) but show decreased synthesis of the products of genes td, 32, 43, 45, (102, 103). Other mutants that map in the mot gene, near far mutations, are tsG<sub>1</sub>, amG<sub>1</sub> and sip (104). These mutants also show effects similar to those mentioned above for far regulatory mutants, except that tsG<sub>1</sub> causes delayed synthesis of rIIB gene product as opposed to overproduction in far mutants (99, 102). The levels of DHFR and TS in tsG<sub>1</sub> and amG<sub>1</sub> have not been tested.

Some of the early and middle promoters have been sequenced (96). Although a few of these promoters have been mapped, e.g., rIIB (105), others have been identified by their proximity to Shine-Dalgarno sequence (106, 107) and by sequence homology to the consensus sequence of E. coli promoters (39). All of these promoter sequences show good homology in the region corresponding to the Pribnow Box (-10 region) of the E. coli promoter.

Reasonable similarities are observed in the -35 region of the early promoters to the corresponding E. coli consensus sequence, but the middle promoter sequences in this region are quite different. The possibility of a new consensus sequence " A  $\begin{matrix} T \\ A \end{matrix}$  G C T T " as the gp mot recognition site has been discussed (96).

##### 5. Deletion Mutants in Genes 63-32 Region

It was observed that phage carrying a duplication of the rII region (108, 109, 110, 111) often contained deletions (109, 112). Homyk and Weil (113) used this knowledge and isolated a number of deletion mutants in nonessential regions of the T4 genome. They crossed two deletion mutants, r1589 (114) and r1236 (115), with overlapping deletion in the rII region. The mutant, r1589, although deleted in part of rIIA and rIIB cistron, retains cistron B activity, whereas r1236 has a deletion in cistron B. Although both mutants complement each other functionally, no wild type phage recombinant can be obtained by such crosses. Thus the mutant phage with rII duplications were selected by their ability to form plaques on E. coli lambda lysogen that restricts the growth of rII mutants. Phage particles with abnormal plaque morphology were selected as the recombinant phage with possible deletions.

Several deletion mutants in the non-essential regions between genes 39-56 and between genes 63-32

regions were isolated. The size and locations of the deletions were identified by heteroduplex mapping. Possible functions of the deleted genes were further analyzed by enzymatic assays.

The 63-32 region spans a cluster of early genes (Figures 3, 5) frd, td, nrdA, nrdB and denA. All except denA duplicate pre-existing activities in host cells. As mentioned earlier, both frd and td gene products are also structural components of the phage tail baseplate. Three of the deletion mutants in the 63-32 region have been studied extensively. These mutants, del1, del7, del9 (113), although viable, grow poorly in general and are restricted completely by several wild type E. coli strains in the Cal. Tech. Collection (71). These Cal. Tech. strains do not restrict either wild type T4D or the parent mutant, r1589. These three deletion mutants lack TS and rNDP reductase activities (86, 71, 113, 117). In addition del 1 lacks endonuclease II (113) whereas del7 and del9 both are deficient in DHFR activity (86, 71).

The lengths of these deletions of about 4,800 base pairs for del1 and approximately 3,000 base pairs each for del7 and del9 have been determined by heteroduplex mapping (113). By these estimates the genes coding for TS from del1 and both TS and DHFR from del7 and del9 should be deleted. However, all these three mutants synthesize peptides that cross react with antibodies raised against

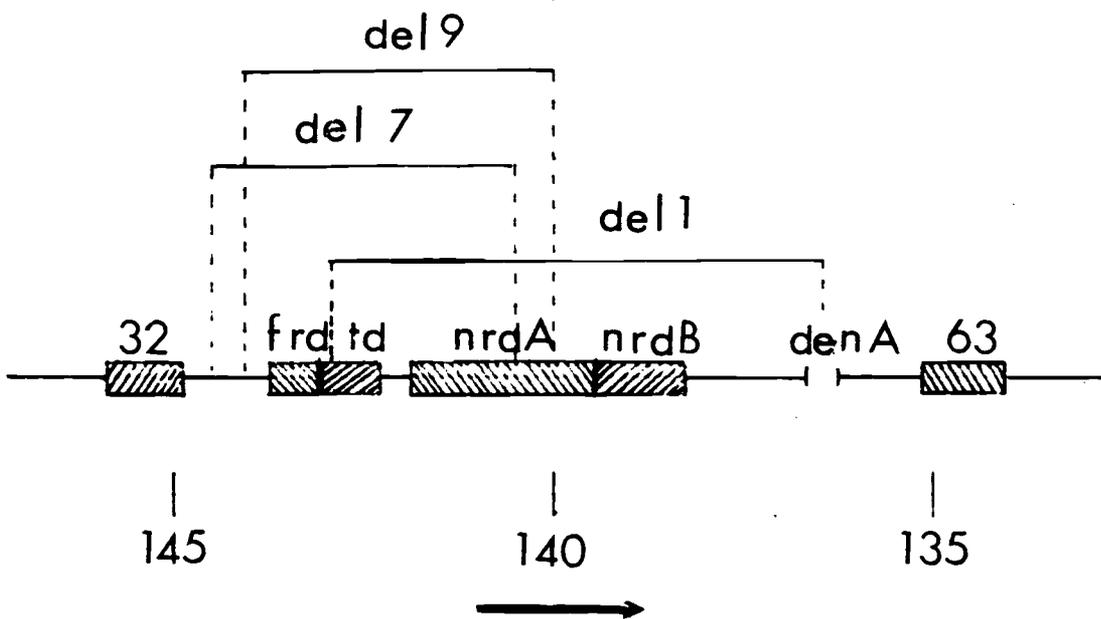


Figure 5. Map of 63-32 region of T4 showing deletions in del1, del7 and del9 mutants. The end points of each deletion (dotted lines) are taken from the data of Homyk and Weil (112). The positions of the genes are taken from Kutter and Rüger (116). All genes except denA have been mapped by restriction analysis or analysis of the cloned fragment (116). The arrow points the direction of transcription from l strand. Numbers indicate kilobase pairs from rIIA/rIIB cistron divide.

purified TS (86) and DHFR (81). The immunologic cross reactivity was demonstrated both by immunodiffusion and by inactivation of phage infectivity (86, 71). That these mutants synthesize peptides similar to DHFR was also shown by heat inactivation of the phage particles (227).

The possibilities that td and frd genes are not deleted from these mutants have been extensively discussed (71, 117).

#### K. Mutants of Bacteriophage T4 that Synthesize Cytosine Containing DNA

As mentioned earlier, the DNA of bacteriophage T4 contains a modified base, hydroxymethylcytosine, instead of cytosine. The presence of this modified base renders the phage DNA resistant to most restriction endonucleases. By combined efforts of several investigators (118, 119), multiple mutant phage containing deoxycytidine instead of HMdC DNA has been constructed. The sequence of reactions necessary for the synthesis of HMdC DNA and the sites of mutations to construct the multiple mutant phage are shown in Figure 6. In wild type T4, for the synthesis of HMdC DNA, the phage codes for several enzymes. The products of gene 56, dCTPase (120, degrades both dCTP and dCDP to dCMP and thus indirectly inhibits the incorporation of deoxycytidine in the phage DNA. The enzyme dCMP HMase, product of gene 42 (121), converts dCMP to HMdCMP. Methylene tetrahydrofolate is the one carbon donor in this reaction.

Figure 6. Generation of viable bacteriophage T4 containing dCDNA.

- (1) Ribonucleoside diphosphate reductase
- (2) Nucleoside diphosphate kinase
- (3) DNA polymerase
- (4) dCTPase
- (5) dCMP hydroxymethylase
- (6) dNMP kinase
- (7) Endonucleases II and IV

= indicates the site of mutation

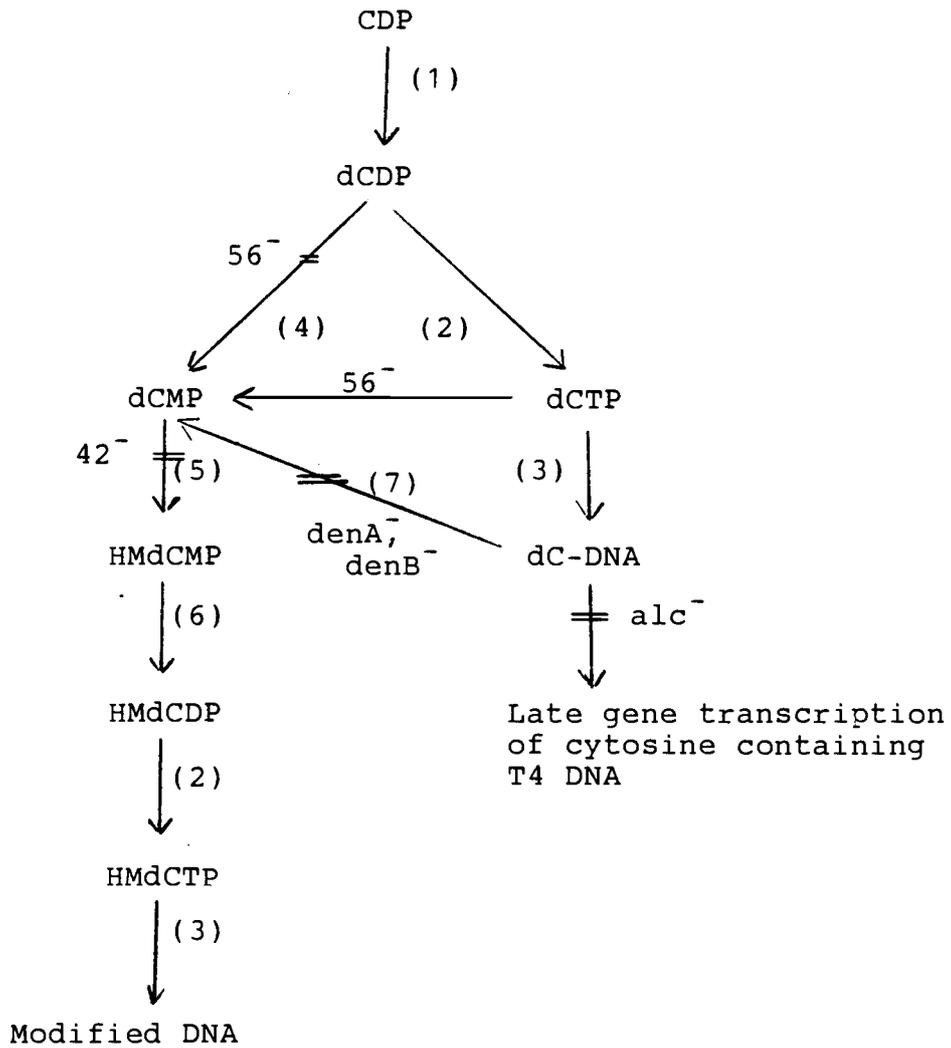


Figure 6. Generation of viable Bacteriophage T4 containing dCDNA.

Endonucleases, endo II and endo IV, coded by genes denA and denB respectively, degrade either host or phage dC DNA (122). However, denB is more active on phage dC DNA than that of host (123, 123) and the reverse is true for denA (125, 126). Mutants lacking in functions of genes 56, 42, and/or denA and denB, however, are capable of synthesizing cytosine-containing DNA, and do not produce viable phage particles because the product of another gene, the alc gene, does not allow late gene transcription of cytosine-containing T4 DNA (118), and hence no late proteins are synthesized. Thus a multiple mutant, dec8, that lacks the functions of 56, denA, denB and alc, constructed by Snyder et al., was used as the source of dC DNA in the present work. This mutant, because of the lack of mutation in gene 42, contains dC DNA as well as HMdC DNA, and thus, although not completely, can be digested with restriction endonucleases (118).

#### L. Present Work

The work presented in this dissertation deals with the investigation of the primary structure of frd, the gene for bacteriophage T4 dihydrofolate reductase and its flanking regions, which contain part of td, the gene for thymidylate synthase. The investigation includes the isolation and cloning of a 1.1-kb fragment (59) and determination of its nucleotide sequence. During the

course of this investigation, I improved the technique to isolate restriction fragments from agarose gel, and also standardized the conditions for selection of clones carrying the active gene for T4 dihydrofolate reductase. The cloned fragment was inserted in a hyperexpression vector that resulted in greater than 300-fold overproduction of DHFR. Also, the possible regulatory role of the flanking sequence 5' to frd was investigated by cloning of MbO II/Hind III fragment which lacks most of these sequences into pBR322. In addition, DNAs from deletion mutants, del7 and del9, were probed for sequences complementary to the cloned 1.1-kb fragment (127, 128).

## II. MATERIALS

### A. Reagents

All restriction enzymes used in this work were purchased from Bethesda Research Laboratories except EcoR V which was purchased from New England Biolabs. The commercial sources of other enzymes and reagents are as listed below.

<u>Reagents</u>	<u>Commercial Source</u>
<u>DNA Modifying Enzymes</u>	
DNA Polymerase I	New England Nuclear
DNA Polymerase I (large fragment)	New England Nuclear
Pancreatic DNase I	Worthington Biochemical
Bacterial Alkaline Phosphatase	Worthington Biochemical
T4 Polynucleotide Kinase	P-L Biochemicals
T4 DNA Ligase	Bethesda Research Labs.
<u>Reagents for Electrophoresis</u>	
Agarose	Bio-Rad
Acrylamide	Bio-Rad
Bis-acrylamide	Bio-Rad
Ammonium Persulfate	Bio-Rad
TEMED	Bio-Rad
Urea	R-Plus Laboratories

ReagentsCommercial SourceReagents for Chemical Modifications and Cleavage

Hydrazine, 95%	Eastman Organic Chemicals
Piperidine, 99%	Fisher Scientific
Dimethyl Sulfate, 99%	Aldrich Chemical Company

Radioactively Labelled Nucleotides

( $\gamma$ - $^{32}\text{P}$ ) ATP 7,000 Ci/m mole	New England Nuclear, Radioisotope Division
( $\alpha$ - $^{32}\text{P}$ ) dCTP 600 Ci/m mole	New England Nuclear, Radioisotope Division

The reagents for media such as Nutrient Broth, Bacto Agar, Bacto Tryptone, Bacto Yeast, etc., were purchased from DIFCO.

All other chemicals used were of reagent grade.

Dihydrofolate was synthesized in the laboratory by the procedure of Futterman as modified by Friedkin et al. (129).

B. Bacterial Strains

E. coli B 201 ( $\text{thy}^- \text{Su}^-$ ) is the permissive host used for growth of bacteriophage T4 amber mutants. E. coli B 40 ( $\text{Su}^+$ ) is the nonpermissive host for phage amber mutants. E. coli B 834 ( $\text{r}_\text{B}^- \text{m}_\text{B}^- \text{met}^- \text{Su}^- \text{rgl}^+$ ) is the permissive host for the growth of dec8, a T4 multiple mutant used for making cytosine-substituted phage DNA. E. coli CR 63 ( $\text{Su}^\pm$ ) is the permissive and E. coli CT526 is the

restrictive host for the deletion mutants of T4 in 63-32 region. E. coli K38 ( ), a K12 strain restrictive for the growth of rII mutants is  $r_k^- m_k^- rgl^-$ . E. coli RRI ( $Su^- R^- M^- thr^- leu^- pro^- bio^-$ ) is used for transformation of plasmid DNA.

### C. Bacteriophage T4 Strains

All strains used are in T4 D background. These include BL 292 (amber mutant in gene 55), dec8, the T4 multiple mutant which is 56<sup>-</sup> (am E51 dCTPase<sup>-</sup>), denA (nd 28), denB (rII H23B) alc8; and deletion mutants, del (63-32)1; del (63-32)7; and del (63-32)9.

### D. Plasmid Vectors

During the course of this research two plasmid vectors were used, pBR322 constructed by Bolivar et al. (130), and the hyperexpression vector pGW7 in E. coli RR1. pGW7 was kindly supplied by Geoff Wilson, now at New England Biolabs.

The T4 multiple mutant dec8 and its host E. coli B 834 were obtained from Larry Snyder (Michigan State University), and all three deletion mutants in 63-32 region from L. Homyk and J. Weil (Vanderbilt University). These strains have been maintained in C. K. Mathews's laboratory, Department of Biochemistry and Biophysics, Oregon State University.

The compositions of media used are listed in the Appendix.

### III. PROCEDURES

#### A. Preparation of High Titer Phage Stocks

The bacterial culture was incubated at 37° with aeration and was allowed to reach a cell density of  $2-3 \times 10^8$  cells/ml (131). L-tryptophan was added to give a concentration of 20 µg/ml and the cells were infected with the phage at a multiplicity of infection of 0.1-0.5. The culture was incubated for about 5 hours and complete lysis of the culture was achieved by adding a few drops of chloroform. Cellular nucleic acids were degraded by incubating the lysed culture overnight at room temperature with DNase I and RNase A, each at 1 µg/ml, and in the presence of 0.04%  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ . The flask was chilled in ice and the opalescent liquid was centrifuged at  $4,000 \times g$  for 5 minutes. The titers of phage preparations varied from  $10^{11}$  to  $10^{12}$  plaque forming units per milliliter. The phage stocks were stored at 4°: all centrifugations were done at 4° unless otherwise specified.

#### B. Purification of Phage Particles

The phage particles in a crude phage lysate were concentrated by centrifugation at  $23,000 \times g$  for 2 hours. The pellet containing the phage particles was suspended

in one tenth of the original volume in M9 salts solution containing 0.5%  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  and the suspension was again centrifuged at 35,000 x g for 30 minutes. The pellet was resuspended in M9 salts as above and centrifuged at 4,000 x g for 5 minutes to remove the remaining cellular debris. Phage particles from the supernatant were collected by centrifugation at 35,000 x g for 30 minutes and were suspended in M9 salts solution to give one twentieth of the original volume.

### C. Isolation of DNA from Phage Particles

The purified phage suspension was mixed with an equal volume of redistilled phenol which had been pre-equilibrated with 1 M Tris-HCl, pH 8.0. The phenol and aqueous phases were separated by centrifugation at 3,000 x g for 5 minutes at room temperature. The top aqueous phase containing DNA was transferred to another tube, leaving the white interphase containing proteins undisturbed. This procedure was repeated about 4 - 5 times until the interphase was no longer turbid. The contaminating phenol was removed by extraction with a mixture of chloroform: isoamylalcohol (24:1), followed by a few extractions with diethyl ether. Ether was removed from the aqueous phase by gently blowing in nitrogen or argon. DNA was precipitated by adding sodium acetate (NaOAc) to 0.3 M and three volumes of 95% ethanol and collected by

centrifugation at 27,000 x g for 15 minutes. The DNA at this stage can be stored by resuspending in 0.3 M NaOAc and 3 volumes of ethanol at -20° or can be used for further studies. For restriction enzyme digestion the DNA suspension was centrifuged and rinsed with 95% ethanol. After removing the ethanol, it was suspended in DNA suspension buffer (TE buffer; 10mM Tris-HCl, 1mM EDTA, pH 7.6). Approximately 20 mg of cytosine containing DNA (T4 dC-DNA) was obtained from one liter of B834 culture infected with bacteriophage T4 strain dec8.

The ratio of absorbances at 260/280 nm of the DNA solution was 1.94 and the concentration of DNA was calculated by assuming that 50 µg DNA/ml gives an absorbance of 1.0 at 260 nm with a light path of 1 cm. All glass and plastic wares used for DNA isolation were autoclaved and all solutions were sterilized either by autoclaving or by filtering through bacteria proof membrane filters (Millex-Gv 0.22 µ,). The ethanol precipitation of DNA was done in an ethanol-dry ice bath for 5 - 10 minutes, at -80° for one hour or at -20° for overnight.

#### D. Isolation and Purification of Plasmid DNA

The plasmid DNA was isolated in large scale by the method developed by Summerton et al. (132).

Small scale plasmid isolations were done by several different procedures, as described below.

### 1. Large Scale Plasmid Isolation

One liter of medium containing ampicillin (50 µg/ml) was inoculated with 1-2% overnight culture of E. coli carrying the plasmid. The cells were grown with aeration to an  $OD_{590}^{1\text{ cm}}$  of 0.1, and uridine to 1 mg/ml was added (133). The culture was further incubated to an absorbance of 0.5 - 0.7 at 590 nm, at which time chloramphenicol to a final concentration of 100 µg/ml, for the amplification of plasmid, was added (134) and the culture was incubated for 15-16 hours with vigorous aeration. The culture was chilled in ice and the cells were harvested by centrifugation. The cell pellet was suspended in 25 ml of resuspension buffer (RB:100 mM Tris-HCl, 100 mM EDTA, 30% sucrose, pH 8.0) and pelleted again. The cell pellet was suspended in 25 milliliters of RB. The degradative enzymes were inhibited by 50 µl of diethylpyrocarbonate, and the cell suspension was kept in an ice bath for 5 minutes. The cells were lysed by incubation with a freshly prepared solution of lysozyme, 1.5 mg/ml, for 20 minutes at ice bath temperature followed by addition of 25 milliliters of prewarmed lysis solution (2.0 M NaCl, 10 mg/ml Brij 58, 4 mg/ml deoxycholate) and incubation at 25° for 20 minutes. Chromosomal DNA and cell debris were removed by centrifugation at 40,000 x g for 30 minutes. The plasmid DNA present in the supernatant was precipitated by adding

four volumes of sodium trichloroacetate (2.25 M, pH 7.0, in 50% ethanol). The mixture was kept in ice for one hour and the plasmid DNA was harvested by centrifugation at 12,000 x g for 20 minutes. The pellet was suspended in 30 milliliters of 50 µg/ml DNase I and RNase A solution and incubated at 37° for 10 minutes. Plasmid DNA was recovered by ethanol precipitation as described earlier.

Further purification of plasmid DNA was done by cesium chloride (CsCl) density gradient centrifugation in the presence of ethidium bromide (EtBr) (135), and gel filtration on Sephadex G-75.

## 2. Cesium Chloride-Ethidium Bromide Density Gradient Centrifugation

Partially purified plasmid DNA solution (1 mg/ml) in TE buffer was mixed with one gram of solid CsCl per milliliter of DNA solution. Ethidium bromide was added to give a final concentration of 600 µg/ml and the refractive index was adjusted to 1.2870. Centrifugation was done at 20° for 48 hours in a Beckman Ti 50 rotor model L2-658 ultracentrifuge. Fractions, 0.25 ml, were collected from the bottom of the tube, and each fraction was assayed for the presence of plasmid DNA by electrophoresis on 1.2% agarose gel. Phage lambda DNA that had been digested with Hind III was used as the size marker. The fractions containing closed circular plasmid DNA

were pooled and ethidium bromide was removed by extraction with n-butanol. The plasmid DNA was recovered by precipitation with ethanol without adding sodium acetate.

### 3. Purification by Gel Filtration

Closed circular plasmid DNA, after purification by CsCl-Ethidium bromide gradient centrifugation is contaminated with oligonucleotides. For efficient labelling of the plasmid DNA in "nick-translation" reactions and also for the end labelling of the restriction fragments for sequence determinations, further purification was necessary. This was done by passing the DNA from the CsCl-EtBr step through a column of Sephadex G-50 or G-75, preequilibrated with TE buffer. About 2 mg of DNA was loaded on a 18.5 x 0.8 cm column. When 0.3 ml fractions were collected, two overlapping peaks which absorb at 260nm were usually observed. Fractions containing plasmid DNA were identified in the first peak by agarose gel electrophoresis. The plasmid DNA was recovered by ethanol precipitation and stored at -20°.

### E. Large Scale Purification of Plasmid Vector pGW7

Plasmid pGW7 was isolated from E. coli RR1 carrying the plasmid after growing the culture in Pen assay broth at 30° and amplifying in the presence of 100 µg/ml of chloramphenicol. The DNA was further purified by CsCl-ethidium bromide gradient centrifugation and stored

at -20° in ethanol.

This vector was initially constructed by G. Wilson of New England Biolabs by inserting a 4.7 kb, EcoR I-BamH I fragment from the immunity region of Lambda DNA in pBR322 which had been restricted with the EcoR I and BamH I. This plasmid, designated as pGW6, carried the lambda cro,  $ci^{857}$ , the temperature sensitive  $ci$  repressor gene, rex and N genes in addition to the strong  $P_L$  promoter. It was further engineered by deleting a Bgl II fragment carrying the cro gene, and designated as pGW7. It contains a 4.41 kb lambda DNA fragment (Figure 14).

#### F. Large Scale Purification of Recombinant pSP19 DNA

The DNA of pSP19 was isolated from 2 l of amplified culture of E. coli carrying the plasmid by the procedure of Summerton as described earlier. The partially purified DNA was further purified by two CsCl-EtBr density gradient centrifugations. About 8.5 mg of DNA was obtained and was stored in small volumes in ethanol at -20°.

Before end labelling, the plasmid DNA was further purified by gel filtration on a Sephadex G-75 column, pre-equilibrated with TE buffer.

#### G. Small Scale Isolation of Plasmid DNA

Plasmids in small scale were isolated, mainly for quick scanning of the recombinant plasmids for size by

any of the following three methods.

1. Scaled down version of the large scale procedure described. This procedure, as expected, allows isolation of partially purified plasmids.

2. Boiling Method (136). One and a half milliliters of an overnight culture grown in the presence of the antibiotic ampicillin (AP) at 50 µg/ml were pelleted in an Eppendorf tube by centrifuging in a Microfuge for one minute. The tube was drained and the pellet was suspended in 0.35 ml of lysis buffer (8% sucrose, 0.5% Triton x-100, 50 mM EDTA, 20 mM Tris-HCl, pH 8.0). A large colony obtained by spotting about 2 µl of bacterial suspension on a plate containing the antibiotic was also sometimes used. Twenty-five microliters of freshly prepared solution of lysozyme (10 mg/ml) in 10mM Tris-HCl pH 8.0 were added and mixed by vortexing. The cells were lysed by heating in boiling water for 40-60 seconds. The chromosomal DNA was pelleted by centrifugation in an Eppendorf centrifuge for 10 minutes at room temperature, and was removed with a toothpick. Plasmid DNA was precipitated by adding 40 µl of 2.5 M NaOAc and 420 µl of isopropanol. The tube was kept in a dry ice/ethanol bath for 15 minutes and the DNA was recovered by centrifugation. The pellet was treated with RNase as described earlier and the DNA was recovered by alcohol precipitation.

3. Alkaline Lysis Method. This method is similar to the boiling method except that the cells are lysed by alkali. The cell pellet was suspended in 100  $\mu$ l of a buffer containing 50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0 and 4 mg/ml lysozyme which was added just before use. The suspension was kept at room temperature for 5 minutes and 200  $\mu$ l of lysis solution containing 1% sodium dodecylsulfate (SDS) in 0.2 N NaOH was added. The contents were mixed gently by inverting the tube. After allowing the lysis to proceed for 5 minutes in an ice bath, the suspension was neutralized with 150  $\mu$ l of ice cold 3 M potassium acetate buffer 4.5. The tube was kept standing in an inverted position for a few seconds and then allowed to stand on ice for 5 minutes. The tube was centrifuged and the supernatant containing plasmid DNA was transferred to another tube. The protein was removed by phenol: chloroform (1:1) extraction and the residual phenol was extracted 2 - 3 times with chloroform. The plasmid DNA was precipitated with ethanol and RNA was removed by RNAase treatment as described earlier.

#### H. Digestion of DNA with Restriction Enzymes

The DNA solutions containing 1-10  $\mu$ g DNA per 50  $\mu$ l of digestion buffer were digested at 37° for one to two hours. Each digestion buffer was as specified by the manufacturer. Two units of restriction enzyme were used

per microgram of DNA. After adding half of the total enzyme and incubating for 30-60 minutes, the remaining enzyme was added and the incubation was continued for another 30-60 minutes. During restriction digestion the concentration of glycerol was kept below 4 per cent. The reaction was stopped by heating at 65° for 10-15 minutes. The progress of the digestion reaction was monitored by agarose gel electrophoresis in the presence of 0.8 µg/ml EtBr.

Agarose gels of 4-5 mm thickness were poured in a flat-bed gel electrophoresis apparatus. The agarose concentration varied from 0.7-1.2 per cent. The same buffer (50 mM Tris-acetate, 20 mM NaOAc.3H<sub>2</sub>O, 2 mM EDTA, 10 mM NaCl, pH 7.8) was used for preparation of the gels and for electrophoresis. The samples were made in the electrophoresis buffer containing 10% Ficoll and 0.01% bromophenol blue. The electrophoresis was done at 150 volts to monitor the progress of digestion and at 44 volts to separate the fragments. During electrophoresis, the buffer was kept circulating between the two chambers, and over-heating of the gel was prevented by running cold tap water under the gel.

#### I. Digestion of DNA Isolated from T4 Deletion

##### Mutants with EcoR V

Restriction endonuclease, EcoR V, does not dis-

tinguish between the modified base of bacteriophage T4, hydroxymethyl cytosine, and the unmodified cytosine. Hence it is the enzyme of choice to restrict HmC DNA deletion mutants del1, del7 and del9 to compare their restriction patterns. DNA from each deletion mutant as well as dC DNA from dec8 was suspended in EcoR V digestion buffer containing 50mM Tris-HCl pH 9.4, 10 mM MgCl<sub>2</sub>, 10 mM DTT and 100 µg/ml BSA at 1 µg/µl. Ten units of EcoR V per microgram of DNA was added and the reaction mixture was incubated at 37° for two hours.

#### J. Isolation of Restriction Fragments

The restriction fragments were separated by electrophoresis on agarose gels containing EtBr. The bands were identified by illumination with 300 nm ultraviolet light. The DNA fragment(s) was isolated by adsorption on DE81 paper.

Slits were made on both sides of the band and pieces of DE81 paper were inserted into each of the slits. As much as 100-150 volts were applied for about 30 minutes until all DNA of the band was transferred to the paper towards the anode. This could be monitored by illuminating the gel with UV light. DE81 paper containing the DNA was placed in a 0.5 ml Eppendorf tube with a small hole made by a 22 gauge needle. This tube was placed in another 1.5 ml Eppendorf tube, and the excess

liquid from the paper was removed by centrifuging at the maximum speed of a bench top centrifuge. DNA from DE81 paper was eluted by equilibrating the paper with 50-100  $\mu$ l volumes of elution buffer (100 mM Tris-HCl, 5 mM EDTA, 1 M NaCl, pH 8,0). The buffer containing DNA was collected by centrifugation as described earlier. The procedure was repeated 2-3 times. The eluants were pooled and EtBr was removed by extraction with n-butanol, until the lower aqueous phase, which contains DNA, was colorless. The DNA solution was deproteinized by extraction with phenol:chloroform as described earlier. The DNA was recovered by precipitation with three volumes of ethanol. The salts were removed by repeating the ethanol precipitation two more times. The DNA was dissolved in 0.3 M sodium acetate after each precipitation. The precipitate was washed once with 70% and then with 95% ethanol. It was dried in air or under vacuum and was suspended in TE buffer to give a concentration of about 1  $\mu$ g DNA/ml.

#### K. Transformation of Cells by Calcium Chloride

##### Procedure

The original procedure developed by Cohen et al. (138) has been slightly modified (139).

Fifty ml of Pen assay broth was inoculated with 1 ml of overnight culture of E. coli RR1. The culture

was incubated at 37° for 1-1.5 hours, until the cell density reached  $5 \times 10^7$  cells/ml. The culture was chilled on ice and the cells were collected by centrifugation at 4,000 x g for 5 minutes. The pellet was suspended in 25 ml of ice cold 10 mM Tris.HCl, pH 8.0, containing 50 mM CaCl<sub>2</sub>. The cell suspension was kept on ice for 15 minutes, and was then centrifuged. The pellet was suspended in 3.3 ml of ice-cold Tris-CaCl<sub>2</sub> solution.

The CaCl<sub>2</sub> treated cell suspension, 0.2 ml, was mixed with the ligation mixture containing 0.5 µg of the vector and the fragment DNA. The suspension was kept on ice for 30 minutes and was then heat-shocked at 42° for two minutes. One milliliter of L broth was added and the culture was incubated without shaking for one hour. The temperature of incubation varied with the vector used, 37° for pBR322 and 32° for pGW7. About 50-100 µl of the culture was spread on Davis plates containing ampicillin (50 µg/ml) for primary screening.

#### L. Molecular Cloning of T4 frd Gene

Two hundred micrograms of T4 dC-DNA isolated from the multiple mutant dec8 were digested with Sma I at 37° in 800 µl Sma I digestion buffer (15 mM Tris-HCl, pH 8.0, 6 mM MgCl<sub>2</sub>, 15 mM KCl) for 60 minutes. About 140 micrograms of the digested DNA were electrophoresed on 0.4% agarose gel for 16-18 hours at 44 volts to sep-

arate the fragments. The 16 kb fragment containing the gene for DHFR (140) was isolated by transferring the DNA to DE81 paper. Approximately one microgram of 16 kb fragment was further digested with Hind III at 37° for one hour in 50  $\mu$ l of digestion buffer (20 mM Tris-HCl, pH 7.4, 7mM MgCl<sub>2</sub>, 60 mM NaCl). The digest was precipitated with ethanol and the pellet was suspended in 10  $\mu$ l of TE buffer. This digest, containing 0.56  $\mu$ g of DNA, was ligated with 0.5  $\mu$ g of Hind III cleaved pBR322, in the presence of 0.1 unit of T4 DNA ligase in 20  $\mu$ l of ligation buffer (66.6 mM Tris-HCl, pH 7.6, 6.6 mM MgCl<sub>2</sub>, 10 mM DTT, 0.4 mM ATP). The ligation was carried out overnight at 12° (141). Eight microliters of this ligation mixture were used to transform E. coli RR1 by the calcium chloride method (142). The cells were spread on Davis plates for primary screening and the plates were incubated overnight at 37°. About 1,000 ampicillin-resistant (AP<sup>R</sup>) colonies were transferred to Davis plates containing AP, 50  $\mu$ g/ml, and tetracycline (Tc), 25  $\mu$ g/ml. Twenty-seven recombinants showing AP<sup>R</sup>Tc<sup>R</sup> phenotypes were transferred to plates containing AP and 0.025 mM TMP. One of the recombinants which grew at TMP concentration 10 times higher than that required to inhibit the growth of E. coli carrying pBR322 was selected and the plasmid was designated as pSP19.

M. Molecular Cloning of 1.1-kb Hind III Fragment Isolated from pSP19 into Hyperexpression Vector pGW7

Recombinant plasmid pSP19 was digested with Hind III and the 1.1 kb Hind III fragment (see Results) was isolated by electrophoresis on 1.2% agarose gel and adsorption on DE81 paper as described earlier. The 3' recessed ends of this fragment were filled by polymerization using *E. coli* polymerase I, large fragment (Klenow fragment). The reaction mixture contained 3 µg 1.1-kb fragment, 0.5 units of Klenow fragment, 250 µM each dGTP, dTTP, dATP and 80 µM ( $\alpha$ -<sup>32</sup>P) dCTP (about 10 µCi) in 25 µl DNA polymerization buffer (66.6 mM Tris-HCl, pH 7.4, 6.6 mM MgCl<sub>2</sub>, 6.6 mM DTT and 50 mM NaCl). After incubating for 20 minutes at room temperature, the reaction was terminated by adding EDTA, pH 7.2, to 40 mM. The volume was made to 200 µl and the sample was deproteinized by phenol-chloroform and the DNA was recovered by ethanol precipitation.

Twenty micrograms of plasmid vector pGW7 was digested with 20 units of Bam HI at 37° in Bam HI digestion buffer (20 mM Tris-HCl, pH 7.0, 100 mM NaCl, 7 mM MgCl<sub>2</sub> and 2 mM 2-mercaptoethanol) for 30 minutes. After heat inactivation at 65° for 10 minutes the linearized DNA was recovered by ethanol precipitation. The 3' recessed ends of 5 µg of the linear pGW7 DNA were filled with Klenow fragment, by the same procedure as described

above.

About 1.24  $\mu\text{g}$  of the vector DNA was blunt end ligated to 1.2  $\mu\text{g}$  of the 1.1-kb Hind III fragment in a total volume of 20  $\mu\text{l}$  of ligation buffer. The ligation mixture was incubated with 1.0 unit of T4 ligase at room temperature for 1 hour (142).

Four microliters of the ligation mixture were used to transform E. coli RRI. The cells after transformation and heat shock at 42° for 2 minutes were incubated at 32° for 1 hour in 1 ml of L-broth. One hundred microliters of the cell suspension was plated on 4 Davis plates containing AP, 50  $\mu\text{g}/\text{ml}$ . The plates were incubated at 32° overnight. The recombinants were selected for AP<sup>R</sup> TMP<sup>R</sup> on AP/TMP plates. The concentration of TMP was 4  $\mu\text{M}$ .

About 50% of the transformants grew under these conditions. Twenty colonies from AP plates were suspended in 50  $\mu\text{l}$  each of Davis medium and 2  $\mu\text{l}$  from each were spotted, in triplicate on AP and AP/TMP plates. One plate from each set was incubated at 32° , 37° and 43°. All except two clones grew on both, AP and AP/TMP plates. Transformants that showed good growth at all temperatures tested are classified as group I and those that did not grow well at 43° as group II. These two transformants from group II were selected as the possible overproducers, because at 37° and 43° the lambda c<sub>1</sub> repressor is partially inactivated and results in the constitutive expression of the gene(s) downstream to the P<sub>L</sub> promoter. This selection was based on the assump-

tion that overproduction of a protein might be detrimental to the cell. One transformant from each group was further analyzed.

#### N. Molecular Cloning of 0.9-kb Mbo II-Hind III Fragment

The nucleotide sequence of 1.1-kb Hind III fragment of T4 predicts the presence of two Mbo II recognition sites (see Results Section) on the 5' flanking region of frd. So the complete digestion of the fragment with Mbo II should result in fragments of 0.887, 0.156, and 0.03 kb, respectively.

Fifteen micrograms of purified 1.1-kb Hind III fragment were digested with 16 units in Mbo II of 75  $\mu$ l of digestion buffer (10 mM Tris-HCl, pH7.9, 10 mM MgCl<sub>2</sub>, 6 mM KCl, 1 mM DTT and 100  $\mu$ g/ml BSA) at 37° for 30 minutes. The incubation was continued for an additional 90 minutes after adding 16 units of the enzyme. The enzyme was inactivated by heat and the DNA was precipitated with ethanol. The large Mbo II-Hind III fragment was separated from the small ones by electrophoresis on 0.75% agarose gel and was purified by the DEAE adsorption technique as described earlier.

The recessed and protruding 3' ends at the Hind III and Mbo II cleavage sites, respectively, were converted into blunt ends by using Klenow fragments. The reaction conditions were similar to those used for polymerization reaction and have been described earlier.

The vector pBR322 used for this cloning was digested with Hind III. The 3' recessed ends were filled in by polymerization reaction using Klenow fragment and the 5' was dephosphorylated by bacterial alkaline phosphatase, as described in the section on nucleotide sequence determination.

Dephosphorylated vector pBR322 DNA and Mbo II-Hind III large fragment, each 0.1  $\mu\text{g}$ , were mixed in 16  $\mu\text{l}$  ligation buffer and incubated with 1 unit of T4 DNA ligase for three hours at room temperature. E. coli RR1 cells were transformed with 12  $\mu\text{l}$  of the ligation mixture containing about 0.15  $\mu\text{g}$  of DNA by the  $\text{CaCl}_2$  treatment procedure (138). The cells were plated on Davis minimal medium containing 50  $\mu\text{g}/\text{ml}$  AP and incubated at 37° for 48 hours. Forty colonies showing AP<sup>R</sup> phenotype when further screened also showed TMP<sup>R</sup> and Tc<sup>S</sup> phenotype.

#### O. Growth of Transformed Bacterial Cells for

##### DHFR Assay

E. coli RR1 carrying the plasmids pBR322, pGW7 or their recombinants were grown in Davis medium containing ampicillin at 50  $\mu\text{g}/\text{ml}$ . The cultures were grown to late log phase with aeration. Cultures containing the vector pBR322 were grown at 37° and those which contained pGW7 were cultured at 32°. The cells were har-

vested by centrifugation at 4,000 x g for 5 minutes and washed with 10 mM Tris-HCl, pH 7.6. The cell pellet was stored at -80°.

P. Preparation of Phage-Infected Cultures

A culture of E. coli RR1 was grown at 37°, with aeration to a cell density of  $3 \times 10^8$  cells /ml in SM 9 medium. After adding L-tryptophan to 20 µg/ml, the cells were infected with T4 am BL 292 at a multiplicity of infection of 4-5. The cells were harvested 14 minutes after infection, washed and stored as described earlier.

Q. Culturing Cells Containing pSP104 for

Overproduction of DHFR

Two flasks containing 50 ml of nutrient broth were inoculated with E. coli RR1 containing the recombinant plasmid pSP104. The cells were grown to late log phase at 32° with aeration. One of the cultures was then shifted to 43° and the other flask was incubated at 32°. After incubation for another 3-3.5 hours, the cells were harvested, washed with 10 mM Tris-HCl, pH 7.6, and stored at -80°. Two cultures with the plasmid pSP19 were also grown identically for comparison.

R. Preparation of Crude Extract

Infected or uninfected cells from 50-100 ml of culture were thawed in 3-5 ml of 10 mM Tris-HCl, pH 7.6.

The cell suspension was sonicated at setting 5 of a Branson sonicator, using the macro tip. The cells were disrupted by 3-4 bursts, each of 30 seconds duration. The cell suspension was kept in ice during sonication. The translucent disrupted cell suspension was centrifuged at 20,000 x g for 20 minutes and the supernatant was used for DHFR assays. For some experiments the enzyme was further purified.

#### S. Fractionation of E. coli and T4 DHFRs

Nucleic acids and some proteins were removed from the 20,000 x g supernatant by dropwise addition of 0.3 volumes of 8% freshly prepared solution of streptomycin sulfate. The suspension was kept stirred during streptomycin addition. The precipitate was removed by centrifugation at 10,000 x g for 15 minutes and the supernatant was subjected to ammonium sulfate fractionation (13, 59).

#### T. Dihydrofolate Reductase Assay

Dihydrofolate reductase was assayed essentially as described by Mathews et al. (13). Dihydrofolate concentration was decreased to 0.03 mM to avoid enzyme inhibition which has been observed at higher levels of the substrate. The enzyme activity was assayed at 37° in a Beckman spectrophotometer, model UV 5230, equipped with a temperature control device. The reaction mix-

ture contained 50 mM potassium phosphate buffer, pH 7.0, 10 mM 2-mercaptoethanol, 0.03 mM dihydrofolate, 10 mM Tris-HCl, pH 7.5, and 0.1 mM NADPH, and the enzyme extract. The final volume was 1.0 ml. After thermal equilibration, the reaction was initiated by the addition of dihydrofolate, and the decrease in absorbance at 340 nm was measured. A molar absorptivity change of  $12,000 \text{ M}^{-1} \text{ cm}^{-1}$  at 340 nm was used for the calculations of activity (143). Protein determinations were made by the dye-binding method of Bradford (144), or by the modified Lowry method (145).

#### U. Nucleotide Sequence Determination

##### 1. Dephosphorylation of DNA

The DNA sample, 50-150  $\mu\text{g}$ , and trace amount of marker DNA from the same source which had been labelled with  $^{32}\text{P}$  at the 5' end and contained  $1 \times 10^6$  CPM, were suspended in 250  $\mu\text{l}$  of 0.1 M Tris-HCl, pH 8.0. The mixture was incubated at  $37^\circ$  with 0.1 unit of bacterial alkaline phosphatase. The progress of the reaction was monitored by spotting 1-2  $\mu\text{l}$  on PEI-cellulose thin layer sheet and chromatographing with 0.75 M potassium phosphate, pH 3.5 (146). The position of  $^{32}\text{Pi}$  was ascertained by autoradiography of the developed sheet. In this chromatography system, DNA stays at the origin and orthophosphate moves with the solvent front. Under the

conditions described, complete dephosphorylation was achieved in about 30 minutes. The reaction was quenched by addition of EDTA, pH 7.2, to 5 mM and heating at 50° for 10 minutes. To assure complete denaturation and removal of alkaline phosphatase, the sample was deproteinized by phenol:chloroform extraction.

## 2. 5' End Labelling of Restriction Fragments

Endonuclease digestion of pSP19 with any of the restriction enzymes Hind III, Sal I, or EcoR I produced two fragments. The fragments generated in each digestion were dephosphorylated. For 5' end labelling (147, 146), 20 µg of these fragments were mixed with 20 µl of kinase buffer containing 50 mM Tris-HCl, pH 7.6; 10 mM MgCl<sub>2</sub>, 5mM DTT, 0.1 mM EDTA, 0.1 mM spermidine and about 20 pmoles ( $\gamma$ -<sup>32</sup>P) ATP (specific activity more than 7,000 Ci/m mole). The reaction was initiated with 5 units of T4 polynucleotide kinase and incubated for 30 minutes at 37 . The incorporation of <sup>32</sup>P into the DNA fragments was monitored by thin layer chromatography and autoradiography which has been described earlier.

The reaction was stopped by adding ammonium acetate to 0.625 M. and the DNA was precipitated with two volumes of ethanol. The precipitate was suspended in 0.3 M NaOAc and ethanol precipitation was repeated two more times. This procedure almost completely removed

the unreacted ( $\gamma$ - $^{32}\text{P}$ ) ATP.

3. Isolation of Fragments Labeled with  
 $^{32}\text{P}$  on One End

Two fragments of DNA which resulted from digestion of pSP19 were dephosphorylated and labelled with  $^{32}\text{P}$  at the 5' ends. These fragments were restricted with a second enzyme so that four fragments of different size and each labelled at one end only were obtained.

When Hind III was used as the enzyme for first digestion, the enzyme for second digestion was EcoR I. However, after digestion with Sal I or EcoR I, the second digestion was done with Hind III.

The DNA after second restriction digestion was precipitated with ethanol and the pellet was dissolved in the fragment separation buffer (FSB) containing 10% glycerol, 0.05% xylene cyanol, 0.05% bromophenol blue. The FSB was made up with 50 mM Tris-borate, 1 mM EDTA, pH 8.3. The fragments were separated on 0.5 x 12.5 x 16 cm vertical slab gels polymerized from 7.5% acrylamide at 150-200 volts for 2 - 4 hours. The electrophoresis was terminated when the slower-moving tracking dye, xylene cyanol, reached the bottom of the gel. The separated fragments were electroblotted onto DEAE paper (148) by using a Bio-Rad Trans-Blot Cell and the procedure described by the vendor.

The electroblotting was done at 0.17 A and at less than 20 volts for about 16 hours at room temperature. The DEAE paper was transferred onto a glass plate and the positions of the radioactive bands were located by exposing an X-ray film Kodak (X -AR 5). The identified bands of DNA were excised from the DEAE paper and eluted with high salts solution and 5 - 10 micrograms of carrier DNA was added. The DNA was recovered by ethanol precipitation as described earlier.

#### 4. Determination of the Nucleotide Sequence

The DNA sequence was determined by the chemical modification method of Maxam and Gilbert (146, 149). The fragment under investigation, which had one 5' end labelled, after 95% ethanol wash was dissolved in water. Five microliters of the one end labelled fragment containing 50,000 to 300,000 CPM were transferred to three siliconized Eppendorf tubes. These samples were used for limited cleavage at guanine (G), cytosine (C) and adenine and cytosine (A > C). Ten-microliter volumes were used for limited cleavage at adenine and guanine (A + G) and thymine and cytosine (T + C). The temperatures and the durations of the modification reactions are summarized in Table III.

#### 5. Limited cleavage at Guanines (G)

Two hundred microliters of 50 mM sodium cacody-

Table III. CONDITIONS FOR DNA MODIFICATION  
REACTIONS

Base	Fragment Size			
	Up to 400 base pairs		400 to 800 base pairs	
	Time (min)	Temperature (°C)	Time (min)	Temperature (°C)
G	5	20	10	0
A+G	15	37	10	37
A C	15	75	10	75
T+C	5	20	10	0
C	5	20	10	0

late buffer, pH 8.0 containing 1 mM EDTA were added to the tube for the limited cleavage at guanines. The tube was chilled in ice and 1  $\mu$ l of dimethyl sulfate (DMS) was added. The tube was then incubated either at 20° for 5 minutes or in ice for 5 - 10 minutes.

The reaction was stopped by adding 50  $\mu$ l of 1.5 M NaOAc containing 1.0 M 2-mercaptoethanol and 100  $\mu$ g/ml tRNA. Nucleic acids were precipitated by adding 750  $\mu$ l of ethanol. After 5 - 10 minutes in a dry ice-ethanol bath, the nucleic acids were collected by centrifugation for 10 minutes in an Eppendorf centrifuge at 4°. The pellet was dissolved in 0.3 M NaOAc and reprecipitated with 750  $\mu$ l of 95% ethanol. The pellet was rinsed with 70% followed by 95% ice cold ethanol and dried in air or under vacuum.

#### 6. Limited Cleavage at Cytosines (C)

The labelled DNA fragment was mixed with 15  $\mu$ l of 5M NaCl. The reaction was started by adding 30  $\mu$ l of 95% hydrazine and incubated for 5 - 10 minutes (Table III). The reaction was stopped by adding 200  $\mu$ l of 0.3 M sodium acetate containing 0.1 mM EDTA and 25  $\mu$ g/ml RNA. The DNA was precipitated as described under the reaction for guanines.

### 7. Limited Cleavage at Adenines and Cytosines

(A > C)

The DNA solution was mixed with 100  $\mu$ l of 1.2 N NaOH and incubated at 75° for 10 - 15 minutes. The reaction was terminated by adding 150  $\mu$ l of 1 N acetic acid and the DNA was precipitated with 750  $\mu$ l of 95% ethanol after adding 5  $\mu$ g of tRNA. The pellet was washed with 1 ml 70% ethanol and rinsed with 1 ml 95% ethanol.

### 8. Limited Cleavage at Adenines and

Guanines (A + G)

For cleavage at A + G, the original method of Maxam and Gilbert (146, 149) was modified. The labelled DNA solution was diluted to 20  $\mu$ l with water, and the reaction was started with 3  $\mu$ l 10% formic acid instead of piperidine formate. The temperature of incubation was raised to 37° (150) and incubation was for 10 - 15 minutes. The reaction was terminated with 5  $\mu$ l 1.2 N NaOH. The DNA was precipitated with three volumes of 95% ethanol after adding 5  $\mu$ g of RNA and 217  $\mu$ l of sodium acetate solution. The pellet was washed with alcohol as described under A > C.

### 9. Limited Cleavage at Thymines and Cytosines

(T + C)

The labelled DNA was diluted with 10  $\mu$ l water and the reaction was started with 30  $\mu$ l 95% hydrazine and

the mixture incubated for 5 - 10 minutes. The reaction was stopped by adding 200  $\mu$ l of the same quenching solution used in limited cleavage at cytosines.

10. Cleavage of Chemically Modified DNA Fragments with Piperidine

The pellets after chemical modifications were dried in air or under vacuum. Each pellet was suspended in freshly prepared 0.1 M piperidine (149, 146). The lids were tightly closed after placing conformable tapes on the lip of the tubes. The tubes were heated at 90° for 30 minutes in a water bath. The lids were kept secure during heating by placing a lead brick on top. The tapes were removed and the tubes were centrifuged for a few seconds. Piperidine was removed by lyophilization. The residue was suspended in 10  $\mu$ l water and lyophilized again. This procedure was repeated once again.

An alternative procedure developed by Smith and Calvo (150) was also occasionally used. In this procedure the quantity of piperidine was decreased to 50  $\mu$ l. The reaction tubes were heated at 90° for one minute uncapped and for 30 minutes more after closing the caps. The tubes were centrifuged for a few seconds and the contents were transferred to another tube. The DNA was precipitated by adding 50  $\mu$ l of 0.3 M sodium acetate and 0.4 ml of ethanol. The precipitated DNA was pelleted by centrifugation and rinsed with ethanol as

described above.

#### V. Sequencing Gels

The cleavage products after base-specific reactions were separated on thin (0.3-0.4mm) urea polyacrylamide gels (157). Gels of two sizes were used. The 40 x 40 cm gels were run at 1,000 volts whereas the longer 40 x 80 cm gels were run at 1,800 to 2,800 volts (150). The composition of reagents to prepare gels at different concentrations of acrylamide for nucleotide sequence determination is given in Table IV.

Urea was dissolved in water containing acrylamide and bis acrylamide by warming to 37° with occasional mixing. The solution was treated with 5 g of mixed bed resin, Amberlite MB1, and stirred slowly for 30 minutes. The solution was filtered through a Whatman #3 filter paper and then through a Whatman #1 filter. After adding the 10 X buffer, the solution was degassed under vacuum. Freshly prepared solution of ammonium persulfate and TEMED were added and mixed by swirling the flask. After pouring, the gel was allowed to polymerize for at least one hour at room temperature and pre-electrophoresed overnight before loading the samples. The same buffer (0.1 M Tris-borate, 2 mM EDTA, pH 8.3) was used for preparing and running the gels. The conditions of electrophoresis varied with the number of bases

Table IV. PROTOCOL FOR PREPARATION OF  
UREA-POLYACRYLAMIDE GELS

REAGENT	PER CENT ACRYLAMIDE			
	6	8	12	20
Urea	90 g	90 g	90 g	90 g
Water	69 ml	60 ml	42 ml	6 ml
Acrylamide 38%				
Bis Acrylamide 2% Soln.	27 ml	36 ml	54 ml	90 ml
Buffer <sup>1</sup> 10X :	18 ml	18 ml	18 ml	18 ml
Ammonium Persulfate <sup>2</sup> (10%)	1.0 ml	0.72ml	0.72 ml	0.9 ml
TEMED <sup>3</sup>	100 ul	72 ul	72 ul	50 ul

<sup>1</sup> Tris base, 121 g; boric acid, 61.83 g; EDTA.2Na.2H<sub>2</sub>O  
7.44 g; H<sub>2</sub>O to 1l, pH 8.3.

<sup>2</sup> Freshly prepared.

<sup>3</sup> N,N'-Methylenebisacrylamide

sequenced and the concentration of gel used. These conditions are summarized in Table V. The wet gels were covered with a plastic sheet and autoradiographed at  $-20^{\circ}$  for 1 - 8 days, but better results were obtained by autoradiography at  $-80^{\circ}$  of gels that had been dried under vacuum. A Bio-Rad Slab Gel Dryer was used for drying the gels.

#### W. Nick Translation

DNA for hybridization probe was nick translated by the method of Rigby et al. (152, 139). DNA, 1  $\mu$ g, was dissolved in 50  $\mu$ l of nick translation buffer (50 mM Tris-HCl, pH 7.4, 5 mM  $MgCl_2$ , 5 mM DTT and 50  $\mu$ g/ml bovine serum albumin). About 12 pmoles (50  $\mu$ Ci) of ( $\alpha$ - $^{32}P$ ) dCTP was added. The concentration of dCTP was brought up to 2  $\mu$ M with cold dCTP. All other deoxyribonucleoside triphosphates were added to a final concentration of 20  $\mu$ M. DNase I, 100 ng, was added to the ice-cold reaction mixture and the reaction was started by the addition of 5 units of E. coli DNA polymerase I. The incubation period was 60 minutes at  $16^{\circ}$ . The reaction was stopped by adding EDTA to 20  $\mu$ M and the  $^{32}P$  labelled DNA was separated by gel filtration on a 0.5 cm x 20 cm column of Sephadex G-50 which had been pre-equilibrated with TE buffer. Fractions of 0.2 ml were collected and those eluting in the first radioactive peak

Table V. ELECTROPHORESIS CONDITIONS FOR  
SEQUENCING GELS

Acrylamide concentra- tion (per cent)	Number of loadings	Electro- phoresis conditions	Dye** Migration	Number of nucleo- tides
<u>40x40x0.4 cm Gel</u>				
20	1	1,000V/ 4h	BP-2/5 XC-1/5	3-40
8	2	1,000V/ 4h	BP-7/8	20-110
<u>40x80x0.4 cm Gel</u>				
12	2	3,000V/ 4h 2,000V/ 9h	BP-2/3 BP-7/10	5-84 20-120
8	1	1,800V/10h	BP-14/15	24-200
6	3	2,800V/ 7h 2,000V/12h 2,200V/14h	XC-7/8 XC-7/8 XC-off gel	50-240 70-258 120-298

\*\*BP = bromphenol blue; XC = xylene cyanol. Fractions indicate the migration distance relative to the length of the gel.

were pooled and precipitated in the presence of 10  $\mu$ g carrier DNA.

#### X. Hybridization

T4 mutants del1, del7, and del9 have deletions in the 62-32 region. The DNAs isolated from these mutants were probed for sequence complementarity to the cloned 1.1 kb fragment from T4 dec8 by dot blot hybridization as described by Thomas (153).  $\lambda$ C DNA from T4 dec8 was used as the positive control lacking a deletion in the 63-32 region.

Five micrograms of DNA from each of these mutants and also from T4 dec8 were dissolved in 30  $\mu$ l of 0.33N NaOH and were denatured in boiling water for 10 minutes and chilled in ice-ethanol bath. The alkali was neutralized with 20  $\mu$ l of buffer containing 6 x SSC, 0.5 M Tris-HCl, pH 8.0, and 1 M HCl (154). Twelve microliters of each neutralized solution were applied to nitrocellulose filters in triplicate. The nitrocellulose filters were presoaked in 20 x SSC (3 M NaCl, 0.3 M sodium citrate) and dried under a heat lamp. The spot sizes were kept as small as possible by intermittent drying during the application. The dried filters were washed two times in 6 x SSC and were baked for 2 hours at 80° under vacuum.

The hybridization protocol and procedure have been

described by Wahl et al. (155). The baked filters were soaked in 6 x SSC and were transferred to sturdy, heat-sealable plastic bags. The filters were prehybridized in sealed bags at 65° overnight in 0.2 ml prehybridization buffer per cm<sup>2</sup> of the filter containing 6 x SSC, 0.5% sodium dodecylsulfate (SDS), 1 µg/ml BSA, 100 µg/ml salmon sperm DNA and 5 times Denhardt's solution (0.1% Ficoll 400, 0.1% polyvinyl pyrrolidone and 0.1% BSA pentax fraction V). The prehybridization buffer was removed and replaced by 50 µl hybridization buffer (prehybridization buffer containing 0.01 M EDTA) per cm<sup>2</sup> of the filter to which the denatured, labelled probe had been added. The hybridization was done at 65° for 24 hours in a sealed bag. The filters were then washed two times with 2 x SSC containing 0.5% SDS, 5 minutes each time at room temperature, and then with 2 x SSC containing 0.1% SDS for 15 minutes. The final wash was with 0.1 x SSC containing 0.5% SDS for two hours with shaking at 55°. The filters were dried at room temperature, wrapped with thin plastic and were autoradiographed. The amount of label retained on each spot was quantitated by cutting the spot and counting in a scintillation spectrophotometer with a Triton x-100 scintillation cocktail.

#### IV. RESULTS

##### A. Molecular Cloning of frd Gene

Digestion of bacteriophage T4 dC DNA with Sma I produces five fragments of 140.72, 15.58, 3.43, 3.32, and 2.45 kb respectively (140). The 15.58-kb fragment (16-kb) can be separated by electrophoresis on 0.4% agarose gel (Figure 7). This fragment contains the gene for DHFR (91). Digestion of this fragment with Hind III resulted in six fragments of 5.18, 4.42, 3.5, 1.1, 0.76 and 0.45 kb size respectively (91, 156). Restriction mapping of the T4 genome (91, 92) and cloning of T4 DNA fragments into phage lambda vectors (93) had located the frd gene in the 1.11-kb(1.1-kb) Hind III fragment, about 143 kb clockwise from the rIIA/rIIB cistron divide. This fragment, obtained by digesting the 16-kb Sma I fragment with Hind III, was cloned in E. coli RRI by ligating the Hind III digest with Hind III-digested vector pBR322, and transforming CaCl<sub>2</sub>-treated E. coli RRI with this ligation mixture. Plasmid pBR322 has a single Hind III restriction site in the tetracycline resistance gene (130). Insertion of foreign DNA at this site inactivates the tetracycline resistance gene and renders the plasmid to the Tc<sup>S</sup> phenotype, but the resistance to AP is retained. Thus, out of 1,000 AP<sup>R</sup> transformants, 27 recombinants were selected for AP<sup>R</sup>Tc<sup>S</sup>

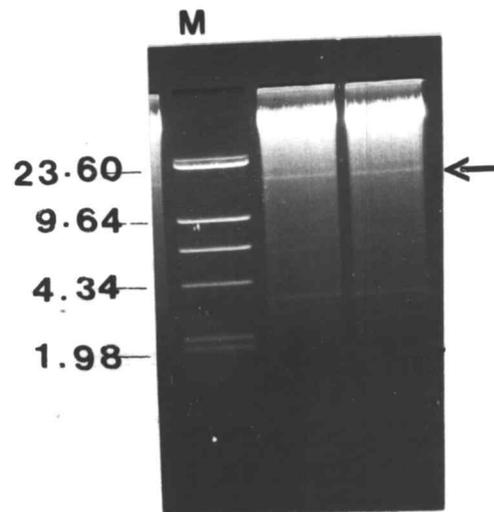


Figure 7. Preparative scale separation of 16-kb fragment of T4 dC-DNA from multiple mutant, dec8 digested by Sma I, on 0.4% agarose gel. The arrow on the right points to 16-kb fragment. Numbers on the left designate the size, in kilobase pairs, of lambda DNA digested by Hind III (lane M).

characteristics. These recombinants were further screened for the presence of the recombinant plasmid containing an active frd gene. For this, advantage was taken of the relative insensitivity of T4 DHFR to trimethoprim. Cells that synthesize T4-specific enzyme should be able to grow at trimethoprim levels which block normal cellular growth of E. coli. One such recombinant strain that grew at 0.025 mM TMP was selected and was designated as pSP19 (Figure 8).

#### B. Digestion of pSP19 with Restriction Enzymes

DNA from E. coli RRI carrying the plasmid pSP19 (RRI/pSP19) was isolated. Digestion of this DNA with Hind III produced two fragments. The large fragment of 4.36-kb corresponds to the plasmid vector pBR322 (157, 130) and the second fragment of 1.1-kb represents the insert (Figure 9). The 1.1-kb fragment containing frd in the T4 genome has a single internal Sal I site and a single EcoR I site (91, 92). Cleavage of pSP19 by these enzymes gives patterns (Figure 10) which are consistent with the existence of these sites (Table VI, Figure 9) and with the insert orientation as depicted in Figure 8.

Although pSP19 was selected by TMP resistance, it is important to establish that the cloned T4 segment contains the entire frd gene. This was confirmed by showing that E. coli cells carrying pSP19 produce an active func-

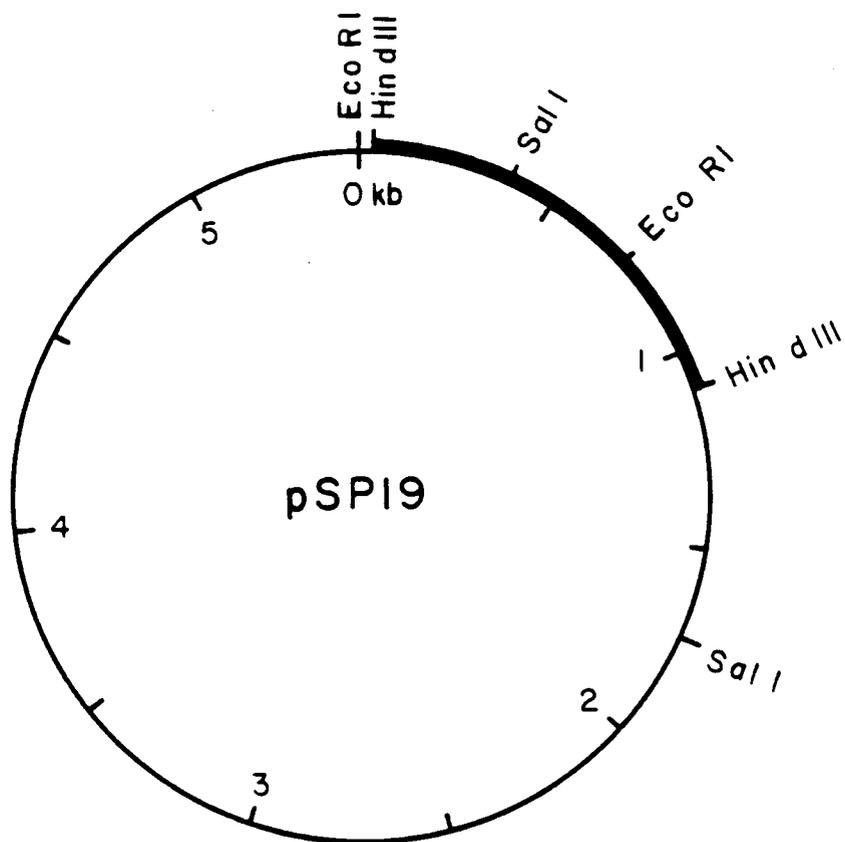


Figure 8. Restriction map of pSP19. Heavy line designates the insert from T4 and light line, pBR322 DNA. Restriction sites are based on the map of pBR322 (130) and the T4 mapping data (116 and Figure 10) and the fragment sizes resulting from restriction cleavage of pSP19 (Figure 9 and Table VI).

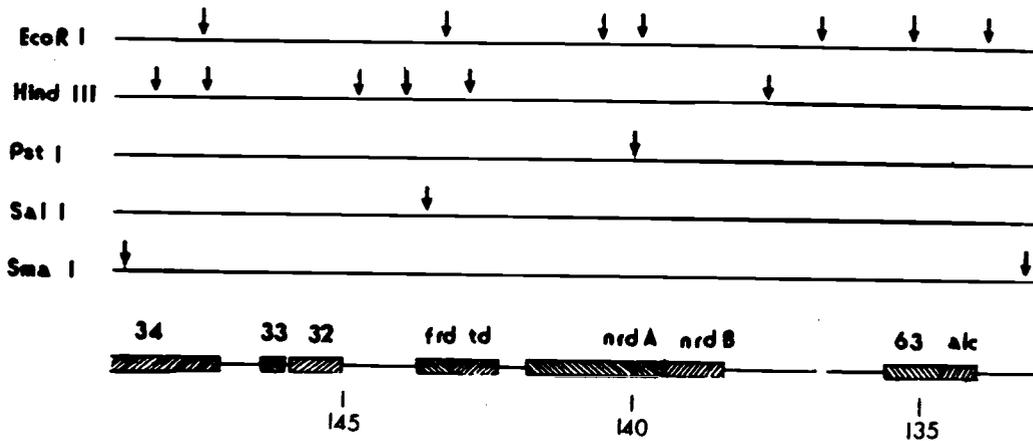


Figure 9. Partial restriction map of T4 16-kb Sma I fragment. All genes except denA and gene 59 have been mapped by restriction analysis or analysis of the cloned fragment (116). Distances in kilobase pairs are from the rIIA/rIIB cistron divide. Vertical arrows represent the restriction sites for enzyme listed on the left.

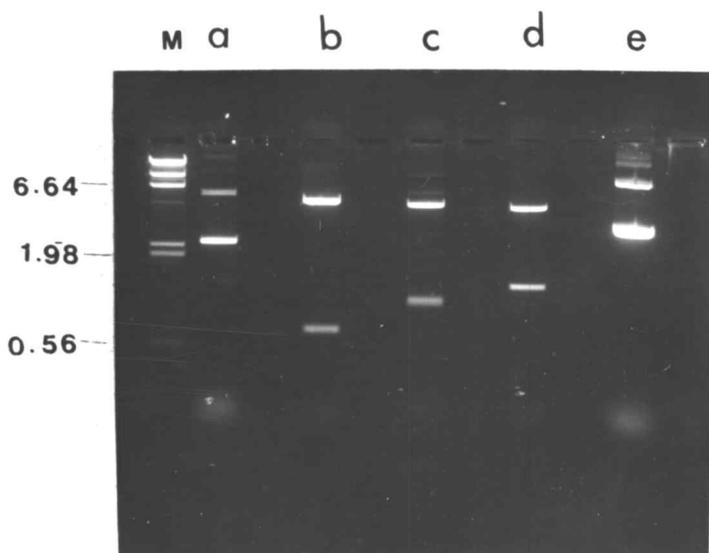


Figure 10. Determinations of fragment sizes produced by digestion of pSP19 by Hind III, EcoR I and Sal I. pSP19 was digested by each enzyme separately and the digests were electrophoresed on a 1.2% agarose gel for 16 - 18 hours at 46 volts in the presence of ethidium bromide. The fragment sizes of marker lambda DNA digested with Hind III are listed on the left. Lane M, Hind III digested lambda DNA; lane a, pBR322 undigested; lane b to d, pSP19 digested with EcoR I, Hind III and Sal I, respectively; lane e, undigested pSP19.

Table VI. FRAGMENT SIZES RESULTING FROM  
RESTRICTION DIGESTION OF pSP19

	DNA Fragment Size kb	
	<u>Expected</u>	<u>Found</u>
Hind III	1.06,4.36	1.10, 4.36
EcoR I	0.73,4.67	0.78,4.69
Sal I	1.35,4.07	1.34,4.15

tional enzyme which is indistinguishable from DHFR produced during T4 infection by several different ways.

C. Heat Stability of Dihydrofolate Reductases from pSP19

Bacteriophage T4 DHFR is more heat labile than the E. coli enzyme (13). Sonicated extracts of RRI/pSP19, RRI/pBR322 and RRI infected with T4 BL292 were incubated at 43°. Aliquots of about 300 µg protein/assay for cells carrying the plasmids and about 50 µg protein/assay for T4-infected cells were transferred at different time periods and DHFR activity was determined. The specific activities are listed in Table VII. Residual activities plotted as per cent of initial activities are shown in Figure 11. The extract of E. coli carrying pSP19 shows intermediate heat lability as compared to the enzymes of T4 and E. coli, suggesting that it contains both E. coli and T4 DHFRs in roughly comparable amounts. This was confirmed by fractionation of these enzymes.

D. Fractionation of Enzymes by Ammonium Sulfate

E. coli and T4-coded DHFRs can be resolved from one another by ammonium sulfate fractionation (13); the phage enzyme salts out at 50% saturation whereas the bacterial enzyme does so at 60-80% saturation with respect to ammonium sulfate. As shown in Figure 12 and Table VIII, cells carrying the plasmid pSP19 contain two types of

Table VII. HEAT STABILITY OF DIHYDROFOLATE  
REDUCTASES AT 43°C

Enzyme Source	SPECIFIC ACTIVITY*					
	0	2	5	12	30	80
	minutes					
RRI/pSP19	21.3	21.4	22.5	20.9	12.7	10.7
RRI/pBR322	16.3	16.8	21.8	21.8	21.6	---
RRI/BL292	131.9	137.4	95.4	28.1	17.2	17.2

\*  
nmoles of dihydrofolate reduced/min/mg protein

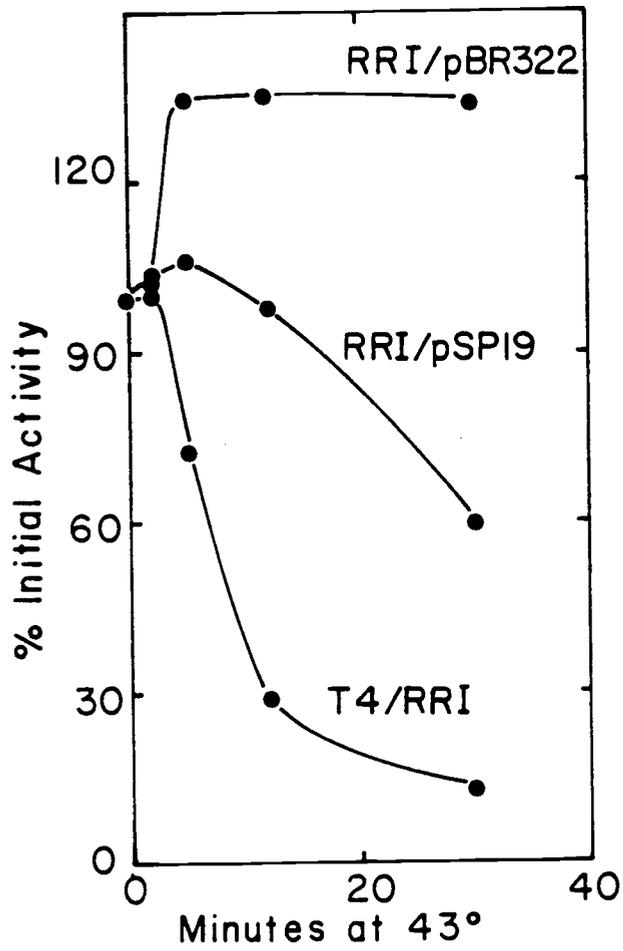


Figure 11. Heat stability of dihydrofolate reductase activity in crude extracts.

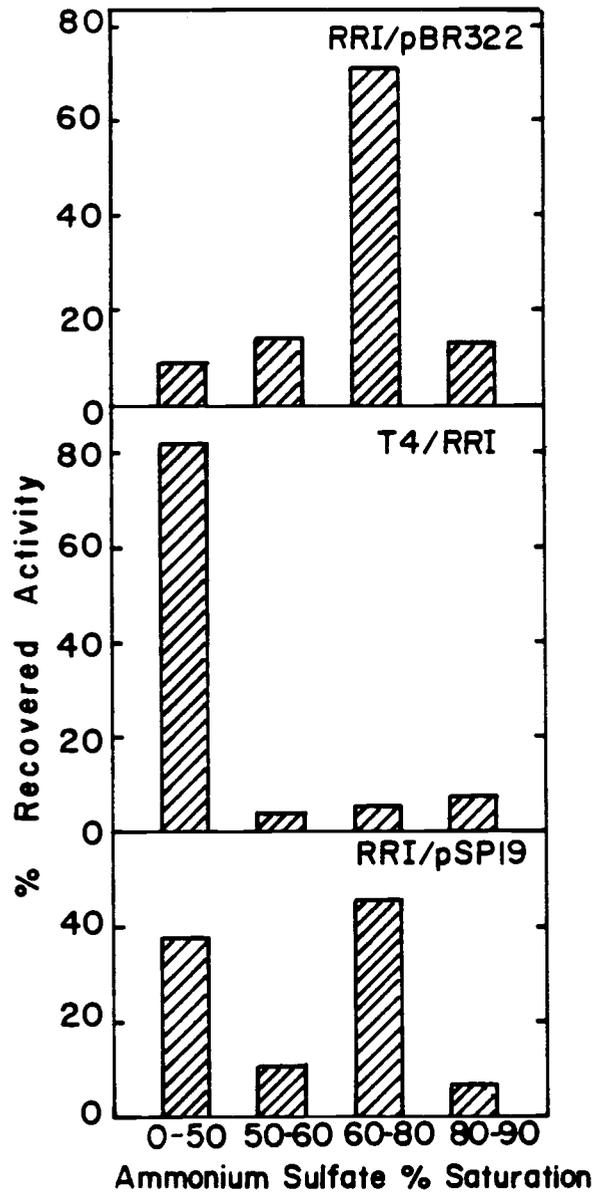


Figure 12. Ammonium sulfate fractionation of dihydrofolate reductase activity in crude extracts.

Table VIII. AMMONIUM SULFATE FRACTIONATION OF  
DIHYDROFOLATE REDUCTASES

Source of Enzyme	Crude Extract	Specific activity*				
		Ammonium sulfate per cent saturation				
		0-40	40-50	50-60	60-80	80-90
RRI/pSP19	14.6	39.9	5.7	11.1	96.7	8.3
RRI/pBR322	7.83	4.84	5.48	17.55	95.4	9.3
RRI/BL292	25.2	80.9	57.5	14.9	29.6	10

\*  
nmoles of dihydrofolate reduced/min/mg protein

DHFR in roughly equal amounts. The enzyme precipitating at low salt corresponds to the T4 enzyme and that precipitating at high salt represents E. coli DHFR.

E. Sensitivity of Dihydrofolate Reductases to Trimethoprim

E. coli and bacteriophage T4 dihydrofolate reductases can be distinguished, even in crude extracts, by their trimethoprim inhibition patterns. E. coli enzyme is more sensitive to trimethoprim than T4 enzyme (14). In Figure 13 and Table IX the trimethoprim sensitivities of 0-40% and 60-80% ammonium sulfate fractions from the extract of RRI/pSP19 and those of corresponding fractions from the extracts of T4-infected cells (1-40% ammonium sulfate fraction) and uninfected cells RRI/pBR322 (60-80% ammonium sulfate fraction) are presented. The 0-40% ammonium sulfate fraction from the RRI/pSP19 extract is indistinguishable from the crude T4 enzyme, whereas the 60-80% fraction is identical to the E. coli enzyme. The bacterial enzyme and the fraction corresponding to it are completely inhibited by 1.0  $\mu$ M trimethoprim, but the T4 enzyme and its plasmid-coded counterpart show insignificant inhibition at this concentration of the inhibitor.

F. Sensitivity of Dihydrofolate Reductases to Antiserum

Earlier results of Mathews et al. and Moser et al.

Figure 13. Trimethoprim sensitivity of dihydrofolate reductase. A, 0-40% ammonium sulfate fraction from T4-infected E. coli RRI; B, 60-80% ammonium sulfate fraction from E. coli RRI/pBR322; C, 0-40% ammonium sulfate fraction from E. coli RRI/pSP29; D, 60-80% ammonium sulfate fraction from E. coli RRI/pSP19.

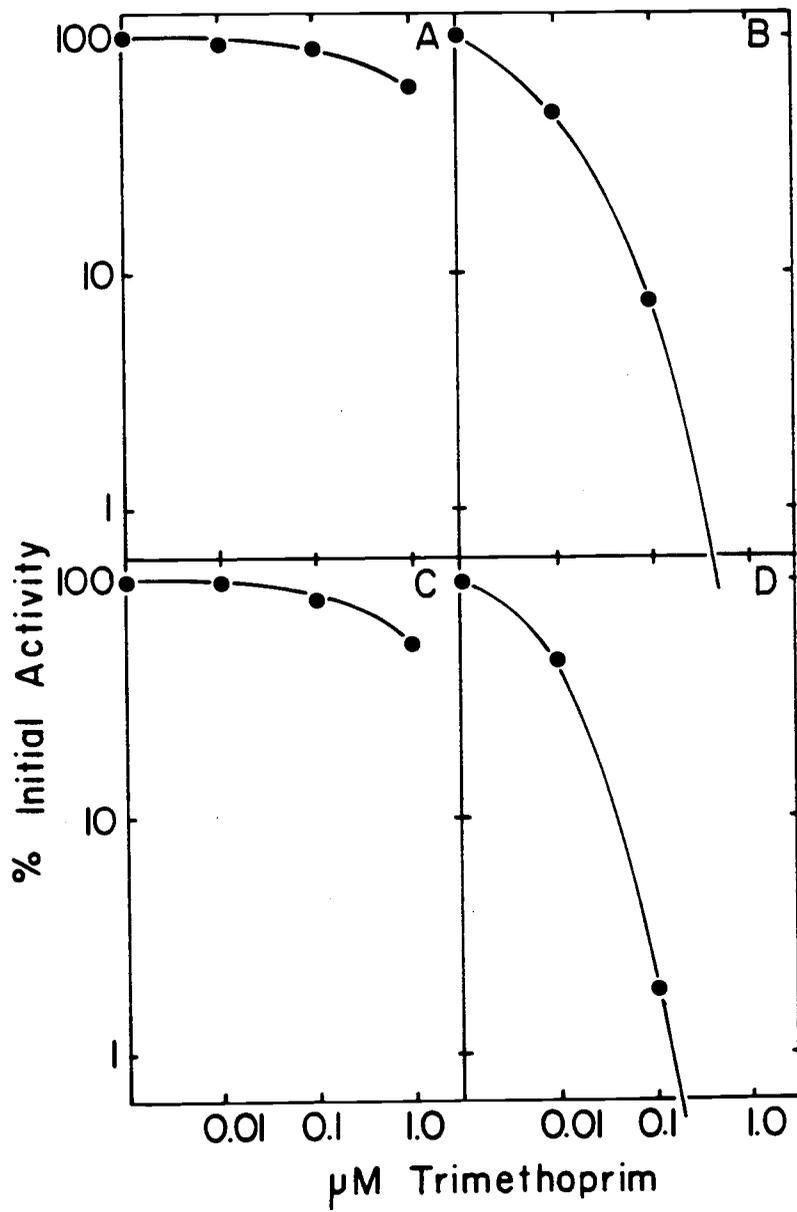


Figure 13. Trimethoprim sensitivity of dihydrofolate reductase.

Table IX. SENSITIVITY OF AMMONIUM SULFATE  
FRACTIONS TO TRIMETHOPRIM

Source of Enzyme		Specific Activity*			
		2	0.01	0.1	1.0
RRI/pSP19	( 0-40 )	15.2	15.2	13.3	7.6
RRI/pSP19	( 60-80 )	39.1	19.6	3.6	1.8
RRI/pBR322	( 60-80 )	51.7	24.7	3.87	0
RRI/BL292	( 0-40 )	30.5	28.8	27.3	19.2

\* nmoles of dihydrofolate reduced/min/mg protein. The numbers in parenthesis designate per cent saturation with respect to ammonium sulfate.

(81, 71) have demonstrated that E. coli dihydrofolate reductase is not inhibited by antiserum raised against homogeneous T4 dihydrofolate reductase. Ammonium sulfate fractions similar to those described in the previous section were incubated at 37° with 1:100 dilution of anti-T4 DHFR immune serum and the enzyme activities were assayed over ten minutes. The per cent inhibition of the activities is shown in Table X. These results confirm that the 0-40% ammonium sulfate fraction corresponds to T4 dihydrofolate reductase.

From these data it is logical to conclude that the enzyme activity in the 0-40% ammonium sulfate fraction of RRI/pSP19 represents bacteriophage T4 dihydrofolate reductase which is encoded by the frd gene borne on the 1.1-kb insert of T4 DNA in pSP19.

#### G. Recloning of 1.1-kb Fragment in the Hyperexpression

##### Vector pGW7

Cells carrying pSP19 produce T4 specific DHFR at a level lower than that of T4 infected cells (59). Therefore, to obtain large quantities of purified enzyme needed for the determination of amino acid sequence of N and C termini, the 1.1-kb fragment was inserted into the hyperexpression vector pGW7. Two groups of AP<sup>R</sup>TMP<sup>R</sup> transformants were selected. Group I transformants were insensitive and group II were sensitive to higher tempera-

Table X. ANTISERUM SENSITIVITY OF  
DIHYDROFOLATE REDUCTASES

Minutes of incubation	Per cent Initial Activity		
	T4	<u>E. coli</u>	pSP19
0	100	100	100
0.2	40	99	38
2.5	24	84	28
10	16	111	26

Antiserum to purified T4 dihydrofolate reductase (18) was incubated at 1:100 dilution with a 0.40% ammonium sulfate fraction from T4-infected bacteria, a 60-80% ammonium sulfate fraction from E. coli RRI carrying pBR322, and a 0-40% ammonium sulfate fraction from E. coli RRI carrying pSP19. Incubations were at 37°, and reactions were started by addition of NADPH.

ture (43°). It is obvious that this difference in these two groups of transformants is the consequence of the orientation of the insert. The nucleotide sequence suggests that the 1.1-kb fragment carries the structural and regulatory regions of frd. Hence the expression of this gene should be independent of the orientation of the insert. However, under conditions that inactivate the lambda repressor, frd will be overexpressed when inserted downstream to P<sub>L</sub> promoter and the increased synthesis of a single protein may adversely affect normal cell division and eventually cell growth. Such is the case with group II transformants. This effect of higher temperature will not be observed if the gene is not controlled by the P<sub>L</sub> promoter, i.e., if the orientation of the insert is reversed, as is possibly the case with group I transformants.

DNAs from pSP104 and pSP38, the representative plasmids from group II and group I, respectively (Figure 14), were restricted with Sal I. The sizes of these fragments are comparable to those predicted with the proposed orientation of the 1.1-kb insert and are shown in Figure 15 and Table XI.

Crude extract of cells carrying pSP104 after temperature shift from 32° to 43° and incubation for 3 - 4 hours showed more than 300-fold increase in specific activity of DHFR as compared to the cells which were

Figure 14. Cloning of the T4 frd gene into a high-expression vector. pGW7 contains a portion of the  $\lambda$ phage genome cloned into part of pBR322. This region includes the mutant c<sub>I</sub>857 gene, encoding a thermolabile repressor, plus the early leftward promoter and gene N. pSP19 contains a 1.1-kb Hind III fragment of T4 DNA cloned into pBR322. This fragment, which includes the T4 frd gene, was cloned in both orientations into the Bam HI site of pGW7, as indicated. Solid lines, pBR322 DNA; wavy lines, phage DNA; heavy lines, T4 phage DNA.

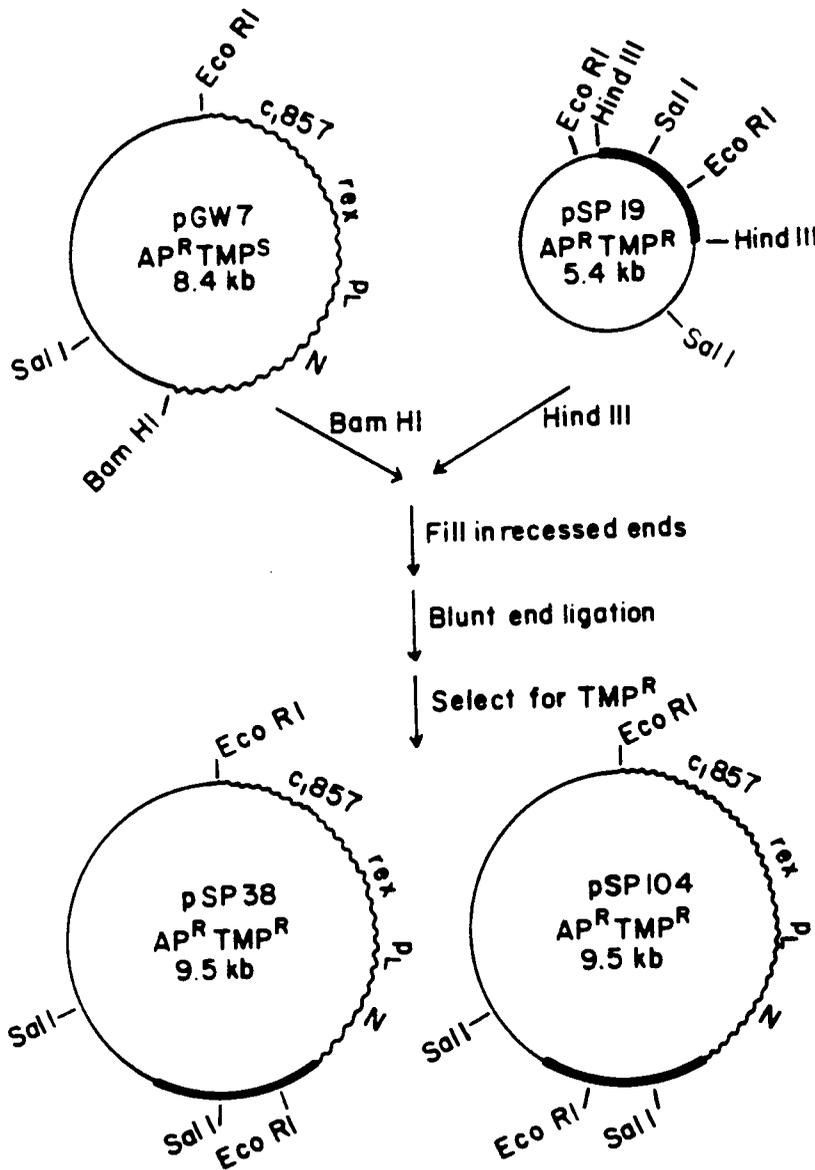


Figure 14. Cloning of the T4 *frd* gene into a high-expression vector.

Figure 15. Separation of Sal I restriction fragments of pGW7-1.1-kb recombinant plasmids on agarose gel. Plasmids were isolated on small scale by boiling method (132) and were phenol extracted. DNA fragments after digestion with Sal I were separated on 0.75% agarose gel for 16-18 hours at 44 volts. Lane M<sub>1</sub>, lambda DNA digested with Hpa I; lane M<sub>2</sub>, lambda DNA digested with Hind III and EcoR I. Fragment sizes of digested lambda DNA are listed on the left. Lanes a, e, f, g represent group I and lanes b and c group II recombinants. Lanes d and h show DNA isolated from transformants that showed growth similar to E. coli RRI on plates containing TMP at 4 μM but are AP<sup>R</sup>; lane i represents pSP19 digested with Sal I.

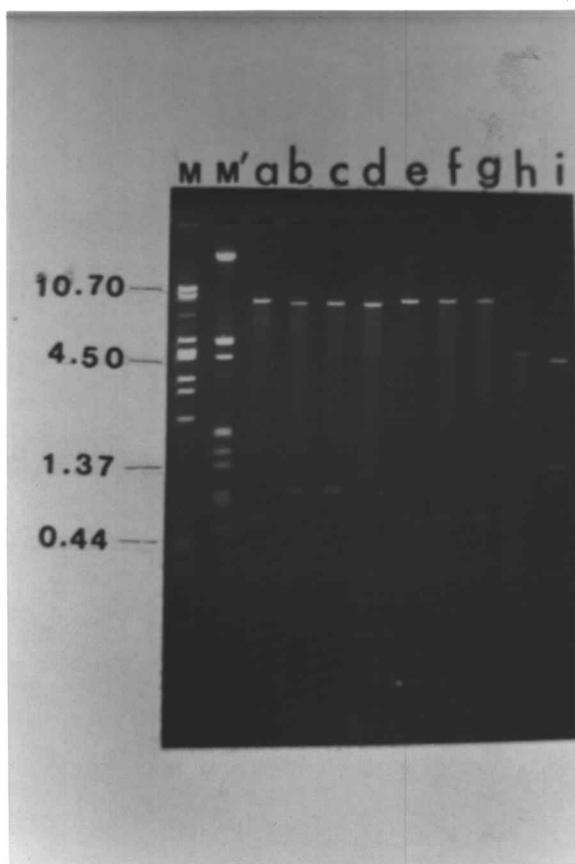


Figure 15. Separation of Sal I restriction fragments of pGW7-1.1-kb recombinant plasmids on agarose gel.

Table XI. FRAGMENT SIZES OF pGW7-1.1-kb RECOMBINANT  
PLASMIDS DIGESTED WITH Sal I

Recombinant plasmid	DNA Fragment size (kb)	
	Expected	Found
PSP19	1.07, 4.36	1.37, 4.1
pSP104 (Group II)	0.98, 8.5	0.94, 8.5
pSP38 (Group I)	0.65, 8.8	0.62, 9.0

Plasmid DNA was digested with Sal I and electrophoresed on 0.75% agarose gel at 44 volts for 16-18 hours. Lambda DNA digested with Hpa I and also EcoR I and Hind III was used as the size marker.

grown at 32° (Table XII). Other strains including pSP38 did not show any significant increase of DHFR activity under similar conditions. These results strongly suggest that the overexpressed gene in plasmid pSP104 is under the control of the lambda P<sub>L</sub> promoter.

That the overproduced DHFR is indistinguishable from T4 enzyme encoded by pSP19 was demonstrated by various methods. For these experiments, strains carrying pSP19, pSP38 and pSP104 and E. coli RRI were grown at 32° until late log phase and were then shifted to 43° for 3 - 4 hours. Crude extracts from these cells were assayed for inhibition by TMP and inactivation of DHFR by T4 DHFR-specific antiserum. At a concentration of 0.1 μM TMP, the enzyme overproduced after temperature shift by pSP104 was inhibited only by fifteen per cent. As expected, at this concentration of the drug, complete inhibition of E. coli DHFR was observed. Also when crude extracts were incubated with the antiserum produced against electrophoretically pure T4 DHFR at 37° for 5 minutes, the overproduced enzyme from pSP104 showed greater than seventy per cent reduction in activity. Under these conditions, E. coli DHFR was not affected.

Results which support that the overproduced enzyme is indistinguishable from the one encoded by pSP19 were also obtained from the subunit molecular weight as determined by SDS polyacrylamide gel electrophoresis, and

Table XII. DIHYDROFOLATE REDUCTASE ACTIVITY  
 IN E. coli RRI CARRYING  
 RECOMBINANT PLASMIDS

Plasmid	Specific Activity* of Cells Grown at		
	32°	37°	32°, shifted to 43°
None	n.d.	7.8	n.d.
pSP19	31.2	n.d.	26.5
pSP38	13.5	24.9	17.3
pSP104	24.7	n.d.	8200

\* nmoles of dihydrofolate reduced/min/mg protein. Cultures were grown at 32° or 37° to late log phase, then shifted as indicated to 43° and incubated for 3 - 4 hours before harvesting for enzyme assay; n.d., not determined.

also by the amino acid composition of the overproduced enzyme.

Crude extracts of pSP 104 (15  $\mu$ g) and pSP19 (40  $\mu$ g) were electrophoresed on 16% acrylamide-0.1% SDS gels. Purified T4 DHFR and extracts from pSP19 and pSP104 grown at 32° were used as the controls. A strong Coomassie brilliant blue R positive band, corresponding to T4 DHFR was observed only with the extract of pSP104 culture that had been shifted to 43° (Figure 16), although DHFR activity could be assayed in all extracts used for electrophoresis. The amino acid composition of DHFR purified from cells carrying pSP104 corresponds to that predicted from the nucleotide sequence of frd gene (Table XIII).

#### H. Nucleotide Sequence Determination of the 1.1-kb

##### Hind III Fragment from T4

The nucleotide sequence of the 1.1-kb Hind III fragment from T4 cloned in E. coli was determined essentially by Maxam and Gilbert's technique (146, 149). The conditions for modification cleavage reactions of A + G and A + C were modified and have been described in the previous section.

##### 1. Strategy of Sequencing

Recombinant plasmid pSP19 was purified by CsCl-EtBr density gradient centrifugation followed by gel filtration through a Sephadex G-75 column. The 1.1-kb

Figure 16. SDS polyacrylamide gel stained by Coomassie brilliant blue R, showing overproduced T4 dihydrofolate reductase. E. coli RRI carrying pSP19 and pSP104 were grown at 32° in nutrient broth and were shifted, in late log phase, to 43° for 3 - 4 hours. Crude extracts from these cultures were electrophoresed on 15% SDS polyacrylamide slab gel (14 x 15 x 0.4 cm) until the tracking dye (bromophenol blue) reached the bottom.

Lane a: extracts from pSP19 cultured at 32°

Lane b: same as a after temperature shift

Lane c: extracts from pSP104 cultured at 32°

Lane d: same as c after temperature shift

Lane f: dihydrofolate reductase purified from T4  
infected gels (from G. Lasser)

Arrow points to the band corresponding to T4 dihydrofolate reductase visible in lanes d and f only.

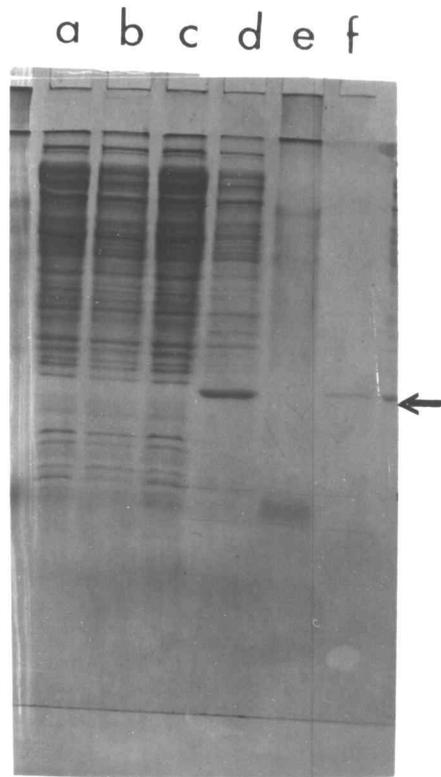


Figure 16. SDS polyacrylamide gel stained by Coomassie brilliant blue R, showing overproduced T4 dihydrofolate reductase.

Table XIII. AMINO ACID COMPOSITION OF  
T4 DIHYDROFOLATE REDUCTASE

Residue	Moles per subunit (21,603 daltons)		
	Predicted from Sequence	Experimental from infected bacteria <sup>a</sup>	Experimental from pSP104
alanine	12	14.4	14.3
valine	18	16.8	16.4
leucine	17	16.0	17.2
isoleucine	12	12.0	11.0
proline	9	9.6	8.9
phenylalanine	8	7.2	7.7
tryptophan	3	3.2 <sup>d</sup>	3.0 <sup>d</sup>
methionine	5	4.0	5.1
glycine	13	16.8	14.1
serine	13	10.4	11.7
threonine	15	12.8	14.4
cysteine	1	1.6	1.4
tyrosine	9	7.2	8.2
aspartate + aspar- agine	18 <sup>b</sup>	18.4	18.9
glutamate + gluta- mine	16 <sup>c</sup>	21.6	18.8
lysine	12	10.4	10.9
arginine	8	8.8	9.8
histidine	4	4.0	4.2

<sup>a</sup>calculated from data in reference 21

<sup>b</sup>13 asp + 5 asn

<sup>c</sup>10 glu + 6 gln

<sup>d</sup>determined spectrophotometrically (165)

Hind III insert has two internal restriction sites, one for Sal I and another for EcoR I. These restriction sites have been mapped (158-161, 89-91). The vector pBR322 also contains one restriction site each for Sal I and EcoR I. A partial linear map of pSP19 showing the relative positions of these restriction sites is presented in Figure 17.

For nucleotide sequence determinations, purified pSP19 DNA was digested with each of these three restriction enzymes in separate experiments. Digestion with any one of these enzymes results in two fragments, each of which has 5' protruding ends. The 5' ends were labelled with ( $\gamma$ - $^{32}\text{P}$ ) ATP and T4 polynucleotide kinase. These labelled fragments were then digested with a second enzyme to obtain fragments labelled only at one end. In each case the second enzyme was so selected that the labelled fragments obtained were of unequal sizes. The resulting one end labelled fragments were resolved by electrophoresis on gels polymerized from 7.5% acrylamide. A representative radioautograph of the resolution of labelled fragments resulting from the sequential digestions of pSP19 by EcoR I and Hind III is shown in Figure 18. The fragments were electroblotted on DEAE paper and eluted with high salt, then precipitated with ethanol, as described earlier. In the digestion described above, the fragments are of 0.36 and 0.78 kb. These purified frag-

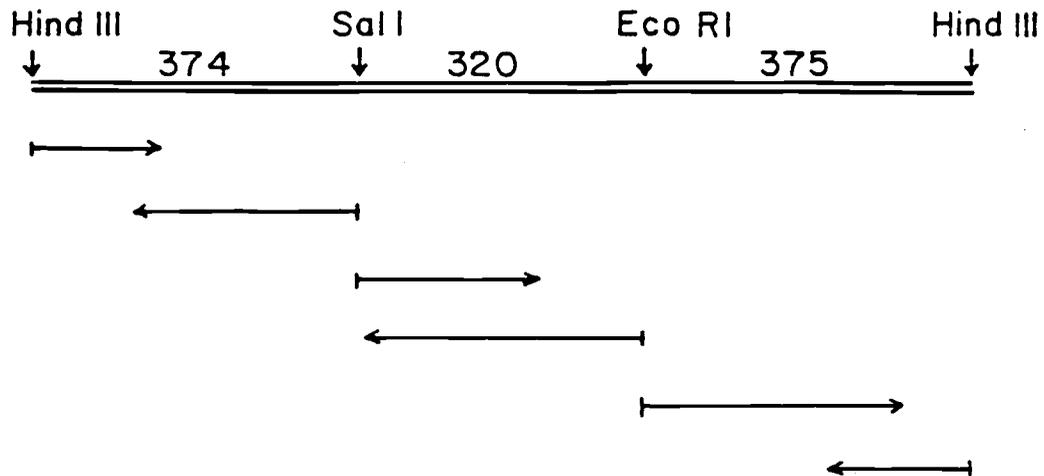


Figure 17. Strategy for sequencing the T4 *frd* gene. A 1.1-kb fragment of the T4 genome, bounded by Hind III sites, was sequenced from the 5' ends of these sites, and of internal Sal I and Eco RI sites, by the Maxam-Gilbert procedure. The length of each segment is given in numbers of nucleotides between cutting sites on the coding strand. Each arrow depicts the direction and approximate number of bases sequenced from each site. For regions where both strands were not sequenced, each determination was carried out at least three times.

Figure 18. Separation of fragments that are labelled with  $^{32}\text{P}$  on one end. pSP19, after digestion with EcoR I was labelled on 5' ends. Labelled fragments were digested with Hind III and separated by electrophoresis on 7.5% polyacrylamide gel (14 x 15 x 0.4 cm) for 2 - 3 hours at 150-200 volts, until the faster dye (bromophenol blue) moved to the end. The gel was then radio-autographed wet for 1 - 2 minutes. Numbers on the side indicate the fragment sizes in kilobase pairs; o, origin.

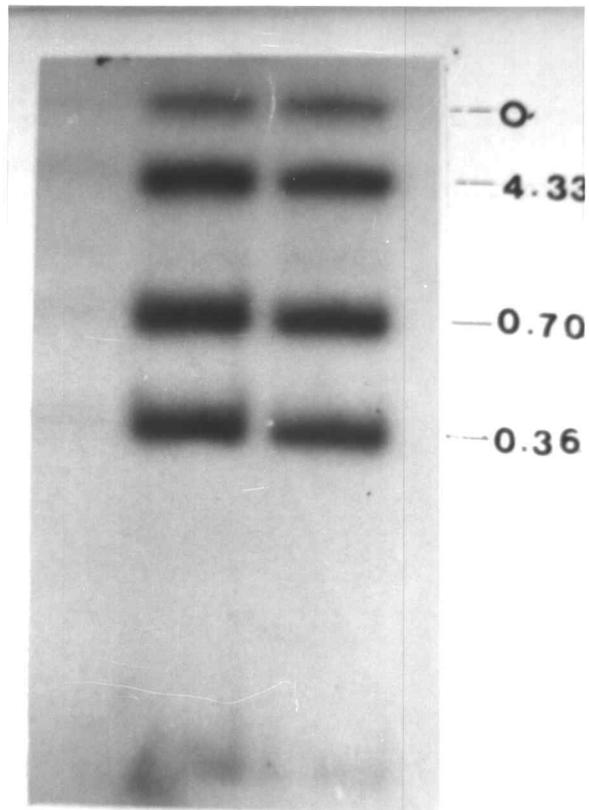


Figure 18. Separation of fragments that are labelled with  $^{32}\text{P}$  on one end.

ments, labelled on one end, were used for chemical modification cleavage reactions (as described in Procedures). The direction and number of bases sequenced from each labelled end is shown in Figure 16, and representative sequencing gels are shown in Figure 19. Although only 30% of the sequence has been determined from both strands, each determination was repeated at least three times.

## 2. Nucleotide Sequence

The nucleotide sequence of the cloned 1.1-kb fragment (127, 128) is presented in Figure 20. Only the anti-sense strand is shown. This sequence contains at least four reading frames.

The first open reading frames (ORF) starting at position 21 and terminating at 261, upstream to frd, could encode a peptide of only 80 amino acids. Dr. Gary Stormo, of the University of Colorado, has estimated the correlation coefficient correspondence between potential codons in the ORF and codons actually used in T4 messages of 0.84 for this frame, which suggests that this region is translated and that it may be an as yet unknown gene. This interpretation is supported by the presence of a potential ribosome-binding site upstream from the ORF (underlined in Figure 20). Both of these observations suggest that this region is probably expressed in vivo on infection. However, SDS polyacrylamide gel electrophoresis of extracts from overproducers showed no extra peptide

Figure 19. Polyacrylamide-urea gel showing part of the sequence of the 360-base pairs Sal I-Hind III fragments. pSP19 was digested with Sal I, labelled on 5' ends with  $^{32}\text{P}$ , and redigested with Hind III. The 360-bp Sal I-Hind III fragment was isolated and used for sequence determination reactions. Panel A shows sequence pattern after electrophoresis on 400 x 400 x 0.4 mM-0,8% polyacrylamide-urea gel at 1800 volts for 1.5 hours; panel B, same as A after 3.5 hours.



Figure 19. Polyacrylamide-urea gel showing part of the sequence of the 360-base pairs Sal I-Hind III fragments.

Figure 20. Nucleotide sequence of the non-coding strand of the Hind III fragment encompassing the frd gene. The 5' end is at the upper left-hand corner. Restriction cutting sites are identified, as are the translational start and stop signals for the open reading frame (ORF) and the leader sequence (L) upstream from the DHFR coding sequence. Sd1 to Sd4 identifies Shine and Dalgarno sequences. Numbers at the right identify distances, in nucleotides, of the rightmost nucleotide on each line from the 5' end of the fragment. For the deduced DHFR amino acid sequence, every tenth residue from the N-terminus is identified. The deduced amino acid sequence for the 24 N-terminal residues is also presented for thymidylate synthase (TS). Horizontal arrows indicate complementary inverted repeats.



band other than the one corresponding to T4 dihydrofolate reductase subunit (Figure 16).

The second reading frame starts at position 913 and extends through the 3' end of the non-coding strand. This frame has been identified as the gene for thymidylate synthase by aligning the translated amino acid sequence with the first twenty N-terminal amino acid residues of TS sequenced by Belfort et al. (162). Also the 380 nucleotides on the 3' side of the EcoRI site are included in a 2.7-kb EcoRI fragment known to include td gene (162, 93).

As reported earlier (59), T4 DHFR is a dimer and the subunit molecular weight is about 23,000. Accordingly, the size of the gene coding for T4 dihydrofolate reductase should be approximately 600 bp. Sequence data show only one open reading frame, the third one, which can code for a peptide of 193 amino acid residues. This reading frame starts at position 335 of the 1.1-kb fragment, and, centered around ten nucleotides upstream to the start codon -ATG-, is the ribosome binding site -AGGA. The first 20 N-terminal amino acid residues predicted by the nucleotide sequence of this frame are identical to the experimentally determined N-terminal sequence of DHFR purified from pSP104 (128). Also, the molecular weight of this encoded peptide of 21,603 is in good agreement with the subunit molecular weight of 23,000 as determined by SDS polyacrylamide gel electro-

phoresis (59). For technical reasons, the C-terminal amino acid sequence has not yet been determined. Determination of the C-terminal amino acid sequence is important to show that translation terminates at the opal codon represented by the TGA at nucleotides 914-916, because Engelberg-Kulka (163) has shown that UGA, particularly when followed by A, need not terminate translation, but that it can be translated as a tryptophan codon. Were that to occur in this case, then translation could occur through six additional codons, with termination at an in-phase ochre site at nucleotides 932-935, with a much greater frd-td overlap than the putative four nucleotides. However, this possibility is unlikely, partly because it would increase the number of tryptophan residues per subunit from 3 to 4. In Table XIII I have compared three amino acid compositions: that predicted from nucleotide sequence with the assumption that translation terminates at 914-916 opal site; that reported earlier (164) for the enzyme purified from T4-infected bacteria; and that determined by C. K. Mathews and T. Bailey for the enzyme hyperexpressed in pSP104. The experimental values are 3.2 and 3.0, respectively, tryptophan residues per DHFR subunit. The general agreement among the three sets of data supports my conclusion that I have identified the structural gene of DHFR and that translation terminates at nucleotides 914-916.

The fourth reading frame between nucleotides 274-322, immediately upstream to frd, is the putative leader region (Figure 20). The Shine and Dalgarno sequence (SD2) for this reading frame overlaps with the opal codon of the upstream ORF and the amber triplet of the leader sequence overlaps with the ribosome binding site (SD2) of DHFR. The leader region can code for a small peptide of 16 residues which is rich in isoleucine. Three of a total of four isoleucine residues are clustered in the N-terminal region of the peptide. Some of the secondary structures are also underlined. This analysis was done by the program "hairpins," reported by Staden (166). Most of these potential secondary structures are AT rich and the segments between 189-345 and 600-734 nucleotides show the highest concentration of these structures. Interestingly, these are also the regions that contain most of the proposed regulatory sequences.

Both 5' and 3' flanking regions of the initiator triplet for DHFR have a series of unusual inverted complementary repeat sequences. In Figure 20, two pairs 1,2 and 3,4 of complementary inverted repeats flanking this region are shown by arrows pointing in opposite directions. Formation of hairpin-loop structures by these complementary inverted repeat pairs results in secondary structures as shown in panel A of Figure 21. By the association of complementary inverted repeats, 3 and 4,

Figure 21. Alternative secondary structures of the segment between 259 to 358 nucleotides from the 5' end of the fragment. SD2 and SD3 are Shine and Dalgarno sequences for the leader (L) and dihydrofolate reductase (DHFR). A, Suggested inactive conformation with SD3 sequestered in the stem formed by the association of the complementary inverted repeat pair, 3,4. Loops (a) and (b) are formed by the association of inverted repeat pairs, 1,2 and 3,4. B, Suggested active conformation with SD3 and the initiator codon, ATG, of DHFR exposed in the loop formed by the association of inverted repeat pair, 2,4. Numbers on either end of the nucleotide sequences identify distances, in nucleotides, of the terminal nucleotide from the 5' end of the fragment.

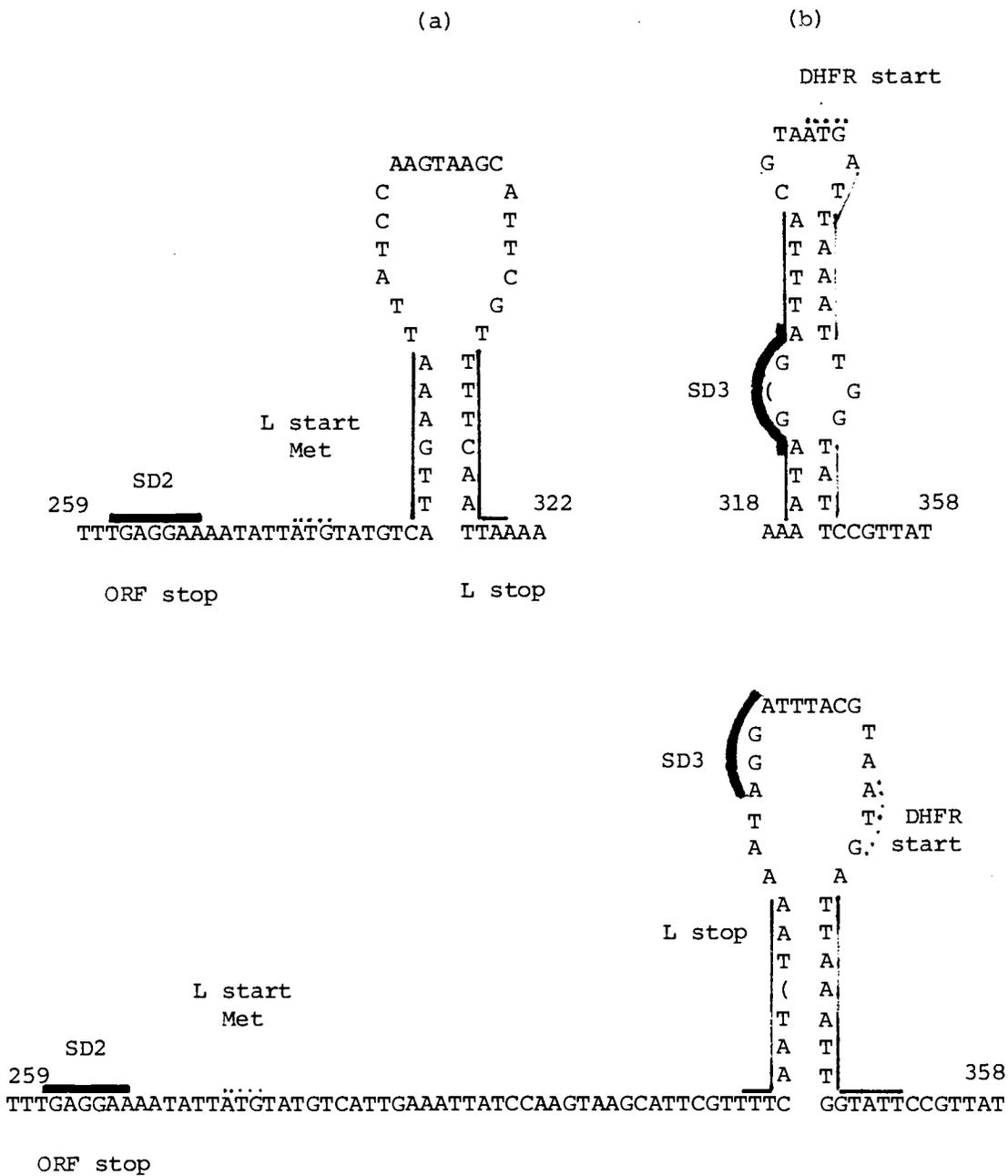


Figure 21. Alternative secondary structures of the segment between 259 to 358 nucleotides from the 5' end of the fragment.

the presumed ribosome binding site (SD3) for DHFR, sequestered in the stem and the initiator triplet, is exposed in the loop. In panel B, a stem-loop structure formed by the association of complementary inverted repeat pair 1,4 is also shown. In this structure both the elements of translation, the initiator codon, and the ribosome binding site, are present in the loop region and are exposed.

There are two restriction sites for the enzyme Mbo II in the 1.1-kb fragment. Both of these sites are present in the ORF 5' to the DHFR structural gene. According to the sequence information, restriction with this enzyme should result in one large (887-bp) and two small fragments of 30 and 156-bp. The large fragment contains the entire coding sequence for T4 DHFR as well as its 3' flanking sequences. It also contains the 5' segment that harbors the presumptive regulatory sequences for frd.

Another enzyme, Hinf I, restricts this cloned 1.1-kb fragment at three positions producing a large, 806-bp long and three small fragments of 167, 48 and 45 bp. The large fragment has most of the 5' flanking sequences but cuts the structural gene of DHFR at its 3' end. These two enzymes, Mbo II and Hinf I were used to test-digest the 1.1-kb fragment and the restriction patterns are consistent with the determined nucleotide sequence. Results of Mbo II digests are shown in Figure 22.

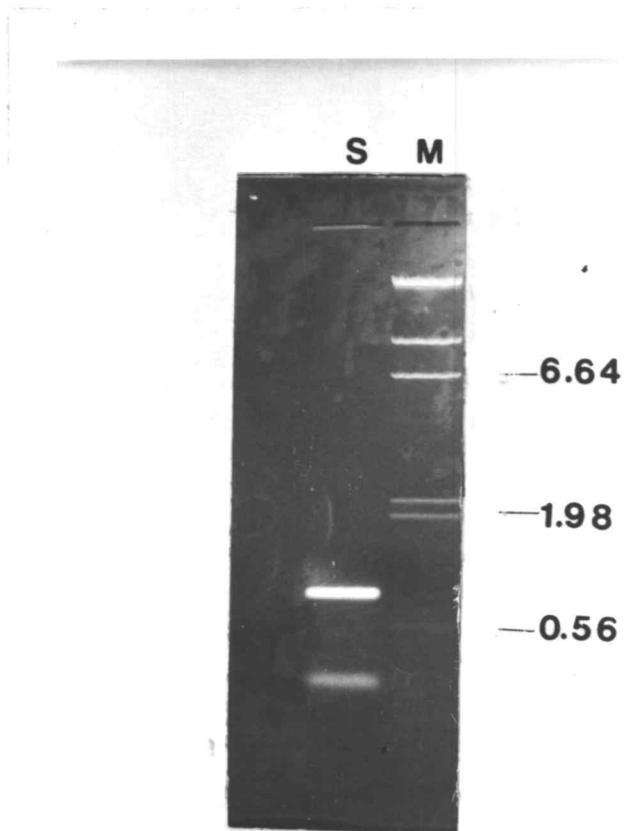


Figure 22. Determinations of sizes of fragments produced by digestion of 1.1-kb fragment by Mbo II and electrophoresis of the digest on 0.75% agarose gel for 16-18 hours at 44 volts. Lane M, lambda DNA digested with Hind III. Numbers on the right show the fragment sizes of the marker. Lane S. represents the fragments 0.88 and 0.16-kb generated by digestion of 1.1-kb fragment by Mbo II.

### 3. Regulatory Sequences

The first seven bases of the sequence, -TAAGGAGAAT-, are identical to the complementary sequence of E. coli B 16S ribosomal RNA, the Shine and Dalgarno sequence (SD1, Figure 20; 105, 107). This sequence, as mentioned above, may serve as the ribosome binding site for the translation of the first ORF. The initiation of translation may be at position 21. Sequence, -AGGA- (SD3), starting at position 323 is the probable candidate for ribosome binding during the translation of frd. Another sequence, -AGGA-, starting at position 263 is the proposed Shine and Dalgarno sequence (SD1) for the leader sequence. This sequence overlaps the opal termination codon of the upstream ORF.

There are two putative promoter sequences ( $P_1$  and  $P_2$ ) upstream of frd.  $P_1$  lies in the ORF region, whereas part of  $P_2$  overlaps with the termination of ORF. The invariant 'T' of the -10 region of E. coli promoter for  $P_1$  and  $P_2$  is at positions 82 and 275 respectively. Either or both of these sequences could serve as the transcriptional regulatory region for frd and possibly for td. The concept of a single transcript for both frd and td is supported by the earlier observations of Trimble and Maley (97). They have reported a 1600 nucleotide long active transcript for T4 DHFR.

Three additional sequences,  $P_3$ ,  $P_4$ ,  $P_5$ , that bear considerable homology to the Pribnow box sequence are present in the structural gene of DHFR. Any or all three of these sequences may be regulatory sequences for the expression of td. The invariant 'T' of -10 region of these sequences is at positions 483, 607, and 876, respectively. Although sequences corresponding to -10 region of these putative promoters are highly homologous to the corresponding region of E. coli consensus sequences, sequences corresponding to -35 region show varying degree of homology. Comparison of the highly conserved sequence, -TTGACA-, in the -35 region of E. coli promoter with the corresponding sequence of these presumptive promoters show that  $P_2$  has the most and  $P_4$  the least homology to -35 region. Interestingly, -35 region of  $P_3$  and  $P_5$  also shows homology to -35 region of the promoters of  $rII_B$  and gene I of bacteriophage T4. Both of these genes show middle mode characteristics (96). The sequences of these consensus sequence of E. coli promoters are listed in Figure 23. The sequence -<sup>895</sup>TACGGA- may function as the ribosome binding site (SD4) for the translation of TS.

#### 4. Relationship to Other DHFR Amino Acid Sequences

I was interested in a possible evolutionary relationship between phage-specific dihydrofolate reductases and those recently characterized as Type I plasmid-specified DHFRs, particularly one encoded for the trans-

Figure 23. List of putative promoter sequences of frd and td genes of T4. The top column shows the consensus sequences of -10 and -35 regions of E. coli promoters (35) and promoter sequences of T4 genes rIIB and gene I (94) are listed in the bottom two rows. P<sub>1</sub>-P<sub>5</sub> represent the presumptive promoter sequence for frd and td. These sequences have been identified by their similarity to the consensus sequences of E. coli promoters. Overlined segments show sequence similarity to -35 region of E. coli promoters and underlined segments show similarity to those of rIIB and gene I middle promoters. Numbers on left and right identify distances in base pairs, from the 5' end of the fragment. \* Identifies the invariant "T" of -10 region.

<u>E. coli</u> consen- sus sequence		-35 region		-10 region
		TTGACA		TATAAT*
P <sub>1</sub>	50	<u>TT</u> CAGAAAGTTAG		TAATAT 82
P <sub>2</sub>	243	ATT <u>GTG</u> AAAAAG		TATTAT 275
P <sub>3</sub>	451	TTT <u>TAAAG</u> CTCG		TATGAT 483
P <sub>4</sub>	575	ACTAAAGACGGC		TATATT 607
P <sub>5</sub>	844	GCGT <u>GAAAT</u> GGTTG		TAAAAT 876
RIIB		ATCAAAT <u>TAATGCTTCA</u>		TAAAAT
gene 1		AGAAGT <u>TTAATGCTTC</u>		TATAAT

Figure 23. List of putative promoter sequences of frd and td genes of T4.

poson Tn 7 (165, 167). Comparison of amino acid sequence of T4 DHFR derived from the nucleotide sequence shows approximately 24% and 23% sequence homology to E. coli and R-plasmid type I enzymes, respectively. The sequences were aligned using the program "align" of the Prophet computer system (168). During this alignment, although insertions are allowed in both sequences, they are controlled and there is a "fair cost" for each insertion. In addition, two consecutive insertions cost much less than the insertions that are not consecutive.

The deduced amino acid sequence of T4 DHFR is aligned alongside the amino acid sequence of five other DHFRs (Figure 24). These sequences include three bacterial enzymes, E. coli, L. casei, and S. faecium, one vertebrate enzyme from chicken liver and one R-plasmid Type I enzyme that is present on the transposon T<sub>n</sub>7 (70). This alignment is based upon the X-ray structural equivalence of three of these enzymes, from E. coli, L. casei and chicken liver, as described by Volz et al. (75). These authors aligned amino acid residues from four animal and three bacterial DHFRs, based upon the occurrence of each residue in a known portion of the three-dimensional structures of the enzymes. Irrespective of the amino acid sequence, overall backbone foldings of bacterial and vertebrate DHFRs are similar (73, 79, 74). Hence this alignment can be made with some degree of certainty. Using the struc-

Figure 24. Sequence alignment of six dihydrofolate reductases according to structural equivalence. T4 DHFR sequence, deduced from nucleotide sequence reported in this paper; Tn 7 DHFR sequence, deduced from nucleotide sequence reported in references 70 and 167; other sequences, from sources cited by Freisheim and Matthews (22). Alignments were made as shown by Freisheim and Matthews. The numbering system is that used for chicken liver DHFR.

```

          5          10          15          20
Chicken Liver V R S L N S I V A V C Q N M - C I C - K D C N - - - - -
E.coli MB1428 - - M I S L I A A L A V D R - V I C - M E N A - - - - -
Type I Tn7 - - M K L S L M V A I S K N C - V I C N G P D - - - - -
T4 - - M I K L V F A Y S P T K T V D C F N E L A F C L C D C
S.faecium - - M F I G M W A Q D K N C - L I C - K D C L - - - - -
L.casei - - T A F L W A Q N R D G - L I C - K D C H - - - - -

          25          30          35          40          45          50
Chicken Liver L P W P P L R N E Y K Y F Q R M T S T S H V E C K Q N A V
E.coli MB1428 M P W - N L P A D L A W F K R N T L - - - - - N K P V
Type I Tn7 I P W - S A K C E Q L L F K A I T Y - - - - - N Q W L
T4 L P W C R V K K D L Q N F K A R T E - - - - - C T I M
S.faecium L P W - R L P N D M P F F R E H T M - - - - - D K I L
L.casei L P W - H L P D D L H Y F R A Q T V - - - - - C K I M

          55          60          65          70          75
Chicken Liver I M G K K T W F S I P E K N R P L K D P I N I V L S - - R E
E.coli MB1428 I M G R H T W E S I G - - - R P L P C R K N I I L S S Q P -
Type I Tn7 L V G R K T F E S M C - - - - A L P N R K Y A V V T R S S F
T4 I M G A K T F Q S L P - - - T L L P C R S H I V V C D L - -
S.faecium V M G R K T Y E C M G - - K L S L P Y R H I I V L T T G K D
L.casei V V G R R T Y E S F P - - K R P L P E R T N V V L T H Q E D

          80          85          90          95          100          105
Chicken Liver L K E A P K C A H Y L S K S L D D A L A L L D S P E L K S K
E.coli MB1428 - - G T D D R - V T W V K S V D E A I A A C C N V P - - - -
Type I Tn7 - - T S D N E N V L I F P S I K D A L T N L K K I T - - - -
T4 - - A R D Y P V T K D C D L A H F Y I T W E Q Y I T - - - -
S.faecium - - F K V E K N A E V L H S I D E L L A Y A K D I P - - - -
L.casei - - Y Q A Q C - A V V V H D V A A V F A Y A K Q H L D - - - -

          110          115          120          125          130
Chicken Liver V D M V W I V G C T - - A V Y K A A M E K P I N H R L
E.coli MB1428 - - E I M V I G C G - - R V Y E Q F L - - P K A Q K L
Type I Tn7 - - D H V I V S G C G E I Y K S L I - - - D Q V D T L
T4 - - - - Y I S C C E I Q V S S P N A - - P F E T M L
S.faecium - E D I Y V S G C S - - R I F Q A L L - - P E T K I I
L.casei - Q E L V I A G C A - - Q I F T A F K - - D D V D T L

          135          140          145          150          155          160
Chicken Liver F V T R I L H E F E S D T F F P E I D Y K D F K L L T E Y P
E.coli MB1428 Y L T H I D A E V E C D T H F P D Y E P D D W E S V F S - -
Type I Tn7 H I S T I D I E P E C D V Y F P E I - P S N F R P V F - - -
T4 D G N S K V S V I G C P A L L Y A A L P Y A D E V V V S - -
S.faecium W R T L I D A E F E C D T F I G E I D F T S F E L V E E H E
L.casei L V T R L A G S F E C D T K M I P L N W D D F T K V S S - -

          165          170          175          180          185          189
Chicken Liver C V P A D I Q E E D C - - - I Q I K F E V Y Q K S V L A G
E.coli MB1428 - - - E F H N A D A Q N S H S Y C F K I L E R R
Type I Tn7 - - - T Q D F A S N - - - I N Y S Y Q I W Q K C
T4 - - - R I V K R H R - - - V N S T V Q L D A S F L D D I
S.faecium C I - - V N Q E N G Y - P H R F Q K W Q K M S Y V V
L.casei - - - R T V E D T N - P A L T H T Y E V W Q K K A

T4 S K R E M V E T H W Y K I D E V T T L T E S V Y K

```

Figure 24. Sequence alignment of six dihydrofolate reductases according to structural equivalence.

ural alignment scheme of Volz et al. (74) for several DHFR sequence alignments, Fling and Richards (70) aligned the amino acid sequence of Type I DHFR to the known sequences of other DHFRs. For the amino acid sequence alignment of T4 DHFR (Figure 24), I have used the homologies in the Fling-Richards scheme. As has been observed for several DHFRs, homologies in the C-terminal region are not substantial and hence the assignments are not very precise. However, from this alignment it can be concluded that the region between nucleotides 335 to 916 is the coding sequence for dihydrofolate reductase, and that the T4 enzyme, like the other known prokaryotic DHFRs, is not particularly closely related to the R-plasmid enzyme or to any other known DHFRs. Some features unique to the T4 enzyme, such as insertion between residues 22 and 27, gap between 98 and 99 and a long C-terminal tail, have no apparent counterparts among the sequences of five other DHFRs.

In Figure 25, the number of identities in ten-residue segments of T4 DHFRs is plotted against the number of identities in the corresponding regions of the other sequences. This analysis was made by C. K. Mathews. The identities are clustered near the N-terminal region and are mainly among those residues that are involved in the substrate and/or inhibitor binding (reviewed in Reference 22). Sixteen of a total of nineteen

Figure 25. Sequence homologies among dihydrofolate reductases. The height of each bar denotes the number of identities within a ten-residue segment between T4 DHFR and each of the indicated enzymes. The vertical ticks on the top line identify residues conserved among the six sequences being compared (either identical or nearly so, e.g., valine and isoleucine). The next two lines of ticks represent residues in L. casei DHFR known to be in contact with NADPH or methotrexate, respectively, as determined crystallographically (22, 73, 74, 75). The line of ticks in the Tn 7 panel depicts identities between Tn 7 and T4 DHFRs, between residues 92 and 118 on the T4 sequence.

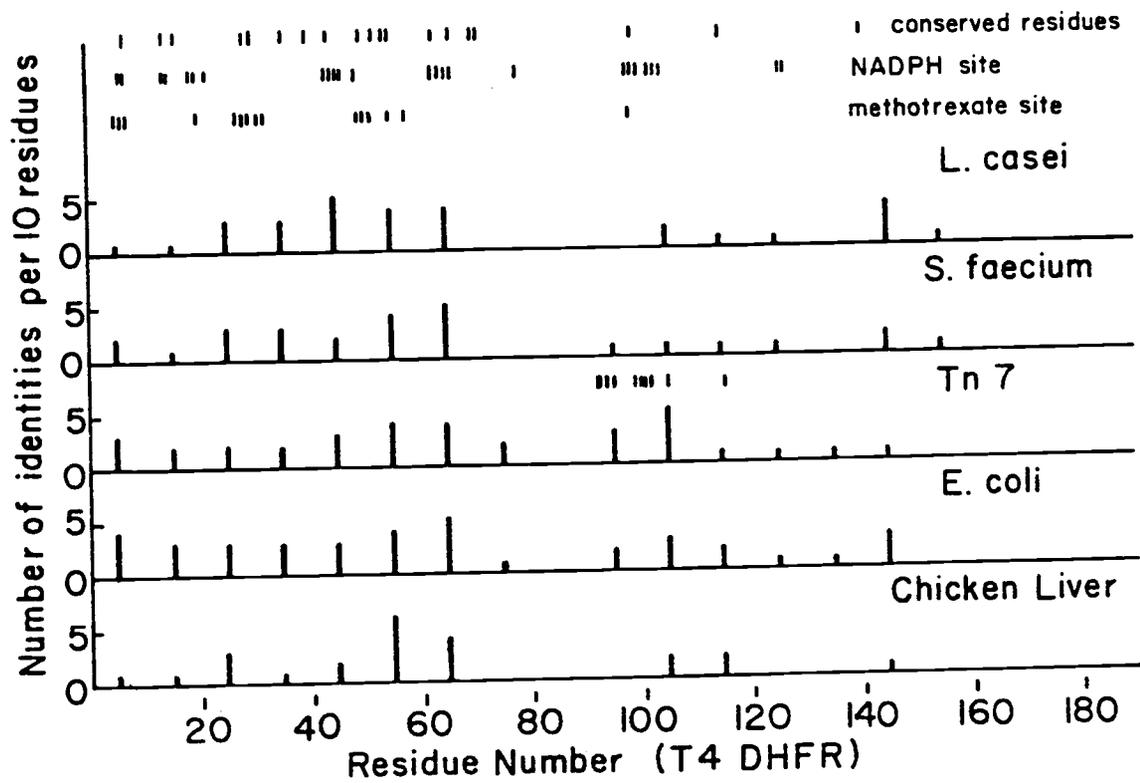


Figure 25. Sequence homologies among dihydrofolate reductases.

residues of T4 DHFR, which are conserved among all six DHFRs compared, are found before the residue 71. This confirms my earlier conclusion that this sequence does indeed encode T4 dihydrofolate reductase. This analysis also suggests that the overall sequence of T4 DHFR is quite different from the rest of the known DHFRs, including the R-plasmid Type I encoded enzyme. The overall homology between the T4 and any of the five compared sequences is not more than 20%. The total number of identities between 193 amino acid residues of T4 enzyme and E. coli, L. casei, S. faecium, chicken liver and Type I enzymes are 38, 30, 29, 23 and 36 respectively.

As mentioned earlier, both T4 and Type I enzymes are distinctive among all bacterial and animal dihydrofolate reductases in being dimers, each containing two identical subunits (59, 167). Comparison of T4 and Type I DHFR sequences shows unusually high homology between residues 92 and 118, and several of these identities are present in these two enzymes only. Is it possible that these residues are involved in the subunit-subunit interactions which lead to dimerization? Crystallographic studies now under way for the T4 enzyme in Brian Matthews laboratory at the University of Oregon might answer this question.

## I. Study of Deletion Mutants in the 63-32 Region

### 1. EcoR V Restriction Analysis of DNA from Deletion Mutants

In Figure 26, band patterns of DNA from various T4 mutants restricted with EcoR V are shown. As expected, the EcoR V restriction pattern of dec8 dC DNA is different from those of HmDC DNA of deletion mutants. However, the restriction patterns of del1, del7 and del9 are alike except one band of approximately 11-kb. This band is present in both del7 and del9 but is absent from del1 lane. This observation is in accord with published results (71, 86, 117), which suggest that del7 and del9 have similar deletions.

### 2. Dot Hybridization of 1.1-kb Fragment to DNA from Deletion Mutants

DNAs from deletion mutants del7 and del9, when hybridized with  $^{32}\text{P}$ -labelled, either purified 1.1-kb fragment or with pSP19, gave positive results (Figure 27). In this experiment  $^{32}\text{P}$ -labelled pBR322 DNA was used as the negative control. It suggests that these deletion mutants contain sequences that are complementary to 1.1-kb fragment.

Figure 26. Comparison of EcoR V restriction patterns of T4 mutants deleted in 63-32 region. DNA from each deletion mutant was digested with EcoR V as described in Procedures, and the fragments were separated by electrophoresis on 0.75% agarose gel for 16 - 18 hours at 44 volts. Lambda DNA digested with Hind III was used as markers (lane e) and the numbers on the side represent the sizes of these fragments. Lane a, del1; lane b, del7; lane c, del9. In lane d, dC DNA from dec8 digested under conditions similar to those used for deletions were used. The arrow points to the 11-kb fragment that is absent from del1 digest.

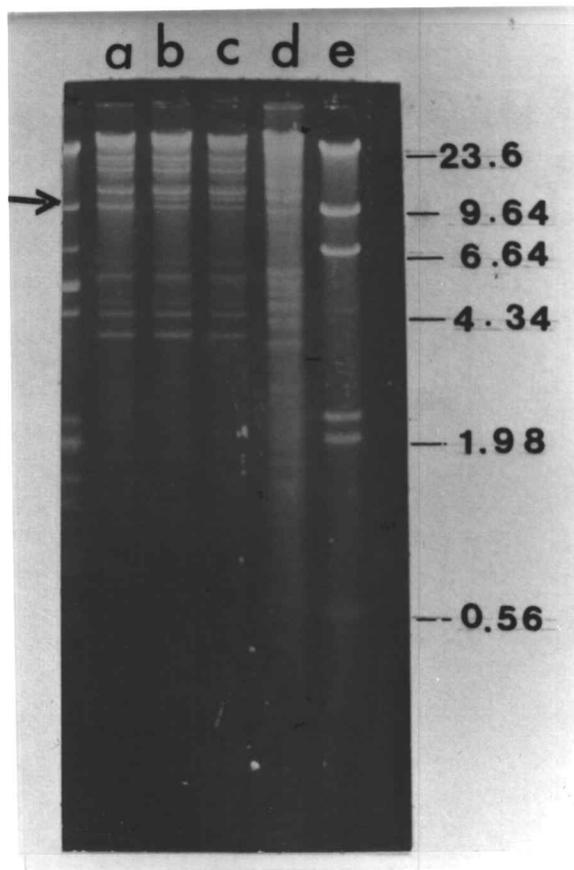


Figure 26. Comparison of EcoR V restriction patterns of T4 mutants deleted in 63-32 region.

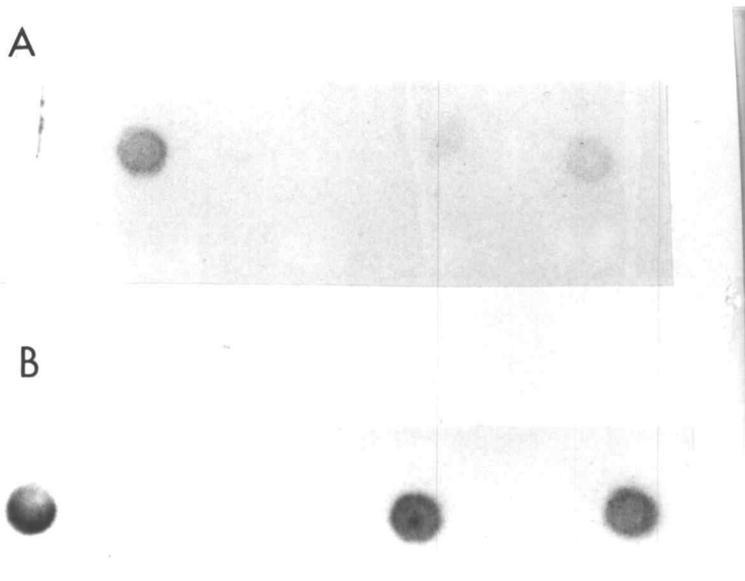


Figure 27. Hybridization of DNA from T4 mutants deleted in 63-32 region. Panel A, dots were hybridized to  $^{32}\text{P}$ -labelled pSP19 and in panel B to  $^{32}\text{P}$ -labelled 1.1-kb fragment.

## V. DISCUSSION

The molecular cloning and nucleotide sequence determinations presented in this dissertation have demonstrated that the 1.1-kb Hind III fragment from bacteriophage T4 contains the entire gene for dihydrofolate reductase and that there is at least a four-base pair overlap in the structural genes of dihydrofolate reductase and the functionally related enzyme thymidylate synthase. This work has also shown that the 5' flanking region of the dihydrofolate reductase structural gene has sequences to which RNA polymerase and ribosomes can bind to initiate transcription and translation, respectively. These sequences are most likely used for the expression of dihydrofolate reductase in pSP19 (Figures 8, 14) and also in pSP38 (Figure 14). Evidence is also presented that shows that the deletion mutants which have deletions in the 32-63 region, the region which contains the genes for DHFR and TS, have sequences that are homologous to the 1.1-kb Hind III fragment.

### A. Molecular Cloning of 1.1-kb Hind III Fragment

As stated earlier, T4 dihydrofolate reductase is of interest for a variety of reasons. To study this enzyme and its gene frd at the molecular level, attempts have been made in the past to clone the gene frd. Mainly two

restriction enzymes, Pst I and EcoR I were used to digest the T4 DNA for these cloning experiments. However, it was later shown by restriction mapping (88) that frd contains one internal restriction site for EcoR I, and that the Pst I fragment used in earlier cloning experiments (169) contains gene 32 (Figure 28) which codes for single stranded DNA binding protein. These clones were unstable.

Genetic evidence (89) and restriction mapping (86) have demonstrated that there is no internal restriction site for Hind III in frd; hence, in the present experiments, I used the restriction enzyme Hind III to digest the T4 dC DNA. Digestion of T4 dC DNA with Hind III results in a large number of fragments of varying sizes. Therefore, to facilitate screening of the transformants, a 16-kb Sma I fragment that contains frd and its flanking regions was isolated. Digestion of this fragment with Hind III produces a 1.1-kb fragment containing frd and five other fragments of varying sizes (Figure 10). Selection of the specific transformant carrying the gene for T4 DHFR was achieved both by its resistance to high concentration of TMP as compared to the host enzyme, and also by the size of the insert of 1.1-kb into the recombinant plasmid, pSP19. I have demonstrated, by comparing properties of DHFR from T4 infected cells and their pSP19 specified counterparts, that these two are identical (59). Production of active T4 specific DHFR in E. coli

carrying pSP19 shows that the insert carries the entire structural gene for T4 DHFR. These results are supported by genetic evidence (89), which also suggests that this 1.1-kb Hind III fragment contains, in addition to the structural gene, the regulatory sequences required for its expression. However, from the available data, it cannot be concluded whether frd transcription in pSP19 initiates from its own self or from a plasmid promoter. Interestingly, this fragment was found inactive in an in vitro transcription-translation system (88).

The Hind III site of pBR322 where the 1.1-kb fragment has been inserted is close to two partially overlapping promoters,  $P_A$  and  $P_B$  (corresponding to  $P_1$  and  $P_2$  of the reference 170). These promoters initiate transcription in opposite directions (170). In pBR322, the transcription of the tetracycline resistance ( $Tc^R$ ) gene is initiated at the  $P_B$  which lies on the left of the Hind III site and that of beta-lactamase (*bla*) gene is initiated at  $P_A$ , located on the right of this site. Transcription of frd from  $P_B$  would be clockwise on pSP19 as depicted in Figure 8. Likewise, transcription from the resident promoter would also be clockwise; this is deduced from the orientation of transcription of frd in the T4 genome, as shown by the alignment of physical and genetic maps in this region (87, 88, 89). DNA inserted at the Hind III site in pBR322 can be transcribed by  $P_A$ ,

the leftward promoter that transcribes bla gene in pBR322; however,  $P_B$ , the  $Tc^R$  promoter, is known to be inactivated by Hind III cleavage. The 1.1-kb fragment in pSP19 is inserted in the orientation such that it will be transcribed in the rightward direction like  $Tc^R$ . However, I was unable to obtain any transformants with the insert in the reverse orientation when pSP19 was digested with Hind III, religated and selected for TMP resistance after transformation by this ligated DNA. Similar results were obtained independently by Hanggi and also by Sargent and Hall (personal communication) even though Dr. D. Hall selected the transformants that were  $Ap^R Tc^S$  and was selecting only for the insert and not for the expression of the gene. It is likely that the 1.1-kb fragment, when inserted in the reverse orientation, is transcribed by bla promoter,  $P_A$ , which is more efficient than  $P_B$  (170). Failure to obtain the plasmids with the insert in the orientation opposite to that in pSP19 may be a consequence of overproduction of DHFR. This interpretation is supported by the experiments presented in this dissertation in relation to cloning of frd in the high expression vector, pGW7. Under conditions of overproduction, the cells carrying the plasmid pSP104 were unable to grow. The high expression vector, pGW7, was constructed (by Dr. G. Wilson of New England Bio-Labs) by inserting a fragment of phage lambda in EcoR I-BamH I-

cut pBR322 and inserting the 1.1-kb fragment at BamH 1 site in this vector. Synthesis of active T4 specific DHFR in transformants carrying the recombinant plasmid, pSP38 with insert in orientation opposite to that of pSP104, suggests that the 1.1-kb fragment contains the regulatory sequences necessary for the expression of frd. However, from the available information it cannot be concluded whether or not these same sequences also regulate the synthesis of DHFR in vivo.

#### B. Nucleotide Sequence of the Cloned 1.1-kb Fragment

A stretch of 579 nucleotides between positions 335 and 913 has been identified as the structural gene for dihydrofolate reductase. The open reading frame consists of 579 nucleotides and can encode a peptide of 193 residues. The estimated molecular weight of this peptide, 21,603, is in good agreement with the experimentally determined value 23,000 (70), and the amino acid composition of the encoded peptide is comparable to that of purified DHFR from T4-infected cells as well as from hyperproducers (127, 128). In addition, several residues which are implicated in inhibitor and substrate binding are also present in the deduced amino acid sequence. Finally, the most significant evidence came from the N-terminal sequence of the enzyme purified from the over-producer, pSP104. The first 20 amino acid sequence is

identical to the deduced sequence. Although C-terminal sequence of the purified enzyme has not yet been determined, comparison of the composition of amino acids and the similar rates of synthesis of both DHFR and TS after T4 infection (14) suggests that the translation of DHFR terminates at the first termination codon -TGA- that overlaps with the initiation of TS -ATGA- (SD4) between nucleotides 914-916.

The -ATGA- overlap between two coding sequences has been observed in a chain of interlinked genes in the nin<sup>R</sup> region of bacteriophage lambda (171) and also near T4 gene 30 (172). The authors suggest that this structure plays a role at the translation level of gene expression. According to their interpretation, the genes with -ATGA- overlap should be expressed in stoichiometric amounts. They also observed polarity effect in the expression of these genes. They reported that the untranslated segments of proximal mRNA of these interlinked genes lead to a reduction in transcription rates of distal DNA segments. Although there is some evidence suggesting that both DHFR and TS may be synthesized in similar amounts upon T4 infection (14), there are three major differences in regulation of expression of the frd and td genes of T4. First, DHFR activity appears earlier in infection than TS activity (16), with synthesis of DHFR polypeptide chain occurring entirely within

the first ten minutes (59). Second, while td behaves like any other early genes in being hyperexpressed after infection by a T4 DNA negative amber mutant, DHFR activity is not overproduced when DNA replication is blocked (16). Third, while DHFR is overproduced in some of the far mutants, TS is underproduced (26, 97). As mentioned earlier, these far mutants are regulatory mutants and map in the mot gene, the product of which is required for the readthrough at IE/DE junctions and also for the recognition of middle promoters (95, 96). It is conceivable that the two enzyme proteins are translated from the same mRNA, since Trimble and Maley (94) showed the T4 DHFR message to be at least 1600 nucleotides long. Although this would seem to be enough information to encode polypeptides of 22,000 and 29,000 daltons (the molecular weights of T4 DHFR and TS, respectively), it is not yet known where transcription initiates for frd expression. Upstream from the DHFR coding sequence, I find in the pSP19 insert a potential promoter with -35 and -10 sequences of TTGTGA and TATTAT, respectively. These are reasonably close to the consensus sequences for E. coli RNA polymerase of TTGACA and TATAAT, respectively (89, 123). I believe that this represents the promoter for frd transcription, in part because the frd gene embedded in the 1.1-kb fragment is expressed in either orientation after cloning into pGW7 (Table XII and Figure 13).

When the gene is not under lambda  $\rho_L$  control, the level of DHFR activity is comparable to that seen in pSP19, suggesting that the gene is transcribed from its own promoter in both cases (compare SP19 with SP38, Table XIII).

A transcript originating at the putative frd promoter and extending 1600 bases downstream would be expected to include the entire coding sequence for thymidylate synthase. However, Chu et al. (174) have made the surprising finding of a 1,107-base pair intervening sequence in the td gene. A 1600-nucleotide mRNA could encode both enzymes only if the intervening sequence had first been processed out.

It is possible that td is also transcribed from the presumptive regulatory sequences that are present in the DHFR structural gene. There are at least three such sequences (Figures 20 and 22), P3, P4 and P5. As mentioned earlier, the -10 regions of these sequences are highly homologous to the corresponding region of E. coli promoters. However, the -35 region shows homology to a lesser degree. Two of these presumptive regulatory sequences, P3 and P5, have sequences similar to -35 regions of rIIB and gene I (Fig. 22). Both rIIB and gene I are expressed from middle promoters (96) and, like td, are gp mot sensitive (102, 103). The sequence -ATTGCTTC- of rIIB and gene I is the proposed consensus sequence for

gp mot interaction (96). Another regulatory element that is required in the expression of a gene is the region where ribosomes bind for the initiation of translation. The purine rich segment -TACGGAATCA- upstream from the initiation codon of TS constitutes a weak ribosome binding site (106, 107). Possibly, the overlap between the coding regions of two genes, frd and td, may facilitate the translation of TS. Ribosomes, after completing the translation of DHFR, may start the translation of TS before they fall off the transcript. The presence of regulatory sequences in the segment between nucleotides 694, which constitutes the EcoR I restriction site, and 913, the 5' terminus of the structural gene for TS, are supported by the genetic data reported by Mileham et al. (93). They could complement a  $thy^-$  host by a recombinant phage lambda that carries a 1.7-kb E. coli fragment. The 5' end of this fragment originates at position 694 of the 1.1-kb insert in pSP19. In addition, under conditions where lambda promoters are repressed, a polypeptide similar in size to bacteriophage T4 thymidylate synthase subunit was detected on SDS polyacrylamide gels.

The four-base pair overlap in the structural genes of DHFR and TS is also intriguing when compared to DHFRs from parasites. These enzymes are dimers and are bifunctional, so that DHFRs have also associated TS

activity. The subunit molecular weight of DHFR from Crithidia fasciculata (25) of 58,000 is approximately the same as the sum of the subunit molecular weights of bacteriophage T4 dihydrofolate reductase and thymidylate synthase (59, 86).

In 1970, Burchall classified DHFRs into three phylogenetic classes: vertebrates, invertebrates, and bacteria. This classification was based on the comparison of the ratios of concentration of inhibitors, trimethoprim and triazines, that are required to inhibit the enzyme activity by 50% (the binding ratios). As expected, these ratios are highest for vertebrate DHFRs and are clustered together. The lowest value of the TMP/triazines binding ratio is for the bacterial enzymes; however, small differences are observed among the individual values. For invertebrates, which also include protozoal enzymes, the binding ratio is intermediate and shows a wide spread in these values (23). I made a similar analysis for type I and T4 dihydrofolate reductases using published data (14, 23, 25, 60, 24, 26), and surprisingly, the TMP/triazines binding ratios for both of these enzymes fell in the same range as those of invertebrate enzymes. However, these binding ratios are on the two extremes within this class. The binding ratio for T4 DHFR is of the same order of magnitude as that of Plasmodium berghei and represents the upper limit, and

that of type I DHFR is similar to Trypanosome equiperdum, representing the other extreme of this class (Table I).

The possible overlap -ATGA- between the structural genes of bacteriophage T4 DHFR and TS, the dimeric nature of both of these enzymes, the sum of the molecular weights of T4 DHFR and TS, approximately 100,000, that is similar to that of DHFR-TS bifunctional enzyme of protozoa and the similar TMP/ triazine binding ratios of T4 and protozoan enzymes are interesting coincidences. Simple mutational events could have created overlapping genes from a gene encoding a fused polypeptide chain, or vice versa. It will be intriguing indeed to learn whether the T4 sequences reported here and also determined by Chu et al. (174) bear close relationships to corresponding sequences in protozoal genomes.

Two additional possible overlaps have also been identified in the 5' region of frd. The first overlap is between the termination codon of the upstream open reading frame and the Shine and Dalgarno sequence of the leader sequence. From the available data, it cannot be concluded whether the open reading frame is translated or whether it has any functional significance in the expression of neighboring gene(s). However, a strong ribosome-binding site upstream from the ORF (Figure 20) and a close correspondence between potential codons in the ORF and codons actually used in T4 messages (Gary Stormo,

personal communication) suggest that this region is translated. Another interesting feature of this ORF is evident when the segment between 85 to 105 nucleotides from the 5' end of the 1.1-kb fragment is translated. This deduced heptapeptide ile-gly-lys-asp-ser-glu-leu, is highly homologous to a conserved sequence in the active site of all known DHFRs. It is possible that this region might represent a quasi-pseudogene.

Although the cloned 1.1-kb Hind III fragment directs the synthesis of active DHFR and has regions that can function as the regulatory sequences for both transcription and translation of frd, this fragment was found to be inactive in the in vitro transcription-translation system (92). In this same system, a larger Hind III fragment which has additional sequences upstream to the 5' end of the cloned fragment directed the synthesis of active DHFR (92). This observation suggests that there might be some sequences in the larger fragment that facilitate the expression of frd. It is likely that due to the lack of promoter sequence upstream to ORF on the 1.1-kb fragment, this ORF is not transcribed, which in some way may be responsible for the lack of syntheses of the enzyme protein in the in vitro system. This ORF terminates in an opal codon which overlaps with the Shine and Dalgarno sequence (SD2) of another small reading frame, the leader sequence. The segment between nucleo-

tides 280 to 352, which includes this entire leader region and twenty nucleotides of the 5' end of frd contains the complementary inverted repeats (Figures 20 and 22). In panel A of Figure 22, although the initiation codon of DHFR is exposed in loop b, the proposed Shine and Dalgarno sequence (SD3) is buried in the stem formed by the association of the complementary inverted repeats, 3 and 4. In panel B of this Figure, both elements of translation, the initiation codon and the Shine and Dalgarno sequence are present in the loop formed by the association of the complementary repeats, 2 and 4. This rearrangement of the RNA may be achieved either by the interaction of charged fmet-tRNA at the initiation codon in loop b (175) that may facilitate the isomerization of the RNA helix to expose the ribosome binding site as shown in panel B, or disruption of stem a, either by stalling of ribosomes at ile codons present in tandem in the sequence 1 of this stem, or by the interaction with some factor at the 5' region of the leader sequence (176). Either the ORF and/or the leader sequence may be involved in the regulation of frd. As a preliminary approach to the question of whether the ORF is related to the downstream frd gene, I deleted most of the information in this ORF but kept the potential secondary structures intact, by restricting the 1073 base-pair fragment with restriction endonuclease Mbo II. As noted in

Figures 18 and 20, the ORF contains two sites for cleavage by this enzyme. When the purified large fragment was recloned into pBR322 at Hind III site, no apparent effect upon T4 DHFR activity was observed in the transformed cells (data not shown). However, these transformed cells have unusual properties. They grow very slowly even in the rich medium containing 50 µg/ml AMP. Also efforts to isolate plasmids even from the cultures that have been amplified by chloramphenicol treatment were unsuccessful. Perhaps presence or expression of this ORF is necessary for maintaining some higher order structures to facilitate the expression of frd, and due to the selection pressure, integration in the genome may somehow have bypassed this requirement. It will be interesting to probe the chromosomal DNA for the presence of sequences complementary either to the Mbo II large fragment and/or to pSP19. Deletion of or mutation in the leader sequence that disrupts the secondary structures but leaves the regulatory sequences intact will be useful in further exploring the possible role(s) of these upstream reading frames in the regulation of expression of frd and/or td.

Insertion of the 1.1-kb fragment into a hyperexpression vector was of great advantage for the present work and will be useful for future experiments where large quantities of purified DHFR will be needed. As a result of the high concentration of DHFR protein in the cell

extract, the purification of the enzyme became very simple. Gerry Lasser in Dr. Mathew's laboratory was able to purify this enzyme protein in two simple steps, namely streptomycin sulfate followed by ammonium sulfate precipitation and dialysis. Depending upon the requirements, a passage through hydroxylapatite column may also be needed. The protein purified from these cells was used to determine the N-terminal sequence of T4 DHFR, and the determination of the C-terminal sequence is underway. The DHFR protein so obtained was also subjected to amino acid analysis. The results are listed in Table XIII and are comparable to those obtained from the analysis of the enzyme purified from T4-infected cells as well as to the peptide translated from the nucleotide sequence. This indicates that all these manipulations of DNA have not altered or modified the protein.

Finally, the hybridization data obtained for del7 and del9 suggest that the sequences complementary to 1.1-kb fragment are present in the genome of these mutants. The 11-kb band which is present in del7 and del9 but absent from del1 EcoR V restriction patterns is of interest. It is possible that the segment deleted from 32-63 region of the T4 genome (112) is duplicated/inserted somewhere else in the genome and is represented by this extra band in del7 and del9. Should this be the case, the 11-kb band should hybridize to  $^{32}\text{P}$  labelled 1.1-kb fragment.

## VI. REFERENCES

1. Wood, W. B., and Revel, H. R. (1976) *Bacteriol. Rev.* 40, 847-868
2. Mosig, G. (1983) in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G. Berget, P. B., eds.) American Society for Microbiology, pp. 362-374
3. Cohen, S. S. (1968) Virus-induced Enzymes, Columbia University Press, New York
4. Mathews, C. K. (1971) Bacteriophage Biochemistry, Van Nostrand Reinhold co., New York
5. Mathews, C. K. (1977) *Compr. Virol.* 7, 179-294
6. Mathews, C. K., and Allen, J. R. (1983) Bacteriophage T4, 59-70
7. Reddy, G. P. V., and Mathews, C. K. (1978) *J. Biol. Chem.* 353, 3461-3467
8. Wovcha, M. G., Chiu, C. S., Tomich, P. K., and Greenberg, G. R. (1976) *J. Virol.* 20, 142-156
9. Allen, J. R., Lasser, G. W., Goldman, D. A., Booth, J. W., and Mathews, C. K. (1983) *J. Biol. Chem.*, 5746-5753
10. Flaks, J. G., and Cohen, S. S. (1957) *Biochem, Biophys. Acta* 25, 666-669
11. Flaks, J. G., and Cohen, S. S. (1959) *J. Biol. Chem.* 234, 2981-2986
12. Mathews, C. K., Cohen, S. S. (1963) *J. Biol. Chem.* 238, PC853
13. Mathews, C. K., and Sutherland, K. E. (1965) *J. Biol. Chem.* 240, 2142-2147
14. Mathews, C. K. (1967) *J. Biol. Chem.* 242, 4083-4086
15. Warner, H. R., and Lewis, N. (1966) *Virology* 29, 172-175
16. Mathews, C. K. (1966) *Biochemistry* 5, 2091-2100

17. Hall, D. M., Tessman, I. and Karlstrom, O. (1967) *Virology* 31, 442-448
18. Mathews, C. K. (1967) *J. Virol.* 1, 963-967
19. Hall, D. H. (1967) *Proc. Nat. Acad. Sci. U.S.A.* 58, 584-559
20. Blakley, R. L. (1969) in The Biochemistry of Folic Acid and Related Pteridines (Neuberger, A., and Tatum, E. L., eds.) North-Holland Publishing Company, Amsterdam, London
21. Bertino, J. R., and Johns, D. (1972) in Cancer Chemotherapy (Bradsky, I., ed.) Grune and Stratton, New York
22. Freisheim, J. H., and Matthews, D. A. (1983) in Folate Antagonists as Therapeutic Agents, Vol. 1 (Sirotnak, F. M., Burchall, J. J., Ensminger, W. B., and Montgomery, J. A., eds.) Academic Press, New York, in press
23. Burchall, J. J. (1970) *Ann. N.Y. Acad. Sci.* 1, 143-152
24. Pattishall, K. H., Acar, J., Burchall, J. J., Goldstein, F. W., and Harvey, R. J. (1977) *J. Biol. Chem.* 252, 2319-2323
25. Ferone, R., and Roland, S. (1980) *Proc. Nat. Acad. Sci. U.S.A.* 77, 5802-5806
26. Trimble, R. B., Galivan, J., and Maley, F. (1972) *Proc. Nat. Acad. Sci. U.S.A.* 69, 1659-1663
27. Hryniuk, W. A., and Bertino, J. R. (1969) *J. Clin. Invest.* 48, 2140-2155
28. Hakala, M. T., Zakrzewski, S. F., and Nichol, C. A. (1961) *J. Biol. Chem.* 236, 952-958
29. Alt, F. W., Kellems, R. E., Bertino, J. R., and Schimke, R. T. (1978) *J. Biol. chem.* 253, 1357-1370
30. Morandi, C., and Attardi, G. (1981) *J. Biol. Chem.* 256, 10169-10175
31. Coderre, J. A., Beverley, S. M., Schimke, R. T., Santi, D. V. (1983) *Proc. Nat. Acad. Sci. U.S.A.* 80, 2132-2136

32. Nunberg, J. H., Kaufman, R. J., Schimke, R. T., Urlaub, G., and Chasin, L. A. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 5553-5556
33. Dolnick, B. J., Berenson, R. J., Bertino, J. R., Kaufman, R. J., Nunberg, J. H., and Schimke, R. T., (1979) J. Cell. Biol. 83, 394-402
34. Kaufman, R. J., Brown, P. C., and Schimke, R. T. (1979) Proc. Nat. Scad. Sci. U.S.A. 76, 5669-5673
35. Haber, D. A., and Schimke, R. T. (1981) Cell. 26, 355-362
36. Sheldon, R., and Brenner, S., (1976) Molec. Gen. Genet. 147, 91-97
37. Smith, D. R., and Calvo, J. M. (1979) Molec. Gen. Genet. 175, 31-38
38. Smith, D. R., and Calvo, J. M. (1982) Mol. Gen. Genet. 187, 72-78
39. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353
40. Sirotnak, F. M., and McCuen, R. W. (1973) Genetics, 74, 543-556
41. Goldie, J. M., Krystal, G., Hartley, D., Gudouskas, G., and Dedhar, S. (1980) Eur. J. Cancer, 16, 1539-1546
42. Goldie, J. H., Dedhar, S., and Krystal, G. (1981) J. Biol Chem. 256, 11629-11635
43. Haber, D. A., Beverley, S. M., Kiely, M. L., and Schimke, R. T. (1981) J. Biol. Chem. 256, 9501-9510
44. Sirotnak, F. M. (1973) Biochem. Biophys. Acta 302, 13-23
45. Freisheim, J. H., Smith, C. C., and Guzy, P. M. (1972) Arch. Biochem. Biophys. 148, 1-9
46. Ferone, R. (1970) J. Biol. Chem. 245, 850-854
47. Sirotnak, F. M., Hachtel, S. L., Williams, W. A. (1969) Genetics 61, 313-326
48. Goldberger, R. F. (1974) Science 183, 810-816

49. Gronenborn, A. M., and Davies, R. W. (1981) *J. Biol. Chem.* 256, 12152-12155
50. Goldman, I. D. (1971) *Ann. N. Y. Acad. Sci.* 186, 400-422
51. Bertino, J. R., Nixon, P. F. and Nahas, A. (1977) in Food and Nutrition Board Research Council (eds.) Folic Acid: Biochemistry and Physiology in Relation to the Human Nutrition Requirements, 178-187, Nat. Acad. of Sci., Washington, D. C.
52. Huennekens, F. M., Vitols, K. S., and Henderson, G. B. (1978) *Advan. Enzyme* 47, 313-346
53. Sirotnak, F. M., Kurita, S., and Hutchinson, D. J. (1968) *Cancer Res.* 28, 75-80
54. Hill, B. T., Bailey, B. D., and Goldman, I. D. (1978) *Proc. Am. Assoc. Cancer Res.* 19, 49
55. Lindquist, C. A., Moroson, B. A., and Bertino, J. R. (1978) *Proc. Am. Assoc. Cancer Res.* 19, 165
56. Sköld, O., and Widh, A. (1974) *J. Biol. Chem.* 249, 4324-4325
57. Amyes, S. G. B., and Smith, J. T. (1974) *Biochem. Biophys. Res. Commun.* 58, 412-418
58. Baccanari, D. P., Philips, A., Smith, S., Sinski, D., and Burchall, J. J. (1975) *Biochemistry* 14, 5267-5273
59. Purohit, S., Bestwick, R. K., Lasser, G. W., Rogers, C. M., and Mathews, C. K. (1981) *J. Biol. Chem.* 256, 9121-9125
60. Fling, M. E., Walton, L., Elwell, L. P. (1982) *Antimicrob. Agents Chemother.* 22, 882-888
61. Gutteridge, W. E., Jaffe, J. J., and McCormick, J. J., Jr. (1969) *Biochem. Biophys. Acta* 191, 753-755
62. Ferone, R., Burchall, J. J., and Hitchings, G. M. (1969) *Mol. Pharmacol.* 5, 49-59
63. Wang, C. C., Stotish, R. L., and Poe, M. (1975) *J. Protozool.* 22, 564-568

64. Fleming, M. P., Datta, N., and Grüneberg, R. N. (1972) *Br. Med. J.* 1, (5802) 726-728
65. Fling, M. E., and Elwell, L. P. (1980) *J. Bacteriol.* 141, 779-785
66. Smith, S. L., Stone, D., Novak, P., Baccanari, D. P., and Burchall, J. J. (1979) *J. Biol. Chem.* 254, 6222-6225
67. Stone, D., and Smith, S. L. (1979) *J. Biol. Chem.* 254, 10587-10861
68. Swift, G., McCarthy, B. J., and Heffron, F. (1981) *Mol. Gen. Genet.* 181, 441-447
69. Zolg, J. W., and Hänggi, U. J. (1981) *Nuc. Acids Res.* 9, 697-710
70. Fling, M. E., Richards, C. (1983) *Nuc. Acids Res.* 11, 5147-5158
71. Mosher, R. A., Di Renzo, A. B. and Mathews, C. K. (1977) *J. Virol.* 23, 645-658
72. Gleisner, J. M., Peterson, D. L., and Blakley, R. L. (1974) *Proc. Nat. Acad. Sci. U.S.A.* 71, 3001-3005
73. Matthews, D. A., Alden, R. A., Bolin, J. T. Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M., and Hoogsteen, K. (1977) *Science* 197, 452-455
74. Matthews, D. A., Alden, R. A., Bollin, J. T., Filman, D. J., Freer, S. T., Hamlin, R., Hol, W. G. J., Kisliuk, R. L., Pastore, E. J., Plante, L. T., Xuong, N., and Kraut, J. (1978) *J. Biol. Chem.* 253, 6946-6954
75. Volz, K. W., Matthews, D. A., Alden, R. A., Freer, S. T., Hansch, C., Kaufman, B. T., and Kraut, J. (1982) *J. Biol. chem.* 257, 2528-2536
76. Erickson, J. S., and Mathews, C. K. (1971) *Biochem. Biophys. Res. Comm.* 43, 1164-1170
77. Erickson, J. S., and Mathews, C. K. (1972) *J. Biol. Chem.* 247, 5661-5667
78. Kozloff, L. M., Verses, C., Lute, M., and Crosby, L. K. (1970) *J. Virol.* 5, 740-753

79. Kozloff, L. M. (1983) in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B., eds.) pp. 25-31, American Society for Microbiology, Washington, D. C.
80. Mathews, C. K. (1971) *J. Virol.* 7, 531-533
81. Mathews, C. K., Crosby, L. K., and Kozloff, L. M. (1973) *J. Virol.* 12, 74-78
82. Dawes, J., and Goldberg, E. B. (1973) *Virology* 55, 380-390
83. Kozloff, L. M., Lute, M (1965) *J. Mol. Biol.* 12, 780-792
84. Kozloff, L. M., Lute, M., Crosby, L. K., Rao, N., Chapman, V. A., and Delong, S. S. (1970) *J. Virol.* 5, 726-739
85. Mosher, R. A., and Mathews, C. K. (1979) *J. Virol.* 31, 94-103
86. Capco, G. R., Krupp, J. R., and Mathews, C. K. (1973) *Archives of Biochem. and Biophys.* 158, 726-735
87. Capco, G. R., Mathews, C. K. (1973) *Archives of Biochem. and Biophys.* 158, 736-743
88. Kozloff, L. M., Crosby, L. K., and Lute, M. (1975) *J. Virol.* 16, 1409-1419
89. Kozloff, L. M. (1981) in Bacteriophage Assembly (Du Bow, M. ed.) pp. 327-342, Alan R. Liss, Inc., New York
90. Kozloff, L. M. and Zorzopulos, J. (1981) *J. Virol.* 40, 635-644
91. O'Farrell, P. H., Kutter, E., and Nakanishi, M. (1980) *Mol. Gen. Genet.* 179, 421-435
92. Hänggi, U. J., and Zachau, H. G. (1980) *Gene* 9, 271-285
93. Mileham, A. J., Revel, H. R., and Murray, N. E. (1980) *Mol. Gen. Genet.* 179, 227-239
94. Rabussay, D., and Geiduschek, E. P. (1977) *Compre. Virology* 8, 1-196

95. Rabussay, D. (1982) in Molecular Aspects of Cellular Regulation (Cohen, P. and Van Heynigan, S., eds) 2, 219-331, Elsevier/North-Holland Biomedical Press, New York
96. Brody, E., Rabussay, D., and Hall, D. H. (1983) in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B., eds.) pp. 174-183, American Society for Microbiology, Washington, D. C.
97. Trimble, R. B., and Maley, F. (1973) *Biochem. Biophys. Res. Comm.* 52, 1026-1033
98. Mattson, T., Richardson, J., and Goodin, D. (1974) *Nature* 250, 48-50
99. Mattson, T., Van Houwe, G., and Epstein, R. H. (1978) *J. Mol. Biol.* 126, 551-570
100. Johnson, J. R., and Hall, D. H. (1973) *Virology* 53, 413-426
101. Johnson, J. R., and Hall, D. H. (1974) *J. of Virology* 13, 666-676
102. Chace, K. V., and Hall, D. H. (1975) *J. of Virology* 15, 929-945
103. Hall, D. H., and Snyder, R. D. (1981) *Genetics* 97, 1-9
104. Homyk, T. A., Rodriguez, A., and Weil J. (1976) *Genetics* 83, 477-487
105. Singer, B. S., Shinedling, S. T., and Gold, L. (1983) in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B., eds.) pp. 327-333, American Society for Microbiology, Washington, D. C.
106. Shine, J., and Dalgarno, L. (1974) *Proc. Nat. Acad. Sci. U.S.A.* 71, 1342-1346
107. Shine, J., and Dalgarno, L. (1975) *Nature* 254, 34-38
108. Weil, J., Terzaghi, B., and Craseman, J. (1965) *Genetics* 52, 683-693
109. Parma, D. H., Ingraham, L. J., and Snyder, M. (1972) *Genetics* 71, 319-335
110. Parma, D. H. and Snyder, M. (1973) *Genetics* 73, 161-183

111. Symonds, N., van den Ende, P., Durston, A., and White, P. (1972) *Mol. Gen. Genet.* 116, 223-238
112. Weil, J. and Terzaghi, B. (1970) *Virology* 42, 234-237
113. Homyk, T., Jr., and Weil, J. (1974) *Virology* 61, 505-523
114. Champe, S. P. and Benzer, S. (1962) *J. Mol. Biol.* 4, 228-292
115. Barnett, L., Brenner, S., Crick, F. H. C., Shulman, R. G. and Watts-Tobin, R. J. (1967) *Phil. Trans. Royal Soc. London* 252, 487-560
116. Kutter, E. M., and R uger, W. (1983) in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B., eds.) pp. 277-190, American Society for Microbiology, Washington, D. C.
117. Kozloff, L. M., Lute, M. and Crosby, L. K. (1977) *J. Virol.* 23, 637-644
118. Snyder, L., Gold, L., and Kutter, E. M. (1976) *Proc. Nat. Acad. Sci. U.S.A.* 73, 3098-3102
119. Wilson, G. G., Tanyashin, V. D., and Murray, N. E. (1977) *Mol. Gen. Genet.* 156, 203-214
120. Wiberg, J. S. (1967) *J. Biol. Chem.* 242, 5824-5829
121. Wiberg, J. and Buchanan, J. (1964) *Proc. Nat. Acad. Sci. U.S.A.* 5, 421-428
122. Warner, H. R., Snustad, D. P. (1983) in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B., eds.) pp.103-109, American Society for Microbiology, Washington, D. C.
123. Bruner, R. D., Souther, A. and Suggs, S. (1972) *J. Virol.* 10, 88-92
124. Kutter, E. M., Beug, A., Slurs, R., Jensen, L. and Bradley, D. (1975) *J. Mol. Biol.* 99, 591-607
125. Sadowski, P., Warner, H., Hercules, K, Munro, J., Mendelsohn, S., and Wiberg, J. (1971) *J. Biol. Chem.* 246, 3431-3433
126. Snustad, D. P., Snyder, L., Kutter, E. M. in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B., eds.) pp. 41-55, American Society for Microbiology, Washington, D. C.

127. Purohit, S. and Mathews, C. K. (1983) Federation Proc. 42, 1759
128. Purohit, S. and Mathews, C. K. (1984) J. Biol. Chem. in press
129. Friedkin, M., Crawford, E. S., and Misra, D. (1962) Fed. Proc., Fed. Amer. Soc. Exp. Biol. 21, 176
130. Bolivar, F., Rodriguez, R. L., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. M., and Falkow, S. (1977) Gene 2, 95-113
131. Mathews, C. K., Brown, F., and Cohen, S. S. (1964) J. Biol Chem. 239, 2957-2963
132. Summerton, J., Atkins, T. and Bestwick, R. (1983) Anal. Biochem. 133, 79-84
133. Norgard, M. Emigholz, K., and Monahan, J. (1979) J. Bact. 138, 270-272
134. Clewell, D. B. (1972) J. Bacteriol. 110, 667-676
135. Bauer, W., and Vinograd, J. (1968) J. Mod. Biol. 33, 141-171
136. Holmes, D. S., and Quigley, M. (1981) Anal. Biochem. 114, 193-197
137. Birnboim, H. C. and Doly, J. (1979) Nuc. Acids Res. 7, 1513-1523
138. Cohen, S. N., Chang, A.C.Y., and Hsu, L. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 2110-2114
139. Maniatis, T., Fritsch, E. S., Sambrook, J. (1982) in Molecular cloning: A Laboratory Manual, 250; Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
140. Kiko, H., Niggemann, E., and Rürger, W. (1979) Mol. Gen. Genet. 172, 303-312
141. Dugaiczky, A., Boyer, H. W., and Goodman, H. M. (1975) J. Mol. Biol. 96, 174-184
142. Ferretti, Luca and Sgaramella (1981) Nuc. Acids Res. 9, 85-93

143. Mathews, C. K., and Huennekens, F. M. (1963) *J. Biol. Chem.* 128, 3436-3442
144. Bradford, M. (1976) *Anal. Biochem.* 72, 248-254
145. Peterson, G. L. (1977) *Anal. Biochem.* 83, 346-456
146. Maxam, A. M. and Gilbert, W. (1980) *Methods in Enzymology* 65, 499-560
147. Richardson, C. C. (1965) *Proc. Nat. Acad. Sci. U.S.A.* 54, 158-165
148. Winberg, G. and Hammarskjold, M. L. (1980) *Nuc. Acids Res.* 8, 253-264
149. Maxam, A. M. and Gilbert, W. (1977) *Proc. Nat. Acad. Sci. U.S.A.* 74, 560-564
150. Smith, D. R. and Calvo, J. (1980) *Nuc. Acids Res.* 8, 2255-2274
151. Sanger, F. and Coulson, A. R. (1978) *FEBS Lett.* 87, 107-110
152. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251
153. Thomas, P. S. (1980) *Proc. Nat. Acad. Sci. U.S.A.* 77, 5201-5205
154. Kafatas, F. C., Jones, C. W. and Efstratiadis, A. (1979) *Nuc. Acids Res.* 7, 1541-1552
155. Wahl, G. M., Stern, M., and Stark, G. R. (1979) *Proc. Nat. Acad. Sci. U.S.A.* 76, 3683-3687
156. Kutter, E., and R ger, W. (1983) in Bacteriophage T4 (Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B., eds.) pp. 277-290, American Society for Microbiology, Washington, D. C.
157. Sutcliffe, J. G. Cold Spring Harbor Symposium on Quantitative Biology (1979) Vol XLIII, 77-90
158. Carlson, K. and Nicolaisen, B. (1979) *J. Virol.* 31, 112-123
159. R ger, W., Neumann, M., Rohr, U. and Niggemann, F. (1979) *Mol. Gen. Genet.* 176, 417-425

## APPENDIX

160. Marsh, R. C. and Hepburn, M. L. (1981) *J. Virol.* 38, 104-114
161. Niggemann, F., Green, I., Meyer, H. P. and Rüger, W. (1981) *Mol. Gen. Genet* 184, 289-299
162. Belfort, M., Moellekan, A., Maley, G. F. and Maley, F. (1983) *J. Biol. Chem.* 258, 2045-2051
163. Engelberg-Kulka, H. (1981) *Nuc. Acids Res.* 9, 983-988.
164. Erickson, J. S. and Mathews, C. K. (1973) *Biochemistry* 12, 372-380
165. Edelhoch, H. (1967) *Biochemistry* 6, 1948-1954.
166. Staden, R. (1978) *Nuc. Acids Res.* 5, 1013-1015
167. Novak, P., Stone, D. and Burchall, J. J. (1983) *J. Biol. Chem.* 258, 10956-10959
168. Raub, W. F. (1974) *Federation Proc.* 33, 2390-2392.
169. Meade, L. C., Direnzo, A. B., Bestwick, R. K., Mosher, R. A. and Mathews, C. K. (1979) (Kisliuk/Brown, eds.) Chemistry and biology of Pteridines, 321-326, Elsevier North Holland, Inc.
170. Stuber, D. and Bujard, H. (1981) *Proc. Nat. Acad. Sci. U.S.A.* 78, 167-171
171. Kröger, M. and Hobom, G. (1982) *Gene* 20, 25-38
172. Armstrong, J., Brown, R. S., and Tsugita, A. (1983) *Nuc. Acids Res.* 11, 7145-7156.
173. Siebenlist, U., Simpson, R. and Gilbert, W. (1980) *Cell* 20, 269-276
174. Chu, F., Maley, F., Maley, G. F., and Belfort, M. (1984) *Proc. Nat. Acad. Sci. U.S.A.* in press
175. Gold, R. Pribnow, D. Schneider, T. Shinedling, S., Singer, B. S. and Stormo, G. (1981) *Ann. Rev. Microbiol.* 35, 365-402
176. Horinouchi, S. and Weisblum, B. (1980) *Proc. Natl. Acad. Sci. U.S.A.*, 77, 7079-7083

## VII. APPENDIX

## COMPOSITION OF MEDIA

Nutrient Broth (NB)

Per Liter:

Nutrient Broth.....	8g
NaCl.....	5g

Nutrient Agar Plates (NB Plates)

Per Liter:

Nutrient Broth.....	4g
Nutrient Agar.....	23g
NaCl.....	5g

Nutrient Agar Slants (ND Slants)

Per Liter:

Nutrient Broth.....	15g
Nutrient Agar.....	20g
NaCl.....	5g

Hershey Soft Agar (Top Agar)

Per Liter:

Nutrient Broth.....	8g
Bacto Agar.....	7g
NaCl.....	5g

Luria-Bertani Broth (LB)

Per Liter:

Bacto Tryptone.....	10g
Bacto Yeast Extract.....	5g
NaCl.....	5g
Glucose.....	1g

Penassay Broth

Per Liter:

Nutrient Broth.....	8g
Yeast Extract.....	1.5g
Dextrose.....	1g
NaCl.....	3.5g
$K_2HPO_4$ .....	3.6g
$KH_2PO_4$ .....	1.32g

M9 Salts

Per Liter:

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

Autoclaved

M9 Medium

Per Liter:

M9 Salts.....	880 ml
30% Glucose.....	17 ml

M9 Medium (continued)

20% MgSO <sub>4</sub> ·7H <sub>2</sub> O.....	0.85 ml
1% CaCl <sub>2</sub> .....	0.85 ml

SM9 Medium

Per Liter:

M9 Medium .....	990 ml
20% Casamino Acids.....	10 ml

Davis Medium

Per Liter:

K <sub>2</sub> HPO <sub>4</sub> .....	14g
KH <sub>2</sub> PO <sub>4</sub> .....	4g
Sodium Citrate.....	1g
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> .....	2g

Autoclaved, cooled and added:

20% MgSO <sub>4</sub> .....	1ml
30% Glucose.....	6.7ml
1% Thiamine-HCl.....	1ml
20% Casamino Acids.....	20ml

### Davis Plates

Five hundred milliliters of 2 x Davis Medium were mixed with 500 ml of hot, sterile 2.4% solution of Bacto Agar.

### Antibiotics

Stock solutions of Ampicillin (25 mg/ml) and Tetracycline (5 mg/ml) were prepared by dissolving the required amount in water and filter sterilizing by passing through a milipore filter.

Trimethoprim stock solution (0.1 M) was prepared by dissolving the required amount (29 mg) in one milliliter of DMSO. Further dilutions were made in water.