INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
DEOXYGUANOSINE AND DEOXYCYTIDINE KINASES ARISE FROM A SINGLE GENE IN

LACTOBACILLUS ACIDOPHILUS R-26

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By

Ning Ma

* * * * *

The Ohio State University

1996

Dissertation Committee:
D.H. Ives, Adviser
S. Patel
G.A. Marzluf
R.P. Swenson

Approved by

Adviser,
Biochemistry Department
Copyright by

Ning Ma

1996
To my dear wife, Xin
ACKNOWLEDGMENTS

I like to express my sincere appreciation toward Dr. David H. Ives for his kindness, guidance, advices, support and patience. I know deeply from my heart that without his help I could not finish this study.

I also like to thank Dr. Seiichiro Ikeda for his generous help in this study and valuable advice on many other things. I appreciate it very much for the help and suggestions from many of my friends, Grace Ma, Young Soo Hong, and Shenyuan Guo.

Especially I like to thank my dear wife and son for their patience and firm support.
VITA

May 9, 1955  Born in Shenyang City, China

1978  Graduated from Shenyang Agriculture College, China

1981  M.S. of Entomology Northwest Agriculture College, China

1989  Ph.D. of Entomology The Ohio State University

PUBLICATIONS


**FIELDS OF STUDY**

Major field: Biochemistry

Study in Protein Biochemistry  
Professor David H. Ives

Study in Molecular Genetics  
Professor David H. Ives
# TABLE OF CONTENTS

DEDICATION ........................................................... ii  
ACKNOWLEDGMENTS ....................................... iii  
VITA ................................................................. iv  
LIST OF TABLES ....................................................... x  
LIST OF FIGURES .................................................. xi  
ABBREVIATIONS .................................................. xiv  
ABSTRACT ................................................................. xvi  

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>A. A general review of deoxynucleoside kinases from <em>Lactobacillus acidophilus</em> R-26</td>
<td>1</td>
</tr>
<tr>
<td>B. Previous cloning study of the kinases</td>
<td>11</td>
</tr>
<tr>
<td>C. The original aims and overall thinking about the dCK gene in <em>L. acidophilus</em> R-26</td>
<td>15</td>
</tr>
<tr>
<td>II. Materials and Methods</td>
<td>17</td>
</tr>
</tbody>
</table>
A. Materials ................................................................. 17

B. Methods ................................................................. 19

1. PCR .............................................................. 19
2. Culturing L. acidophilus R-26 .................... 20
3. Purification of genomic DNA ......................... 21
4. Plasmid DNA preparation ............................... 22
5. Competent cells and transformation ............. 23
6. Genomic libraries of Lactobacillus acidophilus R-26 ........................................................... 24
7. Preparation of hybridization probes ............. 25
8. Southern hybridization .................................. 26
9. Cloning by PCR ............................................... 27
10. Cloning by colony hybridization ................. 28
11. Normal DNA Sequencing ................................ 29
12. Cycle DNA sequencing .................................. 30
13. Designing and making oligonucleotides ...... 31
14. Extraction of enzymes from E. coli ............. 32
15. Enzyme activity assay ................................. 33
16. Protein assay ................................................. 34
17. Steady state kinetics ...................................... 34
18. SDS electrophoresis of proteins .................. 35
19. Gel filtration chromatography ....................... 36
20. Total RNA purification ............................... 37
21. Northern Hybridization ................................ 38
22. RT-PCR .......................................................... 38
23. Site directed mutagenesis ........................... 39

III. The evidence that there is no distinct dCK gene in L. acidophilus R-26 genome ......................... 41
A. Statistical improbability that dCK can be cloned from \textit{L. acidophilus} R-26 genome ..................... 45

B. Southern Hybridization ......................... 46

1. Hybridized with dAK probes ................. 49
2. Hybridized with dCK/dGK probes .......... 54

C. Direct genomic sequencing .................... 61

1. Direct genomic sequencing of intact genomic DNA from \textit{L. acidophilus} R-26 .......... 63
2. Direct genomic sequencing of fragmented genomic DNA from \textit{L. acidophilus} R-26..66
3. Direct genomic sequencing on PCR products (dCK/dGK probes) of genomic DNA of \textit{L. acidophilus} R-26 ......................... 69

D. Peptide sequence comparison of wild type dAK, dGK and dCK revealed that dGK and dCK are too similar to be from separate genes ...................... 75

E. Molecular mass comparison of dAK, dGK and dCK suggested that dGK and dCK are too similar to be from separate genes ....................... 79

IV. dCK-like proteins can be produced from two mutants of the dGK gene ............................. 85

A. Mutant I, its gene product and its characterization and comparison with LBA wild-type dCK ........ 88
B. Mutant II, its gene product and its characterization and comparison with LBA wild-type dCK ........... 95

C. Similarities of the Mutant II gene product and the Human dCK ........................................... 105

D. Expression of the mutants in E. coli .............. 107

V. Conclusions, discussion and recommendations ..... 115

A. Conclusion and discussion on the hypotheses I proposed ...................................................... 115

B. Discussion on the possible mechanisms behind The one-gene-two-proteins phenomenon ................................. 126

BIBLIOGRAPHY ................................................................. 139
LIST OF TABLES

Table 1  Deoxynucleoside kinase comparison between purified LBA and cloned dAK/dGK gene product ............... 15

Table 2  Molecular mass comparison of dAK, dGK and dCK ................................................................. 80

Table 3  N-termini of dAK, dGK and dCK of L. acidophilus R-26 and N-termini of cloned dAK and dGK from E. coli ................................................................. 86

Table 4  Specific activity comparison of crude wild-type enzymes, cloned enzymes and the Mutant I products ....... 90

Table 5  Properties of the crude wild-type enzymes, cloned dAK/dGK and Mutant I gene products .................. 92

Table 6  Properties of the pure deoxynucleoside kinases from LBA, cloned dAK/dGK, and Mutant II gene product .... 97

Table 7  The similarities between the Mutant II and human dCK ................................................................. 106

Table 8  Specific activity comparison of the mutants and the unmodified enzymes ............................................ 112

Table 9  List of oligo-nucleotides used in this study .......... 138
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Putative modes of binding of substrates and multiple substrate analogs at the active sites of deoxynucleoside kinases (Figure from Ikeda (36), used with permission from the author)</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Partial N-terminal amino acid sequence of LBA dCK/dAK (Figure from Ikeda 38, used with permission from the author)</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Cloned dAK/dGK genes in pBluescript vector</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Sequences of the cloned dAK/dGK genes from <em>L. acidophilus</em> R-26 (Figure from G. Ma, used with permission from the author)</td>
</tr>
<tr>
<td>Figure 5</td>
<td>The dAK and dGK/dCK back-to-back primers used in probing similar region flanking the dAK/dGK genes in LBA genome</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Southern hybridization analysis of genomic DNA of <em>L. acidophilus</em> digested by restriction enzymes (Figure from G. Ma, used with permission from the author)</td>
</tr>
<tr>
<td>Figure 7</td>
<td>dAK probe, 375 bps, from 345 to 699 bp on the dAK gene, made by PCR and internally labeled by $^{35}$S or biotin</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Southern Hybridization with a dAK probe labeled with biotin</td>
</tr>
<tr>
<td>Figure 9</td>
<td>dGK/dCK probe, 675 base-pairs from the beginning to the end of the dGK gene, made by PCR and labeled with $^{35}$S</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Southern hybridization with the dCK/dGK probe, labeled</td>
</tr>
</tbody>
</table>
With $^{35}$S ......................................................... 59

Figure 11  Restriction map for the cloned dAK/dGK genes .... 60

Figure 12  Results of direct genomic sequencing with whole genomic DNA of LBA ................................. 65

Figure 13  Direct genomic sequencing with the fragmented genomic DNA of _L. acidophilus_ R-26. .......... 68

Figure 14  Sequencing of dGK/dCK probes made by PCR. . . 73

Figure 15  Result of amino acid sequencing of the dCK subunit from LBA. ................................................. 77

Figure 16  Result of molecular mass measurement of the native dCK by Nanospray method ....................... 81

Figure 17  Mutant I from the cloned dAK/dGK genes in the vector pBluescript ............................................. 88

Figure 18  Measurement of apparent K for dCyd for Mutant I gene product ............................................... 93

Figure 19  Mutant II from isolated dGK gene in the vector pBluescript .......................................................... 96

Figure 20  Cross inhibition and activation of the gene products of the Mutant II, isolated dGK gene, Mutant I and cloned dAK/dGK genes ................................................................. 98

Figure 21  Rate plot of the three kinase activities of the Mutant II gene product, showing the interaction of the two active sites. ................................................................. 99

Figure 22  Rate plot of the dCK activities of the Mutant II gene product under different protein concentrations, to show the cooperation between the subunits for the three activities of the
Mutant II gene product is not caused by concentration changes ........................................ 100

Figure 23 Hill plot for the three activities of the Mutant II gene product, to show strength of the cooperations between subunits ................................................................. 101

Figure 24 Molecular weight measurement of the gene products of Mutant I, II, isolated dGK gene and cloned dAK/dGK genes by gel filtration chromatography ................. 102

Figure 25 SDS polyacrylamide gel electrophoresis of the crude extracts from the cells of cloned dAK/dGK genes and their mutants ........................................................................ 110

Figure 26 SDS polyacrylamide gel electrophoresis for the purified Mutant II gene product with the g/a change at the S/D region ........................................................................ 111

Figure 27 Northern hybridization of wild time LBA, cloned dAK/dGK gene and its mutants with the dGK/dCK probe labeled with $^{32}$P ...................................................... 114

Figure 28 Design of RT-PCR. Purpose of the RT-PCR is to determine if any editing occurs at the 5' end area of the mRNA of the dGK gene ....................................................... 130

Figure 29 RT-PCR result for the total RNA preparations from the time-course of the LBA culturing, labeled with $^{32}$P ...... 131
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>dAdo</td>
<td>2'-deoxyadenosine</td>
</tr>
<tr>
<td>dAK</td>
<td>deoxyadenosine kinase</td>
</tr>
<tr>
<td>dATP</td>
<td>2'-deoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>dCK</td>
<td>deoxycytidine kinase</td>
</tr>
<tr>
<td>dCTP</td>
<td>2'-deoxycytidine 5'-triphosphate</td>
</tr>
<tr>
<td>dCyd</td>
<td>2'-deoxycytidine</td>
</tr>
<tr>
<td>dGK</td>
<td>deoxyguanosine kinase</td>
</tr>
<tr>
<td>dGTP</td>
<td>2'-deoxyguanosine 5'-triphosphate</td>
</tr>
<tr>
<td>dGuo</td>
<td>2'-deoxyguanosine</td>
</tr>
<tr>
<td>dTTP</td>
<td>2'-deoxymethyldine 5'-triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acids</td>
</tr>
<tr>
<td>HPLC</td>
<td>high pressure liquid chromatography</td>
</tr>
<tr>
<td>KM</td>
<td>Michaelis-Menten constant</td>
</tr>
<tr>
<td>LBA</td>
<td><em>Lactobacillus acidophilus</em> R-26</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acids</td>
</tr>
</tbody>
</table>
RT-PCR  reverse transcriptase polymerase chain reaction
SDS    sodium dodecyl sulfate
BRL    Bethesda Research Laboratories
USB    United Stated Biochemical Corp.
G-A    the mutant with the entire dAK gene removed and dGK gene intact
DEPC   diethyl pyrocarbonate
DTT    dithiothreitol
TK     thymidine kinase
MALDI  matrix-assisted laser desorption/ionization
ABSTRACT

Three deoxynucleoside kinases, deoxyadenosine kinase (dAK), deoxyguanosine kinase (dGK) and deoxycytidine kinase (dCK) previously purified and characterized from *Lactobacillus acidophilus* R-26, exist as highly specific heterodimeric pairs, dAK/dGK and dAK/dCK. The tandem genes for one of these pairs, encoding dAK/dGK, have been previously cloned and expressed in *E. coli* (1). The original goal of this study was to clone the dCK gene of the second pair. After thorough research, I will present the evidence, such as Southern hybridization and direct genomic sequencing, which strongly indicates that actually there is no distinct gene encoding dCK in the *L. acidophilus* R-26 genome. Furthermore, the peptide sequencing of the dCK from LBA and molecular mass comparison among the deoxynucleoside kinases shows that dCK and dGK are too similar to be from separate genes. In addition, a protein having dCK and dAK activities, but little dGK activity, has been produced from a dAK/dGK gene modified by simple deletion of the second and third codons only from
the dGK gene. In another mutant, having the entire dAK coding region of
the tandem genes deleted, the mutant dGK gene (the 2nd and 3rd codons
also deleted) is now over expressed as a broadly-specific homodimeric
kinase with dCK, dAK and dGK activities. These kinase activities all
exhibited large $K_m$ values, resembling the properties of human dCK toward
dGuo and dAdo. Based on all the evidence, we propose that there is no
distinct dCK gene in the LBA genome and that the dCK is produced from
the dGK gene.
CHAPTER I

INTRODUCTION

A. A GENERAL REVIEW OF DEOXYNUCLEOSIDE KINASES

The bacterial deoxynucleoside kinases have been the primary project in our lab for more than 10 years now, and since their discovery there has been much progress in study of their distribution, functions and regulations. The deoxynucleoside kinases consist of four functional kinases: thymidine kinase (TK), deoxyadenosine kinase (dAK), deoxyguanosine kinase (dGK) and deoxycytidine kinase (dCK). Deoxynucleoside kinases catalyze the transfer of a phosphoryl group from MgATP to deoxynucleosides to make deoxynucleoside 5'-monophosphates, each of which will be further phosphorylated by two additional kinases into deoxynucleoside triphosphates (dNTP) for DNA synthesis.
Bacterial thymidine kinase was first partially purified and characterized by Okazaki and Kornberg group from *E. coli* (2). The other three deoxynucleoside kinases were first discovered and partially purified and characterized by Durham and Ives, from *L. acidophilus* R-26 (3). Since there is no ribonucleotide reductase in this organism (4), all four deoxynucleoside kinases and a source of deoxyribose, therefore, are required for its growth.

The thymidine kinase has also been reported from many other organisms, including viruses, bacteria and most animals, including mammals. However, the other deoxynucleoside kinases (dAK, dGK, and dCK) are less widely distributed, and in particular, are reported in only a few bacterial species. All four deoxynucleoside kinase activities were reported only from *Lactobacillus leichmannii* (5) and *Bacillus subtilis* (6), in addition to *L. acidophilus* R-26. However, in contrast to *L. acidophilus* R-26, the presence of all four kinases for *Bacillus subtilis* are not obligatory since a spontaneous mutant lacking dCK and dAK activities continues to grow (6). As far as the prokaryotic organisms are concerned, deoxycytidine kinase was reported from *Bacillus megaterium* KM (7) and in class
Mollicutes (8) but only Spiroplasme ciri has dGK activity; Pneumococci were reported to have dAK and dGK activities (9), and the similar activities also was discovered from Salmonella typhimurium (10). In E. coli, only TK, but not the other three kinases, was discovered (11). Besides TK, there are no reports of the presence of these deoxynucleoside kinases from viruses, fungi, or plants. All three kinase activities, in addition to TK, were reported from calf thymus (12-14), murine neoplasm P815 (15), rat liver mitochondria (16), human lymphoblasts and myeloblasts (17-19), human leukemic spleen (20), and leukemic T-lymphoblasts (21,22). However, the \( K_m \) value of deoxycytidine is always the lowest among the substrates in all the instance. The deoxyguanosine kinase activity has been reported from calf thymus mitochondria (14), pig skin (23), neonatal mouse (24), and human placenta (25).

It is worth mentioning again that among the organisms from which deoxynucleoside kinases were found, the Lactobacillus acidophilus R-26 is a quite unique organism because all four deoxynucleoside kinases are essential for its survival (3,26). This organism cannot use the de novo dNTP production pathway for its DNA synthesis since it does not have
ribonucleotide reductase, which is the bridge between ribonucleoside- and deoxyribonucleoside diphosphate precursors for dNTP production for most organisms. However, the provision of only one deoxynucleoside is sufficient for its growth, since the bacterium can transfer deoxyribosyl group between bases (26).

After they were partially purified and characterized by Durham and Ives, the functions and mechanisms of these three nucleoside kinases (dAK, dGK and dCK) have been further studied. Actually, except for the detailed early work on TK from *E. coli* by Kornberg group and some work done for the enzymes from *B. Subtilis* (6), all of the work on the characterization of those deoxynucleoside kinases were done by the people from this lab with enzymes from *L. acidophilus* R-26. Therefore, it is worthwhile to generally summarize many people's work done here.

Deibel and Ives (27-29) first discovered that these deoxynucleoside kinase activities were associated in a pairwise fashion, that is, dAK/dGK and dAK/dCK respectively. Chakravarty *et al* (30) discovered that dAK and dGK activities were located on separate sites but on a single protein,
which she suggested was a single peptide with multiple functions; later, however, those enzymes were proved to have a heterodimeric construction.

Since it was realized that the nucleoside binding sites were exquisitely specific, using unaltered deoxynucleoside analogs to make affinity column had been proposed and investigated (31). The bi-substrate analog dNp4A, was used to build different affinity columns by attaching them to a hydrazide linker chain. By using these columns, the existence of separate binding sites for two deoxynucleoside substrates for the kinases was confirmed, and the concept of subunits (dAK/dGK and dAK/dCK) was established (32). It was also showed that this organization of deoxynucleoside kinase activities is very different from that of B. subtilis, since the latter has TK, dGK, dAK/dCK and this pair is catalyzed by a common active site, so there is no positive interaction between the active sites. Moreover, their activities are not inhibited strongly by the end products (6).
The inhibition and activation patterns, and substrate binding of those Lactobacillus deoxynucleoside kinases were studied by Ikeda *et al* (33), using some multiple-substrate-type analogs. They found that dNTPs, which were potent end product inhibitors of the corresponding kinases, exhibited kinetics identical to the bi-substrate inhibitors. This observation also confirmed that the kinetic mechanisms of dCK and dGK followed a rapid equilibrium random Bi Bi mechanism, while dAK followed an ordered Bi Bi mechanism. Therefore, for each pair of kinase activities, there are two different patterns of end product inhibitions. In addition, based on their investigations, they proposed a multiple-substrate binding model, shown in Figure 1.

Later, a new and more efficient affinity column with ligand of dCp4-linker or dAp4-linker (phosphate group was attached to the linker), instead of dNp4A-linker, was developed (34), which further facilitated the purification of those kinases. After the development of the new columns it became possible to produce some pure deoxynucleoside kinases for N-terminal amino acid residue determination. Before the dAK/dGK genes
from LBA were cloned, up to 28 amino acid residues from the N-termini for those kinases were determined (35), which is shown in Figure 2.

**Figure 1.** Putative modes of binding of substrates and multiple substrate analogs at the active sites of deoxynucleoside kinases. (Ikeda et al. (34), used with author's permission)
Figure 2. Partial N-terminal amino acid sequence of LBA dCK/dAK (36), used with permission from the author.
Ives (37) also developed the assay method for these deoxynucleoside kinases, which has become a routine tool for study of these kinases in today’s laboratories.

Based on all the previous studies, we can clearly see that besides thymidine kinase (TK), there are three other nucleoside kinases in *L. acidophilus* R-26, which have been well defined and characterized. In addition to their precise substrate specificities, these three enzymes have some distinctive characteristics to show they are REALLY individual enzymes (in the later chapters we will deal with the identity problem of dGK and dCK). For example, each is inhibited by its corresponding end-product and each subunit can be activated by its partner’s end product or nucleoside substrate.

On the other hand, these three kinases have many properties in common, which were used in their gene cloning and in other aspects of this study:
They have similar subunit molecular mass: about 26 kDa

Each enzyme has both an ATP site and a deoxynucleoside site

$K_m$ values for the deoxynucleosides are low for all of them (1 - 3 μM)

There is no cysteine residues in any enzyme, and only 1-2 tryptophans

The enzymes primarily form heterodimers

Positive allosteric effects exist between subunits and the substrates of the dGK and dCK toward dAK producing a much larger effect on dAK velocity than does dAdo in the reverse direction.
B. PREVIOUS WORK ON GENE CLONING OF THE DEOXYNUCLEOSIDE KINASES

Since the discovery of the nucleoside kinases, to clone their genes had become the next step in the study of these kinases. The first gene that was cloned and sequenced is the thymidine kinase gene from virus (38) in 1983. After that, the TK gene has been cloned and sequenced from other kind of viruses (39-43), human (44,45), chicken (46), and Chinese hamster cells (38), and from *E. coli* (47). Human dCK gene was cloned and sequenced from human leukemic T-lymphoblast MOLT-4 cells (48) in 1991.

Two of the kinase genes from the *L. acidophilus* R-26, dAK and dGK genes, have been recently cloned, expressed in *E. coli* and characterized by G. Ma and Y. S. Hong in our laboratory (1,49). The dAK gene contains 648 base pairs in the open reading frame, encoding 215 amino acid residues, and dGK gene contains 675 base pairs in the reading frame, encoding 224 amino acid residues. The reading frames of the two
genes are arranged as a tandem pair, separated by a 21-bp space. The relationship of the promoter and terminator sites of the two genes with their cloning vector, pBluescript (KS +) are shown in Figure 3. The DNA sequence and amino acid sequence inferred from those two genes can be found in Figure 4.

The dAK/dGK protein expressed in *E. coli* has most of the characteristics of the wild-type enzyme from *L. acidophilus* R-26, except that the $K_m$ value for dGuo is increased about 6.5 times, perhaps suggesting a folding or processing error (49). A comparison of the characteristics of the cloned dAK/dGK gene product and wild-type enzymes from LBA is shown in Table I. From this comparison it is clear that the cloned genes are expressed as dAK/dGK in *E. coli*. 
Figure 3. Cloned dAK/dGK genes in pBluescript vector, ligated at Kpn I site: P = promotor; T = terminator.
Figure 4. Sequence of the cloned dAK/dGK genes from LBA. Lower case letters indicate the nucleotide sequence; upper case letters indicate amino acid sequence. (Figure from Grace Ma (1), with the permission from the author)
Table 1: A comparison of the deoxynucleoside kinases from LBA and the cloned dAK/dGK gene product. All the enzymes are purified.

<table>
<thead>
<tr>
<th></th>
<th>LBA(dAK/dGK, dAK/dCK)</th>
<th>cloned dAK/dGK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Substrates</strong></td>
<td>dAdo and dGuo, or dCyd</td>
<td>dAdo, dGuo</td>
</tr>
<tr>
<td><strong>$K_m$ of dAdo</strong></td>
<td>2.2 $\mu$M</td>
<td>2.8 $\mu$M</td>
</tr>
<tr>
<td><strong>$K_m$ of dGuo</strong></td>
<td>10 $\mu$M</td>
<td>65 $\mu$M</td>
</tr>
<tr>
<td><strong>$K_m$ of dCyd</strong></td>
<td>2.2 $\mu$M</td>
<td>-</td>
</tr>
<tr>
<td><strong>Activators of dAK</strong></td>
<td>dGuo/dGTP, or dCyd/dCTP</td>
<td>dGuo/dGTP</td>
</tr>
<tr>
<td><strong>Inhibitors</strong></td>
<td>dATP, dGTP, or dCTP</td>
<td>dATP, dGTP</td>
</tr>
</tbody>
</table>

C. ORIGINAL AIM AND GENERAL THINKING ABOUT 

A PUTATIVE dCK GENE IN *L. ACIDOPHILUS* R-26

It is well-known that in prokaryotic organisms such as bacteria that if there is a polypeptide there should be a co-linear gene from which that polypeptide is produced. Because the polypeptide subunits of dAK, dGK, and dCK have been purified as heterodimeric proteins and characterized, but only the dAK/dGK genes from *L. acidophilus* R-26 have been cloned, the next logical step in the deoxynucleoside kinase study should be to clone
and express the third in this family, the dCK gene. However, this has proven to be a very difficult task. When the dAK/dGK genes were cloned, the original goal in our laboratory actually had been to clone the dCK gene. This was successfully accomplished for the tandem dAK/dGK genes, but not for the original target, the dCK gene.

After two years of cloning efforts and examining evidence from many aspects of the study of the possible dCK gene, such as Southern hybridization, direct genomic sequencing, amino acid sequencing, molecular mass comparison and site-directed mutagenesis, I present evidence for the following hypotheses regarding the putative dCK gene:

1. There is no distinct dCK gene in *L. acidophilus* R-26 genome.
2. dCK is produced from the dGK gene.
CHAPTER II

MATERIAL AND METHODS

A. MATERIAL

All the radioactive nucleotides, \(\alpha^{32}\text{P}\) dATP, \(\gamma^{32}\text{P}\) ATP, and \(\alpha^{35}\text{S}\) dATP were from Amershan Corp. or ICN. The radioactive nucleosides, \(^{3}\text{H}\) dAdo, dGuo and dCyd were from Moravek Biochemicals, Inc. The Photogene Nucleic Acid Detection kit and biotin-7-dATP were purchased from Bethesda Research Laboratories (BRL). DNA sequencing kits were from BRL, Epicenter Technology Inc. and United States Biochemical Corp. (USB). BioClean kit for DNA preparation was purchased from USB. The mutagenesis kit (Mutagene) was from Bio-Rad, and the reverse transcription kit was from BRL. Taq polymerase for PCR or RT-PCR was from BRL and USB. The restriction enzymes, T4 ligase and polynucleotide kinase were from Boehringer Mannheim (BM) and BRL. The photogene membranes for hybridization were from BRL. The culture media, such as Lactobacillus broth, and MRS medium were from Difco Corp. All the primers used for sequencing, PCR, and mutagenesis were from the Bio-technology Center at the Ohio State University. Qiagen
columns were purchased from the Qiagen Corp. The seed stock of the
*Lactobacillus acidophilus* R-26 was from American Type Culture
Collection (ATCC- 11506). *E. coli* strains, MV1190 and CJ236 were from
Bio-Rad, and HMS and XL1-blue was from Stratagene. Plasmid
pBluescript (+) KS and pUC-19 was from Stratagene. Reagents for gel
electrophoresis were supplied by Bio-Rad Laboratories, AT-Biochemicals
and Sigma Corp. The Sephacryl S-200 HR for gel filtration was a product
of Pharmacia. PCR instruments were the Perkin Elmer Cetus DNA
Thermal Cycler and the MJ Research PTC-100 Programmable Thermal
Controller. Electroporation was carried out on the Gene Pulse from Bio-
Rad. Certain special material and instruments will be mentioned under the
individual methods in the Methods section.
B. METHODS

1. PCR

The method of polymerase chain reaction (PCR) developed by Saiki et al. (50) involves the use of a set of primers and thermostable DNA polymerase (Taq polymerase) in repeated cycles to amplify the DNA region enclosed by the primers. Some primers were made to be degenerate to ensure that a certain DNA region can be amplified by the PCR. Purification of the primers was made by following the instruction provided by the Biotechnology Center at the Ohio State University. The reaction mixture usually contained 2.5 mM MgCl$_2$, 50 mM KCl, 10 mM Tris-HCl, pH 8.2 at room temperature; 100 μM each of dNTP (dATP, dTTP, dGTP and dCTP); 1.0 unit of Taq polymerase. The components were varied according to the purpose of the PCR, such as adding radioactive material to make a radioactive probe. PCR was carried out in a Perkin Elmer Cetus DNA Thermal Cycler or in an MJ Research PTC-100 Programmable Thermal Controller. Cycle operation usually consisted of three phases: the denaturing phase, the annealing phase and the elongation phase. The length of each phase was varied depending upon the kind, amount and purity of the templates, the purpose of the PCR, and the characteristics of the primers.
At the end of PCR there was usually an extension phase at 70 or 72 °C for about 10 minutes, and the samples were kept at 4 °C afterward.

2. Culturing \textit{L. acidophilus} R-26

The primary cell stock of \textit{L. acidophilus} R-26 was obtained from American Type Culture Collection (ATCC 11506). Then the secondary cell stock was made by 1:100 inoculations in Lactobacillus Broth from Difco Co. Additional cultures were grown either in Lactobacillus Broth or the medium listed below (4). In one liter:

\begin{itemize}
  \item Bactopeptone \hspace{1cm} 20 g
  \item Bacto yeast extract \hspace{1cm} 10 g
  \item Tween 80 \hspace{1cm} 1.5 ml
  \item Sodium citrate \hspace{1cm} 0.25 g
  \item Manganese sulfate \hspace{1cm} 0.002 g
  \item \( \text{KH}_2\text{PO}_4 \) \hspace{1cm} 4.76 g
  \item \( \text{K}_2\text{HPO}_4 \) \hspace{1cm} 6.09 g \hspace{1cm} \text{Both results in pH 6.8}
  \item Dextrose \hspace{1cm} 20 g (autoclaved separately)
\end{itemize}
3. Purification of genomic DNA of \textit{L. acidophilus} R-26

The cells of \textit{L. acidophilus} R-26 were collected by centrifugation (all the centrifugation was done at 4 °C in this method) at 9000 \( \times \) g for 5 minutes. The cells were broken by a Bead-Beater containing glass beads (0.2 mm) with the maximum speed (500 rpm) for 2 minutes and the buffer contained 20 mM Tris, 1 mM EDTA and 10 mM NaCl. The lysate was centrifuged at 8000 \( \times \) g for 10 minutes and the precipitate was discarded. Phenol/chloroform (1:1) was used to extract DNA 3 times or until no white protein boundary could be seen at the interface (being very careful with the solution, avoiding vortexing). Then, 1/10 volume of 3 M sodium acetate (pH 5.2) and 3 volumes of 100 % ethanol were added and the mixture was centrifuged at 10,000 \( \times \) g for 20 minutes. The precipitate was washed once by 80% ethanol and dissolved in water. RNase (DNase free) was added to 10 \( \mu \)g/ml, kept at room temperature for 20 minutes, then extracted once with phenol/chloroform and precipitated with ethanol. The pellet was
washed briefly by 80% ethanol and dissolved in 1 X TE buffer and stored at -20°C. The yield of DNA by this method was very good, and usually several mg of genomic DNA could be obtained from about 500 ml of culture.

4. Plasmid DNA preparation (miniprep)

The preparation of the plasmid DNA for analysis or sequencing was performed by one of two methods. The first was the Qiagen column (Tip 5 or 20) method following the manufacturer’s manual exactly, and from which the quality of the DNA was usually excellent. The second method was a modified alkali method. Bacterial cells were collected by centrifugation and lysed by adding 300 µl of TENS buffer (20 mM Tris, pH 8.0; 0.5 % SDS and 100 mM NaCl). Then they were vortexed for 10 seconds, and 150 µl of 3 M sodium acetate was added, following by vortexing again. After centrifuged for 5 minutes, the precipitate was discarded, and phenol/chloroform was used once to extract lipids and to remove proteins from DNA. Then 900 µl of 100 % ethanol (-20 °C) was added and the mixture was centrifuged for 20 minutes. The pellet was washed once with 70% ethanol and dissolved in water. Then RNase A
(DNase free) was added to a level of 10 μg/ml, and the mixture was kept at room temperature for 20 minutes. Finally, the solution was extracted once with the phenol/chloroform and the DNA was precipitated with ethanol. It was then dissolved in 1X TE buffer, and the plasmid DNA was stored at -20 °C.

5. Competent cell preparation and transformation

In the preparation of the competent cells for transformation by the calcium method, the desired strain of *E. coli* or other bacterial cells were cultured to $A_{600} = 0.8$ and harvested by centrifugation at 9000 x g for 10 minutes. The cells were suspended in Mg** solution (0.2 M of MgCl₂ in DDW), and left on ice for 5 minutes. The cells were collected by centrifugation at 9000 x g for 10 minutes and resuspended in Ca** solution (0.2 M of CaCl₂ in DDW). After 1 hour on ice, the cells were collected by centrifugation and resuspended in Ca** solution containing 15% glycerol, and divided into small aliquots, and stored at -76 °C. In preparation of the competent cells for electroporation, the cells of the desired strain were cultured and collected as above, and then washed at least 3 times with 10%
glycerol in DDW. Then the cells were stored at -76°C. To transform plasmids into the competent cells by the calcium method, several nanograms of DNA were added to 100 µl of competent cells, which were kept on ice for 1 hour. Then they were incubated at 42 °C for 2 minutes without any shaking, and placed on ice for 3 minutes immediately after this heat-shock treatment. After adding 500 µl LB medium and holding at 37 °C for 5 minutes, the desired amount of the cells were streaked onto the appropriate plates. To transform cells with the plasmids by electroporation, the Bio-Rad manual was followed. Usually the calcium method was efficient enough for most purposes, but the efficiency of the electroporation method may be 10 times higher than the transformation by the calcium method.

6. Making genomic libraries of *L. acidophilus* R-26

The genomic fragments on the agarose gel, which corresponded to the bands identified by Southern hybridization, were purified by means of a Bio Clean kit (USB). The plasmid, pBluescript (K/S), was also linearized by the same restriction enzyme as in the Southern hybridization, and the enzyme was removed by phenol/chloroform extraction and ethanol
precipitation after the digestion. Then the linearized plasmid and the
genomic DNA fragments were ligated by T4 ligase. The libraries were
usually stored as ligated plasmids until transformation, since the DNA is
stable after the ligation and is much easier to handle than the colonies on
plates.

7. Preparation of hybridization probes

    a. End-labeling by $[\gamma^{32}\text{P}]$ATP: Synthetic primer was mixed with
10 μCi of $[\gamma^{32}\text{P}]$ATP, 1 μl 10 X kinase buffer and 2 units of T4
polynucleotide kinase(BRL). The final volume was 10 μl, and the mixture
was incubated at 37 °C for 45 minutes. The labeled primer was purified by
ethanol precipitation, and the specific activity of the primer was determined
by liquid scintillation counting. The primer was then used with the desired
template, such as the genomic DNA of L. acidophilus, in PCR to synthesize
the probe. After PCR, the probe was denatured, precipitated with ethanol,
and dissolved in DDW.

    b. Internal labeling by $[\alpha^{35}\text{S}]$-dATP, or biotin-labeled
deoxynucleotides: A pair of primers was used in PCR to amplify a
particular region of DNA, which could be genomic DNA, cDNA or plasmid
DNA. In place of one of the normal deoxynucleoside triphosphates, usually dATP, the desired isotope or biotin-labeled compound was partially substituted so that the isotope or biotin could be incorporated internally into the probe. After PCR, the probe was denatured, precipitated and dissolved in DDW.

8. Southern hybridization

This method was adapted from the one developed by Southern (51, 52). The genomic DNA was digested with restriction enzyme(s) and resolved by electrophoresis on 1% agarose gel. After denaturation, the DNA fragments were transferred to a Photogene membrane (BRL) by capillary action. DNA fixation to the membrane was accomplished by either UV light exposure for 30 seconds or baking one hour at 80 °C under vacuum. Hybridization was carried out in 6X SSC/ 5 X Benhardt’s solution/ 1% SDS at temperatures from 47 to 57 °C, depending on the template and probe, for at least 20 hours. Then the membrane was washed with 0.3 X SSC/0.5 % SDS at room temperature for 5 minutes, rinsed twice with 2 X SSC. Signals were detected by autoradiography on X-ray film (Kodak), either isotopically or photochemically (if the probe was labeled
with biotin). The manufacturer's manual (BRL) was followed in the photochemical detection procedure.

9. Cloning by PCR

The *L. acidophilus* R-26 libraries after ligation were transformed into particular strain of *E. coli* by the calcium method, monitoring with X-Gal. The white colonies were then transferred to master plates prepared with LB medium. Then groups of 20 colonies from the master plates were mixed and cultured together in the liquid LB medium. DNA minipreps were carried out with the mixed contents of 12 culture tubes so that a total 240 colonies were handled every time. The plasmid DNA was then subjected to PCR to amplify the sequence enclosed by the primers. The primers (20-mer) were designed for the DNA region that encodes the first 28 amino acid residues of dCK, and this region is identical to the amino acid sequences of dGK inferred from their DNA sequences. The PCR products were then resolved by 8% polyacrylamide gel electrophoresis, and the positives contained an 84-mer band. The positive and negative controls for PCR were the cloned dAK/dGK genes and DDW respectively. After identifying the positives, the 20 colonies comprising a positive culture were
individually cultured, miniprepped and subjected to PCR to single out positive clone(s). The final confirmation of a positive clone was carried out by DNA sequencing and kinase activity assays.

10. Cloning by colony hybridization

The *L. acidophilus* R-26 libraries were transformed to particular strains of *E. coli* by the calcium method and were selected by means of X-Gal. The white colonies were transferred to the master plates with LB medium. After 15 hours of incubation at 37 °C, the colonies were lifted from the master plates onto Photogene membrane (BRL), and the master plates were incubated again for eight hours to restore the colonies, and the plates then were stored at 4 °C. The colonies on membrane were lysed and the DNA was denatured with 1.5 M NaOH for 5 minutes and renatured with 1 M Tris, pH 8.0. The DNA then was fixed to the membrane by baking for 1 hour at 80 °C under vacuum, after the cell debris was washed off with 2 X SSC. A probe generated by PCR was internally labeled by [α-35S]-dATP and hybridization took place in a sealed hybridization bag; usually there were two membranes per bag (membranes were positioned back-to-back). After hybridization and washing, the membranes were applied to X-ray
film, and the positive and negative controls in this experiment were the cloned dAK/dGK genes and the original pBluescript vector alone, respectively. The final identification of positive (s) was carried out by sequencing and kinase activity assays.

11. Normal DNA sequencing

The method developed by Sanger (35) was used. The sequencing template could be plasmid, linearized plasmids, PCR products, or single-stranded DNA. The Sequenase Version II kit (USB) was used in this procedure. The first step was the alkali denaturation of the template (except the single-stranded DNA). After recovery of the template (about 3 μg in DDW), the desired primer (1 pmol) was annealed to the template by heating (up to 75 °C) for 3 minutes, followed by cooling down to room temperature over a period of 40 minutes. The synthesis reaction mix, which contained the desired isotopically-labeled nucleotide, generally [α-35S]dATP, along with dGTP, dCTP, dTTP, DTT and DNA polymerase, was added to the annealing mix. After vortexing, the mixture was divided into four tubes each containing didoxynucleotide termination mix, and they were immediately incubated for 5 minutes at room temperature and 5 minutes at
42 °C. Stop solution was added to each tube, mixed well and applied to the sequencing polyacrylamide gel which was prepared from Long Ranger sequencing gel solution (AT Biochem). The polyacrylamide concentration in sequencing gel usually depending on the DNA region to be read, was from 5 to 8%. After running, the gel was dried by using a gel dryer (Bio-Rad, Model 583), and subjected to autoradiography on X-ray film (Kodak X-omat AR).

12. Cycle DNA sequencing

Thermocycle Sequencing kit from Epicenter Technology (EP) was used in this method, with some modifications. The templates could be the plasmids, linearized plasmids, PCR products, fragmented genomic DNA of *L. acidophilus* R-26, or intact genomic DNA. The purity of the template preparation was very critical in this method, so the Qiagen columns (Qiagen Scientific) and BioClean kit (USB) were used. The amount of the primer was usually double (up to 10 pmol/sample) that recommended in the manual, and the number of cycles was usually more than 30 (generally from 40 to 50). The sequencing-gel preparation and running conditions were the same as the normal didoxynucleotide sequencing procedure mentioned
previously, except the volumes of the loaded samples were usually 5 μl instead of 3 μl.

13. Design and synthesis of the oligonucleotide primers

The first concern in primer design was to find a region that was suitable for the experimental purpose. However, sometimes there was not much choice. Typically, a good region should have a high GC/AT ratio, no repeats, no secondary structures and a relatively high annealing-temperature. The second point of concern was the length of the primers. For PCR and sequencing, the primers were usually about 20 nucleotides long. For mutagenesis, especially for deletion or insertion, length of the primers was usually more than 30 nucleotides (15 nucleotides or more on each side of the mutation). The last consideration in primer design was the 3' end of the primers. A good 3' end should have a 100% match with the template for at least 3 residues, and the last nucleotide should be either C or G. Following designing with the aid of the ‘Primer Designer” program, all the oligonucleotides were made by the Biotechnology Center at the Ohio State University, and the manual for the purification of the primers, provided by the Center, was followed. The primers for mutagenesis were
further phosphorylated by polynucleotide kinase (BRL), and the reaction conditions were set as described by the manufacturer's manual. All the oligomers we used were listed in Table 9.

14. Extraction of enzymes from *E. coli* or *L. acidophilus R-26*

*E. coli* cells were cultured in LB medium and *L. acidophilus* cells were grown in either Lactobacillus Broth (Difco) or home-made broth (method 1). The cells were collected at an *A*₆₀₀ of about 1.3 to 1.5, and lysed by glass beads with a Bead Beater for small scale preparation (usually under 20 ml) or lysed by ultrasonication (Cell Disruptor 350 from Branson Sonic Power Co. at following setting: power = 6, output = 20%, 15 seconds) for larger scale preparation (usually from 20 ml to several liters). The cell suspension buffer usually contained 100 mM Tris, pH 8.0 and 40 mM NaCl in DDW. The cell debris was removed by centrifugation. The crude extraction could be loaded onto a particular chromatography column for purification, subjected to the ammonium sulfate fractionation, or used directly in the enzyme activity assays. The 30 - 65% cut of the ammonium sulfate fractionation contained the deoxynucleoside kinases.
15. **Enzyme activity assay**

The deoxynucleoside kinase assays were developed by Dr. Ives (37). The only modification was that the final assay volume was 40 µl instead of 80 µl. The final assay conditions are as follows: ATP 10 mM; MgCl₂ 12 mM; Tris-HCl (pH 8.0) 0.1 mM; deoxynucleoside (dAdo, dGuo or dCyd) 20 µM; [³H] deoxynucleoside (dAdo, dGuo or dCyd) 0.2 mM. The assay was started by adding the enzyme to the mixture. After incubating 30 minutes at 20 °C, 100 µl of formic acid (0.1 M) was added to stop the reaction. Then a 20-µl aliquot was spotted on a square of Whatmen DE-81 anion exchange paper, measuring 1 cm by 1 cm, and the papers were washed free of un-reacted nucleoside by immersion for 30 minutes in a water bath recycled through a charcoal cartridge. The radioactivity remaining on the papers was counted in a liquid scintillation counter.

Protein concentrations were determined by the protein assay described below, and specific activity was defined as nanomoles of product found per minute per mg of protein.
16. Protein assay

Amounts of protein were determined by the Mini Protein Assay method, developed by Bio-Rad, a modification of the Bradford Method (53). Concentrated Bio-Rad dye reagent (0.2 ml) was added to 0.8 ml protein solution, and mixed well. After 5 minutes at room temperature, the absorbance at 595 nM was measured, and the protein concentration was determined by interpolation from a standard curve, using bovine serum albumin as the standard protein.

17. Steady state kinetics

The study of the enzyme kinetic mechanisms used the fixed-time assay (37) in the presence or absence of end-products which mimic multi-substrate analogs. Since each enzyme reaction involved two substrates, $K_m$ values for each enzyme were measured by fixing one substrate at several concentration and varying the other substrate concentration. The range of the substrate concentrations varied according to each enzyme and the preliminary estimation of its $K_m$ values. In some cases the active sites on
subunits, especially for some homodimers, exhibited cooperative effects
and apparent $K_m$ values were estimated from the Michaelis-Menten plots.

18. Protein electrophoresis on SDS polyacrylamide gel

SDS-gel electrophoresis was carried out according to the method
developed by Laemmli (54). The stacking gel was 4% polyacrylamide
buffered at pH 6.8, and the separating gel was 12% polyacrylamide,
buffered at pH 8.8. Protein samples were mixed with SDS sample buffer at
4:1 ratio, boiled at 100 °C for four minutes, and then 1 μl of 1%
bromophenol blue (BPB) was added. The electrophoresis running buffer
was prepared as 5X stock solution by dissolving 4.5 g Tris base, 21.6 g
glycine and 1.5 g SDS in 100 ml of DDW, and diluted to the proper
concentration before use. After running, the SDS gel was stained and fixed
by soaking for 30 minutes in the staining solution consisting of 40%
methanol, 10% acetic acid and 0.1% Coomassie Blue (Sigma) R-250.
Destining was carried out by soaking the gel in the destining solution (10%
methanol and 7.5% acetic acid) until background color was gone. After the
destining, the SDS gel was dried in a gel-drying frame between two pieces
of cellophane membrane from Novex.
19. Gel filtration chromatography

Sephacryl S-200 HR resin from Sigma was used in the gel filtration chromatography. To prepare the column, the fine particles were carefully removed after overnight soaking in DDW, and the gel beads were packed into the column (1.5 X 100 cm), avoiding air bubbles. When the gel material settled, the column was washed with buffer, 20 mM Tris, pH 8.0, 10 mM NaCl, and 5% glycerol (optional), and calibrated. The sample size was usually less than 0.5 ml and column flow rate was kept constant by using a peristaltic pump at speed of 25 ml/hr. The fractions were collected by a fraction collector and each fraction size was usually about 2.5 ml. The exact position of each fraction was monitored on a strip chart recorder. The void volume was monitored by tracing blue dextran through the column, and the standard curve for molecular weight was obtained by monitoring the eluting positions of four common proteins: BSA, ovalbumin, carbonic anhydrase, and myoglobin (Sigma), with a UV monitor which was attached to the strip chart recorder. The correlation coefficient of the linear regression on the standard curve was 0.997. Each of the samples was run in
exactly same fashion and at the same flow rate, and each molecular weight was calculated by interpolation from the standard curve.

20. Total RNA purification

Bacterial cells were cultured to an $A_{600}$ of 0.9 in liquid medium and collected by centrifugation at 9000 x g for 5 minutes. The cells were washed and resuspended in the buffer containing 50 mM Tris, pH 8.0, and 10 mM NaCl. The buffer was prepared by using DEPC-treated water and every container used was also washed by DEPC-treated water. The cells were lysed by extraction with phenol/chloroform at volume ratio of 1 to 1. After vortexing for 15 seconds, the mixture was centrifuged for 5 minutes and the supernatant was placed into new tubes. The total polynucleotide (DNA and RNA) fraction was precipitated with ethanol and 0.1 volume of 3 M sodium acetate, pH 5.3. The pellet was dissolved in DNase digestion buffer, RNase free DNase (BM) was added to the nucleotide solution and the tubes were incubated at 37 °C for 50 minutes to remove DNAs. The phenol/chloroform extraction and ethanol precipitation were performed twice to remove DNase and some fragmented DNAs. Then the total RNA
was dissolved in DEPC water containing 5 mM EDTA and 10 mM Tris, pH 8.0, and stored at -20 °C.

21. Northern Hybridization

The protocol listed in Sanbrook’s manual of Molecular Genetics (51) was followed with a few changes. The denatured agarose gel was 1.4% instead of 1.0%. After briefly rinsing briefly in DEPC-treated water twice, the RNA blotting was carried out by using the Turbo Blotter from Schleicher & Schuell, with a blotting solution of 3 M NaCl, and 8 mM NaOH. Then the membrane was baked at 80 °C under vacuum for one hour to fix the RNAs, and hybridized with a probe made by PCR and labeled with [α-32P] dATP. The probe used here was the dCK/dGK probe, the same as the one used in the Southern hybridization. The hybridization conditions and the steps in washing were also the same as for the Southern hybridization.

22. RT-PCR

The SuperScript kit from BRL was used in this procedure, with a few modifications. Four μl of the total RNA purified from various sources
was mixed with 1 μl of the desired primer, and the mixture was incubated at 75 °C for 10 minutes and put on ice immediately for annealing. Then 1 μl DTT (50 mM), 2 μl of the first incubation buffer, and 2 μl dNTP (1 mM each) were added to the mixture. The reaction tubes were preheated at 42 °C for 2 minutes and 1 μl (5 units) SuperScript II was added and mixed well by pipetting. After incubating at 42 °C for 50 minutes, 10 μl TE buffer was added to stop the reaction. The normal PCR procedure (Method 1) was carried out for each cDNA produced from the reverse transcription with desired isotope, and the volume of the RT mixture was no more than 1/20 of the final PCR volume. The PCR bands were resolved by a denaturing polyacrylamide gel (12%) or polyacrylamide DNA sequencing gel (5%).

23. Site-directed mutagenesis

The materials and protocols of the Mutagene kit (version II) from Bio-Rad were used. After phosphorylation by polynucleotide kinase, the primer with the desired mutation was annealed to the single-stranded template (pBluescript KS(+) with the dAK/dGK gene insert). After polymerization and ligation, the closed circular double stranded DNA was
transformed to *E. coli* strain MV1190 by the calcium method. The transformants were grown on LB plates containing ampicillin, and the mutations were confirmed by DNA sequencing, using the dideoxynucleotide termination method (35) and the Sequenase Version II kit.
CHAPTER III

EVIDENCE THAT THERE IS NO DISTINCT GENE FOR dCK IN THE GENOME OF L. ACIDOPHILUS R-26

The first step towards the cloning of the dCK gene was to design the experiment. Among the three deoxynucleoside kinases, two of their substrates contain purines (dAdo and dGuo). Since the genes for dAK and dGK have been cloned in our lab, the similarities and the differences between those two genes and between the amino acid sequences have been analyzed. Therefore, when we tried to clone the putative third gene, dCK gene, we were expecting that its sequence would differ from those of either dAK or dGK, since the substrate of dCK is deoxycytidine, a pyrimidine, which should have a different binding site than the purine deoxynucleosides. Further, it seemed likely that the putative dCK gene could be associated with the dAK gene in the same way as the dGK gene.
did since these three enzymes are found in nature in the heterodimeric forms, dAK/dGK and dAK/dCK.

Here are the general ideas on which we based our cloning experimental design:

1) The first task for cloning dCK gene is to choose a region on the dAK/dGK genes to make a probe for colony hybridization. Since in *L. acidophilus* R-26 only dAK/dGK and dAK/dCK heterodimers were reported (apparently there are also smaller concentrations of dGK and dCK monomers, dGK/dGK and dCK/dCK homodimers, and dAK/dGK and dAK/dCK trimers or tetramers, data not shown), and the dGK gene was associated with the dAK gene, it was quite likely that the dCK gene would be associated with a dAK gene as well. Therefore, a dAK probe (375 base-pairs) might be a good candidate for cloning dCK gene by the colony hybridization method. It is obvious that using some regions on the dGK gene to make probe is not proper because dGK is a purine enzyme and dCK is a pyrimidine enzyme, so their genes could be very different, therefore.
2) Because the first 28 amino acid residues (84 base-pairs of DNA) of dCK were sequenced (Figure 2), and are identical to those of dGK, and the DNA sequence between the N-terminal portions of dGK and dAK genes differs in only one base-pair, whereas overall identity between dAK and dGK is 78% (21/27 amino acid residues), it was reasonable to speculate that the DNA sequences of dCK and dGK at that region should be very similar or even the same. Therefore, using this region for PCR cloning could be another reasonable path to the putative dCK gene cloning. Because the region was only 84 base-pairs long, it might not be as reliable for colony hybridization as for PCR cloning.

3) There was some doubts about whether a distinct dCK gene exists in the LBA genome even before we started the cloning experiment. First of all, the Southern hybridization of the genomic DNA of *L. acidophilus* R-26, conducted by G. Ma (1), revealed only one band for each restriction enzyme digestion, which could be further made into libraries, after she tested 6 common restriction enzymes. Grace Ma’s Southern hybridization results are shown in Figure 4. The double bands at EcoRI lane could be explained perfectly by the DNA sequence of the cloned dAK/dGK genes. Therefore,
the possible explanations for the one-band or one-region phenomenon were that either her probe, which extended between codons for the first 28 amino acids of dCK to the space sequence between the dAK and dGK genes, could not hybridize with the dCK gene under her experimental conditions, or there is no such a gene in the LBA genome. Since we knew at that time that the dGK and dCK shared the same stretch of the 28 amino acid residues at the N-termini and that the DNA sequences of dAK and dGK were very similar (83/84 base pairs), the first explanation seemed quite questionable, and the idea that there is no distinct dCK gene in the LBA genome might be true.

Although we had those doubt about the existence of a dCK gene, we still started the experiment as planned. However, after two years of intensive searching and studying, we have accumulated many pieces of solid evidence that supported our hypothesis that there is really no distinctive dCK gene on the *L. acidophilus* R-26 genome.
A. THE STATISTICAL IMPROBABILITY OF A DISTINCT
dCK GENE IN L. ACIDOPHILUS R-26 GENOME

Two cloning methods (PCR cloning and Colony Hybridization
cloning) were used. More than 10 libraries and 100,000 colonies were
checked. Since the sizes of the cloned fragments of the genomic DNA were
from 3 kb to 5 kb, the following equation:

\[ N = \frac{\ln (1 - P)}{\ln (1 - f)} \]

\( N = \) colony number
\( P = \) probability expected, such as 0.99
\( f = \) size of inserts / size of genomic DNA

predicts that the putative dCK gene should be present at least once for every
5,000 to 8,000 colonies. However, only dAK/dGK genes were encountered
(a total of 8 times) in these experiments, which was about as predicted by
the equation above for a random event. Since the number of the colonies
examined surpassed the expectation for the putative dCK gene many times,
we had to think that the appearance of a positive dCK colony might not be a
random event, or that the distinct dCK gene might not exist in our libraries
after all.
B. FURTHER SOUTHERN HYBRIDIZATION EVIDENCE

Southern hybridization is a powerful tool for detecting the existence of a specific sequence within genomic DNA. We tried to determine whether there is more than one region on the LBA genome which can hybridize with our probes. The whole idea is that if there are two or more binding regions for the probes on the genomic DNA, the hybridization results after a specific restriction enzyme digestion should show a pattern with more than one band, simply because the enzyme would likely cut somewhere between the potential regions and the areas would be resolved on the agarose gel. If there was only one region on the genome to which the probe could bind, there should be only one band for each digestion.

It is also possible that the potential binding regions may be so close to each other that they show up on the same band, or the two region are not too close but there is no enzyme cutting site between them so that two binding regions show up on a single band. First, I used a simple PCR distance measurement test to check if closing-by there is another region that is similar to the cloned dAK/dGK genes on the LBA genome. The idea is
very simple: when a pair of “back-to-back” primers for PCR is used, and if there are two similar areas on the DNA template within 10 kb from each other, PCR will amplify the area between them and a resolving gel will define the distance between the two regions. In our experiment, however, when two pairs of “back-to-back” primers were used for dAK and dGK/dCK (since the primers covered the known amino acid residues common to dGK and dCK) and *L. acidophilus* R-26 genomic DNA was taken as the template, nothing was amplified by PCR, which means either there is only one dAK gene and one dGK gene in the genome, or the two regions were too far apart to be amplified by PCR. This experiment was repeated many times for each pair of primers and the results were the same. This experiment was repeated multiple times because if the dCK gene was nearby, this method could reveal it instantly and the PCR product could be cloned or sequenced. Therefore, we concluded that either there is only one region which includes the cloned dAK/dGK genes, or that the other potentially probe-binding regions are more than 10 kb away from the dAK/dGK genes on the LBA genome.
As for the absence of a cutting-site between two regions, a site might be found by trying several other kinds of restriction enzymes. About 20 different restriction enzymes were tried, but none of them gave any pattern that could not be explained by the sequence of the cloned dAK/dGK gene. It seemed improbable there isn’t even one different cutting site between the dAK/dGK genes and the target, the putative dCK gene. Moreover, dAK/dGK sequence has a pair of Kpn I sites flanking these two genes (by which the dAK/dGK genes were cloned). However, the Southern hybridization pattern for Kpn I showed only one band, which was consistent
with Grace Ma's results (Figure 6). These observations together show there is only one sequence which could bind to our probes in the LBA genome. It is true that some results of Southern hybridization did show some double bands, even triple bands, such as for *EcoRI* (triple) and *Bcl I* (double), but in each case, they could be explained perfectly by the known DNA sequence of the cloned dAK/dGK genes (having restriction enzyme cutting sites within the dAK/dGK genes) (Figure 11).

1. Hybridization with a dAK probe

   The reason for using a dAK probe was described above. The dAK probe was 375 base-pairs long, within the reading frame of the dAK gene (Figure 7)
Figure 6. Southern hybridization analysis of restriction digested genomic DNA from LBA. Probe = biotinylated 117 mer PCR fragment. Hybridization at 42°C; 0.1 X SSC, 1% SDS wash at 60°C. Restriction enzymes used are: lane 1, Kpn I; lane 2, Sac I; lane 3, DNA markers (λ/Hind III); lane 4, Xba I; lane 5, EcoR I; lane 6, DNA marker (λ/Hind III/EcoR I); lane 7, Hind III; lane 8, Pst I. Figure from Grace Ma (1), used with permission from author. See restriction map on Figure 11.
The dAK probe was made by PCR. The 5' primer was shared by the dGK gene, but the 3' primer had only 40% homology with the dGK gene. Therefore the PCR should only amplify the dAK region under our PCR reaction conditions. The annealing temperature for PCR was set at 55 °C. The dAK probe was labeled internally by biotin-labeled dATP.

Figure 7. dAK probe, 375 bps, from 345 to 699 bp on the dAK gene, made by PCR and internally labeled by $^{35}$S or biotin. P = promotor; T = terminator.
Hybridization with the dAK probe was carried out at 53 °C, and results are shown in Figure 8. For the dAK probe, many kinds of restriction enzymes or their combinations were tested and results shown here are only the representatives. All of them shown the patterns to be expected from the known dAK/dGK genes, strongly indicating that there is only one region that can hybridize with the dAK probe. Since we have demonstrated that there is no dAK-like region within 10 kb of the known dAK/dGK genes, the Southern hybridization results show that the possibility of another dAK gene in addition to the cloned dAK gene in the LBA genome is very remote. It is very unlikely that the putative dAK gene associated with the dCK gene has such a different sequence from the cloned dAK gene that it can not be detected by the dAK probe, since there is no indication of the existence of a different kind of dAK in LBA. It is also entirely possible, of course, that the dCK gene is not associated with a dAK gene at all. In that case we would have to find a probe which can detect the putative dCK gene itself.
Figure 8. Southern hybridization results with the dAK probe labeled with biotin. Lanes: 1 = EcoRI; 2 = EcoRI/PstI; 3 = XbaI/PstI; 4 = HindIII/PstI; 5 = XbaI/ApaI; 6 = XbaI/HindIII. Check restriction map on Figure 11.
2. Southern hybridization with the dGK/dCK probes

When we started the cloning work, we only knew the first 28 amino acid residues of dCK, which was too short for making an effective probe. When the question as to whether there is a distinct dCK gene in the genome was raised, more peptides of the dCK were sequenced, which will be mentioned in detail later. The N-terminus was resequenced and the C-terminus of dCK was also sequenced and found to be exactly the same as that of the dGK. Therefore, we had all the information to make a dGK/dCK probe by PCR. The 5' and 3' primers were designed in a way so that they should anneal not only to the dGK gene but also to the dCK gene on the genomic DNA, and the two primers covered entire the dGK gene (675 bp).

5' primer (12-8-5'): 5'-ATGACAGTTATTGTATTAAGCG-3'

This primer was designed for the first 22 nucleotides of the dGK gene and the primer sequence was also shared with the dAK gene. Since the amino acid sequences of dAK, dGK and dCK were identical at this region, and the DNA sequences of dAK and dGK genes were identical as well, it is very
likely that the DNA sequence of the primer is also shared by the putative dCK gene (if there is a distinct dCK gene). Therefore, this primer should anneal to the dAK and dGK genes and the putative dCK gene. The annealing temperature of the primer 12-8-5' is about 67 °C according to the analysis from Primer program (this primer, but not the 3' primer should anneal to the dAK gene too, but in order to simplify the Southern hybridization pattern, we shortened the PCR extension time so that almost no dAK gene was amplified).

3' primer (p121694): 5'-AACTAAATTAGGTTTCCAGTCGTTCTTC-3'
   (A) (A) (G)

The sequence of the 3' primer (p121694) was based upon the 3' end of dGK gene (the extended “tail”, amino acid sequence = 5'-EDDWKPNLKV, which is not shared by dAK, Figure 4). As already mentioned, the C-terminal amino acid sequence of dCK is identical to that of dGK. In order to increase the likelihood that this primer would anneal to the putative dCK gene, some degeneracy at certain positions on the primer was included. Since the first (E), second (D) and third (D) residues have two possible codons each, with variation only at the third base pair, only \( \frac{1}{2} \) of degeneracy was used on each third base pair. The fourth residue (W) has
only one codon, so there can be no degeneracy there. So far, we have covered four amino acid residues and 12 base pairs. If we also consider that the first two base pairs of the fifth residue (K), which has no codon variation, 14 base-pairs are covered so far. Since a 15-oligomer is long enough for many PCR applications, this primer should anneal to the putative dCK gene already. However, in order to ensure the annealing to the putative dCK gene, 13 additional nucleotides were added (without further degeneracy). The annealing temperature of the primer p121694 (27 mer) is about 73 °C according the analysis of the program Primer (the GC content of the first 14 nucleotides is 50%) and the annealing-temperature of the PCR reaction was set at 47 °C, a 26 °C difference. Therefore, we can safely anticipate that primer p121694 should anneal to the putative dCK gene and amplify it by PCR.
Figure 9. dGK/dCK probe; 675 bps from the beginning to the end of the dGK gene. Made by PCR and labeled with $^{35}$S. P = promotor; T = terminator
The Southern hybridization results by using the dGK/dCK probe are shown in Figure 10, which is only part of the hybridization results we got in the past experiments. Six different restriction enzymes were used, and the hybridization temperature was 53 °C. These results are similar to the Southern hybridization by the dAK probe (Figure 6, and 8), and only one band was found for each lane, including Kpn I digestion. This implies that there is only one region on the *L. acidophilus* R-26 genome with which the dGK/dCK probes can hybridize. Those instances where more than one band was observed can be explained by known restriction sites within the dAK/dGK DNA sequences (Figure 11).

The tentative conclusion drawn from the Southern hybridization results was that there is only one region, namely the tandem dAK/dGK genes, on the genomic DNA of the *L. acidophilus* R-26 which can hybridize with either the dAK probe or the dGK/dCK probes.
Figure 10. Southern hybridization with dCK/dGK probe, labeled with $^{35}$S. Lanes: 1 = cloned dAK/dGK genes in pBluescript; 2 = EcoR I; 3 = Xba I; 4 = Hind III; 5 = Kpn I; 6 = Apa I; 7 = Bcl I. Molecular weight markers = $\lambda$/Hind III. Also check the gene map of the cloned dAK/dGK (Figure 11)
Figure 11. Restriction map for the cloned dAK/dGK genes. $A = Apa I$; $E = EcoR I$; $X = Xba I$; $B = Bcl I$. There are no cutting sites for $Hind III$, or $Pst I$. 

$dAK$  

$dGK$
C. EVIDENCE FROM DIRECT GENOMIC SEQUENCING

One of the most direct and convincing methods in detecting the putative dCK gene is to sequence the genomic DNA directly. Recently some methods for sequencing large stretches of DNA by cycle sequencing techniques have been reported, such as for lambda DNA (36). Nevertheless, so far no one has been able to “directly” sequence the genomic DNA of any organism because of the complicity of the DNA. Here, I present some improvement on the direct genomic DNA sequencing techniques and the results of the sequencing of the *L. acidophilus* R-26 genome by using modified cycle sequencing technology.

The term, direct genomic DNA sequencing, is sometimes confusing because it could indicate at least three different methods:

1) knowing some information about genomic DNA from cDNA or amino acid sequence,
1) ▼ using this information to amplify the desired region on the genomic DNA by PCR (two primers),
▼ using normal sequencing methods to sequence the PCR products (PCR product = template).

2) ▼ knowing some information about genomic DNA from cDNA or amino acid sequence,
▼ using this information to amplify the region of the genomic DNA by PCR (two primers),
▼ using cycle sequencing methods to sequence the PCR products (PCR product = template).

3) ▼ knowing only enough information about the genomic DNA sequence to make a single primer (at least 20 base-pairs or 7 amino acid residues),
▼ using cycle sequencing methods and a SINGLE primer to DIRECTLY sequence the whole genomic DNA or genomic DNA fragments isolated from agarose gel of a Southern hybridization (genomic DNA = template).
So far no one has yet been able to perform the third method very well. It is easy to see the usefulness of the third method, since despite knowing very little about a genomic DNA, it allows you to sequence into unknown areas, making a gene search much easier. I used the second and third methods extensively to do genomic sequencing of LBA, whose total size is only 4,000 kb. The third method is the REAL genomic DNA sequencing method which means that the template is THE genomic DNA (not fragmented genomic DNA or its PCR product), and only one primer is involved.

1. Direct genomic sequencing with the whole genomic DNA of *L. acidophilus* R-26

Genomic DNA of *L. acidophilus* R-26 was purified with complete removal of RNAs as described in Methods and Material. The Cycle Sequencing kit was purchased from BRL. The genomic DNA was denatured by the alkali method and primer ("7", made by Grace Ma) was labeled at the 5' end by [γ-32P] ATP and polynucleotide kinase from BRL. The primer (+7) is a 20-mer oligonucleotide, designed from the 466th base-
pair to the 485th base-pair on the dGK gene. The annealing temperature of this primer was about 60 °C. In this region, the amino acid residues from dGK and dCK are identical, and the DNA sequence of dGK is 70% homologous with that of the dAK gene, (however, the last 3 base-pairs at the 3' end of the primer were totally different from that of the dAK gene, which usually makes the amplification impossible to initiate). The degeneracies of the possible codons for the amino acid residues (from 3': KYYDELT) in this region are 2, 2, 2, 2, 2, 2, 2 (the only variation being at the third base-pair), and 4 respectively. Since the annealing temperature in the cycle sequencing was 45 °C (a 15 °C difference), it seems likely that this primer could anneal to both the dGK gene and the putative dCK gene. The sequencing results are shown in Figure 11. The control was the cloned dAK/dGK genes in the pBluescript vector. This method obviously works well, and only the dGK gene sequence was revealed by this method (sequence can be read clearly from the 562th base-pair up to the end of the coding region) both for the plasmid control and the genomic DNA. Although the amino acid sequences of dGK and dCK are the same in the primer region, the possibility that +7 might not be able to anneal to the putative dCK gene could not be
Figure 12. Result of direct genomic sequencing with the whole genomic DNA. A. Sequence from cloned dAK/dGK genes; B. Sequence from whole genomic DNA of LBA. Sequences are the same (from bottom up, and lanes = ATGC): cttgattcatggcaagtgaagaagatcgtcaagaagtcctaaatcaaattgtggctaa
gctcaagaatggtaaacttgtagacgactggaaacttaatttagttaaat........
eliminated totally, since no degeneracy was provided in the primer, +7 (a primer with any degree of degeneracy failed to work well in this method). Several other primers, such as +6 and 10-6, were also tried, and the results were similar.

2. Direct genomic sequencing of fragmented genomic DNA of *L. acidophilus* R-26

Because we know the Southern hybridization patterns for many restriction enzymes, linearized genomic DNA was electrophoresed and specific bands were purified by the glass powder method ("Bio-clean", USB), and subjected to direct-genomic sequencing. The fragmented genomic DNA was denatured and the Cycle Sequencing kit from Epicenter Technology was used in this experiment. The isotope used was [α-35S]dATP for internal labeling. Many bands were purified and sequenced, and the result shown in Figure 12 was the fragment from Hind III digestion. The primer was 4-28-R, a 22 mer with an annealing temperature of 66 °C, designed from the 416th bp to 395th bp (reverse) of the dAK gene, without
degeneracy (again, primers with any degeneracy did not work well). Under the annealing condition of cycle sequencing (52 °C, 14 °C lower than primer’s annealing temperature), this primer should pick up the dAK gene and any dAK-gene-like region on the genome, but not dGK or dCK since the homology between dAK and dGK in this area was only 62% and the 3' ends were totally different. The purpose of this sequencing was to see if there is another dAK gene in the LBA genome, which could be associated with the putative dCK gene and might be very similar to the cloned dAK gene but different at certain positions. From these results, only the sequence of the cloned dAK gene was revealed by this method. Therefore, considering results of the Southern hybridization with the dAK probe, it appears to be very unlikely that there is another dAK gene which is either the same as, or a little different from, the cloned dAK gene in the LBA genome.
Figure 12. Direct genomic sequencing with the fragmented genomic DNA of LBA. Lanes: A = genomic DNA; B = cloned dAK/dGK genes. Sequences are the same (from bottom, lanes = TACG):
3'-gtgtgacaatcgcgtctgtatcgaaccaattaaactcgtgttgccaccaaatctttctttcatattttctttacgtgtggtggtggtttaaccaatagatgtggag ......
3. Direct sequencing of PCR products (dGK/dCK probes) of genomic DNA of *L. acidophilus* R-26

In the section on the Southern hybridization by the dCK/dGK probe in Chapter II, it was explained why the dCK/dGK probe used there should contain sequences from both dGK and dCK. Assuming that primer p121694 could anneal to copies of both the dGK gene and the putative dCK gene (for detailed argument, see Southern hybridization section), the dCK/dGK probe, the PCR products, should contain not only the sequence of the dGK gene but also that of the putative dCK fragment. Therefore, direct sequencing of genomic DNA with the dCK/dGK probe as the template will be desirable to further support the result shown by the Southern hybridization with the dCK/dGK probe.

The sequencing template preparation was the same as was used for the preparation of the dGK/dCK probe from the genomic DNA with some modifications: 1) the same 5' primer (12-8-5') (which could anneal to the dAK gene, the dGK gene and the putative dCK gene) with much longer
PCR extension time to ensure the inclusion of the dAK fragment in PCR as control; 2) without any radioactive material but using the normal concentration of dNTP. Since the primer 12-8-5' could anneal to the dAK gene, the dGK gene and the putative dCK gene, and the primer p121694 could anneal to dGK gene and dCK gene, the PCR products might be a mixture of copies of fragments from the  

1) dGK gene only  
2) dAK/dGK genes  
3) dCK gene only and  
4) dAK/dCK genes  

The PCR products were resolved by 8% polyacrylamide gel electrophoresis, and there were only two bands on the gel. These two products were each isolated by Bio-clean kit (USB). If there is a distinct dCK gene and it can be annealed to primer p121694, the larger PCR product should reflect both the genes of dAK/dGK and dAK/dCK, and the smaller product should contain the fragment of dGK and dCK. If there is no distinct dCK gene, or if the dCK gene could not be annealed by
p121694, which is very unlikely, the larger band should contain dAK/dGK genes and the smaller one should have the dGK gene only.

Cycle sequencing was performed on those two bands with the 5' primer 12-8-5' which could anneal to the dAK gene, the dGK gene and the putative dCK gene, and the following results should be expected:

For the larger PCR product:
1) dAK/dGK genes: some double bases at certain positions
2) dAK/dGK and dAK/dCK genes: some double, triple, and quadruple bases at certain positions

For the smaller PCR product:
1) dGK gene only: only single base at all positions
2) dGK and dCK gene: some double bases at certain positions

The sequencing results are shown in Figure 13. It was clear that for the larger PCR product there were single and double bases, which are all at the expected positions, but absolutely no triple or quadruple bases, and for the smaller PCR product there were only single bases. This strongly suggested that no distinct dCK fragment was amplified, even though the
primers should anneal to both the dGK gene and the putative dCK gene, thus indicating that there is no distinct dCK gene on the genomic DNA of the *L. acidophilus* R-26, (unless the primer p121694 failed to pick up the distinct dCK gene, which, as has been explained before, is very unlikely).
Figure 14. Sequencing of dCK/dGK probe made by PCR. Lane = ATGC. A = larger band on PCR gel, and only dAK/dGK genes have been revealed; B = smaller band on PCR gel, and only dGK gene has been revealed.
In this chapter so far we have examined the negative evidence for a unique dCK gene obtained from the cloning attempts, Southern hybridizations, and direct genomic sequencing. Based on these lines of evidence, it is quite obvious that the possibility that there is a distinct dCK gene in the genomic DNA of \textit{L. acidophilus} R-26 is very remote. However, some reservations are still in order: for example, whether or not the primers used in making the probes and for direct genomic sequencing could anneal to the putative dCK gene, since the whole Southern hybridization and genomic sequencing experiments are based on the assumption that those primers will anneal the putative dCK gene. Another reasonable observation is that all the evidence so far is negative, and it is well known that it is always difficult to prove a concept (as in there being no distinct dCK gene in the LBA genome) by using negative evidence. Therefore, I had to seek more evidence from other directions.
D. PEPTIDE SEQUENCE COMPARISON OF WILD TYPE
dAK, dGK AND dCK REVEALS THAT dGK AND dCK ARE
TOO SIMILAR TO BE FROM SEPARATE GENES

So far, all the indications have pointed to there being no distinct dCK
gene in the LBA genome. To further confirm these findings, we used
another powerful and direct tool to prove that theory, the peptide
sequencing of dCK. If we could find even one residue in dCK that is not
encoded in the cloned dAK/dGK genes, our hypothesis that there is no
distinct dCK gene in the genome of the *L. acidophilus* R-26 could be
proved wrong. The peptide sequencing was performed by Dr. Ikeda in our
lab (36), and his data are presented here with his permission.

The dCK/dAK heterodimeric protein was purified from *L.
acidophilus* R-26 by dCTP affinity chromatography, developed by Ikeda et
al (9). Then the subunits dAK and dCK were separated by HPLC
(Waters/Millipore) on a C-4 column. After confirming the purity and
subunit separation by SDS polyacrylamide gel electrophoresis, the dCK was
digested by trypsin and the peptides were separated by HPLC, using a C-18 column. The sequencing of each peak was performed by the Biochemical Instrument Center of the Ohio State University. The sequencing results are shown diagrammatically in Figure 14. There were several interesting points to note from these results:

1) except for the first 3 amino acids, all the residues of dCK that have been sequenced (159 out of 225 residues, about 63%) were identical to those of dGK, as deduced from the DNA sequence of the dGK gene (a 100% match). Not even ONE residue in dCK was found that clearly was not encoded in the cloned dGK gene. This is, therefore, a very strong indication that dGK and dCK may be encoded by a single structural gene, and that there is no distinct dCK encoded in the L. acidophilus R-26 genome.
Figure 15. Result of amino acid sequencing of the dCK subunit from LBA. The underline portions of the dGK sequence represent dCK peptides which have been sequenced. The peptide sequences of dCK and dGK are identical, so far. The dAK and dGK sequences are inferred from the cloned dAK/dGK genes.
2) the amino acid residue identity between dAK and dGK (both purine deoxynucleoside kinases) was only about 50 % (80/159), and the identity between dGK (a purine deoxynucleoside kinase) and dCK (a pyrimidine deoxynucleoside kinase) was 100 % (159/159) of those residues sequenced, out of a total of 222 residues.

3) since dCK and dGK were identical so far (except difference for the first three residues, attributed to mRNA or protein editing?), it was easily to speculate that dCK and dGK were encoded by the same gene.

In the following chapters, evidence will be presented that some dCK-like proteins can be produced from mutated dGK genes.
E. MOLECULAR MASS COMPARISON OF WILD TYPE
dAK, dGK AND dCK SUGGESTS THAT dGK AND dCK ARE
TOO SIMILAR TO BE FROM SEPARATE GENES

It is obvious that it would be very difficult and expensive to
sequence every amino acid residue of dCK. However, there is another way
to detect a difference between peptides or proteins, and that is the molecular
mass measurement. If the mass of a peptide or protein doesn’t match up
with the one inferred from a known gene, the peptide or protein might not
be from the gene after all, or protein modifications may have occurred,
depending on how large the difference is.

Like the peptide sequencing work, the molecular mass measurements
was performed by Dr. Ikeda in our lab with the help of Ms. Angela Fieno
and her colleagues at Proctor & Gamble’s Miami Velley Research Facility.

First of all, the native dAK/dCK and dAK/dGK heterodimers from L.
acidophilus R-26 and the cloned dAK/dGK heterodimers from E. coli were
all purified on affinity columns, and the subunits were separated on a C-4
column by HPLC. After confirmation by SDS polyacrylamide gel
electrophoresis as to their identity and purity, masses of the dAK, dGK and
dCK subunits were measured by two separate methods, Nanospray and
MALDI. The result of the native dCK measurement is showed in Figure
16, and the results of all the measurements are listed in Table 2.

Table 2: The molecular mass comparison of dAK, dGK and dCK. Δ =
difference between the actual measurement and the calculated one. Unit =
Daltons (36). Figure 17 shows typical results of mass determination by the
Nanospray method.

<table>
<thead>
<tr>
<th>Kinase subunits</th>
<th>N-Terminal a.a.</th>
<th>Calculated Mass</th>
<th>Nanospray MH+</th>
<th>Δ</th>
<th>MALDI MH+</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>w.t. dAK</td>
<td>MIVLSG..</td>
<td>24,449</td>
<td>24,500.3</td>
<td>+1.3</td>
<td>24,504.0</td>
<td>+5.0</td>
</tr>
<tr>
<td>W.t. dGK</td>
<td>TVIVLSG..</td>
<td>26,185</td>
<td>-</td>
<td>-</td>
<td>26,195.8</td>
<td>+10.8</td>
</tr>
<tr>
<td>w.t. dCK</td>
<td>MIVLSG..</td>
<td>26,116</td>
<td>26,121.6</td>
<td>+5.6</td>
<td>26,137.5</td>
<td>+21.5</td>
</tr>
<tr>
<td>Cloned dAK</td>
<td>TVIVLSG..</td>
<td>24,568</td>
<td>24,568.4</td>
<td>+0.4</td>
<td>24,560.9</td>
<td>-7.1</td>
</tr>
<tr>
<td>Cloned dGK</td>
<td>TVIVLSG..</td>
<td>26,185</td>
<td>26,186.8</td>
<td>+1.8</td>
<td>26,194.5</td>
<td>+9.5</td>
</tr>
<tr>
<td>Mutant II*</td>
<td>MIVLSG..</td>
<td>26116</td>
<td>26,124.7</td>
<td>+8.7</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- Sample did not give the result because of slight salt content

* It will be described in the Chapter IV; it has the same primary amino acid
sequence as native dGK but with the same N-terminal sequence as native
dCK (so it can be seen as the control for w.t. dCK). If the only difference
between w.t. dGK and dCK are the two residues (TV), the Mutant II should
have exactly the same molecular mass as w.t. dCK (the difference here is
only 3.1 Daltons).
Figure 16. Result of molecular mass measurement of the native dCK by nanospray method. The final value is the average of 20 estimates. Performed by Proctor & Gamble’s Miami Valley Research Facility at Cincinnati, Ohio.
From the molecular mass measurements we could find several interesting things:

1) The difference between dGK and dCK, taking N-terminal variations into account, is only about 10 Daltons by the MALDI method, which was 1/10 of the mass of ONE amino acid residue (average = 110 Daltons). The difference is even smaller, merely 3.1 Daltons, between the wild-type dCK and its control (the Mutant II) by Nanospray method. The difference between the wild-type dCK and the calculated value is 5.6 Daltons. We were told there was trace of salt in our samples, which could shift the measured values slightly to the heavier side. Therefore, the wild-type dCK may well be exactly the same as its control or its calculated value if measured in the absence of salt.

2) The accuracy for MALDI method was about +/- 10 Daltons, therefore the maximum probable difference between dGK and dCK would be no more than 20 Daltons at most. The accuracy of Nanospray method is about +/- 1.5 Daltons according to the people who did our samples.
Therefore, dCK and dGK are indeed very similar since a 5 or even 3 Dalton difference is even smaller than the change caused by methylation (14 Daltons), the smallest candidate for peptide modifications for dCK (deamidation of Asn and Gln results in a change in mass of only one Dalton, but will change the charge, whereas there is no charge difference between wild-type dGK/dAK and dCK/dAK (36). Also a disulfide bond formation is -2 Daltons but there is no cysteine residue in any of these deoxynucleoside kinases. There is a remote possibility that dCK has a very similar molecular mass but a different amino acid sequence from dGK. However, considering that more than 63% of the dCK amino acid residues have been sequenced and they are identical to the sequence of the dGK residues, the likelihood that dCK has a alternative amino acid sequence which has almost exactly identical molecular mass as dGK will be very remote. Therefore, the molecular mass comparisons among wild type dAK, dGK and dCK strongly suggested that dGK and dCK are too similar to be from separate genes, and that there might be no distinct dCK gene in the genomic DNA of *L. acidophilus* R-26.
So far, the Southern hybridization results, direct sequencing results, cloning frequency, peptide sequencing results and the results of molecular mass comparison have been examined, and all the evidence strongly points in one direction: our hypothesis that there is no distinctive dCK gene in the LBA genome is correct. In the next chapter, evidence is presented which supports this hypothesis and demonstrates that indeed some dCK-like protein CAN be produced from the dGK gene by manipulation of the N-terminus.
CHAPTER IV

dCK-LIKE PROTEINS CAN BE PRODUCED
FROM TWO MUTANTS OF THE dGK GENE

My second hypothesis is that dCK and dGK are produced from the same gene, the dGK gene. In this chapter, I will demonstrate by site-directed mutagenesis that dCK-like proteins can be produced from mutated dGK genes. The N-termini of dAK, dGK and dCK of *L. acidophilus* R-26 are different at the first three amino acid residues (36), and this could be important in determining the specificity of the gene product. Since after sequencing more than 63% of dCK residues and finding that only the three N-terminal residues differ between dGK and dCK, it seemed reasonable to speculate that these three residues should play a key role in determination of the enzyme specificity of the dGK gene product. The N-termini of the subunit peptides responsible for dAK, dGK and dCK activities, are shown in Table 3.
Table 3. N-termini of dAK, dGK and dCK of *L. acidophilus* R-26 and N-termini of cloned dAK and dGK from *E. coli* (36)

<table>
<thead>
<tr>
<th></th>
<th>aa sequence inferred from gene</th>
<th>Actual aa sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>dAK from LBA</td>
<td>MTVIVLSG...</td>
<td>M__IVLSG...</td>
</tr>
<tr>
<td>dGK from LBA</td>
<td>MTVIVLSG...</td>
<td>_TVIVLSG...</td>
</tr>
<tr>
<td>dCK from LBA</td>
<td>-</td>
<td>M__IVLSG...</td>
</tr>
<tr>
<td>Cloned dAK in <em>E. coli</em></td>
<td>MTVIVLSG...</td>
<td>TVIVLSG...</td>
</tr>
<tr>
<td>Cloned dGK in <em>E. coli</em></td>
<td>MTVIVLSG...</td>
<td>TVIVLSG...</td>
</tr>
</tbody>
</table>

From Table 3, it is clear that some kind of mRNA editing or N-terminal post-translational processing or editing events occur for the first three residues (MTV) in LBA and for the first residue (M) in *E. coli*. While the N-terminus of dAK was the same as that of dCK in LBA, the total residue identity between the rest of dAK sequence and those of dCK residues, which have been identified, is merely 59%, assuming that its associated dAK is a product of the same gene which is associated in tandem
with that of dGK. Therefore, there might be many residues in dAK other
than the first three that can determine its specificity, and its N-terminal
processing may not be important in determination of its specificity.

However, between wild-type dGK and dCK the only difference so
far identified is at the first three residues. Accordingly, I created some
mutations, in which only the second and third codons were deleted from the
dGK gene, without any change in other parts of the gene, to determine what
effect these the loss of two residues would have on the dGK gene product.
The mutated genes were ligated into the pBluescript (KS +) vector and
expressed in *E. coli*. In the following sections I present experimental results
on the mutation designs, characteristics of the gene products of these
mutants, and comparisons with the wild-type dAK, dGK and dCK from
LBA.
A. MUTANT I, AND CHARACTERIZATION OF ITS GENE PRODUCT

The first mutation was designed to keep everything intact within entire dAK/dGK gene reading frames, except for elimination of the second and third codons only from the dGK gene. The design of Mutant I is shown in Figure 17.

Figure 17. Mutant I from the cloned dAK/dGK gene in pBluescript (KS+); P = promotor; T = terminator; \( \Lambda \) mark = removal of the 2nd and 3rd codons.
The mutagenesis was carried out by using the Mutagene kit from Bio-Rad, and the detailed steps can be found in the Method 23 (Material and Methods section). After mutagenesis, the synthesized double-stranded DNA (dAK/dGK with pBluescript vector) was transformed to the MV1190 strain of *E. coli*, and the mutation was confirmed by sequencing of the plasmid. Then Mutant I was transformed to *E. coli* strain HMS for expression and enzyme assays.

As I mentioned in the introduction, the deoxynucleoside kinases from *L. acidophilus* R-26 each have unique distinguishing characteristics, such as substrate specificity, $K_m$ values, activation effects by certain compounds upon its dimeric partner, inhibition effects of end-products, and its molecular weight. To examine if Mutant I produced a dCK-like protein, I made detailed comparisons between the mutant product and the wild-type dCK and dGK.

Specificity comparisons of the wild type, cloned dAK/dGK and the mutant I product are shown in Table 4. The reason I compared only the crude extract is that the Mutant I gene product is not very stable, and after
ammonium sulfate fractionation only about 1/5 of the total activity remained. The only way found to purify the mutant product at all successfully, so far, was gel filtration chromatography with the protection of 3% glycerol. However, during this purification process the product also lost activity rapidly even when ATP (5 mM) was used as an additional protecting reagent.

Table 4: Specific activity comparison of the wild-type enzymes, cloned enzymes and the Mutant I product. All the enzymes are crude extracts. Unit = nol/min/mg.

<table>
<thead>
<tr>
<th>Spec. Activity</th>
<th>Crude of LBA WT</th>
<th>Cloned dAK/dGK</th>
<th>Mutant I product</th>
</tr>
</thead>
<tbody>
<tr>
<td>dAK</td>
<td>0.6</td>
<td>3.6</td>
<td>0.02</td>
</tr>
<tr>
<td>dGK</td>
<td>0.33</td>
<td>22.0</td>
<td>0.01</td>
</tr>
<tr>
<td>dCK</td>
<td>0.32</td>
<td>0.1</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Although the Mutant I product was not very stable, from Table 4 it is very obvious that the deletion of the 2nd and 3rd codons of the dGK gene did switch the enzyme specificity from deoxyguanosine to deoxycytidine. The cloned dGK activity is 220 times higher than dCK activity, but the dCK
activity of the Mutant I product is about 40 times higher than dGK activity, with some dAK activity.

In order to further demonstrate that the protein with the higher dCK activity from the Mutant I is a dCK-like protein, we compared its basic characteristics with those of the wild type deoxynucleoside kinases. Four major characteristics of the kinases: the $K_m$ values of the substrates, the activators for dAK, the inhibitors and quaternary structures (by measuring the molecular weight of the mutant) were examined.

The $K_m$ values of the native kinases (dAK, dGK and dCK from LBA) are from the previous work by Drs. Hong, Ma and Ives (49), and the $K_m$ value of the Mutant I product was obtained as described in the Material and Method section (Figure 18). This value is close to true $K_m$ of dCyd since different concentrations of ATP were used (2.0 mM, 2.5 mM and 6.0 mM respectively). However, since rates at only 3 concentrations for each substrate could be measured reliably, I still call it an apparent $K_m$. The dAK activation, dAK, dGK and dCK inhibitions and cross inhibitions were performed as mentioned in the Material and Method section, and the results
are shown in Table 5 and Figure 20. Its molecular mass measurement, which indicated the organization of heterodimer, is shown in Figure 24.

Table 5: Properties of Wild-Type (dAK, dGK and dCK), cloned dAK and dGK, and Mutant I gene product (Wild type = purified enzymes; cloned = purified enzymes; Mutant I = partially purified enzymes). For activation and inhibition data see Figure 20; for IV° structure data see Figure 24. Km values of the wild-types and cloned dAK/dGK see (49); for Km value of dCyd of the Mutant I, see Figure 18.

<table>
<thead>
<tr>
<th></th>
<th>Wild-type (LBA)</th>
<th>Cloned dAK/dGK</th>
<th>Mutant I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Km of dAdo</td>
<td>2.2 µM</td>
<td>2.8 µM</td>
<td>-*</td>
</tr>
<tr>
<td>Km of dGuo</td>
<td>10.0 µM</td>
<td>65 µM</td>
<td>-</td>
</tr>
<tr>
<td>Km of dCyd</td>
<td>2.2 µM</td>
<td>--</td>
<td>98 µM**</td>
</tr>
<tr>
<td>dAK activators</td>
<td>dGuo, dGTP/</td>
<td>dGuo, dGTP</td>
<td>dCyd, dCTP</td>
</tr>
<tr>
<td></td>
<td>dCyd, dCTP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibitors</td>
<td>dATP, dGTP,</td>
<td>dATP, dGTP</td>
<td>dATP, dCTP</td>
</tr>
<tr>
<td></td>
<td>dCTP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV° Structure</td>
<td>Heterodimer</td>
<td>Heterodimer</td>
<td>Heterodimer</td>
</tr>
</tbody>
</table>

* The measurement of dAdo Km value failed for the Mutant I because of low activity and instability of the gene product.

** apparent Km (Figure 18)
Figure 18. Measurement of apparent $K_m$ for dCyd for Mutant I. X-intercept = 1.01, and $K_m$ value of dCyd = 98 μM. Three different concentrations of ATP were used: 2.0 mM, 2.5 mM and 6.0 mM.
By comparison of characteristics among the wild type (LBA), the cloned dAK/dGK and the Mutant I product, we concluded that the latter is very similar to the wild type dCK. The only differences between native dCK and the product of Mutant I are:

1) Mutant I product is very unstable (only about 1/5 total activity left after ammonium sulfate fractionation)

2) $K_m$ value of Mutant I product for dCyd is higher (98 μM) than that of the wild type dCK (2.2 μM), which seems to be reasonable in comparison with the relatively higher $K_m$ value of the cloned dGK for dGuo (60 μM) vs the value of the wild type dGK (10 μM), because their folding in *E. coli* may not be as ideal as in LBA.

3) the dAK activity of the Mutant I product is low in comparison with the cloned dAK/dGK gene product in *E. coli*. A possible reason is that the removal of the 2nd and 3rd codons from the dGK gene might affect its product’s allosteric effect upon dAK, which normally is seen as an increase in the turnover at the dAK subunit, when dGK is saturated with
dGuo or dGTP. Alternatively, the altered dGK gene product might have an effect upon dAK folding.

B. Mutant II, its product and characterizations

Since Southern hybridization with the dAK probe and direct genomic DNA sequencing indicate there is only one dAK gene in the LBA genome, it is very possible that the putative dCK gene is not associated with a dAK gene. In order to mimic that kind of situation, Mutant II, eliminating the second and third codons from the dGK gene, and the entire dAK reading frame as well, was designed. The design of the Mutant II is shown in Figure 19.
Figure 19, Mutant II from the isolated dGK gene in vector pBluescript. P = promotor; T = terminator, \( \wedge \) mark = position of mutation (deletion of the 2nd and 3rd codons)

Since the expression of the Mutant II, with an additional single base-pair modification at the Shine-Dalgarno region (to be described in detail in a later section) is excellent, amounting to about 40% of the total soluble proteins, the product of the Mutant II was purified to about 80% purity by a single-step purification (gel filtration chromatography). The characteristics of the purified Mutant II product were examined, and the results of
comparison with the wild type kinases are shown in Table 6. The control for Mutant II is the isolated, unmodified, dGK gene product (G-A, dAK not included).

Table 6. Properties of the pure wild type enzymes (LBA), cloned dAK/dGK, isolated dGK and Mutant II gene products. The data of wild type and cloned dAK/dGK were from Hong (49), Datta (19), and Bohman (20). The detailed data of molecular weight measurement, cross inhibitions, specific activities and $K_m$ study of Mutant II are shown in Figures 20, 21, 22, 23 and 24. The unit of specific activity is $\text{nol/min/mg}$

<table>
<thead>
<tr>
<th></th>
<th>Wild type (LBA)</th>
<th>Cloned dAK/dGK</th>
<th>G-A</th>
<th>Mutant II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp. Act., dAK</td>
<td>280</td>
<td>65</td>
<td>-</td>
<td>950</td>
</tr>
<tr>
<td>Sp. Act., dGK</td>
<td>2200</td>
<td>1000</td>
<td>850</td>
<td>1000</td>
</tr>
<tr>
<td>Sp. Act., dCK</td>
<td>2500</td>
<td>-</td>
<td>-</td>
<td>&gt;10,000</td>
</tr>
<tr>
<td>$K_m$ of dCyd</td>
<td>2.2 $\mu M$</td>
<td>-</td>
<td>-</td>
<td>cooperative</td>
</tr>
<tr>
<td>$K_m$ of dAdo</td>
<td>2.2 $\mu M$</td>
<td>2.8 $\mu M$</td>
<td>-</td>
<td>cooperative</td>
</tr>
<tr>
<td>$K_m$ of dGuo</td>
<td>10.0 $\mu M$</td>
<td>65 $\mu M$</td>
<td>60 $\mu M$</td>
<td>cooperative</td>
</tr>
<tr>
<td>Inhibitors</td>
<td>dATP, dGTP, dCTP</td>
<td>dATP, dGTP</td>
<td>dGTP</td>
<td>dCTP</td>
</tr>
<tr>
<td>IV$^0$ Structure</td>
<td>Heterodimer</td>
<td>Heterodimer</td>
<td>Homodimer</td>
<td>Homodimer</td>
</tr>
</tbody>
</table>
Figure 20. Cross activation and inhibition of the gene products of the Mutant II, G-A, Mutant I and cloned dAK/dGK with 2 mM dNTP. Y axis = percentage activity of enzyme with inhibitor or activator; X axis: 1 = dAK; 2 = dAK+dATP; 3 = dAK+dGTP; 4 = dAK+dCTP; 5 = dGK; 6 = dGK+dATP; 7 = dGK+dGTP; 8 = dGK+dCTP; 9 = dCK; 10 = dCK+dATP; 11 = dCK+dGTP; 12 = dCK+dCTP. * some dAK activity but it was too low to be reliable. In Mutant II, inhibition by dCTP is not as great as in Mutant I, which is similar to human dCK.
Figure 21. Rate plot of the three kinase activities of the Mutant II gene product, showing the interaction of the two active sites on the subunits. All three curves for dAK, dGK and dCK shown cooperative interaction between the active sites on the subunits. Y axis = reaction rate, v (nmol converted /30 min); X axis = each substrate concentration.
Figure 22. Rate plot of the dCK activities of the Mutant II under different protein concentrations. Shows the cooperation between the subunits for the three activities of the Mutant II is not caused by concentration changes. A = first degree plot; B = second degree plot.
Figure 23. Hill plot for the three activities of the Mutant II, showing degree of cooperations between subunits.
Figure 24. Molecular weight measurement of the products of the Mutant I gene, Mutant II gene, isolated dGK gene and cloned dAK/dGK genes by gel filtration chromatography (Method 19) to show the IV₀ structure of each protein. The molecular weight of monomers of dAK, dGK or dCK is about 26 kd. Therefore, all four species are dimers, but the mutant products appear a little larger than the unmodified cloned dAK/dGK, which may indicate misfolding. \( r^2 = 0.979 \)
The properties of the Mutant II product have been summarized in Table 6, and the cooperation between its substrate binding sites is shown in Figures 21, 22 and 23. Its molecular weight measurement by gel filtration, which indicated the homodimer organization is shown in Figure 24. From these comparisons with the wild type and cloned deoxynucleoside kinases, this mutant shown to have some very distinguishing characteristics:

1) by removing the entire dAK gene and the second and third codons of the dGK gene, the dGK gene now produced a protein with three kinds of deoxynucleoside kinase activities, i.e. for dAdo, dGuo and dCyd. The specific activity of dCK is the highest, much higher even than the wild type dCK from LBA at higher substrate (dCyd) concentrations (Figure 21). In comparison with the isolated and unmodified dGK gene product, it seems that the removal of the dAK gene did not play a key role in the switch of the specificity of the substrate. Since the mutant G-A (unmodified dGK only) showed homodimer organization, it is not likely that dimerization plays a role in determination of the specificity of the enzymes.

2) the $K_m$ values of the substrates are all high (about 50 $\mu$M for dAdo, dGuo and dCyd by estimation from the curves in Figure 21), and the active sites on the subunits of the homodimer appeared to be cooperative, a
phenomenon which is not found in the case of the dAK/dGK or dAK/dCK heterodimers, but was observed for dGK/dGK homodimers from LBA (data not shown). Hill plots have been constructed for dAK, dGK and dCK activities (Figure 23). For dAK, \( n_{app} = 1.8 \); for dCK, \( n_{app} = 2.8 \); for dGK, \( n_{app} = 3.4 \), which is strange but maybe interesting, since it has been found that there might be some higher oligomers, possibly trimers, (data not shown) for all the deoxynucleoside kinases (except TK) from LBA. Sometimes changing protein concentrations can cause a similar effect as cooperation between substrate binding sites. In order to check if it is true cooperation between the substrate binding sites, different amounts of protein were used (from 0.4 ng crude extract per assay to 4.0 ng/assay, a 10 fold increase) in the same experiment (assay for dCK activity only since it is the highest). The results (Figure 22) indicates that the concave kinetics is due to true cooperation, and was not caused by variable subunit association.

3) Inhibition tests (Figure 20) reveal the mutant having the basic qualitative characteristics of the wild type dCK from LBA, even though the inhibition was not as effective as in wild-type enzyme. However, this was to be expected because of its increased \( K_m \) for dCyd (possibly due to misfolding), resulting also in an increase \( K_i \) for dNTP.
From above comparisons, we can see once again that a dCK-like protein is produced from the dGK gene by very small changes at N-terminus.

C. Similarities of the Mutant II gene product and human dCK

Another very interesting characteristic of the Mutant II product is that its properties are very similar to those of human dCK (showing evolution pathway?). The comparison between the Mutant II product and human dCK is shown in Table 7.
Table 7: The similarities between Mutant II and Human dCK,

<table>
<thead>
<tr>
<th></th>
<th>Mutant II</th>
<th>Human dCK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrates</td>
<td>dAdo, dGuo, dCyd</td>
<td>dAdo, dGuo, dCyd</td>
</tr>
<tr>
<td>Inhibitor</td>
<td>dCTP</td>
<td>dCTP</td>
</tr>
<tr>
<td>IV(^0) Structure</td>
<td>Homodimer</td>
<td>Homodimer</td>
</tr>
<tr>
<td>Lowest K(_m) with:</td>
<td>dCyd*</td>
<td>dCyd</td>
</tr>
</tbody>
</table>

* Cooperative and the Michaelis-Menten curve of the Mutant II product was shown in Figure 20, 21 and 22.

It is quite obvious that these two proteins are similar to each other. The homology between the cloned dGK gene and the cloned human dCK gene was about 55%, but the homology between dGK and human dCK (amino acid sequences) was only about 27%. The discovery of the similarity between Mutant II and human dCK could shed some light on the evolutionary pathway of the deoxynucleoside kinases in general.
D. Expression of the Mutants in *E. coli*

Since the expression of dAK/dGK gene amounted to about 3% of the total protein (149), and the expression of Mutant I and the isolated dGK gene was less than 1% of the total proteins, it is desirable to increase the production of the cloned dAK/dGK gene and the mutants in *E. coli*. From the HPLC analysis of the dAK/dGK production from *E. coli* we knew that the ratio between dAK and dGK was about 1:3. There could be some elements in the upstream region, which determine the different expression level of the dAK and dGK (of course, it has to be pointed out that the 1:3 ratio could simply come from the instability of dAK). A check of the respective ribosomal binding sites (S/D regions) reveals that the difference between the dAK and dGK gene as only one base-pair. The S/D region not only can affect the binding of the ribosome, which affects translation directly, but also for some genes the complex plays a certain role in stabilizing mRNA (38) in *E. coli*. Therefore, it is interesting to examine what kind of effect the modification of that single base-pair at the S/D region has on the expression of the genes under studying.
The mutagenesis of a single base pair (8th upstream from the ATG site of the dAK gene) was performed for several genes and mutants by using Method 23 (Material and Methods section), and the mutation was confirmed by DNA sequencing. The wild type sequence at the promotor and S/D regions and the mutated sequence are shown below:

```
-35 -10  S/D
Wild type seq: ....tacactattaagcaaatatatgtgaggacgtaATG.....
Mutant seq:    ....tacactattaagcaaatatatgtgaggacgtaATG....*
```

The following clones were modified in this way to improve the expression:

- dAK/dGK genes
- dGK gene
- Mutant II

All the mutants were inserted into the pBluescript vector (KS +), and the expression conditions were the same as described in the Material and Methods section. *E. coli* host strains were varied, since some strains gave better specific activities for certain mutants. The effects of the single-base-pair change at S/D region on expression are shown in Figures 25 and 26, and the analysis of the resulting specific activities is shown in Table 8. For the cloned dAK/dGK genes, the single base-pair modification at the S/D region increased the protein expression by about 3-fold, from 3% to 10% of
the total proteins, and the specific activity was increased 3-fold as well, from 22 to about 65 nmol/min/mg. For the unmodified dGK gene, isolated from its dAK gene partner, the S/D modification increased the protein expression 30-fold, from less than 1% to about 30% of total proteins, but the specific activity was increased by only 5-fold, which suggested that there was much misfolding during the expression. In comparison with the cloned dAK/dGK modification, it is easy to see that the existence of the dAK gene could play some role in the folding of the dGK gene product. For Mutant II, the modification increased the protein expression by at least 30-fold, from less than 1% to at least 30% of total protein. An interesting point is that the specific activity has been increased by about 100-fold, much more than the increment of the protein production, which suggested the single base-pair modification may have had a positive effect on folding in this instance. Therefore, the S/D sequence and dimerization both could play some role in stabilizing or de-stabilizing the gene products of the dAK/dGK genes in LBA.
Figure 25. SDS polyacrylamide gel electrophoresis of the crude extracts from the cells of the cloned dAK/dGK genes and their mutants. M = markers; Lanes: 1 = pBluescript vector only; 2 = unmodified cloned dAK/dGK genes; 3 = cloned dAK/dGK genes with g/a change at S/D region; 4 = unmodified dGK gene; 5 = dGK gene with g/a change; 6 = Mutant II without g/a change; 7 = Mutant II with g/a change. Kd = kilodaltons.
Figure 26. SDS polyacrylamide gel electrophoresis for purified Mutant II with the g/a change at the S/D region. Lanes: 1 = protein markers; 2 = purified Mutant II product from the gel filtration chromatography; 3 = markers
Table 8. The comparison of specific activities of products of the mutants and of unmodified genes in crude extracts.

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>dGK Specific activity (nmol/min/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dAK/dGK</td>
<td>22</td>
</tr>
<tr>
<td>dGK</td>
<td>1.7</td>
</tr>
<tr>
<td>Mutant II</td>
<td>0.5</td>
</tr>
<tr>
<td>dAK/dGK+g/a</td>
<td>64.7</td>
</tr>
<tr>
<td>dGK+g/a</td>
<td>17.1</td>
</tr>
<tr>
<td>Mutant II+g/a</td>
<td>50</td>
</tr>
</tbody>
</table>

Northern hybridization was also performed on the total RNA from *E. coli* host cells expressing the cloned dAK/dGK genes (no g/a change), cloned dAK/dGK genes with the g/a change, the isolated dGK gene (no g/a change) and the isolated dGK with the g/a change, to see if any alterations
occurred at the mRNA level. The Northern hybridization results are shown in Figure 27, and from these results it is clear that the amount of mRNA was much higher in cells containing the modified S/D sequence. There are two possible reasons: 1) the change at the S/D region increased the production of the mRNA; 2) the change at the S/D region increased the stability of the mRNA so that its rate of turnover was reduced. It is entirely possible that both effects occurred simultaneously.
Figure 27. Results of Northern hybridization of the unmodified cloned dAK/dGK, and its mutants. Probe = dGK/dCK probe, labeled with $^{32}$P. 1 = the isolated dGK gene (from its partner dAK gene) without any S/D modification; 2 = the dGK gene with the g/a change at S/D region; 3 = the cloned dAK/dGK without any modification; 4 = the cloned dAK/dGK with the g/a change at S/D region; 5 = pBluescript vector only. Marker = denatured λ/Hind III.
CHAPTER V

CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS

A. CONCLUSION AND DISCUSSION ON THE HYPOTHESES I PROPOSED

I have presented evidence in the previous chapters for the two hypotheses I proposed: 1) there is no distinct dCK gene in the LBA genome, and 2) the dCK is produced from the dGK gene in this organism. Actually this idea began to develop before we started the cloning experiments as there was some evidence from G. Ma’s Southern hybridization that suggested it (1). Of all the lines of evidence we presented, some are stronger and some are weaker. An attempt to evaluate and compare these lines of the evidence will be presented here.
The piece of first evidence was from Southern hybridization experiments. In the Southern hybridization, the dAK probe (375 base-pairs of the dAK gene) detected only one dAK gene fragment in the LBA genome, and there is absolutely no other evidence to show otherwise. Therefore, a dCK gene, if it exits, cannot be associated in tandem with a dAK gene. In addition, the dGK/dCK probe detected only the dGK gene in the LBA genome, and again there is absolutely no other evidence to the contrary. However, there may be some additional arguments which can be drawn from these experiments.

The first is whether or not the probes can anneal to the target genes they were designed for, that is, the putative second dAK gene and the putative dCK gene. In the case of the dAK probe, since the molecular masses of the cloned dAK gene product from *E. coli* and the wild type dAK from LBA are exactly the same (Table 2, the only difference being due, apparently, to N-terminal processing), it seems clear that there is only one kind of dAK peptide in the LBA. Therefore, if there were another copy of the dAK gene in the LBA genome, it would have to be either exactly the same as, or very similar to, the cloned dAK gene. Thus, the dAK probe
should be able to anneal to any dAK gene in the LBA genome, and the
Southern hybridization results should be conclusive.

In the case of dCK/dGK probe, the key was whether the primer
p121694, by which the probes were made, would anneal to the putative
target dCK gene (it is certain that this primer can anneal to the dGK gene)
in the probe making by PCR. Considering the degeneracy and the
annealing temperature (a 26 °C difference) we used in making the probe,
there is no apparent reason why primer p121694 should not anneal to the
putative dCK gene. Another factor is that since the peptide sequences of
the dGK and dCK are at least 63 % identical, and these two enzymes have
virtually identical molecular weights, if there is a distinct dCK gene its
DNA sequence should also have very high homology to the dGK gene and
even a dGK probe (from the dGK gene) should hybridize with it.
Therefore the Southern hybridization result with the dGK/dCK probe in
revealing only one gene (the dGK gene) should be conclusive.

The second argument which could be raised from the Southern
hybridization experiment is that the dGK gene and the putative dCK gene
are either too close together or have no cutting site between them. This argument is not valid because:

1) the PCR with the “back-to-back” primers (primer 4-28-R and 4-28-L for dAK, +7 from Grace Ma and p121694 for dGK, shown in Table 9) yielded no product, so that there is not likely to be a dAK/dGK-like sequence near-by (within about 10 kb)

2) Even we assume the PCR by the “back-to-back” primers failed to reveal the near-by putative dCK gene, we knew there are, at least, two Kpn I sites closely flanking the dAK/dGK genes (the dAK/dGK gene was cloned by cutting at these sites), which will have eliminated the possibility of “too-close” or “no-cutting-site” theory. However, the Southern hybridization of Kpn I fragments still revealed only one band.

Therefore, the results from Southern hybridization strongly supported the hypothesis that there is no distinct dCK gene in the LBA genome.

Direct genomic DNA sequencing is a new and powerful technique. The result of the direct genomic DNA sequencing of the LBA firmly
supported, from another perspective, the hypothesis that there is no distinct
dCK gene in the LBA genome. Even though there was no degeneracy used
in the design of the primers for the direct genomic DNA sequencing,
considering the difference between the annealing temperature of the primers
and the PCR annealing temperature (usually more than 10 °C) and the 100
% identity of the dGK and dCK amino acid residues at the primers’ regions,
it is reasonable to say that the primers should anneal to the potential dCK
gene as well. However, it has to be pointed out that the probability of the
primers used in the experiment annealing to the potential dCK gene are not
as great as for the probe used in the Southern hybridization experiment.

Someone may ask if the results we showed here were from the
contaminations by the cloned dAK/dGK gene. From the Method section,
we can see that this method, unlike PCR, does not exponentially amplify the
template, and the density of the resulting bands are sensitive to the amount
of the template. Therefore, unless the amount of the contamination is
extremely high, such as reaching 1/5 of that of the template, it won’t show
up on the film (we did the contamination test, and when the amount of the
contamination was less than 1/5 of the template, only the template could be
detected on the film).

The another way to prove our first hypothesis is to show that dGK
and dCK have a common amino acid sequence. Two pieces of evidence
were presented, that is, the peptide sequencing data of the wild type dCK
and the molecular mass measurement data of the wild type dAK, dGK and
dCK, and the cloned dAK and dGK peptides. These data clearly show dGK
and dCK to be identical in the peptide stretches we sequenced and to be
virtually identical in their molecular masses with only a 3-Dalton
discrepancy, about 1/30 of an amino acid residue (taking into account
differences in the first three residues). It would be very hard to imagine that
two such similar peptides could come from two separate genes, especially
considering that the genome of the LBA is fairly small (about 4000 kb).
Another point is that these data really make the Southern hybridization
results and direct genomic sequencing results more convincing because if
the two proteins are very similar or even identical, their genes should be
hybridized by the same probe or amplified by PCR to similar degrees,
whereas no such parallel hybridization or amplification was seen in the case
of LBA. Therefore, the hypothesis that there is no distinct dCK gene in the LBA genome was further supported. The weak points in the argument based on the peptide sequencing and molecular weight data are 1) not all the residues were sequenced, which becomes increasingly difficult to do and, perhaps, not cost-effective; 2) there was a 3-Dalton difference between the dGK and dCK, which as I mentioned before could be caused by slight salt content in our samples. The second point could become significant if dCK turns out to have some small post-synthetic modification(s), which might help to determine the specificity of the dGK gene product.

The evidence from the dCK gene cloning, or rather the lack of it, is not as strong an argument as other evidence, since many unknown factors may be involved such as possible toxicity of a dCK gene product to the host cells or dCK gene location (near a toxic gene), really bad luck, instability of the putative dCK cloned gene construction, or the inadequacy of the methods used to clone the putative dCK gene. The last one is not likely because the dAK/dGK genes have been re-cloned many times by the same methods, and the dGK and dCK genes are supposed to be very similar.
The really positive and direct evidence for the hypotheses that there is no distinct dCK gene in the LBA genome, and that dCK is from the dGK gene, is from the results of the mutagenesis experiments. In two mutants we presented, the simple deletion of the 2nd and 3rd codons from the dGK gene totally switched the specificity of the dGK gene product from dGuo to dCyd, which strongly indicated the possible origin of the dCK in LBA, even though the mechanism of such a conversion is lacking. Of course, the products of the two mutants were not as good as native dCK in terms of stability, $K_m$ values, and maybe folding, which might be the problem of the mutations themselves or of the host ($E. coli$), especially since the unmodified dAK/dGK gene product in $E. coli$ also has similar problem with elevated $K_m$ value for both of its substrates.

In summary, putting all the evidence we presented together, in my opinion, the two hypotheses that there is no distinct dCK gene and the dCK is produced from the dGK gene is proved beyond a reasonable doubt.

Two other possible ways to prove our hypotheses were tried without success, but perhaps they should be considered by future researchers. The
first one was the conversion of the dGK to dCK by denaturation and renaturation, supposing that the dGK and dCK might be two interconvertible conformers. Several means of denaturation of the dGK were tried, including different concentrations of guanidine-HCl (from 0.01 M to 0.4 M); different concentrations of urea (from 0.5 M to 2 M); thermal-denaturation (20 °C, 25 °C, 30 °C and 37 °C). For the chemical denaturation and renaturation, all the procedures were carried out at room temperature.

The dGK and dCK activities were monitored during both the “opening-up” phase and the recovery phase, and no dCK activity was detected even though up to 40% of the dGK activity was recovered in some cases (Ikeda, 38). The reason for this result could be: 1) the folding to dGK or dCK conformers either happened at the early stage of peptide synthesis, or is guided by some other proteins so that, after formation of the final structure, it could not be redirected into different active conformers in vitro; 2) the correct reaction conditions have not been discovered yet.

The second experiment we tried was to transform the cloned dAK/dGK genes back into LBA or its close relatives to see if the transformation could promote an elevation of the dCK activity, which could
be the really positive evidence for our hypothesis. Four kinds of vectors from Gram-positive bacteria (pC194, pGDV1, pHB201, kindly provided by Dr. Donald H. Dean at the Ohio State University, and pLB-10, kindly provided by Dr. Klaenhammer of the North Carolina State University) and several species and strains of *Lactobacillus* were tried without success. The problem was that any plasmids, with or without the cloned dAK/dGK gene insert, once entering the host, would slow down cell growth drastically, increasing growth-time to achieve an $A_{600}$ of 1.0 from 12 hr to 72 hr, and this usually led to the situation that there were so little deoxynucleoside kinase activities left that they could not be clearly distinguished from the background.

In addition to all the evidence and the arguments I presented before, there is another interesting point which may support my hypotheses and needs to be addressed here. There are many reports that simple alteration of the primary peptide sequence, such as a one-residue change at the active site of the enzyme, can switch the substrate specificity of an enzyme. The most classical example is the change at substrate pocket of trypsin (55). In more closely-related field, there are several reports on switching between
purine substrates (A and G) (56,57) and pyrimidine substrates (T, C and U) (58) caused by single amino acid alterations. An example that is even closer to our case is reported by Okajima group (59) in 1993, in which two amino acid residue alterations (T39 by A and L66 by I) caused remarkable increases in CMP or UMP activities and decrease in AMP activity for adenylate kinase. However, it is obvious that those changes all have to be at the active site, but it is not known yet in our case if the amino terminus or the residues involved are located in or near the deoxynucleoside binding site.
B. DISCUSSION ON THE POSSIBLE MECHANISM BEHIND
THE ONE-GENE-TWO-PROTEINS PHENOMENON

So far, we have discussed a very interesting phenomenon in bacterium, Lactobacillus acidophilus R-26: there is only one gene (dGK gene), but two different primary peptide product, one is dGK and the other is dCK, and the only difference between them is the removal of the second and third residues from dGK. Since the change can happen at only two possible places (mRNA level or protein level), there are several mechanisms to be considered:

At mRNA level:

1) mRNA editing;
2) ribosomal hopping;

At protein level:

3) N-terminal processing;
4) protein splicing;
a. At mRNA level

As far as the Ribosome hopping is concerned, it is a relatively rare phenomenon, and only reported in the gene 60 of the bacteriophage T4 (60) and probably in a 1,4-β-D-endoglucanase gene (from *Bacteriodes ruminicola*) expressed in *E. coli* (61). Someone speculated that the control over cat 86 (human) protein expression might be the ribosomal hopping, but later it was proved that the mechanism is still attenuation (62). Considering in all instances the hopping distance is relatively large (50 nt in T4 phage and over 600 nt in the second example) due to the hair-pin structure involved, it is hard to imagine that the 6 base-pairs in our case could form a hair-pin and cause the ribosomal hopping.

mRNA editing concept has been established since 1986 (63) along with other kind of RNA editing. Today, the term mRNA editing is referring to base addition or deletion by RNA cleavage and substitution or modification without cleavage. It doesn’t include multiple RNA processing (splicing, poly A tail formation and capping) and co-transcriptional editing (in many viruses). It seems that tRNA editing is more common than mRNA editing (64). There have been many reports on mRNA editing in various
kinds of organisms, such as protozoa, vascular plants, especially in mitochondria, and mammals (64), but none have been reported in prokaryotic organisms. In most of the instances, the editing has been limited to a single base pair change, and the most common change is from C to U or U to C (64-66). The most-quoted example of mRNA editing is apoB100 (512 kd) and apoB48 (241 kd) proteins from mammals, the latter being the product of tissue-specific mRNA editing from Glu (CAA) to a stop codon (UAA) (67,68). So far two kinds of mechanisms have been proved or proposed for mRNA editing. 1) function of a double stranded RNA adenosine or cytidine deaminase (65,66) and 2) transesterification with help of a guide RNA (gRNA)(69,70), but none of them could explain our case well. It is easy to see that mRNA editing has important function in control of development and regulations.

Among the possible mechanisms mentioned before, mRNA editing is the easiest to work on since there are some proven methods of detection, and they are relatively easy to use. Therefore, mRNA editing has been chosen to start the investigation of the mechanism behind the one-gene-two-protein phenomenon, and the experimental procedure is the following:
The total RNAs from LBA and *E. coli* were isolated according to the method mentioned in Material and Methods section, the only change is: using DNase (RNase free) 3 times instead of once to eliminate DNA or DNA fragments as much as possible. Then, RT-PCR was performed according to the method described in the Material and Methods section for two separated preparations of the total RNA of LBA to ensure reliability of the results. The positive controls were the mRNA from the *E. coli* clones containing the cloned dAK/dGK genes. The first negative control was the total RNA from LBA without going through the RT step (therefore, nothing should be amplified by RT-PCR), to show the positive bands are not from DNA remaining in RNA preparations. The second negative control was the reaction mixture of the RT-PCR without adding any polynucleotide (RNA or RT products) to show the positive bands are not from any contamination (DNA or cDNA) in the RT-PCR experimental procedure. Primers used in RT-PCR are +6 and 12-8-3' (Table 10), which covered a total of 105 base-pairs from the 5' end of the interval sequence between the dAK and dGK genes to the 84th base-pair from the beginning of the dGK gene (Figure 28). The final results are showed in Figure 29.
Figure 28. Design of RT-PCR. Purpose of the RT-PCR is to determine if any editing occurs at the 5' end area of the mRNA of the dGK gene. Template = mRNA (in total RNA), and covering area from the 5' end of the interval sequence between the dAK and dGK genes to the 84th base-pair of the dGK gene.
**Figure 29.** RT-PCR result on denaturing DNA sequencing gel. Lane 1 = negative control (without adding any RNA or RT products); 2 = negative control (adding RNA without going through RT step); 3 = RT-PCR for LBA RNA preparation; 4 = RT-PCR for another LBA RNA preparation; 5 = positive control (RT-PCR for total RNA from the cloned dAK/dGK genes). Primers = +6 and 12-8-3', and product is 105 bp in length.
From the results we can see clearly that there is one species amplified from the mRNA in the isolated total RNA, which was exactly same as the positive control, namely, the one from the unmodified-cloned dGK/dGK genes. In the two negative controls nothing had been amplified by PCR, which indicated that the bands in lane 3 and 4 were indeed from mRNA but not DNA or other contaminations.

It is has to be pointed out that even though we used two negative controls, they do not rule out any contaminations occurring between the RNA preparation and RT-PCR, or RT steps. Therefore, the bands in Figure 29 could still arise from contamination, since this type of contamination problem is a routine “headache” for any PCR operation. The only resolution to this problem is to repeat the above procedure in a new location or lab which has never been exposed to the cloned dAK/dGK genes or their mutants.
b. At the protein level

As far as the change at the protein level (co- or post-translational) is concerned, there are two possible mechanisms as I mentioned before, that is protein splicing, and N-terminal processing. The latter shouldn’t be confused with N-terminal peptide cleavage in which a peptide is cleaved at certain position near the N-terminus to form a mature or functional protein, such as in the case of insulin.

Protein splicing (71), analogous to the term RNA splicing, means the removal of an internal protein sequence, or “intein”, from a precursor molecule and the ligation of the two flanking sequences (exteins) to produce a mature protein product. The mechanism is probably entirely encoded within the inteins and does not require other accessory molecules. From the reported instances in which a protein splicing happened, it is apparent that all the inteins are much larger sequences, than the two amino acid residues to be accounted for in our case.

Protein splicing, discovered in 1986 (72), is the newest member of a class of molecular phenomena that determines the flow or organization of
genetic information. It has been reported from many phylogenetically diverse species, such as from yeast (VMA1 or TFP 1) (73), and bacteria (RecA) (74,75), and certainly more and more instance will be observed from many other organisms in the future. However, in our case the removal of the second and third amino acid residues is not likely due to the protein splicing, because of the size of the sequence removed and the location, so it more likely occurs by N-terminal processing or a related mechanism.

The second possible mechanism operating at the protein level is the N-terminal processing. The N-terminal processing refers to co- or post-translational modifications in which the initiating methionine (and the second residue in some cases) is enzymatically removed (76). The removal of the amino acid residues has to be at the end and in the fashion of one by one. It shouldn't be confused with N-terminal cleavage because the latter is “endo” cleavage and can be at any specific site in the middle of the primary sequence near the N-terminus.

N-terminal processing was first recognized by J. Adams in 1968 (77). Since then, the mechanisms and enzymes involved have been
discovered. There are at least two separate enzymes involved in the N-terminal processing, the first one is methionine aminopeptidase (MAP), which will remove the methionine residue, and the second one is N-acetyltransferase (NAT), which will add an acetyl group to the N-terminal residue. Therefore, the combined activity of NAT and MAP can produce four classes of proteins: those with and without methionine and those with and without N-acetyl groups (76). N-terminal modifications may affect biological function, stability and folding, translocation, and degradation.

As I mentioned before, the modification is not limited to the first amino acid residue (i.e. methionine in most cases). Rubenstein group has found that there are two groups of actins (class I and class II), with only the N-acetylmethionine being removed in class I actins, while the acetylmethionine and acetyl-X (second residue, mostly Cys) are removed in class II actins (78), catalyzed by actin N-acetylaminopeptidase. Therefore, it is very possible that more than one residue is removed in the N-terminal processing. In our case, the first three residues (MTV) could be removed one by one (it cannot be ruled out that the cleavage happens at the site either), and then the methionine is somehow re-attached to the peptide. Of
course, nothing similar has been reported so far, so such a mechanism is purely a speculation here. The questions arising from this speculative mechanism are 1) how the methionine is re-attached to the peptide, and 2) how the removal of the T and V residues causes the switch in specificity. However, it should be pointed out that there are reports that some small changes at the N-terminus can switch the enzyme specificity. Multiple receptor subtypes have been recognized for several human peptide hormone families, and in one hormone receptor family PYY, PYY1-36 receptor was enzymatically modified to remove Tyr-Pro from the N-terminus to form PYY3-36 (79). The test results showed that the specificities of those two receptors are quite different. Even though the length of the receptor is only 36 residues and its specificity depends on the length of its N-terminus, which may make the example less convincing, it did show, to some degree, that it is possible that some small change at N-terminus can cause a switch in binding specificity.

It is worth mentioning that there is another term, protein editing, which can be considered to correspond to the term mRNA editing. It should be pointed out that protein editing has not been reported from any
organisms yet, but is the subject of recent speculation. By analogy with mRNA editing, we can suggest that protein editing could involve deletion, insertion or alteration of one or a few amino acid residues in the middle of a primary peptide sequence. In my opinion, this term, if the phenomenon really exists, should explain our case (removal of two amino acid residues in the middle of the primary peptide sequence) better than protein splicing or N-terminal processing.
Table 9. List of oligo-nucleotides designed in this study

<table>
<thead>
<tr>
<th>Primers</th>
<th>No.</th>
<th>Sequences (all from cloned dAK/dGK genes except marked)</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-22-3' (24)</td>
<td>1</td>
<td>5'- AAAGGATTAGTACCCAATAATTT</td>
<td>For PCR cloning</td>
</tr>
<tr>
<td>12-22-5' (22)</td>
<td>2</td>
<td>5'-ATGACAGTTATTGTATTAAGCG</td>
<td>For PCR cloning</td>
</tr>
<tr>
<td>3-30-3' (18)</td>
<td>3</td>
<td>5'-GTAAAACGACGGCCAGTG</td>
<td>=forward primer of pUC</td>
</tr>
<tr>
<td>3-30-5' (20)</td>
<td>4</td>
<td>5'-AGGAAAACAGCTATGACCAGT</td>
<td>=reverse primer of pUC</td>
</tr>
<tr>
<td>5-25-R (21)</td>
<td>5</td>
<td>5'-CTTITAAACTTGGATGCGGGT</td>
<td>Back-to-back</td>
</tr>
<tr>
<td>5-25-L (22)</td>
<td>6</td>
<td>5'-ATTACTATGCTGCAATTTTTAT</td>
<td>Back-to-back</td>
</tr>
<tr>
<td>12-8-3' (25)</td>
<td>7</td>
<td>5'-ATTAGTACCCAATATTTAGATAAG</td>
<td>For PCR cloning</td>
</tr>
<tr>
<td>12-8-5' (22)</td>
<td>8</td>
<td>5'-ATGACAGTTATTGTATTAAGCG</td>
<td>For PCR cloning</td>
</tr>
<tr>
<td>10-6 (32)</td>
<td>9</td>
<td>5'-CTAACAGGATCTCTATCTAATATTTGGTAC</td>
<td>For direct genomic seq.</td>
</tr>
<tr>
<td>4-28-R (21)</td>
<td>10</td>
<td>5'-GGTGAAAGCATATATGCAAGAG</td>
<td>For dAK probe</td>
</tr>
<tr>
<td>4-28-L (20)</td>
<td>11</td>
<td>5'-CGAACATTTAGGACTGAGG</td>
<td>For dAK probe</td>
</tr>
<tr>
<td>p121694 (27)</td>
<td>12</td>
<td>5'-AACTAAATTAGGTTTCCAGTTCTACCAGGCC (A) (A) (G)</td>
<td>For dGK/dCK probe</td>
</tr>
<tr>
<td>p121794 (35)</td>
<td>13</td>
<td>5'-CCCCGCTTAAATACAAT_CATCAGGTCCTCTTTCCAGCG</td>
<td>For mutation (-TV)</td>
</tr>
<tr>
<td>p12095 (27)</td>
<td>14</td>
<td>5'-AACTAAATTAGGTTTCCAGTTCTC</td>
<td>For dGK/dCK probe</td>
</tr>
<tr>
<td>p32995 (27)</td>
<td>15</td>
<td>5'-CACGTTCTTTTCTACATATATGTTG</td>
<td>For mutation (g/a)</td>
</tr>
<tr>
<td>p33095 (38)</td>
<td>16</td>
<td>5'-CCCCGCTCAATACAATATACGTCGTTCTCTAC</td>
<td>For mutation (g/a)</td>
</tr>
<tr>
<td>p71295 (32)</td>
<td>17</td>
<td>5'-CCCCGCTCAATACAATACGTCGTTCTCTAC</td>
<td>For mutation (-TV)</td>
</tr>
<tr>
<td>p61495 (23)</td>
<td>18</td>
<td>5'-GTTAGGCTAGTTAAATACTCG</td>
<td>For RT-PCR</td>
</tr>
<tr>
<td>p72095 (50)</td>
<td>19</td>
<td>5'-CCATATGGACCCAGTTTTGCT_CGTTAACTGTTTTAATTTCCCT</td>
<td>For mutation (-dGK gene)</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


51. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G.T., Mullis,


