Title
Sequence determination, expression, and site-directed mutagenesis of creatine kinase

Author
Babbitt, Patricia Clement

Publication Date
1988

Peer reviewed|Thesis/dissertation

SEQUENCE DETERMINATION, EXPRESSION, AND SITE-DIRECTED MUTAGENESIS OF CREATINE KINASE

by

Patricia Clement Babbitt

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

Acknowledgments

In the ordinary course of events, a graduate student has the opportunity in these acknowledgments to thank his or her mentor, the person who has been most influential in the student's development. How extraordinarily fortunate I have been in having in effect, three such mentors, each of whom has contributed to my education in a special way.

First, I would like to thank my two research advisors, for whom I worked jointly during my graduate career. Dr. George L. Kenyon not only taught me enzymology and protein chemistry but excitement about science. His example of intensity and focus has become my own sense about what being a scientist is really about. Dr. Irwin D. Kuntz ("Tack") not only taught me a great deal about the physical chemistry of proteins but also to be thoughtful and patient. He corrected my thinking when it wandered into dubious speculation and responded to the problems which occasionally came my way with concern and gentleness. The third scientist I include as my mentor is Dr. Brian L. West. He spent two years teaching me molecular biology when it was neither his job nor particularly to his advantage to do so. He gave his time to me because of his excitement about our project and his non-egotistical attitude about doing research. His example of openess and cooperation in scientific inquiry is as great a gift as his on-going intellectual and technical help.

There are many others who helped me during the course of my graduate career. I would like to thank in particular Drs. Phyllis Kosen, Margaret Phillips, and Fred Cohen for many helpful and stimulating discussions on my work and on theirs. I would especially like to thank my friend and colleague, the "soon-to-be-Dr." Vincent Powers, who talked and learned with me during the five years of our friendship. Vince, I couldn't have chosen a better bench-mate or a colleague from

whom I could learn more. In addition, I would like to thank a few of the many other scientists whose helpful discussions made my work smoother and more fruitful: Drs. Norman Oppenheimer, Charles Craik, Ken Dill, C.C. Wang, Alice Wang, Tom Marschner, Nariman Nabor, George Garcia and Mr John Vasquez.

I would also like to acknowledge and thank those non-scientists in my life whose patience, understanding and support have made my graduate student days one of the truly best times of my life. First, my husband David has been supportive when I was down, proud over each small and large success, and patient over what must have seemed to him an interminable length of time to be a student. Second, my parents have been proud of me and encouraging throughout. Third, I would like to thank my good friends, Sharon and Charles Gustavson, Sue Sherk, Kathy Clement, my sister, and again my husband, David, for their interest in my work when they really had no idea what I was doing, for the celebration of each milestone along the way, from my acceptance into the program to my graduation, and for being "in my corner" during all the times I was unsure, insecure, or just plain tired.

Finally, I would like to thank both the Regents of the University of California for their financial support through the Graduate Opportunity Fellowships and the Graduate Division of the University of California at San Francisco for making available emergency student loans which I used several times during financial emergencies.

Sequence Determination, Expression, and Site-Directed Mutagenesis of Creatine Kinase

by Patricia C. Babbitt

Abstract

Creatine Kinase has been extensively studied by enzymologists since its discovery by Kuby in 1954. It is an important enzyme in the maintenance of intracellular energy, and its isozymic composition in plasma has become an important diagnostic indicator of cardiac damage. Nevertheless, until recently, almost no structural information has been available about this important enzyme. As part of the work described in this thesis, we have sequenced a full-length cDNA clone coding for creatine kinase from Torpedo californica electric organ. We have compared the inferred amino acid sequence with recently published amino acid sequences from nine other species and have analyzed the relationships among them. In addition, we have searched available sequence databases for proteins which might be related to creatine kinase by primary structure similarity. We found no significant similarities with any other such proteins except for lombricine kinase and taurocyamine kinase which are evidently primordial forms of creatine kinase. We have also expressed this creatine kinase cDNA in E. coli using a tac-related promoter and have reactivated the resulting aggregated protein to nearly normal specific activity. Site-specific mutagenesis of an iodoacetate-sensitive cysteine residue to alanine has shown that this cysteine residue is not essential to creatine kinase activity since the mutant shows 20% of wild-type activity.

George L. Kenyon

John D. Kuntz Jr.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Since the early discovery of enzymes, enzymologists have strived to understand the structure and function of catalytic proteins at increasingly sophisticated levels. Ranging from the elegant work in deducing the Krebs cycle (1) to the integrated understanding of structure and function in the serine proteases (2, 3), our understanding of enzymes has developed into an enormous and detailed body of knowledge which is increasingly useful to the well-being of our society as a whole.

Just as technical innovations in the past have sparked sudden progress in the development of new tools for the enzymologist (such as pH studies, isotope effect studies, kinetic analysis, affinity labeling and chemical modification studies), recent technical advances have been combined with new insights to develop one of the most powerful tools of all, recombinant DNA technology. In the few years since the groundbreaking work by Cohen, et al., (4), these techniques have become the methods of choice for deducing protein primary structures (inferred from DNA sequencing) (5), for producing large amounts of previously difficult to obtain proteins such as insulin (6) and human growth hormone (7), and for developing new protein therapeutic agents such as interferon (8).

Enzymologists and investigators of protein structure are applying these technologies in other ways as well. In addition to chemical modification studies to deduce active site structures or the role of particular residues in catalysis, protein modification experiments are creating mutant proteins whose study can be directed to similar goals. It is now becoming routine to see interdisciplinary efforts which involve molecular biologists to clone, sequence, and express wild type and mutant proteins, crystallographers or solution structure scientists to deduce their structures at atomic levels, and enzymologists

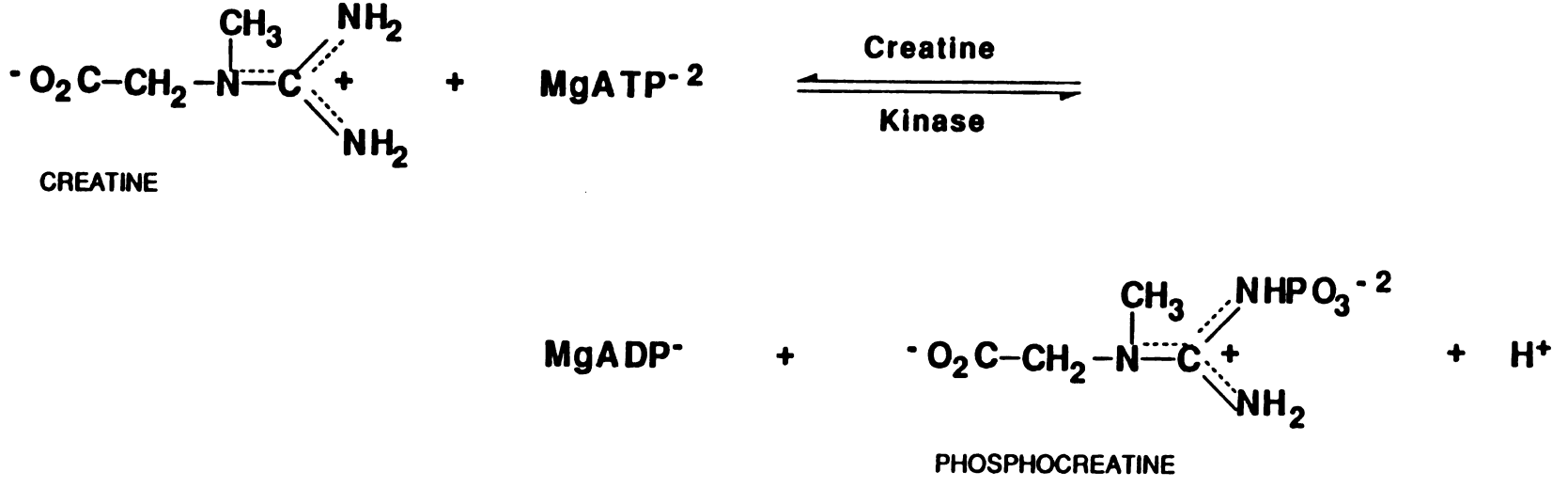
to study and compare the steps in catalysis. In addition to its contribution to classical modes of understanding, the ability to make and express such mutant proteins allows us to explore the interplay of structure and function in enzymatic catalysis in ways that are entirely new and which promise to expand our understanding enormously. The work described in this thesis is based on this new field of protein research and the question I have tried to answer was conceived in the spirit of these new modes of inquiry.

The enzyme we are studying is creatine kinase (CK). It is an important and interesting protein which we can understand better through the use of these new technologies Because a great deal is already known about it, it is also a good model system with which to develop these tools in our laboratory. A short description of CK and some of its characteristics follows.

Creatine kinase (E.C. 2.7.3.2, adenosine-5' triphosphate:creatine phosphotransferase) was first isolated in 1954 by Kuby et al. (9). It is a key enzyme in the bioenergetics of muscle, heart, and brain and acts to provide an ATP "reservoir" in energy-requiring tissues (10). It catalyzes the reversible phosphorylation of creatine by ATP to produce phosphocreatine and ADP as shown in Figure 1 (11).

Because this enzyme has been well-characterized by classical enzymological methods and there are several good reviews which summarize the extensive information available (11-14), it is unnecessary to discuss the full range of CK literature in this thesis. Therefore, only a brief review of some creatine kinase studies pertinent to this project is included here.

Figure 1: The Reaction Catalyzed by Creatine Kinase

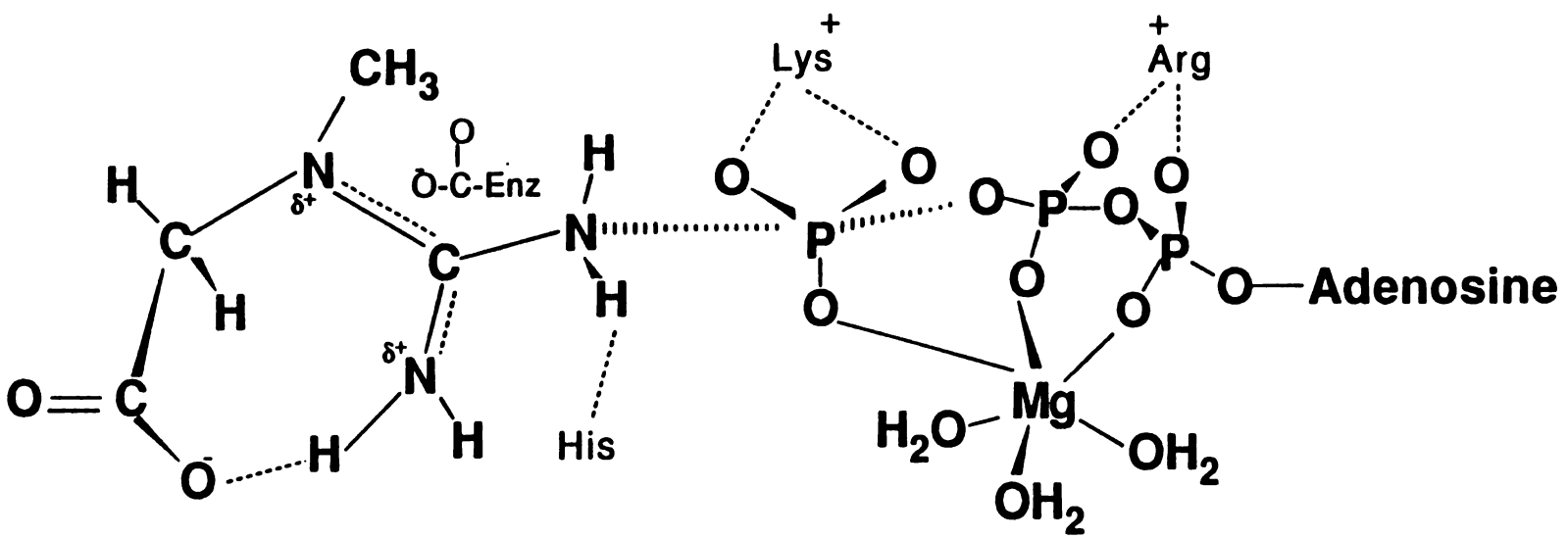


In the higher phyla in which it occurs (12), CK is normally a cytoplasmic enzyme although exceptions may occur for some membrane associating forms (15, 16). The enzyme has no intramolecular disulfide bonds (17). It functions physiologically as a dimer and the monomer exhibits only about 20% of normal activity (18). There is no evidence for allosteric regulation.

The kinetic mechanism at pH 8.0 is of the rapid equilibrium, random, bimolecular, bimolecular type wherein phosphoryl transfer is the rate-limiting step (19). The enzyme possesses two substrate binding sites and has an absolute requirement for a divalent metal cation complexed to the nucleotide substrate and coordinated to three waters in the active site (20).

As reviewed by Kenyon and Reed (11), a number of mechanistic and conformation studies have identified amino acid residues which are likely to be involved in substrate binding or conformational interactions necessary to catalysis. A histidine has been implicated as the general acid-base catalyst. A carboxylic acid has been postulated to anchor the guanidinium group and withdraw electrons from its tertiary nitrogen. The transferring phosphoryl moiety may interact with lysines or arginines conferring a partial positive charge on the phosphorus atom, thus rendering it susceptible to nucleophilic attack. The conformation of the substrates bound in the active site have also been determined. Taken together, the results of these and other studies led Kenyon and Reed to propose a possible model for the CK active site as shown in Figure 2.

Figure 2: Model for the Possible Active Site Structure of Creatine Kinase



It has been suggested that a conformational change occurs in the enzyme after substrate binding but prior to catalysis. McLaughlin et al. first proposed that the specificity of the enzyme must rest on the ability of the substrates to shift towards the enzyme conformation with the ligand in its most favorable position for substrate activation (21). Distance calculations showed that this shift brought substrates closer by about 3 angstroms and resulted in the loss of waters from metal coordination in the active site (22).

A candidate for involvement in this conformational change is an iodoacetate-sensitive cysteine residue (23) which, when modified by a number of reagents, leads to loss of enzyme activity. This loss of activity ranges from about 80% to 100% depending on the blocking agent used. Nuclear magnetic resonance studies have shown this residue to be about 10 angstroms away from bound substrates (24). Other studies have shown that the nucleotide substrate could bind to the enzyme even after sulfhydryl inactivation by iodoacetic acid and the loss of enzyme activity (24). Further discussion of the role of this cysteine residue occurs later in this Introduction.

In spite of the wealth of mechanistic information about CK that has accumulated since its first isolation and characterization, until recently, there has been little structural information available. No primary structure for any CK had been published at the beginning of this thesis work. Obviously, the starting point for asking structure/function questions about this enzyme required that we have at least a primary structure available. It was at this point that I entered the CK project and as part of the work of this thesis, we have sequenced a full-length cDNA clone from the electric organ of Torpedo californica (26) and compared primary sequences for nine other recently-sequenced creatine kinases

(27). The sequencing work was a project primarily done by Dr. Brian West.

Using the information obtained from these experiments and from the wealth of physical and enzymological data available for this protein, we asked the question, "How can we further probe structure/function relationships in creatine kinase?" Because no crystal structure is yet available, our ability to infer specific models for such relationships is severely limited. We decided, therefore, to focus on one of the simplest, most discrete questions we could ask.

This question concerns the role of one cysteine residue, Cys_{283}. This residue is identical in all nine CK sequences compared in our study, as are 12 of the 13 residues immediately surrounding it (27). Early experiments on rabbit muscle CK showed that this residue is uniquely modified by a variety of reagents, two of which are iodoacetamide and methyl methanethiolsulfonate (MMTS). Modification of CK by iodoacetamide results in the loss of detectable enzyme activity and Cys_{283} was thus concluded to be "essential" to CK catalysis (28). In 1974, Smith and Kenyon labeled this residue with MMTS and reported about 20% enzyme activity (29). They claimed that this cysteine was not therefore essential and suggested that the addition of larger modifying reagents such as iodoacetamide produced an artifactual loss of activity due to steric hindrance, possibly in connection with the conformational change that occurs after substrate binding. Fawcett et al. countered with the argument that the CH_3S- group delivered by MMTS exchanges between Cys_{283} and another free thiol and suggested that this is the basis for the residual activity in the modified CK (30). Their conclusion was that the essentiality of active site thiols in creatine kinases is still open to question.

Having obtained the primary sequence for the CK from Torpedo californica electric organ, we have used the techniques of molecular biology to try to determine conclusively either the essentiality or non-essentiality of this cysteine residue. We have changed Cys_{283} to alanine using site-directed mutagenesis. Also, the alanine mutation was chosen because such a change would preclude any cysteine-like chemistry at position 283. The alanine mutation does not introduce steric bulk to interfere with the postulated movements which may occur around position 283 following substrate binding (22). The activity of the mutant enzyme was assessed as the primary test for Cys_{283} essentiality. Since the alanine mutation is not vulnerable to the criticism of Fawcett et al., we consider this approach to be superior to earlier chemical modification experiments to study this particular question.

While our experimental design is conceptually simple, its execution was difficult. The primary prerequisite for this study was the ability to express our CK wild type and mutant clones in a system which would provide us with sufficient amounts of active, pure protein. The system we chose to use was E. coli. Our success with that system has been only marginal in fulfilling our primary prerequisite and the bulk of the work in this thesis went toward achieving suitable expression rather than to the study of the mutant CK. Nevertheless, we have been able to come to some conclusions in answer to our major question. Our results show that Cys_{283} is not essential to CK activity since the alanine mutant showed 20% of wild type activity.

This thesis has been divided into four chapters. The first two represent work that has already been published. Chapter 1, "Sequence Determination of Creatine Kinase

from Torpedo californica" was published as "Creatine Kinase Protein Sequence Encoded by a cDNA made from Torpedo californica electric organ mRNA" (26). While it has been reproduced here in its entirety as Chapter 1, most of the work it represents was not performed by this author. Dr. Brian West had already obtained the clone and had completed a large portion of the sequencing when I joined the project. It is included in this thesis because I was actively involved in the sequencing and homology searches and because it represents a real portion of my thesis work.

The second chapter, "Comparison of Creatine Kinase Primary Structures" was published under that title (27). It is also included here in its entirety. While the major part of this work was mine, other scientists made significant and original contributions, particularly Drs. Irwin D. Kuntz and Fred Cohen in the secondary structure prediction work. Drs. George Kenyon and Brian West also made major conceptual contributions. Others of the authors provided us, prior to their publication, with CK primary structures of various species. Some portions of the work represented by Chapters 3 and 4 were executed by others, primarily Dr. Brian West.

The construction of the expression vectors for wild-type and mutant CKs is discussed in Chapter 3 as are the manipulations of the bacterial hosts to produce expressed protein. Processing the expressed proteins is discussed in Chapter 4. Because the purification and activation of expressed protein was very difficult to execute successfully, it has been accorded a separate chapter so that the problems encountered could be more fully discussed. Chapter 4 also includes the comparisons of wild type and mutant CKs expressed in E. coli.

Chapter 1: Sequence Determination of Creatine Kinase from Torpedo californica

Abstract

Creatine kinase (E.C. 2.7.3.2, adenosine-5'-triphosphate: creatine phosphotransferase) ((CK) is important in the maintenance of ATP levels in high energy-requiring tissues such as muscle and brain. A complete understanding of its function requires knowledge of its amino acid sequence. In order to obtain cDNA clones encoding creatine kinase sequences, a cDNA bank was contructed using mRNA from the electric organ of Torpedo californica and screened by comparing differential colony hybridization of electric organ and liver-derived ^{32}P-cDNA's. Cloned DNA's have been isolated which can arrest the abundant synthesis of 40,000 to 43,000 molecular weight material seen after in vitro translation of electric organ mRNA. One of the clones, CK52g8, was sequenced by the dideoxy M13 method, and found to encode a 42,941 molecular weight protein which is 68% homologous to a known partial sequence of rabbit muscle creatine kinase and which has a composition similar to creatine kinases from chick and rabbit tissues. By contrast, no significant homology was found with the known sequences of kinases which use other substrates. Northern blot hybridization analysis indicated that CK52g8 is complementary to a 1600 base pair mRNA. Primer extension analysis indicated that CK52g8 is only 5 nucleotides short of a full-length cDNA, implying that it encodes a complete protein sequence. The availability of this complete sequence should be useful in further studies of creatine kinase structure and function using techniques such as site-specific mutagenesis.

Introduction

Creatine kinase (CK, EC 2.7.3.2, adenosine-5'-triphosphate: creatine phosphotransferase), helps maintain necessary ATP levels in high energy-requiring tissues, such as muscle and brain, by catalyzing the reversible transfer of a phosphoryl group between ATP and creatine. Several features of the enzyme have been observed which require a more complete explanation in molecular terms. These include a specific localization of a portion of CK with the cell (31), conformational changes within the enzyme (32-34), and, of course, the catalytic activity. Concerning the catalytic activity, functional roles have been suggested for certain amino acids in or near the active site: a lysine and an arginine are thought to interact with the transferring phosphate group (35-37), a tryptophan is thought to interact with the adenine group (38), a histidine is presumed to act as the acid-base catalyst (39), and a cysteine can be chemically modified to cause changes in the enzyme activity (11, 29). In order to create molecular models for how these amino acids contribute toward catalysis, conformational changes, and cellular localization of the enzyme, it is imperative to know the structure of CK and to know its primary amino acid sequence. Although amino acid compostion data have been obtained for CK from several species (12, 40), and a partial amino acid sequence has been reported for the rabbit muscle enzyme (41), no complete amino acid sequence has yet been reported.

In earlier studies of the in vitro translation of mRNA from the electric organ of Torpedo californica (42, 43), the presence of heavy ^{35}S-methionine incorporation in the region of 40,000 to 43,000 molecular weight was observed. Since a single subunit of CK

has this molecular weight and since the electric organ is rich in Na^+/K^+-ATPase (44) and presumably requires high energy reservoirs to regenerate its ionic balance after discharge, it was reasoned that this tissue might contain a high abundance mRNA coding for CK. In this report we describe the cloning and sequencing of a cDNA to electric organ mRNA which does encode a protein sequence highly homologous to a previously known partial sequence of rabbit muscle CK. Other groups have reported the isolation of CK cDNA clones from the chick muscle, but no sequence for these were given (45, 46).

Materials and Methods

1. RNA Preparation and cDNA Library Construction

To obtain RNA, frozen electric organ tissue (Biomarine Laboratories, Venice, CA) was pulverized and homogenized in guanidine thiocyanate, followed by lithium chloride precipitation as was described by Cathala et al. (43). Poly(A) RNA was prepared by oligo(dT) cellulose chromatography (47).

To synthesize full-length double-stranded (ds) cDNA, each first strand of the cDNA was tailed with approximately 15 dCMP residues using terminal deoxynucleotide transferase (PL Biochemicals, Milwaukee, WI) and synthesis of the second strand was primed with oligo$(dG)_{10}$ (Collaborative Research, Lexington, MA), as described by Cooke et al. (48) and Land et al. (49). The tailing was done after the first strand was heated at 70 °C for 20 min. in 0.1 N NaOH to remove the RNA. The ds cDNA (12ug) was then treated with 2000 units of S1 nuclease (Miles, Elkhart, IN) in 100 ul of 300 mM NaCl, 30 mM NaOAc, pH 4.5, 3 mM $ZnSO_4$ at 37 °C for 1 hr. The ds cDNA was then fractionated on a 10% polyacrylamide gel (50), selecting only the material longer than 500 base pairs (determined by DNA length markers), for recovery by electroelution. Approximately 10 ng of this material were tailed with dCMP and cloned by annealing to Pst1-cut, dGMP-tailed pBR322. A total of 2600 ampicillin-sensitive, tetracycline-resistant clones were obtained after transforming E. coli strain RR1 with this DNA and plating onto LB agar plates containing 5 ug/ml tetracycline.

2. Hybrid-arrested translation

For hybrid-arrested translation analysis, 0.4 ug of poly(A) RNA and 2.5 ug of HindIII-cut plasmid DNA were hybridized using the conditions of Paterson et al. (51). In vitro translation of the RNA was performed using reticulocyte lysates (52) in the presence of 100 uCi L-[^{35}S]- methionine (>600 Ci/mM, Amersham, Arlington Heights, IL) in 30ul. Protein samples were fractionated by sodium dodecyl sulfate ($NaDodSO_4$) polyacrylamide gel electrophoresis according to Laemmli (53). Two cycles of CsCl gradient purification of the DNA were necessary to avoid nonspecific inhibition of the translation, as previously reported by Kronenberg et al. (54).

3. DNA Sequencing

DNA sequencing was done by the dideoxy method (55), after subcloning overlapping fragments into the M13 phage vector, MP10 (56), using E. coli strain JM101 as host. Dideoxynucleotides were from PL Biochemicals, Milwaukee, WI, Klenow fragment of DNA Polymerase I from Boehringer Mannheim, Indianapolis, IN, and alpha-^{32}P dCTP, 400 Ci/mM from Amersham, Arlington Heights, IL.

Results

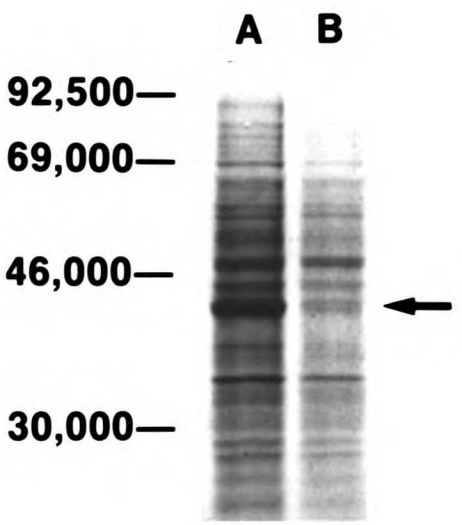
The first clone used for hybrid-arrested translation analysis, CKABh9, was chosen because it gave a very strong signal when probed by colony hybridization (57) with ^{32}P-cDNA to poly(A) RNA from the electric organ, but little or no signal with a probe made from Torpedo liver poly(A) RNA (not shown). The CKABh9 clone specifically arrested the in vitro synthesis of the major proteins(s) at 40,000 to 43,000 molecular weight (Figure 3), and hence was a candidate for a creatine kinase-encoding cDNA.

We ^{32}P-labeled the 500 base pair insert of CKABh9 by nick- translation (58) and screened the 2600 clones by colony hybridization. Forty-eight clones were positive by this assay, and forty-one of these were chosen for further analysis after purification of the DNA by the method of Holmes and Quigley (59). Seven of the largest clones appeared to be very close in size, having inserts approximately 1500 base pairs in length. The restriction map of one of these clones, CK52g8 is shown in Figure 4.

Six other positive clones were close in size to CK52g8, and all had identical restriction maps, except that one or both of the vector Pst1 sites were not reconstructed in some. The CKABh9 insert was found to have the same Bgl-II, ClaI, and Pst1 sites present in the 3' portion of CK52g8. CK52g8 DNA also arrested the in vitro synthesis of the 40,000 to 43,000 molecular weight protein(s) (not shown).

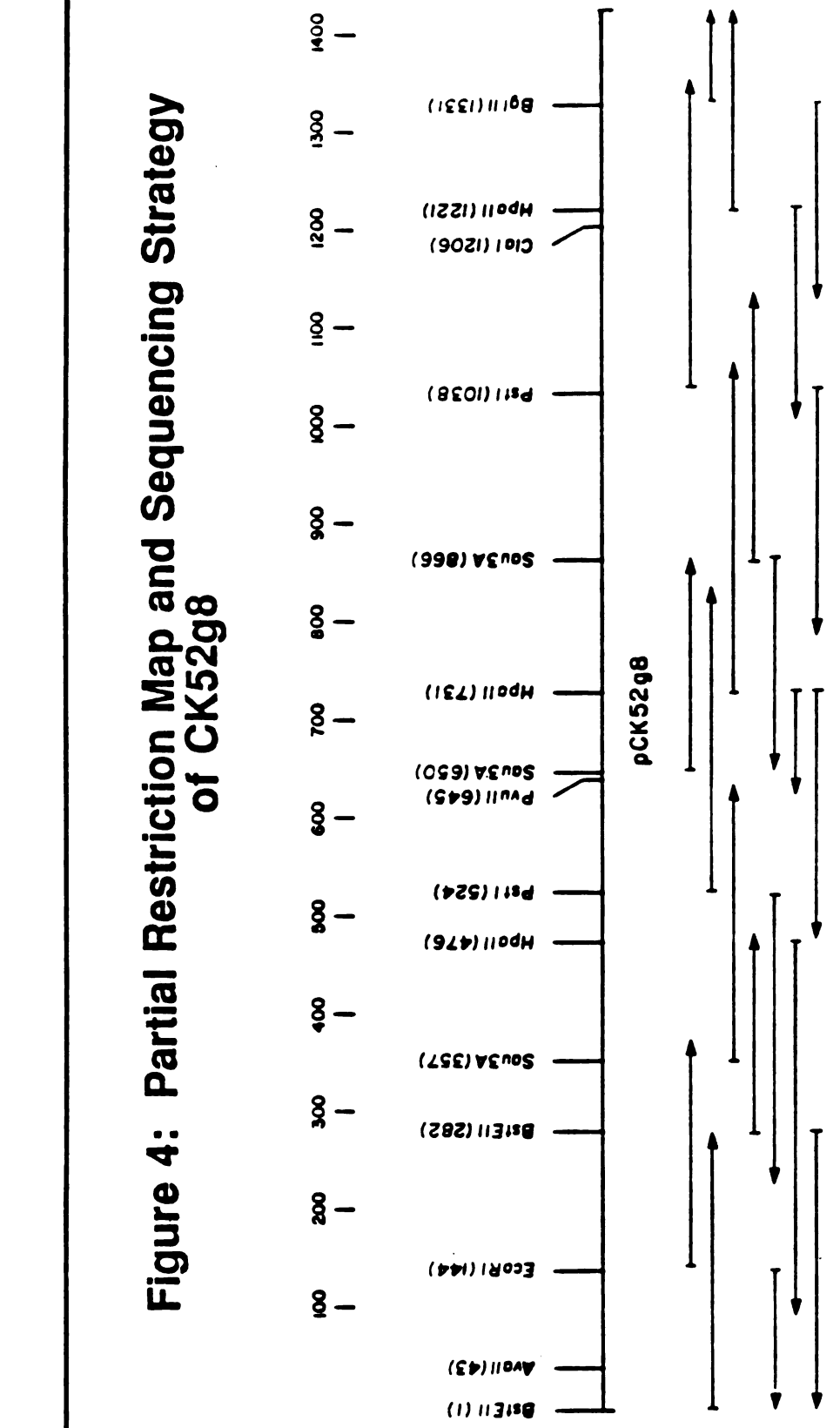
The DNA sequence obtained for the CK52g8 insert is shown in Figure 5. There are 1428 bases in this insert, followed by a stretch of poly(A) and bounded by the GC tails. The sequencing strategy is shown in Figure 4. All portions of the sequence were derived from at least two phage templates, and 92% was sequenced on both strands.

Figure 3: Hybrid-arrested <u>in vitro</u> Translation of <u>Torpedo</u> Electric Organ mRNA



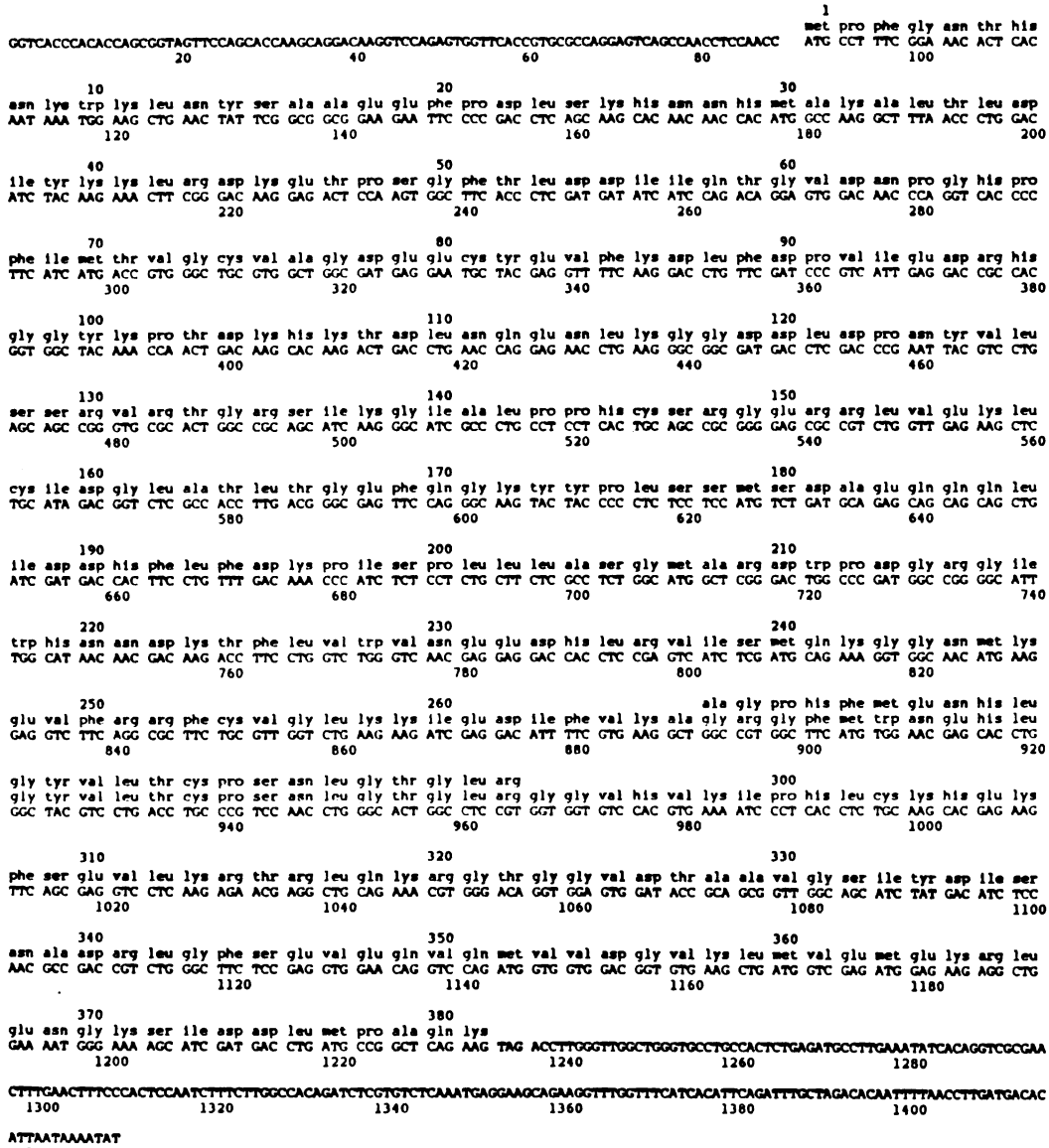
The arrow indicates the position of 40,000 to 43,000 molecular weight protein(s). After hybridization of mRNA to CKABh9 DNA, ^{35}S-methionine incorporation into the 40,000 to 43,000 molecular weight protein(s) is blocked (lane b) whereas incorporation is strong after control hybridization without CKABh9 DNA (lane a). Five μL of each translation mixture was loaded on the gel (2100 cpm lane a, 1600 cpm lane b). Molecular weights indicated on the left were taken from ^{14}C-methylated markers (phosphorylase b, bovine serum albumin, ovalbumin, and carbonic anhydrase) run on the same 12.5% polyacrylamide gel. Autoradiography was done for 21 days.

Figure 4: Partial Restriction Map and Sequencing Strategy of CK52g8



The position from the 5' end of each restriction endonuclease site listed is given in parentheses. Each arrow represents at least one dideoxy sequencing reaction using separate DNA fragments of CK52g8 subcloned into bacteriophage M13. The length of each arrow indicates the length of sequence determination for each reaction.

Figure 5: DNA Sequence of CK52g8 and Inferred Amino Acid Sequence



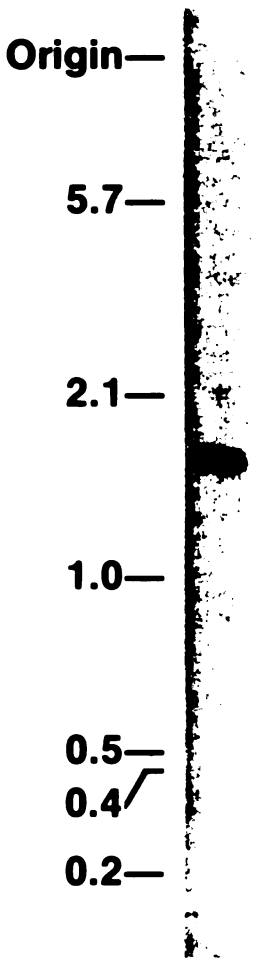
The nucleotide numbering, beginning from the 5' end, appears in the right margin next to the DNA sequence. The inferred amino acid sequence with its own numbering appears above the nucleotide sequence. To localize the homology between the predicted protein sequence and the CK sequence from rabbit muscle, the 25 amino acid sequence reported by Atherton et al. (43) for rabbit muscle CK is shown above the Torpedo sequence between amino acid numbers 268 and 292. Nucleotides 1-89 and 1233-1428 represent inferred 5' and 3' noncoding regions.

Translation of the sequence gives an open reading frame after the first ATG at nucleotides 90-92 that extends for 381 amino acids to a stop codon at position 1233-1235. The other two reading frames contain either 17 or 14 stop codons distributed over the same span. In Figure 5, the amino acids at positions 268-292 are compared with the amino acid sequence reported by Atherton (41) for rabbit muscle CK. There is 68% homology (with no gaps introduced) for the entire sequence, and 100% homology for a 17-amino acid stretch following the Glu at amino acid 275. The inferred carboxyl terminal dipeptide, Gln-Lys, is identical to the carboxyl terminal dipeptide for rabbit muscle CK (60). The molecular weight for this inferred protein is 42,941. C, G, U, and A are used in 47%, 29%, 17%, and 7% respectively, of the third positions of degenerate codons. The presumed polyadenylation signal sequence AATAAA (61) appears at nucleotide position 1419. The precise start of polyadenylation after this signal was not determined due to interference by the GC and AT stretches in the sequencing reactions.

Northern blot hybridization analysis of electric organ poly(A) RNA, using ^{32}P-labeled CK52g8 insert DNA as the probe, indicates only one band migrating in the 1600 base pair position of glyoxylated DNA markers (Figure 6). In another northern blot experiment, using ^{32}P-labeled CKABh9 insert DNA as the probe, a single band of the same size was obtained (not shown).

To determine whether most of the 5'-noncoding sequence was present in CK52g8, the 105 base pair AvaII-EcoRI fragment was used as a primer for ^{32}P-cDNA synthesis, with electric organ mRNA serving as the template. The longest ^{32}P-cDNA

Figure 6: Northern Blot Hybridization Analysis of Electric Organ mRNA



Five μg of electric organ poly(A) RNA was prepared for hybridization according to Thomas (67). Hybridization was done for 15 hrs at 42 °C in a 13 ml volume containing: 5X Denhardt's solution (68), 0.75 M NaCl, 0.075 M Na_3citrate, 20 mM $NaPO_4$, ph 6.5, 200 μg/ml sheared, boiled calf thymus DNA, and 2 x 10^6 cpm boiled ^{32}P-labeled EcoR1/BgIII fragment DNA. The blot was then washed in 30 mM NaCl, 3 mM Na_3 citrate, pH 7, 0.1% $NaDodSO_4$ at 42 °C and autoradiographed. Length markers on the left are in units of base pairs x 10^3 and are derived from 9000 cpm ^{32}P-labeled HindIII-cut phage PM2 DNA (69), which was glyoxalated (70) and electrophoresed in another lane of the same gel.

synthesized extended only 5 nucleotides beyond the 5' end of CK52g8 (not shown). Thus it is probable that CK52g8 contains a nearly complete cDNA insert. The difference between the mRNA size (1600 base pairs and the cDNA size (1428 base pairs) is likely due to additional poly(A) in the mRNA that is not complete in the cDNA (62).

No significant homology with the CK52g8 encoded protein could be found with any of the 2,222 proteins listed for the Protein Data Bank of the National Biomedical Research Foundation, except for the partial sequence to rabbit muscle CK, using the programs of Martinez (63). Kinases listed in the library included Bacillus stearothermophilus phosphofructokinase, herpes simplex thymidine kinase, Escherichia coli aspartokinase I /homoserine dehydrogenase I, Escherichia coli homoserine kinase, Rous sarcoma transforming protein, bovine cAMP-dependent protein kinase, pig and human adenylate kinases, and horse and human phosphoglycerate kinases. Saccharomyces cerevisiae pyruvate kinase (64), not listed in the library, was also examined and found to have no significant sequence homology.

Discussion

The northern blot hybridization results and the primer extension results indicate that CK52g8 is a nearly complete cDNA. This would imply that the protein sequence encoded by CK52g8 is also complete. Although it cannot yet be determined whether or not this inferred protein has CK activity, its amino acid sequence is 68% homologous to the reported partial sequence of rabbit muscle CK (41) and its composition (Table 1) is very close to the compositions of rabbit and chick CK's reported by Eppenberger, et al. (40).

The similarities in composition of these sequences are reflected in low difference index values (65) of 9.1, 7.6, 9.7, and 8.6 for comparisons of the CK52g8-inferred protein composition with the compositions of chick brain, chick muscle, rabbit brain, and rabbit muscle, respectively. Thus, the CK52g8-inferred protein is certainly related to CK proteins.

The inferred amino terminus of CK52g8 occupies a position consistent with the observation that most eukaryotic mRNAs studied begin translation at the first 5' AUG (66). The nucleotide sequence in this region contains elements (A at 87, C at 85, 88, and 89) which have been shown (71, 72) to be part of a consensus sequence for an initiating AUG. Because the amino terminal protein sequence of Torpedo electric organ CK is not known, it is not possible to know whether translation does begin at this Met nor is it possible to know whether post-translational processing alters the amino terminal structure. Useful information concerning these questions might be derived from the microsequence analysis (73) of the first 40 amino terminal residues of rat brain CK, which indicate the

Table 1: Comparison of Amino Acid Composition with Chick and Rabbit CKs

Amino acid	Torpedo electric Organ (CK52g8)	Chick[a] Brain	Chick[a] Muscle	Rabbit[a] Brain	Rabbit[a] Muscle
Lys	31	24	32	26	30
His	14	10	17	15	16
Arg	20	20	20	11	16
Asp	31	44	36	41	42
Thr	17	18	14	17	17
Ser	19	18	16	16	22
Glu	24	38	42	37	39
Pro	17	16	20	20	18
Gly	36	34	30	32	32
Ala	15	18	16	18	13
Val	26	24	24	24	24
Met	12	10	10	10	8
Ile	18	14	12	15	11
Leu	36	41	36	40	36
Tyr	9	10	8	10	10
Phe	16	15	16	20	16
Cys	7	n.d.[b]	n.d.	n.d.	n.d.
Trp	5	n.d.	n.d.	n.d.	n.d.
Asn	17	n.d.	n.d.	n.d.	n.d.
Gln	11	n.d.	n.d.	n.d.	n.d.

[a]Composition of chick and rabbit creatine kinases are calculated from Eppenberger, et al. (40) assuming that dimer subunits are identical. Units for all compositions are moles/mole of subunit.

[b]n.d.: not determined.

presence of Leu residues at positions 11 and 22 with no Met. If the first Met encoded by CK52g8 is removed post-translationally, then Leu occupies positions 11 and 22 of the CK52g8-encoded protein as well. Since the carboxyl terminal dipeptide, Gln-Lys, predicted for CK52g8 is identical to that for rabbit muscle CK (60), it is unlikely that the carboxyl terminus is post-translationally modified.

The inferred protein molecular weight for CK52g8 (42,941) is close to that of CK purified from the electric organ of Torpedo marmorata (74) and to the molecular weight of the in vitro translation products which can be arrested by hybridization of CKABh9 and CK52g8 DNAs. Barrantes et al. (74) report immunological evidence for the presence in the electric organ of two isozymes, referred to as the "brain" and "muscle" forms. It is clear that all of the cDNAs isolated in our experiments are complementary to a single mRNA species since all of these clones had inserts with the same restriction sites as CK52g8 and since a single band appears in the northern blot hybridizations with both CK52g8 and CKABh9 insert probes. The "brain" form of CK shown by Barrantes et al. (74) to be associated with the acetylcholine receptor-enriched membranes of the electric organ has a pI in the range of 6.0 to 6.5, but a pI value of 7.5* can be predicted for the CK52g8 encoded protein by calculating the weighted average of pK values for the 31 Asp, 24 Glu, 31 Lys, and 20 Arg (pK = 3.86, 4.25, 10.53, and 12.48, respectively) residues.

* The method for calculation of pI values described above is incorrect as published. My colleague, Vincent Powers, has recalculated the pI for this CK as 6.31. The computer program he wrote for this purpose determines the pI by calculating the net charge of a protein of given amino acid composition starting with an initial guess pH of 1.8 (more acidic than all possible pI's). The guess pH is iteratively increased by .01 increments until the net charge becomes less than zero for the first time. The pH where this occurs is rounded to the nearest tenth, and is taken as the pI. All ionizable groups are assumed to contribute and pK's are assumed to be identical for each side chain type regardless of environment.

Identification of isozyme type possibly must await expression of the CK52g8 encoded protein and an experimental determination of its pI, as well as further screening of the cDNA bank for other CK encoding clones.

The partial fragment from rabbit muscle CK chosen by Atherton et al. (41) for protein sequencing contains a Cys which selectively reacts with sulfhydryl-modifying reagents to affect enzyme activity. This Cys lies in a 17-amino acid stretch that has 100% homology (Figure 5) to the sequence derived here. Since this 17-amino acid stretch is so highly conserved, even in an evolutionarily divergent specie such as Torpedo (class Chondrichthyes), it probably is important for some aspect of structure or function. The Pro-Asp-Ser sequence after the cysteine strongly resembles sequences in other proteins where beta-turns occur (75), but knowing whether there is a beta-turn at this position may require elucidation of the crystal structure of CK (76-79).

Assuming that CK activity can be confirmed for the protein encoded by CK52g8, the availability of this full-length sequence should greatly facilitate further studies on the enzyme's structure and function. By deriving the protein sequence of some mammalian CK and comparing its sequence with that obtained here for Torpedo, useful insights into imporant conserved regions should be obtained. By expressing the CK52g8 sequence in a heterologous cell, studies of the enzyme's function should be possible using techniques such as site-specific mutagenesis.

Chapter 2: Comparisons of Creatine Kinase Primary Structures

Abstract

Comparisons on nine creatine kinase sequences show that 67% of the protein sequence is identical among rabbit, rat, mouse, and chicken muscle, rabbit, rat and chicken brain, and electric organ sequences from two species of the electric ray (Torpedo). The extensive homology precludes a facile prediction of active-site residues based on sequence conservation. The sequences are more similar within isozyme types than are the different isozymes from any one species. There are 35 positions in the muscle and brain sequence pairs for three species which differentiate the two forms. The Torpedo sequences do not fall completely into either of these patterns. Except for homology with partial sequences of other ATP-guanidino phosphotransferases, no significant homology with other protein or nucleic acid sequences in available databases was found. Preliminary secondary structural predictions suggest that the C-terminal half of the protein is likely an alpha/beta-type protein. Placement in the sequence of two peptides found in previous cross-linking studies reveals two stretches of primary structure that are presumably close in space to the reactive Cys_{283} and hence close to the active site.

Introduction

Creatine kinase (E.C. 2.7.3.2, adenosine-5' triphosphate: creatine phosphotransferase) catalyzes the reversible phosphorylation of ADP and creatine and is important in the maintenance of the intracellular energy supply. It is most abundant in muscle and brain, but it also exists in other tissues. In humans, plasma creatine kinase measurement is an important tool in the medical diagnosis of early cardiac damage (80). For several species, considerable information about CK mode of action, composition, and isozymic variety has been compiled (11, 12, 81). Until recently, however, basic structural information regarding this important protein has been unavailable. Nine complete amino acid sequences of muscle- and brain-type creatine kinases from several species have now been deduced (26, 82-90) either from direct protein sequencing or by inference from nucleotide sequences.

In this study, both the amino acid and nucleic acid identities among the sequences have been established. Homology comparisons of CK protein and nucleotide sequences with other sequences in several nucleic acid and protein databases have also been performed. In addition, regions corresponding to peptide fragments obtained from cross-linking studies have been identified. Using these correlations and the primary sequence data, we have developed preliminary predictions of the enzyme's secondary structure.

Materials and Methods

The CK sequences used in this study were from rabbit muscle (82) and brain (83), rat muscle (84) and brain (85), chicken muscle (86, 87) and brain (88), mouse muscle (89), and Torpedo californica (26) and Torpedo marmorata (90) electric organs. Sequences used for homology analyses between CK and other proteins were obtained from four databases: Pdayhoff (2784 amino acid sequences as of August, 1984) (91), Ndayhoff (1422 nucleic acid sequences as of September, 1984) (92), Newat 83 (658 amino acid sequences as of April, 1983) (93), and Genbank (5198 nucleic acid sequences as of August, 1984) (94). The nucleic acid databases contain noncoding sequences and probable coding sequences not yet assigned to known proteins, as well as sequences confirmed as coding for known proteins. The kinase library database was created from amino acid sequences for all kinases in Pdayhoff.

Sequence alignments (Figure 7) were generated using programs developed by Martinez et al. (63). The comparison of the degree of homology in these CK sequences with that in three other proteins, alpha- and beta-hemoglobins, cytochrome c, and actin (Figure 8), was done using the method of Dickerson (95).

Several programs were used to search for homology between chicken CK sequences and these databases. The dbalign program developed by Martinez et al. (63) was used on Pdayhoff, Ndayhoff, and Genbank databases, as were the National Biomedical Research Foundation (96) search, align, and relate programs. Another program, dfastp, based on an algorithm by Lipman and Pearson (97, 98), was also used to

search the protein sequence databases. In this program, scoring is based both on direct amino acid homology and on homology between "equivalent" amino acids as determined by frequency of such changes over evoutionary time scales. Newat 83 was also analyzed by dfastp and a program developed by Doolittle (99).

The kinase library, along with the highest scoring sequences from the database searches, was analyzed with another alignment program developed by Martinez et al. The malign program (100) is based on the same algorithm as dbalign but has greater flexibility. Based on the lengths of the two sequences being compared, malign computes the expected longest repeat occurring "by chance" and gives the standard deviation for its occurrence. The analysis is independent of whether amino acids or nucleic acids are being compared. End gaps were not counted toward the final score and internal gaps were not weighted. Alignment scores for sequence pairs subject to randomization analysis were averaged over ten randomization runs. This average score was then compared to the scores for the real sequences to see whether possible reported homologies reflected homology due to composition and size or to linear sequence. Only a few of the highest scoring sequences were subjected to this randomization analysis.

In order to check for homology based on chemical classes of amino acids, the Torpedo CK, phosphoglycerate kinase, and yeast pyruvate kinase sequences were converted to a four-letter reduced alphabet in which each amino acid residue was replaced with a designation corresponding to positively charged, negatively charged, polar but not charged, and hydrophobic (63). Malign was used to compare CK and each of these sequences, as well as to compare muscle and brain CK sequences.

Results and Discussion

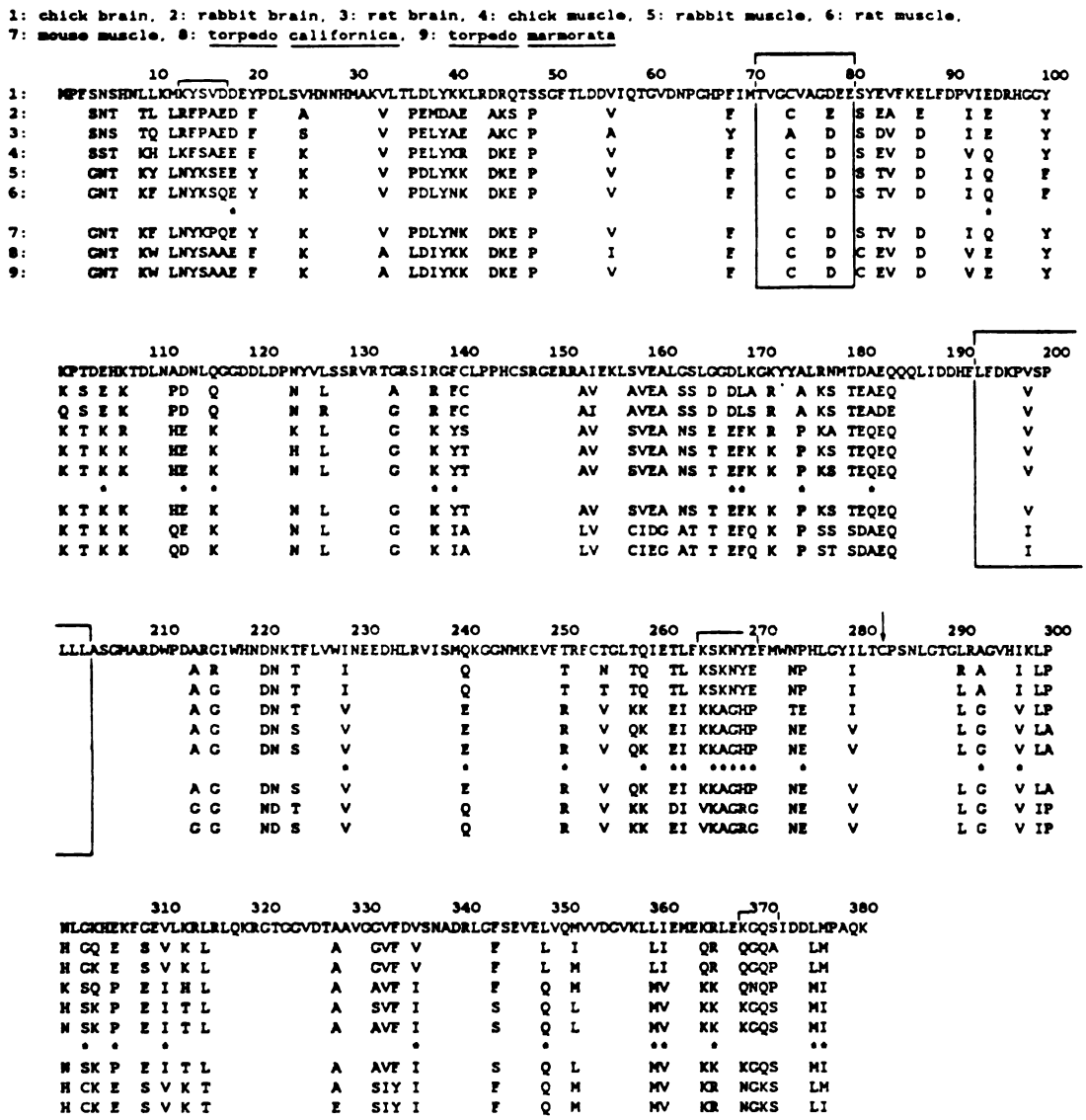
1. Homology among Creatine Kinases

As shown in Figure 7, alignments among nine CK protein sequences from six species reveal a high degree of similarity in their primary structures. Sixty-seven percent of the amino acids are identical among all nine sequences and neither gaps, insertions, nor deletions occur in any of the sequences. The sequence identity among the three mammalian muscle CK (M-CK) sequences, rabbit, rat, and mouse, is 98%; and between the two mammalian brain CK (B-CK) sequences, rabbit and rat, the identity is 93%. The identity between the two Torpedo sequences is also 98%.

The sequence similarity is greater among all muscle or brain isozymes from different species than between the two different isozymes in any one species. Exact sequence identity among brain-type CKs from rabbit, rat and chicken is 88%. Among muscle types the analogous value is 90%. Sequence identity between an unpublished canine myocardial muscle-type CK and the M-CK sequences shown in Figure 7 is reported to be similarly high (101). By contrast, the homology between muscle and brain sequences for any one of the species of chicken, rat or rabbit ranges from 78-80%.

Nucleic acid homology in the protein-coding portions among the nine CK sequences is 55%. As with the protein sequences, neither gaps, insertions, nor deletions occur. When the protein homology was compared to the nucleic acid homology, of the 515 mismatched base positions at which at least one of nine species is nonidentical with the others, 40% were found to be silent changes.

Figure 7: Alignment of Nine CK Sequences

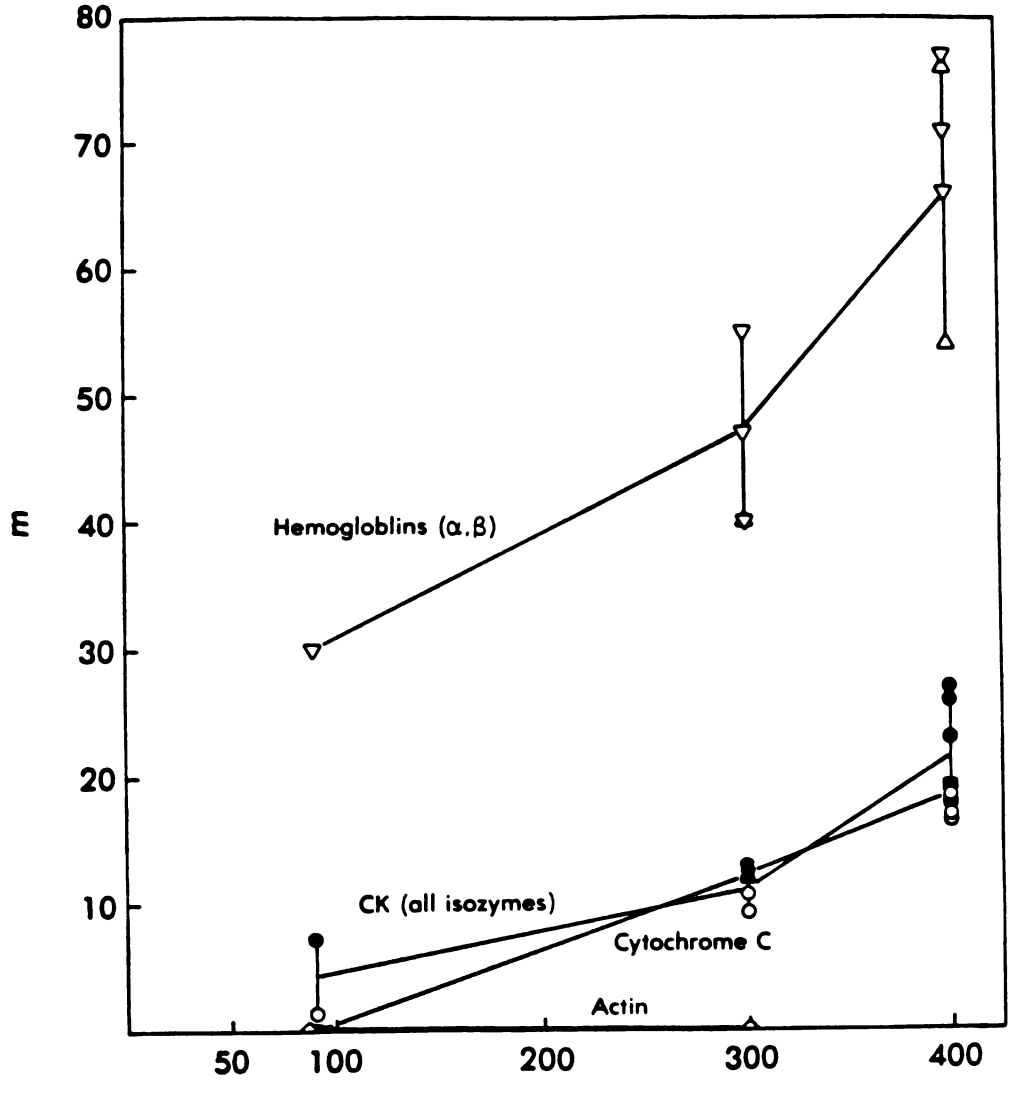


(1) Chicken brain, (2) rabbit brain, (3) rat brain, (4) chicken muscle, (5) rabbit muscle, (6) rat muscle, (7) mouse muscle, (8) Torpedo californica electric organ, (9) Torpedo marmorata electric organ.

Residues with exact sequence identity among all nine sequences are represented by line 1 only. For every position in which one or more residues do not match in all sequences, all nine residues are printed. Starred (*) positions indicate residues that show consistent differences between muscle and brain isozymes for chicken, rabbit, and rat sequences and the mouse muscle sequence. Boxed regions (residues 71-80 and 193-203) represent the peptides in the rabbit muscle enzyme found to cross-link with Cys_{283} (marked with an arrow) via 1,5-difluoro-2,4-dinitrobenzene (102). The three segments with bars above them represent the regions in which four or more consecutive residues are nonidentical in one or more sequences when the nine sequences are converted to the reduced alphabet and aligned. The reduced alphabet alignment is not shown.

Protein divergence among these CKs increases in an approximately linear fashion as evolutionary distance increases (Figure 8). We examined divergence in three other proteins for a range of species similar to that represented by the nine CK sequences. Sequence identity for human, rabbit, rat, and chicken sequences of skeletal and smooth muscle actin is 100%. When sea urchin actin is added to the analysis, the identity among all five species is 93%. Homology among cytochrome c sequences for rabbit, rat, chicken, and lamprey eel is 80%, whereas identity for beta-hemoglobin sequences of rabbit, rat chicken, and carp is only 42%. CK sequences are identical at 79% of the positions in muscle enzymes from rabbit, rat, mouse, chicken, and the electric organ enzymes from Torpedo. The identity between rabbit, rat and chicken brain sequences and the Torpedo sequences is 72%. These proteins are compared in Figure 8.

Figure 8: Sequence Conservation in CK Compared to Three Other Proteins



Here m is the number of changes per 100 residues (corrected for multiple hits at the same locus) calculated as m = 100 { -ln [1 - (n/100)]}, where n is the number of nonidentical residues per 100 residues in the alignment of two present-day sequences.

The nine CK sequences show sequence conservation similar to that for cytochrome c.

Alignments were generated using the malign program. No gaps or insertions occurred except for hemoglobins. The values for "millions of years since divergence" (MY) for each species pair were taken from Dickerson (95). These values are only approximate. The Torpedo californica sequence was used for both the brain and muscle CK data points. The lines are only a visual aid. (○) Muscle CK. Aligned sequence pairs for Torpedo californica/chicken, T. californica/rat (mouse), and T. californica/rabbit (400 MY); rabbit/chicken and rat/chicken (300 MY); and rabbit/rat (90 MY). (●) Brain CK pairs for all of the sequences listed above. (∧) Skeletal muscle actin. Aligned sequence pairs for rabbit/chicken and rat/chicken (300 MY); and rabbit/rat (90 MY). (■) Cytochrome c. Sequence pairs were chicken/lamprey and rabbit (rat)/lamprey (400 MY); rabbit (rat)/chicken (300 MY); and rabbit/rat (90 MY). (▽) alpha-hemoglobin. (The carp hemoglobin sequences had one to three one-residue gaps in alignment with other sequences and in each case the carp sequence had one more residue than the sequence to which it was compared. The number of amino acids in the shorter sequence was used in calculations.) Sequence pairs were rabbit/carp, A-chicken/carp and D-chicken/carp (400 MY); rabbit/A-chicken, rabbit/D-chicken, and rat/A-chicken (300 MY); and rabbit/rat (90 MY). (△)beta-Hemoglcbin. Sequence pairs were chicken/carp and rabbit/carp (400 MY); and rabbit/chicken (300 MY).

It appears that CK is among the more highly conserved of known proteins (95). From consideration of the high degree of homology among the mammalian CK protein sequences, one expects that the human enzymes will have similar sequences to rabbit, rat

and mouse CKs. This prediction has been verified by preliminary partial sequence data on human M-CK (103).

Table 2 shows that amino acid compositions are also similar among the nine sequences. As would be expected from the evolutionary distances represented by these species, the Torpedo compositions show the greatest differences from the other species listed. For example, both Torpedo CKs have seven cysteine residues, compared to four in the muscle sequences and five in the brain sequences.

Since cysteine and tryptophan residues have been implicated as being spatially near the active site in CK (11), it is interesting that cysteine residues are conserved at three positions among all nine sequences. All four tryptophan residues found in the mammalian and avian species are positionally conserved in all nine sequences. Either an aspartic acid or glutamic acid residue has been implicated as being in the active site. Approximately 55 of these acidic residues occur in each of the nine proteins, and 38 of these positions are conserved in all of the sequences.

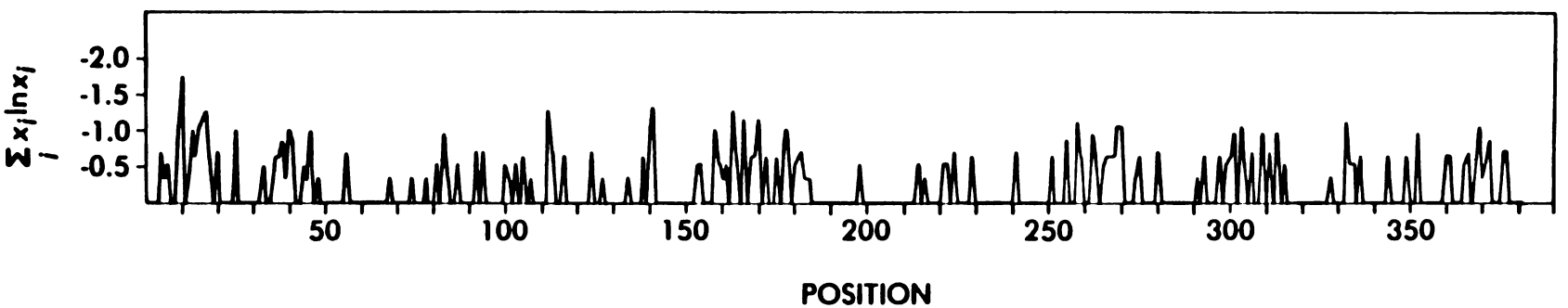
The variability among these sequences is not clustered, but occurs over the entire length (Figure 9). Further comparisons show that this variability tends to give an overestimate of the differences among the nine CK sequences, since many of these changes can be considered chemically conservative. When the sequences are converted to the reduced alphabet described above, the overall homology increases from 67% to 82% (data not shown). For this reduced alphabet there are only three stretches in the sequence in which four to six consecutive amino acids are not identical in all nine species. These stretches are marked with a bar over them in Figure 7.

Table 2: Molecular Weights and Amino Acid Compositions for Nine CK Sequences

	hick brain	Rabbit brain	Rat brain	Chick muscle	Rabbit muscle	Rat muscle	Mouse muscle	Torpedo californica E. organ	Torpedo marmorata E. organ
	42,878	42,822	42,970	43,335	43,119	43,025	43,049	42,941	42,981
Amino acid composition									
Lys	26	20	19	32	34	33	33	31	32
His	13	14	13	17	17	15	15	14	14
Arg	21	23	21	21	18	18	18	20	20
Asp	32	29	31	27	28	28	28	31	30
Thr	18	18	17	16	18	18	18	17	16
Ser	21	21	21	20	22	21	21	19	22
Glu	23	28	23	30	27	26	26	24	26
Pro	16	20	19	21	19	19	20	17	17
Gly	35	30	32	31	33	33	33	36	36
Ala	15	23	19	15	13	14	14	15	15
Val	27	14	23	28	28	28	28	26	27
Met	12	11	10	12	11	11	11	12	11
Ile	14	15	15	14	14	14	14	18	18
Leu	43	40	39	36	37	37	37	36	36
Tyr	10	7	9	8	10	9	10	9	9
Phe	16	18	14	18	16	17	16	16	16
Cys	5	5	5	4	4	4	4	7	7
Trp	4	4	4	4	4	4	4	5	5
Asn	18	17	14	14	16	19	19	17	16
Gln	12	14	13	13	12	13	13	11	11

Molecular weights and amino acid compositions were calculated directly from the sequences and in some cases have been reported previously. Initiating Met is included in these determinations for all of the sequences, but evidence suggests that the initiating Met has been removed in the mature proteins from rabbit muscle (82), rat muscle (84), and mouse muscle (89). Such evidence is unavailable for the sequenced CK proteins from other sources.

Figure 9: Plot of Sequence Variability among Nine CKs



The x-axis represents the residue position in the sequence 1-381. The y-axis represents $\Sigma_i\ x_i\ \ln\ x_i$ where x_i is the number of occurences of a particular residue at position i divided by the number of sequences compared.

Because no regions in the sequences stand out as being significantly more conserved than other regions, we cannot, from primary structure comparisons alone, suggest an active site or other specific residues involved in substrate binding or catalysis. These comparisons may be more helpful, however, in considering the differences between isozymes. With muscle and brain isozyme sequence pairs available for three species (rabbit, rat, and chicken), patterns which differentiate the two forms appear.

2. Differences between Brain and Muscle Isozymes

Table 2 shows that there are about 20% more positively charged residues in the muscle sequences than in the brain sequences. The number of negatively charged residues is about the same for both isozymes. There are 35 positions in the muscle and brain sequence pairs that consistently differentiate between these two forms (starred in Figure 7). The mouse muscle sequence fits the muscle pattern as well. At these positions, the sequences representing each isozyme have identical residues, but the two isozymes differ. At 13 of these 35 positions there are charge differences between the isozymes. Five of these positions show size differences with either an isoleucine or a valine corresponding to each isozyme. These size differences could be compensatory in the packing of the secondary structure into the tertiary structure. At positions 266-270 a stretch of five amino acids is consistently different between the two isozymes, and four of these positions exhibit charge differences between the isozymes. In addition to the 35 positions starred in Figure 7, at ten other positions (not marked in Figure 7) all of the residues are identical for one isozyme but vary in the other isozyme.

Table 3 shows three fragments of sequence in which major differences between isozymes are clustered. All of these regions contain positions in which the differences between brain and muscle sequences are accompanied by a difference in charge.

Table 3: Three Regions in which Differences between Brain and Muscle Isozymes are Clustered

	94																								120
1		E	D	R	H	G	G	Y	K	P	T	D	E	H	K	T	D	L	N	A	D	N	L	Q	G	G	D	D
2		E						Y	K		S		E		K						P	D		Q				
3		E						Y	Q		S		E		K						P	D		Q				
4		Q						Y	K		T		K		R						H	E		K				
5		Q						F	K		T		K		K						H	E		K				
6		Q						F	K		T		K		K						H	E		K				
7		Q						Y	K		T		K		K						H	E		K				
		*											*								*			*				

	251																											276
	T	R	F	C	T	G	L	T	Q	I	E	T	L	F	K	S	K	N	Y	E	F	M	W	N	P	H		
	T			N				T	Q			T	L			S	K	N	Y	E				N	P			
	T			T				T	Q			T	L			S	K	N	Y	E				N	P			
	R			V				K	K			E	I			K	A	G	H	P				T	E			
	R			V				Q	K			E	I			K	A	G	H	P				N	E			
	R			V				Q	K			E	I			K	A	G	H	P				N	E			
	R			V				Q	K			E	I			K	A	G	H	P				N	E			
	*							*				**				**	*	**	*					*				

	292																											311
	R	A	G	V	H	I	K	L	P	N	L	G	K	H	E	K	F	G	E	V									
	A					I		L	P	H		G	Q		E		S		V									
	A					I		L	P	H		G	K		E		S		V									
	G					V		L	P	K		S	Q		P		E		I									
	G					V		L	A	H		S	K		P		E		I									
	G					V		L	A	N		S	K		P		E		I									
	G					V		L	A	N		S	K		P		E		I									
	*		*									*			*				*									

1) Chicken brain, 2) rabbit brain, 3) rat brain, 4) chicken muscle, 5) rabbit muscle, 6) rat muscle, 7) mouse muscle

These regions also have a high proportion of polar residues and are likely to contain turns in the secondary structure. We suggest that one or more of these regions may be involved in the antigenic differences between brain and muscle CKs. The primary structure differences evidently determine the documented (104, 105) differences in kinetic parameters between muscle-muscle (MM) and brain-brain (BB) dimeric CK forms.

Both Torpedo sequences are identical to the muscle pattern at 21 of the 35 positions described above. Both are identical to the brain-type CK at seven of the positions. At four positions the Torpedo sequences are unlike either muscle or brain CKs, and at the remaining three positions the two Torpedo sequences are different from each other, but identical to either muscle- or brain-type enzyme. At the 13 positions showing charge differences between isozymes, both Torpedo sequences have nine residues with like charge to the muscle form and four residues with like charge to the brain form. Thus the Torpedo sequences show more sequence identity with the muscle forms. But neither Torpedo electric organ sequence falls definitively into either the muscle or brain pattern. The fact that these two CKs have both brainlike and musclelike features in their primary sequences may help to rationalize the contradictory isozyme identifications published for Torpedo electric organ CK (13, 106-108).

Identification of CK isozymes from Torpedo has been attempted by homology analyses based on a small number of partial sequences. Perryman et al. have concluded from immunologic, biochemical, and homology analyses that Torpedo electric organ cytoplasmic CK should be classified as MM-CK (108). Their homology analysis is based on the similarity among the 20 N-terminal amino acid residues in the Torpedo CK and

muscle tissue CKs from human, pig, and chicken. Although the Torpedo CKs are considerably more like the muscle isozyme overall, analysis of 20-residue segments from Figure 7 reveals regions in which the Torpedo sequences are more similar to the brain than the muscle isozyme. Due to the high degree of conservation among known CK sequences, homology analysis of such short partial sequences is clearly inadequate to identify isozyme type. While future sequence information may reveal a definitively muscle- or brain-type CK derived from some Torpedo tissue, it is also possible that the categorizations of CK into brain or muscle isozymes may not be valid for species divergent from the mammalian or avian species from which such categorizations arose. Thus, isozyme identification based on homology analysis of even complete sequences should be approached with caution.

The reasons for the existence of tissue-specific isozymes in CKs from some species remains unclear. It is possible that these isozymes perform some different functions which are reflected in the consistent sequence differences we found. Some evidence including differential isozymic association with cellular structures, exists to justify this conclusion (13, 106, 109-111). On the other hand, Pickering et al. (83) have suggested that CK isozymes may have no major distinguishable functions and that the occurrence of different isozymes may be due to the regulation of their expression in a tissue-specific manner. Such differences in regulation are well-documented (81, 112-114). It is not possible to resolve this issue from sequence data alone. Furthermore, both regulatory and functional differences could be present simultaneously.

3. Homology with Other Proteins

Except for CK fragments, none of the sequences in the kinase library, Pdayhoff, Newat 83, Ndayhoff, or Genbank had significant homology with CK. As reported by Buskin et al. (89), CK can be aligned with a partial sequence of lobster arginine kinase (115,116), which is not included in the databases listed above. CK also appears to be related to a short fragment of earthworm lombricine kinase (117). Both of these proteins are thought to perform a similar function to that of CK in the invertebrates (12). Homology comparisons among the reduced alphabet sequences of Torpedo californica CK, phosphoglycerate kinase, and pyruvate kinase showed only 35% of the CK amino acids aligning (with eight gaps) with pyruvate kinase and 30% (with five gaps) with phosphoglycerate kinase. Given the use of the reduced alphabet, these homologies are not considered to be significant. We might expect CK to be most related to the kinases because all of them bind ATP and have a common function. Yet, except for the partial sequences mentioned above, our analysis has revealed no significant primary structural relationships between CK and the other kinases. Buskin et al. (89), however, report that in a limited region, the vertebrate muscle CKs share some sequence identity with several nucleotide-binding proteins. It is possible that significant similarities to other kinases may exist at a structural level not obvious from primary structure comparisons. Whether these similarities reflect an evolutionaly relationship or merely a common structural motif has not yet been resolved.

4. Structural Models

To explore secondary structural possibilities, we used a pattern-matching algorithm described by Cohen et al. (118, 119) and Abarbanel (120). Preliminary results suggest that the C-terminal half of the CK sequence (residues 150-381) is consistent with alpha-helix/beta-sheet secondary structure. Several kinases of known structure follow the alpha/beta motif, including pyruvate kinase and adenylate kinase (121). Since a tryptophan residue has been implicated in the binding of the nucleotide substrate (11), it is interesting to note that all four conserved tryptophan residues in the enzyme are in the carboxyl-terminal half of the protein. These models predict no important secondary structure differences between the isozymes. It is possible that sequence differences are reflected on a three-dimensional level related to charge distribution or packing of secondary structure rather than to differences in secondary structure per se.

5. Protein Chemistry

Correlations of some early cross-linking and chemical modification studies with the primary sequence are now possible. Previously, a reactive cysteine has been identified as Cys_{283} (26). In rabbit muscle CK, Mahowald (102) cross-linked this reactive sulfhydryl group with neighboring polypeptide chains using 1,5-difluoro-2,4-dinitro-benzene. After digestion, two colored peptides were obtained and their compositions reported. We have located these two peptides in the sequence. The first peptide includes the ten amino acids from residues 193-202 (boxed in Figure 7). This peptide would be linked to Cys_{283}

through the epsilon-amino group of Lys_{196}. The other peptide is linked to Cys_{283} through Cys_{74} and spans a region in the rabbit CK sequence including amino acids 71-80. This region is also boxed in Figure 7. Both peptides are highly conserved in the nine sequences. The nonidentical residues in these two regions can be considered primarily conservative in their effect on secondary and tertiary structure. Since Cys_{283} has been shown to be close to the bound substrates in the active enzymes (24), it may be inferred that these two peptides are close in space to the active site of the enzyme. Modeling studies are currently underway to check the feasibility of this suggestion.

Conclusion

It is clear that CK is a highly conserved protein and that even intraspecies muscle and brain isozymes are strikingly similar. When sequence data become available for the mitochondrial CK, it will be interesting to see whether it is equally conserved. Further experiments are required to enable us to interpret the importance of the sequence differences between brain and muscle isozymes. While we can only speculate on the nature of the constraints that have required such a high degree of sequence conservation in CK, the information provided by this analysis may be useful in guiding future approaches to answer these questions. The secondary structure predictions and the correlation with the sequence information obtained in cross-linking studies should also be helpful in the design of experiments directed at the nature of structure/ function relationships in this important enzyme.

Chapter 3: Expression of Wild Type and Mutant Creatine Kinase cDNA's from Torpedo californica in E. coli

Introduction

We chose to express the cDNA clone for CK in E. coli for several reasons. Since E. coli has no endogenous CK (12), we expected to simplify protein purification problems using this system. E. coli has traditionally been used to prepare large amounts of expressed protein and has a distinct advantage in this regard over either yeast or tissue culture systems. Because CK is normally transcribed as a soluble cytoplasmic protein with no intramolecular disulfide bonds, we hoped that expressing a cDNA clone in the bacterial system would provide us with active, soluble protein with no need for post-translational processing.

The promoter we used was designed to produce a normal translational product with no preceding fused peptide which would require processing during purification to obtain a normal, wild-type creatine kinase. A similar system has been used successfully to express a cDNA clone of chicken triosephosphate isomerase as active, soluble protein (122). The system we used incorporates a modification of the trc promoter developed by Amann (123). The sequence of the promoter region is shown in Figure 10.

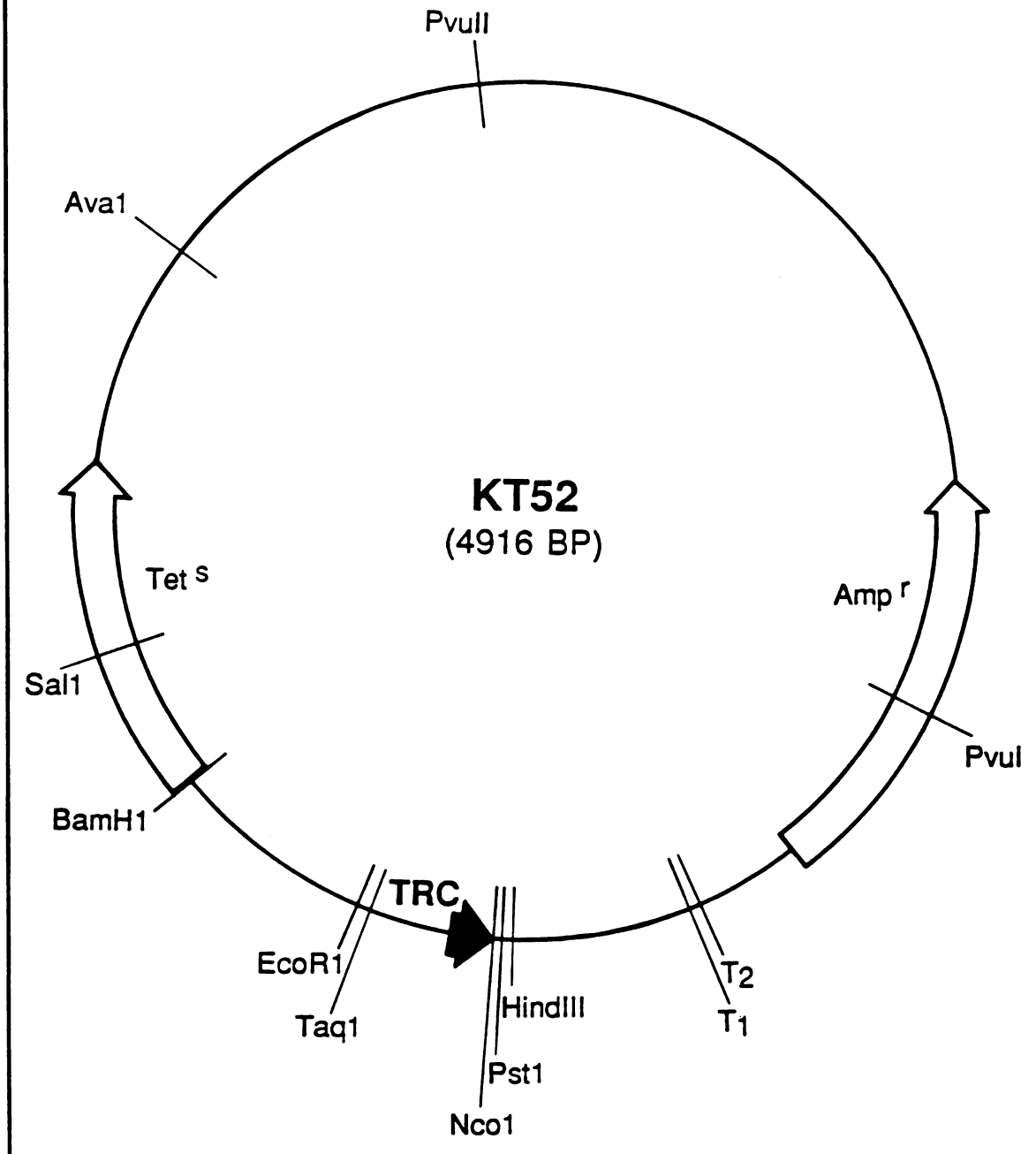
A variation of the earlier tac promoter (124), the trc promoter utilizes portions of both the lac promoter and the trp promoter and includes the Shine-Dalgarno sequence. The region labeled "-10" in Figure 10 represents the consensus RNA polymerase binding site, the "Pribnow Box," (125) from the "-10" region in the E. coli lac operon. It is regulated in nature by lactose (126). The region labeled "-35" in Figure 10

Figure 10: Sequence of the Trc Promoter

represents the regulatory binding site from the "-35" region of the E. coli trp operon (125). The Shine-Dalgarno sequence is the consensus ribosomal binding site (127) and it occurs 8 bases upstream of the initiating ATG of creatine kinase. This spacing between the Shine-Dalgarno sequence and the initiating ATG has been shown to be critical for translation. Variations of this distance to less than or greater than 8 bases results in a 10-1000-fold reduction of translated protein (123).

The promoter was obtained already cloned into a PBR322-derived plasmid called KT52 (124). It has tetracyline sensitivity and ampicillin resistance. There is an NcoI site at the initiating ATG (as shown in Figure 10). About 420 bases downstream from this site are two transcriptional stop signals, called T_1 and T_2. This vector was designed to produce large amounts of protein from the cloned insert since the trc promoter is considered to be among the "strongest" yet developed (128). When transfected into a bacterial host possessing the lac repressor gene, it can be induced to produce the translation product by addition of the lactose analog, isopropylthiolgalactoside (IPTG) (126). A simplified model of this vector is shown in Figure 11.

Figure 11: Expression Vector KT52



Materials and Methods

Creatine kinase clone CK52g8 was obtained from Dr. Brian West (26). The KT52 expression vector was obtained from Dr. Karen Talmadge of California Biotechnology Company . E. coli strains JA221, JA221-I, D1210, JM101, RR1 were provided by Dr. Brian West and strain dg98, a quick-growing mutant of JM101 was provided by Dr. Charles Craik of the Department of Biochemistry and Biophysics and the Department of Pharmaceutical Chemistry, respectively. Phage M13 Mp10, Mp11, Mp18 and MP19 were obtained from Dr. Brain West. Oligonucleotides for sequencing and mutagenesis primers were synthesized at the UCSF Biotechnology Support Center. The double-stranded sequencing kit was obtained from International Biotechnologies, Inc. Reagents and enzymes were exclusively molecular biology grade and were obtained from a variety of suppliers. ^{32}P- and ^{35}S- radionucleotides were purchased from Amersham or ICN. All DNA and bacterial manipulations were done either in disposable labware or in glass which had been treated in a dichromate bath followed by rinsing with double-distilled deionized water. All water used in the experiments was double-distilled and sterilized, as were all buffers. Solutions which could not be sterilized by autoclaving were sterilized by filtration through 0.2 or 0.4um filters obtained from Nalgene.

General methods for DNA and bacterial manipulation were patterned after Maniatis (68) or as referenced. Agarose gels were 1% agarose from SeaChem run in Tris, EDTA, Acetic acid buffers as specified by Maniatis.

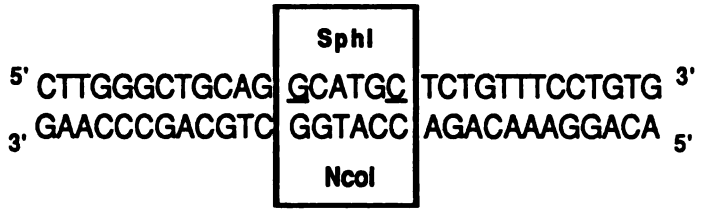
Results and Discussion

1. Construction of the Expression Vector for Creatine Kinase

A. Modification of expression vector KT52:

In order to obtain translation of a native-type CK, expression vector KT52 was modified to replace the original NcoI endonuclease site (8 bases downstream of the Shine-Delgarno sequence) with an SphI site (see Figure 11). Both sites contain an initiating ATG in the proper position (relative to the promoter) for transcption. The mutation was done so that the CK cDNA could be cloned into the initiating ATG site in such a way that no amino acids in the translated protein would be changed from the native sequence. This change would have occurred had we cloned the CK insert directly into the NcoI site rather than the SphI site. The modification was done by site-directed mutagenesis (129) using a 31-base oligonucleotide synthesized to complement the KT52 sequence at all positions around the NcoI site except those required to change the NcoI site to Sph I. The pertinent sequences are shown in Figure 12.

Figure 12: Sequences Involved in Mutagenesis of KT52: NcoI to SphI



In Figure 12, the upper sequence represents the 31-base oligonucleotide used for the mutagenesis. The lower sequence represents the complementary DNA of vector KT52. The SphI and NcoI sites are boxed and the noncomplementary bases involved in the mutation are underlined.

To prepare single-stranded (ss) KT52 template for mutagenesis, we cloned the EcoRI/PstI insert from KT52 (see Figure 11) into replicative form (Rf) M13 MP10 by cutting both Rf M13 vector and KT52 DNA with restriction endonucleases EcoR1 and Pst1. [Rf M13 vector is the double-stranded (ds) form of the phage. In this form it can be treated, for these purposes, as any double-stranded plasmid.] We then purified the KT52 insert, removed the 5' phosphate of the M13 Rf vector with calf intestinal phosphatase, ligated the cut M13 vector with KT52 insert and transformed the ligation mix into JM101 E.coli in the presence of helper phage. The transformation mix was plated on Luria Broth (LB) agar in the presence of 20 mM IPTG and XGal (4-bromo-5- chloroindol-3-yl-beta-D-galactoside) to produce blue-colored plaques from clones without KT52 insert and clear plaques from clones containing the KT52 insert. Single-stranded phage templates to be used in the mutagenesis reaction were prepared from several of the clear plaques that resulted.

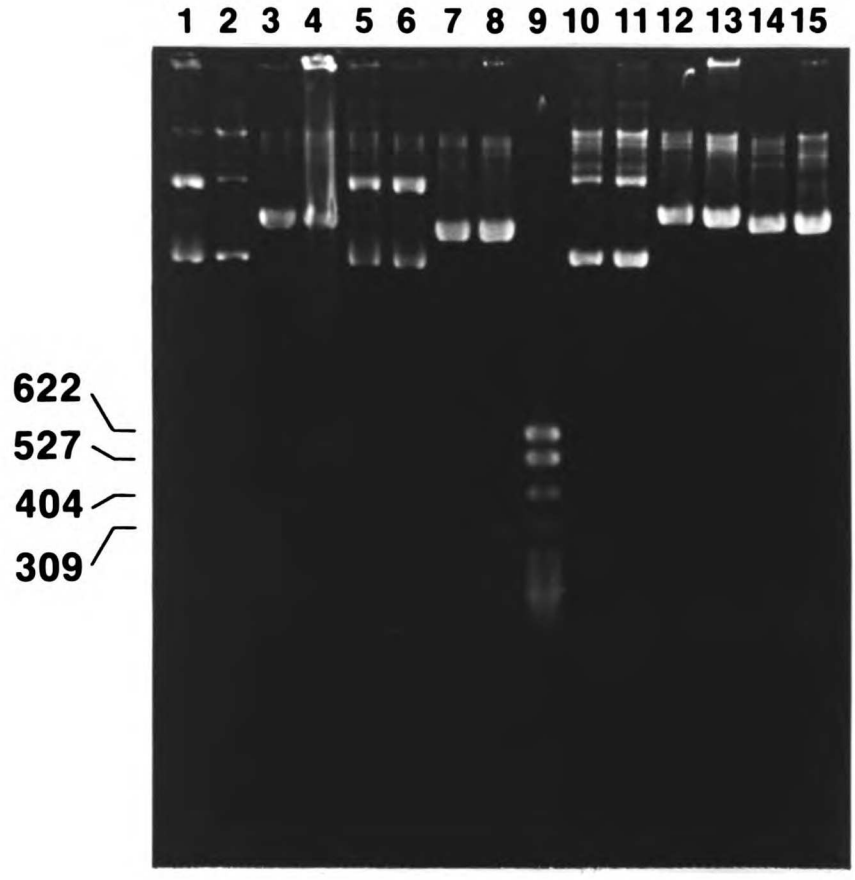
The 31-mer primer was purified by polyacrylamide gel electrophoresis followed by extraction and purification of the DNA from the gel. Mutagenesis conditions were optimized using the 31-mer oligonucleotide as primer in dideoxy-sequencing reactions. Mutagenesis was performed using double primers by the method of Craik (129) and the mixture was transformed into JM101 as described above.

Following mutagenesis, plaques likely to contain the KT52/SphI mutant were identified by colony hybridization using the mutagenesis primer as the labeled probe and plaques containing wild-type DNA (KT52) as the negative control. The likely positives were then purified by restreaking them on new plates. Several clear plaques from each plate were then grown as individual cultures in 96-well plates. A second round of hybridization experiments led to identification of purified clones likely to contain the mutant ss DNA. The positive clones from these experiments were prepared as ss template and sequenced. One clone, KT52/Sph19, proved to have an authentic SphI site replacing the Nco1 site in the original KT52 EcoR1/Pst1 insert.

This clone was given to Dr. Brian West who subcloned the mutant EcoR1/Pst1-Sph1 insert back into KT52, removing the analogous Nco1 insert. He also changed the HindIII endonuclease site in this vector to Bgl-II (130). This was done to provide the expression vector with a site compatible to CK52g8 in the 3' noncoding portion of the clone. The Bgl-II site spans residues 1330-1336 (see Figure 5) in CK52g8 and occurs about 110 bases in the 3' direction from the end of the coding region for CK. Inserting the 3' end of the CK clone into the vector at the new Bgl-II site of the expression vector would place the 3' end of the CK noncoding region between the translational stop codon (bases 1236-1238) in CK52g8 and the transcriptional stop signals, T_1 and T_2 in the expression vector.

The complete expression vector, named KT52/SphBgl, was analyzed by restriction endonuclease cutting (59). The cut fragments were fractionated on agarose gels to verify the construction. Figure 13 is a photograph of such an analytical gel.

Figure 13: Restriction Analysis of KT52/SphBgl



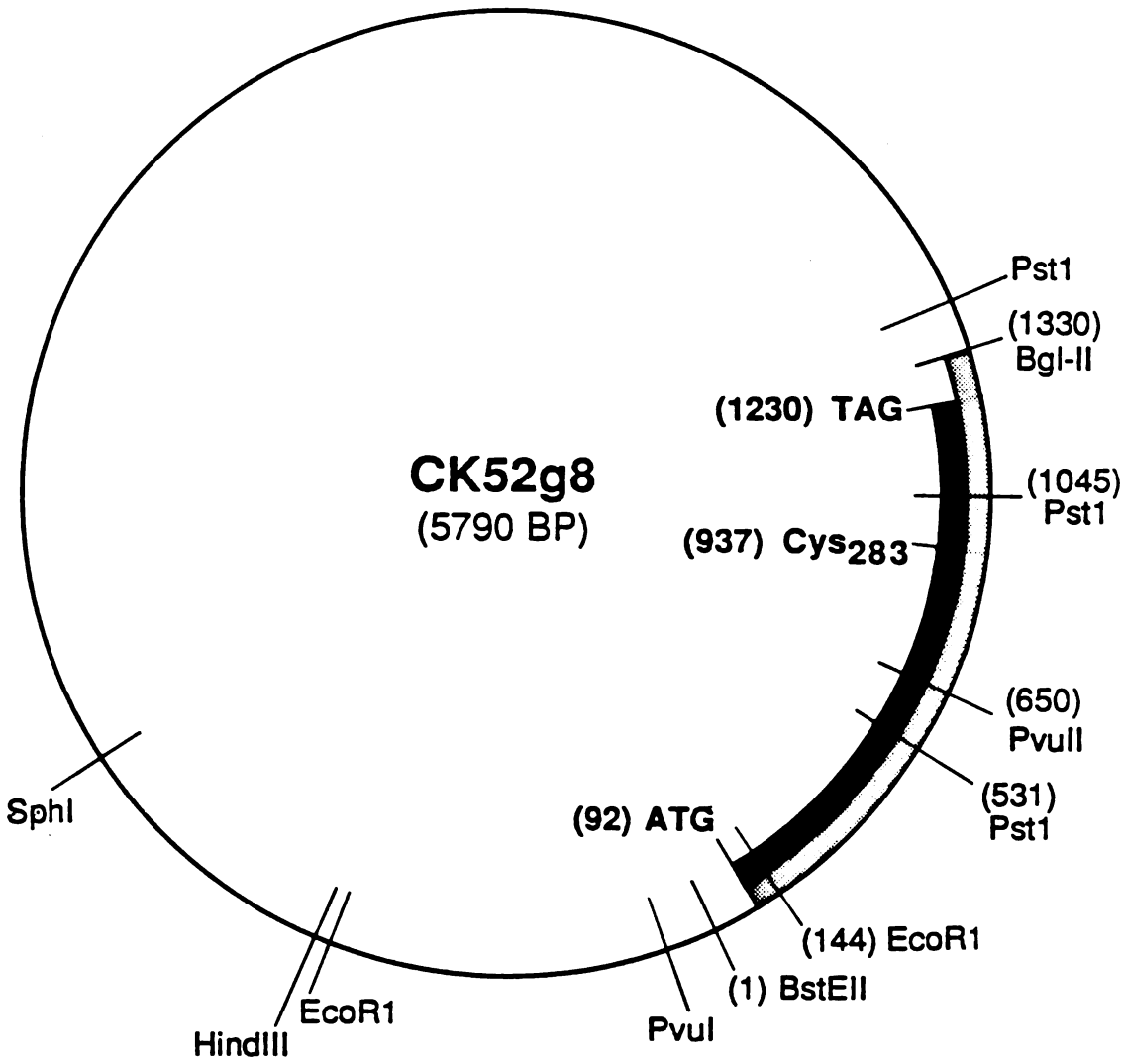
Lane 1) Supercoiled (SC) KT52; Lane 2) SC KT52SphBgl; Lanes 3-8 are KT52 cut with: Lane 3) NcoI; Lane 4) HindIII; Lane 5) SphI; Lane 6) Bgl-II; Lane 7) EcoR1/NcoI; Lane 8) EcoR1/HindIII; Lane 9) PBR322 cut with HpaII (BP lengths numbered on the left of the gel); Lanes 10-15 are KT52SphBgl cut with: Lane 10) NcoI; Lane 11) HindIII; Lane 12) SphI; Lane 13) Bgl-II; Lane 14) EcoR1/SphI; Lane 15) EcoR1/Bgl-II.

It was found, as expected, that the supercoiled form (SC) of the mutated expression vector, KT52/SphBgl, was the same size as the SC form of the original vector, KT52 (Lanes 1,2). The control DNA, KT52, should cut with NcoI and HindIII to give linearized plasmid (Lanes 3,4), and should not cut with SphI or Bgl-II (Lanes 5,6), as can be seen from Figure 13. In contrast, KT52/SphBgl should not cut with NcoI or HindIII (Lanes 10, 11) but should cut with SphI and Bgl-II (Lanes 12,13). Figure 13 confirms these expectations and shows as well that the KT52 DNA cut with either EcoR1 + NcoI (Lane 7) or EcoR1 + HindIII (Lane 8) gives fragments of the same size (~ 300 BP) as KT52/SphBgl DNA cut with either EcoR1 + SphI (Lane 14) or EcoR1 + BglII (lane 15). The fact that these inserts are all the same size (Lanes 7, 8, 14, 15) implies that no major accretions or deletions occurred in the KT52/SphBgl vector during the cloning and mutagenesis manipulations. It also confirms that all of the sites involved in those experiments remain intact. KT52/SphBgl was transformed into D1210 cells for large-scale preparation and purificiation.

B. Modification of CK52g8:

The Torpedo CK cDNA clone, CK52g8, was also modified so that it could be cloned into the expression vector at the correct position, the SphI site, for in-frame translation of the protein. From the sequencing experiments, we had an EcoR1 insert of CK52g8 spanning the initiating ATG (CK52g8/EcoR1) already cloned into M13 MP10. We inserted an Sph1 site at the initiating ATG of the coding strand of the CK insert, CK52g8/EcoR1. Figure 14 is a graphic of CK52g8.

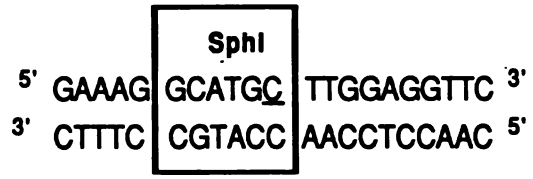
Figure 14: CK52g8



The portion of the clone transferred to the expression vector KT52SphBgl is designated by the shaded area while the portion coding for CK protein is designated by the black area. Restriction sites are numbered according to Figure 5 with base #1 at the BstEII site. The initiating ATG is labeled at base #92 and the translational stop codon, TAG, is labeled at base #1230.

For this mutagenesis procedure, we used a 21-base oligonucleotide containing the Sph1 site and complementary to the noncoding strand of this CK insert. This region is shown in Figure 15.

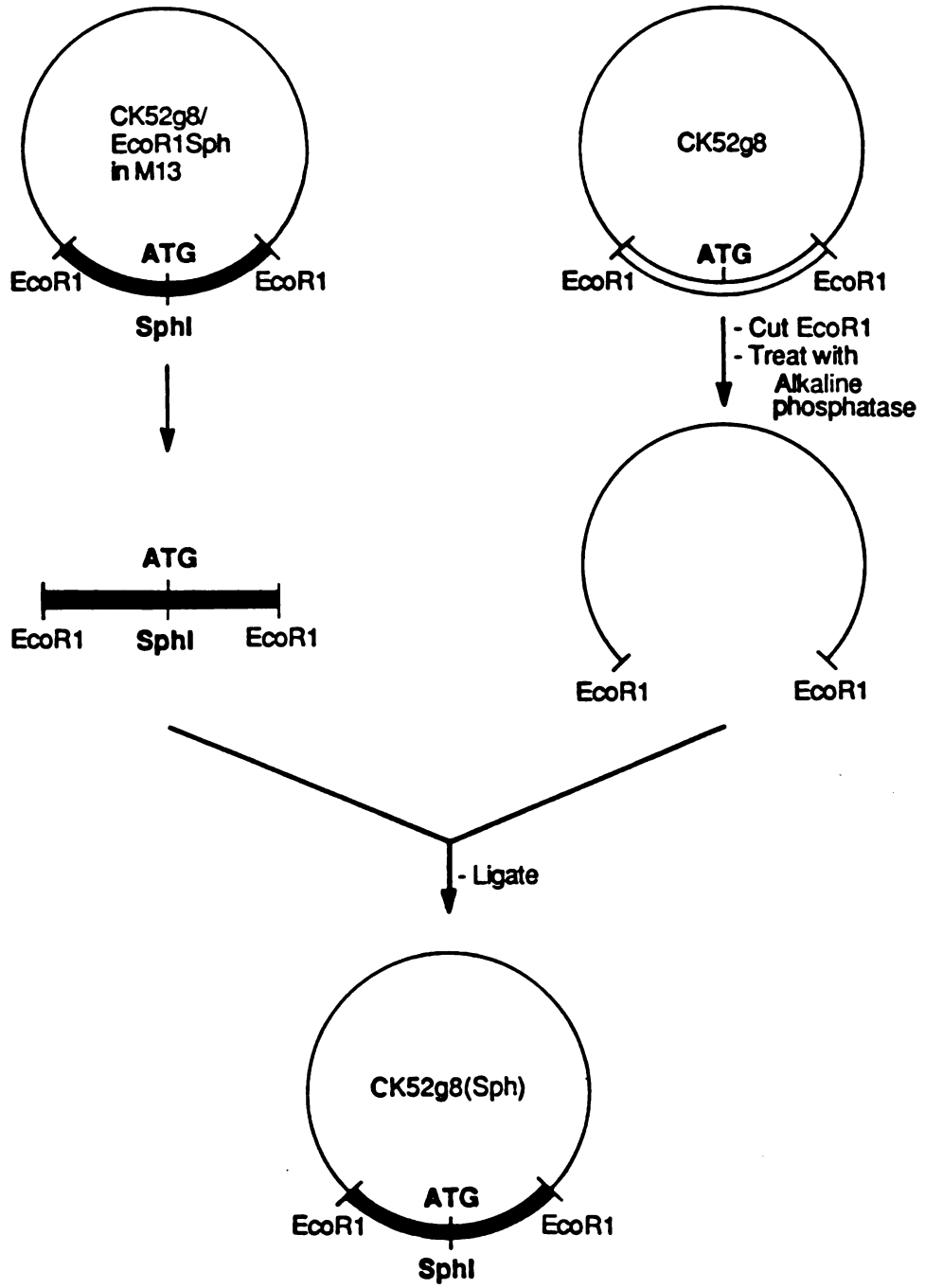
Figure 15: Sequences Involved in Mutagenesis of CK52g8 to CK52g8/Sph



The upper sequence in Figure 15 represents the 21-base oligonucleotide used in the mutagenesis experiment. The lower sequence is the complementary stretch of DNA on the coding strand of CK cDNA. The initiating ATG can be seen within the boxed area outlining the SphI site in the upper sequence. The noncomplementary base in the oligonucleotide primer which produced the mutation and created the SphI site is underlined. The original CK DNA had no known restriction site in the boxed region.

The mutagenesis conditions were optimized by sequencing reactions using the 21-mer as primer. Following mutagenesis, likely mutants were identified by the same cycle of hybridization experiments used for KT52/Sph19. Mutants were verified by sequencing. One correct clone was grown and purified in quantity as the Rf M13 form and named CK52g8/Eco-Sph. Figure 16 shows the scheme used to replace the CK52g8/EcoR1 insert in CK52g8 with the mutant CK52g8/EcoSph insert.

Figure 16: Replacement of CK52g8/EcoR1 Insert with CK52g8/EcoSph

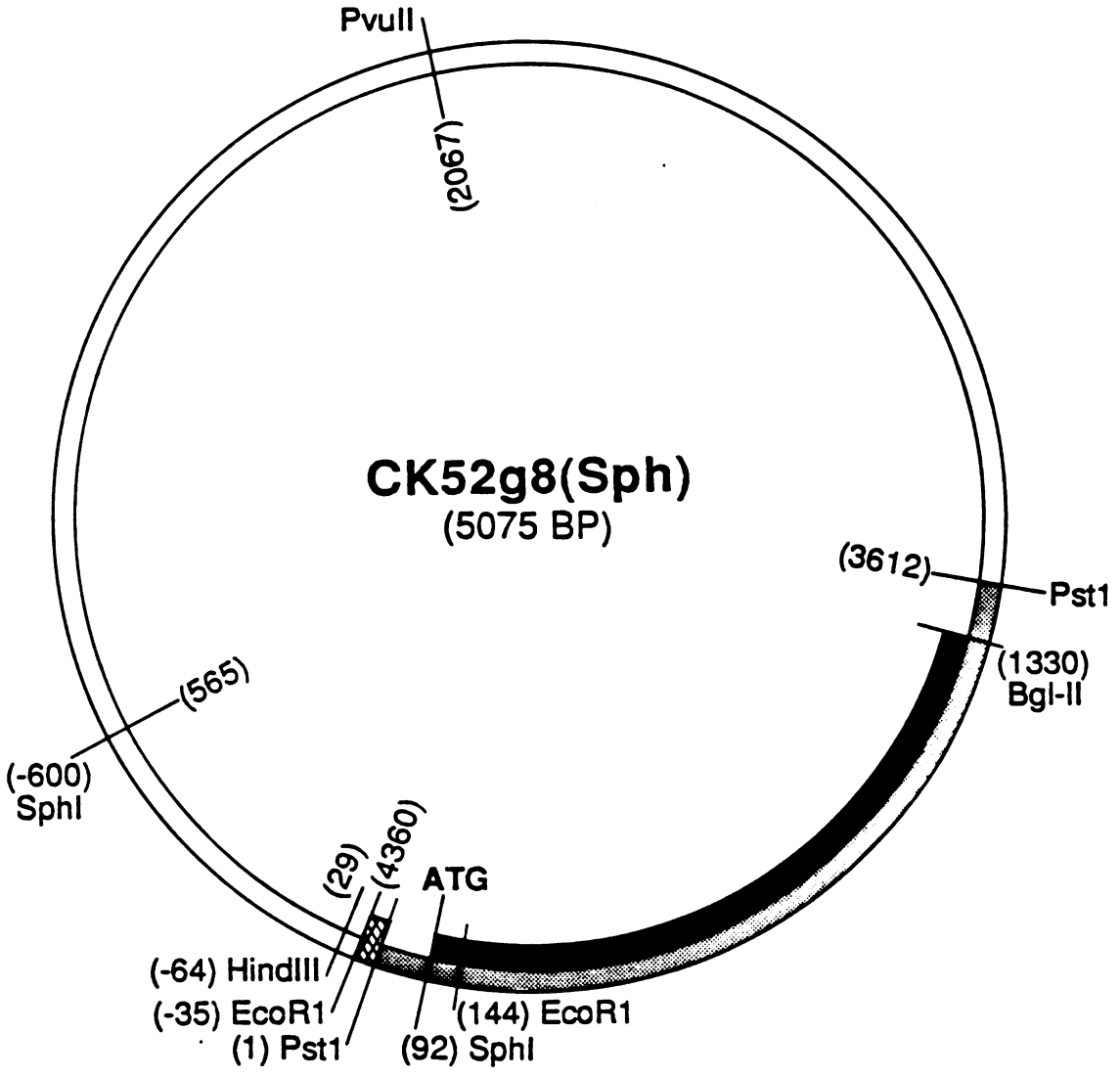


The new CK clone with the SphI site inserted at the initiating ATG of CK52g8 was named CK52g8(Sph). The preliminary construction, CK52g8(Sph), was treated with restriction endonucleases and the fragments were analyzed on agarose and acrylamide gels to be sure that an SphI site had been added at the initiating ATG. This data is not shown since a later gel of the completed expression vector, with this mutated CK cloned into KT52SphBgl, is shown in Figure 21. The data generated from the restriction analysis shown in Figure 21 confirms the correctness of the CK52g8(Sph) construction.

The new CK52g8(Sph) construction includes a small portion of the multiple cloning site from M13 MP10. This fragment is not present in CK52g8. In order to avoid confusion about the sizes of inserts generated from restriction analysis of CK52g8(Sph) (discussed below), Figure 17 has been included. This Figure is a partial restriction map of CK52g8(Sph) identifying the sources of the DNA inserts which were combined to make this clone.

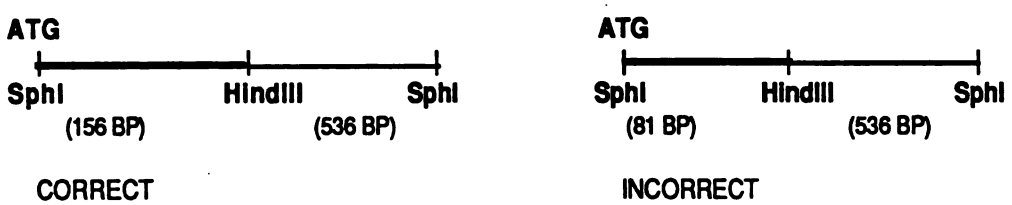
Because the mutated insert had an EcoR1 site at both the 5' and 3' ends, it was likely that some proportion of clones of proper size and which cut with Sph1 would have the wrong orientation for the CK52g8/ EcoSph insert. This was checked on a number of these clones by cutting them with Sph1 and HindIII simultaneously. [There is a HindIII site 5' upstream of the CK insert in CK52g8(Sph) which is derived from the PBR322 vector used for the original CK52g8 CK insert. It is located between the PBR322 SphI site and the CK52g8(Sph) SphI site (Figure 17).] Figure 18 is a graphic of results expected from such cuts for both the correct and incorrect orientations.

Figure 17: CK52g8(Sph)



CK52g8(Sph) represents a full-length CK insert cloned into PBR322 with an SphI site added at the initiating ATG of the CK clone. The two Pst1 sites represent the original PBR322 site into which the tailed CK clone was inserted (see Chapter 1). The open area represents the original PBR322 vector and the position numbers inside the plasmid circle are PBR322 positions. The shaded area represents the CK insert, including noncoding regions. The portion of the CK insert cloned into the expression vector is represented by the black area. The crosshatched area is DNA from M13 MP10 carried over in the exchange of the EcoR1/Sph insert with with the wild-type EcoR1 insert from CK52g8. This exchange deleted approximately 750 BP of PBR322-derived DNA, leaving CK52g8(Sph) about 715 BP smaller than CK52g8. The numbering on the outside of the plasmid circle is negative for PBR322-derived sites 5' of the CK insert. The numbering is positive for CK DNA and position #1 (the 5' Pst1 site) is analogous to the #1 position (the 5' BstEII site) in CK52g8 (Figure 14).

Figure 18: Orientations Possible for CK52g8(Sph)



After cutting several clones with SphI and HindIII, the fragments were end-labeled with ^{32}P-CTP and run on an 8% polyacrylamide gel along with ^{32}P-labeled PBR322 length markers. Autoradiography of this gel revealed several clones with SphI/HindIII inserts of the correct size (156 BP) as well as several with the smaller, incorrect insert size (81 BP). One of the correct clones, CK52g8/Sph #18, was renamed CK52g8Sph and transformed into RR1 cells for large-scale preparation and purification.

C. Assembly of the expression vector: CK52g8Sph into KT52SphBgl:

With both CK52g8Sph and KT52SphBgl modified to be compatible, construction of the actual expression vector was straightforward as shown in Figure 19. Figure 20 shows this final construction, KTCK3F. Figure 21 shows the restriction analysis of KTCK3F compared KT52SphBgl and Table 4 catalogues the expected and observed insert sizes obtained.

Figure 19: Assembly of the Expression Vector KTCK3F

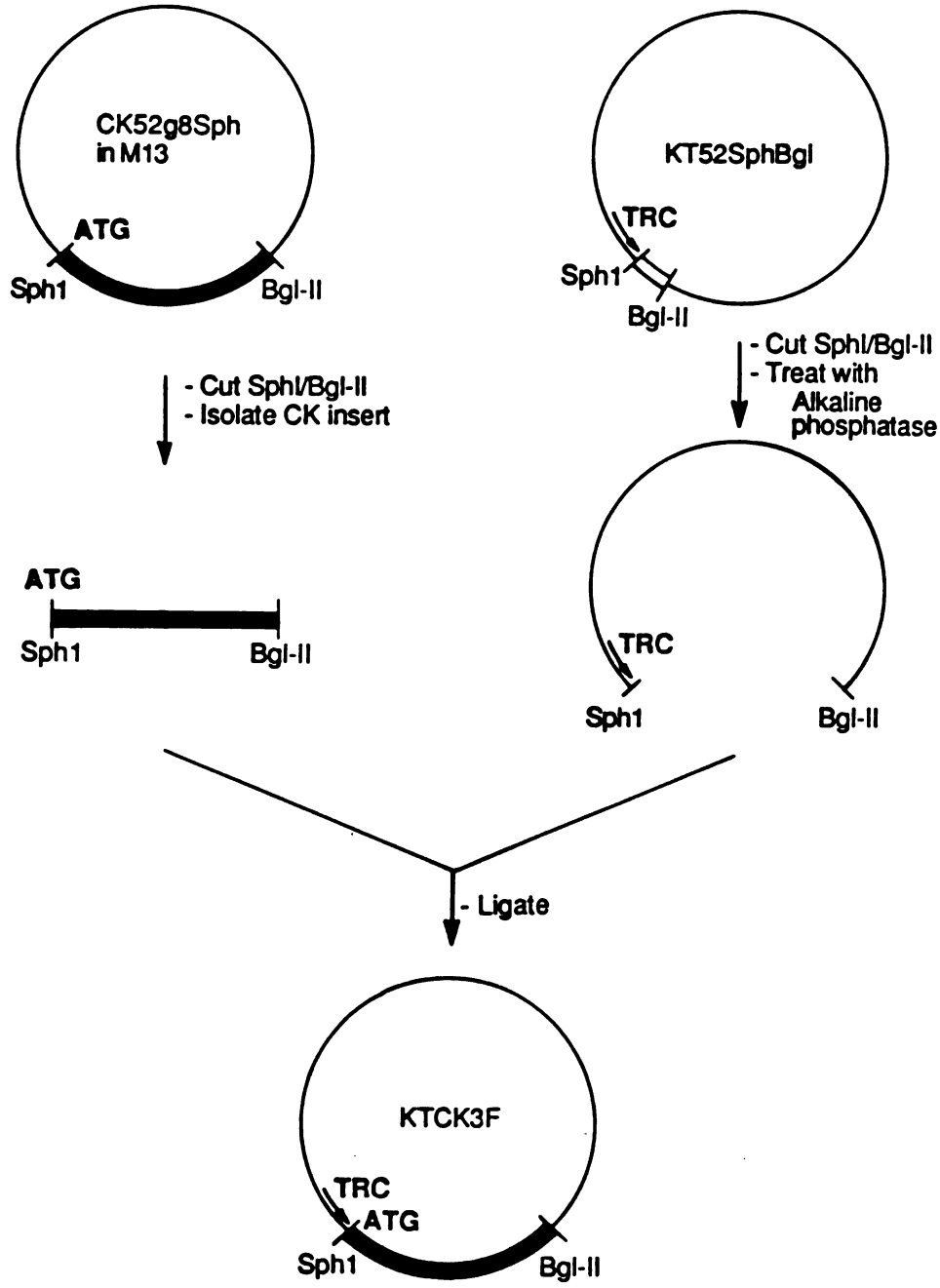
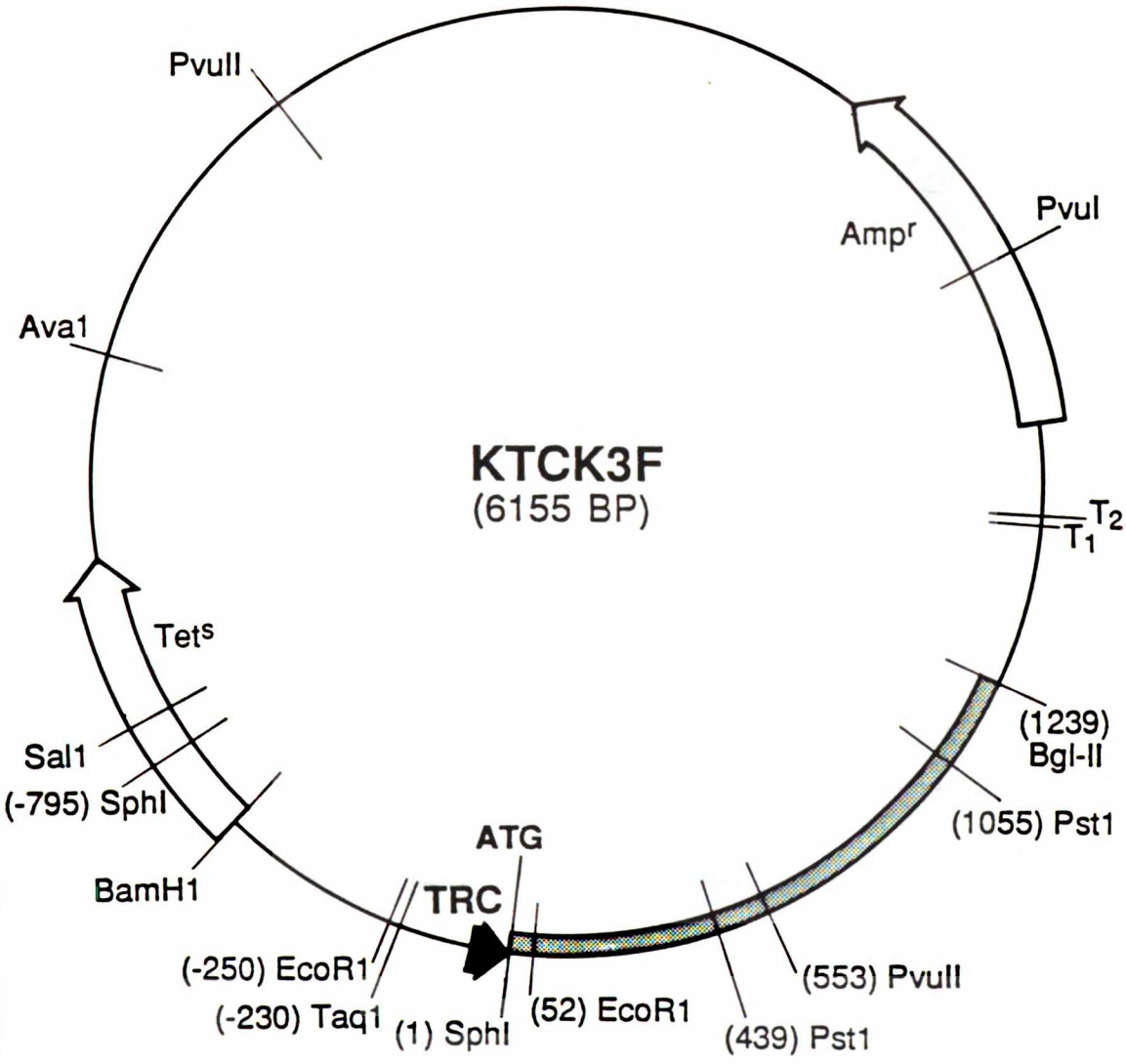
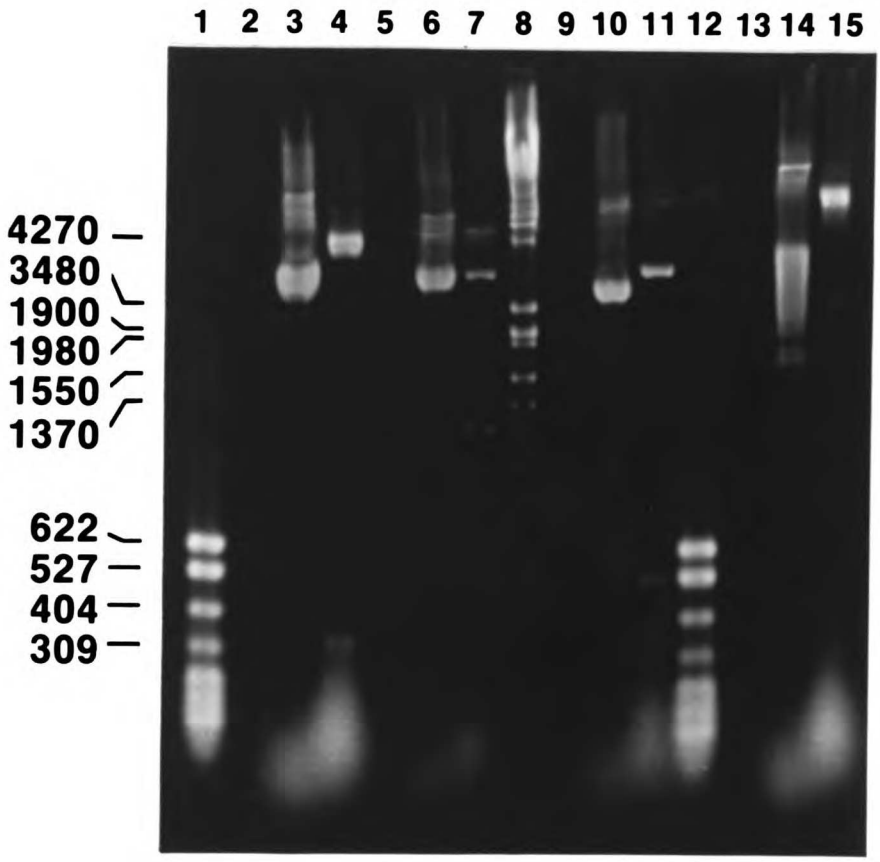


Figure 20: Expression Vector for Wild-type CK: KTCK3F



The position numbers accompanying site labels are in relation to the initiating ATG of CK-coding DNA. The 5' terminus of the SphI site spanning this ATG is designated as base #1. The negative numbering is approximate in the portion of the DNA 5' of base #1 and is meant to give the distances of these sites relative to the 5' start of the CK-coding portion. The CK insert is shown as the shaded area of the plasmid.

Figure 21: Restriction Analysis of KTCK3F Compared to KT52SphBgl



Lane 1) PBR322 cut with HpaII (BP lengths numbered on the left of the gel); Lanes 2, 5 9 13) Blank; Lane 3) KT52SphBgI, EcoR1 cut; Lane 4) KTCK3F, EcoR1 cut; Lane 6) KT52SphBgI, SphI/BgI-II cut; Lane 7) KTCK3F, SphI/BgI-II cut; Lane 8) λ DNA cut with HindIII/EcoR1 (BP lengths numbered on the right of the gel); Lane 10) KT52SphBgI, EcoR1/Pst1 cut; Lane 11) KTCK3F, EcoR1/Pst1 cut; Lane 12) PBR322 cut with HpaII; Lane 14) KT52SphBgI, SC; Lane 15) KTCK3F, SC.

Table 4: Observed Sizes of Inserts from Restriction Analysis of KTCK3F and KT52SphBgl

Restriction Endonuclease cuts	Observed Insert Sizes (BP)a KT52SphBgl	KTCK3F
EcoR1	linearized	300
SphI/Bgl-II	<100, 795	795, 1240
EcoR1/Pst1	300	300, 387, 616

aThe insert sizes for fragments including vector DNA (rather than exclusively CK coding and noncoding DNA) are only approximate because we do not have exact sequence information on the original expression vector, KT52. Large DNA fragments can also be observed near the top of the gel. These are not diagnostic for the analysis of KTCK3F.

The EcoR1 cuts (lanes 3, 4) show that there is a new EcoR1 site in KTCK3F as compared to the KT52SphBgl control. It is about 300 BP in length, showing that the EcoR1 site involved in construction of CK52g8(Sph) and therefore in KTCK3F is intact and at the correct position in the clone. The cuts with SphI/Bgl-II (lanes 6, 7) show a fragment of approximately 795 BP in both DNA samples. These span a region of the KT52SphBgl vector with the 3' end as the SphI site in which the CK was inserted at its 5' end. The result is important in that it shows the SphI site to be intact at the initiating ATG for CK transcription. The small (<100 BP) fragment expected for KT52SphBgl cannot be seen on this gel and is unimportant since we had already shown that there is an intact Bgl-II site in KT52SphBgl. The 1240 BP fragment seen in lane 7 of Figure 21 represents the entire CK

clone inserted into KT52SphBgl. It is approximately the correct size as can be seen from the DNA length markers in lane 8. (See the Figure legend for the sizes of these length markers). It also shows that the SphI and Bgl-II cloning sites used to make the final KTCK3F construction are intact. The EcoR1/Pst1 cut in Lane 10 shows a 300 BP fragment for KT52SphBgl. Reference to Figure 11 shows that this region spans the promoter region of the vector from the EcoR1 site to the Pst1 site in KT52. The short piece of DNA between the Nco1 and HindIII sites in KT52 (between the SphI and Bgl-II sites in KT52SphBgl) containing this Pst1 site has been removed in KTCK3F. In lane 11, the EcoR1/Pst1 cut shows a 616 BP insert representing the Pst1-Pst1 fragment spanning bases 553-1055 in Figure 20. This data is confirmed in Lane 12, which shows the same size insert when KTCK3F is cut with Pst1 alone. The 387 BP insert in lane 11 represents the region in the CK clone from base 52 (EcoR1) to base 439 (Pst1). Lane 11 shows a 300 BP insert as well, but this represents the EcoR1 insert spanning bases -250 to 52 as numbered in Figure 20. The cuts represented by lanes 10-12 were done to confirm in a general way that the sizes of the inserts were those expected if the clone had been constructed correctly. The SphI, EcoR1, and Bgl-II cuts also show that the integrity of these sites had been maintained during the manipulations involved in the CK52g8(Sph) and KTCK3F constructions. The DNA shown in lanes 14 and 15 is supercoiled KT52SphBgl and KTCK3F respectively and shows, as expected, that the clone containing the CK insert (KTCK3F) is larger than the control DNA (KT53SphBgl) containing no CK.

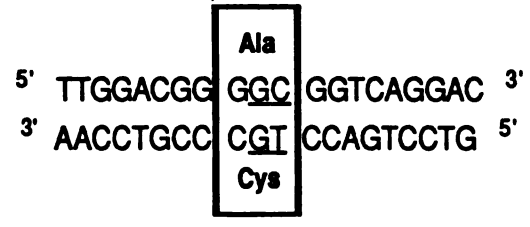
2. Construction of the Expression Vector for Ala_{283} Mutant Creatine Kinase.

A. Modification of CK52g8-520A:

We performed site-directed mutagenesis on the CK cDNA so that it would encode for alanine at amino acid position 283 rather than the cysteine in the wild type. This was done on a subclone of CK52g8, CK52g8-520A, which had been cloned into M13 MP10 for the sequencing experiments (130). This Pst1 insert spans bases 525-1045 in CK52g8 (see Figure 14). The DNA to be modified, the Cys_{283} codon, spans bases 937-939 in CK52g8. Following mutagenesis, we interchanged this subclone with the analogous Pst1 insert in the wild-type expression vector, KTCK3F.

Figure 22 shows the 20-base oligonucletide synthesized to mutate the DNA coding for Cys_{283} to Ala_{283} as the upper strand. The lower strand shows the complementary DNA from CK52g8-520A. The relevant codon in the upper strand is labeled "Ala" and its cysteine-encoding complement in the lower strand is labeled "Cys." Both codons are boxed. The bases involved in the mutagenesis are underlined.

Figure 22: Sequences Involved in Mutagenesis: Cys_{283} to Ala_{283}

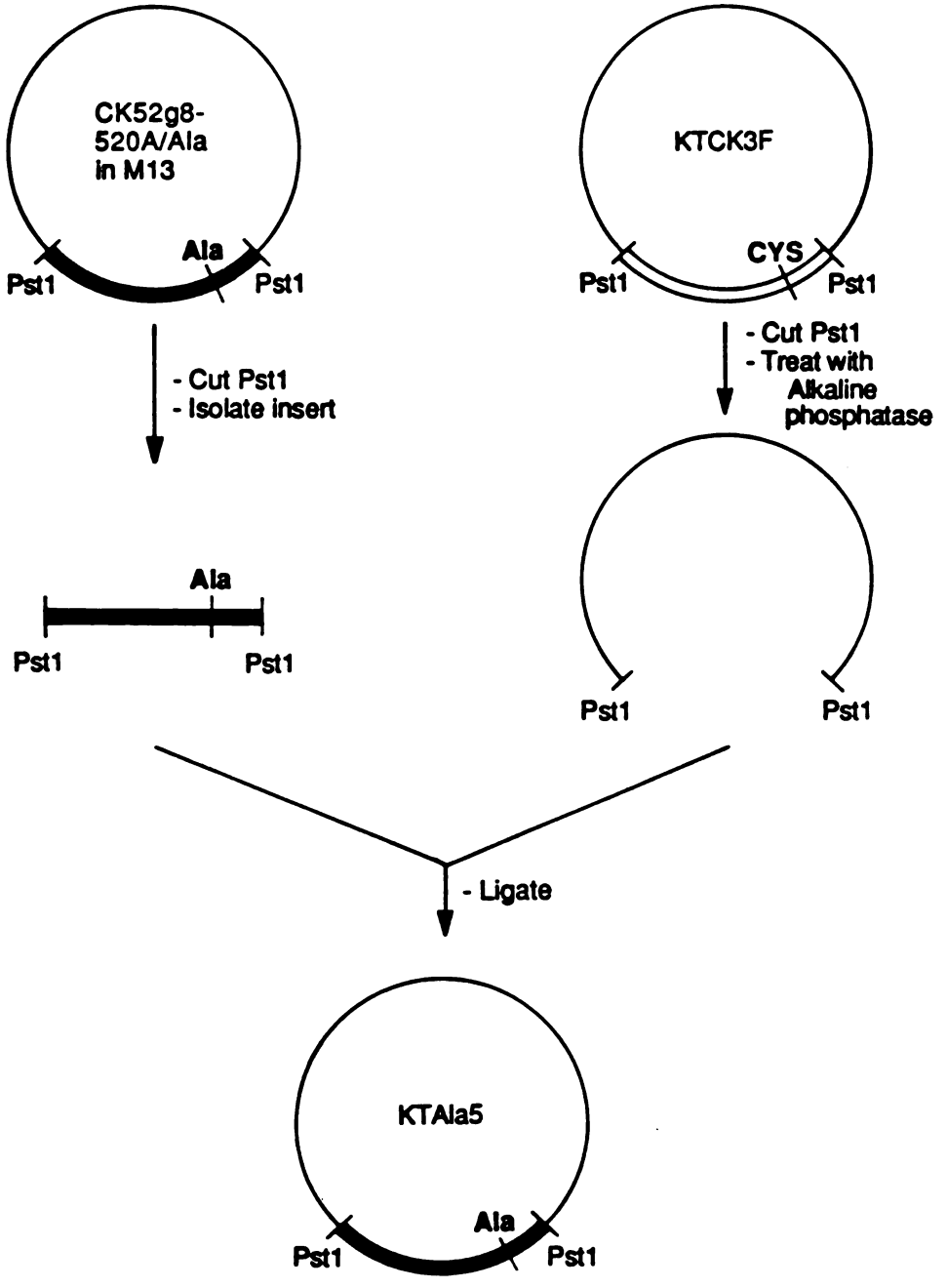


Template preparation, oligonucleotide purification, and mutagenesis was performed as described earlier. The mutant clones were identified by two cycles of hybridization experiments using the mutagenesis olionucleotide as the labeled probe. Mutagenesis was verified by sequencing.

Figure 23 shows the scheme for cloning the CK52g8-520A/Ala subclone into the expression vector KTCK3F. The transformation of this new clone, KTCK3F/Ala, was done in E. coli strain D1210 which does not express the protein.

Verification of this clone was originally done by colony hybridization to pick out the clones with the alanine mutation. This method had to be used because the mutation had not introduced a new restriction site (as had the mutations involved in the construction of KTCK3F) which could be used for preliminary screening to separate primary transformants with and without the mutation. These experiments were done by hybridization of the ^{32}P- labeled 20-base oligonucleotide used as the mutagenesis primer to the double-stranded putative KTCK3F/Ala mutants. Colony lifts were not done in this case, but only the second cycle of hybridization experiments as described above were done on single colonies from the primary plates. Several clones gave positive results from this screening and these were checked for orientation of the CK52g8-520A/Ala insert by cutting with PvuII and EcoR1. The correct orientation should have given (for the fragments of interest) an insert of approximately 500 base pairs and the wrong orientation an insert of approximately 800 base pairs (see Figure 20). While both orientations were found from the comparison of the relative sizes of the PvuII/EcoR1 inserts, the absolute sizes of the inserts looked incorrect. To investigate this problem, one clone of the

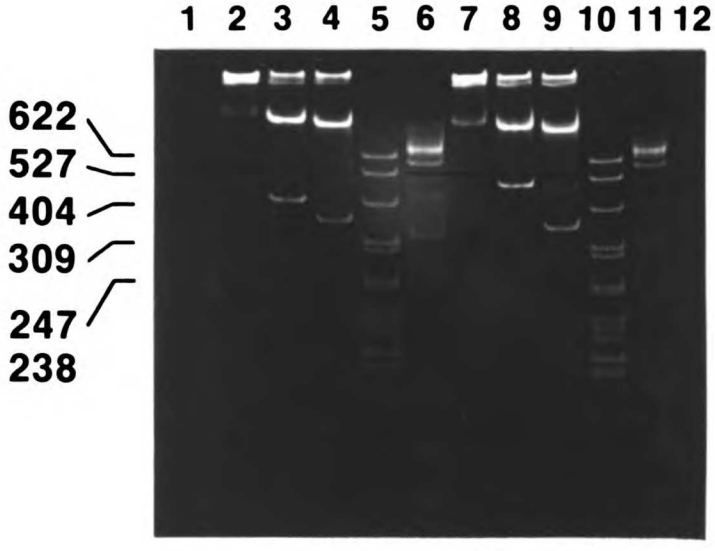
Figure 23: Cloning Scheme: CK52g8-520A/Ala into KTCK3F



apparently correct insert orientation, KTAla5, was amplified in the D1210 strain and purified over a cesium chloride gradient for further analysis. This clone was checked on agarose gels after cutting with several restriction endonucleases and it was found that the Pst1 insert, CK52g8-520A/Ala, was too short by approximately 100 base pairs. Figure 24 shows the gel from this analysis and Table 5 shows the deviation of KTAla5 from expected results. (Note: The DNA in lanes 6 and 11 of Figure 24 represents PBR322 DNA cut with MboII. These markers were not fully cut by the enzyme and should be ignored. The PBR322 DNA cut with HpaII shown in lanes 5 and 10 give correct BP lengths and are numbered at the left of the gel.)

As can be seen in Figure 24, the Pst1 insert is short in comparison with its analogous insert from KTCK3F. In Lane 3, the Pst1 insert is about 400 BP long when compared to the DNA length markers in lane 5. In contrast, the Pst1 insert from KTCK3F, shown in lane 8, is slightly over 500 BP in length when compared to the length markers. The size expected from analysis of the sequence is 514 BP. Further consideration of this gel reveals that the error is in the 5' portion of the Pst1 insert. The EcoR1/PvuII cuts (lane 4 for KTAla5 and lane 9 for KTCK3F) give this information. This insert should be approximately 500 BP long, as it is in the KTCK3F sample. It is only about 400 BP long in the KTAla5 sample. Since the 5' region of this insert (EcoR1-Pst1) was not manipulated in any way during the construction, this data shows that it is the region from the 5' Pst1 site to the PvuII site that is short by about 100 bases.

Figure 24: Restriction Analysis of KTAla5



Lane 1) Blank; Lane 2) KTAla5, SC; Lane 3) KTAla5, Pst1 cut; Lane 4) KTAla5, PvuII/EcoR1 cut; Lanes 5, 10) PBR322 cut with HpaII; Lanes 6, 11) PBR322 cut with MboII; Lane 7) KTCK3F, SC; Lane 8) KTCK3F, Pst1 cut; Lane 9) KTCK3F, PvuII/EcoR1 cut, Lane 12) Blank.

Table 5: Inserts Found from Restriction Analysis of KTCK3F and KTAla5

Restriction Endonuclease cuts	Sizes of Inserts found on Gel (BP) KTCK3F	KTAla5
Pst1	514	400 (approx.)
EcoR1/PvuII	503	400 (approx.)
	350[a]	350[a]

[a]This fragment represents an EcoR1-EcoR1 insert the 5' end of which is from an EcoR1 site from the original KT52. It is 5' upstream of the CK clone and spans the promoter region and the CK initiating ATG to its 3' end at the EcoR1 site designated at base #144 in CK52g8 (see Figure 14). It is also marked as base #52 in Figure 20, KTCK3F.

Further confirmation of this conclusion was provided by sequencing the original CK52g8-520A template. We synthesized an oligonucleotide primer designed to give us the sequence of the original CK52g8-520A template through the PvuII site and into the region 5' (upstream) of it where we expected the error to be. The sequencing gel showed both that the PvuII site was indeed intact and correct as was the sequence 3' (downstream and toward the alanine mutant site) from it. But the region some 20 bases upstream (5') of the PvuII site was garbled, confirming the location of the problem in the Pst1/PvuII insert (bases 439 to 558 in CK52g8, Figure 14).

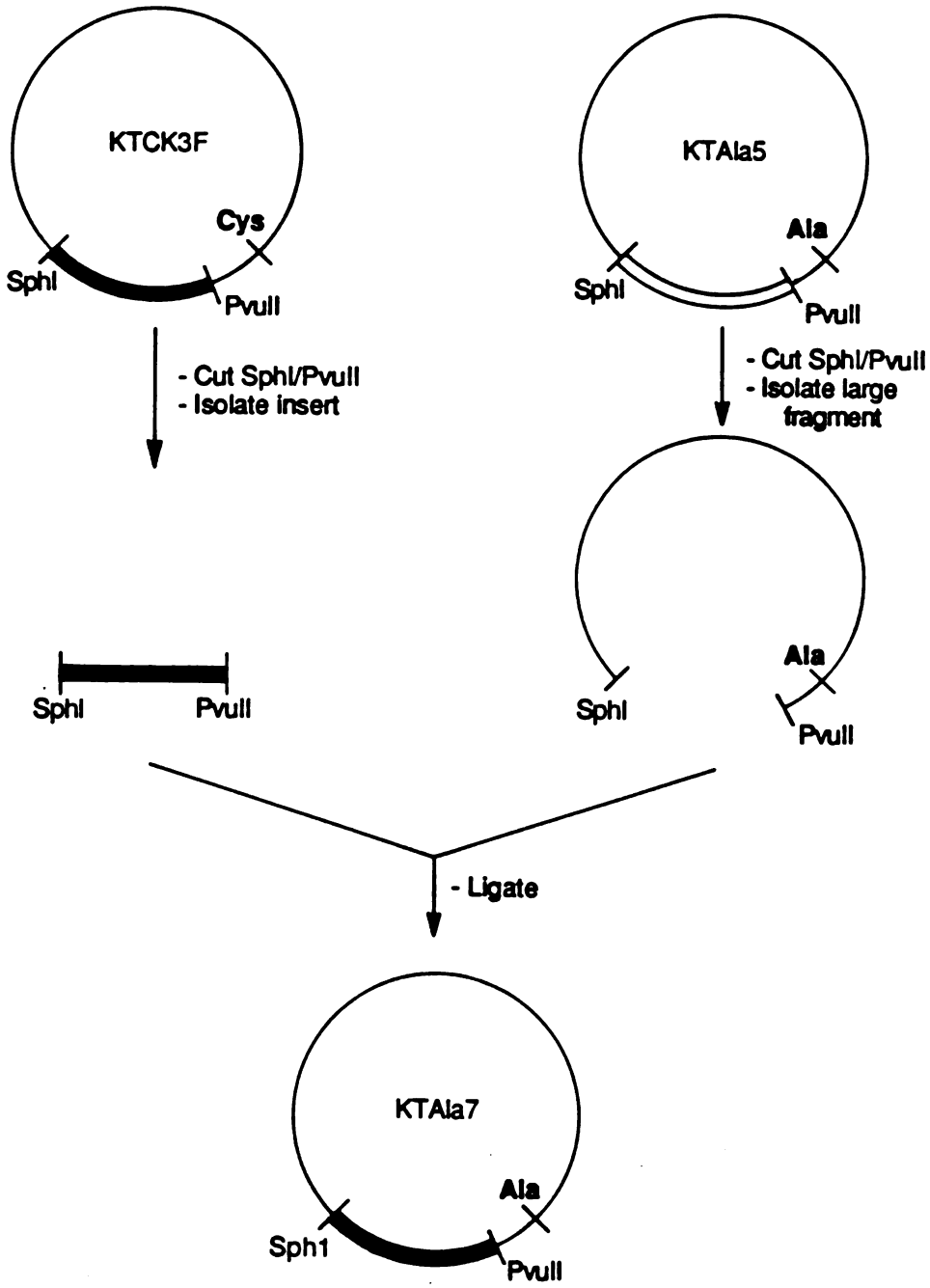
It is difficult to determine when the problem occurred, whether in the original cloning of the insert into M13 or in propagation of the DNA to produce the mutagenesis template. Although this error caused considerable loss of time, it did provide one compensation: it showed that the Pst1 insert in KTAla5 did contain the DNA we had mutated, rather than the original insert from KTCK3F. This extra confirmation was important since our original experimental design had been flawed. We had not isolated the KTCK3F large insert prior to ligation with CK52g8-520A/Ala, but had only treated it with alkaline phosphatase. While this treatment theoretically removed the terminal 5' phosphate of both KTCK3F/Pst1 and CK52g8/Pst1 (see Figure 23), preventing them from simply religating, it is likely that a few of these were not dephosphorylated and did religate. Our control experiments did show that a few such religations did occur. There were always a few transformants from ligation mixes containing only phosphatased KTCK3F/Pst1 and CK52g8/Pst1 inserts, in the absence of any CK52g8-520A/Ala material. Had we purified the KTCK3F/Pst1 large insert before making the construction shown in Figure 23, we would have greatly reduced the possibility that KTAla5 could contain unmutated DNA coding for Cys_{283} instead of Ala_{283}. But we had not done this crucial experiment. Even though we checked the completed KTAla5 for the Ala mutation by hybridization analysis with the mutation oligonucleotide as probe (data not shown), such analysis is only highly suggestive that the mutation is present. It is not nearly so definitive an analysis as is sequencing, for example.

After we had confirmed that CK52g8-520A (and therefore CK52g8-520A/Ala) was of a different size than the analogous insert in KTCK3F, we could be more confident that KTAla5 contained the mutant DNA. This confirmation was especially important because,

at the time we performed these experiments, we could not sequence the double-stranded KTAla5 directly to definitively establish the presence of the alanine mutation. Recently, however, we have sequenced the region of CK including the alanine mutation in the corrected clone, KTAla7 (discussed below) by new double-stranded sequencing techniques (131), and thus have finally confirmed the mutation directly.

However fortuitous the error may have been in the light of the original poor experimental design in making the KTAla5 construction, this error had to be repaired before we could try to express the mutant protein. We theorized that since the region of KTAla5 containing the alanine mutation was of correct size and correct sequence, we could repair the clone by removing the incorrect Pst1/PvuII insert and replacing it with the analogous insert from KTCK3F. This plan would leave intact in our putative CK/Ala expression vector the part of the CK52g8-520A/Ala insert confirmed to contain the alanine mutation. For technical reasons, a larger insert than the 119 BP Pst1/PvuII insert was actually chosen to exchange with its analog from KTCK3F. This would give the same result in correcting the bad stretch of DNA, but would considerably facilitate the DNA manipulations. We chose to use the SphI/PvuII insert from KTCK3F. This insert is contained within the CK coding region and can be identified as spanning bases 1-553 in KTCK3F (Figure 20). This exchange was performed by the scheme shown in Figure 25.

Figure 25: Construction of KTAla7

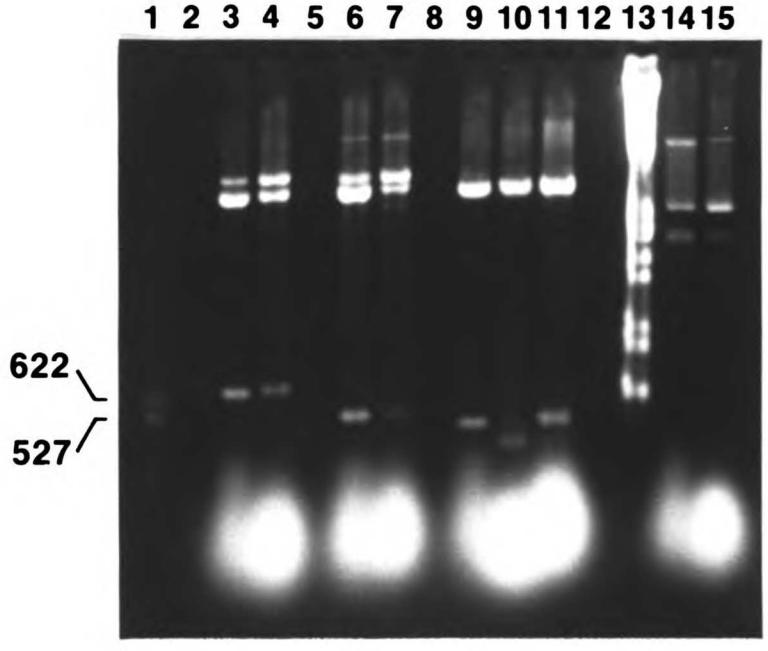


Transformants from this construction were examined by restriction analysis. One "correct" clone, KTAla7, was amplified and purified by cesium chloride centrifugation and eventually used in our expression experiments.

Figure 26 shows the restriction analysis of this clone, KTAla7, and Table 6 catalogues the results. The DNA length markers in lane 1 are labeled at the left of the gel. Unfortunately, they are very light and only the 622 and 527 BP ngths can be discerned. This is sufficient for the analysis, however. The markers in Lane 13 were accidentally run in too high salt concentration and should be ignored. The analysis shows the PvuII/Bgl-II inserts (lanes 3, 4) to be of about the correct size and identical for the two samples. The SphI/PvuII cut (lanes 6, 7) generates an insert about 550 BP in length as expected. This cut also confirms the integrity of both sites used to make the construction. Most importantly, the Pst1 cuts (lanes 9-1, show that this insert is about 500 BP in length for both KTCK3F and KTAla7. The Pst1 cut of the KTAla5 clone is included in lane 10 for comparison. This insert is clearly about 100 bases short of what would be expected for the correct clone.

The overall integrity of KTAla7 is thus verified by the fact that it shows the same size patterns in the analysis as the control DNA, KTCK3F. The only difference between KTCK3F and KTAla7 should be in two bases in the codon coding for Cys_{283} in KTCK3F and Ala_{283} in KTAla7. These base changes do not involve a new restriction site and therefore should not produce any differences between KTCK3F and KTAla7 in the restriction analysis. We did not sequence the entire clones of either KTCK3F or KTAla7 but confirmed their correctness further by using them to express fullength CKs.

Figure 26: Restriction Analysis of KTAla7



Length markers are PBR322 DNA cut with HpaII or λ DNA cut with EcoR1/HindIII. Lane 1) PBR322, Hpa II Markers; Lanes 2, 5, 8, 12) Blank; Lane 3) KTCK3F, PvuII/BgI-II cut; Lane 4) KTAla7, PvuII/BgII-I cut; Lane 6) KTCK3F, PvuII/SphI cut; Lane 7) KTAla7, PvuII/SphI cut; Lane 9) KTCK3F, Pst1 cut; Lane 10) KTAla5, Pst1 cut; Lane 11) KTAla7, Pst1 cut; Lane 12) Blank; Lane 13) λ, EcoR1/HindIII Markers; Lane 14) KTCK3F, Supercoiled; Lane 15) KTAla7, Supercoiled.

Table 6: Inserts Found from Restriction Analysis of KTCK3F and KTAla7

Restriction Endonuclease cuts	Sizes of Inserts found on Gel (BP) KTCK3F	KTAla7	KTAla5
PvuII/BglII	685	685	N.A.
Pst1	514	514	~ 400
SphI/PvuII	550	550	N.A.

With the restriction analyses of KTCK3F and KTAla7, we were thus essentially finished with the phase of the work involving the DNA manipulations except for later attempts to modify the clones to improve expression. We moved on to the next phase described below: expression and analysis of the proteins translated from these clones.

3. Expression of KTCK3F and KTAla7.

Expression of clones KTCK3F and KTAla7 was performed in E. coli strain JA221. We transformed these CK clones in JA221 and selected for colonies containing the plasmid on LB plates containing 30 ug/ml ampicillin. This strain contains an episome and a constitutive gene for the lac repressor, which is necessary for repression of the lac operon as required for the regulation of the trc promoter. When we attempted to express either wild type or Ala mutant CK in this system, however, we found that no observable repression had occurred. In our preliminary small scale cultures, the system apparently

produced similar amounts of CK both in the presence and absence of IPTG. We also had problems obtaining expression in large-scale cultures, as will be discussed later in this section.

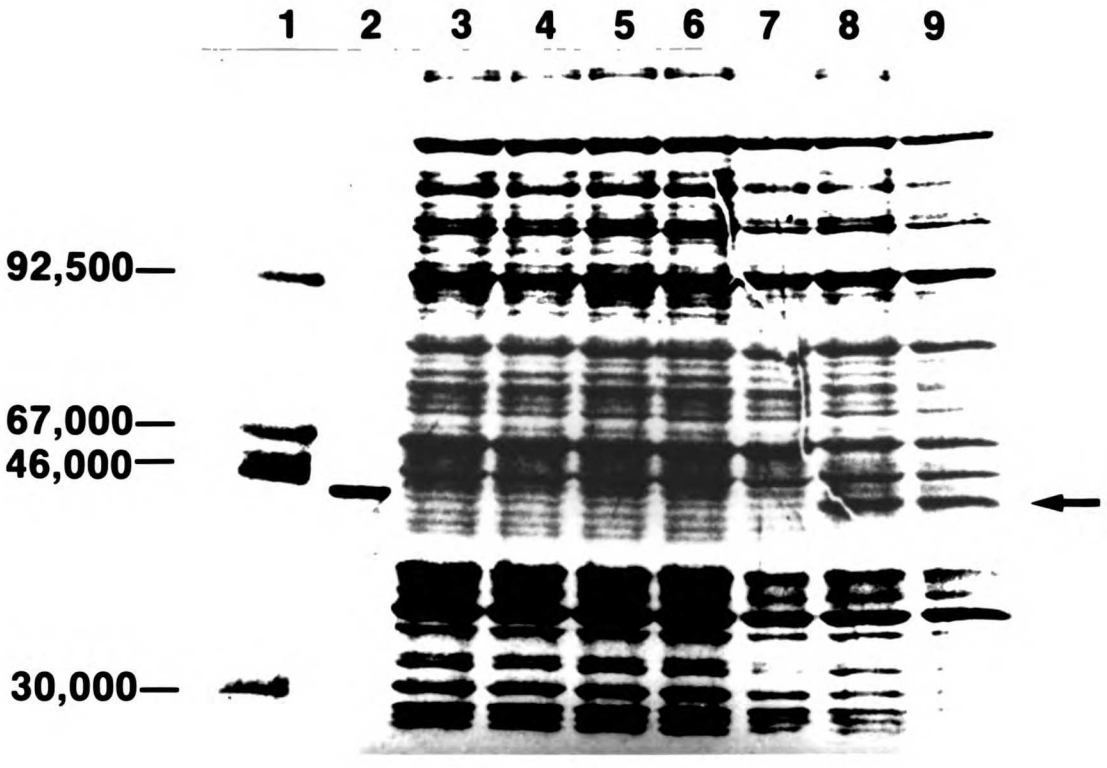
Dr. West tried to correct this regulation problem by making a number of alternative constructions of KTCK3F designed to decrease the number of copies of the expression plasmid in the cell, decrease the strength of the promoter, or increase the amount of lac repressor present in the cell. None of these constructions resulted in an inducible system (130).

An explanation for this phenomenon was not found and speculation about the many possible sources of the problem is beyond the scope of this study. Attempts by Dr. West to correct the problem by using other strains of bacteria were also unfruitful. These strains yielded a variety of results, none of them useful.

We decided, therefore, to complete our experiments on the available system in the JA221 strain. The total protein from JA221 cells transformed with KT52SphBgl (the complete expression vector missing only the CK insert) and other non-CK-producing controls as well as KTCK3F is shown in Figure 27. This figure is a photograph of a 10% SDS-polyacrylamide gel showing the total protein obtained from digestion of whole cells in Laemmli buffer (53). The gel was done by Dr. West (130) and represents one of our first attempts to express CK. The title of the Figure is incorrect. There is no expression of KTAla7 represented on this gel. The title should read "Figure 27: SDS-PAGE of Total Bacterial Proteins Expressed from KTCK3F and non-CK-producing Controls."

Figure 27 shows that JA221 cells transformed with KTCK3F produces a new

Figure 27: SDS-PAGE of Total Bacterial Proteins Expressed from KTCK3F and KTAla7



Lane 1) 5μg Molecular weight markers (Pharmacia) (numbered on the left of the gel); Lane 2) 2 μg rabbit muscle CK (Sigma); Lanes 3-7) Total protein samples from controls that did not produce CK; Lanes 8,9) Total protein, KTCK3F. The arrow at the right designates the CK bands in lanes 8, 9.

protein (lanes 8 and 9) not found in the controls (lanes 3-7). Reference to the molecular weight markers (lane 1) suggests this new protein has a subunit molecular weight slightly smaller than 46,000 kD. The molecular weight of Torpedo californica CK monomers is 42,941 kD. This is very close to the size of the rabbit muscle CK standard (43,119 kD) in lane 2 and the expressed Torpedo CKs have a similar Rf value with this standard. The proteins that correlate with these CK bands can be purified and assayed to give detectable CK specific activity for the wild-type and lower specific activity for the mutant. Nearly normal specific activities were eventually obtained expressing CK from these clones. This aspect of the study will be discussed more fully in Chapter 4.

While this expression system does produce appreciable amounts of wild-type or mutant CK from small volume overnight cultures, there are two major problems with the system that have severely limited its usefulness both for the present study and for future applications. The first problem concerns the lack of inducibility in the system. This problem has caused considerable difficulty through the major course of this thesis work. It was found, early on, that when we tried to grow large-scale cultures expressing gram-quantities of total protein and milligram- quantities of CK, the CK bands could no longer be detected on our gels, nor could detectable amounts be purified from these cultures. When we harvested the plasmid DNA from such cultures, we found that the plasmid was present in normal amounts, and that the plasmids could be analyzed to give identical results to those shown in Figure 26 for both KTCK3F and KTAla7.

From these results and the fact that we did get CK expression from small-scale or short-time large scale cultures, we concluded that the bacteria found the expressed

protein to be harmful. The reasons for this are currently unknown. The fact that JA221 cells transformed with KTCK3F and KTAla7 grew more slowly than those transformed with KT52SphBgl also tends to support this conclusion. We already knew, from our first attempts at expression, that the system is not regulated by IPTG. An explanation that is consistent with all of these findings is that chance mutations in the expression vectors may be occurring, probably in the promoter regions. These mutations in some individual plasmids could produce species carrying the plasmid (thus retaining ampicillin resistance), but not expressing CK protein. Since these individuals grow faster than the CK-expressing individuals, large-scale cultures are quickly overgrown by these "healthier" competitors. This phenomenon, whatever its causes, was complicated by the fact that our system was not regulated and was therefore "turned on" all of the time. The end result was that we could not grow cultures in quantity and therefore could not obtain expressed protein in appreciable amounts. This problem led to grave consequences for the experiments described in Chapter 4. Because we could not obtain more than ug amounts of expressed CK from any one preparation, we could perform only a limited number of very small- scale experiments to optimize purification and activation of expressed protein.

Due to this problem, the expression procedure we were obliged to use was to freshly transform up to 100 20-cm plates with KTCK3F or KTAla7, grow them overnight, and scrape the colonies off the plates for purification of the CK. This procedure worked (as opposed to large-scale liquid cultures) because evidently only some of the colonies from any one plate would have lost the ability to express CK. The awkwardness of this

procedure did, however, severely limit the amounts of protein that could be harvested.

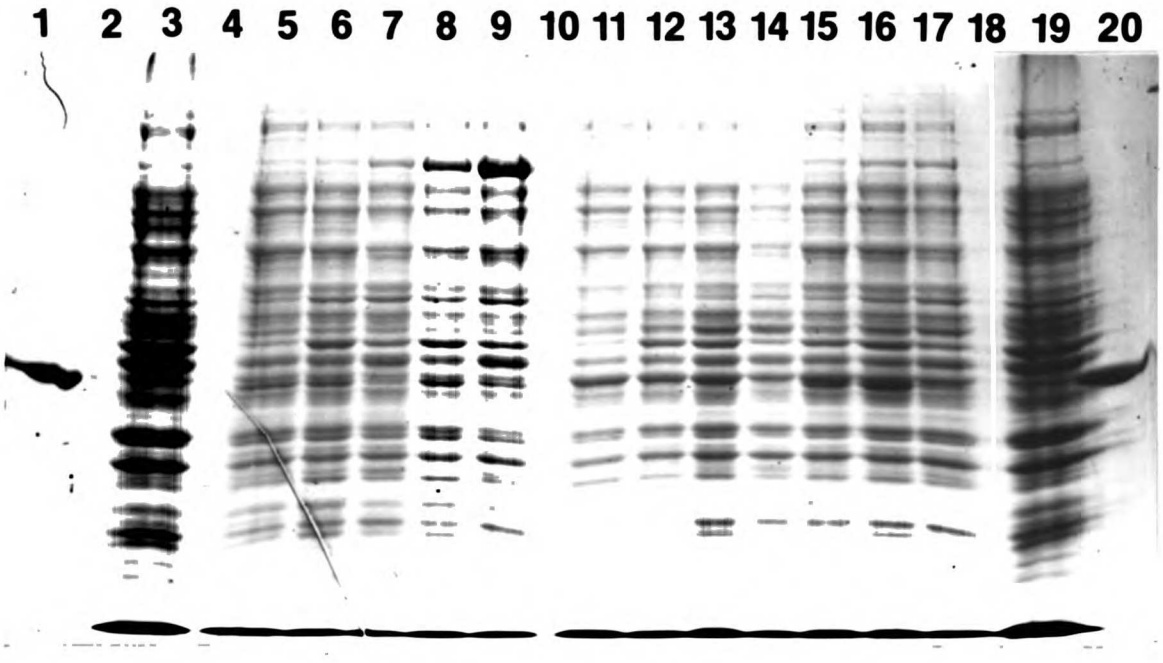
This problem was partially solved near the end of this thesis work when Dr. West obtained a new culture of the JA221 strain, which we designated JA221-I. When we tried to express CK in this new strain, we found it to be at least partially inducible by IPTG as shown in Figure 28.

This figure shows that the new strain, JA221-I, is at least partially regulated by IPTG, in contrast to the strain we had been using previously, JA221. Both strains were transformed simultaneously with KTCK3F and plated as described earlier. One colony from each plate was started in 10 mL LB medium containing 30 ug/ ml ampicillin and grown to an $O.D._{650}$ of 0.6. Five mL of each culture was then transferred to a new tube and IPTG was added to these new tubes to give a concentration in each tube of 1 mM IPTG. One mL was removed from each tube as the "0 hours" timepoint and these cells were pelleted and resuspended as described earlier.

These 100 uL samples were frozen until needed for the gel. Additionally, one mL aliquots were removed from each tube, the $O.D._{650}$ was read, and the cells were processed as above at 2 hours, 4 hours and >10 hours from the zero timepoint. Because the cultures grew at different rates, the volumes of sample applied to the gel varied. The sample amounts used on the gel were calculated from the absorbance readings to allow us to apply equivalent amounts of protein in each lane.

Lanes 5-9 show the results from transformation of KTCK3F into the old strain, JA221, in the presence and absence of IPTG. All of these lanes show a CK protein band with approximately the same Rf value as that for the rabbit muscle CK standard. In

Figure 28: SDS-PAGE of Total Bacterial Proteins Expressed in JA221 and JA221-I



All lanes except 1 and 20 represent fractionation of total bacterial proteins from JA221 and JA221-I cells transformed with either KT52SphBgI or KTCK3F and treated with and without IPTG. Time points are measured from the time the 10 mL cultures were split in half and IPTG added to 50% of them at T = 0 hours. Cells were removed and processed as described below in equivalent sample volumes of 100 μL. All aliquots removed for the gel were brought to 15-20 μL total volume, diluted with 15 μL 2x Laemmli buffer, and run on the gel (see Protocol 17). Lanes 1, 20) 4.5 μg rabbit muscle CK standard (Sigma); Lanes 2, 4, 10, 18) Blank; Lanes 3, 19) 10 μL KT52SphBgI control; Lanes 5-7 are KTCK3F in JA221/- IPTG: Lane 5) 0 hours, 20 μL; Lane 6) 4 hours, 5 μL; Lane 7) > 10 hours, 3 μL; Lanes 8-9 are KTCK3F in JA221/ + IPTG: Lane 8) 4 hours, 6 μL; Lane 9) > 10 hours, 5 μL; Lanes 11-14 are KTCK3F in JA221-I/ - IPTG: Lane 11) 0 hours, 18 μL; Lane 12) 2 hours, 12 μL; Lane 13) 4 hours, 4.5 μL; Lane 14) > 10 hours, 3 μL; Lanes 15-17 are KTCK3F in JA221-I/ + IPTG: Lane 15) 2 hours, 10 μL; Lane 16) 4 hours, 10 μL; Lane 17) > 10 hours, 3 μL.

contrast, the control protein fractionation from transformation of KT52SphBgl (lane 3) shows no CK protein band. The control sample was transformed as KTCK3F and the samples on the gel were from cells treated with IPTG and removed at the 4 hour timepoint. There is no appreciable difference between the CK bands in lanes 5-7 (not treated with IPTG) and the CK bands in lanes 8-9 (treated with IPTG). Clearly, we are not regulating the JA221 cells with IPTG.

On the other hand, the new JA221-I strain shows definite regulation by IPTG. Lanes 11-14 show no CK bands and represent the samples not treated with IPTG. Lanes 15-17 represent samples that were treated with IPTG. All three of these lanes show a CK band with the maximum amount of CK being produced in the culture (lane 16) whose growth was stopped 4 hours after the start of induction by IPTG. Although the data is not shown in Figure 28, we found that JA221-I, IPTG-treated cultures progressively lost CK expression when grown for more than 6 hours after IPTG induction. Large-scale cultures completely lost CK expression if grown for more than 16 hours after induction. This was another indication that our "mutation" theory was correct. Cells expressing CK are less healthy than those not expressing CK. The healthier cells overgrow the CK-producing cells in these cultures and if the cultures are grown to saturation, they show no detectable CK expression. We do not know why the cells are stressed by the expressed CK.

These results indicate that use of the new strain, JA221-I, has considerably improved our ability to regulate CK expression. Later experiments showed, however, that the regulation was imperfect and that cells grown to saturation before the addition of IPTG showed no CK expression. We rationalized this result by concluding that the promoter

was still slightly "leaky" in this new strain, but much less so than in the old JA221 strain. This result meant that we still could not grow large, saturated cultures of cells transformed with our CK-expressing plasmids but had to be content with growing large-scale cultures to early log phase, and then inducing them. Thus, we still cannot obtain as much protein from each preparation as has been possible for other systems using the same promoter (122). But the new JA221-I strain is still an immense improvement over JA221. This new strain should be the same as our original JA221 strain. Evidently, our original strain had mutated early in our experiments in some way that adversely affected its ability to be regulated by the lac repressor. Thus, it was only after we started to use the JA221-I strain that we were able to express CK in large-scale cultures. The details of large-scale CK expression in this system are discussed in Chapter 4.

The second problem we had with this expression system is that the expressed CK forms an insoluble aggregate in the cell. The aggregate must be denatured in 6M guanidine-HCl and refolded before any protein activity can be discerned. The best solutions we have found to this problem will also be discussed in Chapter 4.

Conclusion

We were able to express CK wild-type and mutant proteins in E. coli using a strong bacterial promoter engineered to produce a native cytoplasmic translation product. In order to do this, we had to re-engineer the DNA's of both the expression vector and the CK-encoding insert to be compatible and so that the insert could be placed with correct spacing from crucial promoter elements. This was done by site-directed mutagenesis.

The mutant sites were verified directly by sequencing of the single-stranded DNA templates of the mutant strands. The correct mutants were converted to double-stranded form and inserted into the correct positions in the respective wild-type DNAs. The insertion of these mutant strands into their correct positions and the final constructions were verified indirectly by restriction analyses and by expression of full-length active proteins.

Expression of active protein was complicated both by the low yields obtained from the original JA221 strain of bacteria and by the fact that the CK was expressed as an insoluble aggregate. The yield problem was largely solved by use of the new JA221 strain, named by us as JA221-I. The aggregation problem was solved by denaturing the aggregate followed by refolding of the denatured proteins. This aspect of the thesis problem is discussed in Chapter 4.

Chapter 4: Purification and Refolding of Expressed Creatine Kinases

Introduction

As was stated in Chapter 3, the CKs obtained from our expression system formed an insoluble aggregate in strain JA221-I E coli. Because the methods which have been developed to address this common problem (128) in the expression of eukaryotic proteins in prokaryotic cells are relatively new, a brief review of them follows. Our attempts to solve this problem by purification of the aggregate, followed by denaturation and refolding of the proteins are discussed in the "Results and Discussion" section of this Chapter.

While there have been notable successes in expressing both cytoplasmic and secreted proteins from eukaryotic sources in E. coli (122, 132, 133) many workers have experienced an "aggregation problem," just as we did (134-136). (See Harris (137) and Wetzel and Goeddel (138) for reviews of expression of eukaryotic proteins in prokaryotes.) The general causes of the problem remain unexplained, but a number of investigators have formulated possible explanations (128, 135, 139) such as too high a concentration of expressed protein for the bacterial cell to maintain solubility, or the differences in reducing environment within bacterial and eukaryotic cells, or the differences in compartmentalization and processing of proteins between the two cell types, or the kinetic differences in protein translation between the two types of cells. A host of other differences which may contribute to aggregation problems are either known or suspected between prokaryotic and eukaryotic systems.

Despite our lack of understanding of the causes of this problem,there are a growing number of reports of experimental purification and reactivation methods

successfully used to yield pure, active, soluble product from the initial aggregate. There is mounting evidence in this literature that each aggregate is different. These differences may depend on the composition of the protein, the type of expression vector used, the concentration of the protein expressed, and a variety of other factors not yet understood. Each aggregate thus appears to be susceptible in a different way to reactivation by a variety of methods. But while this area of protein chemistry remains in its infancy, especially in our understanding of the general principles involved, there is an increasing body of literature detailing the types of approaches which have been successful in working with these proteins (139-144). These approaches can be assigned to three general areas, as summarized below.

First, the aggregate must be separated physically from the soluble proteins and other soluble molecules in the bacterial cell. This is done by lysing the cell by a number of methods (145, 146) followed by treatment of the resulting solution with a variety of detergents, reducing agents, and extraction buffers (135, 142, 144) to remove as much soluble material as possible. In some cases, the expressed proteins have been themselves solubilized after such treatment (144).

In other cases, the expressed proteins remain aggregated in particles which are soluble in only very strong denaturing agents. In such cases, the second step of the workup is to treat the aggregate with such a strong denaturing agent to solubilize it. For proteins that can be refolded, such as CK, the methods of choice are to use either guanidine-HCl or urea at a whatever concentration works (11, 140, 144).

It has been found that these aggregates are often composed of a number of

proteins in addition to the protein of interest and techniques have been developed to purify the expressed proteins while in the unfolded state, before they are refolded (140, 144). This step can be important because success in refolding of proteins in vitro is very sensitive to the degree of purity of the denatured protein solution (142, 144, 147, 148).

Two methods have been used recently to purify proteins in their denatured state. The first method involves purification of the unfolded material in 6M guanidine-HCl over a sizing column, usually Sephadex, followed by collection of the peak of interest for refolding. The second method is similar, if more difficult, and requires exchange of the denaturant to urea followed by FPLC with an ion exchange column and again, collection of the purified denatured protein prior to refolding (140, 144).

Third, the protein is refolded by either dialysis or dilution (140, 144). A number of workers have found, when working with aggregates derived from expression systems such as ours, that refolding must be done at very low protein concentrations (< 20 mg/mL) (130). Following refolding, the proteins can be further purified by conventional methods.

The aggregated CK purified from our cultures proved to be highly insoluble under all conditions we tried except for treatment with high concentrations of denaturants. This situation made the further purification and reactivation experiments very difficult. Because of the complexity of these experiments, I have divided the "Results and Discussion" section of this Chapter into 4 parts. The first 3 cover the nature and purification of the aggregate, unfolding the purified aggregate and further purification of the protein in its denatured state, and refolding and characterization of the refolded protein. The overall scheme which produced the most active CK is shown in Figure 31, at the end of the

"Materials and Methods" section. The last part of the "Results and Discussion" section is a comparison of the refolded wild type and mutant CKs. It includes our conclusions about the essentiality of Cys_{283}.

Materials and Methods

1. Reagents, Equipment, and Instrumentation

Buffer materials and dithithreitol, beta-mercaptoethanol, isopropyl thiolgalactoside (IPTG), n-octyl glucoside, Triton X-100, deoxycholate, urea, Sephadex g-200, lysozyme, and EDTA were obtained from Sigma Biochemicals. Ultra-pure guanidine-HCl was obtained from Schwartz-Mann and was not recrystallized before use. Electrophoresis reagents and consumables were obtained from Bio-Rad. Protein molecular weight markers, isoeletric focusing standards and Blue Sepharose CL-6B were from LKB. Enzymes and reagents used in protein concentration and activity assays were from Sigma or Bio-Rad. All other chemicals were reagent grade from Van Waters and Rodgers, Fisher, or American Scientific Products.

Centrifugation was performed in a Sorvall refrigerated centrifuge using bottles and tubes from either Nalgene or Corex. Microcentrifugation was done in a Beckman microcentrifuge (15,000 x g) using 1 mL and 2 mL disposable tubes from Eppendorf. Other glassware was disposable or washed in an acid bath before use. Sonication was performed with a Sonicator Model W-225R sonicator using normal or micro-tips. Concentration of refolded proteins was done in Amicon concentrators with YM-30 membranes or in disposable micro-concentrators (concentration by centrifugation) from Centricon. Gel-scanning was done on a Hoefer GS300 gel scanner equipped with a Shimatzu integrator. Instrumentation for the luciferase assay was a luminescence biometer (Dupont).

2. General methods

Wild-type and mutant CKs were kept scrupulously separate at every step of the work-up, using different columns and column resins, different membranes for concentration, disposable or acid-washed glassware and centrifuge tubes. FPLC was performed only on the wild-type CK.

Lysing and sonication of the cell pellets was done on ice. Centrifugation in the Sorvall was done at 4 °C. Denaturation of the aggregate and initial refolding was done at room temperature. Refolding was continued in the cold room and all subsequent work on the refolded protein was done at 4 °C. The protein was never frozen except for those aliquots saved expressly for analytical electrophoresis.

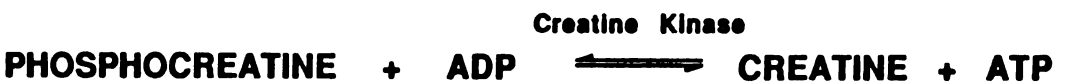
The expressed proteins were analyzed by both native and denaturing polyacrylamide gel electrophoresis and by isoelectric focusing. Protein assays were by the Lowry method (149), Bradford method (150), or by estimation of protein concentration from densitometry. CK activity assays were performed by two methods, a firefly luciferase assay and an NADH-linked assay.

3. Creatine Kinase Activity Assays

A. Firefly luciferase assay:

Figure 29 shows the reaction pathway for this assay. The two key compounds, luciferin and oxy-luciferin, are designated as "L" and "oxy-L" respectively.

Figure 29: Firefly Luciferase Assay for CK Activity



Originally, we used this assay for CK activity because of its greater sensitivity over the NADH-linked assay. The limits of detection for the firefly luciferase assay are 10-20 ng for a solution containing only CK standard and approximately 30 ng for a solution containing expressed CK and other proteins as the crude refolded sample. In contrast, the limits of detection for the NADH-linked assay are 500 ng for CK alone and approximately 1 ug of CK in the mixture. These units are based on CK standards with a specific activity of approximately 20 U/mg of protein as determined by the NADH-linked assay.

In our early refolding experiments, the yield and specific activity of refolded wild-type and mutant CK was so low that we could not detect activity using the NADH-linked assay. Later, as both yields and refolding techniques improved, we used the NADH-linked assay exclusively because it is much easier to perform than the luciferase assay.

Another advantage of the luciferase assay is that it is clearly phosphocreatine dependent. Since the early activity assays were done on samples containing a number of refolded proteins, it was important that we knew we were measuring CK activity only. In contrast, while the NADH-linked assay was creatine-dependent, the baseline response due to other NADH-utilizing enzymes in our crude samples was often so high as to mask CK activity, or least render it difficult to quantitate. This problem ceased in later experiments in which the proportion of CK was much higher in our samples.

B. NADH-linked Assay for CK Activity (152):

Figure 30 shows the reaction pathway for this assay. As can be seen from the figure, the amount of ATP transformed by CK can be directly correlated with the molar loss of NADH, followed at an absorbance of 340 nm.

As described earlier, this assay is easier to perform than the firefly luciferase assay, but it is less sensitive. As we were able to increase the proportion represented by CK in our refolded samples (and as improved activation techniques increased our specific activity), it became the method of choice for our CK assays.

4. Harvest, Purification, and Activation of Wild-type and Mutant CKs

Figure 31 is the final scheme which gave the best specific activities for wild-type and mutant CKs. While it can undoubtedly be improved, especially by the addition of steps to further purify CK in the unfolded state, this scheme yields wild-type CK of nearly normal activity. The development of each aspect of this scheme is discussed in the "Results and Discussion" section.

The scheme begins with freshly transformed JA221-I cells containing KTCK3F or KTAla7 plasmids with the KT52SphBgl plasmid transformation plate used as a control. Control samples were typically grown in 50 ml cultures (rather than the 1 liter used in this scheme) with 1 colony used as infectant rather than the cells from an entire plate. The rest of the control work-up is scaled down appropriately from volumes shown in Figure 31.

Figure 30: NADH-linked Assay for CK Activity

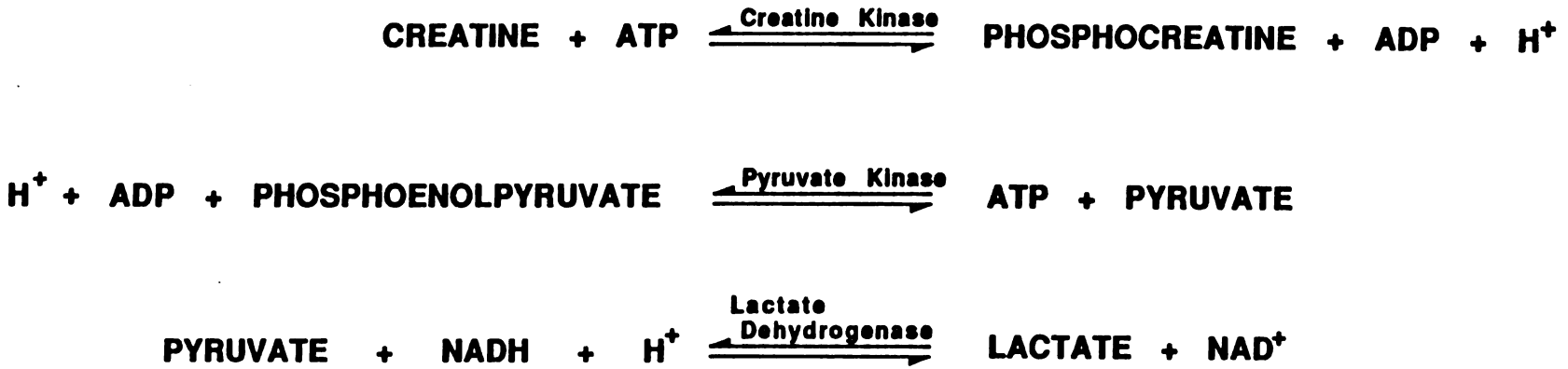
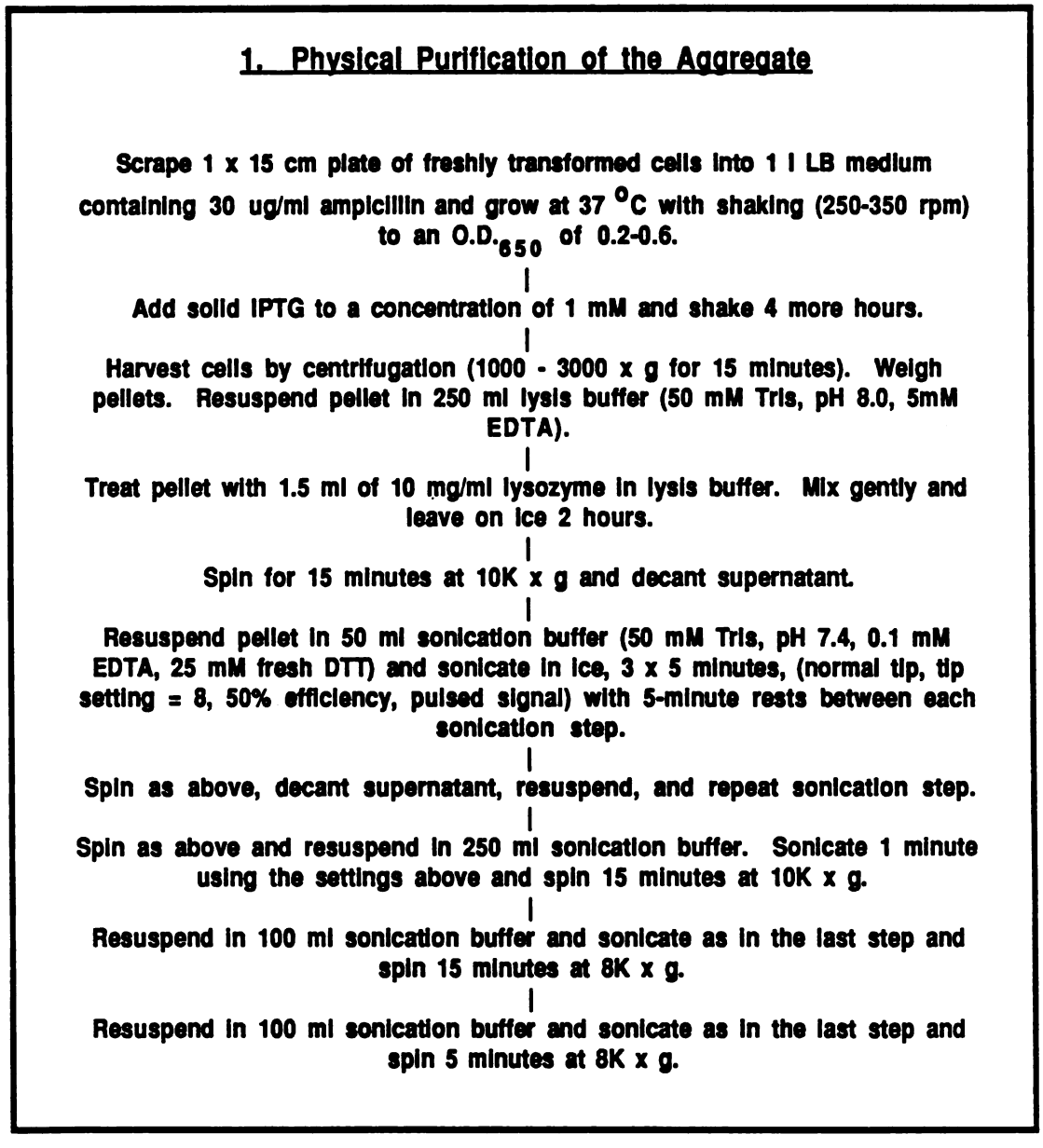


Figure 31: Harvest, Purification, and Activation of Expressed Creatine Kinases



2. Unfolding the Crude Pellet

Resuspend the pellet in sonication buffer at ~1 mg/ml CK (as estimated from PAGE-SDS gels). Add beta-Mercaptoethanol to a concentration of 1% (V/V) and solid n-octyl glucoside to a concentration of 2.5% (W/V). Sonicate as before, 1 minute. Incubate 12-24 hours with slow shaking (<100 rpm) at 37 °C.

|

Spin at 12K x g for 20 minutes and resuspend in sonication buffer in the same volume as above. Repeat sonication and centrifugation as above to wash pellet and remove the majority of the detergent.

|

Resuspend pellet in the volume of sonication buffer needed to bring the estimated CK concentration to 2 mg/ml. Add solid guanidine-HCl, beta-mercaptoethanol, and buffer so that the final guandine-HCl, beta-mercaptoethanol, and CK concentrations are 6 M, 1%, and 1 mg/ml respectively. Leave at room temperature for 2 hours.
The final volume should be such as to give an approximate CK concentration of 17 ug/mL upon a 60-fold dilution with sonication buffer.

3. Refolding the Denatured Creatine Kinase

Refold by dilution by adding 60 x the volume of the unfolded protein solution with refolding buffer (50 mM Tris, pH 8.0, 10 mM fresh DTT). Filter the solution through a 0.2 um Nalgene disposable filter and concentrate about 100-fold in an Amicon concentrator using a YM-30 membrane.

|

Assay for activity within 36 hours or dialyze against buffer containing fresh DTT before assaying.

Results and Discussion

All of the steps in the purification and reactivation scheme were the same for both the wild-type and Ala mutant CKs and are discussed in Parts 1-3 of this section. The only differences we found between these two proteins were in their specific activities and we did not perform the experiments which could help us understand why the wild-type and mutant differed in activity. Comparison of their specific activities is treated in Part 4 of this section. After we had determined that both CKs behaved identically in the initial experiments in the work-up of expressed proteins, further experiments designed to optimize conditions or reveal something of the nature of the aggregated protein were done only on wild-type CK. Therefore, the first four sections of this discussion will only consider the wild-type expressed CK. All of the manipulations (shown in Figure 31) were performed on both wild-type and mutant protein. Results from these experiments were identical for both wild-type and mutant proteins with the exception of the specific activity found for each type.

1. Physical Purification of the Aggregate

A. Cell Growth and Yield of CK

In order to obtain CK expression in large-scale cultures, I had to use freshly transformed cells to inoculate the liquid media. Cells from plates even 24 hours old often had lost the ability to express CK. Additionally, as discussed in Chapter 3, large-scale cultures of cells containing KTCK3F transformed into JA221-I were difficult to grow to

saturation without loss of CK expression. While I was unable to determine the exact $O.D._{650}$ to which these cultures should be grown before addition of IPTG, it was possible to determine a general approach which produced significant if variable amounts of CK in large-scale cultures. Table 7 shows the $O.D._{650}$ values at which IPTG was added along with the wet weight of harvested cells and estimated amounts of CK obtained from several large-scale cultures. The amounts of CK obtained from each culture were estimated from SDS-PAGE gels on the crude purified pellets for each preparation. Since the purification steps which produced the crude pellet involved considerable loss of the total CK (determined by SDS-PAGE analysis of pellets which could be spun out of each supernatant removed during the purification of the pellet), this estimate reflects only the amount of CK recovered as crude pellet, not the total fraction of CK present in the "wet weights of cell pellets" shown in Table 7.

While the data shown in Table 7 is insufficient to allow us to pinpoint the best growth conditions to give optimum amounts of CK in the crude pellet, it does suggest that cells grown only to early log phase before addition of IPTG give the best results. Even though this produces fewer whole cells/Liter, the amount of CK in the crude pellet is greater in these preparations. Purification of the control (cells transformed with KT52SphBgl containing no CK insert) to the crude pellet stage or of CK-containing preparations which have no detectable CK in the crude pellet (Preparations 14 and 17a in Table 7) show that there is a considerable amount of insoluble protein in all of them. By far the greatest recoveries of refolded CK were obtained from preparations showing the

highest proportion of CK in the crude pellets. This correlated in every case with low cell densities at the time of induction by IPTG, even though it represented a considerable loss in total protein (over what could be obtained from saturated cultures).

Table 7: Yield of CK in Large-Scale Cultures of JA221-I

CK Preparation #	$O.D._{650}$ at IPTG Addition	Wet Weight of Cells/L (gms)	Culture Size (L)	Amount of CK in Crude Pellet/L of Cells Processed (mg)
14	1.0	6.30	3.0	N.D.[a]
15	1.7	5.97	6.9	0.9
17a	Saturated	3.68	1.0	N.D.
$17b_{300}$	~.6	6.00	0.3	2.0
$17b_{700}$	~.6	4.30	0.7	0.6
18	0.2	2.40	9.0	2-3

[a]N.D. = Not determined (too low to be detected on an SDS-PAGE gel)

B. Cell Lysis and Sonication

Small-scale cultures treated with and without lysozyme prior to sonication but treated equivalently throughout the rest of the work-up showed similar specific activities in the refolded products. The cells which were not treated with lysozyme, however, gave crude pellets which were slightly less enriched in CK. These cells also required more sonication to give a similar-looking crude pellet as determined by fractionation on

SDS-PAGE (data not shown). I concluded that addition of lysozyme to break down bacterial cell walls prior to sonication did no harm to the CK final activity and helped considerably in making the sonication steps reproducible.

In addition to showing differences in the crude pellets, cells treated with lysozyme also showed differences in the supernatants recovered following this step. For cells suspended in lysis buffer and left in ice for two hours prior to centrifugation, the supernatants recovered were clear and contained little or no protein. These supernatants had similar viscosities to that of the buffer. For cells suspended in lysis buffer with 1.5 mL of lysozyme (10 mg/mL) added per liter's worth of culture and left in ice for two hours prior to centrifugation, the supernatants showed detectable protein bands on SDS-PAGE. In addition, the supernatants were often very viscous, indicating that considerable amounts of intracellular material, especially DNA and RNA, had been released by the lysozyme treatment. The pellets remaining were often also very viscous and difficult to resuspend in sonication buffer, again indicating large amounts of released intracellular material. The pellets obtained from centrifugation following the first sonication step were uniformly whiter and smaller than those not treated with lysozyme.

The resuspension volumes used for sonication gave the most reproducible results (see Figure 31 for these volumes). Sonication was done in ice and care was taken to regulate the length of sonication treatments so that the suspensions would not heat up. Two sonication cycles were required to give crude pellets most highly enriched in CK. Supernatants recovered from centrifugation after each sonication step were turbid, yellow, and contained large amounts of soluble proteins as seen on SDS-PAGE. These

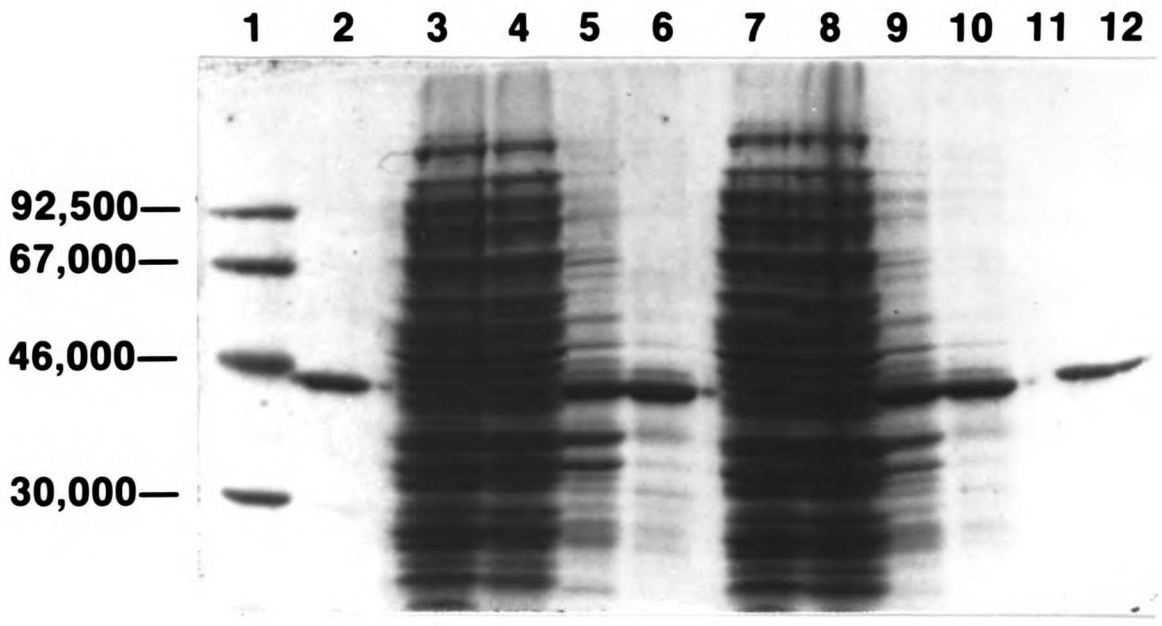
supernatants, when filtered through 0.2 um filters, showed no CK. Nor did these filtered supernatants show detectable CK activity. We found no soluble CK in our preparations.

C. Purification of the Aggregate by Centrifugation

The final steps in the purification of the aggregate were centrifugations done at decreasing speeds and for decreasing lengths of time. The choice of these parameters was made empirically and those finalized in Figure 31 gave the largest amount of CK for the least number of manipulations. I found that increasing the number of centrifugation steps did not significantly improve the purity or proportion of CK in the final crude pellet, but did cause considerable loss in total pellet material recovered. Purification of the CK to the crude pellet used for unfolding and reactivation of the CK is shown in Figure 32. The Figure is an SDS-PAGE gel, (10% polyacrylamide) representing the CK at each major step in the purification and refolding procedures used for both expressed CKs.

The molecular weight markers are phosphorylase b (92,500 kD), bovine serum albumin (67,000 kD), ovalbumin (46,000), and carbonic anhydrase (30,000). The CK protein is barely evident in lanes 4 (wild-type CK) and 8 (Ala mutant CK) compared to the non-CK-containing control (lanes 3, 7). The gel also shows that the CK aggregates purified by the steps described above contain significant amounts of contaminating proteins in the final crude pellets. These pellets have been very difficult to purify further and all of our reactivation experiments have been done on pellets such as these. Lanes 6 and 10 represent the refolded proteins refolded from the crude pellets. These refolded proteins are soluble and show CK activity.

Figure 32: SDS-PAGE of KTCK3F and KTAla7: Purification and Refolding



Lane 1) 5 μg Molecular weight markers (Pharmacia) (numbered on the left of the gel; Lanes 2, 12) 2 μg rabbit muscle CK (Sigma); Lanes 3, 7) Total protein, KT52SphBgl; Lane 4) Total protein, KTCK3F; Lane 5) Crude pellet, KTCK3F; Lane 6) Refolded pellet, KTCK3F; Lane 8) Total protein, KTAla7; Lane 9) Crude pellet, KTAla7; Lane 10) Refolded pellet, KTAla7; Lane 11) Blank.

2. Unfolding the Crude Pellet

Even though I experimented with many more sonication and centrifugation steps in an attempt at further purification of the crude pellet by physical separation, the crude pellets seen in Figure 32 were the purest I could achieve using these techniques. It is obvious that while the CK is greatly enriched and comprises the primary component of the crude pellet, it is still very "dirty" by usual standards of protein purification. I also attempted to further purify this pellet by extraction or solubilization prior to denaturation in Gdn-HCl. While some of these attempts were partially successful in that they yielded a refolded product with higher specific activity than did the untreated crude pellet, none of them produced soluble protein except extraction with SDS. Part of the problem in designing these experiments was that I did not understand the nature of the aggregate or what physical parameters were involved in creating and maintaining the CK in aggregated form rather than the soluble form in which it exists naturally. I still do not have answers to these questions, but the next few paragraphs contain some observations I did make about the nature of the crude pellet and my attempts to purify it further prior to the Gdn-HCl denaturation step.

A. Nature of the Crude Pellet.

The crude pellet is insoluble in the buffers used in the work-up even if left in these buffers for prolonged periods of time. Fortunately, the crude pellet could be refolded to give active CK even after sitting in sonication buffer for several months in the refrigerator. Soluble rabbit muscle CK left in such buffers (these buffers contained at least 5 mM DTT)

lost activity over such periods of time. The soluble rabbit muscle CK was also decomposed after such prolonged storage, producing multiple bands on SDS gels. Evidently, the CK aggregate in the crude pellet is protected in some way from protein breakdown processes which affect soluble CK solutions treated similarly. Other workers have also noted this "protective" aspect of aggregation (144).

It is interesting to note that the control cultures (transformed with KT52SphBgl containing no CK insert) also produced an insoluble pellet when purified in the same manner as the CK-containing aggregates. SDS-PAGE on these non-CK containing pellets look very similar to those containing CK except, of course, for the CK band. Clearly the CK aggregate is made up of many insoluble materials, much of the protein portion of which is present whether or not CK is a component. I do not know to what extent this insoluble crude pellet is an artifact of the purification scheme I used. The obvious question that arises from these observations is "How do these insoluble proteins differ from soluble ones?" Some of the answer is straightforward. Obviously, there are a number of cellular proteins, such as many structural and membrane proteins, which are normally insoluble in polar aqueous buffers. These undoubtedly form a large component of the crude pellets in both the CK-containing and control preparations.

The answer to the question of why the expressed CK is part of this insoluble fraction has not been answered by this thesis work. Nor have I answered the question of how the CK aggregate differs from soluble CK. In order to produce active CK, the immediate question had to be narrowed and became, "Can the insoluble CK aggregate be separated from the other insoluble proteins in the crude pellet?" Our experiments in

purifying the aggregate to the crude pellet stage suggests that further sonication or centrifugation will not provide this separation. It is possible, however, to solubilize membrane proteins, for instance, with the incorporation of detergents or other nonpolar extractants in the purification buffers. We tried a number of these in an attempt to either solubilize the CK or solubilize and remove contaminating proteins from the crude pellet.

B. Purification of the Crude Pellet by Extraction

The first attempt I made to solubilize the crude pellet was to extract it with sonication buffer containing 20% glycerol, 1% Triton X-100, and 2% deoxycholate. These detergents are used routinely in the purification of membrane proteins and I decided to use this mixture after consultation with Dr. Elizabeth Komives at that time in the laboratory of Dr. Paul Ortiz de Montellano (153). The crude pellets were extracted both in ice and at room temperature for periods from 1/2 hour to several days (when the pellet was extracted for more than a few hours, it was stored in the refrigerator). This extraction procedure did not affect the crude pellet, at least insofar as no new soluble proteins could be seen on SDS-PAGE when the detergent supernatants were concentrated and applied to a gel. Specific activities were no better for CKs extracted in this manner than for CKs refolded from the crude pellet directly. While this treatment may well have removed other non-protein nonpolar materials not evident on the gel, its lack of effect in improving the specific activity did not justify continuing the procedure.

I also extracted the pellet with another detergent, n-octyl glucoside. This detergent is a mild one and was used at the suggestion of Dr. Irwin D. Kuntz based on some

conversations he had had with scientists at Merck, Inc., who had been using octyl glucoside to solubilize some of their expressed aggregated proteins (154). I tried a variety of extraction conditions using from 0.1% to 5% octyl glucoside in sonication buffer for time periods ranging from 3 months (stored in the refrigerator) to 16 hours at 37 °C with gentle shaking (~50 rpm). This treatment gave dramatic improvement in specific activity. As will be shown later, in Figure 33, the octyl glucoside did not appear to solubilize the CK pellet, as we expected, but did solubilize other proteins in the crude pellet. These proteins could be removed in the detergent-containing supernatant following centrifugation. The best improvement in specific activity was obtained by extraction of the pellet in sonication buffer containing 1% beta-mercaptoethanol and 2.5% n-octyl glucoside at 37 °C with shaking at 50 rpm for > 16 hours.

C. Solubilization Studies on the Crude Pellet.

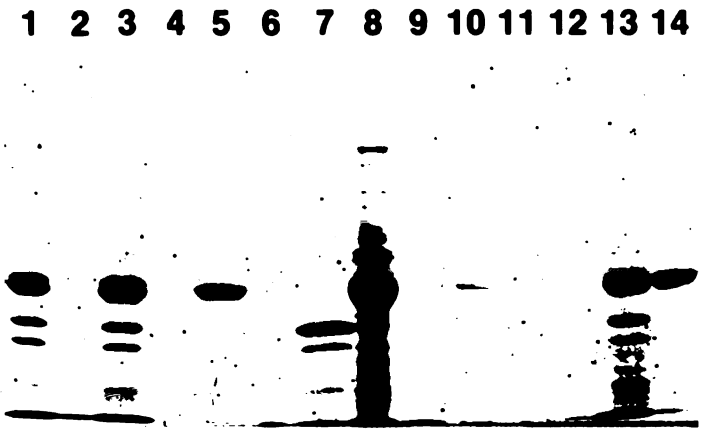
While the octyl glucoside experiments were very encouraging in the improvement of CK specfic activity as mentioned above, they did not solubilize the CK pellet. I performed a series of experiments directed at determining how difficult it was to solubilize the aggregated CK. I already knew, from early refolding experiments, that the CK aggregate was completely solubilized in 6 M Gdn-HCl. From SDS-PAGE, it was obvious that the aggregate is also solubilized in Laemmli buffer containing 1% beta-mercaptoethanol and 2.5% SDS. I performed some experiments to determine whether the CK aggregate could be solubilized under less extreme conditions.

Seventy-five ug each of the crude pellet from CK expression preparation 13 and

rabbit muscle CK (labeled as the Control) from Sigma were suspended or dissolved in 200 ul sonication buffer containing either no additives, 1% SDS + 1% beta-mercaptoethanol, or 2.5% n-octyl glucoside + 1% beta-mercaptoethanol. These tubes were incubated at 37 °C overnight and then spun for 10 minutes at 15,000 x g in the microfuge. The supernatants were transferred to a new tube and spun again for 6 minutes. The supernatants were carefully transferred to a new tube to avoid carrying along any solid material. There was no pellet left in either the expressed CK or control CK tubes after treatment with SDS. Both the expressed CK and the control CK had a pellet left after the octyl glucoside treatment. This was surprising for the control, which was soluble rabbit muscle CK. But the detergent treatment evidently caused this CK to precipitate. There was a pellet left after centrifugation of the expressed CK treated with buffer only and no pellet left for the rabbit CK control after treatment with buffer only followed by centrifugation. Seventy-five ul of sonication buffer was added to the tubes containing the pellets following removal of the supernatants and the pellets were resuspended by vortexing. Amounts of material for the gel were chosen so that equivalent amounts of CK would appear in each lane if it were assumed that all of the CK were contained in each tube. These samples were run on a 10% polyacrylamide gel and the results are shown in Figure 33.

The first two lanes show the crude pellet that was extracted along with the supernatant in which it had been stored for several days. These samples form a baseline control and show, as expected, all of the protein in the pellet (lane 1) and none

Figure 33: SDS-PAGE of Solubilization and Extraction of KTCK3F



Lanes 1, 2, 3, 4, 7, 8, 12, 13 are from CK preparation # 13 and represent the crude pellet treated as described for each lane. Lane 1) Crude pellet; Lane 2) Filtered supernatant from crude pellet; Lane 3) Supernatant from SDS extraction; Lane 4) Pellet left after SDS extraction; Lane 5) Supernatant from SDS extraction of rabbit muscle CK (Sigma); Lane 6) Pellet left after SDS extraction of rabbit muscle CK (Sigma); Lane 7) Supernatant from octyl glucoside extraction; Lane 8) Pellet left after octyl glucoside extraction; Lane 9) Supernatant from octyl glucoside extraction of rabbit muscle CK (Sigma); Lane 10) Pellet left after octyl glucoside extraction of rabbit muscle CK (Sigma); Lane 11) Blank; Lane 12) Supernatant from sonication buffer extraction; Lane 13) Pellet left after sonication buffer extraction; Lane 14) Rabbit muscle CK (Sigma) in sonication buffer treated in the same manner as all pellet extractions, except for final spin to separate pellet and supernatant.

in the supernatant. Lanes 3-5 show the results of the SDS extraction. In this case, all of the expressed CK pellet has been solubilized by 1% SDS/1% beta-mercaptoethanol since it can been seen in the SDS supernatant (lane 3). The tube containing what would have been leftover pellet shows no protein in lane 4. Lane 5 shows all of the rabbit muscle control CK to be in the supernatant tube as expected. I did not apply any control CK from the tube that would have contained a pellet to the gel since there was no pellet to be seen at all. These results show that 1% SDS completely solubilizes the expressed CK pellets.

Lanes 7-10 shows the results of the octyl glucoside extraction. These results are much more interesting than the SDS extraction results and suggest one possible reason that treatment of the crude pellets with octyl glucoside/beta-mercaptoethanol prior to unfolding in Gdn-HCl gave higher specific activity refolded CK. Lane 7 shows that the supernatant removed after extraction of the pellet contains significant amounts of non-CK protein. The pellet extracted by this detergent, shown in lane 8, shows all of the CK, as well as a decrease in contaminating proteins when this pellet is compared to that in Lane 1. The contaminating proteins which have decreased in the pellet correlate with the proteins which have been extracted out into the supernatant in lane 7. Clearly, the pellet has been somewhat purified by extraction with octyl glucoside/beta-mercaptoethanol. This purification has probably contributed to the higher specific activity in the refolded CKs. The control samples extracted with this detergent show no CK in the supernatant and a small amount of CK in the pellet fraction (lanes 9, 10). Even the soluble rabbit muscle CK apparently precipitates in the octyl glucoside-containing buffer. The

precipitated control samples treated in this way showed no loss of activity.

Lanes 12-14 show the results of both expressed CK and control CK extracted with sonication buffer containing no additives. As expected, all of the expressed CK is in the pellet (lane 13) and this pellet was not discernibly purified by extraction with the buffer alone as can be seen by the lack of protein in the supernatant fraction (lane 12). Lane 14 shows the control CK in buffer only and incubated as the other samples. This sample was not centrifuged after extraction, but applied directly to the gel. The gel shows that this incubation treatment did not cause any breakdown of the control CK.

Altogether, the results of these experiments were very promising, giving us some clue as to how the octyl glucoside treatment could improve specific activity in the refolded CK. The SDS extraction results, however, were disappointing. I was hoping to see some purification of the pellet, short of complete resolubilization of it in the 1% SDS extraction. In order to see if SDS could perform a similar or better purification of the crude pellet over that resulting from the octyl glucoside treatment, I performed a further experiment, extracting the crude pellet with lower levels of SDS.

The SDS extraction was performed similarly to the previous extraction except that the incubation period was only two hours. The crude pellet only was extracted (no rabbit muscle control CK was included). The pellet left after purification by octyl glucoside/beta-mercaptoethanol was resuspended in sonication buffer and split into 4 equal aliquots. To each was added a volume of 10% SDS to give a range of final SDS concentrations of 0%, 0.01%, 0.05%, and 0.1%. Following incubation and centrifugation, equivalent volumes from each tube were applied to a 10% polyacrylamide gel and run as

above (data not shown). The results showed no extraction or purification of the pellet by any of the extractants lower than 0.05% SDS. The extraction at an SDS concentration of 0.05% showed a small amount of CK and other proteins in the supernatant, but all of the rest of the protein in the pellet fraction. There was no differential extraction of the pellet that could be seen on the gel. The 0.1% SDS extraction showed all of the material to be solubilized in the supernatant and only some very faint bands, containing CK and other proteins in the pellet fraction. Thus, I had found that the entire pellet is solubilized by SDS at a concentration between 0.05-0.1% but that there had been no differential extraction of the pellet. This experiment should be repeated with the crude pellet resulting from the KT52SphBgl transformation as a control to see if there is a difference in the solubility behavior between pellets with and without CK present.

These results, while encouraging overall, did little to clarify my understanding of the nature of the pellet. Originally, I had expected the pellet to be nearly pure CK, as is the aggregate formed by overexpression of human insulin in bacteria (155). As can be seen from Figure 32, the crude pellet we had purified was not nearly pure CK, but was highly contaminated with other proteins. The extraction data is difficult to interpret insofar as it gives conflicting results as to whether the pellet is simply an accumulation of insoluble proteins which are not physically intertwined (thus being part of the same superstructure aggregate) or whether the pellet is formed by many CK molecules as well as possibly other pellet proteins covalently or otherwise attached to each other to form a highly insoluble aggregate. While I have not been able to answer this question, I have performed some experiments and made some observations which may be useful to those

who might attempt to answer this question in the future.

The primary direction of these experiments has been based on the assumption that the aggregate may be held together by intermolecular disulfide bonds, either between CK molecules alone or between CK molecules and other protein molecules found in the aggregate. There is some evidence that such disulfide bonds are involved in aggregates of some eukaryotic proteins expressed in bacteria. Dr. Steven Anderson from Genentech suggested this possibility to me early in my experiments with the CK aggregate (156) citing this problem as one he had encountered with a number of proteins. The possibility is plausible because while CK does not normally contain disulfide bonds, it does contain 7 free cysteine residues. If the protein had misfolded after translation in the bacterial system, incorrect disulfide bonds could possibly have formed. The fact that I could get differential extraction of the crude pellet in the presence of octyl glucoside/ beta-mercaptoethanol tends to support the idea that the pellet was not one species held together by covalent interactions. The fact that the SDS extraction experiments showed no differential solubilization tends to support the idea that the pellet left after octyl glucoside extraction is held together by relatively strong forces. The second SDS extraction was performed in the absence of reducing agent. This experiment should be repeated with the addition of beta-mercaptoethanol to see if differential extraction could be obtained by breaking disulfide bonds in the presence of low concentrations of SDS. It might also be possible to follow up the disulfide bond theory by trying to run an SDS gel using a sample buffer with and without reducing agent. The appearance of high molecular weight species correlating with the loss of low molecular weight species,

particularly the CK band in the samples without reducing agent would be good evidence for the involvement of disulfide bonds in the aggregate formation. I did try such gels but the results were inconclusive and must be repeated before any conclusions can be drawn from them.

In summary, attempts at further purification of the crude pellet by extraction and differential solubilization experiments were successful in that the octyl glucoside extractions did remove some contaminating proteins and improved specific activities of refolded CK. These extracted pellets were labeled "Final Pellets" and were eventually used for refolding experiments. (Quantitation of specific activities obtained in all of the refolding experiments will be discussed in a later section.) These experiments were not very succussful, however, in providing interpretable results that would help us to understand the nature of the physical forces holding the aggregate together in the crude pellet. These experiments also showed the crude pellet and the final pellet to be highly insoluble, with the CK portion, in particular, resistant to solubilization by any of the detergents we tried. In fact, it appears to be soluble only in SDS and Gdn-HCl. The hardiness of this aggregate appears to protect the CK from degredation even though this results in a pellet very difficult to purify to a high degree of homogeneity prior to the unfolding and refolding steps of the reactivation scheme. Obviously, the fact that we had to perform the reactivation steps on such "dirty" material adversely affected the quality of the final refolded product and makes it difficult to discern whether the lack of native-level specific activity in the final refolded protein is due to the impurity of the pellet or to some modification of the protein arising from expressing it in a foreign expression system.

D. Solubilization of the Final Pellet

Unfolding the final pellet in 6 M Gdn-HCl was straightforward compared to the extraction experiments. As shown in the Scheme detailed in Figure 31, the pellet was solubilized in 6 M Gdn-HCl in a volume that would give an approximate CK concentration of 15 ug/ml upon a 60-fold dilution in the refolding step. This concentration was chosen because it gave the best specific activities. I do not know why such a low concentration was required but it may simply be due to the fact that with so many contaminating proteins present in the denatured and refolded solution, a very low protein concentration allowed the refolding CK molecules to "see" very few other molecules with which they could get entangled during the refolding process (144). The length of time and temperature for unfolding (2 hours at room temperature) was also chosen empirically.

3. Refolding and Characterization of the Denatured Creatine Kinases

Ultimately, refolding the denatured CKs was done as presented in the third part of Figure 31. This final scheme was, however, the result of much trial and error beforel could obtain high CK specific activities reproducibly. The following is a short description of some of the experiments we tried to optimize the refolding procedure.

A. Observations on the Refolding Process.

As mentioned earlier, we found that the best specific activities could be obtained only when the CK concentration during refolding was below 20 ug/mL. Since the unfolding process involved a solution some 60-fold more concentrated than the dilute

refolding solution, we routinely unfolded the CK final pellets at an estimated CK concentration of 1.2 mg/mL. Before we started using the JA221-I cells to give us greater amounts of CK per preparation, we were unfolding and refolding in very small volumes (on the order of 0 .1-0.5 mL of 6M Gdn-HCl solution). Since we suffered tremendous losses of CK during the refolding process (there was always a precipitate formed upon refolding), the total amounts of refolded CK available for study were often only a few ug. Later, as the amount of CK we could produce in any ony preparation increased with the use of JA221-I cells, we could unfold 1 or 2 mg of CK as final pellet in each experiment, recovering up to 200 ug of refolded CK from each attempt.

We tried both dialysis and dilution as refolding methods since both had been used successfully with CK purified from rabbit muscle tissue (11). We found that the dialysis method produced more precipitate and less recovered active CK than did the dilution method. This was probably due to the higher concentration of CK during refolding with the dialysis method.

The refolded CK, by our best methods, still contained contaminating proteins as can be seen in lanes 6 (native CK from KTCK3F) and 10 (mutant CK from KTAla7) in Figure 32. While we made some attempt to further purify the refolded CK (as will be discussed below), we were unable to increase the specific activity by such purification. It will be interesting to see how much the quality of the refolded CK can be improved if we are able to purify the final pellet more completely. It would also be helpful if we could purify the denatured proteins while they are in the Gdn-HCl solution so that we could refold a more homogeneous CK preparation. I have already mentioned in the Introduction

to this Chapter that such techniques have been used successfully for other aggregated proteins. While we did try to purify the denatured proteins by dialyzing the 6 M Gdn-HCl solution into 8 M urea and then applying it to an ion exchange resin on FPLC, but we did not have enough pellet to work out the conditions well enough to give good purification of the denatured CK. Both this technique and purification over a Sephadex column equilibrated with a buffer containing 6 M Gdn-HCl should be tried now that large amounts of expressed CK are available.

In spite of all of the improvements in reactivation techniques that should be tried, we have succeeded in refolding CK (using the techniques described in Figure 31) to nearly normal specific activity. The experiments leading to these results are detailed below. This work is divided into two sections. The section titled "Early Refolding Experiments" details some of the experimentation we did with the refolding procedures in attempt to improve our initially very low specific activites. It also gives specific activity results from these early experiments. The second section, titled "Improvement of CK Activation," describes the improvements we made in the refolding procedures to give us dramatically better specific activities than those reported in the first section.

B. Early refolding experiments

The early refolding experiments were done on crude pellets which had not been further purified by extraction with octyl glucoside/ beta-mercaptoethanol. We had, however, experimented with varying levels of DTT throughout the preparations and had found that our best results correlated with having at least 10 mM DTT present throughout

the preparations.

At the suggestion of Dr. Kuntz, I tried refolding these crude pellets in the presence of one or both substrates or of a mixture of substrates producing a dead-end complex, ADP and creatine, along with the planar anion used to stabilize the dead-end complex, NO_2^- (11). This last mixture was used in the hope that, as a transition-state analog, it might help stabilize the refolding CK into the correct conformation during refolding. The presence of neither the substrates nor the transition state analog complex in the refolding buffer produced improved specific activity in the refolded CKs. Again, these experiments should be repeated with the preparations that do give improved specific activity.

Attempts were also made to purify the refolded and concentrated material in order to improve its specific activity. The refolded CK was applied to a Blue Sepharose column (an affinity resin used for purification of proteins with an ATP substrate) using the conditions with which we had successfully purified CK obtained from rabbit muscle tissue (157). While the CK obtained from this column looked cleaner on SDS-PAGE than unpurified refolded CK, it had no better specific activity.

Thus, in our early refolding experiments, none of our attempts to improve specific activity of the protein refolded from the crude pellet were successful. We were unable to obtain specific activities better than a few percent of that obtained for rabbit muscle CK unfolded, refolded, and assayed under the same conditions as our expressed material. The average specific activities obtained from these experiments are 0.13 U/mg of protein for CK translated from KTCK3F and 0.02 U/mg of protein for CK translated from KTAla7. These values compare to literature values of 4-15 U/mg of protein for CK purified from

Torpedo electric organ. We obtained specific activities of 20 U/mg of protein for rabbit muscle CK made up freshly in refolding buffer and 9.2 U/mg of protein for rabbit muscle CK unfolded and refolded in the same manner as our expressed proteins.

It should be noted that these specific activities are inaccurate numbers insofar as it was very difficult to get accurate protein concentrations on such small amounts of material. Since we had only ug quantities of material from any one refolding experiment and since DTT was absolutely necessary in the buffers in order to maintain CK activity, protein concentrations were done on solutions containing DTT. Unfortunately, both DTT and to a lesser extent, Tris buffers, interfere with all of the protein concentration assays, giving a falsely high reading. Even though standard curves were made using rabbit muscle CK, freshly made up in the DTT-containing buffers as a standard, we never had enough expressed and refolded protein to repeat assays using standards at a low enough protein concentration to correct for the errors. In order to compensate for this problem, we also estimated CK protein concentrations from SDS gels by comparing the intensity of the refolded CK bands with that of known amounts of rabbit muscle CK standard.

Thus, the protein concentrations we report here are only crude estimates. In any case, under any method of protein concentration determination used, the specific activities of our expressed and refolded CKs never rose above 5% of that of soluble rabbit muscle CK in these early experiments. As discussed in the next section, we were able to improve these numbers dramatically in later experiments.

C. Improvement of CK Reactivation

The two methods we found which eventually gave us improved CK activities in the refolded proteins was extraction of the crude pellets with n-octyl glucoside and treatment of both the crude pellet and denatured Gdn-HCl solution with a high concentration of beta-mercaptoethanol. The first experiment with the detergent, CK preparation #10, was stored in sonication buffer containing 50 mM DTT and from 0.1% to 2.5% octyl glucoside in the refrigerator for approximately two months before the pellets were refolded and assayed. The results showed dramatic improvement in the specific activities of the refolded CK (only wild-type CK was expressed for this preparation). With the protein concentrations estimated by densitometry on the gels of the refolded CK compared to known amounts of CK standard on the same gel, the specific activity of the expressed CK was 68% of that of refolded rabbit muscle CK. When these experiments were repeated on CK preparation #15 and #17, under more controlled conditions, the results were similar.

In these experiments, some of the CK crude pellet was extracted in sonication buffer with and without octyl glucoside/beta-mercaptoethanol as specified in Figure 31. These final pellets along with an equivalent amount of crude pellet was then denatured in 6 M Gdn-HCl containing 1% beta-mercaptoethanol and refolded as specified in Figure 31. The CKs refolded from the crude pellet only (no extraction) showed about 60% of the activity of rabbit muscle CK refolded in the same manner. Evidently, the high percentage of reducing agent present during unfolding in Gdn-HCl was sufficient alone to increase dramatically the specific activity of the wild-type CK. The CK refolded after overnight extraction at 37 ^{o}C without octyl-glucoside but including the beta-mercaptoethanol

showed from 20-50% of the activity of refolded rabbit muscle CK. The CKs similarly extracted but including 2.5% octyl glucoside showed specific activities ranging from 40-50% that of refolded rabbit muscle CK.

Obviously, these experiments need to be repeated with better protein quantitation to determine whether the presence of the high concentration reducing agent or the detergent (or both) are primarily responsible for the increases in specific activity in these experiments over that obtained in the earlier experiments. In addition, protein quantitation and activity assays should be done with Torpedo electric organ CK rather than rabbit muscle CK as the control. Nevertheless, the reactivation scheme detailed in Figure 31 does produce soluble CK from the original aggregate with activities approaching what might be expected from normally purified tissue-derived CKs.

4. Comparison of Wild-Type and Mutant CKs

Our most careful early experiments showed the refolded Ala mutant CK to have about 16% of the activity of refolded wild-type CK. The fact that both of these expressed proteins showed such low specific activities compared to CKs purified from tissue is troubling in that it raises the question of whether the expressed CKs are similar enough to native CKs to allow us to consider the results on the activity of the Ala mutant as valid. If the expressed CKs are covalently modified in the expression system so that they cannot refold into native-like conformations, comparisons of the wild-type and mutant expressed CKs are worthless. If, on the other hand, it can be shown that the expressed CKs can be refolded to native-like conformations, then comparisons of wild-type and mutant proteins

are meaningful. The fact that we were able to improve the purification of the crude pellet and the denaturation step to the point that we could obtain nearly normal activity for the wild-type expressed CK suggests that the expressed protein can be used as a valid model for native protein. On the basis of this argument I make the tentative claim that Cys_{283} is not essential to CK activity since the Ala_{283} mutant shows low but appreciable activity.

Obviously, the refolding experiments must be repeated on both the wild-type and mutant CKs using our improved reactivation protocols in order to fully validate this result.

Conclusion

These experiments in the reactivation of CK and comparison of wild-type and mutant proteins represent unfinished work. So much time was invested in learning to reactivate the aggregated CK that there was simply not enough time left to do the controlled studies necessary to complete the story of expression of CK in E. coli using the trc promoter. It is open to question whether further investment of time in this expression system is worthwhile. Work is already underway by others in this laboratory to express rabbit muscle CK in yeast or in bacterial systems using other promoters. Achievement of expression of soluble CK in one of these systems would be vastly superior to the system I have been working on both in terms of yield and of time invested to obtain active protein.

Nevertheless, this work represents the accomplishment of expression of active CK in one prokaryotic system. It also represents the only system yet developed for the expression of cloned CK DNA. While this expression system may not ultimately be a useful one for extensive studies of mutant CKs, I hope that this work contributes some

insights or tools for the use of those studying CK. I also hope that my observations about the CK aggregate may be useful in the guidance of others facing similar problems in expressing proteins in systems foreign to their origin.

References

1. Krebs, H.A. (1970) Perspect. Biol. Med. 14, 154.

2. Blow, D.M., Birktoft, J.J., and Hartley, B.S. (1969) Nature 225, 811.

3. Fersht, A. (1984) Enzyme Structure and Mechanism (W.H. Freeman and Co., New York), pp. 17-21, 27-32, 338-340, 405-413.

4. Cohen, S.N., Chang, A.C.Y., Boyer, H.W., and Helling, R.B. (1973) Proc. Natl. Acad. Sci. USA 70, 3240.

5. As an example, until 1984 there were no published amino acid sequences for creatine kinase in the literature, although at least one laboratory, that of S. Kuby, had been working on it for several years. Their work was done using amino acid sequencing techniques. By 1986, nine CK sequenceshad been published, all inferred from DNA sequences. This patternhas been repeated for many other proteins as well as can be seen by analysis of the databases referenced in Chapter 2. Over the last six years, the number of protein sequences derived from nucleotide sequencing has far outstripped the those derived from amino acid sequencing.

6. Clark, A.J.L., Adeniyi-Jones, R.O., and Knight, G. (1982) Lancet 2, 354.

7. Martial, J.A., Hallewell, R.A., Baxter, J.D., and Goodman, H.M. (1979) Science 205, 602.

8. Goeddell, D.V., Yelverton, E., and Ullrich, A. (1980) Nature 287, 411.

9. Kuby, S.A., Nodal, L., and Lardy, H.A. (1954) J. Biol. Chem. 209, 191.

10. Ganong, W.F. (1977) Review of Medical Physiology (Lange Medical Publications, Los Altos, CA).

11. Kenyon, G.L. and Reed, G.H. (1983) in Advances in Enzymology and Related Areas of Molecular Biology, ed. Meister, A. (Wiley Interscience, New York), Vol. 54, pp. 367-426.

12. Watts, D.C. (1973) in The Enzymes, ed. Boyer, P.D. (Academic Press,) New York, Vol. 8, pp. 383-455.

13. Walliman, T. and Eppenberger, H.M. (1985) in Cell and Muscle Motility, ed. Shay, J.W. (Plenum Press, New York), Vol. 6, pp. 239-285.

14. Kenyon, G.L., Babbitt, P.C., and Reed, G.H. (In preparation) in Advances in Enzymology and Related Areas of Molecular Biology.

15. Witzemann, V. (1985) Eur. J. Biochem. 150, 201.

16. Cheneval, D. and Carafoli, E. (1988) Eur. J. Biochem. 171, 1.

17. Kuby, S.A. and Noltmann, E.A. (1962) Enzymes 6, 572.

18. Bickerstaff, G.F. and Price, N.C. (1976) FEBS Lett. 64, 319.

19. Morrison, J.F. and James, E. (1965) Biochem. J. 97, 37.

20. O'Sullivan, W.J. and Cohn, M. (1966) J. Biol. Chem. 241, 3116.

21. McLaughlin, A.C. Leigh, J.S., Jr., and Cohn, M. (1976) J. Biol. Chem. 251, 2777.

22. Markham, G.L., Reed, G.H., Maggio, E.T., and Kenyon, G.L. (1977) J. Biol. Chem. 252, 1197.

23. Mahowald, T.A., Noltmann, E.A., and Kuby, S.A. (1962) J. Biol. Chem., 237, 1535.

24. Cohn, M. and Reuben, J. (1971) Accts. Chem. Res. 4, 212.

25. O'Sullivan, W.J. and Cohn, M. (1968) J. Biol. Chem. 243, 2733.

26. West, B.L., Babbitt, P.C., Mendez, B., and Baxter, J.D. (1984) Proc. Natl. Acad. Sci. USA 81, 7007.

27. Babbitt, P.C., Kenyon, G.L., Kuntz, I.D., Cohen, F.E., Baxter, J.D., Benfield, P.A., Buskin, J.D., Gilbert, W.A., Hauschka, S.A., Hossle, J.P., Ordahl, C.P., Pearson, M.L., Perriard, J.-C., Pickering, L.A., Putney, S.D., West, B.L., and Ziven, R.A. (1986) J. Prot. Chem. 5, 1.

28. Milner-White, E.J. and Watts, D.C. (1971) Biochem. J. 122, 727.

29. Smith, D.J. and Kenyon, G.L. (1974) J. Biol. Chem. 249, 3317.

30. Fawcett, A.H., Keto, A.I., Mackerrass, P., Hamilton, S.E., and Zerner, B. (1982) BBRC 107, 302.

31. Bessman, S.P. and Geiger, P.J. (1981) Science 211, 448.

32. Hammes, G.G. and Huest, J.K. (1969) Biochemistry 8, 1083.

33. Grossman, S.H. (1983) J. Neurochem. 41, 729.

34. Jacobs, H.K. and Kuby, S.A. (1970) J. Biol. Chem. 245, 3305.

35. James, T.L. and Cohn, M. (1974) J. Biol. Chem. 249, 2599.

36. James, T.L. (1976) Biochemistry 15, 4724.

37. Borders, C.L., Jr. and Riordan, J.F. (1975) Biochemistry 14, 4699.

38. Vasak, M., Nagayama, K., Wuthrich, K., Mertens, M.C., and Kagi, J.H.R. (1979) Biochemistry 18, 5050.

39. Cook, P.F. , Kenyon, G.L., and Cleland, W.W. (1981) J. Biol. Chem. 20, 1204.

40. Eppenberger, H.E., Dawson, D.M., and Kaplan, N.O, (1967) J. Biol. Chem. 242, 204.

41. Atherton, R.S., Laws, J.F., Miles, B.J., and Thomson, A.R. (1970) Biochem. J. 120, 589.

42. Mendez, B., Valenzuela, P., Martial, J., and Baxter, J.D. (1980) Science 209, 695.

43. Cathala, G., Savouret, J.F., Mendez, B., West, B.L., Karin, M., Martial, J.A., and Baxter, J.D. (1983) DNA 2, 329.

44. Duguid, J.R. and Raftery, M.A. (1973) Biochemistry 12, 3593.

45. Rosenberg, U.B., Kunz, B., Frischauf, A., Lehrach, H., Mahr, R., Eppenberger, H.M., and Perriard, J.-C. (1982) Proc. Natl. Acad. Sci. USA 79, 6589.

46. Schweinfest, C.W., Kwiatkowski, R.W., Dottin, R.P. (1982) Proc. Natl. Acad. Sci. USA 79, 4997.

47. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408.

48. Cooke, N.E., Coit, D., Weiner, R.I., Baxter, J.D., and Martial, J.A. (1980) J. Biol. Chem. 255, 6502.

49. Land, H., Grez, M., Hauser, H., Lindenmaier, W., and Schutz, G. (1981) Nuc. Acids Res. 9, 2251.

50. Peacock, A.C. and Dingman, C.W. (1967) Biochemistry 6, 1818.

51. Paterson, B.M., Roberts, B.E., and Kuff, E.L. (1977) Proc. Natl. Acad. Sci. USA 74, 4370.

52. Pelham, H.R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247.

53. Laemmli, U.K. (1970) Nature 227, 680.

54. Kronenberg, H.M., McDevitt, B.E., Majzoub, J.A., Nathans, J., Sharp, P.A., Potts, J.T., and Rich, A. (1979) Proc. Natl. Acad. Sci. USA 76, 4981.

55. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463.

56. Messing, J. and Viera, J. (1982) Gene 19, 269.

57. Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979) Nuc. Acids. Res. 7, 2115.

58. Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113, 237.

59. Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193.

60. Olson, O.E. and Kuby, S.A. (1964) J. Biol. Chem. 239, 460.

61. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211.

62. Brawerman, G. (1981) CRC Crit. Rev. Biochem, 1-38.

63. Martinez, H.M., Katzung, B. and Farrah, T. (1983) Sequence Analysis Program Manual (Biomathematics Computation Laboratory, University of California, San Francisco).

64. Burke, R.L., Tekamp-Olson, P. and Najarian, R. (1983) J. Biol. Chem. 258, 2193.

65. Metzger, H., Shapiro, M.B., Mosimann, J.E., and Vinton, J.E. (1968) Nature 219, 1166.

66. Kozak, M. (1983) Microbiol. Rev. 47, 1.

67. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201.

68. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

69. Streeck, R.E. and Gebhardt, C. (1979) Hoppe-Seyler's Z.

Physiol. Chem. 360, 529.

70. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201.

71. Kozak, M. (1984) Nature 308, 241.

72. Kozak, M. (1984) Nuc. Acids Res. 12, 857.

73. Kaye, A.M. (1983) in Regulation of Gene Expression by Hormones, ed. McKerns, K.W. (Plenum, New York), pp. 103.

74. Barrantes, F.J., Mieskes, G., and Wallimann, T. (1983) Proc. Natl. Acad. Sci. USA 80, 5440.

75. Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 222.

76. McPherson, A. Jr. (1973) J. Mol. Biol. 81, 83.

77. Burgess, A.N., Liddell, J.M., Cook, W., Tweedle, R.M., and Swan, I.D.A. (1978) J. Mol. Biol. 123, 691.

78. Takasawa, T., Fukushi, K., and Shiokawa, H. (1981) J. Biochem. 89, 691.

79. Gilliland, G.L., Sjolin, L. and Olsson, G. (1983) J. Mol. Biol. 170, 791.

80. Fisher, M.L., Carliner, N.H., Becker, L.C., Peters, R.W., and Plotnick, G.D. (1983). J. Am. Med. Assoc. 249, 393.

81. Eppenberger, H.M., Perriard, J.-C., and Walliman, T. (1983) in Isozymes: Current Topics in Biological and Medical Research, eds. Rattozzi, M., Scandillios, J.C., and Whitt, G.S. (Alan R. Liss, New York), Vol. 7, pp. 19-38.

82. Putney, S., Herlihy, W., Royal, N., Pang, H.Y., Aposhian, H.V., Pickering, L., Belagage, R., Biemann, K., Page, D., Kuby, S., and Schimmel, P. (1984) J. Biol. Chem. 259, 14317.

83. Pickering, L., Pang, H., Biemann, K., Monro, H., and Schimmel, P. (1985) Proc. Natl. Acad. Sci. USA 82, 2310.

84. Benfield, P.A., Ziven, R.A., Miller, L.S., Sowder, R., Smythers, G.W., Henderson, L., Oroszlan, S., and Pearson, M.L. (1984) J. Biol. Chem. 259, 14979.

85. Benfield, P.A., Henderson, L., and Pearson, M.L. (1985) Gene 39, 263.

86. Ordahl, C.P., Evans, G.L., Cooper, T.A., Kunz, G., and Perriard, J.-C. (1984) J. Biol. Chem. 259, 15224.

87. Kwiatkowski, R.W., Schweinfest, C.W., and Dottin, R.P. (1984) Nucleic Acids Res. 12, 6925.

88. Perriard, J.-C., Achtnict, U., Cerny, L.C., Eppenberger, H.M., Grove, B.K., Hossle, J.P., and Schafer, B. (1985) in Molecular Biology of Muscle Development, eds. Emerson, C., Fischman, D.A., Nadal-Ginand, B., and Siddigin, M.A.Q. (Alan R. Liss, New York).

89. Buskin, J.N., Jaynes, J.B., Chamberlain, J.S., and Hauschka, S.D. (1985) J. Mol. Evol. 22, 334.

90. Giraudat, J., Devillers-Thiery, A., Perriard, J.-C., and Changeux, J.-P. (1984) Proc. Natl. Acad. Sci. USA 81, 7313.

91. Barker, W.C., Hunt, L.T., Orcutt, B.C., George, D.G., Yeh, L.S., Chem, H.R., Blomquist, M.C., Johnson, G.C., Seibel-Ross, E.I., and Ledley, R.S. (1984) Protein Identification Resource of the NIH, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C.

92. Barker, W.C., Chen, H.R., Hunt, L.T., Orcutt, G.C., Seibel-Ross, E.I., and Dayhoff, M.O. (1984) Nucleic Acid Sequence Database, Version 22, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C.

93. Doolittle, R.F. (1983) Newat83, Department of Chemistry, University of California, San Diego, California.

94. GenBank (1984). Release 23.0, European Molecular Biology Laboratory.

95. Dickerson, R.E. (1971) J. Mol. Evol. 1, 26.

96. Orcutt, B.C., Dayhoff, Mo.O., and Barker, W.C. (1984). Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C.

97. Lipman, D.J. and Pearson, W.R. (1985) Science 22, 1435.

98. Lipman, D.J. and Pearson, W.R. (1985) Proc. Natl. Acad. Sci. USA 80, 726.

99. Doolittle, R.F. (1981) Science 214, 149.

100. Martinez, H.M. and Sobel, E. (1985) Nucleic Acids Res.

101. Roman, D.G., Billadello, J.J., Gordon, J.I., Sobel, B.E., and Strauss, A.W. (1985) Clin. Res. 33, 222A.

102. Mahowald, T.A. (1969) Fed. Proc. 28, 601.

103. Knell, J.D., Perryman, M.B., Bohlmeyer, T.J., and Roberts, R. (1985) Fed. Proc. 44, 1464.

104. Eppenberger, H.M., Dawson, D.M., and Kaplan, N.D. (1967) J. Biol. Chem. 242, 204.

105. Dawson, D.M., Eppenberger, H.M., and Kaplan, N.O. (1967) J. Biol. Chem. 242, 210.

106. Barrantes, F.J., Mieskes, G., and Walliman, T. (1983) FEBS Lett. 152, 270.

107. Gysin, R. and Flanagan, S.D. (1985) Fed. Proc. 44, 2817.

108. Perryman, M.B., Knell, J.D., Ifegwu, J., and Roberts, R. (1985) J. Biol. Chem. 260, 9399.

109. Rosenberg, U.B., Kunz, G., Frischauf, A., Lehrach, H., Mahr, R., Eppenberger, H.M., and Perriard, J.-C. (1982) Proc. Natl. Acad. Sci. USA 79, 6589.

110. Reiss, N.A. and Kaye, A.M. (1980) J. Biol. Chem. 256, 5741.

111. DeLuca, M., Hall, N., Rice, R., and Kaplan, N.O. (1981) Biochem., Biophys. Res. Comm. 99, 189.

112. Walliman, T., Walzthony, D., Wegmann, G., Moser, H., Eppenberger, H.M., and Barrantes, F.J. (1985) J. Cell Biol. 100, 1063.

113. Caravetti, M., Perriard, J.-C., and Eppenberger, H.M. (1979) J. Biol. Chem. 254, 1388.

114. Perriard, J.-C. (1979) J. Biol. Chem. 254, 7036.

115. DerTerrosian, E., Pradel, L.A., Kassab, R., and van Thoai, N. (1969) Eur. J. Biochem. 11, 482.

116. Regnouf, F., Kassals, R., Debuire, B., Richard, C., and Han, K.K. (1981). Int. J. Peptide Protein Res. 17, 143.

117. DerTerrosian, E., Desvages, G., Pradel, L.A., Kassab, R., and van Thoai, N. (1971) Eur. J. Biochem. 22, 585.

118. Cohen, F.E., Abarbanel, R.M., Kuntz, I.D., and Fletterick, R.J. (1983) Biochemistry 22, 4894.

119. Cohen, F.E., Abarbanel, R.M., Kuntz, I.D., and Fletterick, R.J. (1986) Biochemistry 25, 266.

120. Abarbanel, R.M. (1985) PhD Thesis, University of California, San

Francisco, California.

121. Richardson, J.S. (1981) Adv. Protein Chem. 34, 167.

122. Straus, D. and Gilbert, W. (1985) Proc. Natl. Acad. Sci. USA 82, 2014.

123. Amann, E., Brosius, J., and Ptashne, M. (1983) Gene 25, 167.

124. DeBoer, H.A., Comstock, L.J., and Vassar, M. (1983) Proc. Natl. Acad. Sci. USA 80, 21.

125. Glover, D.M. (1984) Gene Cloning (Chapman and Hall, New York), pp. 111-115.

126. Friefelder, D. (1983) Molecular Biology (Science Books International, Boston, MA), pp. 545 ff.

127. Shine, J. and Dalgano, L. (1975) Nature 254, 34.

128. Buell, G. and Panayotatos, N. (1986) in Maximizing Gene Expression, eds. Reznikoff, W. and Gold, L. (Butterworths, New York), pp. 346-361.

129. Craik, C.S. (1985) Biotechniques 3, 12.

130. West, B.L., Private communication.

131. International Biotechnologies, Inc. (1987) Manual for the Sequel-DS/KF Biosystem Double-stranded Sequencing Kit.

132. Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Miozzari, G., Crea, R., and Seeberg, P.H. (1979) Nature 281, 544.

133. Sumi, S., Hasegawa, A., Yagi, S., Miyoshi, K., Kanezawa, A., Nakegawa, S., and Suzuki, M. (1985) J. Biotechnology 2, 59.

134. Kawaguchi, Y., Shimuezu, N., Nishimori, T.U., and Bepper, T. (1984) J. Biotechnology 1, 307.

135. Shoner, G., Ellis, L.F., and Schoner, B.E. (1985) Bio/Technology 2, 151.

136. Panyotatos, N., Fontaine, A., and Truong, K. (1983) J. Cell Biochem. Supplement 7B, 109.

137. Harris, T.J.R. (1983) in Genetic Engineering, ed. Williamson, R. (Academic Press, London, UK), vol. 4, pp. 127-185.

138. Wetzel, R. and Goeddel, D.V. (1983) in The Peptides, eds. Gross, E. and Weienhofer, J. (Academic Press, New York), vol. 5, pp. 1-64.

139. Schoemaker, J.M., Brasnett, A.H., and Marston, F.A.O. (1985) EMBO J. 4, 775.

140. Builder, S.E., Ogez, J.R., Jones, A.J.S., Lin, N.S.C., Olson, K.C., Shire, S.J., Pai, R., and Wetzel, R.B. (1984) European Patent Application, Number 83,307,840.5, filed 21.12.83.

141. Olson, K.C. (1985) US Patent, No. 4518526.

142. Marston, F.A.O., Lowe, P.A., Doel, M.T., Schoemaker, J.M., White, S., and Angel, S. (1984) Bio/Technol. 2, 800.

143. King, J. (1986) Bio/Technol. 4, 297.

144. Marston, F.A.O. (1984) Bio/Technol. 2, (preprint, page # unknown).

145. Fish, N.M. and Lilly, M.D. (1984) Bio/Technology 2, 623.

146. Schneitman, C.A. (1971) J. Bacteriology 108, 553.

147. London, J., Skrzynia, C., and Goldberg, H.E. (1974) Eur. J. Biochem. 47l, 409.

148. Mozhaev, N.M. and Martinek, K. (1982) Enzyme Microb. Technol. 4, 299.

149. Lowry, O., Rosebrough, N., Farr, A., and Randall, R. (1951) J. Biol. Chem. 193, 265.

150. Bradford, M. (1976) Anal. Biochem. 72, 248.

151. Wulff, K., Stahler, F. and Gruber, W. (1981) in Bioluminescence and Chemiluminescence (Academic Press, New York), pp. 209-215.

152. Tanzer, M.L. and Gilvarg, C. (1959) J. Biol. Chem. 234, 3201.

153. Dr. Elizabeth Komives, Private communication.

154. Dr. I.D. Kuntz, Private communication.

155. Williams, D.C., Van Frank, R.M., Muth, R.M., and Burnett, J.P. (1982) Science 215, 687.

156. Dr. Steven Anderson, Private communication.

157. Powers, V.L. and Babbitt, P.C., Unpublished results.