AN ABSTRACT OF THE THESIS OF

Margaret E. Black for the degree of Doctor of Philosophy in Microbiology presented on April 30, 1991.

Title: Structural and Functional Dissection of the Vaccinia Virus Thymidine Kinase Enzyme

Abstract approved: Redacted for Privacy

Dr. Dennis E. Hruby

Thymidine kinase is a key enzyme in the nucleotide salvage pathway, catalyzing the production of dTMP from thymidine and ATP. In order to identify the structural features of the TK protein and/or primary amino acid sequences which contribute to the catalytic and regulatory activities of this enzyme, an *in vitro* transcription and translation system has been used in concert with protein engineering techniques for the production and phenotypic characterization of mutant and wild-type TK enzymes. Because of discrepancies in the literature regarding the quaternary structure of the VVTK, the native molecular weight and quaternary structure was determined to be an 80kDa homotetrameric enzyme by glycerol gradient fractionation, gel filtration and glutaraldehyde cross-linking analyses.

Computer-assisted alignment of the predicted amino acid sequences derived from cellular and poxvirus TK genes identified seven highly-conserved domains distributed throughout the VVTK polypeptide

(domains I-VII). Domain I (amino acid residues 11-18) exhibits a high degree of similarity to both ATP and GTP binding site consensus sequences, although the VVTK utilizes only ATP as a phosphate donor. Site directed mutagenesis and ATP-agarose affinity chromatography techniques were employed to confirm that this region was responsible for ATP binding and to determine which amino acids were essential for efficient binding.

The TK gene (*tdk*) from *E. coli* was isolated and sequenced to serve as a prokaryotic enzyme with which to compare VVTK. The alignment revealed only 23% shared identity with VVTK and, interestingly, the identical and similar residues were clustered into three of the seven domains identified previously (domains I, III and VII).

Preliminary evidence supports domain III (residues 78-90) as a putative magnesium binding site and that a highly conserved cysteine residue (cysteine 170) within domain VII (residues 168-171) may be a component of the catalytic site. Secondary structure alignment between Herpes Simplex Virus (HSV) TK and monkeypox TK (a close relative of VVTK) revealed that the putative nucleoside binding site of HSVTK aligns with residues within domain IV. Replacement of a VVTK residue (Q114) with the corresponding residue of HSVTK (an aspartic acid) greatly alters the substrate specificity and dTTP sensitivity of VVTK.

Structural and Functional Dissection of the Vaccinia Virus Thymidine Kinase Enzyme

by

Margaret E. Black

A THESIS

submitted to

Oregon State University

in partial fulfillment
of the requirements
for the degree of

Doctor of Philosophy

Completed April 30, 1991

Commencement June 1991

APPROVED:

Redacted for Privacy

Professor of Microbiology in charge of major

Redacted for Privacy

Chairman of the Department of Microbiology

Redacted for Privacy

Dean of Graduate School

Date thesis is presented April 30, 1991

Typed by the researcher for Margaret E. Black

ACKNOWLEDGEMENTS

Many thanks to friends and acquaintances who have had a positive influence on me and have supported me over the years. In particular, I would like to thank Norm Boudreaux, my high school Biology teacher who made learning Biology an exciting adventure; Dr. Eugene Cota-Robles, an enthusiastic supporter of my first foray into the world of molecular biology; Dr. Caroline Dean and other members of the AGS Plant Molecular Biology Group, who showed me that only good science is worth doing and, perhaps most importantly, that science can be fun. Special thanks go to my parents whose love and understanding has been unwavering. I cannot express enough gratitude for the love and joy that Carl brings to my life which makes the trials and tribulations of day to day living seem inconsequential.

For all the enthusiastic support and guidance with which Dennis Hruby has aided me through graduate school and into the scientific community I am greatly indebted. I am especially grateful to his openness in discussing how the business science "works" in the academic world, lessons that most graduate students don't receive. Dr. Bill Dougherty has spent many hours discussing varying aspect of my project for which I am extremely grateful. I'd also like to especially thank him for taking the time to listen, a rare trait in scientists these days. Dr. Chris Mathews and Nancy Ray are thanked for helping me get the *E. coli tdk* project off the ground and running and Yea-Huey Yang for making the domain VII mutants.

Chapters II, III and IV have been published by Academic Press, Inc, American Society for Biochemistry and Molecular Biology and Blackwell

Scientific Publications, Ltd., respectively. Permission to publish these manuscripts as part of my doctoral dissertation has been granted by the individual publishing agencies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

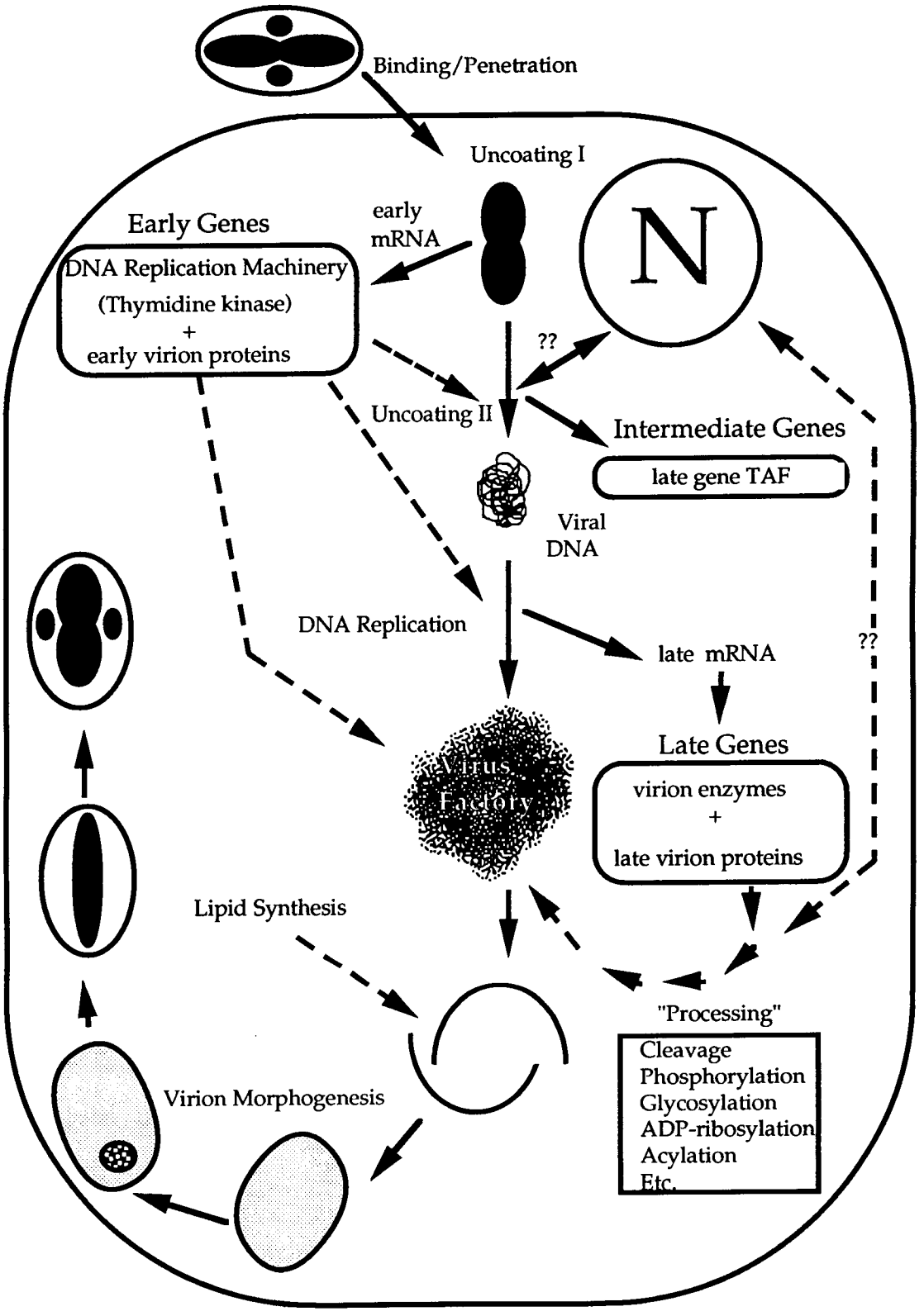
STRUCTURAL AND FUNCTIONAL DISSECTION OF THE VACCINIA VIRUS THYMIDINE KINASE ENZYME

CHAPTER I

INTRODUCTION

Vaccinia Virus Replication Scheme

Vaccinia virus (VV) is a large double-stranded DNA virus with a linear genome of 196 kbp and encodes approximately 190 genes (Goebel *et al.*, 1991). Upon entry into the cell and partial uncoating of the viral particle (uncoating I), a set of genes designated as "early", are actively transcribed and their messages translated (Fig. I.1). Many of these genes encode enzymes involved in DNA replication and DNA precursor synthesis such as DNA polymerase, RNA polymerase, polyA polymerase, capping enzyme, topoisomerases, ribonucleotide reductase (both large and small subunits), thymidylate kinase and thymidine kinase (Moss, 1990). While there is no distinctive sequence at the 5' end of early gene transcripts, the 3' ends of these transcripts contain a sequence comprised of UUUUUNU located approximately 50 nucleotides upstream from the transcriptional termination site (Yuen and Moss, 1987). Recently the RNA polymerase-associated heterodimeric enzyme involved in modifying the first nucleotide of the message with mGppppG, the "capping enzyme", has also been shown to act as a transcriptional terminator of early messages, specifically recognizing the UUUUUNU sequence and terminating



VACCINIA VIRUS REPLICATION STRATEGY

Figure I.1

transcription at a distal site (Shuman and Moss, 1988; Shuman, 1990). Distinct length transcripts then are synthesized with a 5' cap and a 3' poly-A tail.

After early genes are expressed, the viral core completely breaks down (uncoating II) to release the viral DNA (Fig. I.1). Expression of a newly described group of genes, the intermediate genes, occurs after early genes, but prior to DNA synthesis and late gene expression. The lack of any conserved early or late structural features which investigators normally use to distinguish early genes from late genes has made initial classification of these genes difficult. The function of sequences, similar but not identical to the conserved late gene promoter TAAATG (described below), found in the promoter regions of intermediate genes has not been identified. Expression of intermediate genes encoded by transfected DNA in the presence of the DNA replication inhibitor, cytosine arabinoside (AraC), within the infected cell has become the criterion for distinguishing between intermediate and late genes (Keck *et al.*, 1990). The rationale for this distinction is that intermediate genes will be expressed after early gene expression and prior to DNA synthesis if the DNA is naked (Keck *et al.*, 1990). Interestingly, three essential intermediate genes have recently been found to encode late transcription factors (Keck *et al.*, 1990).

At around 4-6 hours post-infection DNA synthesis begins concomitantly with the expression of late genes. The replicating DNA forms large cytoplasmic inclusion bodies called virosomes or virus factories (Morgan, 1976; Moss, 1985). Late genes can be distinguished from early and intermediate genes not only by the time of expression relative to DNA synthesis, but also by a number of structural features. At the 5' end of late gene messages is a capped poly-A head which researchers speculate is

due to polymerase stuttering at distinctive AUUUAC sequences comprising the initiator codon of late gene transcripts (Ink and Pickup, 1990). The lack of discriminant termination in late messages results in transcriptional read-through into the next downstream open reading frame. While the late transcripts are capped, the capping enzyme activity responsible for termination of early messages is either inactive, missing or modified in some fashion which alters recognition of the descriptor sequence UUUUUNU at late times. Late genes encode virion structural proteins and other polypeptides involved in virion morphogenesis and maturation as well as early transcription initiation factors.

Within the virosome maturation of the progeny viruses begins. Lipid membranes form and, with a portion of the virosome inside, fuse together and move away from the virosome. Most of the steps involved in virion morphogenesis are at present unclear. However, viral maturation appears to be a complex biochemical process involving post-translation modification of late gene products by processes such as proteolytic cleavage, phosphorylation, glycosylation, ADP ribosylation and acylation (reviewed in VanSlyke and Hruby, 1990).

The discovery that the late gene transcription factors are products of intermediate genes has greatly aided to clarify the mechanism which regulates the switch from early to late gene expression. One can begin to visualize a cascade of temporal gene expression starting at infection, where transcription of early genes is initiated by the late gene-encoded early transcription factor(s) packaged within the virion. Synthesis of early gene products supposedly then supplies intermediate gene transcription initiation factors required for the production of intermediate gene products. Finally, late gene products are synthesized including early

transcription factors, which are subsequently packaged into virions (Wilson *et al.*, 1988).

Thymidine Kinase Regulation

One of the genes which has been under study in our laboratory for a number of years is that of thymidine kinase (tk, gene; TK, enzyme). Initially chosen as a model gene with which to study gene regulation, thymidine kinase was selected for a number of reasons and, as these studies have expanded and diversified to include structural and functional dissection of TK, so also have the reasons to study this enzyme. Most important, from the perspective of studying gene regulation, TK is regulated at a variety of levels; transcription, mRNA stability, translation (Hruby and Ball, 1981a) and feedback inhibition by dTDP and dTTP (Hruby, 1985). The thymidine kinase gene has been mapped to the 4.8kb HindIII J fragment of the vaccinia virus genome (Hruby and Ball, 1982). It is designated J2R, denoting that it is the second open reading frame in the *Hin*dIII J fragment and reads in the rightward direction. The 5' untranslated end of the tk transcript is relatively short at 6bp in length and is capped with mGpppG at the terminal nucleotide (Weir and Moss, 1983a). At the end of the 3' untranslated region is a polyadenylylated tail.

Upon infection tk message is actively transcribed and has a half life of approximately 8 hours. At 4-6 hours post-infection, concomitant with the onset of DNA replication, the level of distinct tk messages appears to decline and the half life diminishes to about 2 hours (Hruby and Ball, 1981a). Some caution must be exercised when defining levels of early transcripts at late times primarily due to extensive read-through from upstream late genes.

In the past TK protein levels had been deduced from the relative amount of enzyme activity. These studies indicated that TK activity increased steadily after infection until 4-6 hours post-infection. At that point no further increase in activity was detected (Hruby and Ball, 1981a). As an explanation for this it was proposed that since tk mRNA was not synthesized as rapidly at late times as early in the infectious cycle in conjunction with the dramatic drop in message half life, supposedly the plateau of enzyme activity was directly related to reduced message availability (Hruby and Ball, 1981a). Because hybridization experiments demonstrated that tk message at 4 hours post-infection was fully functional but that enzyme activity appeared to be switched off, this was taken as evidence that VV was exerting some kind of translational control (Bajszar *et al.*, 1983). However, we have recently demonstrated by immunoprecipitation of L-[^{35}S] methionine-labeled infected cell extracts that translation of VVTK continues to occur throughout the viral life cycle, even at late times in infection (Black *et al.*, 1991b). Therefore, the supposed translational control that earlier investigators had reported was probably an indirect measurement of dTTP levels.

Thymidine Kinase Function

Thymidine kinase is a key enzyme in the salvage pathway of nucleotide metabolism, catalyzing the production of dTMP from thymidine and ATP. Because dTMP is a precursor for DNA synthesis, TK plays an important role in nucleotide metabolism. dTMP can also be synthesized by the *de novo* pathway by thymidylate synthase. Frearson *et al.* (1965) have reported approximately equivalent amounts of dTMP production from both salvage and *de novo* pathway enzymes. In

eukaryotic cells the levels of cellular TK and many other enzymes involved in DNA precursor synthesis are preferentially expressed during the replicative phase of the cell cycle (Merrill *et al.*, 1984). Because enzymes involved in nucleotide metabolism are required for the production of DNA precursors, their expression just prior to and during DNA synthesis in the S phase is timely. This is especially relevant in the development of chemotherapeutic agents since tumor cells are in a highly proliferative state requiring that nucleotide metabolizing enzymes be present at high levels within the tumor cell (Mathews and van Holde, 1990). To combat the growth of tumor cells, a number of nucleoside analogs have been developed to target these enzymes (Hirsch and Kaplan, 1990).

In the same general fashion as cell cycling, vaccinia virus has evolved to express genes whose products are required for DNA precursor biosynthesis prior to DNA replication during the viral life cycle. As one might expect, the virus-encoded DNA polymerase (Jones and Moss, 1985), ribonucleotide reductase (Slabaugh *et al.*, 1984; Tengelsen *et al.*, 1988), thymidine kinase (McAuslan, 1963) and thymidylate kinase (Smith *et al.*, 1989) have all been shown to be early genes whose gene products are enzymatically active at the onset of DNA synthesis. The role of the VVTK in supporting DNA replication and nucleotide pool levels is of particular interest since vaccinia virus apparently does not encode the gene for thymidylate synthase (M. Slabaugh, personal communication) and therefore is unable to make dTMP via the *de novo* pathway.

While the VV thymidine kinase has been shown to be non-essential in tissue culture, TK^- viruses have an attenuated phenotype in animal models (Buller *et al.*, 1985; Child *et al.*, 1990). Non-essentiality of this gene and the reduced virulence of the virus in the absence of this gene has led to

its development as a target locus for recombinant viruses. The fact that thymidine kinase has been the primary target for the introduction of foreign genes in the production of recombinant vaccines is perhaps also due to the ease of selecting TK-negative viruses in Ltk$^-$ infected cells (Moss, 1990). Selection of recombinant viruses is facilitated by the use of bromodeoxyuridine (BUdR), a thymidine analog. In TK positive viruses, BUdR is phosphorylated initially by TK and then by other kinases to its triphosphate form for incorporation into DNA (Mathews and van Holde, 1990). While the size of the brome atom, comparable to that of the methyl group, makes BUdR function primarily as a thymidine analog, the electronegativity of bromine causes BUdR to transiently resemble deoxycytidine in its base-pairing properties. If this occurs at the time of replication an incorrect base (AT-GC transition) may be incorporated into the newly synthesized strand resulting in a high mutation rate and eventually lead to lethality. Conversely, a TK-positive virus can be selected for by the addition of the folate analog, methotrexate to the culture medium. By inhibiting dihydrofolate reductase, methotrexate interferes with the cellular *de novo* synthesis of dTMP to prevent any rescue from background sources (Mathews and van Holde, 1990). Selection of TK-positive viruses is advantageous when mutant versions of the TK gene are introduced into the viral genome to determine what effect(s) specific mutations have *in vivo*.

The importance of thymidine kinase within the infected cell has been exploited by virtue of its use as a target for antiviral drugs (Hirsch and Kaplan, 1990). Most notably these antiviral drugs have been developed to specifically inhibit replication in members of the *Herpesviridae* family. Acyclovir and gancyclovir are two of the most commonly used anti-

herpetic drugs. These are both guanosine analogs with acyclic sugar groups which, when their triphosphate form is incorporated into DNA, act as chain terminators, thereby inhibiting further extension of the newly synthesized daughter strand (Hirsch and Kaplan, 1990). Before these analogs can be incorporated they must first be initially phosphorylated by the viral thymidine kinase. Recognition and phosphorylation of acyclovir by the herpes thymidine kinases but not by the cellular TK has led to its application as an effective antiviral drug. The key question then arises as to what the difference is between the cellular and the herpes TK enzymes that allows one form to phosphorylate acyclovir while the other cannot. To begin to understand this we need to distinguish these enzymes at the molecular level.

Type I Versus Type II Thymidine Kinases

Thymidine kinases from a wide variety of sources can be divided into two general classes or types (Table I.1). Type I TKs comprise all of the *Herpesviridae* family-encoded enzymes. Type II TKs encompass a large group of enzymes from a wide variety of sources: human, mouse, chicken and members of the *Poxviridae* family. At first glance the genes encoding these two types of TK enzymes are distinct. The type I tk genes encode polypeptides with a predicted molecular weight of around 40kDa while the type II tk genes encode polypeptides ranging in size from 20 to 25kDa. Herpes Simplex Virus (HSV) TK enzymes are active as dimers (Wagner, 1981) yet the type II TKs, VVTK in particular, have been reported to be dimers by some laboratories and tetramers in others (Kit *et al.*, 1977; Hruby and Ball, 1982; reviewed in Traut, 1988). Type II TKs are subject to feedback inhibition by dTDP and dTTP whereas the type I TKs are not (Hruby, 1985).

Table I.1

THYMIDINE KINASE ENZYMES		
Subtype	TYPE I	TYPE II
Monomer Size	40 kDa	20-25 kDa
Active Enzyme	Dimer	Tetramer
Feedback Inhibition	No	Yes
Substrate Specificity	Low	High
Prototype	HSV-1	VV

One of the most distinguishing features between the type I and type II TKs is that of substrate specificity. HSV TK is capable not only of phosphorylating thymidine but also deoxycytidine and therefore is more properly considered a deoxypyrimidine kinase (Robertson and Whalley, 1988). They also have the additional activity of a thymidylate kinase, being able to phosphorylate dTMP to dTDP. Type II TKs, on the other hand, are strict thymidine kinases. This difference in substrate specificity has allowed for the development of anti-herpetic drugs such as acyclovir and gancyclovir (Hirsch and Kaplan, 1990). By understanding the structure-function relationship of VVTK may allow one to design better anti-herpetic drugs with more specificity to Herpes TK and with less toxicity to humans. Because of the close similarity between the VV- and the human-encoded TK, structural and functional dissection of the vaccinia virus thymidine kinase has particular relevance both to the design and development of chemotherapeutic agents and antiherpetic drugs.

Objectives

The structural and functional dissection studies of the vaccinia virus thymidine kinase were initiated to identify and characterize the essential residues or domains of VVTK involved in catalysis and regulation. The specific objectives of this work were to: 1) Define the native molecular weight and composition of VVTK because of the discrepancies in the literature regarding the quaternary structure of this enzyme; 2) Identify and characterize the ATP binding domain of VVTK; 3) Isolate and sequence the *E. coli tdk* gene to serve as a bacterial TK with which to compare VVTK to as a means of identifying conserved motifs and; 4) Provide preliminary evidence regarding the identification and

characterization of domains or residues involved in thymidine binding, divalent cation binding and the sites of feedback inhibition.

CHAPTER II

QUATERNARY STRUCTURE OF VACCINIA VIRUS THYMIDINE KINASE

Authors: Margaret E. Black and Dennis E. Hruby

SUMMARY

Thymidine kinase enzymes isolated from a variety of sources are generally considered to have a native molecular weight of 80-90 kDa composed of two 40-45 kDa subunits. Although these parameters may accurately describe the atypical deoxypyrimidine kinases expressed by members of the *Herpesviridae*, the nucleotide sequences of thymidine kinase genes isolated from human, mouse, chicken and a variety of poxviruses (vaccinia virus, monkeypox virus, variola virus, fowlpox virus and capripoxvirus) predict molecular weights on the order of 20-25 kDa for the derived primary translation products. To resolve this apparent dilemma, velocity sedimentation centrifugation, gel filtration chromatography and protein cross-linking procedures were employed to provide experimental evidence that enzymatically-active vaccinia virus thymidine kinase is a homotetrameric complex of 20 kDa monomers, with a native M_r of 80 kDa.

INTRODUCTION

Thymidine kinase (TK, EC 2.7.1.21) is a central enzyme in the nucleotide salvage pathway. The importance of this enzyme is underscored by its conservation throughout a divergent group of organisms including human (Flemington *et al.*, 1987), mouse (Lin *et al.*, 1985), chicken (Kwoh and Engler, 1984), herpesviruses (Nunberg *et al.*, 1989), bacteriophage T4 (Valerie *et al.*, 1986) and a large number of poxviruses [vaccinia virus (Weir and Moss, 1983a), variola, monkeypox (Esposito and Knight, 1984), fowlpox virus (Boyles *et al.*, 1987) and capripoxvirus (Gershon and Black, 1989)]. Although TK enzymes from a variety of sources have been subjected to detailed biochemical analyses, major discrepancies exist in both the native molecular weights and the subunit compositions which have been reported for these enzymes. Most of the information available in the literature suggests that functional TK enzymes have a native molecular weight of 80-90 kDa composed of dimers of identical 40-45 kDa subunits (Traut, 1988). In contrast, based on the size of the predicted open reading frame of the vaccinia virus TK gene (20 kDa), Hruby and Ball (1982) suggested that the virus-encoded enzyme functions as a tetrameric complex. In order to resolve this issue, the experiments reported in this paper have used velocity sedimentation, protein cross-linking and molecular sieving procedures to provide evidence that VV TK (and by analogy most other TK enzymes) functions as a complex of four 20 kDa subunits.

EXPERIMENTAL PROCEDURES

Expression and analysis of VV thymidine kinase in vitro.

Biologically active VV TK transcripts were produced by transcribing linearized pT7:TKII plasmid DNA with T7 RNA polymerase using conditions previously described by Wilson *et al.* (1989). The capped TK transcripts were translated into protein, in the presence or absence of L-[^{35}S]methionine by using an mRNA-dependent rabbit reticulocyte lysate (Hruby and Ball, 1981b). The synthesis of VV TK was monitored either by the incorporation of L-[^{35}S]methionine into trichloroacetic acid precipitable material or, in the case of unlabeled TK, by measuring the ability of the extract to convert [^{3}H]thymidine to dTMP using a filter binding assay (Hruby and Ball, 1981a).

Glycerol gradient sedimentation.

Radiolabeled or unlabeled pT7:TKII RNA translation products were subjected to velocity sedimentation in 10 to 30% glycerol gradients prepared in TMDT buffer [20 mM Tris-Cl (pH 7.6), 2 mM magnesium acetate, 1 mM dithiothreitol and 50 μM thymidine] in a Beckman SW50.1 rotor at 37,000 rpm for 20 hours at 4°C.

Determination of molecular weight by gel filtration.

A standard curve of protein molecular weights was generated by passing a mixture of protein gel filtration molecular weight markers [1 mg/ml each of ß-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa) and cytochrome c (12.4 kDa),

(MW-GF-200 kit from Sigma, St. Louis, MO.)] through a 1 x 34 cm Sephacryl-300 (Pharmacia LKB Biotechnology, Uppsala, Sweden) gel filtration column equilibrated with TENDT buffer [50 mM Tris (pH 7.6), 1 mM EDTA, 100 mM NaCl, 1 mM dithiothreitol, and 0.015% Triton X-100]. Fractions (250 µl) of the eluate were collected (100 fractions, 25 ml total volume) and alternate samples from number 30 to 100 were subjected to electrophoresis in 12% polyacrylamide gels containing SDS (Studier, 1973). The gels were then stained with Coomassie Brilliant Blue R-250 to visualize the individual proteins. VV TK synthesized *in vitro* in the presence or absence of L-[^{35}S]methionine was chromatographed under identical conditions. The fractions containing VV TK were identified either on the basis of radioactivity or enzyme activity.

Chemical cross-linking of glycerol gradient-purified TK.

VV TK which had been synthesized *in vitro* in the presence of L-[^{35}S]methionine was partially purified by glycerol gradient centrifugation. Aliquots of the peak fraction containing radiolabeled TK were incubated in the presence of 0.01%, 0.05%, 0.1% or 0.5% of glutaraldehyde for one hour (Sigma, St. Louis, MO.) at room temperature. Gel loading buffer was added, the samples heat denatured for 3 min at 100°C and then subjected to electrophoresis in 12% polyacrylamide gels containing SDS (Studier, 1973) followed by autoradiography.

RESULTS AND DISCUSSION

The transcription vector, pT7:TKII (Wilson *et al.*, 1989), was used to produce VV TK mRNA which was translated in a reticulocyte lysate in the presence or absence of L-[^{35}S]methionine to produce enzymatically-active VV TK (Hruby and Ball, 1981b). Radiolabeled and unlabeled VV TK enzymes produced *in vitro* were subjected to velocity sedimentation through glycerol gradients in a buffer designed to maintain enzyme activity. From the data shown in Figure II.1 it is evident that the peaks of radioactivity and enzyme activity are coincident. Furthermore, analysis of the peak fractions of radiolabeled and unlabeled TK enzyme by polyacrylamide gel electrophoresis and autoradiography (top inset of Figure II.1) or immunoblot analysis (bottom inset of Figure II.1), respectively, provided evidence that a 20 kDa protein which was immunoreactive with anti-TK serum co-sediments with enzyme activity. Relative to the internal 64.5 kDa hemoglobin marker in the gradient, the VV TK enzyme complex sediments faster with an apparent molecular weight of approximately 80 kDa, suggesting the complex has a subunit stochiometry of four.

As an alternative means to estimate the native molecular weight of the VV TK enzyme, VV TK synthesized *in vitro* was subjected to gel filtration chromatography through a Sephacryl-300 matrix (Figure II.2). Fractions containing VV TK were identified on the basis of enzyme activity or radioactivity. In both cases, the VV TK enzyme complex chromatographed, eluted in a peak centered around fraction 51. By reference to a standard curve generated by using a mixture of protein gel filtration molecular weight markers, an apparent molecular weight of 80

Fig. II.1 Glycerol gradient separation of VV TK enzyme activity and radiolabeled protein. pT7:TKII-derived capped VV TK RNA was translated in the presence or absence of L-[^{35}S]methionine. The translation products were then subjected to velocity sedimentation in 10 to 30% glycerol gradients as described in "Experimental Procedures". Fractions were collected (200µl) from the bottom and assayed for hot TCA-precipitable radioactivity (white squares) or for enzyme activity (black squares). The sedimentation position of hemoglobin (64.5 kDa) is indicated by the arrow. Radiolabeled cell-free translation products (TK) as well as every other fraction (lanes 1-23) from the glycerol gradient were subjected to discontinuous gel electrophoresis in a 12% polyacrylamide gel (Studier, 1973) and autoradiography (top inset panel). A duplicate gel was subjected to immunoblot analysis using polyclonal antiserum to a *trp*E-TK fusion protein (Black *et al.*, (1991b), bottom inset panel).

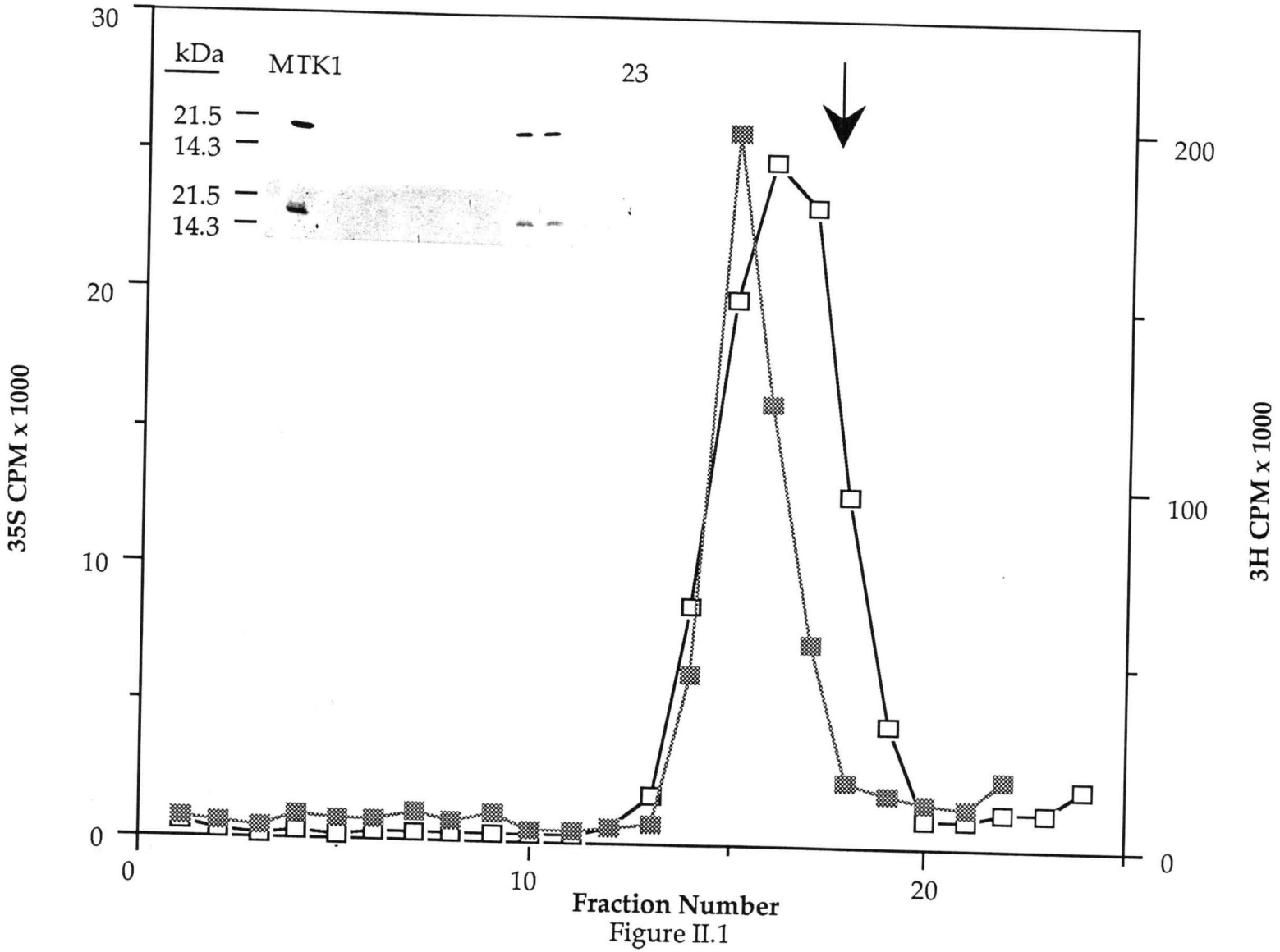
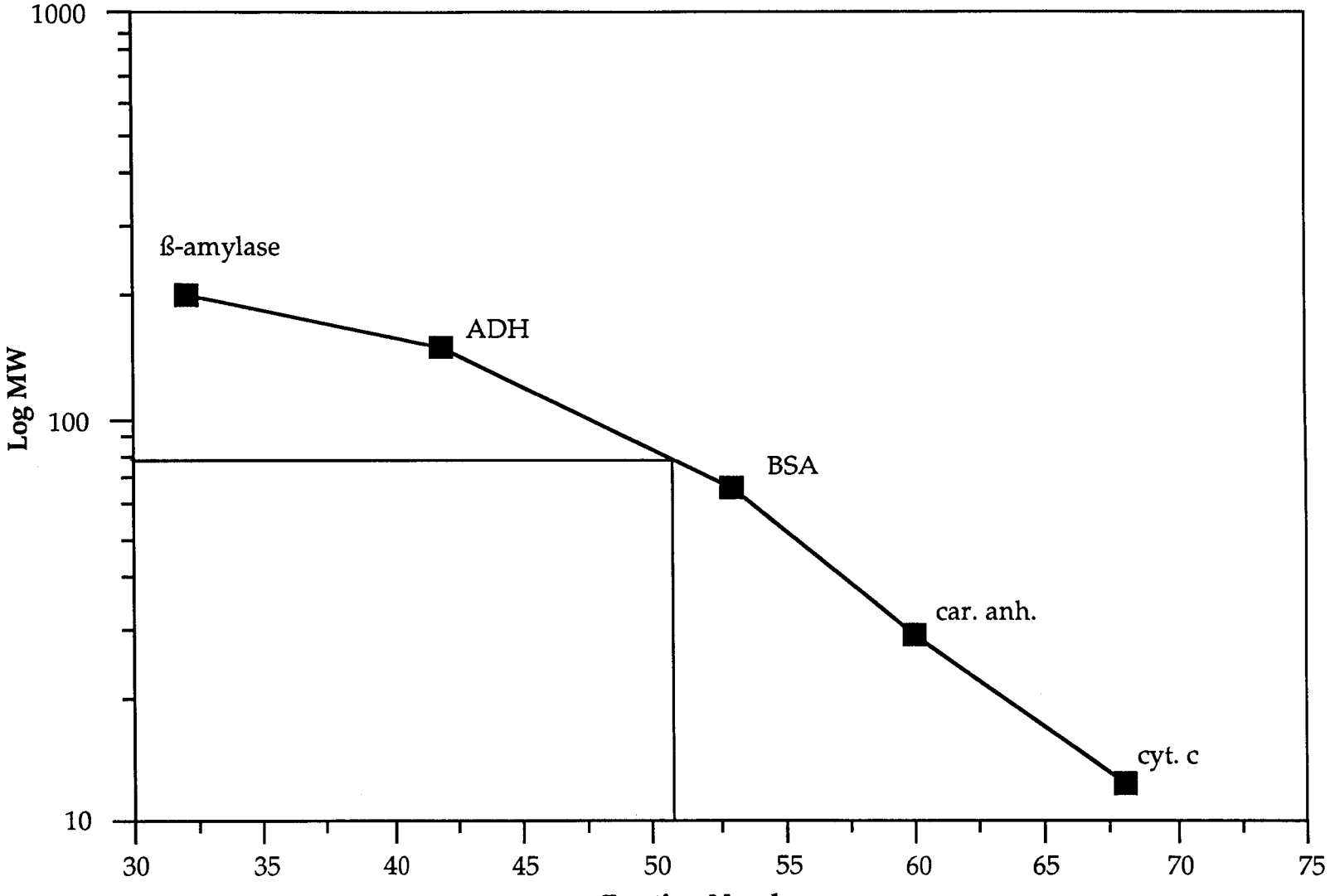

kDa
MTK1
23
21.5
14.3
21.5
14.3
30
20
10
0
0
10
20
200
100
0
35S CPM x 1000
3H CPM x 1000
Fraction Number

Figure II.1

Fig. II.2 Sephacryl-300 gel filtration of TK for native molecular weight determination. A standard curve of protein molecular weights was generated by passing a mixture of protein gel filtration molecular weight markers [ß-amylase (200 kDa), alcohol dehydrogenase (ADH, 150 kDa), bovine serum albumin (BSA, 66 kDa), carbonic anhydrase (car. anh., 29 kDa) and cytochrome c (cyt. c, 12.4 kDa)] through a 1 x 34 cm Sephacryl-300 gel filtration column as described in the "Experimental Procedures". VV TK synthesized *in vitro* in the presence or absence of L-[^{35}S]methionine was chromatographed under identical conditions. The fractions containing VV TK were identified either on the basis of radioactivity or enzyme activity. The thin line corresponds to the chromatographic position of the VV TK complex and its estimated molecular weight.


1000
100
10
Log MW
ß-amylase
ADH
BSA
car. anh.
cyt. c
30
35
40
45
50
55
60
65
70
75
Fraction Number
Figure II.2

kDa for the functional VV TK enzyme was obtained. This is in close agreement with the results of the velocity sedimentation analysis shown in Figure II.1.

There are two potential subunit structures which would explain the apparent 80 kDa molecular weight of the VV TK enzyme. First, and most likely, is that the enzyme is a tetramer of four identical virus-encoded 20 kDa monomers. Alternatively, one or more 20 kDa proteins could associate with a cellular protein present in both the infected cell and the reticulocyte lysate to form the active enzyme complex. As an approach to distinguish between these two possibilities, protein cross-linking procedures were employed. Radiolabeled VV TK was synthesized *in vitro* and partially purified by velocity sedimentation in a glycerol gradient as described in Figure II.1. Portions of the peak fraction were incubated at room temperature in the presence of increasing concentrations of the protein cross-linking agent, glutaraldehyde. After cross-linking, the reaction products were analyzed by polyacrylamide gel electrophoresis (Figure II.3). In the absence of glutaraldehyde, the only protein species evident is the 20 kDa monomer. With increasing concentrations of glutaraldehyde the relative intensity of the 20 kDa protein band decreases and new major (40 kDa and 80 kDa) and minor (60 kDa and 100 kDa) cross-linked products become evident. An immunoblot of a duplicate gel indicated that each of the new bands contain proteins which are immunoreactive with anti-TK serum (data not shown). Since the cross-linked species increase by discrete 20 kDa increments, it is most likely that the 80 kDa VV TK enzyme is a homotetramer and not associated with any other specific ancillary factors. Based solely on amino acid sequence comparisons and substrate specificities it appears that two classes of TK

Fig. II.3 Glutaraldehyde cross-linking of glycerol gradient-purified TK. L-[^{35}S]methionine-labeled TK was partially purified by glycerol gradient centrifugation. Aliquots of the peak fraction containing radiolabeled TK were incubated in the presence of the indicated concentrations of glutaraldehyde (0.01%, 0.05%, 0.1% or 0.5%) for one hour at room temperature. Gel loading buffer was added, the samples heat denatured for 3 min at 100°C and then analyzed by electrophoresis in 12% polyacrylamide gels containing SDS (Studier, 1973), followed by autoradiography. The positions and sizes of protein molecular weight standards (MWM) included on the gel are indicated on the left.

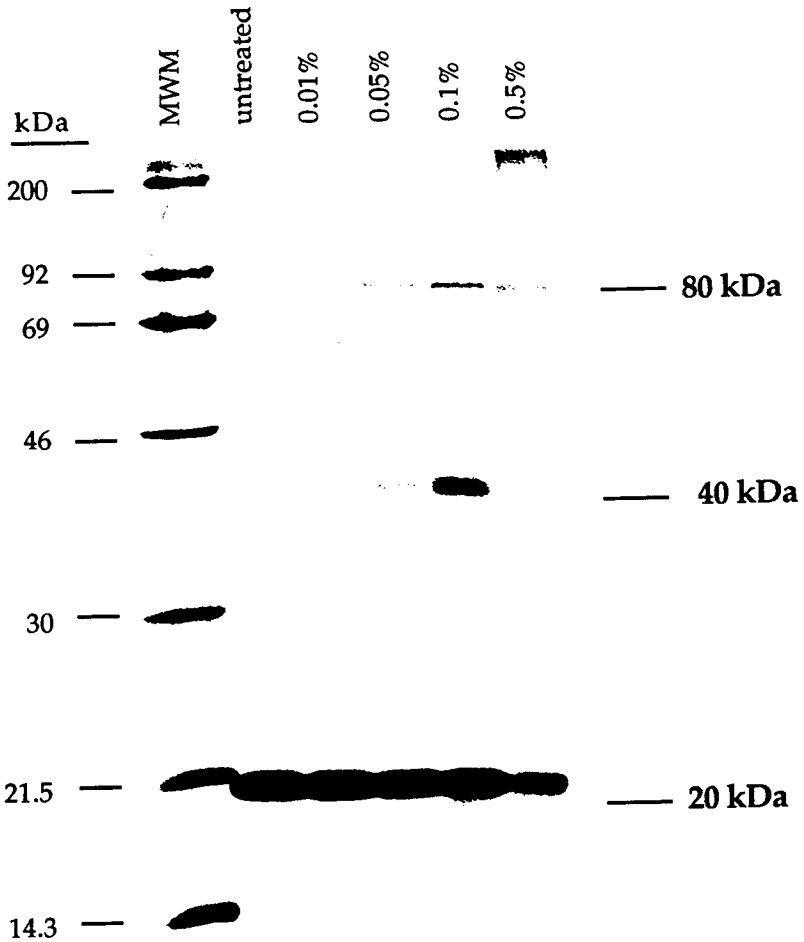

kDa
MWM
untreated
0.01%
0.05%
0.1%
0.5%
200
92
69
46
30
21.5
14.3
80 kDa
40 kDa
20 kDa

Figure II.3

enzymes exist. The first group (type I) contains the Herpes Simplex Virus-like TK enzymes [Herpes Simplex Virus (Wagner *et al.*, 1981), Epstein-Barr Virus (Littler and Arrand, 1988), Feline Herpes Virus (Nunberg *et al.*, 1989), Marmoset Herpes Virus, (Otsuka and Kit, 1984) and Varicella-Zoster Virus (Davison and Scott, 1986)]. The second group (type II), which is quite distinct from HSV-like enzymes, is comprised of thymidine kinases similar to that of VV (Weir and Moss, 1983), namely those expressed by monkeypox (Esposito and Knight, 1984), variola (Esposito and Knight, 1984), fowlpox (Boyles *et al.*, 1987), capripox (Gershon and Black, 1989), mouse (Lin *et al.*, 1985) chicken (Kwoh and Engler, 1984), human (Flemington *et al.*, 1987), bacteriophage T4 (Valerie *et al.*, 1986) and *Escherichia coli* (Black and Hruby, 1991a). Type I TK genes encode an open reading frame with a predicted molecular weight of approximately 40 kDa and the encoded enzyme behaves as a dimer with a molecular weight of 80 kDa (Wagner *et al.*, 1981) Initial studies of the type II prototype TK enzyme encoded by VV suggested that it was also a dimer of 40 kDa subunits (Kit *et al.*, 1977). However, when the VV TK gene was mapped and sequenced (Weir and Moss, 1983a) the nucleotide sequence predicted a protein with a molecular weight of 20 kDa. Since, like the type I enzymes, the functional VV TK enzyme had an apparent native molecular weight of 80 kDa (Hruby and Ball, 1985), this suggested that type II enzymes must function as a tetrameric complex. This hypothesis has been directly tested in the experiments reported in this paper. The results obtained by three independent lines of investigation are all consistent with the notion that VV TK is active as a homotetrameric complex. Given the high degree of amino acid sequence conservation between the different members of the type II TK enzymes (though less so for bacteriophage T4 and *E. coli* TK

enzymes) it is likely that all of these enzymes will function as tetramers and not dimers. This theory is supported by biochemical studies on another type II TK enzyme, namely the human TK (Sherely & Kelly, 1988), purified from tissue culture cells.

As more detailed structural information (e.g., X-ray crystallographic data) becomes available on different TK enzymes it will be of great interest to determine whether the dimeric type I and tetrameric type II enzymes assemble into similar or dissimilar three dimensional structures for binding their substrates and co-factors in order to carry out their catalytic and regulatory functions.

CHAPTER III

IDENTIFICATION OF THE ATP-BINDING SITE OF VACCINIA VIRUS THYMIDINE KINASE

Authors: Margaret E. Black and Dennis E. Hruby

SUMMARY

Although small in size (20 kDa), the vaccinia virus (VV) thymidine kinase protein (TK, EC 2.7.1.21) is a relatively complex enzyme which must contain domains involved in binding both substrates (ATP and thymidine) and a feedback inhibitor (dTTP), as well as sequences directing the association of individual protein monomers into a functional tetrameric enzyme. Alignment of predicted amino acid sequences of the thymidine kinase genes (tk) from a variety of sources was used to identify highly conserved regions as a first step towards locating potential regions housing essential domains. A conserved domain (domain I) near the amino terminus of VV TK protein had characteristics consistent with a nucleotide binding site. Analysis of the nucleotide substrate specificity of VV TK indicated that ATP acts as the major phosphate donor for thymidine phosphorylation while GTP, CTP and UTP were inefficient substrates. Site-directed mutagenesis was performed on domain I to generate eleven mutant enzymes. Comparison of the wild type and mutant proteins with regard to enzyme activity revealed that two of the mutant enzymes, T18 and S19, exhibited enhanced enzyme activity (3.73-fold and 1.35-fold, respectively) relative to the control. The other mutations introduced led to greatly reduced levels of enzyme activity which correlated with a reduced or altered ability of the mutant enzymes to bind ATP as determined by ATP-agarose affinity chromatography. Wild type VV TK bound to an ATP affinity column could also be eluted with dTTP. Glycerol gradient separation of wild type TK in the presence or absence of dTTP indicated that dissociation of the tetrameric complex was not the means by which

enzymatic inhibition was achieved. Taken together, these results suggest that (i) domain I (amino acids 11-22) of the VV TK corresponds to the ATP binding site, and (ii) that dTTP is able to interfere with ATP binding, either directly or indirectly, and thereby inhibit enzymatic activity without dissociating the native enzyme.

INTRODUCTION

Thymidine kinase (TK) plays a central role in the nucleotide salvage pathway by catalyzing the production of dTMP from thymidine and ATP. The importance of this enzyme is evidenced by its expression in a wide variety of organisms including man (Ives *et al.*, 1963), mouse (Littlefield, 1966), chicken (Stubblefield and Murphree, 1967) and a number of viruses including vaccinia virus (McAuslan and Joklik, 1962). The vaccinia virus (VV) tk gene has been mapped to a 531 bp open reading frame within the 4.8 kbp HindIII J fragment of the VV genome (Hruby and Ball, 1982). Tk mRNA is capped and polyadenylylated and is predicted to encode a 177 amino acid polypeptide which has an apparent molecular weight of about 19 kDa on denaturing polyacrylamide gels (Hruby *et al.*, 1983). Previous studies have suggested that the VV TK enzyme is active as an 80 kDa tetrameric complex whose activity is allosterically regulated by feedback inhibition with high levels of dTTP or dTDP (Hruby, 1985; Wilson *et al.*, 1989).

At the amino acid level VV TK shares extensive homology with the predicted amino acid sequences of variola virus (Esposito and Knight, 1984), monkeypox (Esposito and Knight, 1984), mouse (Lin *et al.*, 1985), chicken (Kwoh and Engler, 1984) and human (Flemington *et al.*, 1987) thymidine kinases. Furthermore these enzymes are likely to behave in a similar manner to VV TK based on recent work showing that the human cytosolic TK has a native molecular weight of 96 kDa most likely suggesting a tetrameric structure composed of 24 kDa monomers (Sherley and Kelly, 1988). While none of the catalytic or regulatory domains of the thymidine

kinase enzyme have been formally identified in any member of this group, the extent of amino acid sequence homology between members implies that they possess functional domains in common.

The expression of VV tk is modulated at the transcriptional and post-translational (mRNA stability and feedback inhibition) levels (Hruby and Ball, 1981a, Hruby, 1985, Wilson *et al.*, 1989). Along with an interest in a more detailed study of the mechanisms of post-translational regulation, the identification of the amino acid sequences or secondary structures within the TK polypeptide responsible for catalytic activity, substrate binding, feedback inhibition and subunit interaction are of interest to provide insight into the various structure-function relationships of this enzyme. An *in vitro* transcription system has been developed to generate authentic tk mRNA which can be translated *in vitro* to produce enzymatically active TK (Wilson *et al.*, 1989). This system was used to dissect various domains of the protein in order to address the role these domains play in catalysis and/or regulation as a step towards understanding post-translational expression of vaccinia virus proteins. In the present study, *in vitro* expression of wild type and mutant VV TK enzymes produced by site-directed mutagenesis has been used to identify the ATP binding domain of TK, to determine the specificity of TK for ATP and to analyze the effect of dTTP on the ability of the enzyme to bind ATP.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions.

E. coli strains CJ236 (dut$^-$, ung$^-$) and MV1190 were used. The plasmid pT7:TKII, which contains the T7 RNA polymerase promoter abutted to the VV tk open reading frame in a pUC118 plasmid background (Wilson *et al.*, 1989), was transformed into competent CJ236 cells, which were plated on Luria Bertani (LB) agarose plates containing 75 μg/ml ampicillin (Maniatis *et al.*, 1982). To produce single-stranded plasmid DNA template for mutagenesis, 5 ml of LB containing 75 μg/ml ampicillin was inoculated with a single colony of pT7:TKII (CJ236) and 5 μl of M13K07 helper phage stock. This culture was allowed to grow for 16-18 h before single stranded DNA was isolated.

Computer Analysis.

Deduced protein sequences of VV (Weir and Moss, 1983), variola virus (Esposito and Knight, 1984), monkeypox (Esposito and Knight, 1984), chicken (Kwoh and Engler, 1984), human (Flemington *et al.*, 1987) and mouse (Lin *et al.*, 1985) thymidine kinases were obtained from the Protein Identification Resource bank. Analysis of VV TK, protein sequence comparisons and alignments were prepared using the PEP, QUEST and GENALIGN programs of the IntelliGenetics system.

In Vitro Transcription and Translation Systems.

The method for synthesizing wild type and mutant thymidine kinase enzymes *in vitro* was as previously described (Wilson *et al.*, 1989). In the

present study, all transcripts were capped with m^7GpppG (New England Biolabs) prior to translation in rabbit reticulocyte lysates. Transcription products were analyzed on 1% agarose gels in TBE (89 mM Tris/89 mM borate/2 mM EDTA) buffer, visualized by ethidium bromide staining and quantitated by spectrophotometric measurements (Maniatis *et al.*, 1982). Approximately 0.2 µg of each mutant and wild type tk RNA was used to program a rabbit reticulocyte lysate translation system in the presence of 0.1 mCi/ml L-[^{35}S] methionine (Amersham; 1249 Ci/mmol) in a final reaction volume of 15 µl. To assay TK activity, an identical cell-free translation reaction for each mRNA was performed in the absence of radiolabel. Reaction mixtures were incubated for 90 min at 30°C.

Analysis of Translation Products.

Protein synthesis was measured by spotting 1 µl samples of the reaction mixtures on 2.5 cm Whatman 3MM filter paper discs and determining hot trichloroacetic acid-precipitable radioactivity as measured by liquid scintillation counting in a Beckman LS3801 Scintillation Counter. Radioactively labeled nascent polypeptides were analyzed by electrophoresis on 12% polyacrylamide gels containing SDS (Studier, 1973) and visualized by autoradiography on Kodak X-OMAT film.

An identical SDS:polyacrylamide gel of L-[^{35}S] methionine-labeled translation products was equilibrated in Towbin transfer buffer (methanol/glycine/Tris) and the proteins electroblotted to nitrocellulose at 25V for 40 min at 4°C (Towbin *et al.*, 1979). After transfer the filter was incubated overnight with polyclonal antiserum raised to a *trp*E-TK fusion protein (Black *et al.*, 1991b) followed by incubation with alkaline phosphatase-conjugated goat anti-rabbit antiserum (BioRad), prior to

treatment with color developing reagents according to the manufacturer's directions.

TK activity in rabbit reticulocyte lysate translations was measured as previously described (Hruby and Ball, 1981b) by using a filter binding assay which monitors the conversion of [^{3}H]thymidine to [^{3}H]TMP.

Site-Directed Mutagenesis.

Synthetic deoxyoligonucleotide primers (Applied Biosystems) were purified by electrophoresis in 20% polyacrylamide-urea gels using TBE buffer (Maniatis *et al.*, 1982). After an overnight incubation of the excised gel fragments at 37°C in gel elution buffer (0.5 M NH_4Ac/10 mM $MgAc_2$), the eluate was passed over a Sep-Pak C_{18} column (Millipore, Waters Associates) prepared by passing 10 ml of acetonitrile through the column followed by a 10 ml wash with distilled water. The gel eluate containing the primer was then passed over the column and the column rinsed with 10 ml of distilled water. The primer was eluted off the column with three 1-ml aliquots of 60% methanol. Quantitation of primer concentrations was done by spectrophotometry. The peak fraction was concentrated in a Speed-Vac (Savant) and resuspended in distilled water to a final concentration of 50 pmol/µl. Primers used in this study were **MB1** (5' CATGGGGACGATTATCAAC 3' - lose a HaeIII site); **MB2** (5' CTGTGCTT(T/**A**) TA(C/**A**)CTGAAAAC 3' - lose a RsaI site); **MB3** (5' CTGT**G**CTT**A**TAACTGAAAAC 3' - lose a RsaI site); **MB11** (5' CTTTTACC**NNN**AAACATGGGCCC 3' - create a ApaI site); **MB12** (5' TCCTGTA**G**TTTTACC 3' - lose a RsaI site); and **MB13** (5' CTAATTAATTCT**NN**ACTTTTACC 3' - lose a RsaI site), where N is all four deoxynucleotides and all the mutant bases shown in bold.

The mutagenesis procedure used was a modified version of that described by Kunkel (1985). The standard phosphorylation mixture consisted of 450 pmol of purified primer in kinase buffer [50 mM Tris-HCl (pH 8), 10 mM $MgCl_2$, 2 mM dithiothreitol], 2 mM ATP and 2 units (U) kinase (Boehringer Mannheim) in a 15 µl volume. This mixture was incubated at 37°C for 1 h and then placed on ice. The annealing reaction included ligase buffer [60 mM Tris-HCl (pH 8), 10 mM $MgCl_2$, 10 mM ß-mercaptoethanol], 50 pmol of phosphorylated primer and 150 ng of single stranded template DNA in a 50 µl volume. Single stranded template DNA isolated from pT7:TKII (CJ236) for mutagenesis was prepared as described above. Incubations were carried out at 55°C for 30 min, transferred to 37°C for 15 min followed by incubation for 15 min at room temperature. To the annealing reaction 2.5 mM dNTPs, 20 mM ATP, ligase buffer, 12 U ligase (Boehringer Mannheim) and 5 U polymerase I, large fragment (Klenow) (Boehringer Mannheim) were added. The fill-in/ligation reaction was allowed to incubate at 12°C for at least 2 h. Portions of this reaction mixture were used to transform $CaCl_2$-treated competent MV1190 cells (Maniatis *et al.*, 1982). Plasmid DNA was isolated from transformed cultures by standard methods. Screening transformants was done by cutting isolated plasmid DNA with an appropriate restriction endonuclease as determined by the loss or creation of a restriction site integrated within the mutagenic primer and subjecting the digested fragments to electrophoresis in 1% agarose or 5% polyacrylamide gels using TBE (Maniatis *et al.*, 1982). Single stranded DNA isolated from putative mutant clones was sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977) using the previously described CF-5 primer (Wilson *et al.*, 1989), which is complementary to the coding strand. Once the nature of the

introduced mutations had been determined, mutant plasmids were isolated by alkaline lysis and banded in CsCl density gradients (Maniatis *et al.*, 1982).

Mutant and wild type plasmid DNAs were linearized with BamHI, phenol extracted, ethanol precipitated and resuspended in TE [10 mM Tris (pH 7.2), 1 mM EDTA] buffer. A small fraction of the resuspended DNA was subjected to electrophoresis on 1% agarose (TBE) gels and visualized by ethidium bromide staining to ensure that complete digestion had occurred (Maniatis *et al.*, 1982). Linear plasmid DNAs were then transcribed *in vitro* to generate wild type and mutant capped messages.

ATP Affinity Column.

Adenosine 5' triphosphate agarose (500mg) (attached at C8 with a six carbon spacer, Sigma) was swollen overnight at 4°C in binding buffer [10 mM MOPS, 1 mM EDTA, 10 mM NaF, 1 mM dithiothreitol, 0.1% Triton X-100, 1 mM benzamidine, 50 μM thymidine (pH 7.4)] (Ballou *et al.*, 1989). All of the following procedures were performed at 4°C. The swollen beads were packed into a disposable polypropylene column and rinsed thoroughly with binding buffer. Radiolabeled translation products (10 μl) were diluted into 1 ml of binding buffer and loaded onto the column, which was then washed with two 2.5-ml aliquots of binding buffer. A 0 to 10 mM ATP gradient was used to elute TK from the affinity column. One ml fractions were collected. The column was then rinsed with 2 2.5-ml fractions of binding buffer, pulsed with 1 ml binding buffer containing 100 mM ATP to remove any remaining TK associated with the column. The column was then thoroughly washed with binding buffer and stored at 4°C. For elutions with heterologous deoxynucleotides, 0 to 50 μM gradients were

passed over the column with five additional 1 ml fractions collected after the 100 mM ATP pulse. The amount of radiolabeled protein in 100 µl of each fraction was determined by hot trichloroacetic acid precipitation and scintillation counting.

Glycerol Gradient Fractionation.

Products from *in vitro* translations of wild type tk RNA were layered over 4.8 ml (10-30%) glycerol gradients in TMDT [20 mM Tris-HCl (pH 7.6), 2 mM $MgAc_2$, 1 mM dithiothreitol, 50 µM thymidine] or TMDT containing 50 µM dTTP and centrifuged in a Beckman SW50.1 rotor at 37,000 revolutions/min for 20 h at 4°C. Fractions (200 µl) were collected and assayed (10 µl) for hot trichloroacetic acid precipitable radioactivity.

RESULTS

Amino acid sequence comparison and correlation with reported nucleotide binding site consensus sequences.

The amino acid sequence of VV TK was aligned with corresponding TK sequences from a variety of sources in order to identify conserved regions that might suggest functional domains. As shown in Figure III.1, alignment of the TK sequences reveals 7 domains of homology which are highly conserved among all the enzymes compared. Examination of domain I (amino acid residues 11-18 of VV TK) reveals a strong similarity with purine nucleotide binding site consensus sequences (Fig. III.2). Apparently, the difference in affinity for the specific nucleotide results from the particular amino acid residue in the last position within the consensus sequence. Gly Xaa Yaa Yaa Gly Lys Zaa, where Xaa is any amino acid, Yaa is often a proline or glycine. In the case of GTP/GDP binding proteins, Zaa is a serine whereas in ATP/ADP binding proteins this residue is a threonine (Reinstein *et al.*, 1988). While a single domain is apparently sufficient for binding ATP, two additional domains are thought to be involved in binding GTP (Dever *et al.*, 1987; Lowe *et al.*, 1987; Moeller and Amons, 1985). Domain I of the VV TK appears to be somewhat of a hybrid between the GTP and ATP consensus sequences, suggesting that VV TK might be able to utilize GTP as well as ATP as a phosphate donor. Among thymidine kinase enzymes this is not unprecedented. The herpes simplex virus (HSV) TK enzyme is capable of utilizing all four nucleoside triphosphates as phosphate donors (Littler and Arrand, 1988). Interestingly, the HSV TK nucleotide binding site is completely conserved with the ATP

Fig. III.1 Amino acid sequence of thymidine kinase from vaccinia virus. The predicted amino acid sequence of the VV TK protein is shown by using the standard single letter amino acid code. The numbers refer to residue numbers. From a comparison of thymidine kinase amino acid sequences from vaccinia virus (Weir and Moss, 1983a), variola virus (Esposito and Knight, 1984), monkeypox (Esposito and Knight, 1984), mouse (Lin *et al.*, 1985), chicken (Kwoh and Engler, 1984) and human (Flemington *et al.*, 1987), seven domains (I-VII, boxed and shaded) were identified as having complete identity with VV TK. One of these domains, domain I, corresponds to purine nucleotide binding site consensus sequences. The importance of each of the other domains, II-VII, has not been evaluated. The location of aspartic acid residues (D) which demonstrate periodicity are accentuated by a darkened circle.

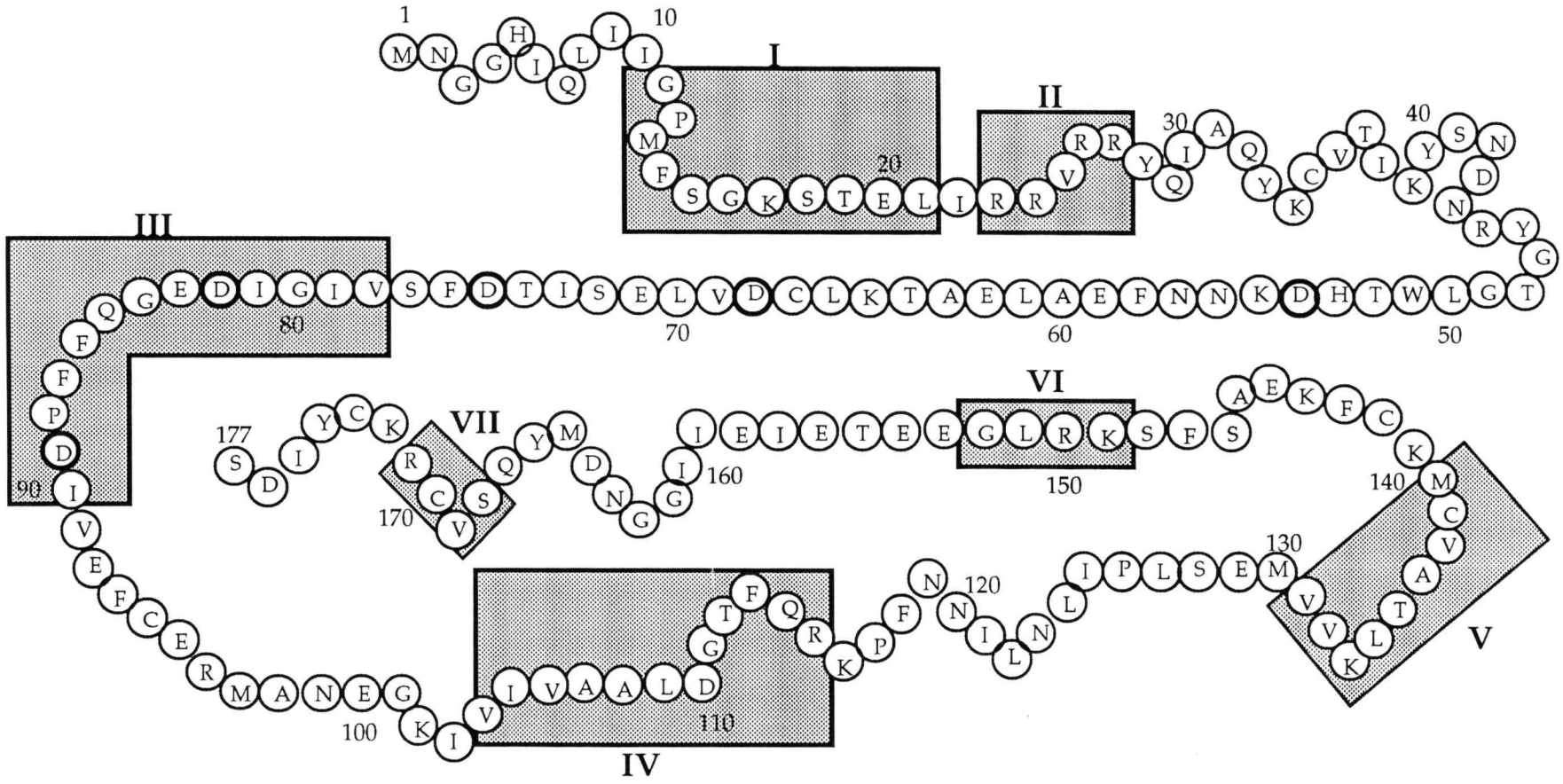
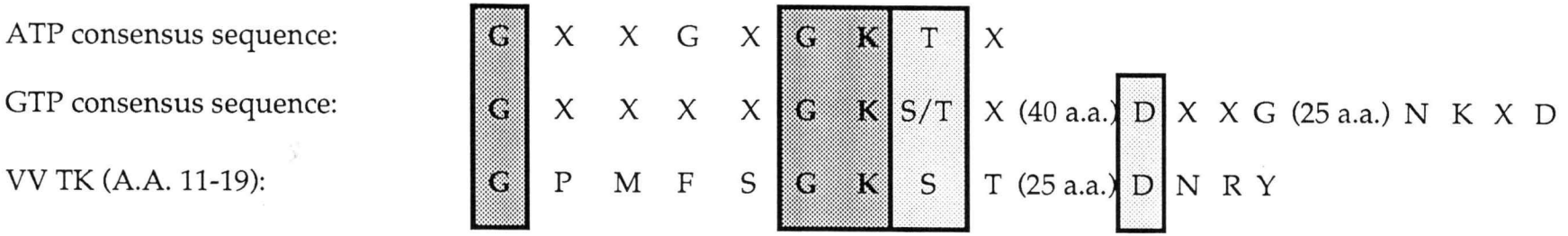

I
II
III
IV
V
VI
VII
1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
177

Figure III.1

Fig. III.2 Alignment of VV TK with ATP and GTP binding site consensus sequences and designation of mutations generated within this site. Comparison of VV TK sequences with ATP or GTP binding site consensus sequences reveals homology to both sequences (darkly shaded residues). The lightly shaded boxes indicate VV TK homology to only one of the nucleotide binding site consensus sequences. The GTP consensus sequence (Liu and Summer, 1988) is comprised of three domains whereas the ATP consensus sequence (Lowe *et al.*, 1987) contains only a single region which to a certain extent both sequences share. At a distinct site distant to this shared region, the VV TK has limited homology with the second domain of the GTP consensus sequence and none with the third region.

Eleven mutations generated in domain I of the VV TK (amino acid residues 11-19) are listed. The nomenclature for the mutant TK enzymes indicates the identity and position of the introduced amino acids. For example, V11 indicates that the eleventh amino acid residue was changed to a valine, P15 indicates that the fifteenth residue was changed to a proline, etc. Standard single letter amino acid code is used where X is any amino acid.



MUTATIONS GENERATED

V11	V								
P15				P					
T15				T					
V16					V				
I17						I			
T18							T		
I19								I	
P19								P	
Q19								Q	
R19								R	
S19								S	

Figure III.2

binding site consensus sequence (Liu and Summers, 1988). Because VV TK shares some homology with both ATP and GTP binding site consensus sequences it was of interest to determine whether this region is indeed the nucleotide binding site, and if so, whether VV TK is able to use GTP as well as ATP as a phosphate donor.

Specificity of the ATP binding site for ribonucleoside triphosphates.

In order to ascertain whether the VV TK is capable of using ribonucleotides other than ATP as phosphate donors, wild type TK synthesized *in vitro* was partially purified by glycerol gradient fractionation to remove any endogenous ribonucleotides (Wilson *et al.*, 1989). The gradient fractions containing VV TK were then used in TK assays containing increasing concentrations of ATP, GTP, CTP or UTP as the potential phosphate donor. A corresponding fraction from a glycerol gradient separation of a cell free translation reaction lacking VV tk mRNA was used as a control in these experiments. As shown in Figure III.3 it is apparent that ATP was used efficiently as a phosphate donor by the VV thymidine kinase *in vitro*. Under the conditions of this assay 5 mM ATP was able to support the maximal levels of enzyme activity. In contrast, GTP, CTP and UTP were not apparently utilized by VV TK *in vitro* even when the concentrations of the nucleotides were raised as high as 25 mM. These results suggested that at least *in vitro*, ATP was the sole phosphate donor used by VV TK to phosphorylate thymidine.

Directed genetics of the putative VV TK ATP binding site.

By using the pT7:TKII vector (Wilson *et al.*, 1989), wild type and mutant thymidine kinase enzymes can be produced *in vitro* transcription

Fig. III.3 Specificity of ribonucleotides to serve as phosphate donors. Partially purified wild type TK translated *in vitro* was assayed for its ability to utilize ATP, GTP, CTP or UTP as the sole phosphate donor for the production of dTMP. As controls water or a fraction from a glycerol gradient separation of a cell free translation carried out in the absence of added RNA which corresponded to the peak TK fraction was used (-RNA). 0, 1, 5, 10 and 25 mM concentrations of each ribonucleotide was assayed for the ability to phosphorylate thymidine.

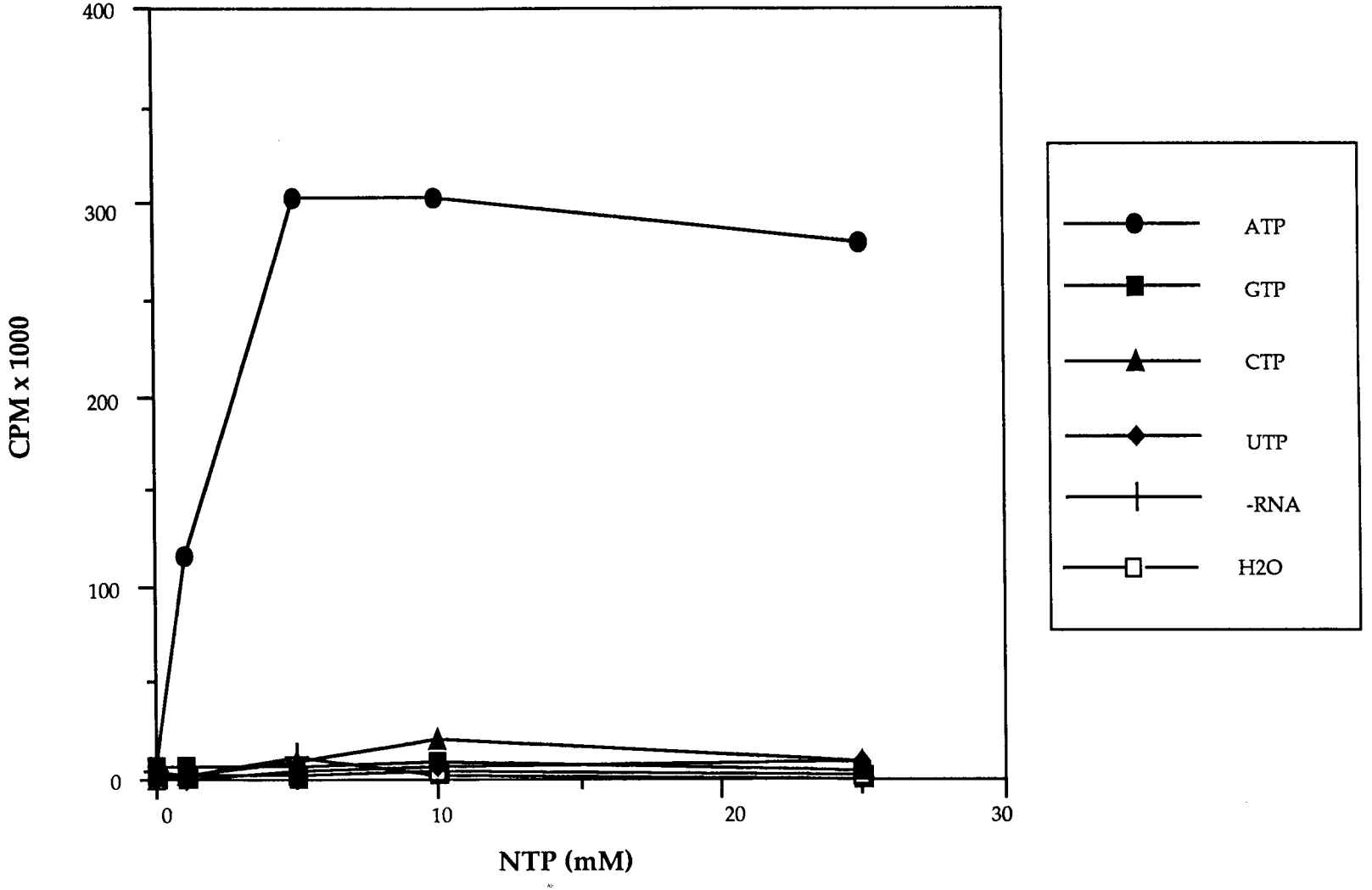

400
300
200
100
0
CPM x 1000
0
10
20
30
NTP (mM)
ATP
GTP
CTP
UTP
-RNA
H2O

Figure III.3

and translation systems. Site-directed mutagenesis was used to introduce a number of mutations within and adjacent to the putative ATP binding site of VV TK. The conserved glycine residues were changed to valine and the conserved lysine to isoleucine primarily because these alterations in analogous positions in the HSV TK ATP binding site had been shown to inactivate the enzyme (Liu and Summers, 1988). The serine to threonine mutation at position 18 was done to make this region more closely resemble the ATP binding site consensus sequence. A number of other mutations were introduced at positions 15 and 19 to analyze the importance of these non-conserved residues. Figure III.2 shows the nature and location of the amino acid substitutions in eleven mutant tk genes. The mutant templates were used to prepare transcripts which were tested for both their ability to serve as a mRNA *in vitro* as well as to direct the synthesis of enzymatically active TK.

The radiolabeled translation products were subjected to SDS polyacrylamide gel electrophoresis in order to determine whether full length 19 kDa polypeptides were being translated (Fig. III.4A). Western blot analysis, using a polyclonal antibody raised to a TK fusion protein (Black *et al.*, 1990a), was performed on a duplicate gel to determine if the major band of 19 kDa proteins was TK (Fig. III.4B). From the data shown in Figure III.5 it is apparent that the major radioactive products from the mutant transcript cell free translations correspond to the predicted molecular mass of wild type TK and are immunoreactive with anti-TK serum (Fig. III.4).

Functional analysis of VV TK ATP binding site mutants.

Analyses of wild type and mutant TK enzymes synthesized *in vitro* are shown in Table III.1 and represent the results of seven independent

Fig. III.4 SDS:PAGE analysis and western blot of wild type and mutant tk translation products. Panel A shows an autoradiograph of L-[^{35}S] methionine-labeled cell free translation products separated by 12% SDS:PAGE. A duplicate gel was used to transfer the proteins to nitrocellulose which was subsequently immunoreacted with a polyclonal antiserum to a *trp*E-TK fusion protein (Panel B). Both panels demonstrate a major radioactive or immunoreactive band at 19 kDa corresponding to the VV TK monomer. MWM corresponds to protein molecular weight marker standards. Equal volumes of translation products from control translations (-RNA) and translations programmed with wild type tk mRNA or the mutant tk mRNAs are represented as noted.

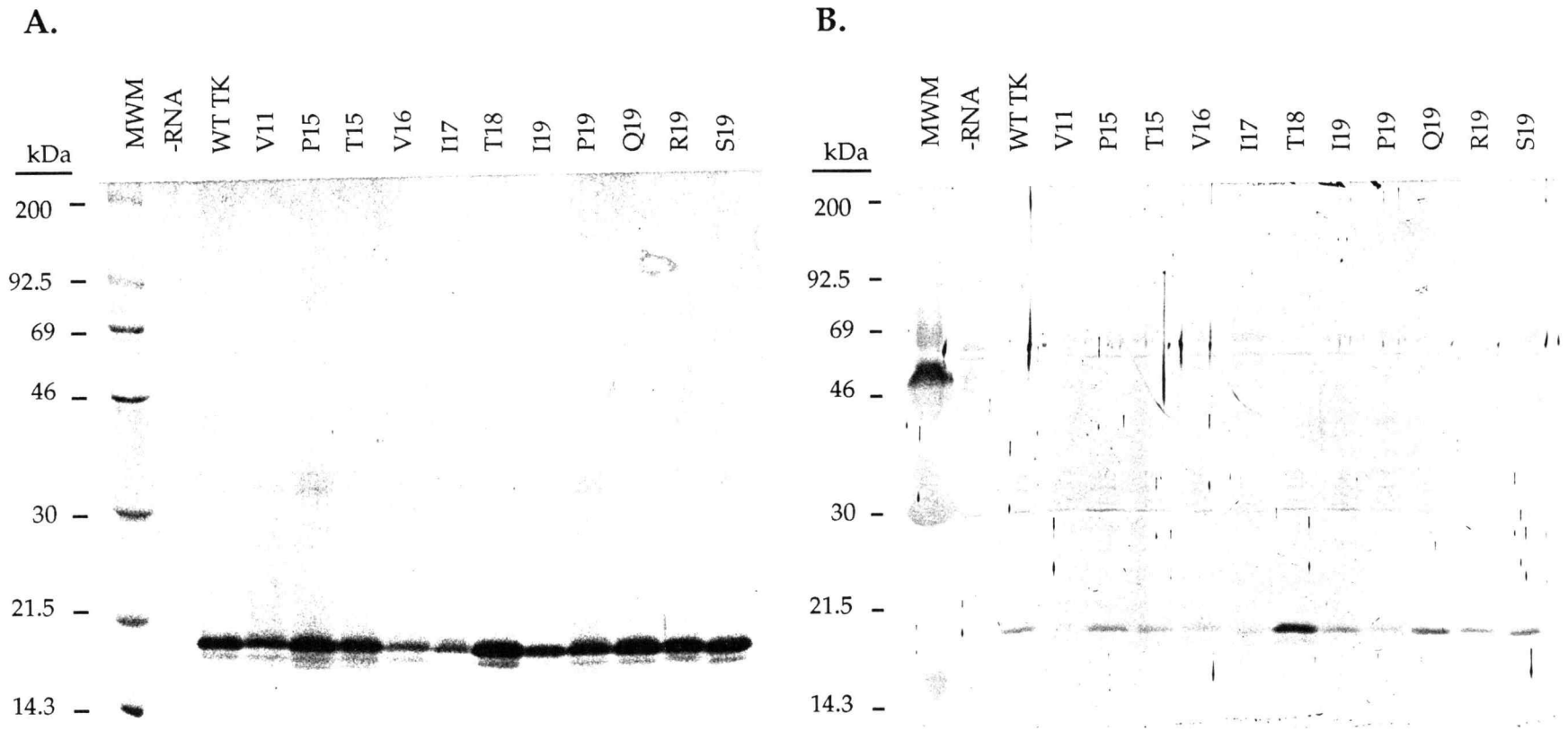

A.
B.
kDa
MWM
-RNA
WT TK
V11
P15
T15
V16
I17
T18
I19
P19
Q19
R19
S19
200
92.5
69
46
30
21.5
14.3

Figure III.4

TABLE III.1

Mutations generated in the putative nucleotide binding site of VV TK

Mutants [a]	WT TK	V11	P15	T15	V16	I17	T18	I19	P19	Q19	R19	S19
Protein [b] Synthesis	100%	49%	68%	83%	46%	56%	100%	65%	86%	86%	90%	100%
TK [c] Activity	100%	26%	14%	17%	19%	17%	373%	17%	10%	28%	27%	135%

[a] Nomenclature for the mutations introduced are as described in Fig. III.2. WT TK is the wild type form of TK.

[b] Hot trichloroacetic acid precipitable counts of radiolabeled cell free translation products listed have been normalized to the level of wild type TK.

[c] Enzyme activity determined was also normalized to wild type TK enzyme levels.

Fig. III.5 Elution of L-[^{35}S] methionine labeled wild type and mutant TKs from an ATP agarose affinity column using ATP gradients. Elution patterns of wild type (WT) and mutant TK enzymes (T18, R19, S19, Q19) are shown. All of the remaining mutant TK enzymes (V11, P15, S15,V16, I17, I19 and P19) demonstrated similar elution patterns which are represented by the pattern of I17 in this graph. The counts per minute (CPM) were adjusted to indicate the percentage of the total amount of CPMs loaded onto the column. For a detailed description of the analysis of wild type and mutant TK enzymes using the ATP affinity column see "Experimental Procedures".

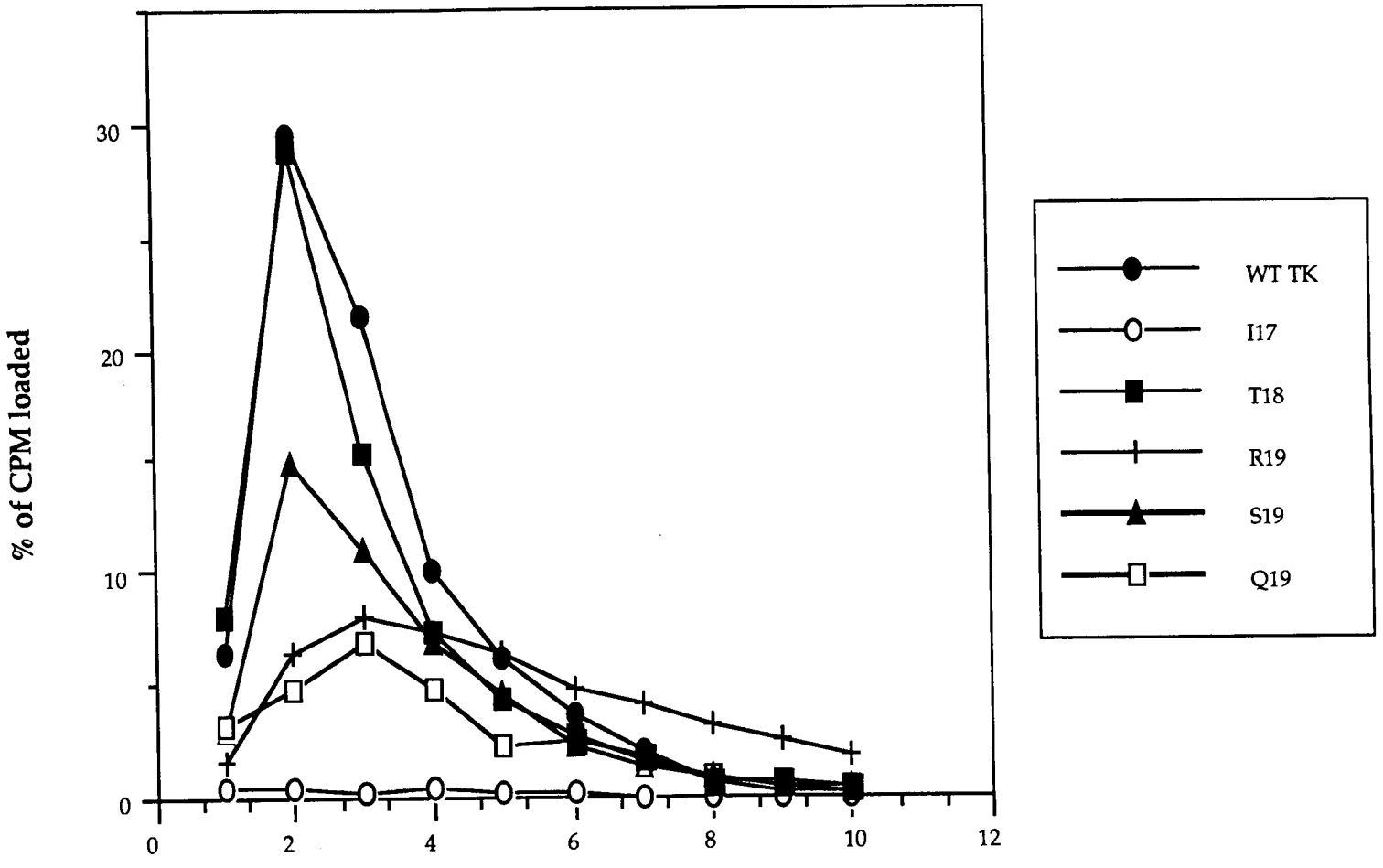

% of CPM loaded
30
20
10
0
0
2
4
6
8
10
12
Fractions
WT TK
I17
T18
R19
S19
Q19

Figure III.5

protein synthesis determinations and TK assays performed on wild type TK and the putative ATP binding site mutants. Mutant and wild type TK enzyme activity levels were normalized by adjusting protein synthesis counts of L-[^{35}S] methionine labeled mutant TKs with that of wild type TK (Table III.1). Based on the incorporation of L-[^{35}S] methionine, it would appear that all eleven mutant transcripts were translated *in vitro*, albeit with some small apparent differences in efficiency. In contrast, comparison of wild type to mutant TK enzyme levels revealed two mutant enzymes (T18 and S19) with elevated levels of activity while all the other mutant enzymes (V11, P15, T15, V16, I17, I19, P19, Q19 and R19) had greatly reduced levels of enzyme activity.

To determine whether the loss of enzyme activity in the mutant proteins was due to the inability to bind substrate ATP or to some other factor, ATP affinity chromatography was implemented. After application of the TK enzymes to the matrix, a gradient of 0 to 10 mM ATP was passed over the column. Figure III.5 shows the elution patterns of the wild type TK enzyme, the two mutants with enhanced activities (T18 and S19), two mutants with intermediate enzyme activity (Q19 and R19) and a representative (I17) of the mutants with background level of TK activity. Clearly T18 and S19 demonstrate a similar elution pattern to that of wild type TK indicative of a correspondingly similar binding affinity for ATP. The elution pattern for R19 and Q19 differ significantly from the patterns exhibited by fully functional TK enzymes. Trichloroacetic acid precipitable counts of both R19 and Q19 fractions demonstrate elution over a range of ATP concentrations resulting in a broad shoulder rather than the distinctive peak seen with wild type TK. This suggests that R19 and Q19 have an increased affinity for ATP which appears to result in a lowered

enzyme activity possibly due to an inability to release or transfer the phosphoryl group of ATP to thymidine. Mutants with a reduced or background level of enzyme activity (V11, P15, T15, V16, I17, I19 and P19) did not bind to ATP agarose to any appreciable level.

Interaction of ATP and dTTP binding sites.

It has been previously reported that dTTP binds noncompetitively with thymidine to allosterically regulate the VV TK enzyme. Hruby (1985) has shown that 10 μM dTTP leads to 50% inhibition of VV TK enzyme activity and that at 25 μM dTTP complete inactivation of enzyme activity results. It is possible that the interaction of dTTP with VV TK might interfere with the binding of ATP by blocking the ATP binding site or by altering the conformation of the ATP binding pocket such that the affinity for ATP is reduced to such an extent that catalysis does not occur. Alternatively, the binding of dTTP could act to disrupt the quaternary structure of TK by causing dissociation of the tetrameric complex such that the ATP binding site is thereby destroyed. To address this issue radiolabeled TK was passed over an ATP agarose affinity column followed by elution with a 0 to 50 μM dTTP gradient. Fractions were analyzed by hot trichloroacetic acid precipitable counts and plotted according to the percentage of counts loaded onto the column (Fig. III.6). The results from this analysis demonstrated that dTTP was able to elute TK from the ATP affinity column, suggesting that dTTP interfered with the ATP binding pocket in some fashion. The capability to elute TK was not a characteristic of deoxyribonucleotides in general. This was shown by the inability of 0 to 50 μM gradients of dGTP, dATP or dCTP to elute TK from the column (Fig. III.6). Bound TK was pulse eluted with 100 mM ATP after the gradient

Fig. III.6 Elution profile of L-[^{35}S] methionine labeled wild type thymidine kinase from an ATP agarose affinity column with deoxynucleotide gradients. Cell free translation products of wild type tk mRNA were loaded onto an ATP agarose affinity column and eluted with 0 to 50 μM gradients of dTTP, dGTP, dCTP or dATP. 100 ul aliquots of the collected fractions (1 ml) were analyzed and adjusted to the percentage of CPMs loaded onto the column as in Fig. III.5. Only the feedback inhibitor, dTTP, was able to elute TK from the ATP column while dGTP, dCTP and dATP did not. In the case where bound TK was not eluted by the deoxynucleotides, radiolabeled TK was released from the column by a 1 ml pulse of 100 mM ATP and 5 1-ml aliquots of binding buffer.

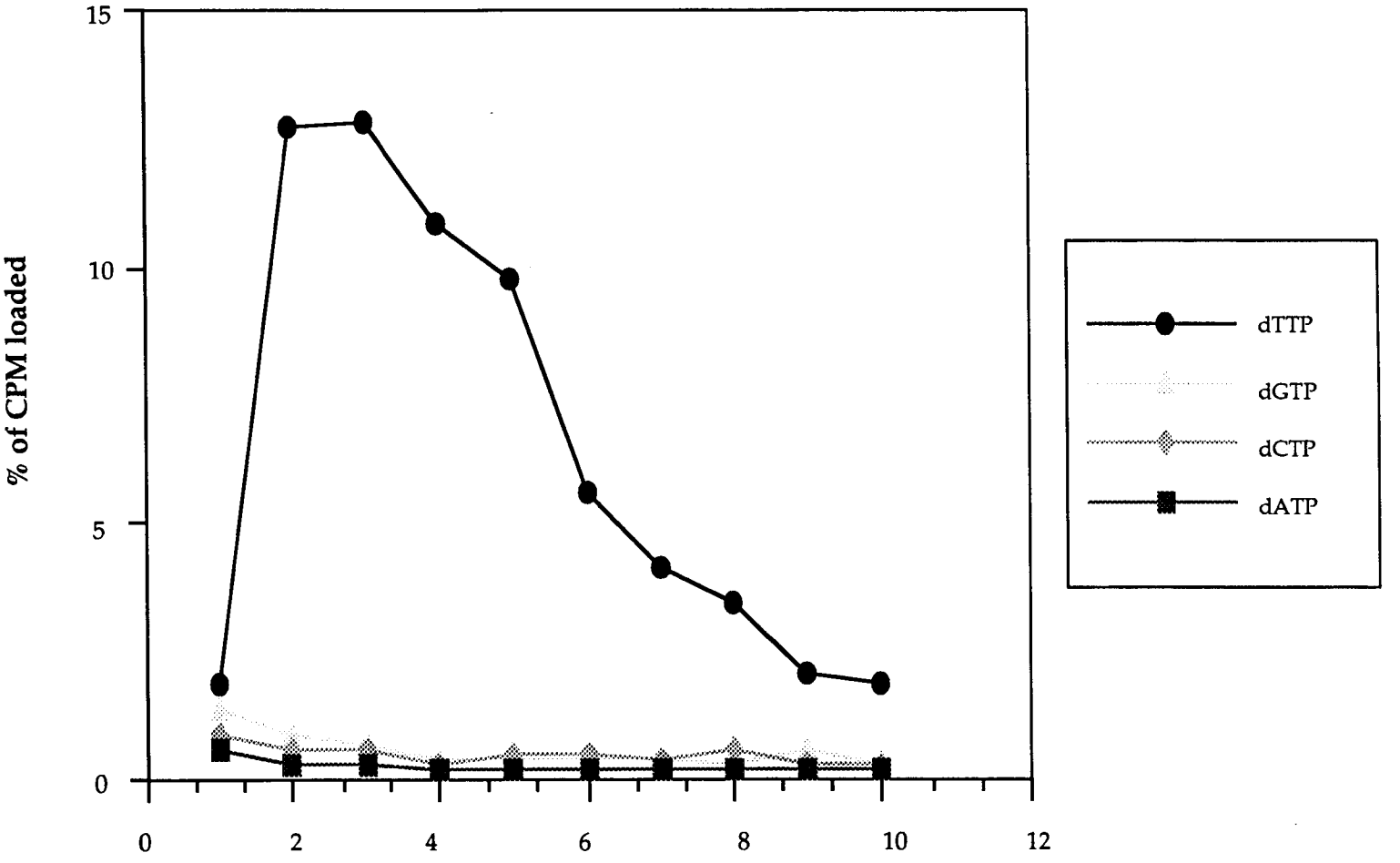

15
10
5
0
% of CPM loaded
0
2
4
6
8
10
12
Fractions
dTTP
dGTP
dCTP
dATP

Figure III.6

elutions (data not shown) to demonstrate this treatment had not destroyed the enzyme.

With regard to whether dTTP may act to alter the tetramer by catalyzing subunit dissociation, the following experiment was performed. L-[^{35}S] methionine labeled wild type TK generated in a cell free translation system was separated in 10-30% glycerol gradients, one containing 50 μM dTTP and the other without dTTP. Analysis of the fractions (Fig. III.7) revealed no difference in sedimentation patterns in the two situations indicating that dTTP at this concentration probably does not cause the tetramers to dissociate into their monomeric components as a means of enzyme inhibition.

Fig. III.7 Sedimentation analysis of wild type TK in 10-30% glycerol gradients in the presence or absence of dTTP. Radiolabeled wild type TK was centrifuged on 10-30% glycerol gradients +/- 50 μM dTTP for 20 h at 37,000 rpm at 4°C. Fractions were collected from the bottom, and the amount of trichloroacetic acid precipitable counts in every other fraction was determined and plotted with fraction 1 corresponding to the bottom of the gradient. The arrows indicate the position of hemoglobin (64.5 kDa), from rabbit reticulocyte lysates which serves as an internal size marker.

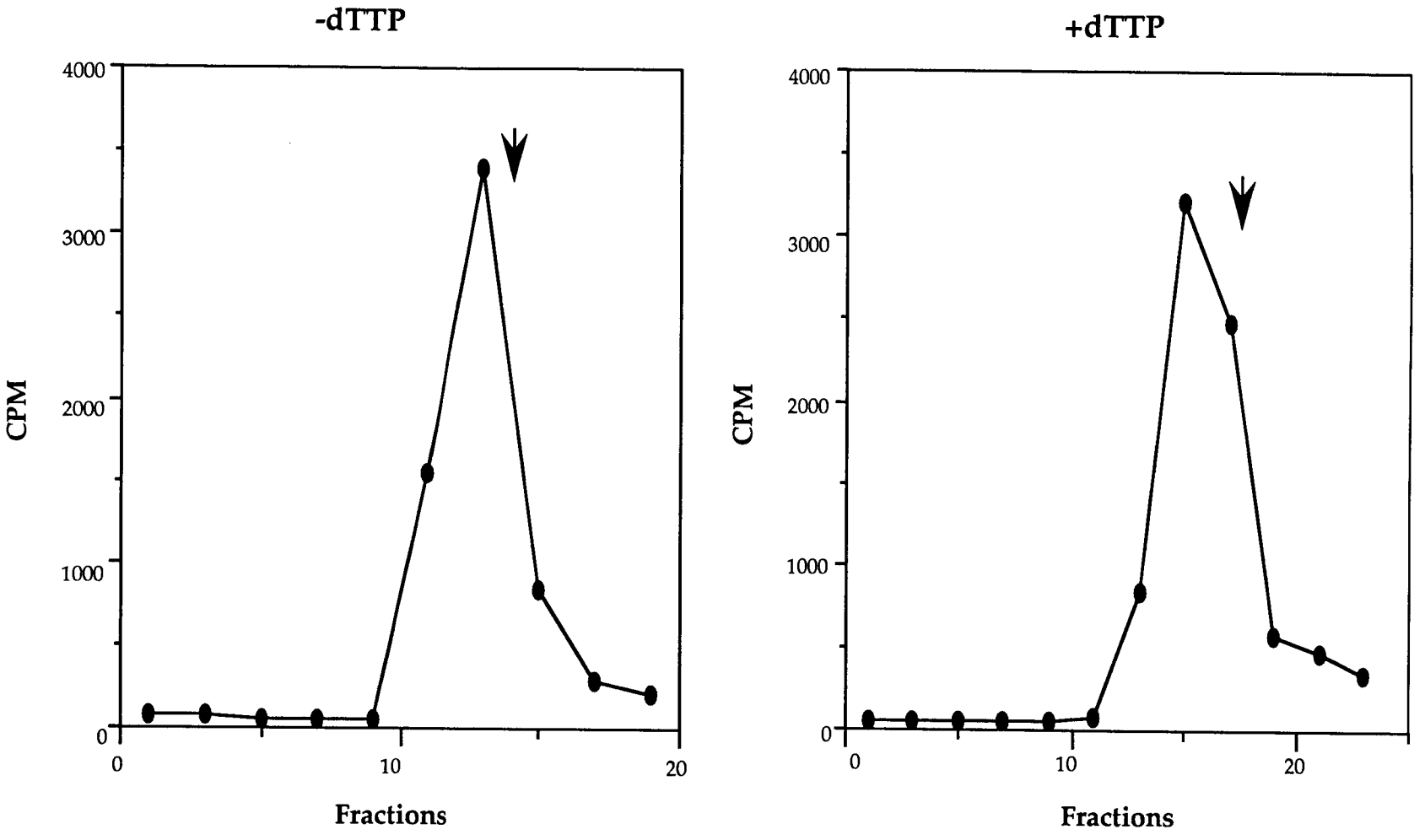


Figure III.7

DISCUSSION

Alignment of the VV TK amino acid sequence with the ATP and GTP binding site consensus sequences reveals a region (domain I) of the VV TK enzyme with high homology to these consensus sequences. While the TK sequence is not identical to either ATP or GTP consensus sequences, there is strong homology with the ATP consensus sequence and with the first domain of the GTP consensus sequence. Because of the relatedness to both sequences we approached identification the putative nucleotide binding site of TK by asking a series of questions: 1) What, if any, is the specificity for phosphate donor (ATP, GTP or other)?, 2) Is domain I involved in nucleotide binding?, and 3) Is the mechanism of feedback inhibition of TK enzyme activity by dTTP due to direct interference with ATP binding or does it act to limit catalysis in a different fashion?

Nucleotide binding site consensus sequences all share a core amino acid sequence of G X X X X G K (Moeller and Amons, 1985; Dever *et al.*, 1987; Liu and Summers, 1988 ; Reinstein *et al.*, 1988). A second region composed of a hydrophobic stretch terminated by an aspartic acid (G X_{4-8} D, segment 3) located approximately 90 residues from the core sequence towards the carboxy terminus, was also identified by homology searches of ATP binding proteins (Fry *et al.*, 1986). X-ray diffraction and NMR analyses of adenylate kinase led Fry *et al.* (1986) to suggest the involvement of Asp-119 in Mg^{2+} binding. In the case of VV TK there is a similar region (Fig. III.1, domain IV) spanning residues glycine 101 to aspartic acid 110. Because a number of ATP binding proteins including the thymidine kinases from Herpes Simplex Virus (McKnight, 1980; Fry *et al.*, 1986), *Escherichia coli*

(Black and Hruby, 1991a) and bacteriophage T4 (Valerie *et al.*, 1986), demonstrate no sequence similarity to the segment 3, its role in nucleotide binding is not clear. Furthermore, Kim *et al.* (1990) recently reported preliminary kinetic analysis of a Asp-119 to Asn-119 mutant of adenylate kinase which suggested that Asp-119 appears not to be involved in substrate binding or catalysis. Comparison of *E. coli* and bacteriophage T4 TK sequences (Black and Hruby, 1991a; Valerie *et al.*, 1986) with those shown in the alignment in Figure III.1 revealed identity with amino acids in domain III (residues 78-89), specifically the sequence comprised of D E X Q F (residues 82-86). Since this sequence is highly conserved among these thymidine kinases it is likely to be functionally important, perhaps as an alternative sequence to segment 3 for Mg^{2+} binding. Genetic analysis of domain III is currently under investigation.

Three domains are thought to be involved in the binding of GTP whereas only a single core domain appears to be necessary for ATP binding. To ascertain the specificity for ribonucleoside triphosphates which can act as donors in the phosphorylation of thymidine, VV TK was assayed with each in turn. In this system, 5 mM ATP was required for maximal TK activity. GTP, CTP or UTP were apparently not utilized even at concentrations as high as 25 mM (Fig. III.3). These results clearly demonstrate that ATP is the main phosphate source for the phosphorylation of thymidine *in vitro* by VV TK. Furthermore, Lowe *et al.* (1987) have proposed that one of the regions in the GTP binding site composed of the sequence N K X D is required for guanine specificity. This region is not present in the VV TK enzyme and therefore, according to this criterion, it is not surprising that VV TK is unable to recognize the guanine moiety of GTP.

Reinstein *et al.* (1988) reported that the amino acid at the carboxyl side of the lysine residue determines whether the protein is a GTP binding protein (serine) or an ATP binding protein (threonine). In contrast, Lowe *et al.* (1987) suggested that either a serine or threonine can be present at that position in a consensus sequence derived from the GTP binding protein, R-*ras*, thereby reducing the distinction between ATP and GTP binding proteins at this particular residue. In their analysis of the nucleotide binding site present in the HSV TK enzyme, Liu and Summers (1988) replaced the wild-type threonine (T64) (corresponds to residue 18 in VV TK) with an alanine or a serine. The alanine mutant resulted in inactivation of enzyme activity and, while the serine mutant retained enzyme activity, it displayed increased K_ms for thymidine and ATP. This suggested that a hydroxyl group at this position is required for transfer of the phosphoryl group from ATP to thymidine. In wild type VV TK the residue following lysine (residue 17) is a serine. Site-directed mutagenesis was used to exchange the serine with a threonine at position 18. Surprisingly, T18 displayed a greater than 3.5 fold increase in enzyme activity over that of wild type TK. Wild type TK and T18 ATP affinity agarose elution patterns were similar. However, as used here, this technique does not allow for quantitation of specific binding affinities. In any case it appears that altering domain I at residue 18 in VV TK to fit more closely with ATP binding site consensus sequences significantly increases enzyme activity. The question of why vaccinia virus encodes a thymidine kinase enzyme with sub-optimal activity is currently under investigation as well as further detailed biochemical analyses of T18.

X-ray crystallographic studies of porcine adenylate kinase suggest that the ATP binding site is comprised of a beta sheet - alpha helix interface

which forms a flexible loop structure (Fry *et al.*, 1986). Two conformations of adenylate kinase indicate that it is capable of opening or blocking the cleft formed by the loop structure (Fry *et al.*, 1985). The lysine at the carboxy end of the consensus sequence is in proximity to the phosphoryl group of ATP in adenylate kinase (Rozen *et al.*, 1989) and Liu and Summers (1988) suggest that the positively charged lysine in the corresponding region of HSV TK acts to neutralize the negatively charged phosphoryl group. Conversion of this lysine to the similar sized amino acid, isoleucine, results in the loss of enzyme activity, further suggesting that the ε-amino group of lysine is involved in binding ATP (Liu and Summers, 1988). Analysis of the glutamine and glutamic acid mutants at lysine 155 in F_1-ATPase shows reduced ATP hydrolysis of 80% and 66%, respectively. The binding affinities of these mutants for Mg·ATP and Mg·ADP further suggests that lysine 155 interacts with the gamma phosphoryl group of ATP (Parsonage *et al.*, 1988). Alteration of lysine 17 to isoleucine (I17) in VV TK leads to inactivation of enzyme activity and the loss of ability to bind ATP-agarose. While Parsonage *et al.* (1988) proposed that the lysine substitutions may interfere with the conformational movement of the loop structure such that product release is hindered, this appears not to be the case with the lysine to isoleucine mutation (I17) of VV TK, as ATP affinity column results indicate that no initial binding of the mutant I17 to ATP occurs.

The glycine residues within the consensus sequence apparently also play an important role. In the HSV TK enzyme, changing any of the three glycines to valine residues results in a loss of enzyme activity (Liu and Summers, 1988). It has been suggested that replacement of glycine with a larger valine residue prevents access of ATP to the binding site. Similarly,

replacement of glycine 10 in adenylate kinase or glycine 12 in p21 (*ras*) with valine leads to an increase in K_m for ATP (Reinstein *et al.*, 1988). While glycine to valine mutations in p21 and HSV TK reportedly maintain their native structure, the same mutation in adenylate kinase causes major structural changes in the conformational state of the protein. The V11 and V16 mutants of VV TK lack the ability to phosphorylate thymidine and are unable to bind ATP. Glycerol gradient separation of cell free translated V11 and V16 TK proteins suggests that these mutant enzymes are not capable of achieving the wild type conformation necessary for tetramer formation (data not shown).

Analysis of four other mutants, T15, P15, P19 and I19 (see Fig. II.2), also reveals a lack of enzyme activity and an inability to bind to an ATP affinity column. As with the V11 and V16 mutants, three of these mutants (T15, P15 and P19) appear to have undergone major conformational changes as suggested by the lack of any distinct radioactive peak in hot trichloroacetic acid precipitations of glycerol gradient fractions (data not shown). Indeed most of the mutants which exhibit background levels of enzyme activity appear to have an altered conformation. This is not surprising with the proline mutants since proline residues invoke protein turns (Creighton, 1984). However, it is less clear why a serine to threonine mutation (T15) would cause such an alteration in conformation. I19 on the other hand appears to maintain the native conformation in glycerol gradient sedimentation experiments. Although position 19 is not a component of the ATP binding site consensus sequence, mutations at that residue lead to enzymes with reduced or enhanced enzyme activity which demonstrate varying degrees of ability to bind ATP, suggesting that the amino acid residue at this site may be required in a nonspecific way for

maintaining proper conformational integrity of the beta sheet - alpha helix loop structure. The elevated level of enzyme activity when threonine is replaced by the smaller serine (S19) and the reduction in activity when threonine is replaced by isoleucine (I19), proline (P19), glutamine (Q19) or arginine (R19) supports the possibility that residues with large or bulky side chains at this position drastically interfere with enzymatic catalysis whereas a smaller residue enhances activity. Whether the specific requirement for a residue with a hydroxyl group is necessary for full activity and normal ATP binding or simply for a residue with a small side chain remains to be determined.

VV TK is feedback inhibited by the distal products of catalysis, dTDP and dTTP (Hruby, 1985). To address the mechanism of feedback inhibition it was of interest to determine whether ATP would still bind to TK in the presence of inhibitor. Studies presented here indicate that dTTP interferes with the binding of ATP, possibly by causing a conformational change in the loop structure. In the case of feedback inhibition of uridine kinase by CTP, Payne, *et al.* (1982) have shown that CTP causes the dissociation of the enzyme quaternary structure and subsequently leads to a loss of activity. In contrast, VV TK appears to remain as a tetrameric complex in the presence of dTTP, further suggesting that dTTP interacts either by sterically hindering the binding of ATP or by alteration of the ATP binding site loop structure, perhaps by binding at the ATP binding site or nearby. Studies to identify the dTTP and thymidine binding sites of VV TK are in progress and should shed light on the mechanism of substrate and regulatory element interactions.

Within the sequence of VV TK lies a repetitive aspartic acid motif (amino acid residues 68 - 89) with four aspartic acid residues positioned at

intervals of every seventh residue, separated by six nonspecific residues. Periodicity of a specific amino acid has recently been described for a number of signal transduction and cell transformation proteins, all of which contain heptad repeats of leucine residues now commonly referred to as the "leucine zipper" (Landschultz *et al.*, 1988; Turner and Tjian, 1988; Kouzarides and Ziff, 1988). The leucine zipper motif is believed to lie in an alpha helix such that all leucine residues align on one face when arranged on an idealized alpha helix (Landschulz *et al.*, 1988). Proteins which contain this motif such as *fos* and *jun*, interact with each other by the formation of a coiled-coil between the individual leucine zipper-containing alpha helices (O'Shea *et al.*, 1989). Replacement of individual leucine residues with glycines residues in *fos* reduced the ability of *fos* to bind with *jun* (Gentz *et al.*, 1989). While glycine residues are not normally found in alpha helices, glycine can function in hydrogen bonding without disrupting the stability of alpha helix structures (Gentz *et al.*, 1989). This is of interest because the chicken, mouse and human thymidine kinases contain a glycine residue at the equivalent site of the second aspartic acid residue (D2, amino acid 75) of this motif in VV TK. Computer generated analysis of VV TK secondary structure by Chou and Fasman algorithm predicts an alpha helix in the region of the aspartic acid repetitive sequence (residues 57 - 79 <Pa> = 1.100, <Pb> = 1.035, residues 81 - 87 <Pa> = 1.077, <Pb> = 1.017, residues 88 - 93 <Pa> = 1.060, <Pb> = 1.023, residues 89 - 100 <Pa> = 1.169, <Pb> = 0.935). The presence of a proline residue at position 88 next to D4 (residue 89) implies that this region is not maintained as a stable helix. However, this aspartic acid motif can be extended towards the amino terminus to include a conserved aspartic acid at residue 54 if one aspartic acid can be skipped in this motif. Chou and Fasman analysis of residues 50

- 55 predicts an alpha helix for this region as well (<Pa> = 1.048, <Pb> = 1.002). This motif would then consist of D1 (54), D2 (68), D3 (75) and D4 (82). When O'Shea *et al.* (1989) searched a protein database for additional proteins containing four consecutive leucine repeats spaced seven residues apart, this leucine zipper motif has been show to be prevalent in ~200 proteins. When the Protein Identification Resource bank was searched for proteins containing this motif with aspartic acid residues in place of leucines only ~25 proteins were selected. In light of fact that thymidine kinase requires a tetrameric quaternary structure for enzyme activity, it seems plausible that the aspartic acid motif could be involved in subunit interaction, either by a means similar to the leucine zipper or, perhaps more likely because of the negatively charged aspartic acid residues, by interaction with a region of positively charged residues such as lysine 133, 140, 143, 149 and arginine 150. Purification of VV TK expressed in *E. coli* is in progress for x-ray crystallographic analysis of protein structure to identify the conformation of this domain and to determine the involvement of it or other domains in subunit interaction, substrate binding or catalysis.

In summary, the studies presented in this paper demonstrate that the 11-18 amino acid domain of VV TK is involved in binding ATP and that no other ribonucleoside triphosphate is efficiently utilized as a phosphate donor. Site directed mutagenesis was used to create mutant TK enzymes with varying levels of enzyme activity. These mutants were further characterized by a novel use of an ATP agarose affinity column to demonstrate ATP binding capability. In addition we present data which suggests that dTTP inhibits enzymatic catalysis by interfering with the ability of TK to bind ATP and that this inhibitory effect is not due to dissociation of the tetrameric complex of TK.

Future studies include identification and characterization of other functional regions of this enzyme including those required for thymidine binding, regulation and subunit interaction in order to correlate amino acid sequences or structures with a specific function.

CHAPTER IV

NUCLEOTIDE SEQUENCE OF THE *E. COLI* THYMIDINE KINASE GENE PROVIDES EVIDENCE FOR CONSERVATION OF FUNCTIONAL DOMAINS AND QUATERNARY STRUCTURE

Authors: Margaret E. Black and Dennis E. Hruby

SUMMARY

Using λ bacteriophage clones from the Kohara *E. coli* library spanning minutes 25.5 to 28.5 on the *E. coli* chromosome (strain W3110), we identified two overlapping DNA fragments which were able to confer thymidine kinase (TK) enzyme activity to a TK^- strain of *E. coli* (KY895). This genetic complementation assay was used in concert with subcloning procedures to identify the minimal region (a 900bp *Eco*RI/*Sal*I fragment) which contained the *E. coli* thymidine kinase gene (*tdk*). The nucleotide sequence of the *Eco*RI/*Sal*I fragment and a small portion of the adjoining downstream fragment were determined. Computer analysis of the derived sequence indicated the presence of a rightward-reading open reading frame of 615bp which could encode a 205 amino acid polypeptide with a predicted molecular weight of 23,458 daltons. The *in vivo* transcriptional activity of this locus was confirmed by northern blot hybridization analysis of RNA isolated from *E coli* JM101 or KY895, which detected a 650 nucleotide RNA transcribed from this region. This places the *tdk* gene at approximately minute 27.35 on the *E. coli* W3110 chromosome, about 15kb downstream from the *nar*G locus and approximately 25kb upstream of the *trp* operon. Although the predicted molecular weight of the *E. coli* TK protein was 23.5kD, gel filtration analyses suggested that, like eukaryotic thymidine kinases, the active form of this enzyme is a multimeric complex. Furthermore, comparisons between the predicted amino acid sequences of the TK proteins encoded by *E. coli*, bacteriophage T4 and a mammalian poxvirus (vaccinia virus) revealed several regions of amino acid identity, one of which appears to be the ATP binding site, thereby suggesting that the

essential functional domains of this enzyme have been well-conserved throughout evolution.

INTRODUCTION

Thymidine kinase (TK, E.C. 2.7.1.21) is a key enzyme in the salvage pathway of pyrimidine nucleotide metabolism, catalyzing the production of dTMP from thymidine and ATP. Because of its importance in nucleic acid metabolism and as a target in anti-viral and anti-tumor chemotherapy (Hirsch and Kaplan, 1990), thymidine kinases from a wide variety of sources including human (Flemington *et al.*, 1987), mouse (Lin *et al.*, 1985), chicken (Kwoh and Engler, 1984), Chinese hamster (Lewis, 1986), bacteriophage T4 (Valerie *et al.*, 1986), a number of representatives of the *Poxviridae* (Esposito and Knight, 1984; Weir and Moss, 1983a; Upton and McFadden, 1983; Boyle *et al.*, 1987; Gershon and Black, 1989) and several members of the *Herpesviridae* (Wagner *et al.*, 1981; Gompels and Minson, 1986; Darby *et al.*, 1986; Kit *et al.*, 1983; Otsuka and Kit, 1984; Littler *et al.*, 1986; Davison and Scott, 1986; Nunberg *et al.*, 1989) have been isolated and sequenced. Regardless of their parental sources, in order to carry out their prescribed metabolic roles, thymidine kinase enzymes must contain protein domains or structures which are essential for catalysis, substrate (ATP and thymidine) and cofactor (Mg^{2+}) binding, responding to feedback inhibitors (dTDP and dTTP) and assembly into multimeric active complexes.

One approach to identifying the essential functional and regulatory domains of TK proteins is to align the predicted amino acid sequences of the enzymes listed above with the assumption that conserved residues will imply similar function (Black and Hruby, 1990b). Interestingly, although many of the TK proteins appear to be highly related to the prototype

vaccinia virus (VV) enzyme, alignment with TK enzymes encoded by any of the herpesviruses reveals little homology, outside a sequence common to enzymes which use ATP (Robertson and Whalley, 1988). This is not entirely surprising since the TK enzymes of the *Herpesviridae* have been shown to be homodimeric deoxypyrimidine kinases which have the additional ability to phosphorylate thymidylate and, unlike VV TK, are not feedback inhibited by dTTP (Robertson and Whalley, 1988). In contrast, the VV and human TK enzymes have been shown to be homotetramers in their native conformation and to be more selective in their substrate utilization (Black and Hruby, 1990b; Sherely and Kelly, 1988). Robertson and Whalley (1988) reported a high degree of homology between the *Herpes simplex* virus (HSV) TK and the *Saccharomyces cdc*8 gene product, a thymidylate kinase, and suggested that the HSV TK may have been derived from a thymidylate kinase progenitor and evolved "backwards" in the metabolic pathway to acquire thymidine kinase activity. Therefore, consideration of the distinct difference in the sizes of the predicted proteins (approximately 40kD for HSV TK versus 20kD for VV TK) and the additional catalytic functions of *Herpesviridae* enzymes prompted the suggestion that two different classes of TK enzymes exist (Black and Hruby, 1990a). The first group (type I) is comprised of TK (deoxypyrimidine kinase/thymidylate kinase) enzymes derived from the members of the *Herpesviridae* family whereas the TK enzymes from human, mouse, chicken, vaccinia virus, variola virus, monkeypox, fowlpox, capripox and bacteriophage T4 sources comprise the type II category (Black and Hruby, 1990a).

Sequence comparison of type II TK enzymes (vaccinia virus, mouse, human, chicken, monkeypox and variola virus) reveals seven regions or

domains of complete identity (Black and Hruby, 1990b). However, since these enzymes share between 68 to 97% overall identity when compared to VV TK, it is not clear whether each of the conserved regions represents an essential domain. It was therefore of interest to compare the predicted amino acid sequence of the mammalian type II TK enzymes with that of a more distantly related enzyme, such as a prokaryotic TK, in order to highlight essential structural features of TK enzymes in general. Unfortunately, examination of the literature and sequence data banks for TK gene sequences from *Escherichia coli* or other bacteria revealed that this information was not available. With this as the starting impetus, the experiments reported in this paper provide the position, nucleotide sequence, deduced amino acid sequence, transcript size and estimated protein native molecular weight of the *E. coli* thymidine kinase gene (*tdk*). The *E. coli tdk* gene encodes a predicted protein with high homology with the bacteriophage T4 TK enzyme but with a somewhat lesser homology to the VV TK. Interestingly, a number of the residues which are conserved between the VV and *E. coli* enzymes fall within three of the seven conserved domains identified in the alignment of viral and vertebrate TKs, suggesting that these regions are likely to be required for enzyme activity.

EXPERIMENTAL PROCEDURES

Bacterial strains, bacteriophage λ clones and enzyme assays.

E. coli strain KY895 [F-*tdk*- 1-*ilv*] is deficient in thymidine kinase activity and was used in situations where TK activity expressed by either λ or plasmid clones were assayed. *E. coli* strain JM101 [*sup*E *thi* Δ(*lac-pro*AB) F' {*tra*D36 *pro*AB+ *lac*Iq *lac*ZΔM15}] was used for general cloning procedures. Bacteriophage λ clones [11G8 (244), 7C10 (245), 12A3 (246), 4D10 (247), 13HP (248), 12HD (249), 4D8 (250), 3D5 (251), 14C4 (252), 4F1 (253), 13F9 (254), 1836 (255)] were from the Kohara library (miniset) of *E. coli* W3110 chromosomal DNA (Kohara *et al.*, 1987). All phage growth was done in NZCYM medium and with growth conditions according to Maniatis *et al.* (1982). Bacterial cultures were grown at 37°C in 2 X YT (Maniatis *et al.*, 1982) and supplemented with 75μg/ml ampicillin where necessary. Thymidine kinase activity was assayed as previously described (Hruby and Ball, 1981a).

Nucleic acid isolation and purification.

λ DNA minipreps were performed as follows. A plate lysate stock of phage was prepared according to Maniatis *et al.* (1982) and 50μl absorbed to 100μl of a saturated bacterial culture for 15 min at 37°C. This mixture was used to inoculate 20ml NZCYM and allowed to grow shaking at 37°C until lysis occured. Cell debris was removed by centrifugation at 10,000 rpm for 10 min in a Beckman J2-21 centrifuge. The supernatant was transferred to a fresh tube and DNAase I and RNAase A added to 20μg/ml and 10μg/ml, respectively. After 15 min at 37°C, 0.2 volumes of 20% PEG 8000/2.5M

NaCl was added and the samples incubated at room temperature for 15 min. The samples were then centrifuged at 10,000 rpm for 15 min and the supernatant discarded. The pellet was suspended in 0.5ml 4% PEG 8000/0.5M NaCl and transferred to a microfuge tube. After a phenol:chloroform and a chloroform extraction, the DNA was precipitated with 1ml 95% ethanol. The precipitate was washed with 70% ethanol, dried and resuspended in 50µl 10mM Tris (pH 7.6)/1mM EDTA (TE). Plasmid DNA minipreps were done using standard procedures and CsCl density gradient purification procedures were as according to Maniatis *et al.* (1982).

Subcloning and sequencing.

The DNA fragment from λ clone 251 was subcloned into the polylinker region of pUC118 vectors restricted with the appropriate enzyme. pUC118 and pUC119 were a gift from J. Vieira. Single stranded DNA template was obtained by extraction of phage particles from 16-18 h old cultures incubated in the presence of the M13 helper virus, M13KO7. Both DNA strands were sequenced using the chain terminating DNA sequencing method (Sanger *et al.*, 1977) with Sequenase™ (US Biochemicals). Alkaline denaturation of CsCl-banded double stranded DNA for sequencing was as specified by US Biochemicals. The 900bp *Eco*I/*Sal*I fragment in pUC118 was gel isolated and cloned into pUC119 to obtain the opposite orientation for sequencing. The universal and reverse primers (New England BioLabs) and four oligonucleotides synthesized at the Center for Gene Research and Biotechnology at Oregon State University (1, 5' CTACTATTCCGCG 3';
2, 5' GCATGCCGCGTTCC 3'; 3, 5' GTCGATCAACTCG 3';

4, 5' CGGGTATATCGAG 3') were used as sequencing primers.

Isolation of total RNA and northern blot analysis.

E. coli strains JM101 and KY895 (100ml each) were grown in 2 X YT to mid-log phase. The cells were pelleted, resuspended in 1ml of 25% sucrose and transferred to a 15ml Corex tube. Six and one-half ml buffer [8% sucrose/0.5% Triton X-100/50mM EDTA (pH 8)/10mM Tris-HCl (pH 8)], 0.4 ml 200mM vanadyl and 0.5ml 10mg/ml lysozyme were added to the cells and the mixture placed on ice for 5 min. The mixture was then boiled for 2 min and centrifuged for 15 min at 8,000 rpm in a Beckman J2-21 centrifuge. The supernatant was transferred to a fresh tube and 0.01 volume 3M $NaAc_2$ and 0.7 volumes isopropanol added. After centrifuging for 10 min at 8,000 rpm, the pellet was redissolved in 2ml 0.3M $NaAc_2$/TE/0.2% sodium dodecylsulfate and extracted with phenol:chloroform three times. An equal volume of 4M LiAc (pH 6) was added and the sample incubated on ice for 1 h. After centrifuging for 10 min at 8,000 rpm, the pellet was resuspended in 1ml H_2O and reprecipitated with LiAc as above and then finally precipitated with 0.01 volume 3M $NaAc_2$ (pH 6) and 2.5 volumes 95% ethanol. The pellet was washed with 70% ethanol, dried and resuspended in 0.5ml H_2O. JM101 and KY895 RNA (5μg and 20μg) and RNA molecular weight markers (0.16-1.77kb RNA ladder from BRL) were subjected to electrophoresis in a 0.8% agarose/formaldehyde gel (Maniatis *et al.*, 1982) and the separated RNA transferred to Nytran and baked according to the manufacturers directions (Schleicher and Schuell). By using the Amersham random primer labeling kit, a gel isolated 629bp *AccI*/*SalI* fragment was labeled with α-[^{32}P]dATP. This probe was used in a rapid hybridization procedure (Amersham) under conditions to allow

hybridization of this probe with the TK transcript. Kodak XOMAT film was exposed to the washed filter at -70°C with one intensifying screen.

Molecular weight determination.

Bacterial extracts of *E. coli* KY895 carrying the 2.5kb *Eco*RI/*Pst*I fragment in pUC118 were prepared. An aliquot (0.5ml) of the culture was pelleted and resuspended in 0.5ml 100 mM Tris-HCl (pH 8) and 40µl lysis buffer [100mM EDTA (pH 8)/100mM dithiothreitol/50mM Tris-HCl (pH 8)]. One drop of toluene was added, the sample vigorously vortexed and incubated for 30 min at 30°C. The extract was then clarified by centrifugation. This extract was mixed with an equal volume of TENDT [50mM Tris (pH 7.6)/1mM EDTA/100mM NaCl /0.015% Triton X-100] and glycerol added to 10%.

A standard curve was generated by passing a mixture of protein gel filtration molecular weight markers (Sigma MW-GF-200 kit) through a Sephacryl-300 (Pharmacia) gel filtration column equilibrated with TENDT buffer. Fractions (250µl) of the eluate were collected (100 fractions, 25ml total volume) and alternate samples from number 30 to 100 were subjected to electrophoresis in 12% polyacrylamide gels containing sodium dodecylsulfate (Studier, 1973). The gels were then stained with Coomassie Brilliant Blue R-250 to visualize individual proteins. The peak TK fraction from the S-300 separation of the bacterial extracts was determined by assaying every fraction from fraction 30 to 100 for TK activity.

Fractions of cell extracts separated in 10-30% glycerol gradients prepared in TMDT buffer [20mM Tris-HCl (pH 7.6), 2mM magnesium acetate, 1mM dithiothreitol and 50µM thymidine] were assayed for enzyme activity to determine the position of TK in the gradient. A mixture of the

MW-GF-200 molecular weight markers was resolved by glycerol gradient fractionation and analyzed as the gel filtration markers above (Studier, 1973). The position of TK from gradients of cell extracts was plotted to this standard curve.

Computer assisted analysis.

Programs from IntelleGenetics, Inc. (Mountain View, CA) were used to analyze sequencing and amino acid information. These included the SEQ, IFIND, PEP and ALIGN programs.

RESULTS AND DISCUSSION

Identification of the tdk gene of Escherichia coli.

The map position of the thymidine kinase gene (*tdk*) has previously been mapped at approximately 27 minutes on the *E. coli* chromosome (Bachman, 1983). In order to more precisely pinpoint the genomic location of the *tdk* gene, genetic complementation procedures were employed. Twelve bacteriophage λ clones (no. 244-255) containing overlapping fragments of the *E. coli* W3110 chromosome spanning minute 25.5 to minute 28.5 were obtained from the miniset library constructed by Kohara (Kohara *et al.*, 1987). Phage lysates were prepared by infecting an *E. coli* strain defective in thymidine kinase expression [KY895 (F⁻ *tdk*⁻ 1-*ilv*) Igarashi *et al.*, 1967] individually with each of the twelve recombinant λ clones. The lysates were subjected to low-speed centrifugation to remove bacterial cell debris and the supernatants assayed for the ability to phosphorylate [*methyl* ^{3}H]-thymidine using a standard assay designed to measure thymidine kinase (TK) activity (Hruby and Ball, 1981a). Of the twelve clones only two, 250 and 251, were capable of conferring a TK$^+$ phenotype to *E. coli* KY895 (Fig. IV.1). The region of shared DNA between the two positive clones was approximately 10kb in length. In order to identify a smaller region amenable to nucleic acid sequencing procedures, the DNA insert from clone 251 was subjected to restriction enzyme cleavage mapping procedures to produce the map shown in the middle of Fig. IV.1. The clone 251 insert was first digested with *Kpn*I and *Eco*RI and the two *Kpn*I/*Eco*RI fragments subcloned into pUC118 and transformed into *E. coli* KY895. Bacterial cell extracts were made from transformants

Fig. IV. 1 Schematic representation of the *E. coli* W3110 chromosomal region which encompasses the *tdk* locus. The top line represents the position on the *E. coli* genetic map in kilobases (kb) (Kohara *et al.*, 1987). The lines numbered 244 to 255 reflect the approximate position of the DNA inserts contained by the bacteriophage λ clones assayed for TK activity. Heavily darkened lines, e.g. 250 and 251, denote that TK activity was detected in *E. coli* KY895 cells infected with recombinant phage or transformed with plasmids carrying this DNA. Restriction sites mapped in the DNA insert of recombinant λ clone 251 are indicated (E, *Eco*RI; S, *Sal*I; P, *Pst*I; H, *Hin*dIII and K, *Kpn*I). The open reading frame of *tdk* and its orientation are denoted by an arrow.

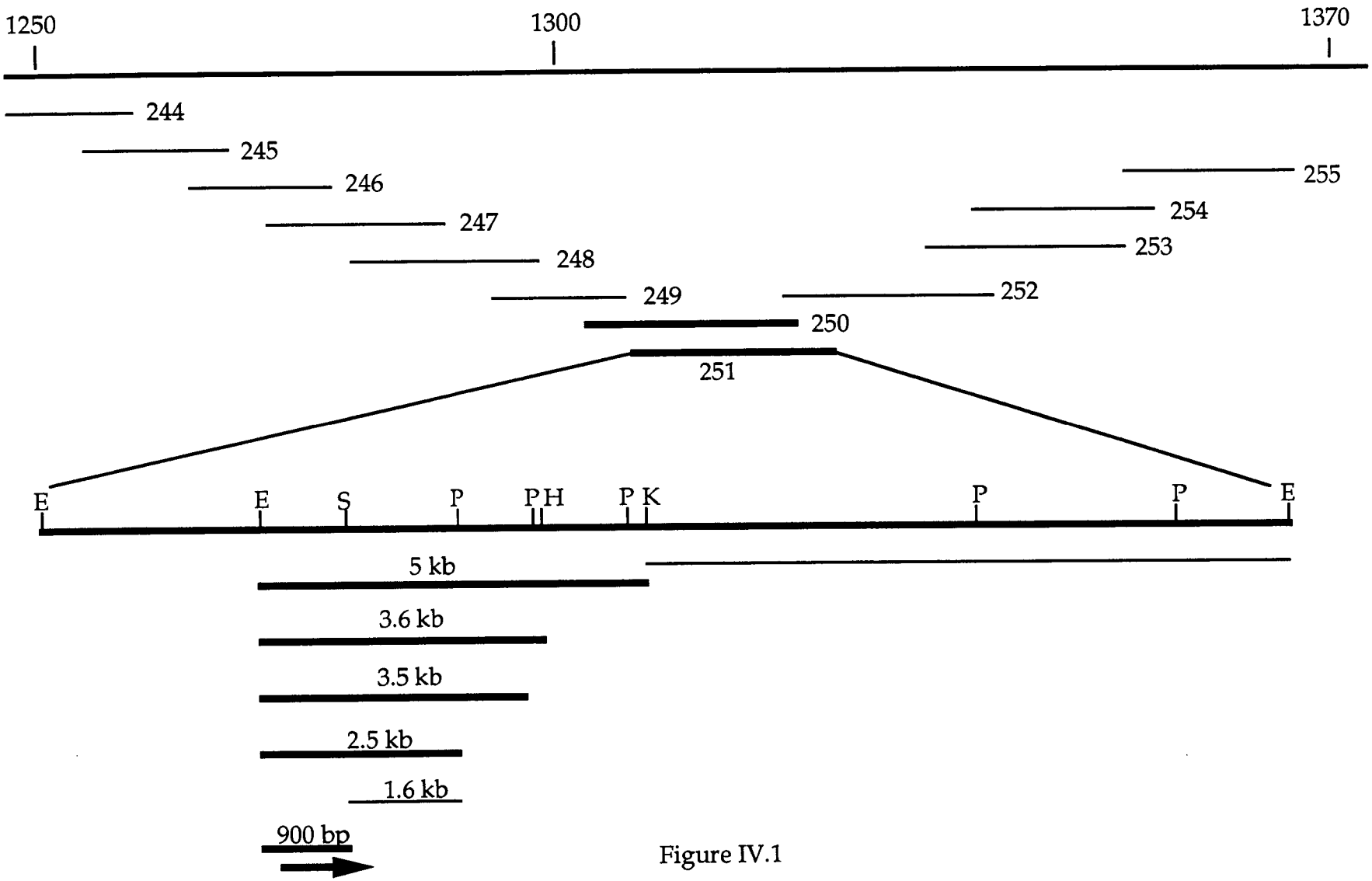

1250
1300
1370
244
245
246
247
248
249
250
251
252
253
254
255
E
E
S
P
PH
PK
P
P
E
5 kb
3.6 kb
3.5 kb
2.5 kb
1.6 kb
900 bp

Figure IV.1

and tested for TK activity. The 5kb *Kpn*I/*Eco*RI fragment conferred TK activity to the cells whereas the larger 7kb *Kpn*I/*Eco*RI fragment did not. Smaller portions of the 5kb *Kpn*I/*Eco*RI fragment were subcloned into pUC118, transformed into *E. coli* KY895 and assayed for TK activity. At the bottom of Fig. IV.1 are shown the position of the subcloned fragments in relation to the original 251 insert and whether or not the clones were able to confer TK activity to *E. coli* KY895. The smallest fragment which was able to confer TK expression to *E. coli* KY895 cells was the 900bp *Eco*RI/*Sal*I subfragment, suggesting that most, if not all, of the *tdk* gene is contained within that insert.

Sequencing of the tdk gene.

Synthetic oligonucleotide primers which were able to anneal to pUC118/pUC119 adjacent to the polylinker, or at sites internal to the insert, were used to sequence both strands of the 900bp *Eco*RI/*Sal*I fragment. The location of the primers and the regions sequenced with each primer are diagrammed in Fig. IV.2. The initial examination of the *tdk* open reading frame (ORF) revealed that an in-frame stop codon was not present on the *Eco*RI/*Sal*I 900bp fragment, suggesting that the 3'-end of the gene was absent. Therefore, a short region (~200bp) of the adjoining 1.6kb *Sal*I/*Pst*I fragment was also sequenced in order to obtain the entire *tdk* ORF (Fig. IV.3).

Analysis of the tdk open reading frame.

Computer assisted analysis of the sequenced region revealed a rightward reading ORF of 615bp beginning at an AUG start codon at nucleotide (nt) 316 and ending with a TAA stop codon at nt 931 (Fig. IV.3).

Fig. IV.2 Strategy for sequencing the *tdk* gene. Arrows denote the direction, location and distance of the determined sequence. Both long and short sequencing runs using the same DNA and primer are depicted by the same arrow. The small bars labeled 1, 2, 3 and 4 represent the position of synthetic oligonucleotides and the number next to the arrows marks the sequence generated using these primers. The position of the start and stop codons are also indicated. The positions of salient restriction sites (E, *Eco*RI; A, *Acc*I; R, *Rsa*I and S, *Sal*I) are also shown.

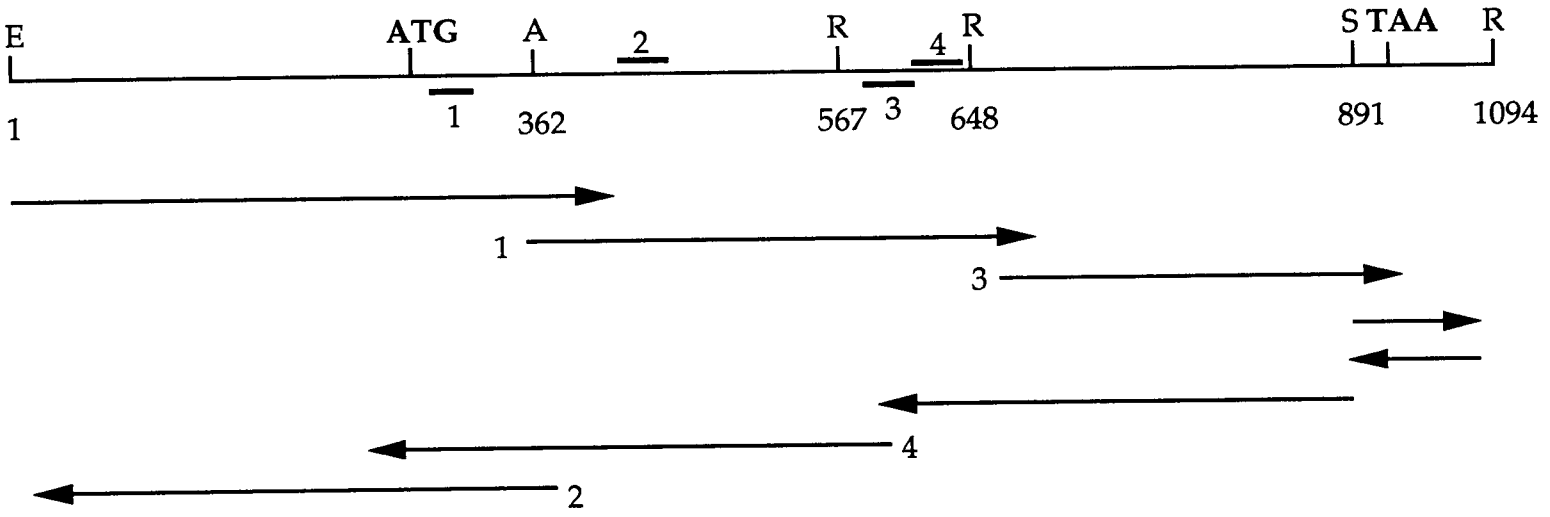


Figure IV.2

Fig. IV.3 DNA sequence of the *tdk* gene and the computer predicted amino acid sequence of the TK protein. Arrows indicate predicted inverted repeats. The *tdk* promoter -35 and -10 sequences and the putative Shine-Dalgarno site (SD) are marked by underlines. Standard single amino acid nomenclature is used. This sequence has been given the EMBL accession number X53733.

```
GAATTCCCTATTAAGCGAATGATGAAAAGTAGAACAGTCGCAATAAGAGCATGGACTTAGTATTGCACT  69

ATCTCCTGGAGGTCAACAGAGGGCTATTACTTGCGCAACAGGTTAAAGATTGTGAATAGTTACCAGCAG 138

TCATTTACCCGCTTATAACAAGCGAGGCAGTTGTAATGATAGCTCAGAAGGATTATGCAAGGCTTCGTA 207
                                          ---------------->
                                    -35                  -10
AGGGAGAACGCATATACCCACTTCTGTGCATACTGTTGAGCTGAAAAACTGACGAATTATGATAAACTC 276
                          <---------------
                            SD          M  A  Q  L  Y  F  Y  Y  S  A   10
CAGCCAACTTTATTTCATATCATTGAGGGCCTGTGGCTGATGGCACAGCTATATTTCTACTATTCCGCA 345

 M  N  A  G  K  S  T  A  L  L  Q  S  S  Y  N  Y  Q  E  R  G  M  R  T   33
ATGAATGCGGGTAAGTCTACAGCATTGTTGCAATCTTCATACAATTACCAGGAACGCGGCATGCGCACT 414

 V  V  Y  T  A  E  I  D  D  R  F  G  A  G  K  V  S  S  R  I  G  L  S   57
GTCGTATATACGGCAGAAATTGATGATCGCTTTGGTGCCGGGAAAGTCAGTTCGCGTATAGGTTTGTCA 483

 S  P  A  K  L  F  N  Q  N  S  S  L  F  D  E  I  R  A  E  H  E  Q  Q   96
TCGCCTGCAAAATTATTTAACCAAAATTCATCATTATTTGATGAGATTCGTGCGGAACATGAACAGCAG 552

 A  I  H  C  V  L  V  D  E  C  Q  F  L  T  R  Q  Q  V  Y  E  L  S  E  102
GCAATTCATTGCGTACTGGTTGATGAATGCCAGTTTTTAACCAGACAACAAGTATATGAATTATCGGAG 621

 V  V  D  Q  L  D  I  P  V  L  C  Y  G  L  R  T  D  F  R  G  E  L  F  125
GTTGTCGATCAACTCGATATACCCGTACTTTGTTATGGTTTACGTACCGATTTTCGAGGTGAATTATTT 690

 I  G  S  Q  Y  L  L  A  W  S  D  K  L  V  E  L  K  T  I  C  F  C  G  148
ATTGGCAGCCAATACTTACTGGCATGGTCCGACAAACTGGTTGAATTAAAAACCATCTGTTTTTGTGGC 759

 R  K  A  S  M  V  L  R  L  D  Q  A  G  R  P  Y  N  E  G  E  Q  V  V  172
CGTAAAGCAAGCATGGTGCTGCGTCTTGATCAAGCAGGCAGACCTTATAACGAAGGTGAGCAGGTGGTA 828

 I  G  G  N  E  R  Y  V  S  V  C  R  K  H  Y  K  E  A  L  Q  V  D  S  194
ATTGGTGGTAATGAACGATACGTTTCTGTATGCCGTAAACACTATAAAGAGGCGTTACAAGTCGACTCA 897

                                  205
 L  T  A  I  Q  E  R  H  R  H  D  *
TTAACGGCTATTCAGGAAAGGCATCGCCACGATTAATAAGAATTTCTTTACTGACAGGGTGAGCAGGGC 966
                                                    -------->

ACTTTTATCCTGTCAGTTCGTTTTACGCACTTCTTCCGGGCTATATACCCTTCTCGGCAGTTTTTTAAC 1035
        <---------                  -------------->

GCCGCTATACGCCTCACAGGGCTCTTAAGCACCGACGTTGACTTGTGACCTGTAAAGTAC  1095
                    <--------------
```

Figure IV.3

This places the *tdk* ORF on the *E. coli* W3110 chromosome at ~27.35 minutes, between the *nar*G locus (~15kb upstream) and the *trp* operon (~25kb downstream). The *tdk* ORF (615bp) is predicted to encode a polypeptide which is 205 amino acids in length with a molecular weight of 23,458 daltons. Directly upstream from the start codon at nts 301 to 305 is a putative ribosome-binding site. Further upstream at nts 243-248 and at nts 265-270 are two regions separated by 16 nt with characteristics typical of -35 and -10 promoter elements (Fig. IV.3). Computer analysis predicts three regions of dyad symmetry with free energies of less than -12. At the 5' end of the ORF, nucleotides 231-249 match nt 199-179 at 72.7% with a $-\Delta G = 13.8$ Kcal. Immediately proximal to the stop codon are two sequences with dyad symmetry. Nucleotides 975-982 will base pair exactly with nts 955-947 with a $-\Delta G = -13.5$ Kcal and nts 1053-1069 match nts 1017-1003 at 72.2% with a $-\Delta G = 13.0$ Kcal.

Transcriptional analysis.

To confirm that the *tdk* locus is expressed *in vivo* and is transcribed into an mRNA of the expected size, total RNA was extracted from log phase *E. coli* JM101 or KY895 cultures and subjected to electrophoresis on a denaturing formaldehyde-agarose gel. The RNA was then blotted onto a nylon membrane and hybridized with a random primer labeled fragment (the 629bp *Acc*I/*Sal*I fragment internal to the *tdk* ORF). A single major band with an approximate size of 650 nt was detected (Fig. IV.4). This size band corresponds to the expected transcript size for the *tdk* ORF. Interestingly, the *tdk* transcript is also evident in the RNA from *E. coli* KY895, which exhibits a TK^- phenotype, suggesting that the loss in enzyme activity in this mutant is likely due to either a missense or nonsense

Fig. IV.4 *In vivo* transcription of the *E. coli tdk* gene. Shown is an autoradiograph of a northern blot analysis of total RNA extracted from *E. coli* JM101 or KY895 and probed with the 629bp *Acc*I/*Sal*I fragment. The numbers at the top indicate the amount (μg) of RNA per lane. Migration pattern and size (kb) of RNA molecular weight markers are shown on the left.

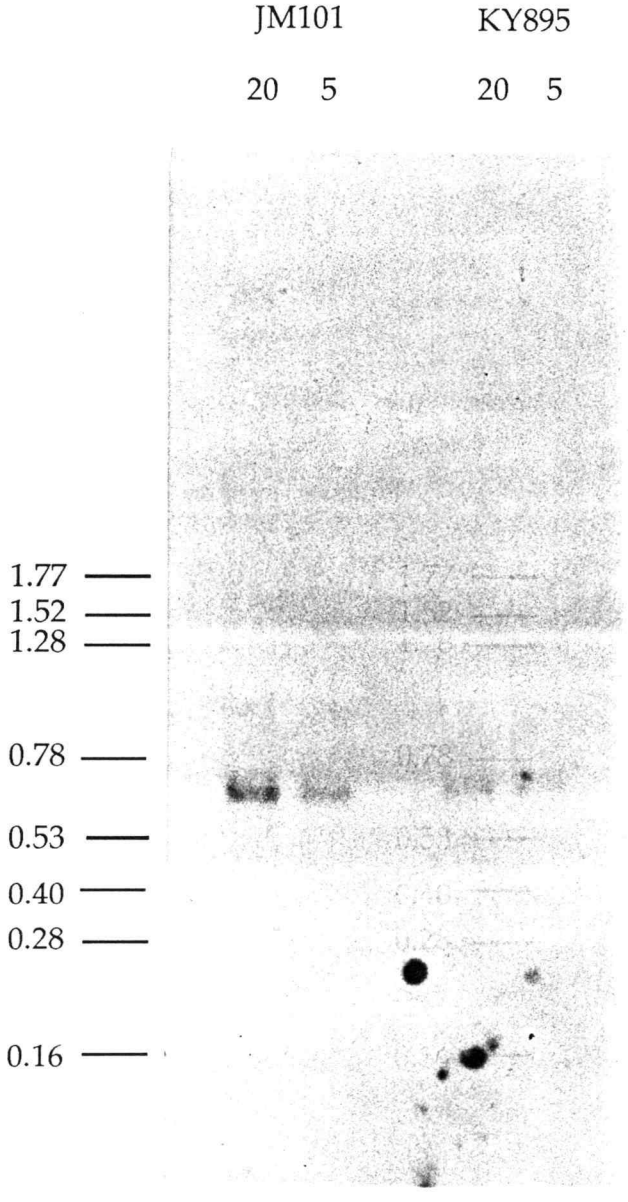


Figure IV.4

mutation rather than a deletion or insertion within the *tdk* locus.

Native molecular weight of E. coli thymidine kinase.

One of the major distinctions between type I (*Herpesviridae*) and type II (*Poxviridae* and vertebrate) thymidine kinases is that of quaternart structure (dimers vs. tetramers). It was therefore of interest to determine the native molecular weight of the *E. coli* TK enzyme. Clarified bacterial cell extracts from *E. coli* KY895 carrying the entire *tdk* gene on pUC118 were subjected to size exclusion chromatography on a Sephacryl-300 (S-300) column. A molecular weight of approximately 82kD was estimated by plotting the fraction containing peak TK activity against a standard curve derived from the analysis of gel filtration markers passed over the S-300 column (Fig. IV.5). Glycerol gradient separation of bacterial cell extracts also revealed a molecular weight of equivalent size (data not shown). From the predicted molecular weight of 23.5kD, a tetramer would be ~94kD and a trimer, ~70.5kD. The estimated ~82kD molecular weight derived from gel filtration and glycerol gradient analysis lies halfway between the predicted sizes for tetramer and trimer conformations, but significantly larger than a possible dimeric conformation (47kD). These results suggest that the native form of the *E. coli* TK is a multimeric enzyme. Although determination of the native conformational state was somewhat inconclusive in these crude extracts, based on the precedents of both the human (Sherely and Kelly, 1988) and vaccinia virus (Black and Hruby, 1990a) thymidine kinases which have been shown to be tetramers, it seems likely that the native *E. coli* TK is a tetrameric complex and therefore should be considered a type II TK enzyme.

Fig. IV.5 Native molecular weight determination of *E. coli* TK by Sephacryl-300 gel filtration. A standard curve of protein molecular weight markers was generated by passing a mixture of protein gel filtration molecular weight markers [ß-amylase (200kD), alcohol dehydrogenase (ADH, 150kD), bovine serum albumin (BSA, 66kD), carbonic anhydrase (car. anh., 29kD) and cytochrome c (cyt.c, 12.4kD) through a 1 x 34 cm Sephacryl-300 gel filtration column in TENDT as described in "Experimental Procedures". Bacterial extracts of *E. coli* KY895 carrying the entire *tdk* gene on pUC118 were passed over the column and the fractions containing TK were identified on the basis of enzyme activity. The dashed line corresponds to the chromatographic position of TK and its estimated molecular weight.

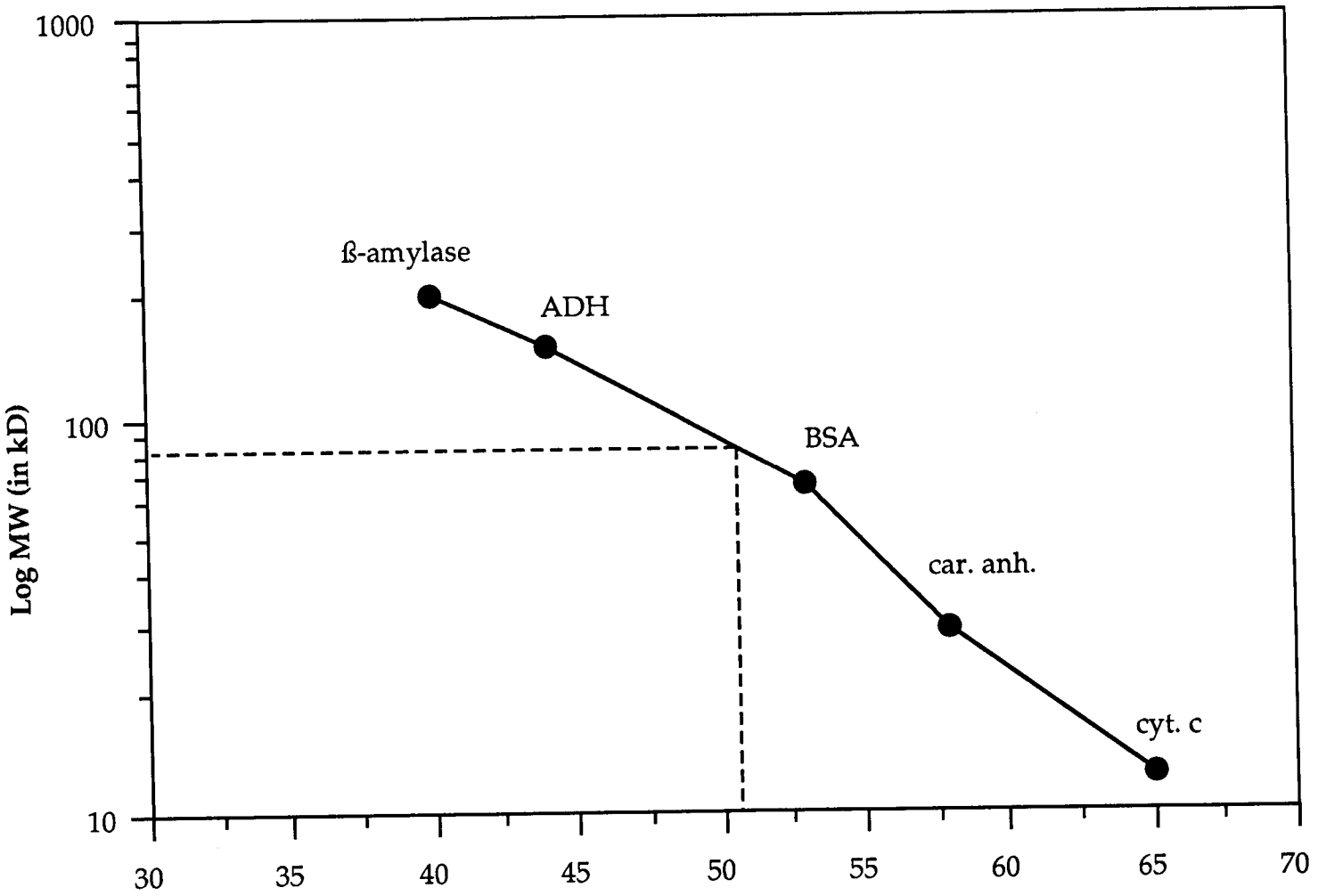

1000
100
10
Log MW (in kD)
ß-amylase
ADH
BSA
car. anh.
cyt. c
30
35
40
45
50
55
60
65
70

Figure IV.5

Evidence for domain conservation.

One of the objectives for mapping and sequencing the *tdk* gene from *E. coli* was to identify potential domains which had been conserved between the TK proteins encoded by *E. coli* and bacteriophage T4, and type II TK enzymes, such as that encoded by vaccinia virus (VV), which should provide information regarding essential domains. At the amino acid level the *E. coli* TK (ETK) shares 23% identity (35.6% similarity) with VV TK and 49.5% identity (65% similarity) with T4 TK, suggesting a closer relationship among T4 and *E. coli* enzymes, as might be expected. Comparison of VV TK with thymidine kinases from human, chicken, mouse, variola virus and monkeypox demonstrates seven domains of complete identity between all members of this group (Fig. IV.6). Interestingly, three of these domains (I, III, and VII) exhibit some conservation when the VV TK and the prokaryotic enzymes are aligned.

Upon inspection, residues in domain I (ETK residues 9-19) share similarity to consensus ATP binding site motifs. ATP consensus sequences generally comprise the sequence G/A X X X X G K where G is glycine, A, alanine, K, lysine and X is any amino acid (Moeller and Amons, 1985). Although the *E. coli* TK sequence [residues 9-16, S A M N A G K (S, serine; M, methionine; N, asparagine)] lacks a glycine or alanine residue at the first position in this motif, the novel use of a serine at this position reflects the substitution of a similar amino acid. In addition, it is located at the same relative position in the polypeptide as the ATP binding site of the vaccinia thymidine kinase (residues 11-18) (Black and Hruby, 1990b). Therefore, this site is likely to be the ATP binding site.

The D E X Q F motif (D, aspartic acid; E, glutamine; Q, glutamic acid; F, phenylalanine; X, any amino acid) in domain III (ETK residues 83-93)

Fig. IV.6 Alignment of the amino acid sequences of vaccinia virus TK (VV), *E. coli* TK (E.c.) and bacteriophage T4 TK (T4) enzymes. Vertical lines (|) mark identical amino acids. Colons (:) indicate similar amino acids. Boxes I through VII represent those domains which have been previously identified as being completely conserved between vaccinia virus, monkeypox, variola virus, chicken, mouse and human thymidine kinases (Black and Hruby, in 1990b). Boxes with little or no similarity to *E. coli* TK sequences are denoted by a hatched box while regions with greater similarity or complete identity are denoted by solid boxes. Standard single amino acid nomenclature is used.

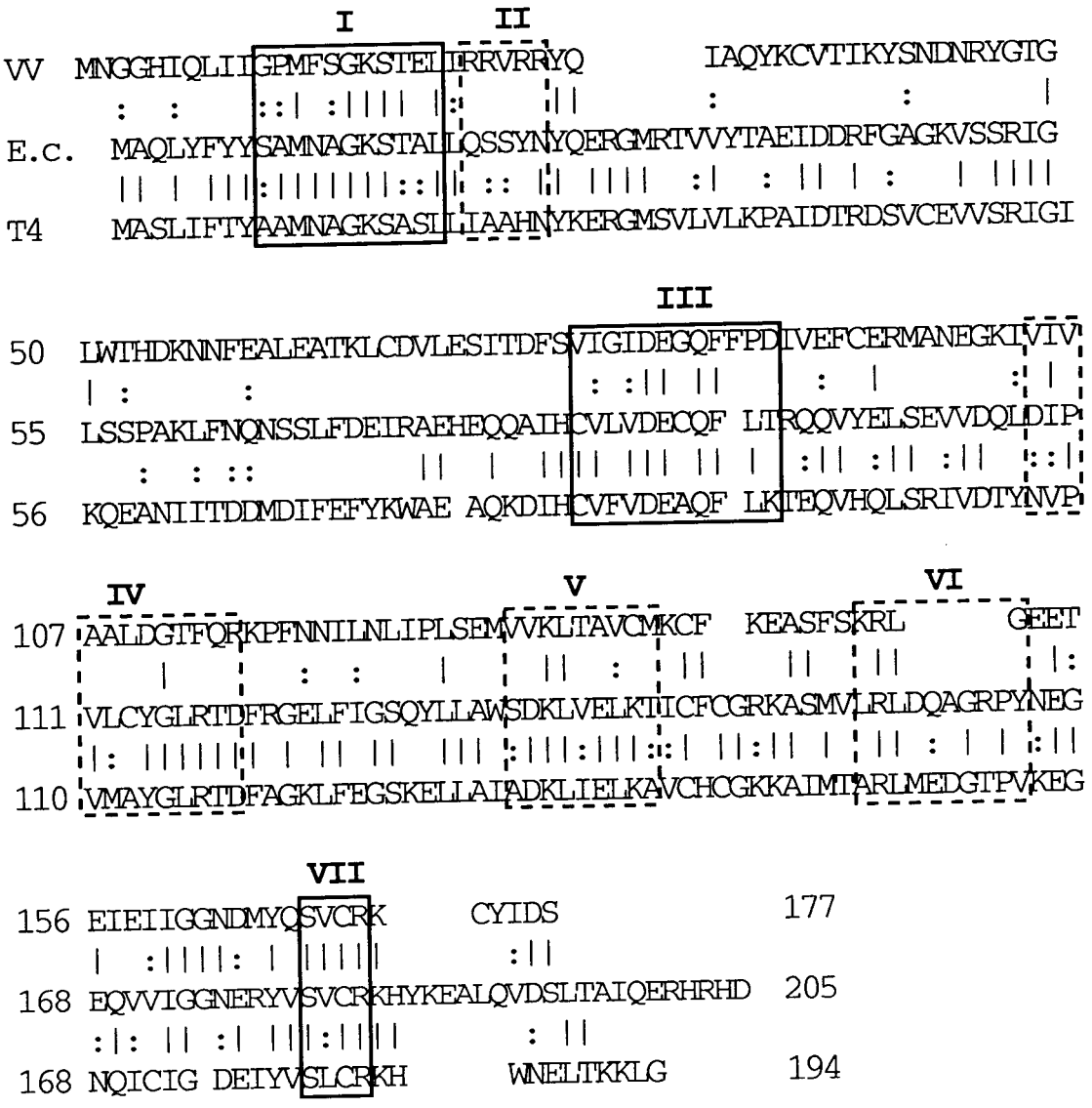

I
II
VV MNGGHIQLIIGPMFSGKSTELIRRVRRYQ IAQYKCVTIKYSNDNRYGIG
E.c. MAQLYFYYSAMNAGKSTALLQSSYNYQERGMRTVVYTAEIDDRFGAGKVSSRIG
T4 MASLIFTYAAMNAGKSASLLIAAHNYKERGMSVLVLKPAIDTRDSVCEVVSRIGI
III
50 LWTHDKNNFEALEATKLCDVLESITDFSVIGIDEGQFFPDIVEFCERMANEGKIVIV
55 LSSPAKLFNQNSSLFDEIRAEHEQQAIHCVLVDECQF LTRQQVYELSEVVDQLDIP
56 KQEANIITDDMDIFEFYKWAE AQKDIHCVFVDEAQF LKTEQVHQLSRIVDTYNVP
IV
V
VI
107 AALDGTFQRKPFNNILNLIPLSEMVVKLTAVCMKCF KEASFSKRL GEET
111 VLCYGLRTDFRGELFIGSQYLLAWSDKLVELKTICFCGRKASMVLRLDQAGRPYNEG
110 VMAYGLRTDFAGKLFEGSKELLAIADKLIELKAVCHCGKKAIMNARLMEDGTPVKEG
VII
156 EIEIIGGNDMYQSVCRK CYIDS 177
168 EQVVIGGNERYVSVCRKHYKEALQVDSLTAIQERHRHD 205
168 NQICIG DEIYVSLCRKH WNELTKKLG 194

Figure IV.6

appears to be highly conserved in all type II TK enzymes. In adenylate kinase, an aspartic acid residue (D119) preceded by a stretch of hydrophobic residues has been proposed to be involved in Mg^{2+} binding (Fry *et al.*, 1986). In fact, in all type II TKs, three of four residues directly preceding the aspartic acid in domain III are hydrophobic. Because of the high degree of similarity at and surrounding the aspartic acid (ETK residues D87) in domain III, it may serve a similar function, although this remains to be determined experimentally.

With the exception of one similar residue (leucine vs. valine) in T4 TK, domain VII (ETK residues 180-183) is identical to all type II TKs as well. While domain I comprises the putative ATP binding site and domain III has some similarity to a possible Mg^{2+} binding site, a putative function has not yet been assigned to domain VII. Likewise, since the 900bp *Eco*RI/*Sal*I fragment was able to confer TK activity to *E. coli* KY895 (Fig. IV.1), but lacked the sequences to encode the last eleven amino acids (Fig. IV.2), these residues appear not to be required for enzyme activity. Interestingly, the region spanning residues 199 to 205 is predicted to be highly hydrophobic. Because neither T4 TK nor VV TK has such a sequence it is difficult to postulate a functional role for this region in regulation and catalysis of enzyme activity.

Summary

The thymidine kinase gene (*tdk*) from *E. coli* has been mapped to ~minute 27.35 on the W3110 chromosome (Figs. IV.1 and IV.2). Sequence analysis of this region reveals a 615 nucleotide open reading frame flanked with the appropriate transcriptional regulatory signals (Fig. IV.3). The *tdk* gene is transcribed *in vivo*, giving rise to a 650 nt transcript (Fig. IV.4). The

E. coli TK enzyme is apparently active as a multimeric complex suggesting that it belongs to the type II class of TK enzymes (Fig. IV.5). Amino acid alignment of *E. coli* TK with VV TK and T4 TK revealed 23% identity (35.6% similarity) and 49.5% identity (65% similarity), respectively (Fig. IV.6). The majority of identical and similar residues were localized to three of seven domains (I, III and VII), which had been previously identified from alignment of vaccinia, monkeypox, variola virus, human, mouse and chicken thymidine kinases (Black and Hruby, 1990b). Domain I shares high homology with ATP binding site consensus sequences, domain III may be a Mg^{2+} binding site and the function of the conserved residues of domain VII remains to be elucidated. Type II thymidine kinases are complex enzymes which, in addition to requiring residues for ATP and Mg^{2+} binding, require those for thymidine and dTTP binding, subunit interaction and catalysis. While domains I, III and VII and other regions may be involved in one or more of these functions, it is likely that other functional domains may be created upon assembly of the enzyme complex itself. Addressing this possibility will await the purification and x-ray crystallographic analysis of a number of type II TK proteins.

CHAPTER V

PUTATIVE ROLES OF DOMAINS III, IV AND VII IN CATALYSIS AND REGULATION OF VACCINIA VIRUS THYMIDINE KINASE

Domain III - Magnesium Binding Site.

Magnesium and ATP bind as a complex (Mg-ATP) to many ATP binding enzymes. When bound to the enzyme, the positively charged magnesium coordinates with the β and γ phosphoryl groups of ATP with magnesium displaying a stronger interaction with the β phosphoryl group. This asymmetric coordination specifically weakens the O_9-Pγ bond causing a "kink" in the phosphoryl groups which subsequently increases the susceptibility of Pγ atoms for nucleophilic attack (Shinohara *et al.*, 1990). The broad description of Mg^{2+} as "structure breaking" (Williams, 1970) presumably then correlates with what is believed to occur in Mg-ATP binding enzymes.

Enzymes which bind ATP contain the general consensus sequence or descriptor consisting of two conserved domains denoted segment A and segment B (Walker *et al.*, 1982). Segment A comprises the G X X X X G K motif (G, glycine, X, any amino acid, K, lysine) which is described in Chapter III of this dissertation. Segment B, alternatively known as the magnesium binding site, has also been implicated in binding the Mg-ATP complex (Fry *et al.*, 1986). While type II thymidine kinases do not contain the exact consensus sequence (G Xh_{4-8} D [Xh, any hydrophobic residue, D,

aspartic acid]) as described by Fry *et al.*, (1986) there is a similar, highly conserved region found in all type II TK enzymes. This region spans domain III and consists of a D E G (A/C) Q (H) F (E, glutamic acid, A, alanine, C, cysteine, Q, glutamine, H, histidine, F, phenylalanine) motif (residues 82-86 in VVTK; parentheses indicate single occurrences of alternate residues), which is preceded by a stretch of hydrophobic residues (Fig. III.1).

Within a protein the Mg-phosphate group is positioned near a negatively charged residue located at the carboxyl end of a β strand (Bradley *et al.*, 1987). There are a number of examples of this structural motif in the literature. In adenylate kinase, Fry *et al.*(1986) have suggested from x-ray crystallography studies that an aspartic acid residue (D119) may be the primary residue involved with magnesium binding. Like the aspartic acid (D82) in VVTK, D119 is preceded by a series of hydrophobic residues. In electron density mapping of EF-TU, a magnesium ion has been located in the pocket formed by the β phosphoryl group of GDP and a loop connecting the β strand and α helix which contains the G X X X X G nucleotide binding motif (La Cour *et al.*, 1985). Situated opposite the β phosphate is the sequence D C P G H (P, proline, residues 80-84) (La Cour *et al.*, 1985). Folkers *et al.* (1989) have aligned this sequence with that of domain III of VVTK on the basis of secondary structure predictions, thereby giving further evidence that these residues may provide a similar function. By the use of a reactive-radiolabeled ATP analogue where the reactive group is covalently bound to the γ phosphate group of ATP, James *et al.* (1990) have reported that a highly conserved acidic region containing a highly conserved aspartic acid (D335) in mitochondrial creatine kinase may be the magnesium binding site. Large tumor antigen (T antigen) from simian

virus 40 (SV40) contains an aspartic acid residue (D474) within a cluster of hydrophobic residues which aligns in computer analyses of T antigen secondary structure predictions with x-ray crystal structures of adenylate kinase and *E. coli* EF-TU at D119 and D80, respectively (Bradley, *et al.*, 1987). This has led them to suggest a functional role for D474 in magnesium binding.

The removal of divalent cations from rabbit reticulocyte lysate cell-free translation products of pT7:TKII (vector is described in Wilson *et al.*, 1989) RNA by the addition of EDTA, by dialysis or gel filtration results in inactivation of thymidine kinase activity (data not shown). Enzyme activity can be regained by the addition of magnesium to the enzyme. Furthermore, the addition of 100mM EDTA during glycerol gradient sedimentation of radiolabeled thymidine kinase does not lead to dissociation of the tetrameric complex of this enzyme. These experiments suggest that magnesium is essential for catalytic activity but is apparently not required for subunit integrity. The question then arises as to whether other divalent cations, such as manganese, can substitute for magnesium in this reaction. The ionic radius of manganese is about 15% greater than that of magnesium, meaning that manganese occupies about 53% greater volumn than magnesium (Martin, 1990). Martin (1990) also suggests that magnesium associates predominantly and often exclusively at the phosphates of all nucleoside phosphates. To examine the role of magnesium in VVTK activity, cell free translation products of pT7:TKII RNA were extensively dialyzed against 20mM Tris, pH 7.6/2mM dithiothreitol/50μM thymidine at 4C and then assayed in the presence of increasing concentrations of magnesium, manganese or calcium. While equivalent concentrations (2-4mM) of both magnesium and manganese

gave maximal TK activity, calcium was unable to substitute for magnesium. Despite the report that magnesium associates predominantly with nucleoside phosphates, the fact that manganese forms stronger complexes than magnesium (Martin, 1990) may compensate for this and thus allow substitution of these cations in thymidine kinase.

To determine whether D82 of vaccinia virus thymidine kinase may be required for magnesium binding a number of site-directed mutations have been introduced into the transcription vector, pT7:TKII (Wilson *et al.*, 1989; Black and Hruby, 1990b). The amino acids replaced are depicted in Table V.1. TK assays of the mutants demonstrated a loss or greatly reduced level of enzyme activity. Only the aspartic acid to asparagine mutation N82 had marginal activity (Table V.1). Because single amino acid changes can lead to major conformational changes (Black and Hruby, 1990b), L-[^{35}S]methionine-labeled domain III mutants were subjected to glycerol gradient sedimentation to determine the presence of a single peak at ~80kDa. All mutants except the glycine 84 to valine mutant (V84) maintained a quaternary structure similar to wild-type TK. Contrary to the results with N82 (VVTK), Kim *et al.*, (1990) reported that exchange of the D119 implicated in magnesium binding in adenylate kinase with an asparagine residue does not alter enzyme activity, suggesting that D119 is not required.

One way to investigate whether cations interact differently between the wild-type and mutant TKs is to analyze the enzymes for their cation requirement. Mutant and wild-type TK cell-free translation products were extensively dialyzed to remove cations and then assayed for activity levels in the presence of increasing concentrations of either magnesium or manganese. The results of these experiments indicate that the divalent

Table V.1

Mutations generated in domain III and their enzyme activity levels relative to wild-type TK

Mutants	V I G I D E G Q F F P D A.A. 78-89	TK Activity
N82	N	+
I82	I	–
L82	L	–
V84	V	–
WT		++++

cation requirement of N82 is approximately 10mM, 3-4 times higher than that required by wild-type TK (data not shown). Replacement of D193 for an N residue in ribulose bisphosphate carboxylase has been demonstrated to alter the affinity of the enzyme for magnesium (Lolis and Petsko, 1990). In addition, this mutant enzyme is able to catalyze the remainder of the reaction, but just not the first step. This raises the question - is the N82 mutant of VVTK then still capable of performing all aspects of catalysis?

Because magnesium is thought to stabilize ATP when it is bound to the enzyme, it was of interest to determine whether the mutants which are apparently dysfunctional in binding magnesium cofactor are also dysfunctional in binding ATP. The ATP binding assay as described in Black and Hruby (1990b) was used to monitor the ATP binding ability of cell free translation products of the domain III mutants. Preliminary results indicate that V84 did not bind, while N82, L82 and I82 bound to the column. However, the elution profiles of L82 and I82 were distinct from wild-type TK and both required more ATP to elute them from the column. N82 apparently maintains a similar ability to bind ATP as the wild-type enzyme. Taken together these results suggest that while N82 is structurally and, at least partly functionally similar to wild-type TK, the loss of a negatively charged residue at position 82 alters the ability of TK to transfer the phosphate moiety of ATP to thymidine, possibly due to an inability to correctly position the γ phosphate for nucleophilic attack.

Domain VII - Putative Role in Catalysis.

One of the other very highly conserved domains in type II thymidine kinases, including the *E. coli* and T4 bacteriophage enzymes, is domain VII (Fig. III.1). This domain consists of S V C R (S, serine, R,

Table V. 2

Mutations generated in domain VII and their levels of enzyme activities relative to wild-type TK

Mutants	S V C R A.A. 168-171	**TK Activity**
V170	V	6.8%
F170	F	8%
G170	G	1.9%
S170	S	250%
A170	A	8.8%

arginine, V, valine; residues 168-171 in VVTK) in type II TKs with the single exception of a conserved substitution of a leucine in place of valine in T4 TK. In an alignment of secondary structure between Herpes Simplex Virus-1 (HSV), monkeypox and human thymidine kinases, the region corresponding to domain VII was identified as similar to A G C R at positions 334-337 in HSV-1 TK (Folkers and Trumpp, 1987). At the primary amino acid level the C R dipeptide aligns with the last two residues of domain VII. Darby *et al.* (1986) reported that an amino acid substitution at C336 (C to Y (tyrosine) in HSV-1 TK) had a dramatic effect on the K_m for ATP, and for the nucleoside substrate. While they did not determine whether this mutation led to a conformational alteration, they suggested that C336 is involved in the catalytic site, interacting with both the ATP and nucleoside binding sites.

Cysteine residues have been implicated to play a role in a number of other purine nucleotide binding enzymes. Nakano *et al.* (1990) reported modification by iodoacetamide at one of three conserved cysteine residues, C306, in glutamine synthetases from *Bacillus subtilis* or *B. cereus.* The finding that C278 of creatine kinase is at or near the substrate binding site proposed by Vandest *et al.* (1980) was supported by studies by James *et al.* (1990) using affinity labeling. This was further supported by NMR and EPR studies by McLaughlin *et al.* (1976), who demonstrated that C278 and the nucleotide and guanidine substrates are only separated by 7-8Å. In myosin, the distance between the ATP binding region and the thiols, C707 and C697, is 2-3 nanometer (nm) apart when the site is occupied. One or both cysteine side chains may be closer when the site is unoccupied, thereby suggesting that the binding of nucleotide causes displacement of one or both thiols. Other investigators have shown that C697 is within 0.3 to

0.45nm from K184 at the nucleotide binding consensus sequence only when the nucleotide binding site is unoccupied (Hiratsuka, 1987; Sutoh and Hiratsuka, 1988). In thymidylate synthases a completely conserved cysteine (C198 in *Lactobacillus* or C146 in *E. coli*) is involved in nucleophilic catalysis. This raises the question as to whether a cysteine residue of VVTK provides nucleophilic attack on the phosphate moiety of ATP.

Preliminary experiments were performed to determine whether a cysteine residue might be involved in catalysis. Cell-free translated VVTK RNA products were incubated in the presence of varying concentrations of the general alkylating agents, iodoacetamide or N-ethylmaleimide (Dougherty *et al.*, 1989), and then assayed for the ability to phosphorylate thymidine. At 0.5mM iodoacetamide, enzyme activity was inhibited about 20% while at 5mM and 50mM iodoacetamide, activity was inhibited 60% and 80%, respectively. Because the rabbit reticulocyte lysates contain many potentially reactive proteins, the concentration required to inhibit TK in this mix probably does not reflect the actual concentration needed to specifically inactivate TK. Furthermore the lysates also contain substrates (thymidine and ATP) which may block or interfere with the accessibility of iodoacetamide. Interestingly, Nakano and Kimura (1987) reported only 19% inhibition of purified glutamine synthetase in the presence of both substrates and 50mM iodoacetamide. Why VVTK is much more sensitive to iodoacetamide even in a crude extract than purified glutamine synthetase is unknown at this time and therefore must await further experimental results before a plausible response can be offered.

Unlike iodoacetamide, N-ethylmaleimide did not appreciably interfere with TK activity even at 50mM levels. The lack of inhibition by

N-ethylmaleimide may be due to the larger size of this compound with respect to iodoacetamide, which prevents its access into the catalytic pocket. Careful examination of this reaction must await purification of thymidine kinase. In any event these preliminary results suggest that a cysteine might be active in the catalytic site and has prompted further investigation.

Six cysteines residues are found in VVTK and the other type II TKs with the exceptions of capripox, fowlpox, *E. coli* and the T4 thymidine kinases. However, only two cysteines are completely conserved in all type II TKs, C141 and C170 in VVTK. The latter (C170) corresponds to C336 from HSV TK which Darby *et al.* (1986) suggest is involved in the catalytic site and has been the target for mutagenesis studies to determine the role, if any, of C170 in catalysis.

Five new residues have replaced C170 within the transcription vector, pT7:TKII, and are shown in Table V.2. Determination of enzyme activity levels revealed that four of the five mutations introduced lead to inactive enzymes whereas a C170 to S170 mutation resulted in a "super TK" phenotype with about 2.5-fold higher enzyme activity than wild-type TK. From glycerol gradient sedimentation patterns apparently only S170 has maintained a quaternary structure similar to that of wild-type TK, while all other mutants display major conformational changes. In ATP binding assays (Chapter III), only S170 bound and eluted with a similar pattern to wild-type TK. As expected from the glycerol gradient results all other mutants did not bind to the column. Furthermore, preliminary results indicate that S170 is still feedback inhibited by dTTP, suggesting that the increase in enzyme activity is not due to a loss of feedback inhibition. This result contrasts with that of a similar mutation made in the *E. coli* thymidylate synthase, where C146 was altered to a serine. While the S146

mutant retained enzyme activity it was severely diminished (Dev *et al.*, 1988). Model studies have demonstrated that thiols are more effective nucleophiles than hydroxyls in nucleophilic displacement reactions (Jencks, 1969). Preliminary experiments of S170 incubated with iodoacetamide suggests that this mutant is inhibited to the same degree as wild-type TK. This results questions the role of C170 in the catalytic site and in nucleophilic attack. The cysteine to serine mutation (C170 to S170) clearly has altered TK somehow to cause such a dramatic increase in enzyme activity, although the manner of this change remains nebulous and is currently being investigated.

The S170 iodoacetamide results indicate that perhaps another cysteine residue is involved in the catalytic pocket. Because C141 is the only other cysteine residue that is conserved in all type II TKs, attention is directed to this cysteine and adjacent residues. Close inspection reveals a conserved motif comprised of C X B X A S beginning with C141 (B, any basic, positively charged residue). The only exceptions to this motif are the presence of a cysteine at the third position in fowlpox virus, an isoleucine in place of the serine in T4 TK and a substitutable glycine for the alanine in Shope Fibroma Virus TK. Whether C141 is the, or one of the, iodoacetamide-reactive residue remains to be determined.

Domain IV - Putative Nucleoside Binding Site.

The predominance of a conserved motif for ATP or GTP binding sites (G X X X X G) has allowed for the identification of a number of such sites in a wide and functionally diverse group of proteins. While the consensus sequence for GTP-binding proteins contains an additional region (N K X D) which has been suggested to be the site of guanosine binding

(Bork and Grunwald, 1990), no comparable sequence has been described for other nucleoside binding sites, e.g., thymidine. The first experimental clue to a possible thymidine binding site came from two mutations in HSV-1 thymidine kinase, which had been selected for by virtue of their resistance to the antiherpetic drugs, E-5-bromovinyl-2'-deoxyuridine (BVdU) or 9-(2-hydroxyethoxy-methyl)guanine (acyclovir) (Larder *et al.*, 1983; Darby *et al.*, 1986). Sequence analysis of these mutants revealed single site mutations at A168 (alanine to threonine in the BVdU resistant mutant) or at R176 (arginine to glutamine in the acyclovir mutant) (Darby *et al.*, 1986). Substrate affinity determinations of the mutant enzymes suggested that these mutants do not have a direct involvement in thymidine binding, but may be positioned in close proximity to the nucleoside binding site (Darby *et al.*, 1986).

In the secondary structure alignment of HSV-1, monkeypox and human thymidine kinases (Folkers and Trumpp, 1987), the site of these mutations corresponds to a region just outside of domain IV, between domains IV and V of VVTK. Because HSV-1 TK is able to phosphorylate the guanosine analog, acyclovir, and the demonstration that an aspartic acid residue is directly involved in guanine binding (La Cour *et al.*, 1985), the highly conserved aspartic acid at position 162, only 4 residues away from A168 described by Darby *et al.* (1986), has been implicated in nucleoside binding (Folkers and Trumpp, 1987). The motif, F D R, containing D162 corresponds in the Folkers and Trumpp (1987) alignment with a similar sequence from monkeypox and human thymidine kinases, that of F Q R. That type II TKs are not susceptible to acyclovir lends further credence to the hypothesis that D162 may be involved in binding the guanine moiety of acyclovir while the presence of a Q at the same relative

position (residue 114) in VVTK defines the strict substrate specificity (thymidine) of most type II thymidine kinases. It is of interest that both the *E. coli* and T4 thymidine kinases lack either the F Q R or F D R motif. However, at the carboxy end of domain IV in these enzymes, these amino acids are represented, albeit in a somewhat jumbled arrangement; R T D F R/A. While the presence of an aspartic acid in this sequence suggests that these TKs would be sensitive to acyclovir, the altered assortment of important residues is likely to translate into differences in the binding pocket which may influence substrate specificities.

Initial studies have been directed towards identifying the role of Q114 in nucleoside binding of VVTK. As a control, the HSV-1 TK open reading frame was obtained from Dr. Gary Merrill (Department of Biochemistry and Biophysics, Oregon State University) and cloned into the transcription vector, pT7:TKII, in place of the VVTK open reading frame. The glutamine at residue 114 in VVTK was replaced by a either histidine or aspartic acid. Thymidine kinase assays of HSV, domain IV mutant and wild-type VVTKs produced *in vitro* indicated that all enzymes had equivalent levels of activity. When ^{3}H [*side chain*] acyclovir replaced ^{3}H-*methyl* thymidine in these assays quite a different story emerged. Setting HSVTK activities at 100%, wild-type VVTK was inactive whereas the H114 and D114 mutants displayed 47% and 44% of HSVTK activities, respectively. Therefore, by altering a single amino acid in the putative nucleoside binding site, the substrate specificity has been dramatically changed. The implication that a single amino acid change can alter the basic function of the target enzyme in the treatment of HSV infections by nucleoside analogs is not a favorable one. Studies are currently underway to determine what other nucleoside these mutants can phosphorylate,

especially those which HSVTK can recognize, e.g., deoxycytidine and nucleoside analogs with demonstrated antiherpetic efficacy such as BVdU and gancyclovir, but which VVTK remains unable to phosphorylate.

The alteration in substrate specificity of these mutants then raises the question as to how else these enzymes resemble HSVTK. Because VVTK is feedback inhibited by dTTP and HSVTK is not, it was of interest to determine the level of dTTP sensitivity in the domain IV mutants. Glycerol gradient partially purified TKs were assayed for their ability to phosphorylate thymidine in the presence of increasing concentrations of dTTP. As had been expected from previous work (Hruby, 1985), wild-type VVTK was completely inhibited at 25μM dTTP while HSVTK was not inhibited even at concentrations as a high as 50μM dTTP. When H114 and D114 were assayed, they displayed similar levels of activity as HSVTK at 50μM dTTP. This strongly suggests that Q114 is involved in feedback inhibition.

Another way to evaluate these mutants for interaction with dTTP is to bind radiolabeled enzymes to an ATP agarose affinity column and then attempt to elute TK from the column with a gradient of dTTP followed by a pulse of ATP to remove any remaining bound enzyme from the column (Black and Hruby, 1990b). This technique has been previously used to demonstrate the interaction of dTTP with ATP binding in VVTK (Black and Hruby, 1990b). Results from this experiment demonstrated that VVTK is largely removed from the column by dTTP, with a small amount eluted with ATP (Fig. V.1). HSVTK was not eluted with dTTP, but the bound enzyme was removed by a pulse of ATP. Interestingly, both mutants show elution profiles somewhere between the VVTK and the HSVTK profiles, suggestive of an altered interaction between ATP and

Fig. V.1 Elution profile of L-[^{35}S]methionine labeled wild-type and domain IV mutant TKs from an ATP agarose affinity column with dTTP. Cell free translation products of HSV TK, wild-type VVTK and the two domain IV mutants mRNA were loaded onto an ATP agarose affinity column. A gradient of 0 to 50μM dTTP was passed over the column. The column was throughly washed with buffer (no nucleotide) and then pulsed with 1ml of 100mM ATP, followed by buffer minus nucleotide. 100ul aliquots of the collected fractions (1ml) were analyzed and adjusted to the percentage of CPMs loaded as in Fig. III.5. The left panel shows the elution pattern of HSVTK and wild-type TK and the right panel shows those of the two domain IV mutants, D114 and H114.

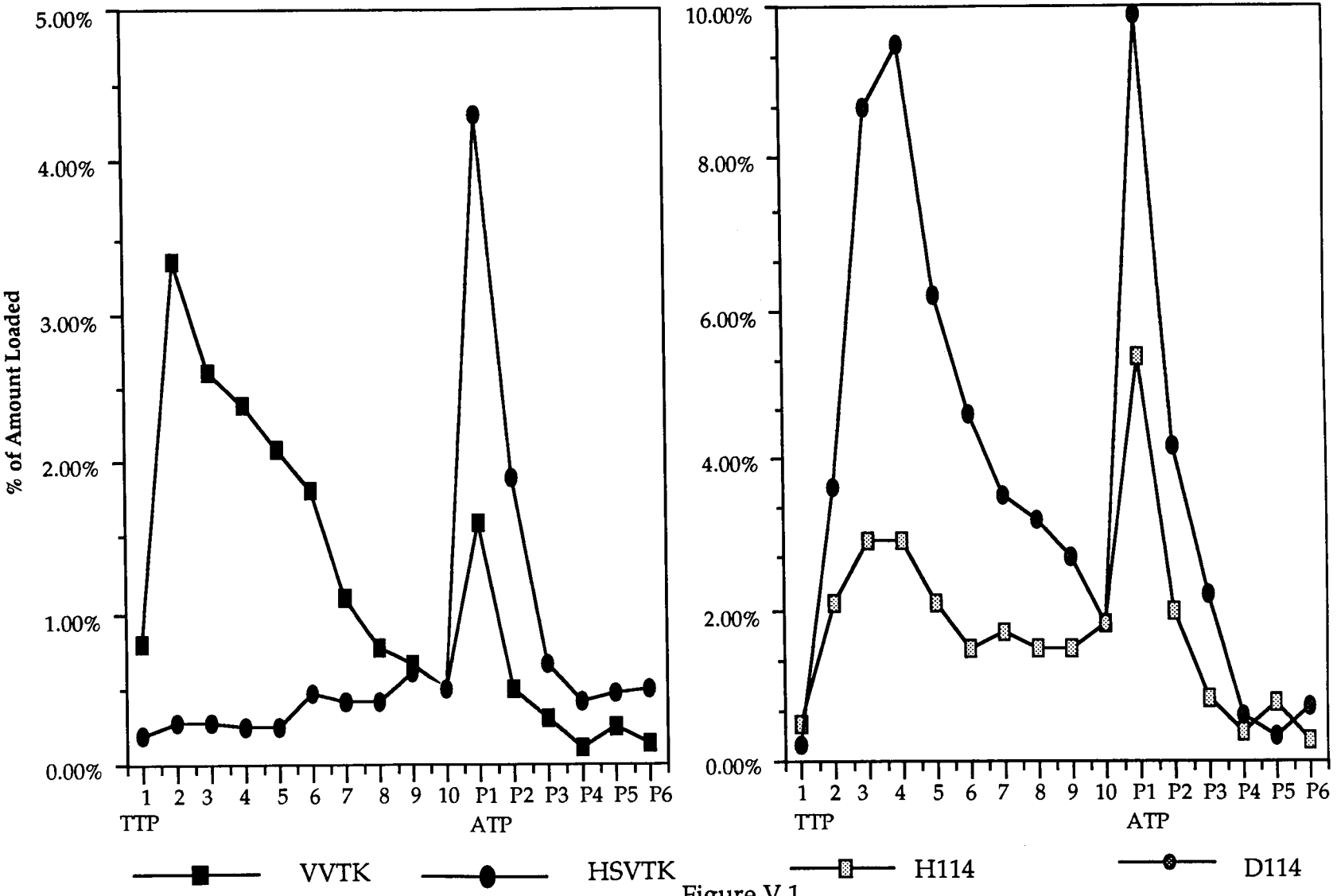

% of Amount Loaded
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
1 2 3 4 5 6 7 8 9 10 P1 P2 P3 P4 P5 P6
TTP
ATP
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
1 2 3 4 5 6 7 8 9 10 P1 P2 P3 P4 P5 P6
TTP
ATP
VVTK
HSVTK
H114
D114

Figure V.1

dTTP in H114 and D114 and providing further evidence that domain IV residues comprise part of the catalytic site. What effects the differences in substrate specificity and sensitivity to dTTP have *in vivo* with regard to nucleotide pool levels and the role of VVTK in viral DNA synthesis and replication, is currently being investigated.

Global VVTK Structure

X-ray crystallographic analysis of a number of nucleotide binding proteins has revealed a common structure comprising four or five β strands, with that number apparently being required for structural stability (Bradley *et al.*, 1987). Chou and Fasman algorithm of VV TK secondary structures predicts at least four and possibly five β sheets. While not exactly matching our computer predictions, Folkers and Trumpp (1987) indicate the location of five β sheets in their analysis of monkeypox and human thymidine kinases. Residues involved in binding substrates or cofactors are generally situated in the loops connecting the sheets with helices. While 4-5 β sheets are a common feature in nucleotide-binding proteins, the order of β strands and α helices may differ from protein to protein (La Cour *et al.*, 1985). Inspection of VVTK secondary structure predictions reveals residues in domains I (ATP), III((Mg^{2+}) and IV (nucleoside) all to be between predicted sheet and helix structures or at predicted turns. The structure encompassing domain VII (catalysis) residues is predicted to have a high propensity for β sheet formation. C198 of *Lactobacillus casei* thymidylate synthase, which has been demonstrated to be required for nucleophilic attack and to bind covalently to the C6 of FdUMP (fluorodeoxyuridine monophosphate), is part of a β sheet conformation, suggesting that the requirement for catalytic residues to be

unstructured is not strict (Montfort *et al.*, 1990). Because VVTK is a tetrameric complex it is difficult at this juncture to envision whether the catalytic site is comprised of essential domains and/or residues contributed by all or some of the subunits or whether each monomer is capable of performing catalysis independently. VVTK is being purified for biochemical and x-ray crystallographic studies which will aid in elucidation of the molecular structure of this complex and important enzyme.

BIBLIOGRAPHY

Bachman, B. J. (1983) *Microbiol. Rev.* **47**:180-230.

Bajszar, G., Wittek, R. Weir, J. B. and Moss, B. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **45**:62-72.

Ballou, L. M., Siegmann, M. and Thomas, G. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**:7154-7158.

Black, M. E. and Hruby, D. E. (1990a) *Biochem. Biophys. Res. Comm.* **169**:1080-1086.

Black, M. E. and Hruby, D. E. (1990b) *J. Biol. Chem.*. **265**:17584-17592.

Black, M. E. and Hruby, D. E. (1991a) *Molec. Micro.* **5**:373-379.

Black, M. E., Wilson, E. M., Hodges, W. M. and Hruby, D. E. (1991b) *Life Sciences Advances-Virology*, in press.

Bork, P. and Grunwald, C. (1990) *Eur. J. Biochem.* **191**:347-358.

Boyle, D. B., Coupar, B. H. E., Gibbs, A. J., Seigman, L. J. and Both, G. W. (1987)*Virology* **156**: 355-365.

Bradley, M. K., Smith, T. F., Lathrop, R. H., Livingston, D. M. and Webster, T. A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**:4026-4030.

Buller, R. M., Smith, G. L., Cremer, K., Notkins, A. L. and Moss, B. (1985) *Nature (London)* **317**:813-8815.

Child, S. J., Palumbo, G. J., Buller, R. M. and Hruby, D. E. (1990) *Virology* **174**:625-629.

Creighton, T. E. (1984) *Proteins: Structures and Molecular Properties*, W.H. Freeman and Co., New York.

Darby, G., Larder, B. A. and Inglis, M. M. (1986) *J. Gen. Virol.* **67**: 753-758.

Davison, A. J. and Scott, J. E. (1986) *J. Gen. Virol.* **67**:1759-1816.

Dev, I. K., Yates, B. B., Leong, J. and Dallas, W. S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **84**:1472-1476.

Dever, T. E., Glynias, M. J. and Merrick, W. C. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **84**:1814-1818.

Dougherty, W. D., Park, T. D., Cary, S. M., Bazan, J. F. and Fletterick, R. J. (1989) *Virology* **172**:302-310.

Esposito, J. J. and Knight, J. C. (1984) *Virology* **135**:561-567.

Flemington, E., Bradshaw, Jr., H. D., Traina-Dorge, V., Slagel, V. and Deininger, P. L. (1987) *Gene* **52**:267-277.

Folkers, G., Krickl, S. and Trumpp, S. (1989) *Arch. Pharm. (Weinheim)* **322**:409-413.

Folkers, G. and Trumpp, S. (1987) *Med. Sci. Res.* **15**:1495-1496.

Frearson, P. M., Kit, S. and Dubbs, D. R. (1965) *Cancer Res.* **25**:737-744.

Fry, D. C., Kuby, S. A. and Mildvan, A. S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**: 907-911.

Gentz, R., Rauscher, F. J. III, Abate, C. and Curran, T. (1989) *Science* **243**:1695-1699.

Gershon, P. D. and Black, D. N. (1989) *J. Gen. Virol.* **70**:525-533.

Goebel, S. J., Johnson, G. P., Perkus, M. E., Davis, S. W., Winslow, J. P. and Paoletti, E. (1990)*Virology* **179**:247-266.

Gompels, U. and Minson, A. (1986) *Virology* **153**:230-247.

Hiratsuka, T. (1987) *Biochemistry* **26**:3168-3173.

Hirsch, M. S. and Kaplan, J. C. (1990) in *Fields Virology* (B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnickand T. P. Monath, Eds.) pp. 441-468. 2nd Edition. Raven Press, New York.

Hruby, D. E. (1985) *Virus Research* **2**:151-156.

Hruby, D. E. and Ball, L. A. (1981a) *J. Virol.* **40**:456-464.

Hruby, D. E. and Ball, L. A. (1981b) *Virology* **113**:594-601.

Hruby, D. E. and Ball, L. A. (1982) *J. Virol.* **43**:403-409.

Hruby, D. E. and Guarino, L. A. (1984) *Virus Research* **1**:315-320.

Hruby, D. E., Maki, R. A., Miller, D. B. and Ball, L. A. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**:3411-3415.

Igarashi, K., Hiraga, S. and Yura, T. (1967) *Genetics* **57**:643-654.

Ink, B. S. and Pickup, D. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**:1536-1540.

Ives, D. H., Morse, Jr., P. A. and Potter, V. R. (1963) *J. Biol. Chem.* **238**:1467-1474.

James, P., Wyss, M., Lutsenko, S., Wallimann, T. and Carafoli, E. (1990) *FEBS* **272**:139-143.

Jencks. W. P. (1969) in *Catalysis in Chemistry and Enzymology*, ed. Jencks W. P. McGraw-Hill, New York, New York, pp.43-84.

Jones, E. V. and Moss, B. (1985) *J. Virol.* **53**:312-315.

Jones, E. V., Puckett, C. and Moss, B. (1987) *J. Virol.* **61**:1765-1771.

Keck, J. G., Baldick, Jr., C. J. and Moss, B. (1990) *Cell* **61**:801-809.

Kim, H. J., Nishikawa, S., Tokutomi, Y., Takenaka, H., Hamada, M., Kuby, S. and Uesugi, S. (1990) *Biochemistry* **29**:1107-1111.

Kit, S., Jorgensen, G. N., Liav, A. and Zaslavsky, V. (1977) *Virology* **77**:661-676.

Kit, S., Kit, M., Qavi, H., Trkula, D. and Otsuka, H. (1983) *Biochem. Biophys. Acta* **741**:158-170.

Kohara, Y., Akiyama, K. and Isono, K. (1987) *Cell* **50**:495-508.

Kouzarides, T. and Ziff, E. (1988) *Nature* **336**:646-6651.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**:488-492.

Kwoh, T. J. and Engler, J. A. (1984) *Nucl. Acids Res.* **12**:3959-3971.

La Cour, T. F. M., Nyborg, J., Thirup, S. and Clark, B. F. C. (1985) *EMBO J.* **4**:2385-2388.

Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988) *Science* **243**:1681-1688.

Larder, B. A., Derse, D., Cheng, Y-C. and Darby, G. (1983) *J. Biol. Chem.* **258**:2027-2033.

Lewis, J. A. (1986) *Mol. Cell Biol.* **6**:1998-2010.

Lin, P. F., Lieberman, H. B., Yeh, D. B., Xu, T., Zhao, S. Y. and Ruddle, F. H. (1985) *Mol. Cell. Biol.* **5**:3149-3156.

Littelfield, J. W. (1966) *Biochim. Biophys. Acta* **114**:398-403.

Littler, E. and Arrand, J. R. (1988) *J. Virol.* **62**:3892-3895.

Liu, Q. and Summers, W. C. (1988) *Virology* **163**:638-642.

Lolis, E. and Petsko, G. A. (1990) *Annu. Rev. Biochem.* **59**:597-630.

Lowe, D. G., Capon, D. J., Delwart, E., Sakaguchi, A. J., Naylor, S. L. and Goeddel, D. V. (1987) *Cell* **48**:137-146.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, New York, USA: Cold Spring Harbor Laboratory Press.

Martin, R. B. (1990) *Metal Ions in Biological Systems.* Marcel Dekker, Inc., New York, USA. Eds. H. Sigel and A. Sigel. **26**:1-13.

Mathews, C. K. and Van Holde, K. E. (1990) *Biochemistry* Benjamin/Cummings Publishing Co. Redwood City, CA pp. 742-778.

McAuslan, B. R. (1963) *Virology* **21**:383-389.

McAuslan, B. R. and Joklik, W. K. (1962) *Biochem. Biophys. Res. Comm.* **8**:486-491.

McLaughlin, A. C., Leigh, J. S. and Cohn, M. (1976) *J. Biol. Chem.* **251**:2777-2787.

McKnight, S. L. (1980) *Nucl. Acids Res.* **8**:5949-5964.

Merrill, G. F., Harland, R. M., Groudine, M. and McKnight, S. L. (1984) *Molec. Cell. Biol.* **4**:1769-1776.

Moeller, W. and Amons, R. (1985) *FEBS* **186**:1-7.

Montfort, W. R., Perry, K. M., Fauman, E. R., Finer-Moore, J. S., Maley, G. F., Hardy, L., Maley, F. and Stroud, R. M. (1990) *Biochemistry* **29**:6964-6977.

Morgan, C. (1976) *Virology* **73**: 43-58.

Moss, B. (1985) In *FieldsVirology"*(B. Fields, D. M. Knipe, R. M. Chanock, J. L. Melnick, B. Roizman and R. E. Shope, Eds.), pp. 685-704. Raven Press, New York.

Moss, B. (1990) In Field's Virology (B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnickand T. P. Monath, Eds.) pp. 2079-2079. 2nd Edition. Raven Press, New York.

Nakano, Y., Itoh, M., Tanaka, E. and Kimura, K. (1990) *J. Biochem.* **107**:180-183.

Nakano, Y. and Kimura, K. (1987) *Biochem. Biophys. Res. Comm.* **142**:475-482.

Nunberg, J. K., Wright, D. K., Cole, G. E., Petrovskis, E. A., Post, L. E., Compton, T. and Gilbert, J. H. (1989) *J. Virol.* **63**:3240-3249.

O'Shea, E. K., Rutkowski, R. and Kim, P. S. (1989) *Science* **243**:538-542.

Otsuka, M. H. and Kit, S. (1984) *Virology* **135**:316-330.

Parsonage, D., Al-Shawi, M. K., and Senior, A. E. (1988) *J. Biol. Chem.* **263**:4740-4744.

Payne, R. C. and Traut, T. W. (1982) *J. Biol. Chem.* **257**:12485-12488.

Reinstein, J., Brune, M. and Wittinghofer, A. (1988) *Biochemistry* **27**:4712-4720.

Robertson, G. R. and Whalley, J. M. (1988) *Nucl. Acids Res.* **23**:11303-11317.

Rozen, F., Pelleteir, J., Trachsel, H. and Sonenberg, N. (1989) *Mol. Cell. Biol.* **9**:4061-4063.

Sanger, F., Milklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci., USA* **74**:5463-5467.

Sherely, J. L. and Kelly, T. J. (1988) *J. Biol. Chem.* **263**:375-382.

Shinohara, Y., Yoshikawa, K., Terada, H. (1990) *Biophys. Chem.* **36**:201-208.

Shuman, S. (1990) *J. Biol. Chem.* **265**:11960-11966.

Shuman, S. and Moss, B. (1988) *J. Biol. Chem.* **263**:6220-6225.

Slabaugh, M. B., Johnson, T. L. and Mathews, C. K. (1984) *J. Virol.* **52**:507-514.

Smith, G. L., deCarlos, A. and Chan, Y. S. (1989) *Nucl. Acids Res.* **17**:7581-7590.

Stubblefield, E. and Murphree, S. (1967) *Exp. Cell Res.* **48**:652-656.

Studier, F. W. (1973) *J. Mol. Biol.* **79**:237-248.

Sutoh, K. and Hiratsuka, T. (1988) *Biochemistry* **27**:2964-2969.

Tengelsen, L. A., Slabaugh, M. B., Bibler, J. K. and Hruby, D. E. (1988) *Virology* **164**:519-527.

Towbin, H., Staehelin, T, and Gordon, J. (1979)*Proc. Natl. Acad. Sci. U.S.A.* **76**:4350-4354.

Traut, T. W. (1988) *CRC Crit. Rev. Biochem.* **23**:121-169.

Turner, R. and Tjian, R. (1989) *Science* **243**:1689-1694.

Upton, C. and McFadden, G. (1986) *J. Virol.* **60**:920-927.

Valerie, K., Stevens, J., Lynch, M., Henderson, E. E. and de Riel, J. K. (1986) *Nucl. Acids Res.* **21**:8637-8654.

Vandest, P., Lappe, J-P. and Kassab, R. (1980) *Eur. J. Biochem.* **104**:433-442.

Van Slyke, J. K. and Hruby, D. E. (1990) *Current Topics in Microb. and Immun.* **163**:18-206.

Vos, J. and Stunnenberg, H. (1988) *EMBO J.* **7**:3487-3492.

Wagner, M. J., Sharp, J. A. and Summers, W. C. (1981) *Proc. Natl. Acad. Sci., U.S.A.* **78**:1441-1445.

Walker, J. E., Saraste, M., Ruswick, M. J. and Gay, N. J. (1982) *EMBO J.* **1**:945-951.

Weir, J. P and Moss, B. (1983a) *J. Virol.* **46**:530-537.

Weir, J. P. and Moss. B. (1983b) *Nucl. Acids Res.* **13**:985-998.

Wilson, E. M., Edbauer, C. and Hruby, D. E. (1988) *Virus Genes* **2**:31-48.

Wilson, E. M., Franke, C. A., Black, M. E., and Hruby, D. E. (1989) *Gene* **77**:69-78.

Williams, R. J. P. (1970) *Quart. Rev. Chem. Soc.* **24**:313-324.

Yuen, L. and Moss, B. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**:6417-6421.