Molecular and Cellular Biochemistry **133/134**: 245–262, 1994.

IV-2 Sequence homology and structure predictions of the creatine kinase isoenzymes

S.M. Mühlebach, M. Gross, T. Wirz, T. Wallimann, J.-C. Perriard and M. Wyss[1]

Institute for Cell Biology, ETH Hönggerberg, CH-8093 Zürich, Switzerland; [1] *Department of Transplant Surgery, Research Division, University Hospital, Anichstr. 35, A-6020 Innsbruck, Austria*

Abstract

Comparisons of the protein sequences and gene structures of the known creatine kinase isoenzymes and other guanidino kinases revealed high homology and were used to determine the evolutionary relationships of the various guanidino kinases. A 'CK framework' is defined, consisting of the most conserved sequence blocks, and 'diagnostic boxes' are identified which are characteristic for anyone creatine kinase isoenzyme (e.g. for vertebrate B-CK) and which may serve to distinguish this isoenzyme from all others (e.g. from M-CKs and Mi-CKs). Comparison of the guanidino kinases by near-UV and far-UV circular dichroism further indicates pronounced conservation of secondary structure as well as of aromatic amino acids that are involved in catalysis. (Mol Cell Biochem **133/134:** 245–262, 1994).

Key words: creatine kinase, arginine kinase, protein sequence comparison, evolution, CK framework, 'diagnostic boxes', secondary structure prediction

Abbreviations: GuaK – guanidino kinase; CK – creatine kinase; B- and M-CK – brain and muscle cytosolic CK isoenzyme; Mi-CK – mitochondrial CK isoenzyme; ArgK – arginine kinase; Cr – creatine; PCr – phosphorylcreatine; PArg – phosphorylarginine

Introduction

Guanidino kinases (GuaKs)[1] in general and creatine kinases (CKs) in particular are found throughout the animal kingdom [1–4]. Vertebrate tissues contain exclusively CK isoenzymes, while tissues of invertebrates mostly display arginine kinase (ArgK) activity. Spermatozoa of many invertebrates, however, comprise CK as the main or even sole guanidino kinase. Annelida are also an exception to the general picture, in as far as a variety of different guanidino compounds and guanidino kinases were found in this phylum, including arginine (Arg), creatine (Cr), glycocyamine (= guanidinoacetate), taurocyamine, hypotaurocyamine and lombricine, always together with the corresponding kinases. Of this whole class of guanidino kinases, the (mostly) complete primary structures of 26 different CK isoenzymes, of lobster tail muscle ArgK, and of a *Schistosoma mansoni* (trematode) guanidino kinase with unknown substrate specificity have been reported until the end of October 1993, together with amino acid sequences of some short protein fragments. The goals of the present article are (i) to align and compare the known guanidino kinase sequences, (ii) to define a 'CK framework', based on the evolutionarily most conserved parts of the molecules, (iii) to identify isoenzyme-specific residues or sequence

Address for offprints: S.M. Mühlebach, Institute for Cell Biology, ETH Hönggerberg, CH-8093 Zürich, Switzerland

blocks, and (iv) to correlate particular residues or sequence stretches with physiological functions of the guanidino kinases. These interpretations are complemented by secondary structure predictions and by the identification by CD spectroscopy of evolutionarily conserved Trp and Tyr residues which are important for catalysis and for the structural integrity of the molecules.

Characterization of the guanidino kinases

In birds and mammals, four different nuclearly encoded CK isoforms are found, all displaying a protomer M_r of approximately 40'000. These isoforms are expressed tissue-specifically and differ in subcellular localization as well. The two cytosolic isoforms B- (for brain) and M-CK (for muscle) form dimeric molecules (MM-, MB- and BB-CK; [5]). In contrast, the two mitochondrial isoforms Mi_a- and Mi_b-CK ('a' for the more acidic and 'b' for the more basic Mi-CK isoform; also called ubiquitous and sarcomeric Mi-CK, respectively) form dimeric and octameric molecules which, depending on the experimental conditions, are readily interconvertible (see [6]). Although heterodimeric and heterooctameric molecules between Mi_a- and Mi_b-CK can be generated *in vit-*

→

Fig. 1. Alignment and comparison of the known CK protein sequences. The primary structures of 26 CK isoenzymes are compared, with each repeat of the triplicated sea urchin tail CK sequence being analyzed separately. The sequences used in this study were from:

Humbck	human B-CK [47, 49, 81]
Dogbck	dog B-CK [82]
Rabbck	rabbit B-CK [83]
Ratbck	rat B-CK [84]
Moubck	mouse B-CK [52]
Chibbck	chicken B_b-CK [85, 86]
Chiback	chicken B_a-CK [8]
Hummck	human M-CK [51, 63]
Dogmck	dog M-CK [87]
Rabmck	rabbit M-CK [88]
Ratmck	rat M-CK [89]
Moumck	mouse M-CK [65]
Chimck	chicken M-CK [90, 91]
Troutck	CK isoenzyme from rainbow trout displaying enhanced testicular expression [21]
Tormarck	CK from the electrocytes of *Torpedo marmorata* (marbeled electric ray; [23])
Torcalck	CK from the electrocytes of *Torpedo californica* (pacific electric ray; [92])
Xenlaeck3	*Xenopus laevis* CK-III [38]
Xenlaeck4	*Xenopus laevis* CK-IV [37]
Humubimick	human ubiquitous Mi-CK [53]
Ratubimick	rat ubiquitous Mi-CK [93]
Mouubimick	mouse ubiquitous Mi-CK [55]
Chimiack	chicken Mi_a-CK (a = more acidic pI; ubiquitous; [39, 40])
Humsarmick	human sarcomeric Mi-CK [94]
Ratsarmick	rat sarcomeric Mi-CK [93]
Chimibck	chicken Mi_b-CK (b = more basic pI; sarcomeric; [62])
Spfckd1,2,3	flagellar CK from the sea urchin *Strongylocentrotus purpuratus*, with part 1 including amino acids 60–433, part 2 including residues 434–807, and part 3 encompassing residues 808–1174 [31]

Dots in the listed sequences stand for inserted gaps, introduced in order to allow an optimal alignment of the different CKs, or for unknown parts of the primary structures (N-termini of *X. laevis* CK-III and CK-IV). Hyphens stand for amino acids which are identical in 24 (residues 1–81) or 25 (residues 82–395) out of the 28 sequences, with the corresponding amino acid being listed in the consensus sequence at the bottom. If at a particular position only two different amino acids are found in all sequences, both of them are listed in the consensus sequence, one above the other. If there is no consensus, or if an amino acid in a given sequence differs from that in the consensus sequence, it is listed in the respective sequence itself. The N-terminal methionine residue in the cytosolic CK sequences is cleaved off after synthesis.

The filled bars (numbered 1 to 6) below the consensus sequence mark the regions with the most pronounced sequence conservation. The open diamonds mark the so-called 'reactive cysteine' Cys-283, the two residues (Cys-74 and Lys-196) which in cross-linking experiments of rabbit MM-CK were shown to be structurally close to Cys-283, as well as Asp-340 which, as suggested by affinity labelling experiments with an alkylating ATP analogue, may be involved in Mg^{2+} binding. Boxed amino acid residues either are isoform-specific or allow a clear-cut distinction between mitochondrial and cytosolic CK isoenzymes. Extended stretches (≥ 5 residues) with isoenzyme-specific sequence patterns are termed 'diagnostic boxes' (see the text) and are designated by A, B, C, . . . I. If a box comprises two (or more) different isoenzymes, and if, within this particular box, residues are conserved in an isoenzyme-specific pattern, they are represented by bold letters. Finally, filled diamonds mark putative phosphorylation sites.

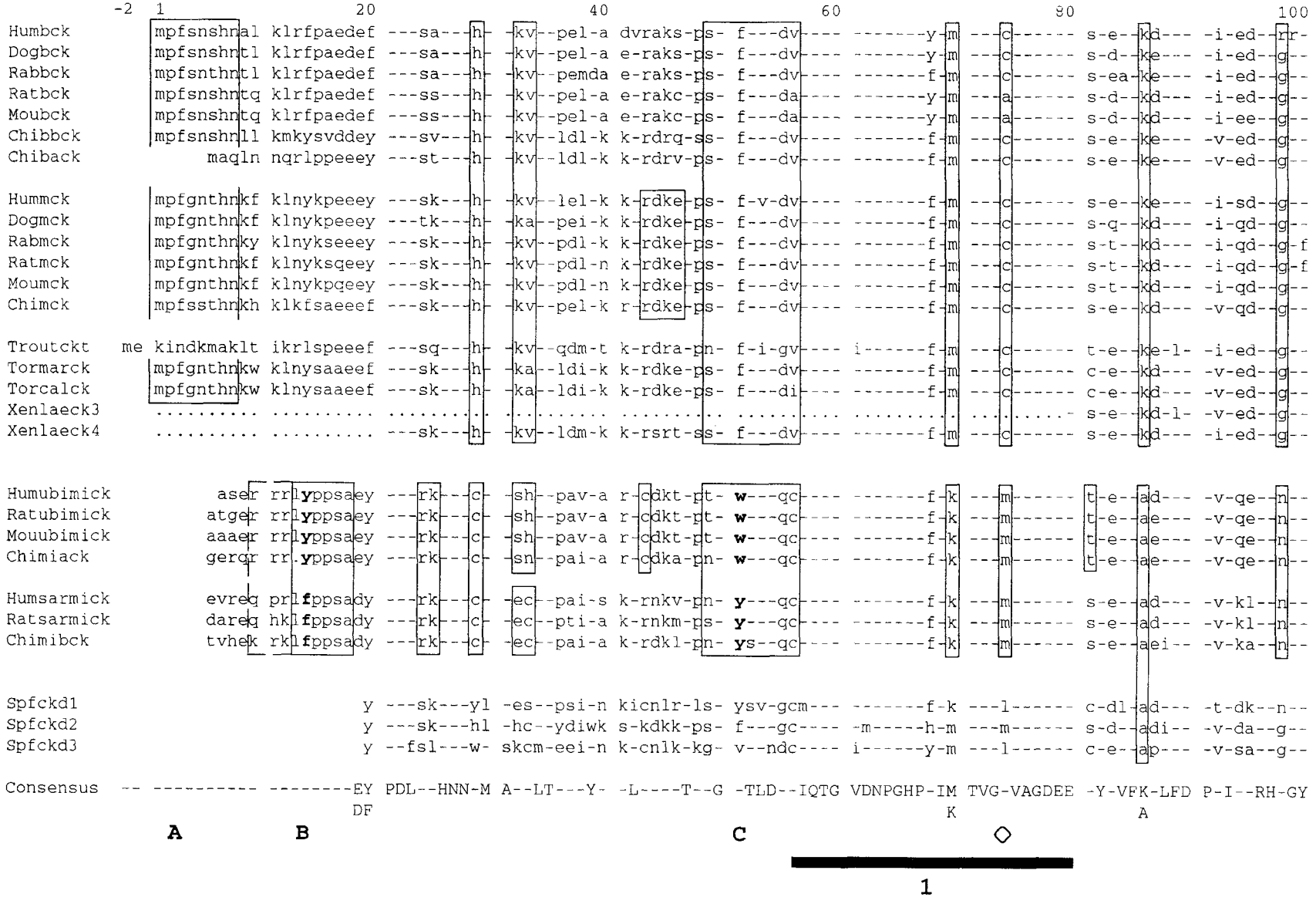
-2 1 20 40 60 80 100
Humbck mpfsnshnal klrfpaedef ---sa---h- -kv--pel-a dvraks-ps- f---dv---- -------y-m ---c------ s-e--kd--- -i-ed--rr-
Dogbck mpfsnshntl klrfpaedef ---sa---h- -kv--pel-a e-raks-ps- f---dv---- -------y-m ---c------ s-d--ke--- -i-ed--g--
Rabbck mpfsnthntl klrfpaedef ---sa---h- -kv--pemda e-raks-ps- f---dv---- -------f-m ---c------ s-ea-ke--- -i-ed--g--
Ratbck mpfsnshntq klrfpaedef ---ss---h- -kv--pel-a e-rakc-ps- f---da---- -------y-m ---a------ s-d--kd--- -i-ed--g--
Moubck mpfsnshntq klrfpaedef ---ss---h- -kv--pel-a e-rakc-ps- f---da---- -------y-m ---a------ s-d--kd--- -i-ee--g--
Chibbck mpfsnshnll kmkysvddey ---sv---h- -kv--ldl-k k-rdrq-ss- f---dv---- -------f-m ---c------ s-e--ke--- -v-ed--g--
Chiback maqln nqrlppeeey ---st---h- -kv--ldl-k k-rdrv-ps- f---dv---- -------f-m ---c------ s-e--ke--- -v-ed--g--
Hummck mpfgnthnkf klnykpeeey ---sk---h- -kv--lel-k k-rdke-ps- f-v-dv---- -------f-m ---c------ s-e--ke--- -i-sd--g--
Dogmck mpfgnthnkf klnykpeeey ---tk---h- -ka--pei-k k-rdke-ps- f---dv---- -------f-m ---c------ s-q--kd--- -i-qd--g--
Rabmck mpfgnthnky klnykseeey ---sk---h- -kv--pdl-k k-rdke-ps- f---dv---- -------f-m ---c------ s-t--kd--- -i-qd--g-f
Ratmck mpfgnthnkf klnyksqeey ---sk---h- -kv--pdl-n k-rdke-ps- f---dv---- -------f-m ---c------ s-t--kd--- -i-qd--g-f
Moumck mpfgnthnkf klnykpqeey ---sk---h- -kv--pdl-n k-rdke-ps- f---dv---- -------f-m ---c------ s-t--kd--- -i-qd--g--
Chimck mpfssthnkh klkfsaeeef ---sk---h- -kv--pel-k r-rdke-ps- f---dv---- -------f-m ---c------ s-e--kd--- -v-qd--g--
Troutckt me kindkmaklt ikrlspeeef ---sq---h- -kv--qdm-t k-rdra-pn- f-i-gv---- i------f-m ---c------ t-e--ke-l- -i-ed--g--
Tormarck mpfgnthnkw klnysaaeef ---sk---h- -ka--ldi-k k-rdke-ps- f---dv---- -------f-m ---c------ c-e--kd--- -v-ed--g--
Torcalck mpfgnthnkw klnysaaeef ---sk---h- -ka--ldi-k k-rdke-ps- f---di---- -------f-m ---c------ c-e--kd--- -v-ed--g--
Xenlaeck3 .... s-e--kd-l- -v-ed--g--
Xenlaeck4 .... ---sk---h- -kv--ldm-k k-rsrt-ss- f---dv---- -------f-m ---c------ s-e--kd--- -i-ed--g--
Humubimick aser rrlyppsaey ---rk---c- -sh--pav-a r-cdkt-pt- w---qc---- -------f-k ---m------ t-e--ad--- -v-qe--n--
Ratubimick atger rrlyppsaey ---rk---c- -sh--pav-a r-cdkt-pt- w---qc---- -------f-k ---m------ t-e--ae--- -v-qe--n--
Mouubimick aaaer rrlyppsaey ---rk---c- -sh--pav-a r-cdkt-pt- w---qc---- -------f-k ---m------ t-e--ae--- -v-qe--n--
Chimiack gerqr rr.yppsaey ---rk---c- -sn--pai-a r-cdka-pn- w---qc---- -------f-k ---m------ t-e--ae--- -v-qe--n--
Humsarmick evreq prlfppsady ---rk---c- -ec--pai-s k-rnkv-pn- y---qc---- -------f-k ---m------ s-e--ad--- -v-kl--n--
Ratsarmick dareq hklfppsady ---rk---c- -ec--pti-a k-rnkm-ps- y---qc---- -------f-k ---m------ s-e--ad--- -v-kl--n--
Chimibck tvhek rklfppsady ---rk---c- -ec--pai-a k-rdkl-pn- ys--qc---- -------f-k ---m------ s-e--aei-- -v-ka--n--
Spfckd1 y ---sk---yl -es--psi-n kicnlr-ls- ysv-gcm--- -------f-k ---l------ c-dl-ad--- -t-dk--n--
Spfckd2 y ---sk---hl -hc--ydiwk s-kdkk-ps- f---gc---- -m-----h-m ---m------ s-d--adi-- -v-da--g--
Spfckd3 y --fsl---w- skcm-eei-n k-cnlk-kg- v--ndc---- i------y-m ---l------ c-e--ap--- -v-sa--g--
Consensus -- ---------- --------EY PDL--HNN-M A--LT---Y- -L----T--G -TLD--IQTG VDNPGHP-IM TVG-VAGDEE -Y-VFK-LFD P-I--RH-GY
DF K A
A B C
1

Fig. 1

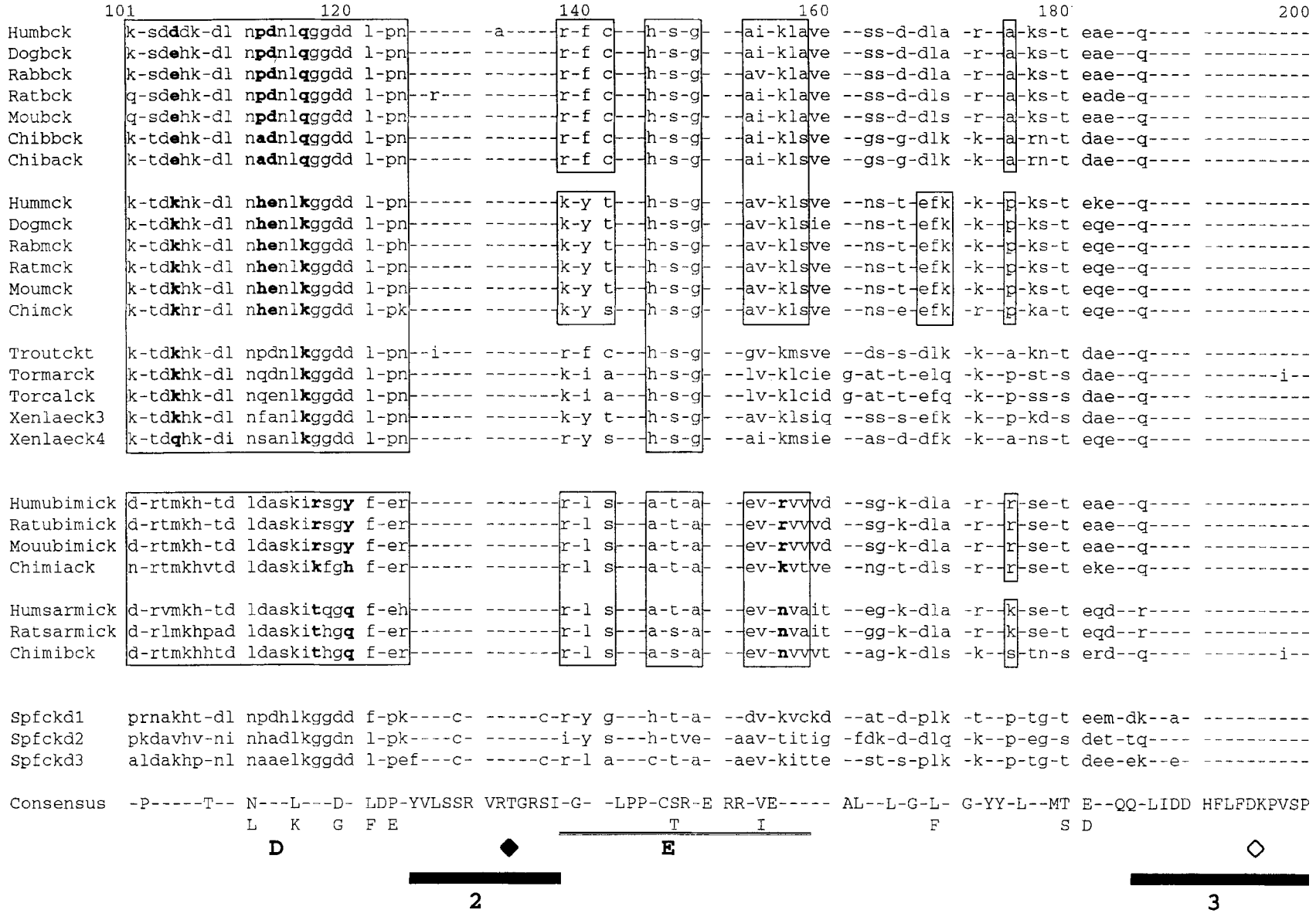
101
120
140
160
180
200
Humbck k-sddk-dl npdnlqggdd l-pn------ -a-----r-f c---h-s-g- --ai-klave --ss-d-dla -r--a-ks-t eae--q---- ----------
Dogbck k-sdehk-dl npdnlqggdd l-pn------ -------r-f c---h-s-g- --ai-klave --ss-d-dla -r--a-ks-t eae--q---- ----------
Rabbck k-sdehk-dl npdnlqggdd l-pn------ -------r-f c---h-s-g- --av-klave --ss-d-dla -r--a-ks-t eae--q---- ----------
Ratbck q-sdehk-dl npdnlqggdd l-pn--r--- -------r-f c---h-s-g- --ai-klave --ss-d-dls -r--a-ks-t eade-q---- ----------
Moubck q-sdehk-dl npdnlqggdd l-pn------ -------r-f c---h-s-g- --ai-klave --ss-d-dls -r--a-ks-t eae--q---- ----------
Chibbck k-tdehk-dl nadnlqggdd l-pn------ -------r-f c---h-s-g- --ai-klsve --gs-g-dlk -k--a-rn-t dae--q---- ----------
Chiback k-tdehk-dl nadnlqggdd l-pn------ -------r-f c---h-s-g- --ai-klsve --gs-g-dlk -k--a-rn-t dae--q---- ----------
Hummck k-tdkhk-dl nhenlkggdd l-pn------ -------k-y t---h-s-g- --av-klsve --ns-t-efk -k--p-ks-t eke--q---- ----------
Dogmck k-tdkhk-dl nhenlkggdd l-pn------ -------k-y t---h-s-g- --av-klsie --ns-t-efk -k--p-ks-t eqe--q---- ----------
Rabmck k-tdkhk-dl nhenlkggdd l-ph------ -------k-y t---h-s-g- --av-klsve --ns-t-efk -k--p-ks-t eqe--q---- ----------
Ratmck k-tdkhk-dl nhenlkggdd l-pn------ -------k-y t---h-s-g- --av-klsve --ns-t-efk -k--p-ks-t eqe--q---- ----------
Moumck k-tdkhk-dl nhenlkggdd l-pn------ -------k-y t---h-s-g- --av-klsve --ns-t-efk -k--p-ks-t eqe--q---- ----------
Chimck k-tdkhr-dl nhenlkggdd l-pk------ -------k-y s---h-s-g- --av-klsve --ns-e-efk -r--p-ka-t eqe--q---- ----------
Troutckt k-tdkhk-dl npdnlkggdd l-pn--i--- --------r-f c---h-s-g- --gv-kmsve --ds-s-dlk -k--a-kn-t dae--q---- ----------
Tormarck k-tdkhk-dl nqdnlkggdd l-pn------ -------k-i a---h-s-g- --lv-klcie g-at-t-elq -k--p-st-s dae--q---- -------i--
Torcalck k-tdkhk-dl nqenlkggdd l-pn------ -------k-i a---h-s-g- --lv-klcid g-at-t-efq -k--p-ss-s dae--q---- ----------
Xenlaeck3 k-tdkhk-dl nfanlkggdd l-pn------ -------k-y t---h-s-g- --av-klsiq --ss-s-efk -k--p-kd-s dae--q---- ----------
Xenlaeck4 k-tdqhk-di nsanlkggdd l-pn------ -------r-y s---h-s-g- --ai-kmsie --as-d-dfk -k--a-ns-t eqe--q---- ----------
Humubimick d-rtmkh-td ldaskirsgy f-er------ -------r-l s---a-t-a- --ev-rvvvd --sg-k-dla -r--r-se-t eae--q---- ----------
Ratubimick d-rtmkh-td ldaskirsgy f-er------ -------r-l s---a-t-a- --ev-rvvvd --sg-k-dla -r--r-se-t eae--q---- ----------
Mouubimick d-rtmkh-td ldaskirsgy f-er------ -------r-l s---a-t-a- --ev-rvvvd --sg-k-dla -r--r-se-t eae--q---- ----------
Chimiack n-rtmkhvtd ldaskikfgh f-er------ -------r-l s---a-t-a- --ev-kvtve --ng-t-dls -r--r-se-t eke--q---- ----------
Humsarmick d-rvmkh-td ldaskitqgq f-eh------ -------r-l s---a-t-a- --ev-nvait --eg-k-dla -r--k-se-t eqd--r---- ----------
Ratsarmick d-rlmkhpad ldaskithgq f-er------ -------r-l s---a-s-a- --ev-nvait --gg-k-dla -r--k-se-t eqd--r---- ----------
Chimibck d-rtmkhhtd ldaskithgq f-er------ -------r-l s---a-s-a- --ev-nvvvt --ag-k-dls -k--s-tn-s erd--q---- -------i--
Spfckd1 prnakht-dl npdhlkggdd f-pk----c- -----c-r-y g---h-t-a- --dv-kvckd --at-d-plk -t--p-tg-t eem-dk--a- ----------
Spfckd2 pkdavhv-ni nhadlkggdn l-pk----c- -------i-y s---h-tve- -aav-titig -fdk-d-dlq -k--p-eg-s det-tq---- ----------
Spfckd3 aldakhp-nl naaelkggdd l-pef---c- -----c-r-l a---c-t-a- -aev-kitte --st-s-plk -k--p-tg-t dee-ek--e- ----------
Consensus -P-----T-- N---L---D- LDP-YVLSSR VRTGRSI-G- -LPP-CSR-E RR-VE----- AL--L-G-L- G-YY-L--MT E--QQ-LIDD HFLFDKPVSP
L K G F E T I F S D
D
E
2
3

Fig. 1

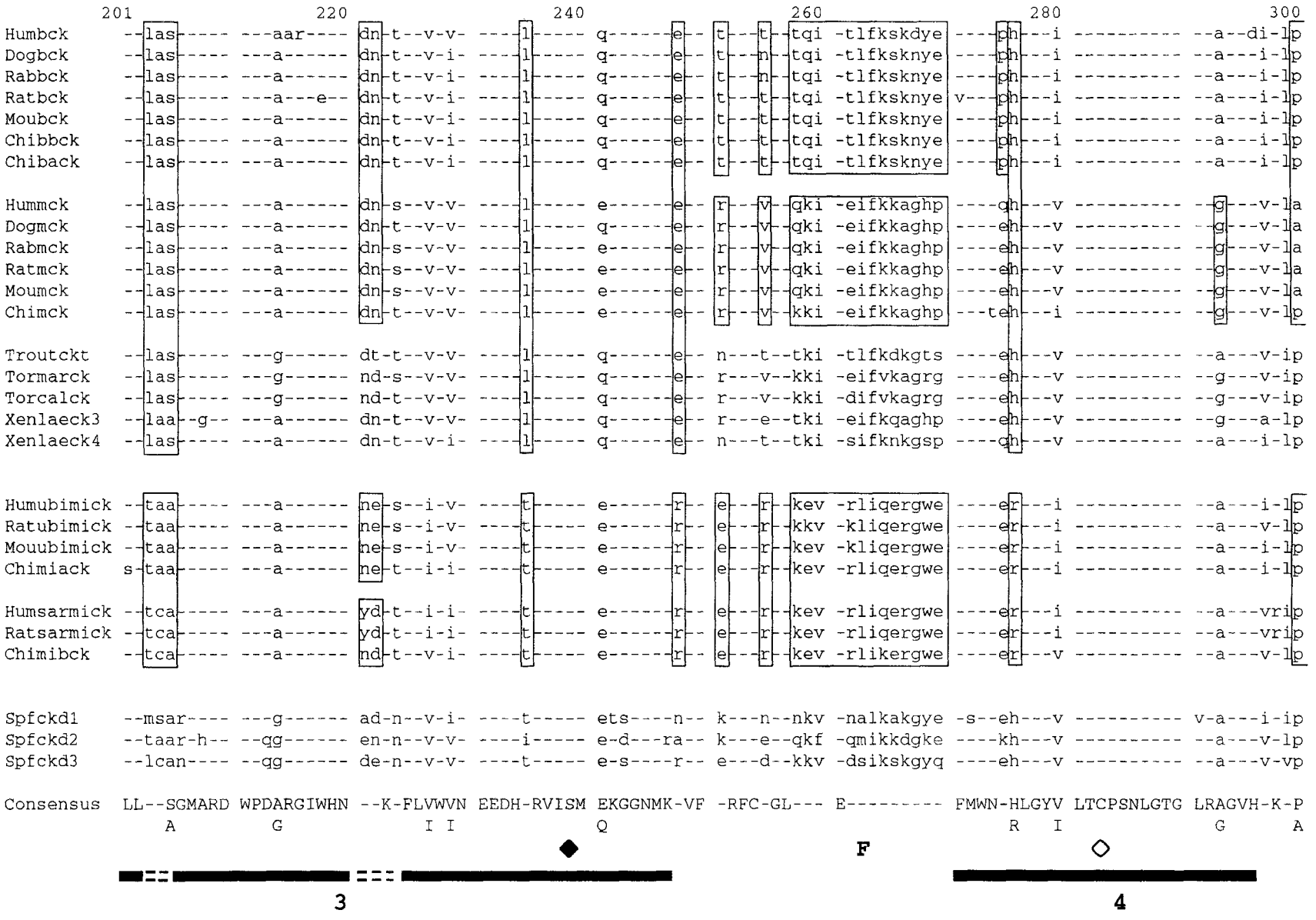
201
220
240
260
280
300
Humbck --las----- ---aar---- dn-t--v-v- ----l----- q------e-- t---t--tqi -tlfkskdye ----ph---i ---------- --a--di-lp
Dogbck --las----- ---a------ dn-t--v-i- ----l----- q------e-- t---n--tqi -tlfksknye ----ph---i ---------- --a---i-lp
Rabbck --las----- ---a------ dn-t--v-i- ----l----- q------e-- t---n--tqi -tlfksknye ----ph---i ---------- --a---i-lp
Ratbck --las----- ---a---e-- dn-t--v-i- ----l----- q------e-- t---t--tqi -tlfksknye v---ph---i ---------- --a---i-lp
Moubck --las----- ---a------ dn-t--v-i- ----l----- q------e-- t---t--tqi -tlfksknye ----ph---i ---------- --a---i-lp
Chibbck --las----- ---a------ dn-t--v-i- ----l----- q------e-- t---t--tqi -tlfksknye ----ph---i ---------- --a---i-lp
Chiback --las----- ---a------ dn-t--v-i- ----l----- q------e-- t---t--tqi -tlfksknye ----ph---i ---------- --a---i-lp
Hummck --las----- ---a------ dn-s--v-v- ----l----- e------e-- r---v--qki -eifkkaghp ----qh---v ---------- --g---v-la
Dogmck --las----- ---a------ dn-t--v-v- ----l----- q------e-- r---v--qki -eifkkaghp ----eh---v ---------- --g---v-la
Rabmck --las----- ---a------ dn-s--v-v- ----l----- e------e-- r---v--qki -eifkkaghp ----eh---v ---------- --g---v-la
Ratmck --las----- ---a------ dn-s--v-v- ----l----- e------e-- r---v--qki -eifkkaghp ----eh---v ---------- --g---v-la
Moumck --las----- ---a------ dn-s--v-v- ----l----- e------e-- r---v--qki -eifkkaghp ----eh---v ---------- --g---v-la
Chimck --las----- ---a------ dn-t--v-v- ----l----- e------e-- r---v--kki -eifkkaghp ---teh---i ---------- --g---v-lp
Troutckt --las----- ---g------ dt-t--v-v- ----l----- q------e-- n---t--tki -tlfkdkgts ----eh---v ---------- --a---v-ip
Tormarck --las----- ---g------ nd-s--v-v- ----l----- q------e-- r---v--kki -eifvkagrg ----eh---v ---------- --g---v-ip
Torcalck --las----- ---g------ nd-t--v-v- ----l----- q------e-- r---v--kki -difvkagrg ----eh---v ---------- --g---v-ip
Xenlaeck3 --laa--g-- ---a------ dn-t--v-v- ----l----- q------e-- r---e--tki -eifkqaghp ----eh---v ---------- --g---a-lp
Xenlaeck4 --las----- ---a------ dn-t--v-i- ----l----- q------e-- n---t--tki -sifknkgsp ----qh---v ---------- --a---i-lp
Humubimick --taa----- ---a------ ne-s--i-v- ----t----- e------r-- e---r--kev -rliqergwe ----er---i ---------- --a---i-lp
Ratubimick --taa----- ---a------ ne-s--i-v- ----t----- e------r-- e---r--kkv -kliqergwe ----er---i ---------- --a---v-lp
Mouubimick --taa----- ---a------ ne-s--i-v- ----t----- e------r-- e---r--kev -kliqergwe ----er---i ---------- --a---i-lp
Chimiack s-taa----- ---a------ ne-t--i-i- ----t----- e------r-- e---r--kev -rliqergwe ----er---i ---------- --a---i-lp
Humsarmick --tca----- ---a------ yd-t--i-i- ----t----- e------r-- e---r--kev -rliqergwe ----er---i ---------- --a---vrip
Ratsarmick --tca----- ---a------ yd-t--i-i- ----t----- e------r-- e---r--kev -rliqergwe ----er---i ---------- --a---vrip
Chimibck --tca----- ---a------ nd-t--v-i- ----t----- e------r-- e---r--kev -rlikergwe ----er---v ---------- --a---v-lp
Spfckd1 --msar---- ---g------ ad-n--v-i- ----t----- ets----n-- k---n--nkv -nalkakgye -s--eh---v ---------- v-a---i-ip
Spfckd2 --taar-h-- --qg------ en-n--v-v- ----i----- e-d---ra-- k---e--qkf -qmikkdgke ----kh---v ---------- --a---v-lp
Spfckd3 --lcan---- --qg------ de-n--v-v- ----t----- e-s----r-- e---d--kkv -dsikskgyq ----eh---v ---------- --a---v-vp
Consensus LL--SGMARD WPDARGIWHN --K-FLVWVN EEDH-RVISM EKGGNMK-VF -RFC-GL--- E--------- FMWN-HLGYV LTCPSNLGTG LRAGVH-K-P
A G I I Q R I G A
F
3
4

Fig. 1

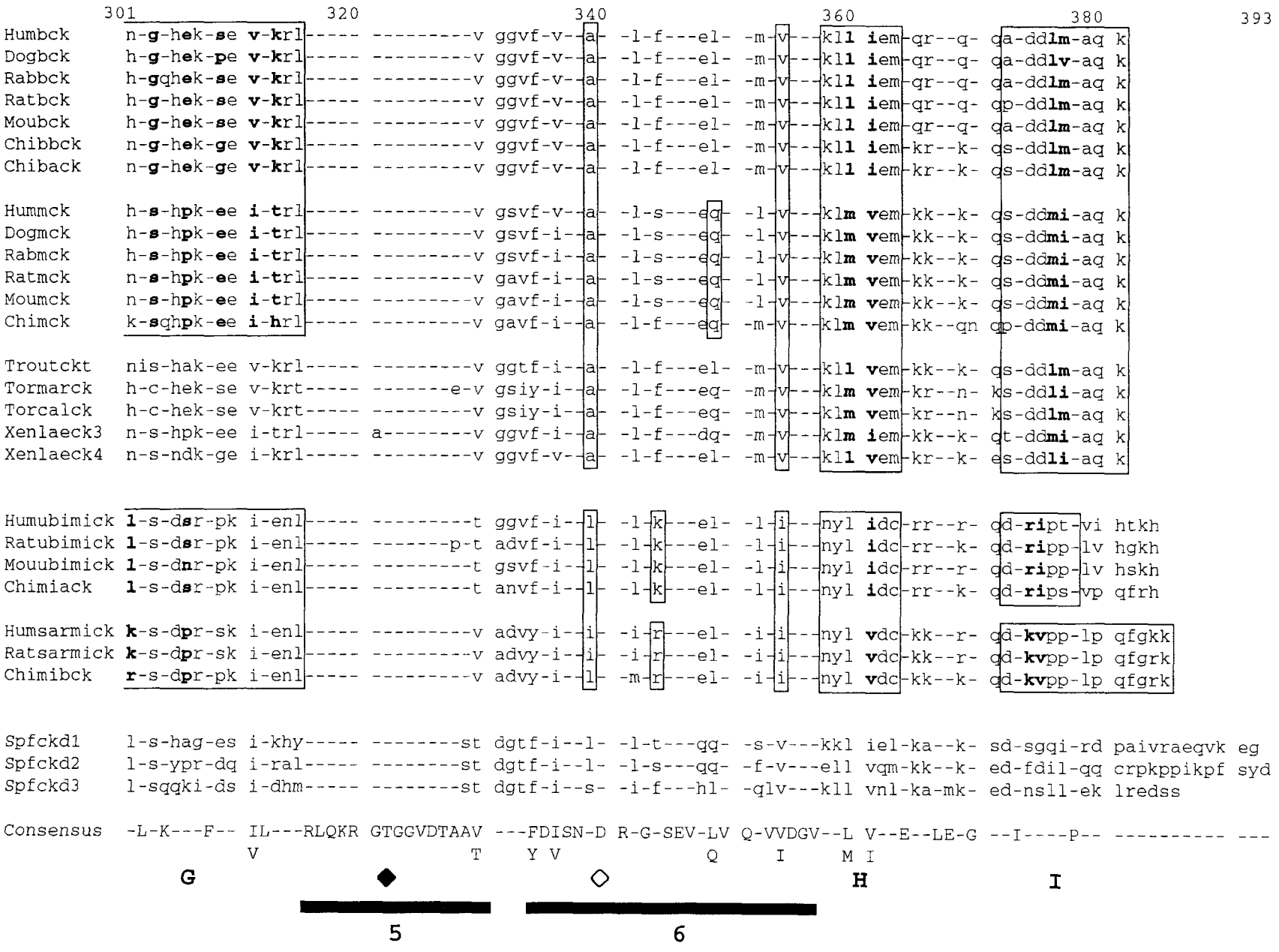
301
320
340
360
380
393
Humbck n-g-hek-se v-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll iem-qr--q- qa-ddlm-aq k
Dogbck h-g-hek-pe v-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll iem-qr--q- qa-ddlv-aq k
Rabbck h-gqhek-se v-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll iem-qr--q- qa-ddlm-aq k
Ratbck h-g-hek-se v-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll iem-qr--q- qp-ddlm-aq k
Moubck h-g-hek-se v-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll iem-qr--q- qa-ddlm-aq k
Chibbck n-g-hek-ge v-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll iem-kr--k- qs-ddlm-aq k
Chiback n-g-hek-ge v-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll iem-kr--k- qs-ddlm-aq k
Hummck h-s-hpk-ee i-trl----- ---------v gsvf-v--a- -l-s---eq- -l-v---klm vem-kk--k- qs-ddmi-aq k
Dogmck h-s-hpk-ee i-trl----- ---------v gsvf-i--a- -l-s---eq- -l-v---klm vem-kk--k- qs-ddmi-aq k
Rabmck h-s-hpk-ee i-trl----- ---------v gsvf-i--a- -l-s---eq- -l-v---klm vem-kk--k- qs-ddmi-aq k
Ratmck n-s-hpk-ee i-trl----- ---------v gavf-i--a- -l-s---eq- -l-v---klm vem-kk--k- qs-ddmi-aq k
Moumck n-s-hpk-ee i-trl----- ---------v gavf-i--a- -l-s---eq- -l-v---klm vem-kk--k- qs-ddmi-aq k
Chimck k-sqhpk-ee i-hrl----- ---------v gavf-i--a- -l-f---eq- -m-v---klm vem-kk--qn qp-ddmi-aq k
Troutckt nis-hak-ee v-krl----- ---------v ggtf-i--a- -l-f---el- -m-v---kll vem-kk--k- qs-ddlm-aq k
Tormarck h-c-hek-se v-krt----- -------e-v gsiy-i--a- -l-f---eq- -m-v---klm vem-kr--n- ks-ddli-aq k
Torcalck h-c-hek-se v-krt----- ---------v gsiy-i--a- -l-f---eq- -m-v---klm vem-kr--n- ks-ddlm-aq k
Xenlaeck3 n-s-hpk-ee i-trl----- a--------v ggvf-i--a- -l-f---dq- -m-v---klm iem-kk--k- qt-ddmi-aq k
Xenlaeck4 n-s-ndk-ge i-krl----- ---------v ggvf-v--a- -l-f---el- -m-v---kll vem-kr--k- es-ddli-aq k
Humubimick l-s-dsr-pk i-enl----- ---------t ggvf-i--l- -l-k---el- -l-i---nyl idc-rr--r- qd-ript-vi htkh
Ratubimick l-s-dsr-pk i-enl----- -------p-t advf-i--l- -l-k---el- -l-i---nyl idc-rr--k- qd-ripp-lv hgkh
Mouubimick l-s-dnr-pk i-enl----- ---------t gsvf-i--l- -l-k---el- -l-i---nyl idc-rr--r- qd-ripp-lv hskh
Chimiack l-s-dsr-pk i-enl----- ---------t anvf-i--l- -l-k---el- -l-i---nyl idc-rr--k- qd-rips-vp qfrh
Humsarmick k-s-dpr-sk i-enl----- ---------v advy-i--i- -i-r---el- -i-i---nyl vdc-kk--r- qd-kvpp-lp qfgkk
Ratsarmick k-s-dpr-sk i-enl----- ---------v advy-i--i- -i-r---el- -i-i---nyl vdc-kk--r- qd-kvpp-lp qfgrk
Chimibck r-s-dpr-pk i-enl----- ---------v advy-i--l- -m-r---el- -i-i---nyl vdc-kk--k- qd-kvpp-lp qfgrk
Spfckd1 l-s-hag-es i-khy----- --------st dgtf-i--l- -l-t---qq- -s-v---kkl iel-ka--k- sd-sgqi-rd paivraeqvk eg
Spfckd2 l-s-ypr-dq i-ral----- --------st dgtf-i--l- -l-s---qq- -f-v---ell vqm-kk--k- ed-fdil-qq crpkppikpf syd
Spfckd3 l-sqqki-ds i-dhm----- --------st dgtf-i--s- -i-f---hl- -qlv---kll vnl-ka-mk- ed-nsll-ek lredss
Consensus -L-K---F-- IL---RLQKR GTGGVDTAAV ---FDISN-D R-G-SEV-LV Q-VVDGV--L V--E--LE-G --I----P-- ---------- ---
V T Y V Q I M I
G
H
I
5
6

Fig. 1

ro [7], they have not been detected so far *in vivo*. Likewise, heterodimer formation between a mitochondrial and a cytosolic CK isoenzyme is also excluded *in vivo*. Among the cytosolic CK isoenzymes of mammals and birds, chicken B-CK seems to be unique, in as far as alternative splicing of the single B-CK gene produces two isoforms, B_a- and B_b-CK (see Fig. 1), differing in approximately 20 of the first 50 N-terminal amino acids [8]. Further heterogeneity of B-CK, as evidenced, for example, by 2D-electrophoresis experiments [9], was shown to be due to alternative ribosomal initiation at internal transcriptional start sites (at Met-12, Met-30, Leu-36 and Met-70 of chicken B_b-CK; [10]) or to post-translational phosphorylation (as demonstrated for chicken, mouse and rat B-CK; [10–14]).

Lower vertebrates like fish and frogs also contain several CK isoenzyme loci, all giving rise to dimeric molecules with an M_r of approximately 80'000. While some frogs were suggested to display a CK isoenzyme system similar to mammals and birds [15–19], the CK isoenzymes of *Xenopus* and other pipid frogs do not fit into the M-/B-/Mi_a-/Mi_b-CK classification. The five CK isoforms observed in pipid frogs are therefore termed CK-I to CK-V and give rise to up to nine different bands on zymograms [20]. CK-I and possibly also CK-V (see [6]) are located within the mitochondria, while CK-II, CK-III and CK-IV are clearly cytosolic.

The four CK isoforms of teleost fish are termed CK-A to CK-D and are all of cytoplasmic origin. CK-A, CK-C and CK-D are expressed predominantly in striated muscle, stomach, and testis, respectively, while CK-B is expressed ubiquitously or is confined to neural tissues [16]. The trout CK listed in Fig. 1, showing enhanced testicular expression [21], thus most probably represents a CK-D isoform. In primitive fish species, only two CK isoforms are found which obviously correspond to CK-A and CK-C of teleost fish. While CK-A is again restricted to striated muscle, CK-C in primitive fish is expressed ubiquitously [16]. Since *Torpedo* electrocytes were shown by isoenzyme and 2D-electrophoresis to contain the same major CK isoenzyme as muscle [22–25], the two *Torpedo* CKs listed in Fig. 1 very likely represent CK-A isoforms.

Based on a comparison of the tissue-specificity of expression of the various isoenzymes, it has been hypothesized that CK-II of frogs and CK-A of fish correspond to M-CK of mammals and birds, while CK-IV and CK-C would correspond to B-CK [20]. The other CK isoenzymes of frogs and fish seem to be the result of (additional) gene duplications. Clearly, more biochemical and molecular genetic work is needed to unequivocally prove the phylogenetic relationships of the various CK isoenzymes.

Sea urchin spermatozoa contain two different CK isoenzymes. The mitochondrial CK is confined to the midpiece of the sperm and is, like the vertebrate Mi-CKs, an octameric molecule composed of subunits with an M_r of 44'000–50'000 [26–29]. The tail CK isoenzyme, however, is located along the sperm tail and is a monomeric protein with an M_r of 140'000–155'000 [26–28, 30] which most likely originates from triplication of an ancestral CK gene ([31]; see Fig. 1). These findings suggest that the separation into mitochondrial and cytosolic CK isoenzymes occurred before the start of divergence between echinoderms and vertebrates.

Relatively little is known about the invertebrate guanidino kinases. While for the lobster tail muscle ArgK shown in Fig. 1 [32], substrate specificity and biochemical properties have been investigated intensively (see [33]), the physiological substrate of the duplicated guanidino kinase from the trematode *Schistosoma mansoni* is not known [34]. Analysis of homogenates simply indicated that this guanidino kinase displays low CK, but no ArgK activity.

Table 1. Amino acid sequence comparisons of the guanidino kinases known so far

B-CKs among each other	88–98
M-CKs among each other	89–99
B-CKs versus M-CKs	77–82
Ubiquitous Mi-CKs among each other	91–98
Sarcomeric Mi-CKs among each other	89–96
Sarcomeric versus ubiquitous Mi-CKs	82–84
Mi-CKs versus B- and M-CKs	60–65
X. laevis CK-III versus B-CKs	79–83
X. laevis CK-III versus M-CKs	87–91
Torpedo CKs versus B-CKs	76–80
Torpedo CKs versus M-CKs	83–86
X. laevis CK-IV versus B- and M-CKs	84–89
Trout CK versus B- and M-CKs	78–87
Trout/*Torpedo*/*X. laevis* CKs versus vertebrate Mi-CKs	61–67
Sea urchin CK repeats among each other	66–70
Sea urchin CK repeats versus vertebrate CKs	60–69
Lobster ArgK versus all CKs	38–44
Lobster ArgK versus Flukeckp1	45
Lobster ArgK versus Flukeckp2	40
Flukeckp1 versus all CKs	33–38
Flukeckp2 versus all CKs	32–35

The degree of identity (in %) in primary structure between any two guanidino kinases was determined using the program 'Distances' of the GCG software package, with the threshold for a match set at 1.5. For discussion of the results see the text.

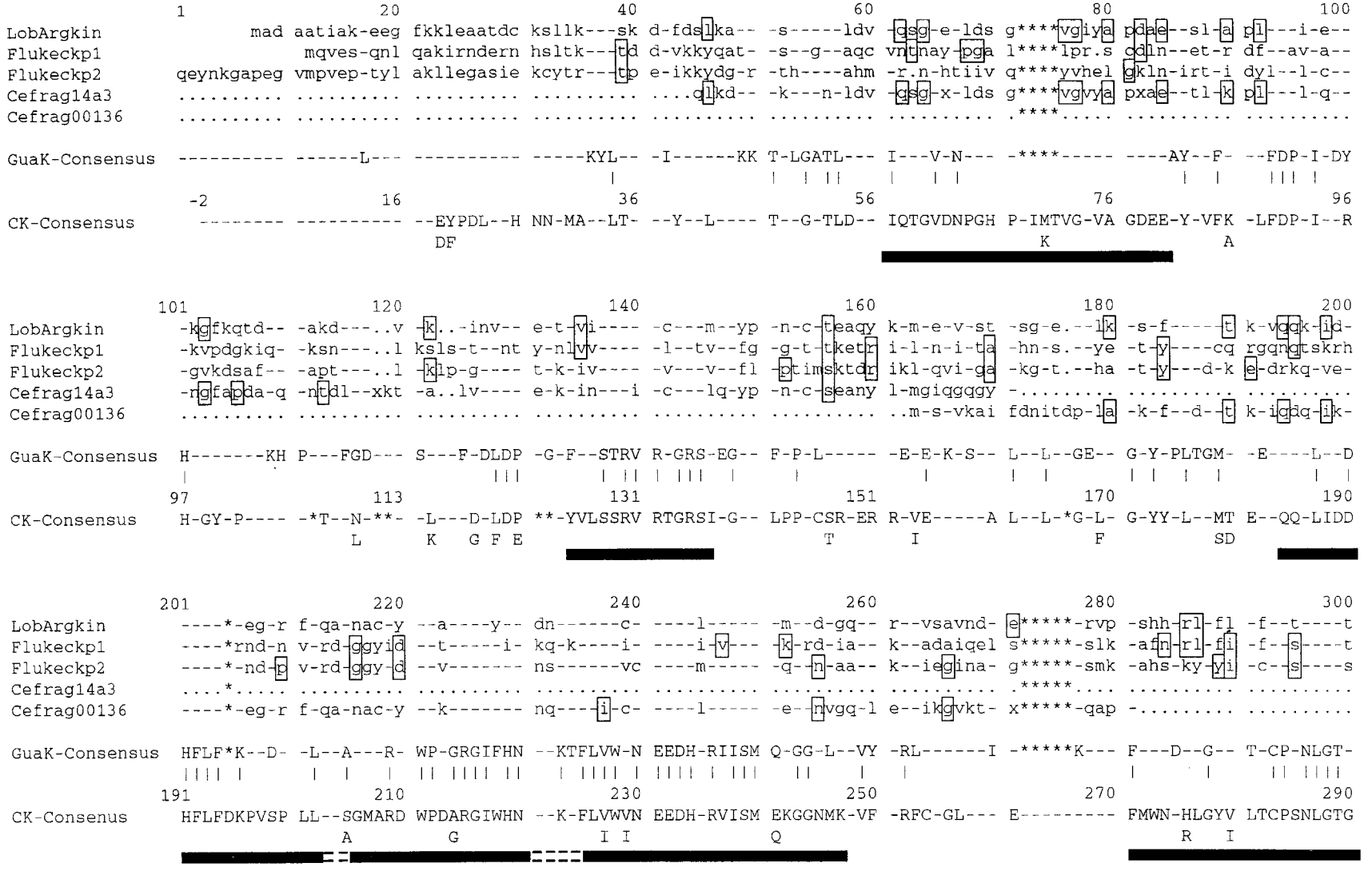
LobArgkin
Flukeckp1
Flukeckp2
Cefrag14a3
Cefrag00136
GuaK-Consensus
CK-Consensus

Fig. 2

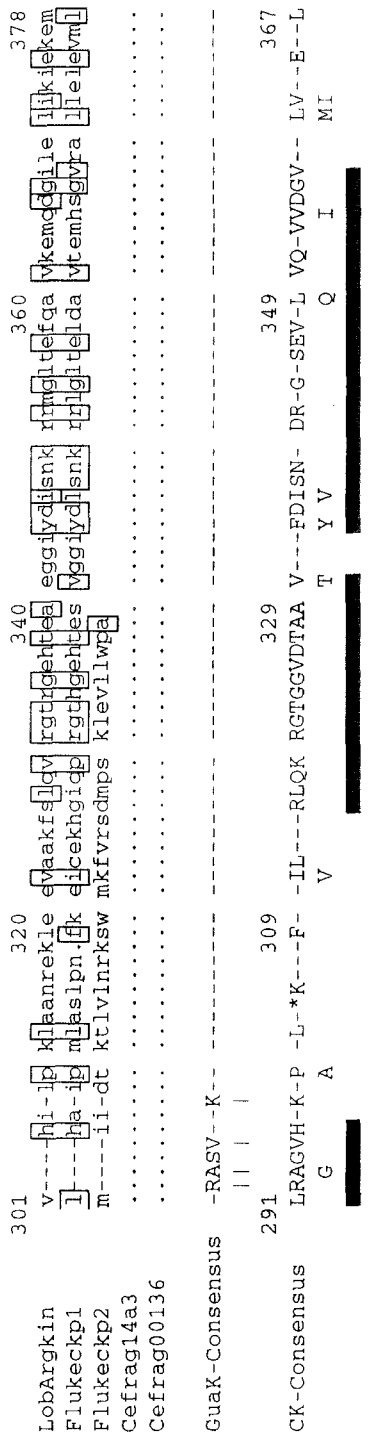

Fig. 2. Comparison of the invertebrate guanidino kinases with the CK consensus sequence. The primary structures of lobster ArgK [32], of a guanidino kinase from the trematode *Schistosoma mansoni* (Flukeckp1 represents residues 1–352 and Flukeckp2 represents residues 353–675 of this 'duplicated' guanidino kinase; [34]), and of two guanidino kinase sequence fragments from *Caenorhabditis elegans* (which were derived from single reads of expressed sequence tags; [95, 96]) were aligned and compared to the CK consensus sequence from Fig. 1.

The dots stand either for gaps that were introduced in order to allow an optimal sequence alignment or for parts of the primary structure that have not been sequenced so far (N- and C-termini of the *C. elegans* fragments). The hyphens in the invertebrate guanidino kinase sequences stand for amino acids that are identical in three out of the five sequences, with the corresponding amino acid being written in the guanidino kinase (GuaK) consensus sequence below. If no consensus is observed, or if an amino acid differs from the consensus sequence, it is written in the respective sequence itself.

Stars represent gaps that were introduced for an optimal alignment of the invertebrate guanidino kinases to the CK consensus sequence. Identical amino acids between the guanidino kinase and CK consensus sequences are indicated by vertical lines, while boxed residues represent additional amino acids in the individual guanidino kinase sequences that are identical to the CK consensus sequence.

Programs and databases

All protein sequences compared in this study are derived from nucleic acid sequences. They were extracted from the GenBank and EMBL nucleotide sequence libraries using the programs of the GCG software package [35, 36], with the exception of the two *Xenopus laevis* [37, 38] and the chicken Mi_a-CK sequences [39, 40]. The alignment was done using the programs 'Bestfit', 'Pileup', 'Lineup' and 'Pretty' available in the GCG software package. The degree of identity between any two protein sequences (Table 1) was calculated using the program 'Distances', with the threshold for a match being set at 1.5 based on the default amino acid comparison table of the GCG software package [41, 42]. The phylogenetic tree of the guanidino kinases (Fig. 3) was constructed using the T.NJM26 program of the MacT (Macintosh's Trees) program package [43]. Tree construction is based on a distance matrix method commonly known as neighbor joining method [44]. Sequences were aligned in the manner shown in Fig. 2, i.e. gaps were allowed, but were not taken into account for the distance calculations. The distance between any two sequences was calculated by summing the *minimal* number of base substitutions required to convert one amino acid in another at each site compared, and by dividing this sum by the total number of amino acids compared ('Fitch' option; see [45]). Therefore, the distance represents the *average* number of mutations per site needed to get from one sequence to the other.

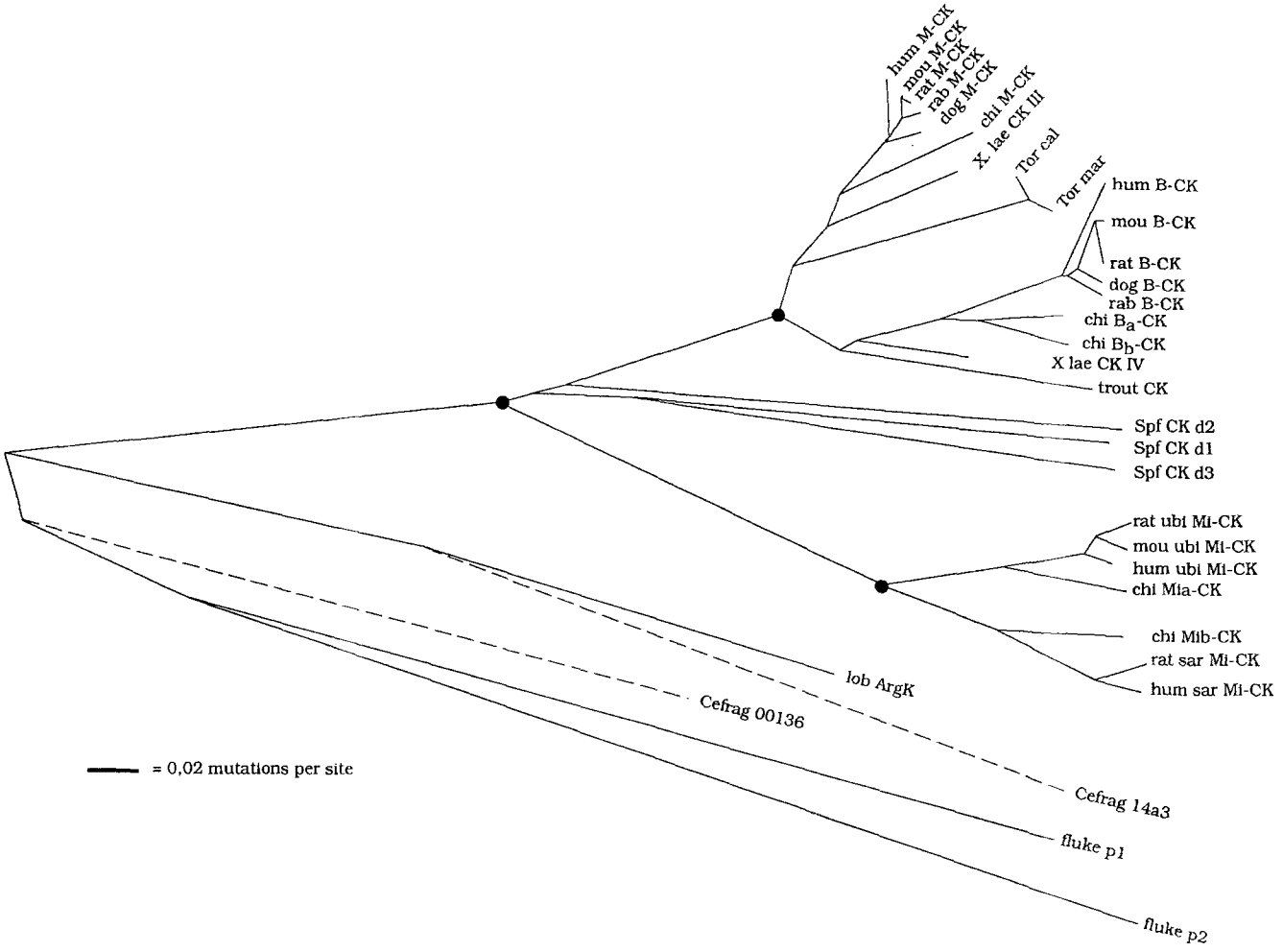

Fig. 3. Evolutionary tree of the guanidino kinases. The phylogenetic tree of the guanidino kinase sequences shown in Figs. 1 and 2 was constructed as described under 'Programs and Databases'. The distance between any two sequences approximates the *average* number of (base) mutations per site needed to convert one sequence into the other. The bar at the bottom represents an 'evolutionary distance' of 0.02 mutations per site. The three dots mark the three principal gene duplication events discussed in the text.

Comparison of the known guanidino kinase protein sequences – evolutionary relationships

Figure 1 shows an alignment of the 28 CK protein sequences or sequence repeats known so far, and Fig. 2 shows an alignment of the invertebrate guanidino kinases to the CK consensus sequence. The three repeats of the triplicated sea urchin tail CK (Spfckd1–3; Fig. 1) and the two repeats of the duplicated guanidino kinase from *Schistosoma mansoni* (Flukeckp1,2; Fig. 2) are analysed separately. Some 20–80 amino acids at the N-termini of the *Xenopus* CKs (Fig. 1) escaped sequencing so far, and for the guanidino kinase(s) of *Caenorhabditis elegans* (Fig. 2), only the listed fragments can be aligned reasonably. The amino acid identity scores between different guanidino kinases are summarized in Table 1, and a tentative evolutionary tree for the guanidino kinases is shown in Fig. 3.

In mammals and birds, the known CK sequences can easily be grouped into the four isoenzyme classes (B-, M-, Mi_a- and Mi_b-CK) that have been postulated earlier on the basis of tissue distribution, electrophoretic behaviour and biochemical characterization. The amino acid identities within each single class range from 88 to 99%. Evidently, a higher degree of homology is observed between the B- and the M-CKs (77–82%) on one hand and between the Mi_a- and the Mi_b-CKs (82–84%) on the other hand than between the cytosolic and the mitochondrial CK isoforms (60–65%), indicating that during evolution, a first gene duplication event resulted in a primordial cytosolic and a primordial mitochondrial CK isoenzyme (Fig. 3). Gene duplications giving rise to the

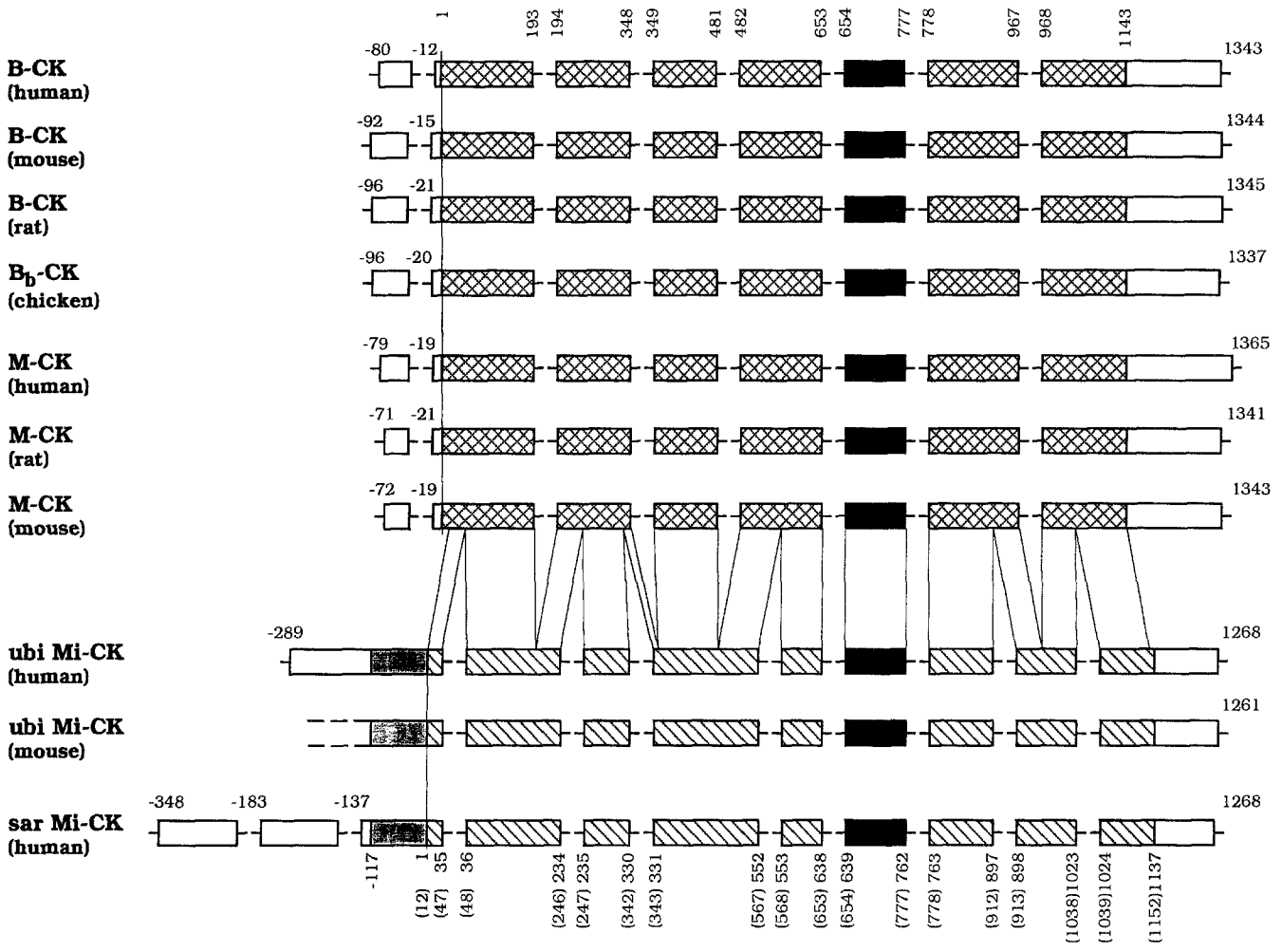

Fig. 4. Schematic representation of the gene structures of the vertebrate CK isoenzymes investigated so far. The lengths of the exons are drawn in scale, while those of the introns are not. The protein coding regions within the exons of cytosolic and mitochondrial CK isoenzymes are cross-hatched and hatched, respectively, while the single exon which is conserved in all mammalian and avian CK isoenzymes (exon 8 of sarcomeric Mi-CK, exon 6 of the other CKs) is represented by a filled box. The coding region for the Mi-CK transit peptide (which is responsible for the import into the mitochondrial intermembrane space) is shaded. Numbers above and below the respective exons refer to the corresponding nucleotide positions within the protein coding (cDNA) sequences. In the case of the cytosolic CK isoenzymes, the first nucleotide of the start codon is given number 1, while for the Mi-CKs, 1 represents the first nucleotide of the codon for the very N-terminal amino acid of the *mature* protein (i.e. lacking the transit peptide). Below the numbers for the Mi-CK isoenzymes, the respective positions within the cDNA sequences of the cytosolic CKs are listed in brackets.

The respective gene structures are from: human B-CK [47–49]; mouse B-CK [52]; rat B-CK [50]; chicken B_b-CK [8]; human M-CK [51]; rat M-CK [50]; mouse M-CK [46]; human ubiquitous Mi-CK [53]; mouse ubiquitous Mi-CK [55]; and human sarcomeric Mi-CK [54]. For simplicity, the additional exon coding for the amino terminus of chicken B_a-CK [8] has not been considered in this diagram.

multiple cytosolic and mitochondrial CK isoenzymes observed in vertebrates must have occurred at a later stage during evolution. This interpretation is favoured by the likely presence of a single cytosolic and a single mitochondrial CK isoenzyme in echinoderm species [27] and by the gene structures of the mammalian and avian CK isoenzymes investigated so far (Fig. 4). The lengths of the exons as well as the location of the splice sites within the coding region are identical for all M- and B-CKs on one hand [8, 46–52] and for all Mi-CKs on the other hand [53–55]. If, however, the gene structure of the Mi-CKs is compared to that of the cytosolic CKs, pronounced differences are observed, with only the location of a single exon (exon 8 of human Mi_b-CK, exon 6 of the Mi_a- and the cytosolic CKs; coding for amino acids 219–259 in Fig. 1) being conserved.

The *X. laevis* isoenzymes CK-III and CK-IV, the two *Torpedo* CKs as well as the trout CK can be assigned to the branch of the cytosolic vertebrate CK isoenzymes (Table 1; Fig. 3), since the amino acid identities to the B-

and M-CKs of mammals and birds (76–91%) are much higher than to the Mi-CKs (60–67%). *X. laevis* CK-III and the *Torpedo* CKs clearly correspond to mammalian and avian M-CKs, while *X. laevis* CK-IV and the trout CK seem to be related somewhat more closely to B-CK than to M-CK (Table 1; Fig. 3). The trout CK is peculiar in as far as the first 17 amino acids at the N-terminus (see Fig. 1) differ completely from those of the other vertebrate cytosolic CK isoenzymes, perhaps with the exception of chicken B_a-CK.

The homologies of the repeats of the triplicated sea urchin CK with each other (66–70% identity) are only slightly higher than those of the sea urchin CK repeats to the vertebrate CKs (60–69%) and of the vertebrate cytosolic CKs to the Mi-CKs (60–67%), implicating that branching into a mitochondrial and a cytosolic CK isoenzyme as well as the triplication of the cytosolic sea urchin CK occurred approximately at the time when echinoderms and vertebrates started to diverge. The two repeats of the guanidino kinase from the parasitic trematode *Schistosoma mansoni* are more homologous to lobster tail muscle ArgK (40–45% identity) than to all CKs known (32–38%). Since extracts of *Schistosoma* cercaria, however, display low CK, but no ArgK activity [34], a different guanidino substrate specificity has to be postulated for this guanidino kinase [32]. The two sequence fragments from *Caenorhabditis elegans* display more pronounced homology to lobster ArgK (63% for the Ce14a3 fragment, 68% for the Ce00136 fragment) than to CK (32–42%, 44-50%) or to the *Schistosoma* guanidino kinase (26 and 33%, 40 and 51%), suggesting that they represent ArgK isoenzymes. Finally, amino acid sequencing of tryptic fragments of a guanidinoacetate (glycocyamine) kinase from the polychaete *Neanthes diversicolor* revealed higher homology to the corresponding parts of the CKs than of the other guanidino kinases [56]. As a matter of fact, guanidinoacetate, among the natural guanidino compounds, also displays the most pronounced structural homology to Cr.

In summary, the evolutionary tree of the guanidino kinases shown in Fig. 3 'visualizes' the homologies observed between the sequences of the various guanidino kinases and, in addition, agrees with most biochemical findings. Nevertheless, it should be regarded as tentative. Clearly, a much larger number of guanidino kinases, especially in the invertebrate phyla, have to be cloned and/or sequenced before a correct tracking of the evolutionary relationships as well as a reliable estimation of the time points in evolution when gene duplication events occurred will be possible. However, one clear-cut postulate of the analyses presented here is that all vertebrates, including fish, amphibia, and reptiles, contain mitochondrial CK isoenzymes.

It has been hypothesized earlier that phosphorylarginine (PArg) and ArgK are phylogenetically older than the other phosphagens and guanidino kinases (see [2, 3]). The facts that (i) ArgK is by far the most widespread guanidino kinase in invertebrate phyla, that (ii) Arg is a key component of basic metabolism, while the other guanidino compounds necessitated the evolution of additional enzymes for their biosynthesis, and that (iii) most ArgKs are monomeric proteins in fact support the notion that a primordial monomeric ArgK represents the common ancestor of all guanidino kinases [2]. However, the postulate that phosphorylcreatine (PCr) represents a functional improvement over PArg, this explaining the *apparent* switch from ArgK to CK at the transition from invertebrates to vertebrates [3], is untenable since PCr and CK are found in a large variety of invertebrate spermatozoa [2]. In these invertebrates, PCr is often the sole phosphagen in the sperm cells, while other tissues contain PArg or other phosphagens. Due to the different thermodynamic properties of the various phosphagens, it is more reasonable to assume that PCr is better suited for *some* species and cell types, while PArg and the other phosphagens are advantagous for others [4, 6].

Recently, it has been postulated, on the basis of proteolysis and small-angle X-ray scattering experiments as well as of sequence comparisons [57–61] that guanidino kinases are structurally similar to 3-phosphoglycerate kinase. Accordingly, each guanidino kinase protomer would consist of two domains with M_r's of 20'000–25'000 which are separated by a deep cleft. The two substrates are bound on either side of the cleft, and binding of the second substrate is likely to result in a closure of the cleft, thus allowing a direct in-line transfer of a phosphate group during catalysis, in an environment excluding water. Whether and how closely the guanidino kinases and 3-phosphoglycerate kinase are also related evolutionarily remains to be established.

CK framework

The amino acid sequences of the CK isoenzymes display six blocks with extensive homologies, flanked by seven regions that are more variable (Fig. 1). Among the latter are the N-terminus and the C-terminus. It has been suggested previously that the highly conserved parts form

the 'framework' of the molecule, being involved in basic functions like substrate binding and catalysis, while the variable segments may be responsible for isoenzyme- or species-specific functions like oligomer formation or interaction of the CK isoenzymes with subcellular structures (either with membranes or with other proteins; [62]). The six highly conserved blocks are indicated by bars below the CK consensus sequence in Fig. 1 and are numbered 1 to 6. Compared to blocks 1–5, block 6 is somewhat less well defined, but still 19 out of 24 amino acids are conserved.

As can be seen in Fig. 1, several residues in blocks 1–6 are only imperfectly conserved. While some of the 'deviations' may be real, others are certainly due to sequencing artifacts. This has, for instance, been demonstrated already [51] for six amino acids (5 of which are in the conserved blocks 2, 3 and 5) of the human M-CK sequence published by Perryman *et al.* [63].

Block 4 contains the highly reactive and absolutely conserved Cys-283, alkylation of which is always paralleled by a very pronounced or even complete loss of enzymatic activity (for references see [64]). In earlier publications, it has therefore often been suggested that this residue is 'essential' for and possibly directly involved in catalysis. Recently, however, site-directed mutagenesis of Cys-283 of chicken Mi_b-CK demonstrated that this residue is not involved in catalysis itself, but that it is necessary for synergism in substrate binding and that it may also provide a negative charge for maximum enzymatic activity [64]. Block 4 also comprises a putative adenine nucleotide binding motif (glycine-rich loop) LGXGXXGXV [65, 66], but it is not yet clear whether it is functional or not. The importance of block 4 for CK and guanidino kinase function in general is further supported by the fact that short peptide sequences around Cys- 283 of taurocyamine kinase, lombricine kinase and glycocyamine kinase also show high sequence conservation [56, 67, 68].

Blocks 1 and 3 contain the two residues Cys-74 and Lys-196 which, on the basis of cross-linking experiments with rabbit MM-CK, have been implicated to be structurally close to the highly reactive Cys-283 [69, 70]. While Lys-196 is absolutely conserved in all CK sequences (Fig. 1), Cys-74 is replaced by an Ala in rat and mouse B-CK, and by a Met or Leu in all Mi-CK and sea urchin CK sequences. This suggests that at least the latter residue is not essential for CK function.

Asp-340 in block 6 of chicken Mi_b-CK has been labelled by an alkylating ATP analogue, and it has been suggested that this residue is involved in the binding of

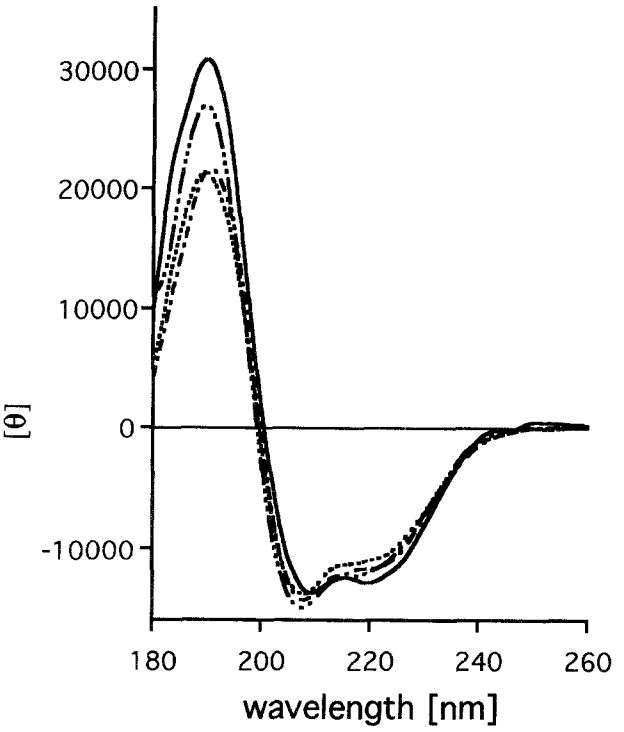

Fig. 5. Far-UV CD spectra of guanidino kinases. Spectra of octameric chicken heart Mi_b-CK (——), chicken brain BB-CK (-···-), rabbit muscle MM-CK (···) and lobster tail muscle ArgK (-··-) were recorded on a Jasco J-710 dichrograph at a protein concentration of 0.5 mg/ml in 50 mM potassium phosphate buffer, pH 7.0. Cylindrical quartz cells with 0.02 cm path length were used. Residual molar ellipticities ([Θ]) are given in deg cm^2 $dmole^{-1}$.

the Mg^{2+} ion [71]. Again, this residue is absolutely conserved in all CK sequences. Finally, no functions have yet been assigned to blocks 2 and 5, except that they contain, together with block 3, three putative phosphorylation sites (Thr-133, Ser-239 and Thr-322) which are absolutely conserved among all CKs and are also found in some of the invertebrate guanidino kinase sequences.

A comparison of the CK framework with the other guanidino kinase sequences reveals that block 1 is 'missing' in the invertebrate guanidino kinases, while in blocks 2–6, pronounced homology is observed (Fig. 2). It is therefore tempting to speculate that block 1 determines the guanidino substrate specificity of the guanidino kinases and, in particular, the Cr specificity of the CKs. In accordance with this notion, it has recently been postulated, on the basis of biochemical evidence, that the N-terminal and C-terminal halves (= domains) of the CK molecule are involved in Cr and MgATP binding, respectively [61]. In the invertebrate guanidino kinases known so far, Cys-283 and Lys-196 (only in Flukeckp1, Lys-196 is replaced by Arg) are also highly conserved, while Cys-74 and Asp-340 are not.

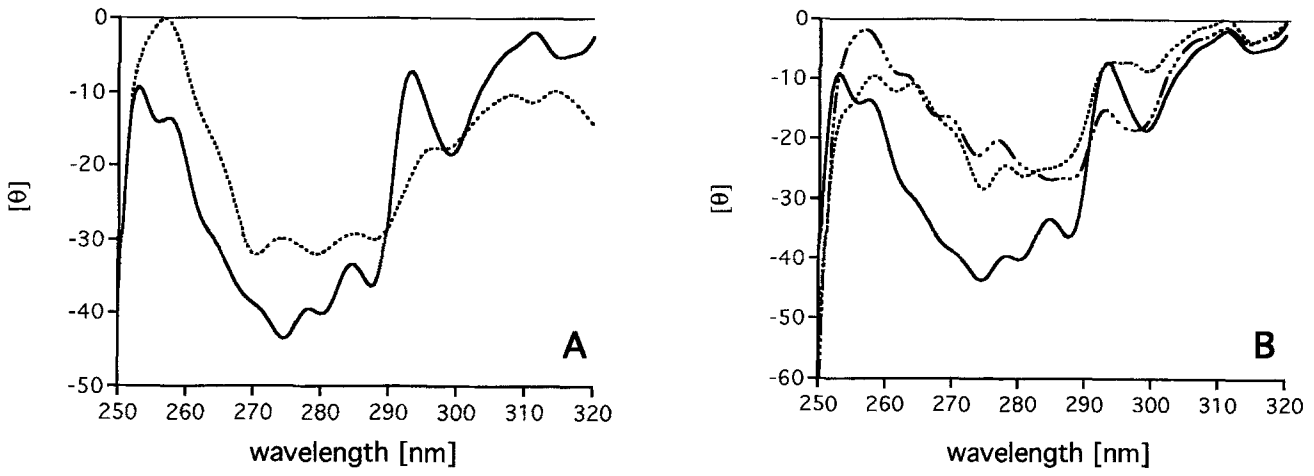

Fig. 6. Near-UV CD spectra of guanidino kinases. (A) Creatine (Mi_b-CK, ——) and arginine kinase (lobster, – – –). (B) Mi_b-CK (——), MM-CK (···), and BB-CK -···-) [same proteins as in Fig. 5]. The spectra were recorded on a Jasco J-710 dichrograph at a protein concentration of 1 mg/ml in 50 mM potassium phosphate buffer, pH 7.0. Cylindrical quartz cells with 1 cm path length were used. Residual molar ellipticities ([(Θ]) are given in degcm^{2}dmole^{-1}.

Isoenzyme-specific boxes

While the six highly conserved blocks of the 'CK framework' often do not allow a distinction between isoenzymes, the more variable segments in between frequently contain single residues or even extended peptide stretches that either allow a clear-cut distinction between cytosolic and mitochondrial CKs or are specific for any one isoenzyme (M-, B-, Mi_a- or Mi_b-CK). The most instructive of these residues or peptide stretches are boxed in Fig. 1, with purely isoenzyme-specific residues within larger boxes being written in bold. Extended stretches (≥ 5 residues) with isoenzyme-specific sequence patterns are termed 'diagnostic boxes' (A, B, C, ... I), since they are likely to allow the correct assignment of a new vertebrate CK sequence to one of the four isoenzyme classes (boxes A, B, D, G, H, I) or at least to either the mitochondrial or cytosolic CK isoenzymes (boxes C, E, F). For instance, screening of the GenEMBL sequence data libraries with blocks A and I yielded exclusively the expected CK sequences with 100% identity, showing that they are absolutely specific for the respective CK isoenzymes.

It is very likely that the 'diagnostic' residues or boxes are responsible for isoenzyme-, cytosolic CK-, or Mi-CK-specific properties, e.g. for octamer formation, membrane binding, interaction with the myofibrillar M-band or other subcellular sites. Accordingly, limited proteolysis and site-directed mutagenesis experiments have shown that the very N-terminus of chicken Mi_b-CK, particularly the positively charged residues 5–7, is important for octamer formation and stability [72]. Further mutagenesis experiments on chicken Mi_b-CK have shown that Trp-269 in block F of Fig. 1 is also important for octamer formation and stability [73], indicating that blocks B and F may form the complementary surface areas on neighboring dimers that interact with each other within the Mi-CK octamer. None of the other diagnostic boxes has been linked with a function of the molecule so far.

Structural homologies in the guanidino kinase family – conservation of secondary structure and of aromatic side chains

The high sequence homology among the members of the guanidino kinase family is paralleled by a number of structural similarities which are revealed by spectroscopic investigations. Circular dichroism (CD) measurements are particularly well suited for a structural comparison of the guanidino kinases.

Secondary structure compositions of proteins can be evaluated and compared by measuring CD spectra in the far-UV wavelength range ($\lambda \leq 240$ nm). Using a library of reference spectra of proteins with known three-dimensional structures, the secondary structure composition of the protein of interest can be computed by a variety of algorithms [74]. The far-UV CD spectra of the guanidino kinases reveal a high α-helical content, the spectra closely resembling those of lysozyme or myoglobin. All guanidino kinases, including the mammalian

and avian octameric Mi-CK and dimeric cytosolic CK isoenzymes, as well as the invertebrate (mostly monomeric) arginine, lombricine and taurocyamine kinases, display nearly superimposable far-UV CD spectra (Fig. 5 and [75]), indicating an almost perfect conservation of secondary structure elements. This leads to the conclusion that all the distinctive properties of the individual members of the family (guanidino substrate specificity, ability to form oligomeric molecules, specific localization within the cell and association with subcellular structures) do not require major adaptations of secondary structure, but can be achieved by relatively small differences in primary sequence (e.g. the isotype-specific sequence boxes of the CKs described above).

Since the far-UV CD spectra of the guanidino kinases are nearly indistinguishable, a representative secondary structure calculation was performed for chicken Mi_b-CK, using the variable selection algorithm introduced by Manavalan & Johnson [76], with 22 proteins in the original reference database. The procedure resulted in a prediction of 37% α-helix, 30% antiparallel β-sheet, no parallel β-sheet, 15% turns, and 18% other structures.

From CD spectra in the UV wavelength range above 240 nm (near-UV), information about the aromatic amino acid residues, in particular Tyr and Trp, can be obtained. Although the near-UV CD spectra of the various guanidino kinases (Fig. 6; see also [75]) show distinctive characteristics, they also exhibit several common, conserved features. The most obvious one is a conserved negative Cotton band at 295–300 nm, which, from its position, can clearly be assigned to (a) Trp residue(s). Since magnetic resonance, fluorescence [77], CD [78], and chemical modification data [79] have shown that in CK, a Trp residue positioned close to the adenine nucleotide substrate is essential for enzyme activity, it might be assumed that the CD band originates from this structurally and functionally conserved Trp. The only two Trp residues common to ArgKs and CKs are amino acids 211 and 228 (numbering according to Fig. 1). They are the only Trp residues present in lobster ArgK, whereas CK has two (cytosolic CKs) or three (Mi-CKs) additional indole side chains. Consequently, either residue 211 or 228 should be the Trp essential for catalytic activity. Site-directed mutagenesis studies on Mi_b-CK have indeed confirmed that Trp-228 is essential for catalysis, with even the conservative replacement by Phe leading to ≥ 99% inactivation. However, these experiments further showed that Trp-211 is also important for the structural integrity of the active site and that in fact this latter residue is the origin of the Cotton band at 295–300 nm [73].

In the region below 290 nm, all CKs and ArgKs show a characteristic pattern of three negative Cotton peaks at 288, 280 and 274 nm, superimposed on a broad negative band extending from 255 to 290 nm. This region mainly reflects the environment of Tyr residues. For CK, an involvement of Tyr in enzyme activity has also been demonstrated [80], suggesting that the common CD pattern reflects the structural and functional conservation of tyrosyl side chains among the guanidino kinases.

Acknowledgement

Prof. H.M. Eppenberger is gratefully acknowledged for continuous support. We are also indebted to Dr. M. Messerli, Prof. R. Rieger (Innsbruck) and D. Gilligan for valuable discussion and comments on the manuscript. This work was supported by research grants from the ETH Zürich (No. 0-20-064-90 to J.-C.P.), from the Swiss National Science Foundation (No. 31-27756.89 to J.-C.P., No. 31-33907.92 to T.W., postdoctoral fellowship No. 823A-037106 to M.W.), from the Swiss Society for Muscle Diseases and from the Helmut Horten Foundation (to T.W.).

References

1. Robin Y: Biological distribution of guanidines and phosphagens in marine annelida and related phyla from California, with a note on pluriphosphagens. Comp Biochem Physiol 12: 347–367, 1964
2. Watts DC: Evolution of phosphagen kinases. In: E Schoffeniels (ed.) Biochemical Evolution and the Origin of Life. North-Holland Publishing Company, Amsterdam and London, 1971, pp 150–173
3. Watts DC: Evolution of phosphagen kinases in the chordate line. Symp Zool Soc Lond 36: 107–127, 1975
4. Ellington WR: Phosphocreatine represents a thermodynamic and functional improvement over other muscle phosphagens. J Exp Biol 143: 177–194, 1989
5. Eppenberger HM, Dawson DM, Kaplan NO: The comparative enzymology of creatine kinases. J Biol Chem 242: 204–209, 1967
6. Wyss M, Smeitink J, Wevers RA, Wallimann T: Mitochondrial creatine kinase: A key enzyme of aerobic energy metabolism. Biochim Biophys Acta 1102: 119–166, 1992
7. Wyss M, Schlegel J, James P, Eppenberger HM, Wallimann T: Mitochondrial creatine kinase from chicken brain. J Biol Chem 265: 15900–15908, 1990
8. Wirz T, Brändle U, Soldati T, Hossle JP, Perriard J-C: A unique chicken B-creatine kinase gene gives rise to two B-creatine kinase isoproteins with distinct N termini by alternative splicing. J Biol Chem 265: 11656–11666, 1990

9. Rosenberg UB, Eppenberger HM, Perriard J-C: Occurrence of heterogenous forms of the subunits of creatine kinase in various muscle and nonmuscle tissues and their behaviour during myogenesis. Eur J Biochem 116: 87–92, 1981
10. Soldati T, Schäfer BW, Perriard J-C: Alternative ribosomal initiation gives rise to chicken brain-type creatine kinase isoproteins with heterogeneous amino termini. J Biol Chem 265: 4498–4506, 1990
11. Mahadevan LC, Whatley SA, Leung TKC, Lim L: The brain isoform of a key ATP-regulating enzyme, creatine kinase, is a phosphoprotein. Biochem J 222: 139–144, 1984
12. Quest AFG, Soldati T, Hemmer W, Perriard J-C, Eppenberger HM, Wallimann T: Phosphorylation of chicken brain-type creatine kinase affects a physiologically important kinetic parameter and gives rise to protein microheterogeneity *in vivo*. FEBS Lett 269: 457–464, 1990
13. Chida K, Kasahara K, Tsunenaga M, Kohno Y, Yamada S, Ohmi S, Kuroki T: Purification and identification of creatine phosphokinase B as a substrate of protein kinase C in mouse skin *in vivo*. Biochem Biophys Res Commun 173: 351–357, 1990
14. Hemmer W, Skarli M, Perriard J-C, Wallimann T: Effect of okadaic acid on protein phosphorylation patterns of chicken myogenic cells with special reference to creatine kinase. FEBS Lett 327: 35–40, 1993
15. Eppenberger ME, Eppenberger HM, Kaplan NO: Evolution of creatine kinase. Nature 214: 239–241, 1967
16. Fisher SE, Whitt GS: Evolution of isozyme loci and their differential tissue expression. J Mol Evol 12: 25–55, 1978
17. Fisher SE, Shaklee JB, Ferris SD, Whitt GS: Evolution of five multilocus isozyme systems in the chordates. Genetica 52/53: 73–85, 1980
18. Klemann SW, Pfohl RJ: Creatine phosphokinase in *Rana pipiens*: Expression in embryos, early larvae and adult tissue. Comp Biochem Physiol 73B: 907–914, 1982
19. Legssyer A, Arrio-Dupont M: Mitochondrial isoenzyme of creatine kinase in frog heart. Comp Biochem Physiol 89B: 251–255, 1988
20. Wolff J, Kobel HR: Creatine kinase isozymes in pipid frogs: Their genetic bases, gene expressional differences, and evolutionary implications. J Exp Zool 234: 471–480, 1985
21. Garber AT, Winkfein RJ, Dixon GH: A novel creatine kinase cDNA whose transcript shows enhanced testicular expression. Biochim Biophys Acta 1087: 256–258, 1990
22. Barrantes FJ, Mieskes G, Wallimann T: Creatine kinase activity in the *Torpedo* electrocyte and in the non-receptor, peripheral v-proteins from acetylcholine receptor-rich membranes. Proc Natl Acad Sci USA 80: 5440–5444, 1983
23. Giraudat J, Devillers-Thiery A, Perriard J-C, Changeux JP: Complete nucleotide sequence of *Torpedo marmorata* mRNA coding for the 43,000-dalton v_2 protein: Muscle-specific creatine kinase. Proc Natl Acad Sci USA 81: 7313–7317, 1984
24. Witzemann V: Creatine phosphokinase: Isoenzymes in *Torpedo marmorata*. Eur J Biochem 150: 201–210, 1985
25. Gysin R, Yost B, Flanagan SD: Creatine kinase isoenzymes in *Torpedo californica*: Absence of the major brain isoenzyme from nicotinic acetylcholine receptor membranes. Biochemistry 25: 1271–1278, 1986
26. Tombes RM, Shapiro BM: Metabolite channelling: A phosphorylcreatine shuttle to mediate high energy phosphate transport between sperm mitochondrion and tail. Cell 41: 325–334, 1985
27. Tombes RM, Shapiro BM: Enzyme termini of a phosphocreatine shuttle: Purification and characterization of two creatine kinase isozymes from sea urchin sperm. J Biol Chem 262: 16011–16019, 1987
28. Tombes RM, Shapiro BM: Energy transport and cell polarity: Relationship of phosphagen kinase activity to sperm function. J Exp Zool 251: 82–90, 1989
29. Wyss M: Biochemical and physiological aspects of mitochondrial creatine kinase. Diss. No. 9777, ETH Zürich, Switzerland, 1992
30. Ratto A, Shapiro BN, Christen R: Phosphagen kinase evolution. Eur J Biochem 186: 195–203, 1989
31. Wothe DD, Charbonneau H, Shapiro BM: The phosphocreatine shuttle of sea urchin sperm: Flagellar creatine kinase resulted from a gene triplication. Proc Natl Acad Sci USA 87: 5203–5207, 1990
32. Dumas C, Camonis J: Cloning and sequence analysis of the cDNA for arginine kinase of lobster muscle. J Biol Chem 268: 21599–21605, 1993
33. Morrison JF: Arginine kinase and other invertebrate guanidino kinases. In: PD Boyer (ed.) The Enzymes. Academic Press, New York, 1973, pp 457–486
34. Stein LD, Harn DA, David JR: A cloned ATP:guanidino kinase in the trematode *Schistosoma mansoni* has a novel duplicated structure. J Biol Chem 265: 6582–6588, 1990
35. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395, 1984
36. Genetics Computer Group, GCG: Program Manual for the GCG Package. Version 7, April 1991. 575 Science Drive, Madison, Wisconsin, 53711, USA
37. Robert J, Wolff J, Jijakli H, Graf J-D, Karch F, Kobel HR: Developmental expression of the creatine kinase isozyme system of *Xenopus*: maternally derived CK-IV isoform persists far beyond the degradation of its maternal mRNA and into the zygotic expression period. Development 108: 507–514, 1990
38. Robert J, Barandun B, Kobel HR: A *Xenopus laevis* creatine kinase isozyme (CK-III/III) expressed preferentially in larval striated muscle: cDNA sequence, developmental expression and subcellular immunolocalization. Genet Res Camb 58: 35–40, 1991
39. Wirz T: Genetische Grundlagen der Heterogenität von zytosolischen und mitochondrialen Kreatin Kinasen. Diss. No. 9409, ETH Zürich, Switzerland, 1991
40. Mühlebach SM, Brändle U, Wirz T, Egli A, Perriard J-C: in preparation, 1994
41. Schwartz RM, Dayhoff MO: Matrices for detecting distant relationships. In: MO Dayhoff (ed.) Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Washington, DC, 1978, pp 353–358
42. Gribskov M, Burgess RR: Sigma factors from *E. coli, B. subtilis*, phage SPO1, and phage T4 are homologous proteins. Nucl Acids Res 14: 6745–6763, 1986
43. Lüttke A, Fuchs R: MacT: Apple Macintosh programs for constructing phylogenetic trees. Comput Applic Biosci 8: 591–594, 1992
44. Saitou N, Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425, 1987
45. Fitch WM, Margoliash E: Construction of phylogenetic trees. Science 155: 279–284, 1967
46. Jaynes JB, Chamberlain JS, Buskin JN, Johnson JE, Hauschka SD: Transcriptional regulation of the muscle creatine kinase gene and

regulated expression in transfected mouse myoblasts. Mol Cell Biol 6: 2855–2864, 1986
47. Mariman ECM, Broers CAM, Claesen CAA, Tesser GI, Wieringa B: Structure and expression of the human creatine kinase B gene. Genomics 1: 126–137, 1987
48. Mariman ECM, Schepens JTG, Wieringa B: Complete nucleotide sequence of the human creatine kinase B gene. Nucleic Acids Res 17: 6385, 1989
49. Daouk GH, Kaddurah-Daouk R, Putney S, Kingston R, Schimmel P: Isolation of a functional human gene for brain creatine kinase. J Biol Chem 263: 2442–2446, 1988
50. Benfield PA, Graf D, Korolkoff PN, Hobson G, Pearson ML: Isolation of four rat creatine kinase genes and identification of multiple potential promoter sequences within the rat brain creatine kinase promoter region. Gene 63: 227–243, 1988
51. Trask RV, Strauss AW, Billadello JJ: Developmental regulation and tissue-specific expression of the human muscle creatine kinase gene. J Biol Chem 263: 17142–17149, 1988
52. Van Deursen J, Schepens J, Peters W, Meijer D, Grosveld G, Hendriks W, Wieringa B: Genetic variability of the murine creatine kinase B gene locus and related pseudogenes in different inbred strains of mice. Genomics 12: 340–349, 1992
53. Haas RC, Korenfeld C, Zhang Z, Perryman B, Roman D, Strauss AW: Isolation and characterization of the gene and cDNA encoding human mitochondrial creatine kinase. J Biol Chem 264: 2890–2897, 1989
54. Klein SC, Haas RC, Perryman MB, Billadello JJ, Strauss AW: Regulatory element analysis and structural characterization of the human sarcomeric mitochondrial creatine kinase gene. J Biol Chem 266: 18058–18065, 1991
55. Steeghs K, Peters W, Wieringa B: Structure of the murine mitochondrial creatine kinase ubiquitous gene. Unpublished/extracted from GenEMBL sequence data libraries under accession number Z13968, 1992
56. Furukohri T, Fujimoto K, Susuki T: Glycocyamine kinase from the polychaete, *Neanthes diversicolor*. Isolation, purification and tryptic digestion of glycocyamine kinase. Mem Fac Sci Kochi Univ Ser D (Biol) 8: 85–94, 1987
57. Pickover CA, McKay DB, Engelman DM, Steitz TA: Substrate binding closes the cleft between the domains of yeast phosphoglycerate kinase. J Biol Chem 254: 11323–11329, 1979
58. Dumas C, Janin J: Conformational changes in arginine kinase upon ligand binding seen by small-angle X-ray scattering. FEBS Lett 153: 128–130, 1983
59. Morris GE, Cartwright AJ: Monoclonal antibody studies suggest a catalytic site at the interface between domains in creatine kinase. Biochim Biophys Acta 1039: 318–322, 1990
60. Morris GE, Jackson PJ: Identification by protein microsequencing of a proteinase-V8-cleavage site in a folding intermediate of chick muscle creatine kinase. Biochem J 280: 809–811, 1991
61. Wyss M, James P, Schlegel J, Wallimann T: Limited proteolysis of creatine kinase. Biochemistry 32: 10727–10735, 1993
62. Hossle JP, Schlegel J, Wegmann G, Wyss M, Böhlen P, Eppenberger HM, Wallimann T, Perriard J-C: Distinct tissue specific mitochondrial creatine kinases from chicken brain and striated muscle with a conserved CK framework. Biochem Biophys Res Commun 151: 408–416, 1988
63. Perryman MB, Kerner SA, Bohlmeyer TJ, Roberts R: Isolation and sequence analysis of a full-length cDNA for human M creatine kinase. Biochem Biophys Res Commun 140: 981–989, 1986
64. Furter R, Furter-Graves EM, Wallimann T: Creatine kinase: The reactive cysteine is required for synergism but is nonessential for catalysis. Biochemistry 32: 7022–7029, 1993
65. Buskin JN, Jaynes JB, Chamberlain JS, Hauschka SD: The mouse muscle creatine kinase cDNA and deduced amino acid sequences: comparison to evolutionarily related enzymes. J Mol Evol 22: 334–341, 1985
66. Taylor SS, Buechler JA, Yonemoto W: cAMP-dependent protein kinase: framework for a diverse family of regulatory enzymes. Annu Rev Biochem 59: 971–1005, 1990
67. DerTerrossian E, Desvages G, Pradel L-A, Kassab R, van Thoai N: Comparative structural studies of the active site of ATP:guanidine phosphotransferases. The essential cysteine tryptic peptide of lombricine kinase from *Lumbricus terrestris* muscle. Eur J Biochem 22: 585–592, 1971
68. Brevet A, Zeitoun Y, Pradel L-A: Comparative structural studies of the active site of ATP:guanidine phosphotransferases. The essential cysteine tryptic peptide of taurocyamine kinase from *Arenicola marina*. Biochim Biophys Acta 393: 1–9, 1975
69. Mahowald TA: Identification of an epsilon amino group of lysine and a sulfhydryl group of cysteine near the reactive cysteine residue in rabbit muscle creatine kinase. Fed Proc 28: 601, 1969
70. Babbitt PC, Kenyon GL, Kuntz ID, Cohen FE, Baxter JD, Benfield PA, Buskin JD, Gilbert WA, Hauschka SD, Hossle JP, Ordahl CP, Pearson ML, Perriard J-C, Pickering LA, Putney SD, West BL, Zivin RA: Comparisons of creatine kinase primary structures. J Protein Chem 5: 1–14, 1986
71. James P, Wyss M, Lutsenko S, Wallimann T, Carafoli E: ATP binding site of mitochondrial creatine kinase. FEBS Lett 273: 139–143, 1990
72. Kaldis P, Furter R, Wallimann T: The N-terminal heptapeptide of mitochondrial creatine kinase is important for octamerization. Biochemistry 33, 952–959, 1994
73. Gross M, Furter-Graves EM, Eppenberger HM, Wallimann T, Furter R: The tryptophan residues of mitochondrial creatine kinase: Trp-223, Trp-206, and Trp-264 in active site and quarternary structure formation, Prot Sci 3, in press, 1994
74. Yang JT, Chuen-Shang CW, Martinez HM: Calculation of protein conformation from circular dichroism. Meth Enzymol 130: 208–256, 1986
75. Oriol C, Landon M-F: Le dichroisme circulaire de diverses phosphagène phosphotransférases. Biochim Biophys Acta 214: 455–462, 1970
76. Manavalan P, Johnson WC: Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Anal Biochem 167: 76–85, 1987
77. Vasak M, Nagayama K, Wüthrich K, Mertens M, Kägi JHR: Creatine kinase. Nuclear magnetic resonance and fluorescence evidence for interaction of ADP with aromatic residue(s). Biochemistry 18: 5050–5055, 1979
78. Kägi JHR, Li T-K, Vallee BL: Extrinsic Cotton effects in complexes of creatine kinase with adenine coenzymes. Biochemistry 10: 1007–1015, 1971
79. Zhou H-M, Tsou C-L: An essential tryptophan residue for rabbit muscle creatine kinase. Biochim Biophys Acta 830: 59–63, 1985
80. Fattoum A, Kassab R, Pradel L-A: The tyrosyl residues in creatine kinase. Modification by iodine. Biochim Biophys Acta 405: 324–339, 1975
81. Villarreal-Levy G, Ma TS, Kerner SA, Roberts R, Perryman MB: Human creatine kinase: Isolation and sequence analysis of cDNA

clones for the B subunit, development of subunit specific probes and determination of gene copy number. Biochem Biophys Res Commun 144: 1116–1127, 1987

82. Billadello JJ, Kelly DP, Roman DG, Strauss AW: The complete nucleotide sequence of canine brain B creatine kinase mRNA: Homology in the coding and 3′ noncoding regions among species. Biochem Biophys Res Commun 138: 392–398, 1986
83. Pickering L, Pang H, Biemann K, Munro H, Schimmel P: Two tissue-specific isozymes of creatine kinase have closely matched amino acid sequences. Proc Natl Acad Sci USA 82: 2310–2314, 1985
84. Benfield PA, Henderson L, Pearson ML: Expression of a rat brain creatine kinase-beta-galactosidase fusion protein in *Escherichia coli* and derivation of the complete amino acid sequence of rat brain creatine kinase. Gene 39: 263–267, 1985
85. Hossle JP, Rosenberg UB, Schaefer B, Eppenberger HM, Wallimann T, Perriard JC: The primary structure of chicken B-creatine kinase and evidence for heterogeneity of its mRNA. Nucl Acids Res 14: 1449–1463, 1986
86. Kwiatkowski RW, Ehrismann R, Schweinfest CW, Dottin RP: Accumulation of creatine kinase mRNA during myogenesis: Molecular cloning of a B-creatine kinase cDNA. Dev Biol 112: 84–88, 1985
87. Roman D, Billadello J, Gordon J, Grace A, Sobel B, Strauss A: Complete nucleotide sequence of dog heart creatine kinase mRNA: Conservation of amino acid sequence within and among species. Proc Natl Acad Sci USA 82: 8394–8398, 1985
88. Putney S, Herlihy W, Royal N, Pang H, Aposhian HV, Pickering L, Belagaje R, Biemann K, Page D, Kuby S, Schimmel P: Rabbit muscle creatine phosphokinase: cDNA cloning, primary structure, and detection of human homologues. J Biol Chem 259: 14317–14320, 1984
89. Benfield PA, Zivin RA, Miller LS, Sowder R, Smythers GW, Henderson L, Oroszlan S, Pearson ML: Isolation and sequence analysis of cDNA clones coding for rat skeletal muscle creatine kinase. J Biol Chem 259: 14979–14984, 1984
90. Kwiatkowski RW, Schweinfest CW, Dottin RP: Molecular cloning and the complete nucleotide sequence of the creatine kinase-M cDNA from chicken. Nucl Acids Res 12: 6925–6934, 1984
91. Ordahl CP, Evans GL, Cooper TA, Kunz G, Perriard J-C: Complete cDNA-derived amino acid sequence of chick muscle creatine kinase. J Biol Chem 259: 15224–15227, 1984
92. West BL, Babbitt PC, Mendez B, Baxter JD: Creatine kinase protein sequence encoded by a cDNA made from *Torpedo californica* electric organ mRNA. Proc Natl Acad Sci USA 81: 7007–7011, 1984
93. Payne RM, Haas RC, Strauss AW: Structural characterization and tissue-specific expression of the mRNAs encoding isoenzymes from two rat mitochondrial creatine kinase genes. Biochim Biophys Acta 1089: 352–361, 1991
94. Haas RC, Strauss AW: Separate nuclear genes encode sarcomere-specific and ubiquitous human mitochondrial creatine kinase isoenzymes. J Biol Chem 265: 6921–6927, 1990
95. McCombie WR, Adams MD, Kelley JM, FitzGerald MG, Utterback TR, Khan M, Dubnick M, Kerlavage AG, Venter J, Fields C: *Caenohabditis elegans* expressed sequence tags reveal gene families and potential disease gene homologues. Unpublished/extracted from GenEMBL data library under accession number M79599, 1992
96. Waterston R, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N, Metzstein M, Hawkins T, Wilson R, Berks M, Du Z, Thomas K, Thierry-Mieg J, Sulston J: A survey of expressed genes in *Caenorhabditis elegans*. Nature Genet 1: 114–123, 1992