The phosphocreatine shuttle of sea urchin sperm: Flagellar creatine kinase resulted from a gene triplication

D. D. Wothe*, H. Charbonneau, and B. M. Shapiro

Department of Biochemistry SJ-70, University of Washington, Seattle, WA 98195

ABSTRACT   TCK, the creatine kinase (ATP:creatine N-phosphotransferase) from sperm flagella of the sea urchin *Strongylocentrotus purpuratus*, is a M_r 145,000 axonemal protein that is employed in energy transport. Its amino acid sequence was obtained by analysis of fragments from cyanogen bromide digestion and by sequencing cDNA clones from two sea urchin testis libraries. TCK contains three complete but nonidentical creatine kinase segments joined by regions of sequence that are not creatine kinase-like and flanked by unique amino and carboxyl termini. Each creatine kinase segment is homologous to vertebrate creatine kinases of both muscle and brain types, and all three repeats contain the essential active-site cysteine. The sequence differences among repeats suggest an ancient gene triplication, around the time of the chordate–echinoderm divergence. The echinoderm, with a unique creatine kinase in sperm, arginine kinase in eggs, and both phosphagen kinases in somatic cells, may represent a preserved branch point in evolution, and TCK may be a relic of this event.

Structurally similar creatine kinases (CKs; EC 2.7.3.2) are found in many invertebrates and vertebrates (1), existing as dimers or octamers of subunit $M_r \approx 40,000$ (2, 3). However, an echinoderm-specific CK is significantly larger: that purified from the flagellum of *Strongylocentrotus purpuratus* sperm is monomeric, with an estimated M_r of 145,000 (4). This CK participates in an energy shuttle that utilizes phosphocreatine to transfer the energy from ATP generated by the mitochondrion in the sperm head to dynein in the distal portions of the flagellum (5). A mitochondrial CK transphosphorylates creatine and ATP, maintaining the high levels of ADP that permit maximum respiration (5, 6). As phosphocreatine diffuses along the flagellum, the flagellar creatine kinase (TCK) regenerates ATP for use as a substrate by dynein. Specific inhibition of CKs by low levels of 1-fluoro-2,4, dinitrobenzene attenuates flagellar beating (6), indicating that the energy shuttle is essential for normal flagellar motion. TCK specifically associates with the axoneme and may bind directly to polymerized microtubules (7). To ascertain the primary structure of this unusually large CK, and to provide tools to examine its association with the axoneme, we have purified TCK, isolated and sequenced six CNBr peptides, and obtained the sequence of clones from two independent cDNA libraries.†

MATERIALS AND METHODS

S. purpuratus were obtained at low tide along the northern shore of the Olympic Peninsula in Washington. Restriction endonucleases were purchased from United States Biochemical and Boehringer Mannheim. All other reagents were of the highest grade available.

Sperm were obtained and TCK was purified as described (4). TCK (46 mg) was dialyzed against 1.44 M Tris (pH 8.6)

and S-carboxymethylated (8); the sample was then dialyzed against 70% formic acid and cleaved at methionine residues with CNBr (9). CNBr peptides were lyophilized, dissolved in 6 M guanidinium chloride buffered to pH 6.0 with 10 mM phosphate, and separated into size classes by size-exclusion HPLC, using two TSK-G3000SW columns (LKB) in series and a buffer containing 6 M guanidinium chloride in 10 mM phosphate (pH 6.0) at an elution rate of 0.5 ml/min. Individual peptides were separated from each size class by loading onto either a SynChropak RP-P C_8 or a SynChropak RP-P C_{18} column (SynChron) and eluting with a 0–60% acetonitrile gradient (0.5%/min) in aqueous 0.085% trifluoroacetic acid at 1 ml/min.

Amino acid analysis of CNBr peptides was performed with a Waters system using the Pico·Tag derivatization protocol (10). Automated Edman degradations were performed in a Beckman 890C sequencer and derivatives were identified by HPLC (11).

A λgt11 cDNA library was generously provided by E. Davidson (California Institute of Technology, Pasadena, CA). This library was constructed by David Nishioka, by ligating inserts from a randomly primed λgt10 library made using *S. purpuratus* testis mRNA (12) into λgt11 vector arms. Plaques (620,000) were screened with antisense oligodeoxynucleotides (GTTGCAGAAACGCTTGAAGAC, GTTGAT-GTCCTCTCCCTTCAT) designed using sequences from TCK CNBr peptides 1 and 6 (Fig. 1). The oligonucleotide sequence synthesized was based on codon-bias estimates for *S. purpuratus* from available cDNA sequences (13). Oligonucleotides were end-labeled with ^{32}P by phage T4 polynucleotide kinase and used to screen filters representing $\approx$25,000 λgt11 plaques each, according to standard techniques (14). A second, oligo(dT)-primed λZAP (Stratagene) cDNA library made using *S. purpuratus* testis mRNA by Dangott et al. (15) was generously provided by D. Garbers (Vanderbilt University Medical Center, Nashville, TN). This library was screened (164,000 plaques) with a TCK cDNA fragment from the λgt11 library labeled to a specific activity of 10^8 cpm/μg by the random priming method (16). cDNA clones that hybridized to both oligonucleotide probes were rescreened until only hybridizing phage were present, and phage DNA was isolated (14). Inserts were cut out with *Eco*RI and cloned into pBluescript KS(−) (Stratagene).

Single- and double-stranded DNA sequencing was performed using a Sequenase 2.0 kit (United States Biochemical), which uses the dideoxynucleotide chain-termination procedure of Sanger et al. (17). The amino acid sequence deduced from the cDNA sequence was compared to the Protein Identification Resource protein sequence data bank (release 21) and the GenBank data base (release 62.0).

CMCK 246-271   M K E V F R R F C V G L K K I E E I F K K A G H P F
Peptide 1   M K N V F K R F C N G L N K V E N A L K A K G Y E F
Peptide 2   M R A V F K R F C E G L Q - F E
Peptide 3   M K R V F E R F C D G L K K V E D S I K S K G Y

CMCK 74-90   C V A G D E E S Y E V F K D L F D
Peptide 4   M V A G D E E S Y D V F A D I F D

CMCK 207-220   M A R D W P D A R G I W H N
Peptide 5   M A R D W P Q G R G I W H N

Peptide 6   M K G E D I N S L L P E K L R

FIG. 1.   Amino-terminal sequences (including the inferred me-thionines) for CNBr peptides from TCK aligned with regions of the chicken muscle CK (CMCK) sequence obtained from the cDNA sequence. The position of the CMCK sequence is indicated before each fragment. Peptides 1–5 are all similar to CMCK sequences. The underlined portions of peptides 1–3 are identical to each other. Each of the residues common to all three repeats is identical to equivalent positions in the CMCK sequence.

RESULTS AND DISCUSSION

Protein Sequencing. CNBr peptides of TCK were separated into nine size classes and each pooled size population was purified further by reverse-phase HPLC (data not shown). Amino acid sequence was obtained from six of the purified peptides and compared to sequences deduced from other CK cDNAs. Fig. 1 shows the peptide sequences aligned with the indicated portion of the protein sequence from chicken muscle CK cDNAs.

TCK peptides 1–5 were similar to regions of vertebrate CKs. In addition, peptides 1–3 were similar to one another, which indicated either a triplication of CK units or polymorphism of TCK. Peptide 6 had no significant sequence similarity to CKs.

cDNA Sequence. Two partial TCK cDNA clones were identified in the λgt11 library by screening with the two synthetic oligonucleotides, and several 4.0-kilobase (kb) clones were identified in the λZAP library by using the two partial cDNA clones. Restriction maps for the TCK cDNA,

a 4.0-kb clone, and the two partial clones are presented in Fig. 2a. The nucleotide sequence and the deduced amino acid sequence are shown in Fig. 3. The longest clone identified was 3985 base pairs (bp), excluding the poly(A) tail, and contained a 3522-bp open reading frame beginning at position 95 with a consensus translation initiation sequence (18, 19). It is preceded by a potential translation initiation sequence beginning at position 70 that is almost certainly not recognized by 40S ribosomal subunits (18, 19). The single long open reading frame is followed by a 366-bp 3' untranslated sequence that includes an imperfect polyadenylylation consensus sequence (AATACA) at position 3961 followed in 20 bases by poly(A) sequence.

Examination of the translation of the 3522-bp open reading frame reveals three nonidentical CK repeats flanked by unique regions (Fig. 2b; regions A and D). Each repeat is joined to the next by unique sequence (regions B and C). The protein sequence predicted by translation of the cDNA sequence contains all six peptide sequences in frame (Figs. 2b and 3). The protein is 1174 amino acids long with a predicted molecular weight of 130,877 and a pI of 5.85. All six peptide sequences are identical to the translations at amino acids 287–312, 488–504, 660–675, 995–1008, 1034–1057, and 1156–1170. However, the translation of clone WB2.1 of Fig. 2 had an amino acid different from that of peptide 6 at position 1166, raising the possibility that this position may be polymorphic.

The triplication of CK units explains the large size of this monomeric CK. To our knowledge, this is the first example of an endotriplication of potentially functional enzyme units. Characterization of other proteins has demonstrated that gene duplication to form homologous domains is a common occurrence in evolution. For example, human CD45 has two phosphatase domains formed by a repeat of exons 17–25 (20). Both repeats are part of the cytoplasmic domain and may be involved in signaling in leukocytes. Precursor proteins such as profilaggrin, which consists of a 250-amino acid filaggrin unit repeated 12 times in the primary transcript, are partially cleaved to form intermediates that, after use, are further
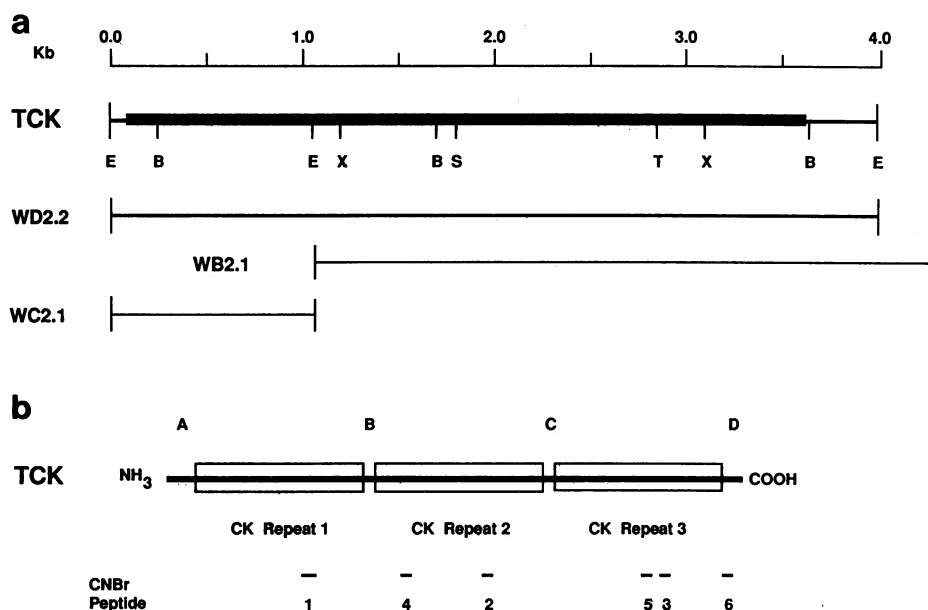
FIG. 2.   (a) Restriction endonuclease map for the TCK cDNA. Heavy line indicates the region of the cDNA that is translated. Restriction endonuclease sites are shown below the line: E, *Eco*RI; B, *Bam*HI; S, *Sst* I; T, *Sst* II; X, *Xho* I. Identical restriction sites were found in the three cDNAs presented below the TCK cDNA. Clones WB2.1 and WC2.1 were selected from the λgt11 cDNA library. Clone WD2.2 was selected from the λZAP cDNA library and contains the entire coding sequence. (b) Line diagram of the TCK protein sequence with the three repeats boxed. The non-CK amino terminus is 60 amino acids long (A). Each repeat is 354 amino acids long. The first and second repeats are separated by a 19-amino acid linking region (B). The second and third are separated by a 20-amino acid linking region (C). The last repeat is followed by a 13-amino acid stretch of non-CK-like sequence (D). The positions of the six CNBr peptides that were sequenced are indicated by the numbered bars.

```
    1  ccctaaacttcaatcacacttctaatcaagaaaggaacaaatttgtattaattatttcattttctacatatgcattgtaatcagtgacccagca

          M  G  C  A  A  S  S  Q  Q  T  T  A  T  G  G  Q  P  A  A  G  E  K  A  N  P  A  P  A  N  N  N  P  N  A  A   35
   95  ATGGGATGCGCAGCATCGTCACAACAAACGACGGCCACAGGTGGTCAACCAGCTGCTGGTGAAAAGGCAAACCCGGCCCCCGCTAATAACAATCCAAACGCAGCA
          N  K  A  E  T  T  G  A  A  E  E  L  T  K  E  S  E  P  F  V  E  P  D  P  N
  200  AACAAAGCAGAAACGACAGGAGCAGCTGAAGAACTCACTAAAGAATCTGAACCATTTGTGGAACCGGATCCCAAT                               60
                                                                      →Y  P  D  L  S  K  H  N  N  Y             70
  275                                                                TACCCTGACCTAAGCAAGCACAACAATTAC
          L  A  E  S  L  T  P  S  I  Y  N  K  I  C  N  L  R  T  L  S  G  Y  S  V  D  G  C  M  Q  T  G  V  D  N  P  105
  305  CTGGCAGAATCTTTAACGCCTTCCATCTACAACAAAATCTGTAACCTTCGCACACTTTCTGGTTACTCTGTGGATGGTTGCATGCAGACCGGCGTGGACAATCCA
          G  H  P  F  I  K  T  V  G  L  V  A  G  D  E  E  C  Y  D  L  F  A  D  L  F  D  P  T  I  D  K  R  H  N  G  140
  410  GGACATCCTTTCATCAAGACTGTCGGATTGGTGGCTGGGGATGAAGAGTGTTACGATCTCTTTGCTGACCTCTTTGATCCAACTATTGACAAAAGACACAACGGA
          Y  P  R  N  A  K  H  T  T  D  L  N  P  D  H  L  K  G  G  D  D  F  D  P  K  Y  V  L  S  C  R  V  R  T  G  175
  515  TATCCCCGCAATGCAAAGCACACAACTGATCTCAACCCTGATCATCTCAAGGGTGGTGATGACTTTGATCCAAAGTATGTCCTTTCATGCAGAGTTCGTACAGGC
          R  C  I  R  G  Y  G  L  P  P  H  C  T  R  A  E  R  R  D  V  E  K  V  C  K  D  A  L  A  T  L  D  G  P  L  210
  620  CGCTGCATTAGAGGGTACGGCCTCCCACCACATTGCACGAGGGCTGAACGTCGAGACGTTGAGAAAGTTTGCAAAGACGCACTGGCAACTCTAGATGGGCCACTC
          K  G  T  Y  Y  P  L  T  G  M  T  E  E  M  Q  D  K  L  I  A  D  H  F  L  F  D  K  P  V  S  P  L  L  M  S  245
  725  AAAGGCACGTATTACCCACTAACGGGGATGACAGAGGAAATGCAGGACAAACTAATTGCAGATCATTTCCTATTCGATAAGCCCGTTTCTCCTTTGCTTATGTCA
          A  R  M  A  R  D  W  P  D  G  R  G  I  W  H  N  A  D  K  N  F  L  V  W  I  N  E  E  D  H  T  R  V  I  S  280
  830  GCTCGCATGGCCCGTGATTGGCCTGATGGACGTGGGATCTGGCACAATGCCGATAAAAACTTCCTGGTTTGGATAAATGAGGAAGATCACACTCGGGTTATTTCC
          M  E  T  S  G  N  M  K  N  V  F  K  R  F  C  N  G  L  N  K  V  E  N  A  L  K  A  K  G  Y  E  F  S  W  N  315
  935  ATGGAAACAAGTGGCAACATGAAGAATGTCTTTAAACGATTCTGTAATGGTCTCAACAAAGTCGAGAATGCTCTCAAAGCAAAAGGGTATGAATTCTCATGGAAT
          E  H  L  G  Y  V  L  T  C  P  S  N  L  G  T  G  V  R  A  G  V  H  I  K  I  P  L  F  S  K  H  A  G  F  E  350
 1040  GAGCATCTTGGGTATGTGCTCACTTGTCCATCCAACCTTGGTACAGGTGTACGTGCAGGAGTCCACATCAAGATTCCTCTTTTCAGCAAGCACGCTGGATTTGAA
          S  I  L  K  H  Y  R  L  Q  K  R  G  T  G  G  V  D  T  A  S  T  D  G  T  F  D  I  S  N  L  D  R  L  G  T  385
 1145  TCCATTCTGAAGCATTATCGTCTCCAGAAACGTGGCACAGGAGGCGTGGATACGGCATCCACTGACGGCACGTTCGACATTTCAAACCTAGATCGTCTTGGCACT
          S  E  V  Q  Q  V  Q  S  V  V  D  G  V  K  K  L  I  E  L  E  K  A  L  E  K  G  S  D  I←S  G  Q  I  P  R  420
 1250  TCTGAAGTACAACAGGTTCAATCCGTTGTTGATGGTGTCAAGAAGCTGATTGAGTTGGAAAAAGCCTTAGAAAAGGGATCAGATATCAGCGGCCAGATTCCCCGG
          D  P  A  I  V  R  A  E  Q  V  K  E  G                                                                   433
 1355  GACCCAGCAATAGTTCGTGCAGAACAAGTGAAAGAAGGA
                                                   →Y  P  D  L  S  K  H  N  N  H  L  A  H  C  L  T  Y  D  I  W  K  S  455
 1394                                             TATCCAGACCTATCCAAACACAACAATCATCTGGCCCCACTGTCTGACATACGACATTTGGAAAAGT
          L  K  D  K  K  T  P  S  G  F  T  L  D  G  C  I  Q  T  G  V  M  N  P  G  H  P  H  I  M  T  V  G  M  V  A  490
 1460  CTCAAGGACAAAAAGACGCCAAGTGGCTTTACTCTGGATGGCTGCATCCAAACGGGAGTCATGAACCCTGGTCATCCTCACATCATGACCGTTGGAATGGTTGCA
          G  D  E  E  S  Y  D  V  F  A  D  I  F  D  P  V  I  D  A  R  H  G  G  Y  P  K  D  A  V  H  V  T  N  I  N  525
 1565  GGAGATGAGGAATCATATGACGTCTTCGCTGATATCTTTGACCCGGTCATTGATGCACGACACGGAGGTTATCCCAAAGATGCCGTCCATGTCACCAACATAAAC
          H  A  D  L  K  G  G  D  N  L  D  P  K  Y  V  L  S  C  R  V  R  T  G  R  S  I  I  G  Y  S  L  P  P  H  C  560
 1670  CATGCAGATCTTAAAGGAGGAGATAACCTGGATCCCAAATATGTTTTGTCTTGCCGTGTTCGCACAGGACGCAGCATCATCGGGTATTCTCTGCCCCCGCATTGC
          T  V  E  E  R  A  A  V  E  T  I  T  I  G  A  L  D  K  F  D  G  D  L  Q  G  K  Y  Y  P  L  E  G  M  S  D  595
 1775  ACCGTGGAAGAGAGAGCTGCCGTAGAGACGATCACGATCGGAGCTCTTGATAAATTTGATGGAGACTTGCAAGGAAAATACTACCCACTGGAAGGAATGTCTGAT
          E  T  Q  T  Q  L  I  D  D  H  F  L  F  D  K  P  V  S  P  L  L  T  A  A  R  M  H  R  D  W  P  Q  G  R  G  630
 1880  GAAACACAAACACAGCTGATAGATGACCATTTCCTCTTCGATAAGCCTGTTTCTCCTCTTCTTACCGCCGCAAGAATGCATCGTGATTGGCCACAGGGAAGAGGA
          I  W  H  N  E  N  K  N  F  L  V  W  V  N  E  E  D  H  I  R  V  I  S  M  E  K  D  G  N  M  R  A  V  F  K  665
 1985  ATTTGGCATAACGAAAACAAGAACTTCTTGGTGTGGGTCAATGAAGAAGATCATATCCGTGTCATCTCTATGGAGAAGGATGGGAATATGCGAGCGGTCTTCAAG
          R  F  C  E  G  L  Q  K  F  E  Q  M  I  K  K  D  G  K  E  F  M  W  N  K  H  L  G  Y  V  L  T  C  P  S  N  700
 2090  CGATTCTGTGAAGGTTTGCAGAAGTTCGAGCAGATGATCAAAAAGGATGGCAAGGAGTTCATGTGGAACAAACATCTGGGTTACGTACTTACTTGTCCGTCTAAC
          L  G  T  G  L  R  A  G  V  H  V  K  L  P  L  L  S  K  Y  P  R  F  D  Q  I  L  R  A  L  R  L  Q  K  R  G  735
 2195  CTGGGAACCGGCCTACGCGCTGGCGTACATGTGAAACTGCCTCTGCTGAGCAAGTATCCGAGATTCGACAGATCCTAAGAGCCCTCCGACTTCAGAAGAGAGGT
          T  G  G  V  D  T  A  S  T  D  G  T  F  D  I  S  N  L  D  R  L  G  S  S  E  V  Q  Q  V  Q  F  V  V  D  G  770
 2300  ACAGGTGGAGTAGATACTGCCTCAACTGATGGAACCTTCGACATCTCGAACCTCGACAGGCTTGGATCATCAGAAGTACAGCAAGTTCAGTTTGTTGTCGATGGC
          V  E  L  L  V  Q  M  E  K  K  L  E  K  G  E  D  I←F  D  I  L  P  Q  Q  C  R  P  K  P  P  I  K  P  F  S  805
 2405  GTTGAACTTCTCGTCCAAATGGAGAAGAAACTTGAAAAGGGCGAGGACATTTTTGATATTCTACCACAGCAATGCCGTCCCAAACCACCAATAAAGCCTTTCAGT
          Y  D                                                                                                    807
 2510  TATGAC
                                            →Y  P  D  F  S  L  H  N  N  W  M  S  K  C  M  T  E  E  I  Y  N  K  L  C  N  L  K  T  K  G  G  V  T  840
 2516                                      TACCCCGATTTCTCTCTGCACAACAACTGGATGTCTAAATGTATGACAGAAGAAATCTACAACAAACTTTGCAACTTGAAGACAAAAGGTGGAGTCACC
          L  N  D  C  I  Q  T  G  I  D  N  P  G  H  P  Y  I  M  T  V  G  L  D  E  E  C  Y  E  V  F  A  P            875
 2615  CTGAATGACTGCATTCAAACTGGCATTGACAATCCTGGACATCCATACATCATGACTGTTGGTTTGGTAGCTGGAGATGAAGAGTGCTACGAGGTTTTCGCTCCT
          L  F  D  P  V  I  S  A  R  H  G  G  Y  A  L  D  A  K  H  P  T  N  L  N  A  A  E  L  K  G  G  D  D  L  D  910
 2720  CTCTTCGACCCGGTGATAAGTGCCAGGCACGGCGGCTACGCTCTTGATGCCAAACACCCTACCAACCTCAATGCAGCAGAACTCAAAGGCGGAGATGATTTGGAC
          P  E  F  V  L  S  C  R  V  R  T  G  R  C  I  R  G  I  A  L  P  P  C  C  T  R  A  E  R  A  E  V  E  K  I  945
 2825  CCTGAGTTCGTCCTATCTTGTCGTGTACGTACTGGTCGTTGCATCCGCGGCTTGGCTCTCCCGCCATGTTGCACCAGAGCAGAACGTGCGGAAGTAGAGAAGATT
          T  T  E  A  L  S  T  L  S  G  P  L  K  G  K  Y  Y  P  L  T  G  M  T  D  E  E  Q  E  K  L  I  E  D  H  F  980
 2930  ACAACAGAAGCATTATCTACCCTAAGTGGACCTCTGAAGGGAAAGTACTACCCACTAACCGGTATGACAGACGAAGAGCAAGAGAAGCTCATCGAAGACCACTTC
          L  F  D  K  P  V  S  P  L  L  L  C  A  N  M  A  R  D  W  P  Q  G  R  G  I  W  H  N  D  E  K  N  F  L  V  1015
 3035  CTCTTTGACAAGCCCGTATCCCCGCTGTTGCTTTGTGCCAATATGGCACGTGACTGGCCCCAGGGTCGTGGCATATGGCACAATGATGAGAAGAATTTCCTTGTC
          W  V  N  E  E  D  H  T  R  V  I  S  M  E  K  S  G  M  K  R  V  F  E  R  F  C  D  G  L  K  K  V  E  D      1050
 3140  TGGGTCAACGAGGAAGATCACACTCGAGTCATCTCCATGGAGAAGAGCGGGAACATGAAGCGAGTGTTTGAGAGGTTCTGCGACGGTCTCAAAAAGGTCGAAGAC
          S  I  K  S  K  G  Y  Q  F  M  W  N  E  H  L  G  Y  V  L  T  C  P  S  N  L  G  T  G  L  R  A  G  V  H  V  1085
 3245  TCCATCAAGTCCAAGGGTTACCAGTTCATGTGGAATGAGCATCTTGGTTATGTACTGACATGTCCATCCAACCTCGGGACGGGACTGCGTGCAGGCGTCCATGTC
          K  V  P  L  L  S  Q  Q  K  I  F  D  S  I  L  D  H  M  R  L  Q  K  R  G  T  G  G  V  D  T  A  S  T  D  G  1120
 3350  AAGGTTCCTTTGTTAAGTCAACAGAAAATCTTTGACTCTATCCTTGATCACATGCGTCTGCAAAAGCGAGGAACAGGAGGTGTAGACACAGCCTCTACTGACGGC
          T  F  D  I  S  N  S  D  R  I  G  F  S  E  V  H  L  V  Q  Q  L  V  D  G  V  K  L  L  V  N  L  E  K  A  L  1155
 3455  ACCTTTGATATCTCCAACTCGGATCGCATCGGTTTCTCCGAAGTCCATCTTGTCCAACAACTCGTAGACGGTGTCAAGCTCCTCGTCAACCTCGAAAAGGCTCTC
          M  K  G  E  D  I←N  S  L  L  P  E  K  L  R  E  D  S  S  *                                                 1174
 3560  ATGAAGGGAGAAGACATCAACAGTCTCCTGCCAGAGAAGCTAAGGGAAGATTCCTCATAA

 3620  agccaggatccacagcttgcagaatagtggccgttgtcgcgcaagtgttataaatgccacctttagaagatgtagcctacacattaaaaaaaat
 3715  acaattagaattgaaatctataaggacgatttgaatataccataaagatttgcccatgatttgaatcaaaataacttcatgctgaacttgtaaa
 3810  tgaaaaagtaatgttagaaaactctgtaaattgggaagatataaaatttcatggccgtttttctttctgatgaagtacactaattatattaattt
 3905  ctcttattaatatgtatggaaactttaatatgtaaaaacaaggaaacgaactgaaatacaattcaaactactttttcattaaaaaaaaaaa
```

FIG. 3.   The entire nucleotide sequence of the longest cDNA clone for TCK, shown with the deduced translation aligned above. Nucleotides are numbered at left and amino acids are numbered at right. The translation initiation codon ATG, the termination codon TAA, and the polyadenylylation signal AATACA are underlined in the nucleotide sequence. Lowercase letters represent 5' and 3' untranslated sequence. The six peptide sequences presented in Fig. 1 are underlined in the protein sequence. The rightward-pointing arrows and offset sequence indicate the start of each of the three repeats. The leftward-pointing arrows indicate the end of each CK repeat, connected by the linking sequence to the next repeat.

cleaved to form monomers (21). The endotriplicated TCK is probably an example of the first sort, creating a protein with three active sites. There is no evidence that the three subunits are cleaved into individual CK units.

Triplicated domains have been seen in two other cases. Ferredoxin, a non-heme iron protein, has three related domains resulting from an ancient gene triplication (22). A rare variant of the hemoglobin-binding protein haptoglobin (Johnson type) contains a partial triplication that may have resulted from the unequal crossing-over of an already duplicated allele (23). A similar mechanism might explain the triplication of this echinoderm CK.

Echinoderm CKs. The only phosphagen kinases of echinoderm sperm are the high molecular weight TCK and a mitochondrial CK (4), whereas eggs have only arginine kinase (24, 25). This situation may have arisen from an early duplication of a primordial phosphagen kinase, which then diverged into creatine and arginine kinases. A later duplication of the CK, followed by unequal crossing-over, could

have resulted in TCK. These several phosphagen kinases might have further evolved for specific functions in each gamete—for example, in producing the phosphocreatine shuttle that facilitates energy transport in sperm.

The high molecular weight form of CK is found in sperm from all echinoderms (24, 25) as well as in some somatic echinoderm tissues. The distribution in somatic tissues varies from one species to another. Thus the echinoderm may represent a preserved branch point in evolution in which the primary phosphagen energy source is in transition from arginine kinase to CK.

Comparisons with Other CKs. The high degree of sequence similarity between TCK and other CKs confirms TCK as a member of the same family. The percent of identical residues when each domain is aligned with the other two domains is presented in Table 1. The percent identity between each repeated domain of TCK and representative muscle (26-28), brain (29), mitochondrial (30), piscine (31), mammalian (26), and avian (27, 28) CKs is also shown (Table 1). For compar-

Table 1. Percent sequence identity among various CKs

CK	% identity					
	TCK1	TCK2	TCK3	HM	HB	HMi
TCK repeat 1 (TCK1)		69	72	65	64	69
TCK repeat 2 (TCK2)	69		71	69	66	65
TCK repeat 3 (TCK3)	72	71		65	69	64
Human muscle (HM)	65	69	65		79	65
Mouse muscle (MM)	67	71	67	95	81	68
Chicken muscle (CM)	65	69	62	89	82	69
Human brain (HB)	64	66	69	79		70
Chicken brain (CB)	65	66	66	80	91	68
Human mitochondrial (HMi)	69	65	63	65	70	
Chicken mitochondrial (CMi)	67	67	64	66	65	86
Torpedo californica electric organ (TC)	65	67	66	83	78	66

The percent sequence identity is shown for each repeat of TCK in comparison with the other two and with 10 vertebrate CK sequences when all are aligned. The alignment was done with the Intelligenetics Suite package of programs and is for the 354 residues comprising each repeat. Similar data are presented for the 381 residues of the human muscle and brain isozymes of CK and for the human mitochondrial CK.

ison, the percent identity to the human isozymes of CK is also presented (32–34). It is noteworthy that while even the human brain and muscle isoforms are 80% identical, no two domains of TCK share more than 72% identity. In fact, the differences between domains of TCK are as great as the differences between TCK and these other CKs. This finding suggests that the gene duplication events took place long ago, around the time of the chordate radiation.

Table 2. A TCK-specific pentapeptide sequence and the reactive cysteine sequence

CK*	Position	Sequence
	Pentapeptide region†	
TCK1	359	QKRGTGGVDTA<u>STDGT</u>FDIS
TCK2	732	. . . . <u>STDGT</u>. . . .
TCK3	1106	. . . . <u>STDGT</u>. . . .
HM	318	. . . . . . A. . . . AVGSV. . V.
MM	318	. . . . AVGAV. . . .
CM	318	. . . . AVGAV. . . .
HB	318	. . . . AVG. V. . V.
CB	318	. . . . AVG. V. . V.
HMi	351	. . . . A. G. V. . . .
CMi	333	. . . . AVADVY. . .
TC	318	. . . . AVGSIY. . .
	Active-site region‡	
TCK1	314	WNEHLGYVLT$\overset{*}{C}$PSNLGTGVR
TCK2	687	. . K. . . . . L.
TCK3	1061	. . . . L.
HM	273	. . Q. . . . . L.
MM	273	. . . . L.
CM	273	. T. . . . . I. . . . . L.
HB	273	. . P. . . . I. . . . . L.
CB	273	. . P. . . . I. . . . . L.
HMi	306	. . . R. . . I. . . . . L.
CMi	288	. . . R. . . . . L.
TC	273	. . . . L.

*See Table 1 for abbreviations.
†A conserved 5-amino acid portion of each TCK repeat differs from vertebrate CKs in the same well-conserved region. The three repeats of TCK were aligned with portions of other CKs. The changed region is underlined.
‡The protein sequences surrounding the putative reactive cysteines (*) of all three repeats are aligned with the same region of vertebrate CKs.

Although the endotriplication was a distant event, features common to CKs are found in each of the three repeats of TCK. Four tryptophan residues exist in conserved positions as determined from previous CK sequences (2, 30) and one of these tryptophan residues is thought to bind the adenosine of ADP at the active site (35). Hossle *et al.* (30) have described a "CK framework" consisting of regions of CK sequence that are conserved among all known sequences. Most regions making up this CK framework are intact in all three TCK repeats, but one region that is characteristic of the vertebrate enzyme differs in TCK and is identical in all three repeated domains (Table 2, upper half).

Although no direct evidence exists as to whether each domain of TCK is catalytically active, residues that are part of the CK active site are conserved in each, including the reactive cysteine (31, 36) and surrounding peptide (Table 2, lower half). If each domain were catalytically active, we would expect the TCK to have a specific activity similar to that of the mammalian enzyme, as well as multiple fluorodinitrobenzene binding sites (37) per molecule of enzyme. Indeed, the specific activity of purified TCK is 235 units/mg, which is comparable to values obtained for mammalian CKs (4). By transfecting cells with appropriately modified cDNA clones, we should be able to directly examine the capacity of each domain to effect catalysis, as well as the sites in the TCK sequence (if any) responsible for association with microtubules (5) of the axoneme.

1. Babbitt, P. C., Kenyon, G. L., Kuntz, I. D., Cohen, F. E., Baxter, J. D., Benfield, P. A., Buskin, J. N., Gilbert, W., Hauschka, S. D., Hossle, H. P., Ordahl, C. P., Pearson, M. L., Perriard, J. C., Pickering, L., Putney, S., West, B. L. & Ziven, R. A. (1986) *J. Protein Chem.* 5, 1–14.
2. Bessman, S. P. & Carpenter, C. L. (1985) *Annu. Rev. Biochem.* 54, 831–862.
3. Watts, D. C. (1973) *The Enzymes*, ed. Boyer, P. D. (Academic, New York), 3rd Ed., Vol. 8, pp. 383–455.
4. Tombes, R. M. & Shapiro, B. M. (1987) *J. Biol. Chem.* 262, 16011–16019.
5. Tombes, R. M. & Shapiro, B. M. (1985) *Cell* 41, 325–334.
6. Tombes, R. M., Brokaw, C. J. & Shapiro, B. M. (1987) *Biophys. J.* 52, 75–86.
7. Tombes, R. M., Farr, A. & Shapiro, B. M. (1988) *Exp. Cell Res.* 178, 307–317.
8. Crestfield, A. M., Moore, S. & Stein, W. H. (1963) *J. Biol. Chem.* 238, 622–627.
9. Gross, E. (1967) *Methods Enzymol.* 11, 238–242.
10. Biddlengmeyer, B. A., Cohen, S. A. & Tarvin, T. L. (1984) *J. Chromatogr.* 336, 93–104.
11. Takio, K., Smith, S. B., Walsh, K. A., Krebs, E. G. & Titani, K. (1983) *J. Biol. Chem.* 258, 5531–5536.
12. Gao, B., Klein, L. E., Britten, R. J. & Davidson, E. H. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8634–8638.
13. Maruyama, T., Gojobori, T., Aota, S. & Ikemura, T. (1986) *Nucleic Acids Res.* 14, 151–197.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
15. Dangott, L. J., Jordan, J. E., Bellet, R. A. & Garbers, D. L. (1989) *Proc. Natl. Acad. Sci. USA* 86, 2128–2132.
16. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* 132, 6–13.

Biochemistry: Wothe *et al.*

Proc. Natl. Acad. Sci. USA 87 (1990)   5207

17. Sanger, F., Miklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

18. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8132.

19. Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241.

20. Hall, L. R., Streuli, M., Schlossman, S. F. & Saito, H. (1988) *J. Immunol.* **141**, 2781–2787.

21. Resing, K. A., Walsh, K. A., Jaugen-Scofield, J. & Dale, B. A. (1989) *J. Biol. Chem.* **264**, 1837–1845.

22. Dixon, G. H. (1966) *Essays Biochem.* **2**, 148–197.

23. Oliviero, S., DeMarchi, M., Carbonara, A. O., Bernini, L. F., Bensi, G. & Raugei, G. (1985) *Hum. Genet.* **71**, 49–52.

24. Tombes, R. M. & Shapiro, B. M. (1989) *J. Exp. Zool.* **251**, 82–90.

25. Ratto, A., Shapiro, B. M. & Christen, R. (1989) *Eur. J. Biochem.* **186**, 195–203.

26. Buskin, J. N., Jaynes, J. B., Chamberlain, J. S. & Hauschka, S. D. (1985) *J. Mol. Evol.* **22**, 334–341.

27. Ordahl, C. P., Evans, G. L., Cooper, T. A., Kunz, G. & Perriard, J. C. (1984) *J. Biol. Chem.* **259**, 15224–15227.

28. Kwiatkowski, R. W., Schweinfest, C. W. & Dottin, R. P. (1984) *Nucleic Acids Res.* **12**, 6925–6934.

29. Hossle, J. P., Rossenberg, U. B., Schafer, B., Eppenberger, H. M. & Perriard, J. C. (1986) *Nucleic Acids Res.* **14**, 1449–1463.

30. Hossle, J. P., Schlegel, J., Wegmann, G., Wyss, M., Bohlen, P., Eppenberger, H. M., Wallimann, T. & Perriard, J. C. (1988) *Biochem. Biophys. Res. Commun.* **151**, 408–416.

31. West, B. L., Babbitt, P. C., Mendez, B. & Baxter, J. D. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7007–7011.

32. Nigro, J. M., Schweinfest, C. W., Rajkovic, A., Pavlovic, J., Jamal, S., Dottin, R. P., Hart, J. T., Kamarck, M. E., Rae, P. M. M., Carty, M. D. & Martin-Deleon, P. (1987) *Am. J. Hum. Genet.* **40**, 115–125.

33. Kaye, F. J., McBride, O. W., Fattey, J. F., Gazdar, A. F. & Sausville, E. A. (1987) *J. Clin. Invest.* **79**, 1412–1420.

34. Haas, R. C., Korenfeld, C., Zhang, Z. F., Perryman, B., Roman, D. & Strauss, A. W. (1989) *J. Biol. Chem.* **264**, 2890–2897.

35. Vasak, M., Nagayama, K., Wuthrich, K., Mertens, M. L. & Kagi, J. H. R. (1979) *Biochemistry* **18**, 5050–5055.

36. Kenyon, G. L. & Reed, G. H. (1983) *Adv. Enzymol.* **54**, 367–426.

37. Mahowald, T. A., Noltmann, E. A. & Kuby, S. A. (1962) *J. Biol. Chem.* **237**, 1535–1548.