Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7007–7011, November 1984
Biochemistry

Creatine kinase protein sequence encoded by a cDNA made from *Torpedo californica* electric organ mRNA

(hybrid-arrested translation/M13 dideoxy sequencing)

BRIAN L. WEST*, PATRICIA C. BABBITT†, BERNARDITA MENDEZ‡, AND J. D. BAXTER*

*Metabolic Research Unit and the Department of Biochemistry and Biophysics, and †Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143; and ‡Department of Cell Biology, Universidad Catolica de Chile, P.O. Box 114-D, Santiago, Chile

Communicated by Nathan O. Kaplan, July 27, 1984

ABSTRACT **Creatine kinase (ATP creatine *N*-phosphotransferase, EC 2.7.3.2) 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. To obtain cDNA clones encoding creatine kinase sequences, a cDNA bank was constructed using mRNA from the electric organ of *Torpedo californica* and was screened by comparing differential colony hybridization of electric organ and liver-derived ^{32}P-labeled cDNAs. Cloned DNAs have been isolated that can arrest the abundant synthesis of M_r 40,000–43,000 material seen after *in vitro* translation of electric organ mRNA. One of the clones, CK52g8, was sequenced by the dideoxy M13 method and was found to encode a M_r 42,941 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 chicken and rabbit tissues. By contrast, no significant homology was found with the known sequences of kinases that use other substrates. RNA 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.**

Creatine kinase (CK; ATP creatine *N*-phosphotransferase, EC 2.7.3.2), 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 that require a more complete explanation in molecular terms. These include a specific localization of a portion of CK within the cell (1), conformational changes within the enzyme (2–4), 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 (5–7), a tryptophan is thought to interact with the adenine group (8), a histidine is presumed to act as the acid-base catalyst (9), and a cysteine can be chemically modified to cause changes in enzyme activity (10, 11). To create molecular models for how these amino acids contribute to 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 composition data have been obtained for CK from several species (12, 13), and a partial amino acid sequence has been reported for the rabbit muscle enzyme (14), no complete amino acid sequence has yet been reported.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

In earlier studies of the *in vitro* translation of mRNA from the electric organ of *Torpedo californica* (15, 16), the presence of heavy [^{35}S]methionine incorporation in the region of M_r 40,000–43,000 was observed. Since a single subunit of CK has this molecular weight and since the electric organ is rich in Na^+/K^+-ATPase (17) and presumably requires high energy reservoirs to regenerate its ionic balance after discharge, it was reasoned that this tissue might contain an abundance of 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 chicken muscle, but no sequence for these was given (18, 19).

MATERIALS AND METHODS

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 described by Cathala *et al.* (16). Poly(A) RNA was prepared by oligo(dT)-cellulose chromatography (20).

To synthesize full-length double-stranded cDNA, each first strand of the cDNA was tailed with ≈15 dCMP residues using terminal deoxynucleotide transferase (P-L Biochemicals) and synthesis of the second strand was primed with oligo(dG)$_{10}$ (Collaborative Research, Waltham, MA), as described by Cooke *et al.* (21) and Land *et al.* (22). The tailing was done after the first strand was heated at 70°C for 20 min. in 0.1 M NaOH to remove the RNA. The double-stranded cDNA (12 μg) was then treated with 2000 units of S1 nuclease (Miles) in 100 μl of 300 mM NaCl/30 mM NaOAc, pH 4.5/3 mM $ZnSO_4$ at 37°C for 1 hr. The double-stranded cDNA was then fractionated on a 10% polyacrylamide gel (23), selecting only the material longer than 500 base pairs (determined by DNA length markers), for recovery by electroelution. Approximately 10 ng of this material was tailed with dCMP and cloned by annealing to *Pst* I-cut, dGMP-tailed pBR322. A total of 2600 ampicillin-sensitive, tetracycline-resistant clones were obtained after transforming *Escherichia coli* strain RR1 with this DNA and plating onto LB agar plates containing 5 μg of tetracycline per ml.

Hybrid-Arrested Translation. For hybrid-arrested translation analysis, 0.4 μg of poly(A)$^+$ RNA and 2.5 μg of *Hin*dIII-cut plasmid DNA were hybridized using the conditions of Paterson *et al.* (24). *In vitro* translation of the RNA was performed using reticulocyte lysates (25) in the presence of 100 μCi of L-[^{35}S]methionine (>600 Ci/mmol; 1 Ci = 37 GBq; Amersham) in 30 μl. Protein samples were fractionated by $NaDodSO_4$/polyacrylamide gel electrophoresis according to Laemmli (26). Two cycles of CsCl gradient purification of the DNA were necessary to avoid nonspecific inhibition of

Abbreviation: CK, creatine kinase.

translation as reported by Kronenberg *et al.* (27).

DNA Sequencing. DNA sequencing was done by the dideoxy method (28), after subcloning overlapping fragments into the M13 phage vector MP10 (29), using *E. coli* strain JM101 as host. Dideoxynucleotides were from P-L Biochemicals, Klenow fragment of DNA polymerase I was from Boehringer Mannheim, and [α-^{32}P]dCTP (400 Ci/mmol) was from Amersham.

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 (30) with ^{32}P-labeled 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 protein(s) at M_r 40,000–43,000 (Fig. 1), and hence was a candidate for a creatine kinase-encoding cDNA. The 500 base pair insert of CKABh9 was ^{32}P-labeled by nick-translation (31) and the 2600 clones were screened 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 (32). Seven of the largest clones appeared to be very close in size, having inserts ≈1500 base pairs long. The restriction map of one of these clones, CK52g8, is shown in Fig. 2. Six other positive clones were close in size to CK52g8, and all had identical restriction maps, except that one or both of the vector *Pst* I sites were not reconstructed in some. The CKABh9 insert was found to have the same *Bgl* II, *Cla* I, and *Pst* I sites present in the 3′ portion of CK52g8. CK52g8 DNA also arrested the *in vitro* synthesis of the M_r 40,000–43,000 protein(s) (not shown).

The DNA sequence obtained for the CK52g8 insert is shown in Fig. 3. There are 1428 bases in this insert, followed by a stretch of poly(A) and bounded by the G·C tails used for annealing to pBR322. The sequencing strategy is shown in Fig. 2. All portions of the sequence were derived from at least two phage templates, and 92% was sequenced on both strands.

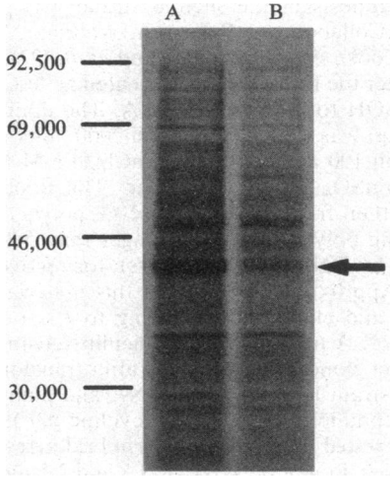

FIG. 1. Hybrid-arrested *in vitro* translation of *Torpedo* electric organ mRNA. Arrow indicates position of M_r 40,000–43,000 proteins(s). After hybridization of mRNA to CKABh9 DNA, [^{35}S]methionine incorporation into the M_r 40,000–43,000 protein(s) is blocked (lane B), whereas incorporation is strong after control hybridization without CKABh9 DNA (lane A). Five microliters of each translation mixture was loaded on the gel (2100 cpm in lane A, 1600 cpm in lane B). Molecular weights indicated on the left were taken from ^{14}C-methylated protein markers (phosphorylase b, bovine serum albumin, ovalbumin, and carbonic anhydrase) run on the same 12.5% polyacrylamide gel. Autoradiography was done for 21 days.

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 positions 1233–1235. The other two reading frames contain either 17 or 14 stop codons distributed over this same span. In Fig. 3, the amino acids at positions 268–292 are compared with the amino acid sequence reported by Atherton *et al.* (14) 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 glutamic acid at amino acid 275. The inferred carboxyl-terminal dipeptide, Gln-Lys, is identical to the carboxyl-terminal dipeptide for rabbit muscle CK (33). The molecular weight for this inferred protein is 42,941. Cytosine, guanine, uracil, and adenine are used in 47%, 29%, 17%, and 7%, respectively, of the third positions of degenerate codons. The presumed polyadenylylation signal sequence A-A-T-A-A-A (34) appears at nucleotide position 1419. The precise start of polyadenylylation after this signal was not determined because of interference by the G·C and A·T stretches in the sequencing reactions.

RNA 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 (Fig. 4). In another RNA 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 *Ava* II/*Eco*RI fragment was used as a primer for ^{32}P-labeled cDNA synthesis, with electric organ mRNA serving as the template. The longest ^{32}P-labeled cDNA 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 (38).

No significant homology with the CK52g8-encoded protein could be found with any of the 2222 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 *et al.* (39). Kinases listed in the library included *Bacillus stearothermophilus* phosphofructokinase, *Herpes simplex* thymidine kinase, *E. coli* aspartokinase I/homoserine dehydrogenase I, *E. 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 (40), not listed in the library, was also examined and found to have no significant sequence homology.

DISCUSSION

The RNA 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 this inferred protein has CK activity, its amino acid sequence is 68% homologous to the reported partial sequence of rabbit muscle CK (14), and its composition (Table 1) is very close to the compositions of rabbit and chicken CKs reported by Eppenberger *et al.* (12). The similarities in composition of these sequences are reflected in low difference index values (41) of 9.1, 7.6, 9.7, and 8.6 for comparisons of the CK52g8-inferred protein composition with the compositions of chicken brain, chicken muscle, rabbit brain, and rabbit muscle, respectively. Thus, the CK52g8-inferred protein is certainly related to CK proteins.

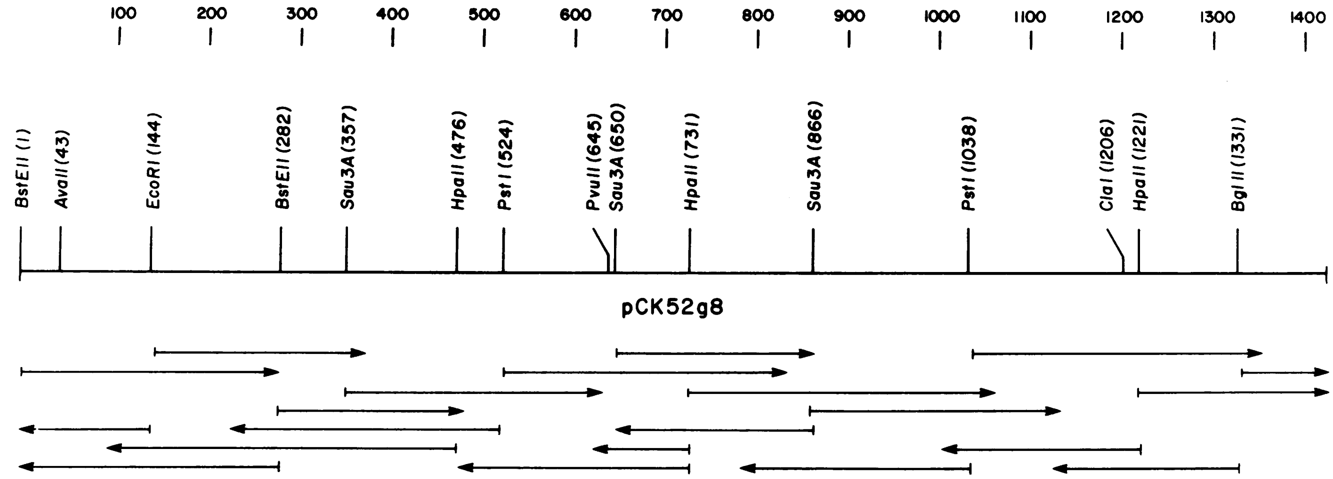

FIG. 2. Partial restriction map and sequencing strategy of CK52g8. The position from the 5′ end of each restriction endonuclease site listed is given in parentheses. Each arrow represents at least one dideoxy sequencing reaction using separate DNA fragments of CK52g8 subcloned into bacteriophage M13. Length of each arrow indicates length of sequence determination for each reaction.

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 (42). The nucleotide sequence in this region contains elements (adenine at position 87, cytidine at positions 85, 88, and 89) that have been shown (43, 44) 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 methionine 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 (45) of the first 40 amino-terminal residues of rat brain CK, which indicate the presence of leucine residues at positions 11 and 22, with no methionine. If the first methionine encoded by CK52g8 is removed post-translationally, the leucine occupies positions 11 and 22 of the CK52g8-encoded protein as well. Since the

```
                                                                                   met pro phe gly asn thr his     7
GGTCACCCACACCAGCGGTAGTTCCAGCACCAAGCAGGACAAGGTCCAGAGTGGTTCACCGTGCGCCAGGAGTCAGCCAACCTCCAACC   ATG CCT TTC GGA AAC ACT CAC   110

asn lys trp lys leu asn tyr ser ala ala glu glu phe pro asp leu ser lys his asn asn his met ala lys ala leu thr leu asp    37
AAT AAA TGG AAG CTG AAC TAT TCG GCG GCG GAA GAA TTC CCC GAC CTC AGC AAG CAC AAC AAC CAC ATG GCC AAG GCT TTA ACC CTG GAC   200

ile tyr lys lys leu arg asp lys glu thr pro ser gly phe thr leu asp asp ile ile gln thr gly val asp asn pro gly his pro    67
ATC TAC AAG AAA CTT CGG GAC AAG GAG ACT CCA AGT GGC TTC ACC CTC GAT GAT ATC ATC CAG ACA GGA GTG GAC AAC CCA GGT CAC CCC   290

phe ile met thr val gly cys val ala gly asp glu glu cys tyr glu val phe lys asp leu phe asp pro val ile glu asp arg his    97
TTC ATC ATG ACC GTG GGC TGC GTG GCT GGC GAT GAG GAA TGC TAC GAG GTT TTC AAG GAC CTG TTC GAT CCC GTC ATT GAG GAC CGC CAC   380

gly gly tyr lys pro thr asp lys his lys thr asp leu asn gln glu asn leu lys gly gly asp asp leu asp pro asn tyr val leu   127
GGT GGC TAC AAA CCA ACT GAC AAG CAC AAG ACT GAC CTG AAC CAG GAG AAC CTG AAG GGC GGC GAT GAC CTC GAC CCG AAT TAC GTC CTG   470

ser ser arg val arg thr gly arg ser ile lys gly ile ala leu pro pro his cys ser arg gly glu arg arg leu val glu lys leu   157
AGC AGC CGG GTG CGC ACT GGC CGC AGC ATC AAG GGC ATC GCC CTG CCT CCT CAC TGC AGC CGC GGG GAG CGC CGT CTG GTT GAG AAG CTC   560

cys ile asp gly leu ala thr leu thr gly glu phe gln gly lys tyr tyr pro leu ser ser met ser asp ala glu gln gln gln leu   187
TGC ATA GAC GGT CTC GCC ACC TTG ACG GGC GAG TTC CAG GGC AAG TAC TAC CCC CTC TCC TCC ATG TCT GAT GCA GAG CAG CAG CAG CTG   650

ile asp asp his phe leu phe asp lys pro ile ser pro leu leu leu ala ser gly met ala arg asp trp pro asp gly arg gly ile   217
ATC GAT GAC CAC TTC CTG TTT GAC AAA CCC ATC TCT CCT CTG CTT CTC GCC TCT GGC ATG GCT CGG GAC TGG CCC GAT GGC CGG GGC ATT   740

trp his asn asn asp lys thr phe leu val trp val asn glu glu asp his leu arg val ile ser met gln lys gly gly asn met lys   247
TGG CAT AAC AAC GAC AAG ACC TTC CTG GTC TGG GTC AAC GAG GAG GAC CAC CTC CGA GTC ATC TCG ATG CAG AAA GGT GGC AAC ATG AAG   830

                                                                    ala gly pro his phe met glu asn his leu
glu val phe arg arg phe cys val gly leu lys lys ile glu asp ile phe val lys ala gly arg gly phe met trp asn glu his leu   277
GAG GTC TTC AGG CGC TTC TGC GTT GGT CTG AAG AAG ATC GAG GAC ATT TTC GTG AAG GCT GGC CGT GGC TTC ATG TGG AAC GAG CAC CTG   920

gly tyr val leu thr cys pro ser asn leu gly thr gly leu arg
gly tyr val leu thr cys pro ser asn leu gly thr gly leu arg gly gly val his val lys ile pro his leu cys lys his glu lys   307
GGC TAC GTC CTG ACC TGC CCG TCC AAC CTG GGC ACT GGC CTC CGT GGT GGT GTC CAC GTG AAA ATC CCT CAC CTC TGC AAG CAC GAG AAG  1010

phe ser glu val leu lys arg thr arg leu gln lys arg gly thr gly gly val asp thr ala ala val gly ser ile tyr asp ile ser   337
TTC AGC GAG GTC CTC AAG AGA ACG AGG CTG CAG AAA CGT GGG ACA GGT GGA GTG GAT ACC GCA GCG GTT GGC AGC ATC TAT GAC ATC TCC  1100

asn ala asp arg leu gly phe ser glu val glu gln val gln met val val asp gly val lys leu met val glu met glu lys arg leu   367
AAC GCC GAC CGT CTG GGC TTC TCC GAG GTG GAA CAG GTC CAG ATG GTG GTG GAC GGT GTG AAG CTG ATG GTC GAG ATG GAG AAG AGG CTG  1190

glu asn gly lys ser ile asp asp leu met pro ala gln lys                                                                 381
GAA AAT GGG AAA AGC ATC GAT GAC CTG ATG CCG GCT CAG AAG TAG ACCTTGGGTTGGCTGGGTGCCTGCCACTCTGAGATGCCTTGAAATATCACAGGTCGCGAA 1295

CTTTGAACTTTCCCACTCCAATCTTTCTTGGCCACAGATCTCGTGTCTCAAATGAGGAAGCAGAAGGTTTGGTTTCATCACATTCAGATTTGCTAGACACAATTTTAACCTTGATGACAC 1415

ATTAATAAAATAT                                                                                                           1428
```

FIG. 3. 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 homology between predicted protein sequence and CK sequence from rabbit muscle, the 25 amino acid sequence reported by Atherton *et al.* (14) 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.

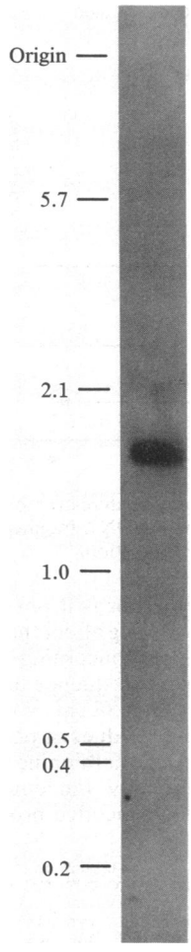

FIG. 4. RNA blot hybridization analysis of electric organ mRNA, using ^{32}P-labeled *Eco*RI/*Bgl* II fragment as probe. Five micrograms of electric organ poly(A) RNA was prepared for hybridization according to Thomas (35). Hybridization was done for 15 hr at 42°C in a 13-ml vol containing the following: 5× Denhardt's solution (36), 0.75 M NaCl, 0.075 M Na_3 citrate, 20 mM $NaPO_4$ (pH 6.5), 200 μg of sheared boiled calf thymus DNA per ml, and 2×10^6 cpm of boiled ^{32}P-labeled *Eco*RI/*Bgl* II fragment DNA. The blot was then washed in 30 mM NaCl, 3 mM Na_3 citrate, pH 7, containing 0.1% $NaDodSO_4$ at 42°C and was then autoradiographed. Length markers on the left represent base pairs $\times 10^{-3}$ and are derived from 9000 cpm of ^{32}P-labeled *Hin*dIII-cut phage PM2 DNA (37), which was glyoxalated (35) and electrophoresed in another lane of the same gel.

carboxyl-terminal dipeptide, Gln-Lys, predicted by CK52g8 is identical to that for rabbit muscle CK (33), 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* (46) and to the molecular weight of the *in vitro* translation products that can be arrested by hybridization of CKABh9 and CK52g8 DNAs. Barrantes *et al.* (46) 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, because all of these clones had inserts with the same restriction sites as CK52g8 and because a single band appears in the RNA blot hybridizations with both CK52g8 and CKABh9 insert probes. The brain form of CK shown by Barrantes *et al.* (46) 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

Table 1. Comparison of amino acid composition with chicken and rabbit CKs

Amino acid	Torpedo Electric organ $CK52_g8$	Chicken Brain	Chicken Muscle	Rabbit Brain	Rabbit 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	ND	ND	ND	ND
Trp	5	ND	ND	ND	ND
Asn	17	ND	ND	ND	ND
Gln	11	ND	ND	ND	ND

Composition of chicken and rabbit CKs are calculated from Eppenberger *et al.* (12), assuming that dimer subunits are identical. Units for all compositions are mol per mol of subunit. ND, not determined.

for the CK52g8-encoded protein by calculating the weighted average of pK values for the 31 aspartic acid, 24 glutamic acid, 31 lysine, and 20 arginine (pK = 3.86, 4.25, 10.53, and 12.48, respectively) residues. Identification of isozyme type possibly must await expression of the CK52g8-encoded protein and an experimental determination of the pI value, 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.* (14) for protein sequencing contains a cysteine residue, which selectively reacts with sulfhydryl-modifying reagents to affect enzyme activity. This cysteine lies in a 17-amino acid stretch that has 100% homology (Fig. 3) 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 β turns occur (47), but knowing whether there is a β turn at this position may require elucidation of the crystal structure of CK (48–51).

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 structure and function of the enzyme. By deriving the protein sequence of some mammalian CK and comparing its sequence with that obtained here for *Torpedo*, useful insights into important conserved regions should be obtained. By expressing the CK52g8 sequence in a heterologous cell, studies of the enzyme function should be possible using techniques such as site-specific mutagenesis.

After completion of this manuscript, a paper by Benfield *et al.* (52) appeared describing the partial protein sequence and partial cDNA sequence for rat muscle CK. There is 83% homology comparing the 295 amino acids obtained by their techniques with the same region of the CK52g8-inferred protein, including a stretch of 28 identical amino acids in the region of the reactive cysteine.

We thank Drs. Peter Seeburg, Charles Craik, and Steve Gardell for helpful suggestions concerning the cDNA library construction and M13 sequencing; Dr. Guy Cathala for suggestions concerning the RNA blot analysis; and Drs. Hugo Martinez and Ari Azhir for providing the sequence analysis computer programs. We thank Mrs. Eva Kuenzli for her dedicated assistance in preparing materials for the sequence analysis and Ms. D. Glaister and Ms. Hoa Nguyen for technical help. We also thank Drs. I. D. Kuntz and George Kenyon for their encouragement and support. This work was supported by U.S. Army Medical Research and Development Command Contract DAMD17-82-C-2143 (J.D.B.), National Institutes of Health Grant GM 31497 (I. D. Kuntz), and U.S. Public Health Service Grant AM17323 (G. L. Kenyon).

1. Bessmaṅ, S. P. & Geiger, P. J. (1981) *Science* **211,** 448–452.
2. Hammes, G. G. & Huest, J. K. (1969) *Biochemistry* **8,** 1083–1094.
3. Grossman, S. H. (1983) *J. Neurochem.* **41,** 729–736.
4. Jacobs, H. K. & Kuby, S. A. (1970) *J. Biol. Chem.* **245,** 3305–3314.
5. James, T. L. & Cohn, M. (1974) *J. Biol. Chem.* **249,** 2599–2604.
6. James, T. L. (1976) *Biochemistry* **15,** 4724–4730.
7. Borders, C. L., Jr., & Riordan, J. F. (1975) *Biochemistry* **14,** 4699–4704.
8. Vasak, M., Nagayama, K., Wuthrich, K., Mertens, M. C. & Kagi, J. H. R. (1979) *Biochemistry* **18,** 5050–5055.
9. Cook, P. F., Kenyon, G. L. & Cleland, W. W. (1981) *J. Biol. Chem.* **20,** 1204–1210.
10. Smith, D. J. & Kenyon, G. L. (1974) *J. Biol. Chem.* **249,** 3317–3318.
11. Kenyon, G. L. & 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. Eppenberger, H. E., Dawson, D. M. & Kaplan, N. O. (1967) *J. Biol. Chem.* **242,** 204–209.
13. Watts, D. C. (1973) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), Vol. 8, pp. 383–455.
14. Atherton, R. S., Laws, J. F., Miles, B. J. & Thomson, A. R. (1970) *Biochem. J.* **120,** 589–600.
15. Mendez, B., Valenzuela, P., Martial, J. & Baxter, J. D. (1980) *Science* **209,** 695–697.
16. Cathala, G., Savouret, J. F., Mendez, B., West, B., Karin, M., Martial, J. A. & Baxter, J. D. (1983) *DNA* **2,** 329–335.
17. Duguid, J. R. & Raftery, M. A. (1973) *Biochemistry* **12,** 3593–3597.
18. Rosenberg, U. B., Kunz, B., Frischauf, A., Lehrach, H., Mahr, R., Eppenberger, H. M. & Perriard, J. C. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6589–6592.
19. Schweinfest, C. W., Kwiatkowski, R. W. & Dottin, R. P. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4997–5000.
20. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
21. Cooke, N. E., Coit, D., Weiner, R. I., Baxter, J. D. & Martial, J. A. (1980) *J. Biol. Chem.* **255,** 6502–6510.
22. Land, H., Grez, M., Hauser, H., Lindenmaier, W. & Schutz, G. (1981) *Nucleic Acids Res.* **9,** 2251–2266.
23. Peacock, A. C. & Dingman, C. W. (1967) *Biochemistry* **6,** 1818–1827.
24. Paterson, B. M., Roberts, B. E. & Kuff, E. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4370–4374.
25. Pelham, H. R. B. & Jackson, R. J. (1976) *Eur. J. Biochem.* **67,** 247–256.
26. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
27. Kronenberg, H. M., McDevitt, B. E., Majzoub, J. A., Nathans, J., Sharp, P. A., Potts, J. T. & Rich, A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4981–4985.
28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
29. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
30. Gergen, J. P., Stern, R. H. & Wensink, P. C. (1979) *Nucleic Acids Res.* **7,** 2115–2136.
31. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
32. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114,** 193–197.
33. Olson, O. E. & Kuby, S. A. (1964) *J. Biol. Chem.* **239,** 460–467.
34. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
35. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.
36. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 131.
37. Streeck, R. E. & Gebhardt, C. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360,** 529–532.
38. Brawerman, G. (1981) *CRC Crit. Rev. Biochem.* **10,** 1–38.
39. Martinez, H. M., Katzung, B. & Farrah, T. (1983) *Sequence Analysis Program Manual* (Biomathematics Computation Laboratory, University of California, San Francisco).
40. Burke, R. L., Tekamp-Olson, P. & Najarian, R. (1983) *J. Biol. Chem.* **258,** 2193–2201.
41. Metzger, H., Shapiro, M. B., Mosimann, J. E. & Vinton, J. E. (1968) *Nature (London)* **219,** 1166–1168.
42. Kozak, M. (1983) *Microbiol Rev.* **47,** 1–45.
43. Kozak, M. (1984) *Nature (London)* **308,** 241–246.
44. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
45. Kaye, A. M. (1983) in *Regulation of Gene Expression by Hormones*, ed. McKerns, K. W. (Plenum, New York), pp. 103–128.
46. Barrantes, F. J., Mieskes, G. & Wallimann, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5440–5444.
47. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 222–245.
48. McPherson, A., Jr. (1973) *J. Mol. Biol.* **81,** 83–86.
49. Burgess, A. N., Liddell, J. M., Cook, W., Tweedle, R. M. & Swann, I. D. A. (1978) *J. Mol. Biol.* **123,** 691–695.
50. Takasawa, T., Fukushi, K. & Shiokawa, H. (1981) *J. Biochem.* **89,** 1619–1631.
51. Gilliland, G. L., Sjolin, L. & Olsson, G. (1983) *J. Mol. Biol.* **170,** 791–793.
52. Benfield, P., Zivin, R., Shearman, C., Graf, D., Henderson, L., Oroszlan, S. & Pearson, M. (1984) *Exp. Biol. Med.* **9,** 187–194.