Evolution of the Creatine Kinases

THE CHICKEN ACIDIC TYPE MITOCHONDRIAL CREATINE KINASE GENE AS THE FIRST NONMAMMALIAN GENE*

(Received for publication, December 20, 1995)

Stephan M. Mühlebach, Thomas Wirz, Urs Brändle, and Jean-Claude Perriard
From the Swiss Federal Institute of Technology, Institute for Cell Biology, CH-8093 Zürich, Switzerland

In both mammals and birds, the creatine kinase (CK) family consists of four types of genes: cytosolic brain type (B-CK); cytosolic muscle type (M-CK); mitochondrial ubiquitous, acidic type (Mi-CK); and mitochondrial sarcomeric, basic type (Mib-CK). We report here the cloning of the chicken Mi-CK cDNA and its gene. Amino acid sequences of the mature chicken Mi-CK proteins show about 90% identity to the homologous mammalian isoforms. The leader peptides, however, which are isoenzyme-specifically conserved among the mammalian Mi-CKs, are quite different in the chicken with amino acid identity values compared with the mammalian leader peptides of 38.5-51.3%.

The chicken Mi-CK gene spans about 7.6 kilobases and contains 9 exons. The region around exon 1 shows a peculiar base composition, with more than 80% GC, and has the characteristics of a CpG island. The upstream sequences lack TATA or CCAAT boxes and display further properties of housekeeping genes. Several transcription factor binding sites known from mammalian Mi-CK genes are absent from the chicken gene. Although the promoter structure suggests a ubiquitous range of expression, analysis of Mi-CK transcripts in chicken tissues shows a restricted pattern and therefore does not fulfill all criteria of a housekeeping enzyme.

A sufficient capacity and balanced regulation of “high energy phosphate” supply and turnover is essential for the proper function of any cell. Large amounts of energy-rich phosphagens can be found in many cells or tissues throughout the animal kingdom. In all vertebrates and also in some invertebrates this phosphagen is phosphorylcreatine (PCr)\(^1\). PCr and ADP are the products of the reversible transfer of \(\gamma\)-phosphate groups from ATP to creatine, catalyzed by the creatine kinases (CKs). Two fundamental types of CKs can be found in vertebrates: cytosolic and mitochondrial CKs. The subcellular localization, the biochemical and kinetic data, and the loss of flagellar motility in spermatozoa upon inhibition of the CK system and other data (for review, see Ref. 2) led to the suggestion of a metabolic PCr circuit with PCr as a transport and storage form of high energy phosphate. The PCr circuit connects sites of high energy phosphate production (glycolysis and oxidative phosphorylation) with those of high energy phosphate consumption. At the producing end of the circuit, CK is thought to have privileged access to ATP generated either by glycolysis or by oxidative phosphorylation in the mitochondrial matrix and uses this ATP to generate PCr. At the receiving end, CK is functionally coupled to various ATPases (for instance myosin ATPase of myofibrils), which use the ATP generated in the reverse CK reaction.

In chicken there are five different CK subunits known so far; three are found in the cytosol, two in mitochondria. The cytosolic subunits are called M-CK (muscle) and B\(_a\)-CK and B\(_b\)-CK (brain, more acidic and more basic, respectively) and can dimerize with each other (3–5). They are found soluble in the cytosol, but fractions are also associated, for instance, with the M-line of the sarcomeres (6), the sarcoplasmatic Ca\(^2+\)-ATPase (7), or the spermatozoan tail.\(^2\) Whereas in mammals there is just one isoform of B-CK, the two B-CK isoforms of the chicken are derived from a single gene by alternative splicing of the second exon (9, 10). Additional heterogeneity of B-CK was shown to be due to alternative initiation of translation (11) or postranslational phosphorylation (12–14). In tissues of adult chicken, M-CK is predominantly found in skeletal muscle, but contrary to mammals, it is absent from heart (3). B-CK is expressed in almost all tissues and found enriched in various regions of the brain, retina, heart, gizzard, gut, and sperm (2).

Evidence for two different isoforms of mitochondrial CK (M1-CK) has been found by comparison of translated cDNA sequences, isolated from a leg muscle cDNA library, to partial amino-terminal protein sequences of Mi-CK purified from brain (15). These two isoforms were termed Mi\(_1\)-CK (a more acidic pl, in mammals it was called ubiquitous Mi-CK by other authors) and Mi\(_2\)-CK (b = more basic pl, sarcomeric Mi-CK in mammals) and are found in vivo exclusively as homodimers and homooctamers (16, 17). The Mi-CKs are synthesized in the cytosol as precursor proteins with distinct leader peptides (15, 18, 19) and get imported into the intermembrane space of mitochondria. Mi\(_1\)-CK is normally coexpressed with M-CK and is present in skeletal muscle and heart (15), but it also has been found in sperm.\(^2\) Mi\(_1\)-CK is distributed more ubiquitously and has been detected, like B-CK, in chicken brain, gut, and retina. Hence the expression of the mitochondrial CKs matches to a certain degree the pattern found for the cytosolic CKs, indicating possible common regulatory mechanisms.

Recently a comparison of the 26 known CK protein sequences...
suggested a highly conserved protein family and allowed the construction of an evolutionary tree (20). This tree predicts three gene duplications at the origin of the four CK isoforms, which is in agreement with observations on the gene structures of the four published human CK genes (21–25). In chicken, only the B-CK gene has been completely analyzed so far, and fragments of the M-CK gene indicate an extraordinarily big size (10).

To gain further insight into the evolution of the CK isoenzyme family, the chicken Mi-CK protein sequence and the M-CK gene structure were elucidated as the first nonmammalian Mi-CK gene. The chicken M-CK leader peptides were analyzed in order to determine if the functional chicken peptides are isoprotein-specifically conserved. Analysis of the promoter elements of the M-CK gene was carried out to investigate if the expression pattern of Mi-CK is coupled to B-CK expression. The so-called gene-specific mechanisms (Boehringer Mannheim and New England Biolabs). Digests of DNA were analyzed by agarose-gel electrophoresis followed by transfer onto nylon membranes (Pall Inc.). After UV cross-linking, the membranes were hybridized according to Church and Gilbert (32), if oligonucleotides were used, or as described in Sambrook et al. (30) for larger DNA probes. For the 64-bp probe, Denhardt’s solution was left out of the hybridization mixture as suggested in recently published protocols (30).

Restriction Mapping and Southern Blot Analysis

Genomic DNA was prepared from chicken erythrocytes with the proteinase K method (30). Restriction digests of genomic DNA or plasmid DNA were done overnight with manufactured restriction endonucleases (Boehringer Mannheim). A total of 3.6 × 10⁶ plaques was screened at a density of 25,000 plaques/plate (100 mm). Lift-offs were performed with Biodyne type A nylon membranes (Pall Inc.) according to the manufacturer’s instructions. DNA (20–40 ng) was labeled with [α-32P]dCTP. Oligonucleotides used were 5'-CTGGAGCAAAGTGGC-3' (derived from the amino-terminal sequence GQQRRYP with two potential start codons at the 5'-end; highly Mi-CK specific) and 5'-GTCACTGACTCCTAGCTGGCC-3' (complementary to amino acids 212–218, WHNNDKT; CK specific). Degen-eracy of these oligonucleotides was reduced by analysis of the cDNA usage of the three available chicken CK sequences (9, 15, 26–28). Total RNA from chicken brain was reverse transcribed with random hexamer primers. PCR on the obtained cDNA was carried out for 60 cycles (94°C for 30 s, 50°C for 30 s, 72°C for 2 min) as described earlier (29). The fragment was gel-purified, reamplified, subcloned into pTZ19U (U.S. Biochemical Corp.) and fully sequenced.

Radioactive Labeling of DNA Probes

Vector-free fragments were isolated from low melting point agarose with Agarase (Boehringer Mannheim). DNA (20–40 ng) was labeled with [α-32P]dCTP (5000 Ci/mmol, Amersham Corp.) by the random primer method (33). PCR labeling by direct incorporation was done exactly as described previously (34). Approximately 0.3 ng of fragment, as obtained in a normal PCR, served as template with 150 μCi of [α-32P]dCTP. Oligonucleotides used were 5'-CTGGAGCAAAGTGGC-3' and 5'-GTCACTGACTCCTAGCTGGCC-3'. The hybridization mixture contained 1500 Ci/mmol, Amersham Corp.) by T4-polynucleotide kinase (New England Biolabs) according to the manufacturer’s conditions.

Sequencing

Sequencing was carried out according to the method of Sanger et al. (35). Cycle sequencing was performed with the fmol kit (Promega); the annealing step in the cycling procedure was done at the Tm of the oligonucleotide. Maxam-Gilbert sequencing was done according to the modified method of Somm and Tomizawa (36).

Construction of Plasmids for RNase Protection and in Situ Hybridization

Mio-CK (pStM5)—PCR on clone UB15-9 was done with the primer 5'-TCTAGGCTTCTTGTCTGACTGACC-3' (HpaI site at 5'-end) and the R-primer of pgEM-3 (Promega). Cycling conditions were 30 s at 95°C, 1 min at 58°C, and 2 min at 72°C. The fragment was digested (HpaI, EcoRI), gel purified, and subcloned into pGEM3-zf(−).

B-CK (pStM6)—The 3'-terminal HaedII fragment of the chicken B-CK clone pUCB-18 (9), comprising most of the 3'-UTR, was cloned into pgEM3-zf(−). It was reexcised with BamHI and HindIII and subcloned into pTZ19U to allow transcription with T7 RNA polymerase. All plasmids were grown in Escherichia coli XL-1 blue (Stratagene), and DNA was isolated either by an alkali lysis-type method (30) or by column preparation (NucleoBond AX kit, Machery-Nagel, Oensingen, Switzerland) according to the manufacturer’s descriptions.
**Conservation of Mi-CK Genes and cDNAs in Mammals and Birds**

**RESULTS**

Cloning of a Mi<sub>1</sub>-CK cDNA Fragment—In chicken, the cDNA sequences of the products of three out of the four CK genes are known. In case of the Mi<sub>1</sub>-CK isoform, only partial amino acid sequences have been published. To amplify a cDNA fragment coding for this isoform, a PCR approach with degenerate oligonucleotides was chosen. As chicken Mi<sub>1</sub>-CK is highly expressed in adult brain (17), total RNA from chicken brain was used in RT-PCR. This RT-PCR produced a fragment of 653 bp (Fig. 1A), which was subcloned (clone UB11-83) and sequenced. Comparison to the human, rat, and mouse Mi<sub>1</sub>-CKs (19, 24, 41) indicated that we had cloned a chicken Mi<sub>1</sub>-CK fragment.

**FIG. 1.** Structure and partial restriction map of the chicken Mi<sub>1</sub>-CK gene locus, of genomic λ clones and of the derived cDNA. A, the nucleotide sequence of chicken Mi<sub>1</sub>-CK cDNA is represented as a bar. The intron-exon boundaries are shown as vertical lines, and the vertical numbers above represent the 5'-ends of the exons. The 5' and 3'-UTRs are represented by open boxes, and the part coding for the leader peptide is hatched. The filled black boxes outline the exon that is conserved in all CK genes (see “Discussion”). The fragment amplified in RT-PCR (UB11-83) is shown with the two degenerate oligonucleotides (boldface), used in the final amplification (also see “Materials and Methods”). The cDNA fragment isolated by library screening (UB15–9) and the extension in the direction of the 5'-end obtained by RACE-RT-PCR (TW61–11) are indicated as well. B, the organization of the chicken Mi<sub>1</sub>-CK gene locus was derived from the analysis of three (gMia-64, -67, -72) of the six genomic λ clones. The localization of these three clones along the gene locus is indicated. Clone 72 is rearranged at a Sau3AI site in intron 7 as verified by direct sequencing of the λ clone. BamHI fragments of clone 72 were subcloned and sequenced in part (see sequencing strategy, indicated by arrows). The region of the BamHI sites and of the 3'-end of the gene were sequenced directly on the corresponding genomic λ clones (indicated by thick bars above the lines of the λ clones). The sequencing strategy is shown by arrows, and intron 1 has not been sequenced in total as indicated. At the ends of the genomic λ clones there are BamHI sites that probably are cloning artifacts from a Mbol-BamHI fusion. At the 5'-end of clone 67 is an additional BamHI site. Because clone 72 contains a upstream HindIII site, which in combination with the HindIII site in clones 64 and 67 gives rise to the same HindIII fragment as the one found in genomic Southern blots, it represents a faithful segment of the chicken Mi<sub>1</sub>-CK locus. B, BamHI; Sa, Sau3AI; H, HindIII; T, TaqI; parts sequenced directly on the λ clones; 4% polyacrylamide-urea gel.

**RNase Protection**

Total RNA was prepared from excised tissue specimen from adult chicken according to the method of Chomczynski and Sacchi (37). Two ml of solution D were used per 100 mg of tissue for homogenization with a Polytron mixer.

Plasmids pStM5 and pStM6 were linearized with HindIII/NdeI and BamHI/EcoRI, respectively. Both were transcribed with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively. [32P]rCTP-labeled cRNA-probes were synthesized with T7 RNA polymerase (Promega), producing probes of approximately 160 and 220 nucleotides, respectively.
cDNA sequence starts only 170 bp downstream of the ATG initiation codon (see Figs. 1 and 2), reaches to the 3'-end at 1383 bp of the mRNA (see Fig. 1A), but contains no poly(A) tail. The RACE-RT-PCR method (31) is a way to amplify unknown cDNA ends, with the advantage of avoiding time-consuming screening procedures. Hence this method was used to verify the 3'-end of the chicken Mia-CK cDNA. In this experiment, a poly(A) tail was found at the expected position. The polyadenylation signal sequence CATAAA was identified 19 bp upstream from the 3'-end (see Fig. 2, boxed nucleotides at the end). The 3'-UTR is 134 bp long, but it lacks the sequence similarities observed by others (42) in the 3'-UTRs of human, rat, and mouse Mia-CKs. For the chicken Mia-CK, a message size of approximately 1700 bp is observed in Northern blot experiments (not shown). Since poly(A) tails usually range between 100 and 200 bp, the Mia-CK cDNA should have a length in the range of 1500 bp. Hence the missing sequence at the 5'-end might be 250–350 bp long.

Fig. 2. Partial sequence of the chicken Mi-CK gene. The sequences were determined as indicated in Fig. 1 (arrows). The length of intron 1, which was not sequenced in total, was determined by analysis of restriction enzyme digests, and only part of the 1 kb that was sequenced is shown. Capital letters represent exon sequences, whereas lower case letters stand for intron and 3'- and 5'-flanking sequences of the gene. All intron-exon boundaries were sequenced, and splice sites were identified in all cases (underlined; points refer to nucleotides that do not match the consensus splice sequences). The polyadenylation signal is boxed. In the promoter region, the two GC boxes are underlined twice (A), and the putative AP-2 binding sites are indicated by a single line above the sequence. The two potential E2A contact sites are shown by arrows below the sequence and the MRE-site by an arrow above the sequence. The derived amino acid sequence is indicated under the respective exon, and the mitochondrial import signal is underlined. The numbering of the protein sequence starts with the first amino acid of the mature protein. Thus the initiating methionine is amino acid 239. The nucleotide sequence is not numbered with the exception of the region upstream of the ATG. There are 120 nucleotides/full line, and every 10 nucleotides there are asterisks on top of the figure.
A uitous; Ch 11924 bp compared with the cDNA clone UB15–9 (Fig. 1 obtained from these experiments provided only 83 additional these missing 5'-terminal sequences. The largest extension chicken Mib-CK (15) was cloned at the same time. There we method was working in our hands, as the missing 5'-end of the RNA, did not yield any further sequence information. The ric hydroxide to melt possible secondary structure elements in the 9-UTR, which we predict to have length of 9 bp of the 5'-end of chicken Mia-CKcDNA, which encodes part of the 9 peptide necessary for import into mitochondria, and 9 of 39 amino acids, necessary for import into mitochondria. To-gether with the chicken Mia-CK sequence, we now have two nonmammalian leader peptides available for analysis of function and evolutionary relationships. Both leaders show the typical tripartite structure of mitochondrial import sequences (43). As can be seen in Fig. 3, the mammalian Mi-CK leader peptides are highly conserved in an isoform-specific manner. The leader peptides of mammalian Mi-CKs display amino acid identities around 90%, and the same is true for the Mi-CKs (Fig. 3, table). If the chicken Mi-CKs are included in this comparison, the values of identity are much lower, and the isoform-specific conservation is no longer striking. The chicken Mi-CK leader peptide shows identities to the ubiquitous Mi-CKs between 56.4 and 61.5%, and the Mi-CK peptide is 54 and 59% identical to the mammalian sarcomeric Mi-CK peptides. These values are barely higher than those obtained in interspecies comparisons, which range from 38.5 to 51.3%.

The amino acids conserved in all of the leader peptides of either the Mia-CK or the Mi-CK are, with one exception, also conserved in the chicken Mia- and Mi-CKs. Preliminary expression studies done with the chicken Mi-CK cDNA have shown that the Mi-CK leader peptide is functional and that Mi-CK gets imported into mitochondria of heterologous mammalian CV-1 cells. 3

Isolation of Genomic Mia-CK Sequences—To investigate the structure and the putative regulatory elements of the chicken Mia-CK gene, the EcoRI/SmaI cDNA fragment of clone UB15–9 was taken to screen a chicken genomic library. In four rounds of screening, six of the initially 17 clones could be purified to homogeneity. The clones were analyzed by Southern blots using oligonucleotides from different regions of the known cDNA fragment. Clone gMia 72, encompassing 12 kb, was chosen for subcloning and further analysis. This clone contains about 5 kb of sequences upstream of the ATG, including the putative promoter region, and reaches as far toward the 3'-end as into intron 7. However, direct sequencing of the λ-clones revealed a rearrangement at the 3'-end (see Fig. 1B). The 3'-region of the gene was therefore directly sequenced on two other, not rearranged, λ-clones (64 and 67 in Fig. 1, thick black bars). In total, close to 5.3 kb have been sequenced (Fig. 2) including 500–600 bp of putative promoter region. About 300 bp of promoter sequences further upstream were analyzed, but they are only sequenced on one strand and therefore are not included. The nonsequenced part is located in intron 1. At the

| Leader peptide | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------------|---|---|---|---|---|---|---|
| 1 HummMcCK     | 92.3| 89.7| 61.5| 41| 38.5| 46.7| |
| 2 RatubMcCK    | 97.4| 92.3| 59| 41| 38.5| 46.7| |
| 3 MouubMcCK    | 96.3| 97.4| 56.4| 43.6| 41| 51.3| |
| 4 ChiMcCK      | 91.3| 90.5| 50.7| 41| 41| 46.2| |
| 5 HumsarMcCK   | 85.1| 83.1| 82.8| 82.8| 87.2| 59| |
| 6 RatubMcCK    | 82.5| 82.5| 82.1| 82.1| 95.5| 59| |
| 7 ChiMcCK      | 82.5| 83.1| 83.4| 83.6| 89.2| 89.8| |

B Mature proteins

|    |    |    |    |    |    |    |
|----|----|----|----|----|----|
| 1 HummMcCK | 92.3 | 89.7 | 61.5 | 41 | 38.5 | 46.7 |
| 2 RatubMcCK | 97.4 | 92.3 | 59 | 41 | 38.5 | 46.7 |
| 3 MouubMcCK | 96.3 | 97.4 | 56.4 | 43.6 | 41 | 51.3 |
| 4 ChiMcCK | 91.3 | 90.5 | 50.7 | 41 | 41 | 46.2 |
| 5 HumsarMcCK | 85.1 | 83.1 | 82.8 | 82.8 | 87.2 | 59 |
| 6 RatubMcCK | 82.5 | 82.5 | 82.1 | 82.1 | 95.5 | 59 |
| 7 ChiMcCK | 82.5 | 83.1 | 83.4 | 83.6 | 89.2 | 89.8 |

Fig. 3. Comparison of the leader peptides of the known Mi-CKs. A, the leader sequences of the known Mia-CKs (or ubiquitous Mi-CKs) and Mi-CKs (or sarcomeric Mi-CKs) were aligned and arranged using the programs Pileup and Pretty of the GCG software package (45). An amino acid was put into the consensus and replaced by a hyphen in the corresponding sequence if it was identical in six of the seven sequences. B, percentages of amino acid sequence identity. The upper right half shows the identities within the leader peptides, whereas the lower left half shows those within mature proteins. Hum-ub, human ubiquitous; Ratub, rat ubiquitous; Mouub, mouse ubiquitous; Ch, chicken; Humsar, human sarcomeric; Ratsar, rat sarcomeric.
3’ end, there are 83 bp of known sequence after the polyadenylation site. The chicken M_i2-CK gene is approximately 7.6 kb long and has nine exons. The defined exon sizes range from 86–247 bp, but the putative exon 1 may be larger. Intron 1 is 4.2 kb long, of which 1.2 kb were sequenced. All other introns are rather small (~520 bp), giving rise to a compact gene structure on the 3’-side. Analysis of the nucleotide sequences around the intron-exon junctions (underlined in Fig. 2) shows that they correspond well to the AG-GT splice junction rule (44) for intron-exon boundaries. The dots in the splice regions indicate deviations from the rule.

The 5’-region of the chicken M_i2-CK gene shows a high GC-content. It is around 80% in the whole BamHI fragment comprising exon 1 (see Fig. 1B), and the same holds true for the entire exon 1, regardless of the size of 5’-UTR proposed. The three mammalian M_i2-CKs (19, 24, 41) also have a higher GC-content in the exon 1 region of the cDNA, but it is 15% below the content found in chicken. Strong RNA secondary structures, as predicted by the Stemloop program in the GCG software package (45, 46), are an additional feature of exon 1. Both the high GC content and the strong secondary RNA structure prevented the mapping of the transcription start site of the chicken M_i2-CK gene (see Figs. 1 and 2), although we tried primer extension analysis and RNase- and S1-protection analysis. From the putative transcription factor binding sites found (see below and Fig. 2), a transcription start site about 100 bp upstream of the ATG initiation codon seems most likely. Alternatively an additional, untranslated exon might exist, but in the sequences upstream of the ATG there was no indication of a further splice site. Therefore it is highly probable that exon 1 is in fact the first exon and that transcription initiates about 100 bp upstream of the ATG.

Analysis of the chicken M_i2-CK gene with the Grail software (47) predicts, around exon 1 (see Fig. 1 and 2), a CpG island with a very high CpG score of 0.98 (for definition of this score, see Ref. 48). The island starts at ~587 bp upstream of the ATG and extends to +506 bp downstream. The human M_i2-CK gene also has such a CpG island from ~198 to +236 bp, with a score of 0.73. In both cases, it has to be investigated if these islands are undermethylated, but their presence in the promoter regions is of interest in the context of their regulation. A lot of so-called housekeeping genes (49) or tissue-specific genes with a broad range of expression (50) do have such CpG islands. In addition, these genes usually lack TATAA and CCAAT boxes. The putative promoter region of the chicken M_i2-CK gene fulfills all of these features and is typical for such a gene type. A search for specific potential binding sites for transcription factors upstream from the ATG revealed, among many SP1 sites (14 on the upper strand), two GC boxes that match the consensus defined by Kadonaga et al. (51) and are marked with a double line in Fig. 2. Further potential binding sites were found for the general factor AP2 (single lines above the sequence) matching the consensus sequence 5’-CCCMNSSS-3’ (52) for the ubiquitously expressed products of the E2A gene (5’-GCAGGGTGCC-3’, arrows below the sequence) and for a metal response element (MRE, at ~82 bp, arrow above), which is identical to the MRE element in the mouse metallothionein-I gene promoter (53). Suzuki and co-workers (54) have identified three sequence elements (M1, M3, and M4) common in the 5’-flanking regions of nuclear genes coding for mitochondrial proteins. The same three stretches are also found in human sarcomeric and mouse ubiquitous Mi-CK. However, none of these “mitochondria-related” binding sites is found in the sequenced upstream region of the chicken M_i2-CK, although they might be located further upstream. Glucocorticoid and estrogen response elements were reported in mouse and human ubiquitous Mi-CK genes in introns or downstream of the gene. In chicken, only some, possibly nonfunctional, half-sites for the glucocorticoid response element are found in intronic regions. There is only one Mi-CK gene—Genomic Southern blot analysis, using high molecular weight genomic DNA prepared from chicken erythrocytes, was performed to verify the gene structure and to rule out the possibility of pseudogenes as were found for rat and human B-CK genes (21–23, 55).

The same 317-bp fragment used for the screening of the genomic library was used as a probe and produced only one signal per lane on a Southern blot of DNA digested with BamHI, HindII, and Taq| (Fig. 4A). The observed hybridizing fragments were 8.5, 11, and 2.9 kb long, respectively, and are in perfect agreement with the restriction map derived from the genomic clones as shown in Fig. 1B.

To further support the gene structure reported here, another Southern blot (Fig. 4B) was hybridized with a 64-bp probe, labeled by PCR, from the 5′-end of the known cDNA sequences (nucleotides 88–151 of the cDNA depicted in Fig. 1A; corresponding to part of exon 1). Again the fragments of 1.1, 11, and 2.3 kb, respectively, were the same as those found in the genomic λ clones. These data therefore indicate that the chicken M_i2-CK gene is a single copy gene. Additionally the 317- and 64-bp probes both hybridize in HindII-digested DNA to a fragment of the same length of 11 kb. This shows that the organization of the λ clones correctly represents the genomic locus.

Expression Pattern of the Chicken M_i2-CK in Adult Tissues—In mammals, the homologue of the chicken M_i2-CK is usually called ubiquitous Mi-CK, thus implicating that it is actively transcribed in all tissues. The expression of chicken M_i2-CK was investigated by RNase protection experiments on total RNA from various tissues of the adult chicken (see Fig. 5). Probes specific for Mi-CK and, as positive control, for B-CK, both derived from the 3′-end of the corresponding cDNAs (see “Materials and Methods”), were used to test RNA from brain, leg muscle, heart, gut, kidney, and testis. This showed that M_i2-CK is highly expressed together with B-CK in brain and gut but only weakly in testis. There was no detectable expres-
Conservation of Mi-CK Genes and cDNAs in Mammals and Birds

**Fig. 6.** **Mi-CK is not localized in smooth muscle tissues of gut and gizzard.** Mi-CK and B-CK mRNA was localized in liver, spinal cord, gut, and gizzard by in situ hybridization with the same, but 35S-labeled cRNA probes as in the RNase-protection experiment (Fig. 5). Sections were from liver (A–C), spinal cord (thoracic vertebrae; D–F), gut (G–I), and gizzard (J–L), A, D, G, and J, bright field micrographs; B, E, H, and K, dark field micrographs of sections hybridized with the B-CK specific probe; C, F, I, and L, dark field micrographs of sections hybridized with the Mi-CK-specific probe. In the three consecutive liver sections shown, the folds in the tissue are due to sectioning artifacts. The folds led to false hybridization signals because of trapping of the probe. The arrows in micrographs H and I give the position of the muscularis mucosae. The level of Mi-CK hybridization in the spinal cord (F) is low on the outside of the neuronal cell bodies, and comparison to the negative control (not shown) revealed that it represents background. nc, neuronal cell body; dh, dorsal horn; lm, longitudinal smooth muscle; cm, circular smooth muscle; mm, muscularis mucosae; v, villi; l, lumen; sm, smooth muscle tissue; ep, epithelial layer. The in situ hybridizations were exposed for 14 d. The bar corresponds to 200 μm.

**Discussion**

Recently several reports described Mi-CK amino acid sequences and the corresponding gene structures of mammalian Mi-CK genes, and regulatory and evolutionary aspects of the Mi-CKs were proposed (20, 24, 25, 54). Among the nonmammalian species, the chicken CK isoenzyme family has already been well documented (2, 10, 11, 57–60) with the exception of the fourth gene, Mi4-CK. Here we present the chicken Mi4-CK amino acid sequence derived from the cDNA as well as the structure of the corresponding gene. The analysis of the chicken system and its comparison with the mammalian CK set allows conclusions to be drawn on functional elements of the protein, like the leader segment, on regulatory properties of the chicken Mi4-CK promoter leading to regulated expression, as well as on evolution of the CK isoforms.

An amino acids sequence comparison (20) of the creatine kinase isoenzymes has shown that the proteins can be arranged into six different groups based on their levels of sequence identities. Two groups consist of the cytosolic CKs of the fishes/amphibians (see ref. 20) and will not be further discussed, while the other four groups are formed by the different isoforms of mammals and birds. The nine exons of the Mi4-CK gene give rise to a precursor protein of 417 amino acids, including an amino-terminal mitochondrial import sequence of 39 amino acids. The mature chicken Mi4-CK fits in the comparison into the group of the Mi2-CKs or ubiquitous Mi-CKs. The levels of identities between the ubiquitous Mi-CKs known so far are around 90% for the mature protein as shown in Fig. 3B.

In addition, an isotype-specific conservation of the leader peptides was noted for mammalian Mi-CKs, and hence it was suggested that these peptides might act as isoprotein-specific import sequences or bind to specific import receptors (19). The observed low identities with the chicken leader peptides (Fig. 3B) are not in favor to extend this hypothesis also to nonmammalian isoforms, unless the isoprotein-specifically conserved residues found between chicken and mammalian leader peptides (amino acid positions 10, 11, 18, 23–25, 27, 28, and 30) were sufficient to ensure the specific import (Fig. 3B). In contrast to the leader peptides, the mature protein Mi1-CK and Mi4-CK sequences are conserved isofrom-specifically to a degree of almost 90%, and no segments can be found in the...
mature proteins where the conservation drops to the low levels of the leader peptides. Hence, the reduced conservation of the leader peptides cannot be explained by the greater phylogenetic distance of the chicken from mammals since leader peptide and mature proteins must have undergone simultaneous evolution. The noted high homologies in the mammalian peptides are therefore a mere fact of their phylogenetic closeness, and the apparent lower evolutionary pressure for conservation of the leader peptides is only observed when the large evolutionary gap, as the one between mammals and birds, is analyzed. Thus, it is unlikely that the conservation of the mammalian Mi-CK leader peptides represent isoform-specific functions. This is supported by the observed import of the chicken M1α-CK into the mitochondria of mammalian fibroblastic CV-1 cells, where endogenous M1α-CK, but no M1α-CK, can be expected.

On the level of the nucleotide sequences, 78% identity is observed if the mature chicken M1α-CK is compared with any of the mammalian ubiquitous Mi-CKs. As conservation on the protein level with 90% is much higher, nucleotide changes occur therefore mainly at wobble positions. The same holds true for the M1β-CKs or any of the cytosolic CKs. In case of the UTRs, the situation is different. Cheng et al. (42) have shown that a sequence in the 5'-UTR of the rat ubiquitous Mi-CK is involved in regulation of its expression/translation. Whereas this sequence stretch of 72 bp is conserved in the mouse and the human sequence, it cannot be found in chicken M1α-CK, suggesting that this mechanism of regulation is not active in chicken.

A schematic representation comparing the four known Mi-CK genes is given in Fig. 7. The organization of the chicken gene is similar to the two other known Mi-CK genes from man and mouse. The localization of intron-exon boundaries in the coding region of the cDNA and the exon-sizes (except number 1) are conserved between these genes. In addition, the chicken gene lacks a noncoding first exon, like the mammalian ubiquitous Mi-CKs but contrary to all other known CK genes. Whereas the mouse and the human genes show a common bipartite gene structure with a clustering into the group of exons 1–6 and 7–9 (41), in chicken M1α-CK, exon 1 is separated from the compact rest of the gene by a rather large intron. The sarcomeric Mi-CK gene, although it is with 37 kb the largest Mi-CK gene known so far and in addition has two noncoding first exons, has exon-intron boundaries at the same positions of the coding region as the ones noticed in the ubiquitous Mi-CK genes (see Fig. 7). Hence, the two gene types might have evolved by a gene duplication event.

As noted earlier (25) the mitochondrial CK gene structure is different from that of the cytosolic CKs. There is only one exon (exon 6, respectively exon 8 in human sarcomeric Mi-CK) that is conserved through all of the CKs in its size and localization in the coding region (black exon in Fig. 7). Other features, for instance the conserved 5’-noncoding exon in case of the cytosolic CK genes, are not found in the Mi-CK genes. Interestingly the region of the conserved exon is also the exon with the highest homology at the amino acid level among the different guanidino kinases (20). This might indicate a critical role of these residues for the structure and function of guanidino kinases, as has already been shown for a tryptophane residue in this region (62).

All of the data mentioned so far show that the CKs form a group of evolutionarily related isoenzymes and can further be embedded into the larger family of the guanidino kinases. The four different CK isoforms most probably evolved by three gene duplication events, with the first of these producing a primordial cytosolic and a primordial mitochondrial isoform. The phylogenetic tree, which can be derived from the protein comparison (20), suggests this first duplication to have occurred before the separation of the echinoderms from chordates, which is in agreement with published data on the expression of Mi-CKs in sea urchins (63, 64). The significant homology, still found if the protein comparison of the CKs is extended to other guanidino kinases like ArgK or guanidinoacetate kinase, is a strong indication for a common ancestor of the guanidino kinases. ArgK has been suggested to be this ancestor on the basis of data on its dimerization capacity with cytosolic CKs, the nature of its substrate arginine as being part of basic metabolism, and especially from the distribution of arginine and ArgK in the animal kingdom (1). However, creatine has been found in sperm of many nonvertebrate taxa as well (1, 65). Taking into account that the recently cloned guanidinoacetate kinase shows higher homology to CKs than to ArgKs (56), it can be
suggested that a duplication event has first produced a guanidinoacetate kinase and ArgK, and that later CK has evolved from guanidinoacetate kinase.

The promoter region of chicken Mi-CK is rather similar to those of the human and mouse genes and displays all of the features attributed to housekeeping genes (49) or tissue-specific genes with a broad range of expression (50). It is at present not known whether any of the transcription factor binding sites identified by sequence analysis are actually used for the regulation of the chicken Mi-CK gene. During rat pregnancy, B-CK as well as ubiquitous Mi-CK (Mi2-CK), are regulated in rat uterus by steroid hormones, and binding sites for steroid receptors are expected. However, in the chicken Mi-CK gene, no such sites have been identified, and only possibly nonfunctional half-sites for glucocorticoid receptors have been found. On the other hand, it has been shown for rat B-CK that the regulation is independent of the binding of an estrogen receptor to the promoter region (66). The three binding sites Mt1, Mt3, and Mt4 (54) identified in nuclear genes coding for mitochondrial proteins have also been found in the mouse ubiquitous Mi-CK gene. These sites are present neither in the chicken nor the human gene, but in both cases the known 5′-located sequences may be short to contain these sequences. Summarizing the data shown, the Mi2-CK gene displays in part the same regulatory elements found in mouse and human. Some of the additional elements reported especially in mouse are not found in the chicken Mi-CK gene, probably due to the limited sequence information at the 5′-end. The “missing” binding sites might, however, be important to explain the observed restricted expression pattern found for chicken mid-CK.

The presence of Mi2-CK in brain tissues is already well documented (2, 60). Our localization in spinal cord shows that Mi2-CK is coexpressed with B-CK in the cell bodies of neural cells of the gray matter but is absent from any other region where only B-CK is found. Hence, neurons seem to rely on a functional PCR shuttle in spinal cord. Whereas other investigators reported minute CK expression in liver, our in situ hybridizations show no Mi2-CK or B-CK expression in general. Either the transcripts are below detection limit or not present at all. The smooth muscle-containing tissues analyzed by in situ hybridization do not express Mi2-CK in their smooth muscle portions, but they display considerable amounts of B-CK. Hence, gut (duodenum) and gizzard smooth muscle in chicken function in the absence of Mi-CK, which is different from vascular and intestinal smooth muscle of guinea pig (8) or smooth muscle from rat (42). For chicken gizzard, the lack of Mi2-CK is due to its peculiar contractile properties (60). Whether this holds true for chicken gut as well is not known. The only portion in gut expressing Mi2-CK is the border region of the villi, which suggests that a functional PCR circuit might be important for cells of the brush border involved in resorption processes. These data show that Mi2-CK expression in tissues of adult chicken is more restricted than that of B-CK. They indicate that the name ubiquitous, given to the mammalian Mi-CK genes, is not justifiable in chicken. The features of the Mi2-CK promoter indicating a housekeeping gene are misleading, and there must be other regulatory elements narrowing its expression. The putative additional regulatory elements will have to determined by future research.

Acknowledgments—We thank Professor H. M. Eppenberger for continuous support. We also thank Drs. M. Wyss and B. Schäfer for valuable discussion and comments on the manuscript. Dr. S. Tanaka for introduction of the technique of Maxam-Gilbert sequencing, and E. Perriard and S. Keller for technical assistance.

REFERENCES

1. Watts, D. C. (1971) Evolution of phosphagen kinases, in Biochemical Evolution and the Origin of Life (Shoffeniels, E., ed) pp. 150–173, North-Holland
50. Edwards, Y. H. (1990) Philos. Trans. R. Soc. London B 326, 207–215
51. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) Trends Biochem. Sci. 11, 20–23
52. Faisst, S., and Meyer, S. (1992) Nucleic Acids Res. 20, 3–26
53. Labbé, S., Prévost, J., Remondelli, P., Leone, A., and Séguin, C. (1991) Nucleic Acids Res. 19, 4225–4231
54. Suzuki, H., Hosokawa, Y., Toda, H., Nishikimi, M., and Ozawa, T. (1990) J. Biol. Chem. 265, 8159–8163
55. Benfield, P. A., Graf, D., Korolkoff, P. N., Hobson, G., and Pearson, M. L. (1988) Gene (Amst.) 63, 227–243
56. Suzuki, T., and Furukohri, T. (1994) J. Mol. Biol. 237, 353–357
57. Perriard, J.-C., Eppenberger, H. M., Hosp, J. P., and Schäfer, B. (1987) inIsozymes (Rattazzi, M. C., Scandalios, J. G., and Whitt, G. S., eds) Vol. 14, pp. 83–101, Alan R. Liss, Inc., New York
58. Schäfer, B. W., and Perriard, J.-C. (1988) J. Cell Biol. 106, 1161–1170
59. Perriard, J.-C., Hosp, J. P., Schäfer, B. W., Soldati, T., Wegmann, G., and Wirz, T. (1989) in Cellular and Molecular Biology of Muscle Development (Stockdale, F., and Kedes, L., eds) pp. 545–554, Alan R. Liss, Inc., New York
60. Wyss, M., Smetink, J., Wevers, R. A., and Walliman, T. (1992) Biochim. Biophys. Acta 1102, 119–166
61. Deleted in proof
62. Gross, M., Furter-Graves, E. M., Walliman, T., Eppenberger, H. M., and Furter, R. (1994) Protein Sci. 3, 1058–1068
63. Tombesi, R. M., and Shapiro, B. M. (1987) J. Biol. Chem. 262, 16011–16019
64. Ratto, A., Shapiro, B. N., and Christen, H. (1989) Eur. J. Biochem. 186, 195–203
65. Ellington, W. R. (1989) J. Exp. Biol. 143, 177–194
66. Sukovich, D. A., Mukherjee, R., and Benfield, P. A. (1994) Mol. Cell. Biol. 14, 7134–7143
Evolution of the Creatine Kinases: THE CHICKEN ACIDIC TYPE
MITOCHONDRIAL CREATINE KINASE GENE AS THE FIRST
NONMAMMALIAN GENE

Stephan M. Mühlebach, Thomas Wirz, Urs Brändle and Jean-Claude Perriard

J. Biol. Chem. 1996, 271:11920-11929.
doi: 10.1074/jbc.271.20.11920

Access the most updated version of this article at http://www.jbc.org/content/271/20/11920

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 60 references, 23 of which can be accessed free at
http://www.jbc.org/content/271/20/11920.full.html#ref-list-1