

Multiple, Closely Spaced Alternative 5' Exons in the *DmCKII β* Gene of *Drosophila melanogaster*

Ashok P. Bidwai,^{*1} Amit Saxena,^{*2} Wenfan Zhao,[†] Richard O. McCann,^{†3} and Claiborne V. C. Glover[†]

^{*}Department of Biology, West Virginia University, Morgantown, West Virginia 26506-6057; and [†]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602-7229

Received May 15, 2000

***Drosophila melanogaster* casein kinase II (CKII) is composed of catalytic α and regulatory β subunits. Using the two-hybrid system, we have isolated a number of cDNAs that are related to a previously published cDNA encoding the β subunit, but exhibit divergent 5' sequences. To determine the source of this sequence variation, we have isolated the gene encoding the β subunit of CKII. The β gene contains five exons encompassing the complete open reading frame, as well as five alternative exons in the 5' untranslated region (UTR). Only one 5' UTR exon is contained in each cDNA, implying five distinct classes of transcript. In addition, the β gene contains at least two poly(A) addition signals which generate additional complexity at the 3' end. The complex pattern of transcription may serve a role in the spatial and/or temporal expression of the β subunit since, with one exception, all transcripts encode the full-length β polypeptide. Phylogenetic comparison of the β genes of *Drosophila*, *C. elegans*, and mammals reveals three invariant introns as well as evidence of recent intron gain/loss. © 2000**

Academic Press

Casein kinase II (CKII) is a highly conserved protein kinase that appears to be ubiquitous in eukaryotes [1, 2]. The enzyme is classified as an acidic-directed Ser/Thr protein kinase [3], but at least one substrate that is phosphorylated on a tyrosine residue *in vivo* has been identified [4]. CKII phosphorylates a broad spectrum of cytoplasmic and nuclear proteins involved in diverse processes, including DNA replication, tran-

scription, translation, cell cycle regulation, and signal transduction [1], suggesting a pleiotropic role. CKII is essential for viability in *Saccharomyces cerevisiae* [5] and probably also in *Schizosaccharomyces pombe* [6] and is capable of functioning as an oncogene in mammals [7]. The enzyme is constitutively active as isolated and may be messenger-independent *in vivo* [2].

CKII of most species is composed of catalytic α and regulatory β subunits that combine to form an $\alpha_2\beta_2$ holoenzyme. The free α subunit is monomeric and active *in vitro*, and the crystal structure of the catalytic subunit of maize CKII has revealed the likely origin of this constitutive activity [8]. The function of the β subunit is complex [1, 2]. On the one hand, the β subunit stabilizes the α subunit against denaturation and stimulates its activity against most substrates *in vitro*; on the other, it exerts a down-regulation of activity against specific substrates, notably calmodulin. The β subunit is also subject to autophosphorylation (catalyzed by α) and mediates activation of the enzyme by polybasic compounds such as spermine and polylysine. The crystal structure of the human β subunit has provided additional insight into the structure and function of the regulatory subunit [9]. The protein is composed of two well defined domains, an α -helical N-terminal domain (domain I, residues 5–104) and a C-terminal domain (domain II, residues 105–161). The latter is composed primarily of β -sheet and contains an invariant Cys-rich motif (CPX₃C-X₂₂-CPXC) that forms a novel zinc-finger. This domain mediates formation of a β dimer that presumably forms the core of the tetrameric holoenzyme. Domain I contains the autophosphorylation site (residues 1–6) as well as the internal acidic region (residues 55–64) implicated in the binding of basic compounds. Both are disordered in the structure but are in close proximity, consistent with cooperation between these two regions in modulation of enzyme activity.

The β gene family of *Drosophila melanogaster* is unusually complex, consisting of two bona fide β sub-

¹ To whom correspondence should be address at Department of Biology, Brooks Hall, West Virginia University, Morgantown, WV 26506-6057. Fax: (304)-293-6363. E-mail: abidwai@wvu.edu.

² Current address: Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73190.

³ Current address: Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205-2185.

unit genes, *DmCKII β* [10] and *DmCKII β'* [11], as well as a large family of β -related genes, *Stellate* (*Ste*) [12], *Suppressor-of-Stellate* [*Su(Ste)*] [13], and *Su(Ste)-like* (*SSL*) [14]. *DmCKII β* and β' appear to be broadly expressed, but expression of the β -related genes is confined to the testis. The β -related genes are believed to have arisen from the β gene [14]. *SSL* contains all of the key motifs required for interaction with and regulation of the *CKII α* subunit, and the *Stellate* protein has been shown to interact weakly with the α subunit *in vitro* [15]. It is thus possible that these β -related polypeptides play a role in regulation of *CKII* activity in testis. The physiological role of *CKII* in *Drosophila* is not well defined, but several well characterized substrates are known, including topoisomerase II [16], dishevelled, a regulator of the frizzled signal transduction pathway [17], and the homeobox protein *Antennapedia* [18].

In the course of carrying out a two-hybrid screen for proteins that interact with *DmCKII α* , we have isolated multiple cDNAs encoding *DmCKII β* . Although homologous to a previously described β cDNA (*Dm98*, see Saxena *et al.*, 1987), these cDNAs exhibit divergent 5' sequences. To determine the basis of this sequence heterogeneity, we have isolated and sequenced the β gene. This analysis demonstrates that the divergent sequences of the available β cDNAs are derived from five closely spaced, alternative 5' exons. All five classes of transcript encode the identical, full-length β subunit, so the alternative exons may reflect developmental and/or tissue-specific regulation of β gene expression. We also compare the intron-exon organization of the *Drosophila* β subunit gene with that of human [19], mouse [20], and *C. elegans* [21].

MATERIALS AND METHODS

Two-hybrid screen for cDNAs encoding DmCKII β . The *DmCKII α* open reading frame was amplified by the polymerase chain reaction (PCR) using 5' primer 5'-GGGAATTCATGACACTTCCTAGTGCGGC-3' and 3' primer 5'-GGGGATCCTTATTGCTGATTATTGGGAT-3'. The PCR product was digested with *EcoRI* and *BamHI* (underlined sites) and subcloned into the two-hybrid bait plasmid, pGBT9 (gift of Dr. S. Fields, now available commercially from Clontech). The resulting plasmid expresses *DmCKII α* as a C-terminal fusion with the DNA-binding domain (amino acids 1–147) of Gal4. The insert was completely sequenced on an Applied Biosystems DNA sequencer 373A using the Prism Dye Terminator Cycle sequencing kit and custom primers. This plasmid was transformed into yeast strain HF7C [22] by electroporation. The resulting strain was used to screen a *Drosophila* 3- to 12-h embryo cDNA library (gift of Dr. S. Elledge, Baylor College of Medicine) in plasmid pACT, where proteins are expressed as fusions with the activation domain

(amino acids 768–881) of Gal4. Positive transformants were identified by plating on minimal medium lacking leucine, tryptophan, and histidine and subsequently counterscreened for the expression of *lacZ*, as described in the Matchmaker Two-Hybrid System protocol (Clontech). The library plasmid, pACT containing the yeast nutritional marker *LEU2*, was selectively recovered from His⁺ and LacZ⁺ transformants via complementation of the leucine auxotrophy of *Escherichia coli* HB101. The 5' and 3' end of each cDNA insert was sequenced on an Applied Biosystems DNA sequencer 373A, using vector-specific primers, 5'-ATACCACTACAATGGATGATG-3' and 5'-ACAGTTGAAGTGAAGTTCGCG-3', respectively.

Isolation of cDNAs encoding DmCKII β from Uni-ZAP XR. Two of the two-hybrid clones isolated using *DmCKII α* as bait encode a novel isoform of the β subunit, β' , encoded by a distinct gene [11]. In an effort to isolate additional β' cDNAs, we screened a *Drosophila* 3- to 18-h embryo cDNA library (Catalog No. 937602) in Uni-ZAP XR (Stratagene). Approximately 200,000 plaques were screened at low stringency using the β' cDNA *DmA15* as a probe. Positive clones were converted to plasmid (pBluescript SK⁻) using the ExAssist helper phage (Stratagene), essentially as described by the manufacturer. The clones were subsequently grouped into classes by restriction analysis. Two clones encoding the β subunit were isolated in this screen by virtue of probe crossreactivity. The 5' and 3' end of both cDNAs was sequenced as described above, using vector-specific primers, 5'-AATTAACCCTCACTAAAGGG-3' and 5'-GTAAAACGACGGCCAGT-3', respectively.

Isolation of genomic clones encoding DmCKII β . To isolate genomic clones encoding *DmCKII β* , cDNA clone *Dm98* [10] was labeled by random hexamer priming (Boehringer-Mannheim Biochemicals) and used to screen a *D. melanogaster* genomic library in lambda gt11 (gift of Dr. J. Wang, Harvard). One clone, *DmBG3*, was subcloned into M13mp18, subjected to the nested deletion procedure of Dale *et al.* [23], and completely sequenced on both strands by dideoxy chain termination sequencing. Although *DmBG3* encoded the complete sequence of *Dm98*, it lacked upstream sequences present on additional cDNAs isolated in the screens described above. To obtain additional 5' sequence, clone *Dm98* was used to screen a *D. melanogaster* cosmid genomic library (gift of Drs. J. Tamkun and M. Scott, University of Colorado, Boulder). Positive clones were mapped using the *SfiI*-linker mapping strategy (Promega) following the manufacturer's instructions. A 6 kb *BamHI* fragment (*DmBG3-4*) containing an additional 2 kb of 5' and 0.1 kb of 3' sequence was subcloned into pBSII-KS⁺ (Stratagene), and the missing sequence was obtained on both strands using an Ap-

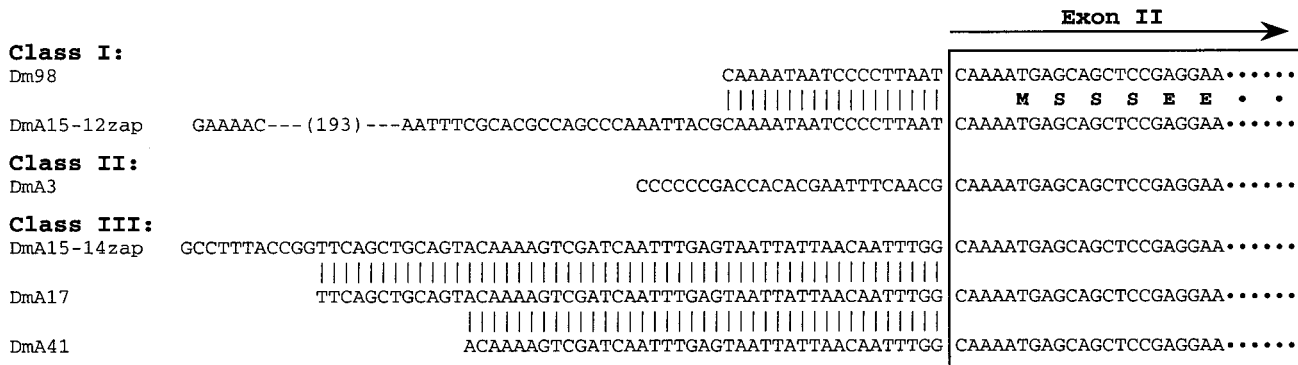


FIG. 1. Alignment of cDNAs encoding DmCKII β . The 5' sequences of cDNAs encoding the β subunit of DmCKII isolated either in a two-hybrid screen (DmA3, DmA17, and DmA41) or from a Uni-ZAP XR embryo cDNA library (DmA15-12zap and DmA15-14zap) have been aligned with the sequence of the previously published DmCKII β cDNA, Dm98 [10]. The 5' UTRs sequences of all the cDNAs are shown in full except for DmA15-12zap which contains 248 nucleotides upstream of the 5' end of Dm98. The boxed area corresponds to exon II, which is common to all of the cDNAs. The deduced amino acid sequence of Dm98 is indicated by single letter code. The 5' sequences of DmA3, DmA15-12zap, and DmA15-14zap have been deposited in GenBank with Accession Nos. AF236850, AF236851, and AF236852, respectively.

plied Biosystems DNA sequencer 373A and custom primers.

In situ hybridization to polytene chromosomes. *In situ* hybridization to polytene chromosomes was carried out essentially as described by Langer-Safer *et al.* [24]. Briefly, either cDNA Dm98 or a cosmid clone containing the β gene was labeled by nick translation with biotin-dUTP (Bio-Rad Laboratories) and hybridized at high stringency to salivary gland squashes prepared from third instar larvae of *D. melanogaster*. Bound probe was visualized with a streptavidin-alkaline phosphatase conjugate (Bio-Rad Laboratories). Slides were examined by phase-contrast microscopy and photographed at 400 \times .

RESULTS AND DISCUSSION

Isolation of multiple cDNAs encoding DmCKII β . This laboratory previously reported the isolation of two cDNAs (Dm98 and Dm107) encoding the β subunit of DmCKII [10]. We have recently conducted a genetic screen using the yeast two-hybrid system to identify proteins which interact with the α subunit of DmCKII [11]. Several clones isolated in this screen encode the β subunit of DmCKII, and two encode a distinct isoform of the β subunit, which we call β' [11]. The isolation of clones encoding β is consistent with the known tight association of the α and β subunits within the $\alpha_2\beta_2$ holoenzyme [25]. Three of the β cDNAs (DmA3, DmA17, and DmA41) contained a full-length open reading frame as well as significant amounts of 5' and 3' UTR. In an effort to isolate additional cDNAs encoding the novel β' subunit, we employed the DmA15 β' cDNA as a probe to screen a *Drosophila* embryo cDNA library in Uni-ZAP XR. No new β' cDNAs were obtained in this low stringency hybridization, but two additional clones encoding the full length β subunit

(DmA15-12zap and DmA15-14zap) were isolated, presumably because of crossreactivity with the β' probe.

Sequence alignment of the five new β cDNAs with Dm98 revealed that only one of the sequences (DmA15-12zap) could be fully aligned with Dm98 in the 5' UTR. Based on the 5' UTR sequences, the six cDNAs could be classified into three groups (Fig. 1): class I (Dm98 and DmA15-12zap), class II (DmA3), and class III (DmA15-14zap, DmA17, and DmA41). We reasoned that the distinct 5' sequences were unlikely to be artifacts introduced during cDNA synthesis since two of the groups were represented by multiple cDNAs from two different libraries. The presence of multiple mRNAs originating from the β gene of *D. melanogaster* is intriguing since the human [19], mouse [20], and nematode [21] β genes apparently do not exhibit such behavior.

Structure and expression of the gene encoding DmCKII β . To ascertain the source of the divergent 5' sequences, we used clone Dm98 to isolate genomic clones encoding DmCKII β . The complete sequence of a 6 kb *Bam*HI fragment (DmBG3-4) was obtained (Fig. 2). Comparison of the available cDNA sequences with the genomic sequence revealed that the β gene is interrupted by 6 introns, five in the protein coding region and one in the 5' UTR. Moreover, the 5' UTR sequences of all three classes of cDNA identified above could be unambiguously identified in the genomic sequence, indicating that the β gene contains at least three alternative 5' exons, Ia (class III), Id (class II), and Ie (class I). Subsequent searches of the *Drosophila* EST database in GenBank identified three other informative β transcripts, one belonging to class III (AI57689) and two defining two additional upstream exons, Ib (AI109297) and Ic (AI109634). Only one of the five upstream exons is present in each of the cDNAs we

1	GGATCCGCTG	CGCGAGAAGC	TCGTGGGCAA	CATCACCTGG	ATGATACCCA	AATTCCAAGA	GCCAGGACAT	TGGCAGCACA	CGGAGGACAT	ATgtgagtga										
101	tggaaatggt	tctacaatca	aattcaaaac	gaaaatctat	tcaaaatgta	gtttgtgatac	gtttggaagta	gttttgtaaat	tactatgcca	aagagtaatt										
201	cgatctgccc	gggtttaata	tatgttttaa	gcagtatatt	cgtttctata	tgtacctctct	tcagatAGTAG	GGCATCTACA	ACACATACGT	TCTGGACACG										
301	GACTACGACA	CTGTGGGCT	GGTGATCGAC	TGCGCCGAGA	AGAAGAAGCA	ACCACGPTAC	CTGTCCGCC	TGCTTTTGTC	CGCGAAAACA	TCGATGGCCG										
401	ACAACGAGAT	CAGCTTTCTG	CGCGGCAAGT	TGCCGACGGA	TATCGACACA	TCCTTTATGT	TTAACATCGG	GCAGGAATCG	TGCGACAATC	TAATGGAATC										
501	CAGTCGCGAC	GATCCACTGG	CCTATGTGT	GAATGGACGC	CAGAGCGACA	AGGAAATCTT	CAAAATTAATC	AATAAGCCTG	ACGGCAGACG	CGCGCTTAGA										
601	AGCGAAGCGA	GAAGCTATGA	GCAAAGCGAG	AAGCTGTGAG	CAAAGCGAGT	CGATGGCAGA	GTAAGAGCTA	AGCGAGAAAC	GAGCTACAT	GCTCTCTTTT										
701	AGTATTATTA	TAACTTCTGA	TGGATTCCGA	CAGGAAATTT	TTTGTATCAA	CGCAACTAT	TAACCTTCTG	TTTAAACTGT	gtgtcaacct	cttaactgaa										
801	atagaaatga	aaaaggcgaa	ttcaataatt	tatataaaag	gttatttaca	gaatggaaca	agacaaaaca	aaataacgtg	gtttttgaaa	ataaacggta										
901	atatagcgca	gaatataaca	aaacaactct	acgcctctgag	cgagttggaag	cgataaagcc	ttgctctctg	tttttctgct	aagctcgcaa	ctcgctctgc										
1001	tctcactctg	acatcttttcg	agcgattgcy	cttaaaatgc	atgcagtggg	ttatctctct	gttcttctg	tattagtcta	atcaatcaac	taagagaaaa										
1101	tattatacga	aaaccggatt	tattttgtca	ataatacaaa	cttaaaatgc	catttcaaac	tttcaacata	tttcaaac	catttcaaac	acaactcaac										
1201	gaaataatac	cttaaaagta	ttgtaagaaa	atattgctag	ctgcaatcgg	taaattattt	gacaaatggc	aaaacagcca	actaaatgct	ggtaataaac										
1301	aatatttttaa	tatttggctt	gatttgccta	tttaacaat	aatctcgcgt	tttgtttggt	taattgcttt	tccgaaccag	gtttgcgaaa	cgcgcaatcc										
1401	cgattgacgc	atacatcgct	tagecatgac	acggcgccac	atcgattggc	aatcgctctg	ggacagcgca	ggacagcggt	aatcgatgat	cgagattgca										
1501	tacctgatat	catCGTCTC	GCTCAATCA	TTGCTTTAC	CGTGTACCT	CAGTCCATC	AGTCGATCAA	TTTGTAGT	TATTACAAAT	TTGGTgagt	Ia									
1601	gctggccact	ggaagatagc	cgacgcgga	aaaatacaga	atagcaagtt	ggtgtgcgtt	ggattgctct	catcggtgta	aaggcgtgag	gtgtgcaag										
1701	tccctatgtg	cgcttctcgg	gtgtcgtggt	tgcttctgtg	cggtgtgtgt	ccccctgtat	atgtgtctct	ttttttctca	atttccaaca	gtgtgtgctt										
1801	<u>CTTCTGTCTG</u>	<u>CTCGGCACAT</u>	<u>TTTgtatgag</u>	cgcttgcTAA	<u>TTGTCACAAT</u>	<u>ACGAgataaa</u>	aaatttattt	ttatgCCCC	<u>CGACCACAGC</u>	<u>AATTTCAACG</u>	Ib, c, d									
1901	gtcggatcac	cgaaaactag	taacattgta	caagcaacgc	gaattaaagc	GAAACCCGAA	AACAAAACAC	CCAAAAGGCC	AATTCACCA	AAAABAAAGA	Ie									
2001	<u>GGAGGAAAGA</u>	<u>AGAAAAAAGC</u>	<u>GGAAAAAGAG</u>	<u>AGTGTGGTTT</u>	<u>GTGCGAGCAA</u>	<u>CAAGAAGCGG</u>	<u>GAACCAAAAG</u>	<u>GAAACGTGAA</u>	<u>GAAAGCAGGG</u>	<u>GGGTGGGGT</u>										
2101	<u>AGACAGAGAA</u>	<u>GTGCCAAGAT</u>	<u>TCCGTCCAA</u>	<u>GATATTTTAA</u>	<u>CAATAAATTT</u>	<u>CGCAGCCGAG</u>	<u>CCCAAATTA</u>	<u>GCAAAATTAAT</u>	<u>CCCCTTAATg</u>	tatgtaaatg										
2201	tgctgtgacg	tttgtcccag	gtgtgtgtgc	atgactgctt	gtgtgtgtga	gagtgctggt	tgcatgtgtg	gtagctgctg	gcaggataat	ttgtcagcaa										
2301	aaaaaaaaacg	atgaaaatat	gaattagcaa	ctccattctg	ttatctttgc	agCAAAATGA	GCAGCTCCGA	GCAGCTCTCC	GGAAGTCTCC	TGGTCTGTGG	II									
						M	S	S	S	E	E	V	S	W	V	T	W	F	C	G
2401	ACTTCGTGGC	AATGAGTTCT	TCTGCGAGgt	gagatatacat	catcatatca	gcatacatcat	atcagatca	acgacgaatt	accocgcttc	ttcgtgtaa										
	L R G	N E F	F C E																	
2501	tcggtatct	aagcgcgcgc	tataaagtca	acaattgcaa	agtgcagccc	tcaatctctg	cttttgcgag	cgctgtgtgt	gtgtgtgtgt	cgctgtgtgc										
2601	gcaaggtggg	gcaataccga	gccccactgt	acacacatgt	ccagataaca	cacaatatag	tgatataatg	acagtataat	ccagtagaca	tcctaattgc										
2701	atacgtcttt	gattccagGT	GGATGAGGAC	TACATACAGG	ATAAAITCAA	TTTAACTGGT	TTAAATGAGC	AGGTACCCAA	CTATCGGCAA	CGGTGGGACA	III									
		V	D E D	Y I Q	D K F N	L T G	L N E	Q V P N	Y R Q	A L D										
2801	TGATCTTGGG	CTTGGAACCG	Ggtaagttta	gaatgggctg	ttaccatcat	ccaaaattgt	aatgtatata	gctaagtcgc	cttaccatcc	acagAGGACG	IV									
	M I L D	L E P								E D										
2901	AGCTCGAGGA	CAATCCACTG	CAGTCCGACA	TGACCAGACA	GGCCGCGAG	ATGCTCTACG	GCCTCATACA	CGCCAGATAT	ATACTAACAA	ATCGCCGCAT										
	E L E D	N P L	Q S D	M T E Q	A A E	M L Y	G L I H	A R Y	I L T	N R G I										
3001	CGCTCAAAATG	ATCGAATCG	ATCAAACTGG	CGATTTCCGA	CAITGTCAC	GCTCTACTG	TGAAAGTCAG	CCCATCTGCG	CATTGGgtgc	gtaaccagat										
	A Q M	I E K	Y Q T G	D F G	H C P	R V Y C	E S Q	P M L	P L											
3101	ccttgatcaa	tcgctcttac	ctatagctaa	caaacatagt	tgcgaatctt	tctgcgcta	attgattgta	atctaaaaaa	aaaaaaaaaa	aaaaaacat										
3201	ctacgaggaa	ctcctgtgctt	ttctgctgat	taattacata	ccagagaaat	atcaacgcga	tatttggcct	aaatttcaat	tgagtttctt	gaatagcata										
3301	tcattggggc	tttccctatga	aatctttacac	caagaaaaaa	accagcatt	cttaatagaa	atcttaccct	aattgtctg	cttggctctg	ccagtagtaa										
3401	accttgcgaa	cgcttccgca	tctccaataa	attcttaccg	aaacttccct	aaacgttgtt	aacgatcaaa	taaagcaatg	ctgaatcaaa	tgcatgcaca										
3501	aatgcaagat	agggtttcac	taacaagagt	ccctcaatct	ctgtgaaaat	taactagggc	aaagtggcta	gacgagcctt	gtaaacattt	ctgctatkaa										
3601	tgataaatga	tcgatgatga	tatcatgttg	agttaaatca	tgttaaatga	gtttgtacca	gatttcaaca	atgaattctg	actttttcta	gttgctagtt										
3701	gctctgcccag	tggtctaatc	ctttgattat	tatcctcagG	CTTGTGGAC	GCCTCCCGCG	AGGCAATGAT	TAAGACCTAT	TGCCCCAAGT	CGATTGACGT	V									
				G	L S D	I P G	E A M V	K T Y	C P K	C I D V										
3801	GTACACACCA	AAATCGTCTG	GTCACCACCA	TACCGATGGC	GCCTAATTCG	GCACITGATT	TCCACACATG	CTCTTCATGG	TGCATCCCGA	ATATCTGTC										
	Y T P	K S S	R H H H	T D G	A Y F	G T G F	P H M	L F M	V H P E	Y R P										
3901	AAGCGTCTTA	CTAATCAGTT	TGTTCCAAGG	taagatttat	accattact	gtatatgcaac	attgtaata	tgctcttttg	ttttatgtgc	ccgatacgac										
	K R P	T N Q F	V P R																	
4001	gtcaacttga	tcatacata	catagGCTAT	ATGGATTTAA	AATACACAGC	TTAGCTTATC	AAATTACAGT	GCAGGCAGCA	GCCAATTTTA	AAATGCCACT	VI									
			L	Y G F K	I H S	L A Y	Q I Q L	Q A A	A N F	K M P L										
4101	ACGAGCGgta	aattttaaac	ttaataacta	gctattagta	tcctttccac	caattacccc	actctctcat	cttttttttt	gttgtatatt	atlttggaa										
	R A																			
4201	ctaaaaatct	atatattgac	taaaaaaaaa	aaacacaacg	ccaccaaggg	atcattccaac	taactaagcat	ataaatatat	tatactcttt	ttatttggtt										
4301	tctctcccca	cacatacaaa	accacaacac	caaccaccac	caccaccaca	aaaaaaaaac	aaaccaaatc	aaatcaaaa	tccataaaat	ccgatcgaca										
4401	accgaacct	gaatatgatt	atgaatacaa	tcattgtctct	gctctgtctga	ataaaaacaa	acataaaacc	aaaatatcga	tatatattcg	actgttccaa										
4501	tcctgctgatt	cttaaaattg	ggtgtgAAA	AACTAATAAA	ATAAATACAC	CACCAACAAC	AACAAATACT	ACAACACAAA	CACATACGCA	CGAATACAAC	VII									
			K	N *																
4601	AACAAACAAT	CCATTTAACT	GCATGTAAAC	GAACAACAAA	CATAAATGTA	AATGATAGTT	TACAACGGGA	AGACCAGTAG	ACAGCAACAC	ACAACAACAA										
4701	CAACACACGC	CACACTGAGA	GACGGAGAAC	CAATTGTATT	GAGGTTTTTA	AACGCTGGGC	GCAAAGTTTC	ATTTATATAA	ATATAAAAAC	TAAAAAAAC										
4801	AACACAAAA	ACAAAAACCA	ACTGCGTTTG	TATTTCAAAA	TGAAAAGAAA	CAGAAACAGA	AGAAATGTGA	TAAAAACAAA	TAGTTAAAAG	GACACACAGC										
4901	AACACGCATA	TTTATTTAGA	GCCCATAAAC	ATTCAAAACA	TGAGAAACTT	CATCGCATGT	TGATTTTTAT	TTAAAACACT	GCAACTTCTT	ATTAATCTAA										
5001	ACTTCATTTCA	GAAACATGTA	TTTATGTTAA	ATTTGAATGT	TACCGCTGAG	CAGATAATCC	TTTCGATTTGA	AGTTTAAAGT	CAAGGACTAC	AAGCATAACT										
5101	CTCAAAATACA	CACACACCCA	CAAAACGCTA	CACACACACG	CTPATATATA	TATATATATA	TTTCTTCAGT	CAATTAACCA	ATPATATATG	TGTTTTTAA										
5201	TGTTCTGTTTC	CTCACATTTT	ATGTATGTAT	GTATGTTTTG	TGTAACATTT	TAAGACACAG	GCAGTGAAC	ACTTGACCAA	AAAAAGATAA	AAAGAAGGAA										
5301	<u>ATAAAAGTAA</u>	<u>CAAAAAGGAT</u>	<u>ACTCGAATTA</u>	<u>GACAGAAAA</u>	<u>GAGAGTGTGT</u>	<u>GAGAGGAGAGA</u>	<u>GAGAGAGTTA</u>	<u>GAAGTAGAGC</u>	<u>GATAGTAAAG</u>	<u>AAGAAACTAA</u>										
5401	TTTAAAGTAG	AGAAAAACAC	GATATAAGCA	ACATAGTACT	TCCAAATTAAT	AAAAGAACACA	AATAACAGTC	AGCATTTATG	TAAACGACAA	CGCAAAATAA										
5501	AACCAATGAT	AAATTCCTATA	GCCGGAAGCG	CCAGTCAAC	TCCGGAAGAT	TCCCAAAATG	AGCAAGTCCC	TAAAGTATCC	TAATATGATA	AAGAAATGAT										
5601	CGAAACCCCA	ACTCCCAAAG	AAGTAGACGA	AAATTAATGA	AACCAACAAA	AACATACAGT	ACAGTCAATC	CCATCCAACC	GAACACCCAC	TCACTCACCC										
5701	ACTGATTTAC	ACACACACAT	CGACGCGCTC	ACACACATAG	ATACCCGATG	ACCCCCCTAA	TACCAACACC	CACCCATACA	CAGAAATGAT	GCAAAATGAC										
5801	GGAAGCAAAA	GCAACCACTT	GGAACATAAG	GATGTGGATC	GAAGAAGAGG	ATGAAGATTTG	GCAGCAGGAC	ACGAGCAACG	AACCGAACGA	ATCAAGCCCC										
5901	AAAATTTCAA	CAGACACACA	CACAATGAAA	CAAGTACACA	TACACACCCA	CACACACACT	CATGAGCATG	CACGGACAAG	CGGAAGGCAA	AACAATGGCA										
6001	GAAGGATCC																			

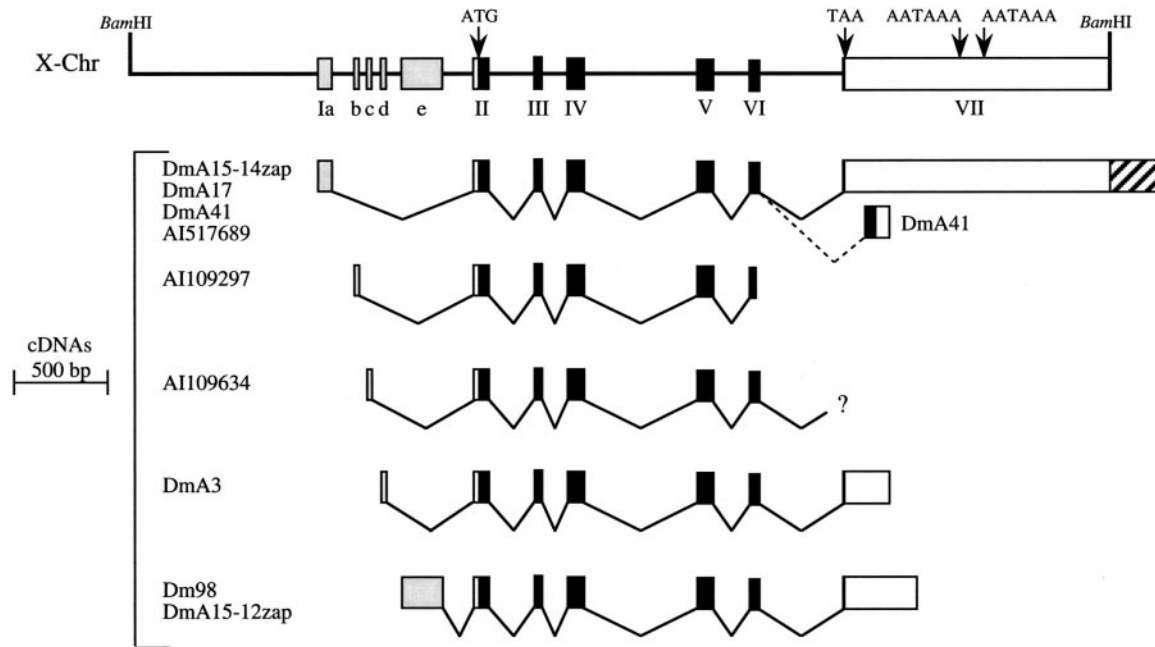


FIG. 3. Alternative transcripts of the DmCKII β gene. The structure of the 6 kb *Bam*HI fragment containing the β gene is shown at the top. Boxes depict the exons contained within the β gene, with the five alternative 5' exons shaded gray and the open reading frame black. The start and stop codons and two poly(A) addition signals, AATAAAA, are indicated. The five classes of cDNA and the clones defining them are indicated below the gene. The composite extent of each class is shown. The cross-hatched box indicates 200 bp of sequence in clone DmA17 that extend beyond the 3' genomic *Bam*HI site. Clone DmA41 contains an alternative splice (dashed line) to an AG dinucleotide located at positions 4657–4658 (Fig. 2). Clone AI109634 contains 11 bp of sequence beyond the end of exon VI that do not match available downstream genomic sequence.

have examined (Fig. 3). None of the alternative 5' UTRs contains an upstream ATG codon.

The occurrence of five distinct 5' UTRs in transcripts of the β gene could be explained either by transcription from multiple promoters or by alternative splicing, or by some combination of the two. Alternative splicing of a single primary transcript provides perhaps the simplest model. One prediction of this model is that transcripts produced by splicing from exons Ib, c, d, or e to exon II will contain the upstream exon(s) and intervening region(s). As just noted, this was not reflected in any of the cDNAs we have examined, though of course it is possible that none of these cDNAs is full-length. In addition, because several of the spacer regions contain ATG codons, all transcripts except those spliced from exon Ia would contain multiple upstream ATGs that could interfere with translation of the DmCKII β coding region and/or generate N-terminal fusion proteins. An

alternative model invokes five distinct promoters, accompanied by obligatory splicing from the donor closest to the 5' end of each transcript. Although several genes in *Drosophila* are known to employ this strategy in the generation of 5' heterogeneity (e.g., *Act5C*, *Adh*, *hb*; [26]), the alternative 5' exons of DmCKII β are unusually close together (maximum separation of 207, 14, 22, and 51 bp). Moreover, only two consensus TATA elements are present in the region (at positions 831–838 and 1856–1863), and the first of these lies well upstream of the sequenced cDNAs. However, TATA-less promoters are common in *Drosophila* [27]. Additional studies will be required to distinguish the relative contributions of alternative promoter utilization versus alternative splicing to the observed expression.

With one exception, all of the cDNAs we have sequenced exhibit an identical pattern of splicing of exons II to VII (Fig. 3). The identical splicing of the

FIG. 2. Sequence of the gene encoding DmCKII β . The sequence of a 6009 bp *Bam*HI fragment (DmBG3-4) containing the DmCKII β gene is shown (GenBank Accession No. U52952). Exons identified by comparison with cDNA sequences are shown in uppercase, with the seven exons of the β gene indicated by Roman numerals to the right. The deduced amino acid sequence of the β polypeptide is shown in the one-letter code below the DNA sequence. The five alternative 5' exons of the β gene and two poly(A) addition signals are underlined. Splice donors (GT) and acceptors (AG) are italicized, as are potential matches to the *Drosophila* branch site consensus sequence (CTAAT). The two exons in the interval from 1 to 780 (defined by EST clones AI534094, AI534574, AI534706, and AI294873) are presumed to derive from an upstream gene. The spliced transcript contains an open reading frame that extends from 2 to 639, but the deduced polypeptide exhibits no significant homology to sequences in GenBank.

coding exons, combined with the absence of upstream initiation codons in all five classes of cDNA, implies that the encoded polypeptide remains identical in each class of transcript. The exception involves alternative splicing from the donor of exon VI. As illustrated in Fig. 3, clone DmA41 contains an alternative splice from exon VI to an AG dinucleotide located at positions 4657–4658 (Fig. 2). This splice is predicted to alter the β open reading frame, resulting in the replacement of the last two amino acids (KN*) with a new 19 residue sequence (FTTGRPIDSNTQQQQQPH*). The β subunits of different species exhibit considerable divergence in both sequence and length of the C-terminal region [28], so the functional consequences of this change may be minimal. A second cDNA (AI109634) exhibited yet a third splice from the same donor (Fig. 3), but the acceptor in this case could not be identified in available downstream sequence.

Alternative splicing from exon VI may reflect the poor quality of downstream acceptors. In general, all of the introns in the β gene exhibited a good match to the 5' splice site, branch point, and 3' splice site consensus sequences for *D. melanogaster*, including a G-poor region between the branch point and the 3' splice acceptor [29]. The lone exception was the 3' splice acceptor of exon VII, which contained a high proportion of G residues upstream of the AG dinucleotide (Fig. 2, residues 4520–4524). The alternative acceptor utilized in clone DmA41 was also poor, failing to exhibit a reasonable match to the branch point consensus (YTAAT). It is possible that there is insufficient evolutionary pressure on the C-terminal sequence of the β polypeptide to insure faithful processing in this region of the gene.

The 3' end of DmCKII β mRNA is also heterogeneous. Two potential poly(A) addition signals, AATAAA, are located downstream of the open reading frame in the available genomic sequence (Fig. 3). The more distal of these appears to be used, as clone Dm107 [10] contains a poly(A) stretch beginning 22 nucleotides downstream of this signal. We have no data which would indicate use of the proximal signal. In addition, one clone, DmA17, extends approximately 200 bp downstream of the *Bam*HI site, consistent with the existence of at least one more poly(A) addition signal distal to the 3' end of DmBG3-4 (additional introns distal to this site may exist as well).

The diversity of mRNAs derived from the CKII β gene, particularly the existence of alternative 5' UTRs, is currently unique to *Drosophila*. What function might this complexity serve? Since the alternative 5' exons are not predicted to affect the encoded polypeptide, generation of alternative β subunit isoforms does not appear to provide a rationale. Two possibilities suggest themselves. First, the distinct 5' UTRs might alter some property of the message, for example its stability, localization, or translation efficiency. Second, the various isoforms may be expressed in a tissue- and/or

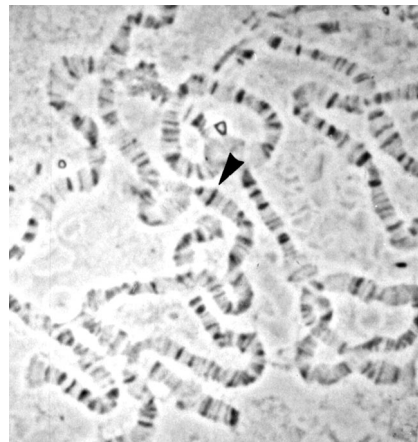
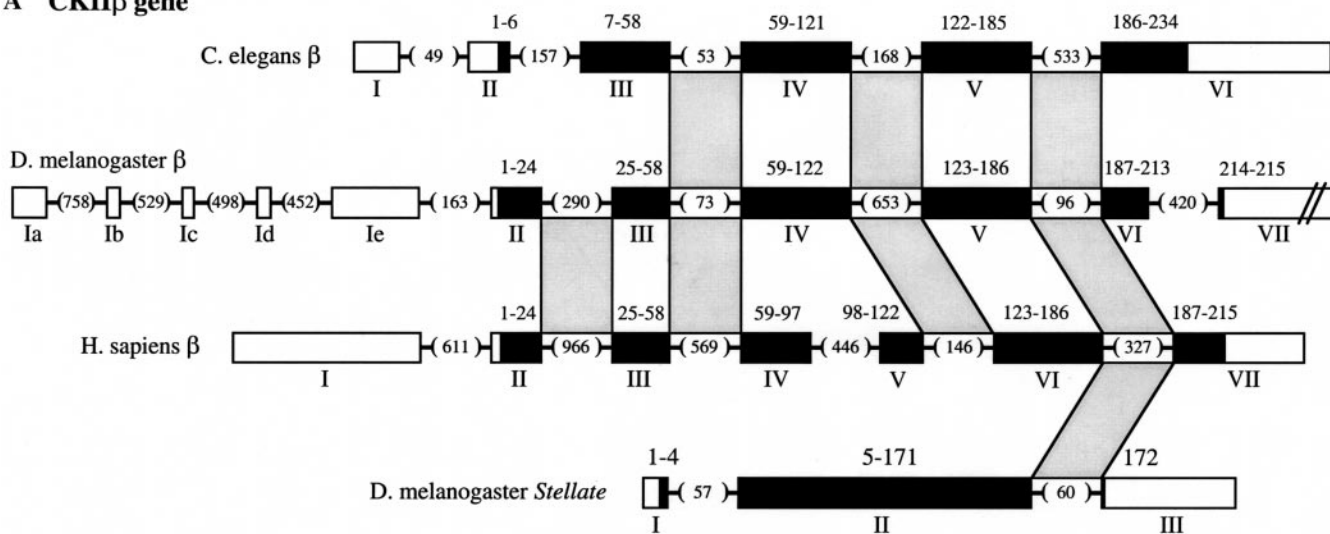


FIG. 4. Cytological location of the DmCKII β gene. *In situ* hybridization to polytene chromosomes was carried out using a biotinylated cDNA probe (Dm98) as described under Materials and Methods. Slides were examined under phase contrast microscopy and photographed at a magnification of 400 \times . The hybridizing band at 10E1-2 on the X chromosome is indicated by the arrowhead.

developmental stage-specific manner. Additional studies will be required to address these possibilities.

In situ hybridization to polytene chromosomes. Both cDNA and cosmid probes were used to localize the β gene to 10E1,2 on the X-chromosome (Fig. 4). The *white* gene, present in the cosmid vector, served as an internal control for hybridization to a single locus (the expected hybridization of the *white* gene was observed at 3C2; data not shown). The single site of hybridization observed for the β gene is consistent with Southern data indicating that the β gene is single-copy [10].

Phylogeny of the intron–exon organization of the β gene. The intron–exon structure of the β gene is identical between mouse and humans [20] and well conserved between *C. elegans* and humans [19]. A comparison of the intron–exon organization of the *C. elegans*, *Drosophila*, and human β genes and *Drosophila Stellate* is shown in Fig. 5A. The location of three introns is precisely conserved (to the base pair) in all three β genes, strongly suggesting that these introns were present in the common ancestor of all three species. A fourth intron (between exons II and III) is precisely conserved in two of the three species (*Drosophila* and humans) but not in the third (*C. elegans*). While it is formally possible that intron sliding accounts for the altered position of this intron in *C. elegans*, current data suggest that intron sliding is rare or nonexistent [30]. Two reasonable possibilities remain: either the intron was present in the common ancestor of all three species, in which case intron loss (and gain) explains the pattern observed in *C. elegans*, or it was acquired after the divergence of *C. elegans* from the lineage leading to *Drosophila* and humans. The latter is the more parsimonious mechanism, as it requires one less

A CKII β gene

B

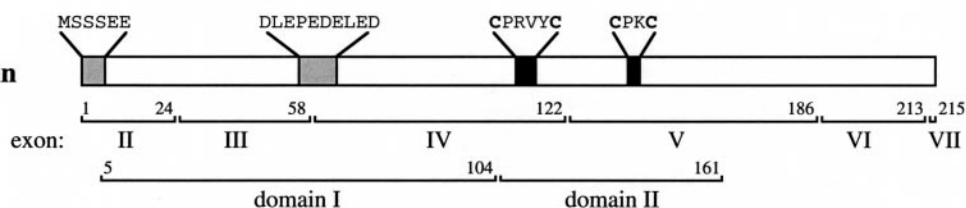
DmCKII β protein

FIG. 5. Intron-exon organization of the CKII β gene. (A) Comparison of the intron-exon structure of the β gene from *C. elegans* (GenBank Accession No. M73827), *D. melanogaster* (U52952), *H. sapiens* (X57152), and *D. melanogaster Stellate* (X15899). Exons are drawn as boxes and designated by Roman numerals, and the open reading frame is shown in black. The numbers above each coding exon indicate amino acid positions, and the numbers in parentheses separating the various exons correspond to the intron size in bp. For the alternative 5' exons of the *Drosophila* gene, the latter numbers represent the distance to the splice acceptor site of exon II. Shaded parallelograms represent introns whose position is conserved among the various genes. Exons are drawn to scale except for exon VII of *Drosophila* (indicated by the double slash). (B) Structural features of DmCKII β . The N-terminal autophosphorylation site, acidic region, and zinc-binding motif are indicated. Regions of the protein encoded by exons II-VII and the two domains defined by X-ray crystallography of the human β subunit [9] are indicated below the figure. Numbers represent amino acid positions.

intron gain or loss (two versus three), but it also requires that nematodes be an outgroup of insects and vertebrates. Although this is the traditional phylogeny, several recent studies indicate that vertebrates may be the outgroup, and the matter remains unresolved [31].

In addition to the shared introns, each of the β genes contains one unique intron in the protein coding region (between exons II and III in *C. elegans*, exons VI and VII in *Drosophila*, and exons IV and V in humans and mouse). Again, the most parsimonious explanation is that each intron was acquired in the line that currently contains it. Each gene also contains a uniquely positioned intron in the 5' untranslated region. These may reflect independent acquisitions in each line, but in a region of so little sequence constraint it is not possible to eliminate the possibility of an ancestral intron whose position has been altered by small deletions or insertions between the intron and the start codon. Finally, *Stellate* (Fig. 5A) as well as the *Su(Ste)* genes and *SSL* [14] share one intron in common with the β

subunit gene. The likely evolution of these genes from the β subunit via a mechanism involving reverse transcription has been discussed [14].

A schematic representation of the exon organization of the *Drosophila* β subunit relative to important structural features of the protein is shown in Fig. 5B. As noted in the introduction, X-ray crystallography of the human β subunit has revealed two structural domains: domain I, residues 5-104, and domain II, residues 105-161 [9]. Although an intron does lie near the junction of these two domains in human β (between residues 97 and 98), its absence in both *Drosophila* and *C. elegans* suggests that it may be of recent origin. Of the three phylogenetically conserved introns, one interrupts domain I (within the acidic region), and a second interrupts domain II, separating the two halves of the zinc-binding motif (CPxxxC and CPxC). The third conserved intron lies downstream of domain II in a region involved in α - β heterodimerization and CKII oligomerization [9]. The possible relationship of this

intron to domain organization in this region is currently indeterminate because the site of this intron lies just beyond the region represented in the X-ray structure. Overall, the available data suggest a relatively poor correlation between known structural domains and ancestral intron locations.

The evolutionary origins of introns continues to be a topic of considerable debate, the two extreme views being denoted as "introns-early" and "introns-late" [30, 32]. The available data on CKII β appear to favor the "introns-late" theory. The positions of the first two phylogenetically invariant introns exhibit no obvious correlation with the domain organization of the β polypeptide, and considerations of parsimony suggest that there have been at least three instances of recent intron acquisition (since the last common ancestor of nematodes, insects, and vertebrates). Nevertheless, the available data cannot exclude the possibility that one or more of the introns is of ancient origin and played a role in exon shuffling to assemble the ancestral β gene. The third conserved intron and/or the unique human intron may fall into the latter category. Sequencing of additional β subunit orthologs and paralogs should further clarify the evolutionary history of this gene.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the excellent technical assistance of Jeffrey S. Beckman, Lisa K. Carlson, Michelle C. Carney, Christopher M. Cerjan, and Michael J. Crawford. This work was supported by National Institutes of Health Grant GM33237 to C.V.C.G. and institutional support from the Eberly College of Arts and Sciences to A.P.B.

REFERENCES

- Allende, J. E., and Allende, C. C. (1995) Protein kinase CK2: An enzyme with multiple substrates and a puzzling regulation. *FASEB J.* **9**, 313–323.
- Guerra, B., and Issinger, O. G. (1999) Protein kinase CK2 and its role in cellular proliferation, development and pathology. *Electrophoresis* **20**, 391–408.
- Meggio, F., Marin, O., and Pinna, L. A. (1994) Substrate specificity of protein kinase CK2. *Cell. Mol. Biol. Res.* **40**, 401–409.
- Wilson, L. K., Dhillon, N., Thorner, J., and Martin, G. S. (1997) Casein kinase II catalyzes tyrosine phosphorylation of the yeast nucleolar immunophilin Fpr3. *J. Biol. Chem.* **272**, 12961–12967.
- Padmanabha, R., Chen-Wu, J. L., Hanna, D. E., and Glover, C. V. (1990) Isolation, sequencing, and disruption of the yeast CKA2 gene: casein kinase II is essential for viability in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **10**, 4089–4099.
- Snell, V., and Nurse, P. (1994) Genetic analysis of cell morphogenesis in fission yeast—A role for casein kinase II in the establishment of polarized growth. *EMBO J.* **13**, 2066–2074.
- Seldin, D. C., and Leder, P. (1995) Casein kinase II alpha transgene-induced murine lymphoma: Relation to theileriosis in cattle. *Science* **267**, 894–897.
- Niefind, K., Guerra, B., Pinna, L. A., Issinger, O. G., and Schomburg, D. (1998) Crystal structure of the catalytic subunit of protein kinase CK2 from *Zea mays* at 2.1 Å resolution. *EMBO J.* **17**, 2451–2462.
- Chantalat, L., Leroy, D., Filhol, O., Nueda, A., Benitez, M. J., Chambaz, E. M., Cochet, C., and Dideberg, O. (1999) Crystal structure of the human protein kinase CK2 regulatory subunit reveals its zinc finger-mediated dimerization. *EMBO J.* **18**, 2930–2940.
- Saxena, A., Padmanabha, R., and Glover, C. V. (1987) Isolation and sequencing of cDNA clones encoding alpha and beta subunits of *Drosophila melanogaster* casein kinase II. *Mol. Cell Biol.* **7**, 3409–3417.
- Bidwai, A. P., Zhao, W., and Glover, C. V. (1999) A gene located at 56F1–2 in *Drosophila melanogaster* encodes a novel metazoan β -like subunit of casein kinase II. *Mol. Cell Biol. Res. Commun.* **1**, 21–28.
- Livak, K. J. (1990) Detailed structure of the *Drosophila melanogaster* *Stellate* genes and their transcripts. *Genetics* **124**, 303–316.
- Balakireva, M. D., Shevelyov, Y., Nurminsky, D. I., Livak, K. J., and Gvozdev, V. A. (1992) Structural organization and diversification of Y-linked sequences comprising Su(Ste) genes in *Drosophila melanogaster*. *Nucleic Acids Res.* **20**, 3731–3736.
- Kalmykova, A. I., Shevelyov, Y. Y., Dobritsa, A. A., and Gvozdev, V. A. (1997) Acquisition and amplification of a testis-expressed autosomal gene, SSL, by the *Drosophila* Y chromosome. *Proc. Natl. Acad. Sci. USA* **94**, 6297–6302.
- Bozzetti, M. P., et al. (1995) The Ste locus, a component of the parasitic cry-Ste system of *Drosophila melanogaster*, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. *Proc. Natl. Acad. Sci. USA* **92**, 6067–6071.
- Cardenas, M. E., Dang, Q., Glover, C. V., and Gasser, S. M. (1992) Casein kinase II phosphorylates the eukaryote-specific C-terminal domain of topoisomerase II *in vivo*. *EMBO J.* **11**, 1785–1796.
- Willert, K., Brink, M., Wodarz, A., Varmus, H., and Nusse, R. (1997) Casein kinase 2 associates with and phosphorylates dishevelled. *EMBO J.* **16**, 3089–3096.
- Jaffe, L., Ryoo, H. D., and Mann, R. S. (1997) A role for phosphorylation by casein kinase II in modulating Antennapedia activity in *Drosophila*. *Genes Dev.* **11**, 1327–1340.
- Wirkner, U., Voss, H., Lichter, P., and Pyerin, W. (1994) Human protein kinase CK2 genes. *Cell. Mol. Biol. Res.* **40**, 489–499.
- Boldyreff, B., and Issinger, O. G. (1995) Structure of the gene encoding the murine protein kinase CK2 beta subunit. *Genomics* **29**, 253–6.
- Hu, E., and Rubin, C. S. (1991) Casein kinase II from *Caenorhabditis elegans*. Cloning, characterization, and developmental regulation of the gene encoding the beta subunit. *J. Biol. Chem.* **266**, 19796–19802.
- Feilotter, H. E., Hannon, G. J., Ruddell, C. J., and Beach, D. (1994) Construction of an improved host strain for two hybrid screening. *Nucleic Acids Res.* **22**, 1502–1503.
- Dale, R. M. K., McClure, B. A., and Houchins, J. P. (1985) A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing. *Plasmid* **13**, 31–40.
- Langer-Safer, P. R., Levine, M., and Ward, D. C. (1982) Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci. USA* **79**, 4381–4385.
- Glover, C. V., Shelton, E. R., and Brutlag, D. L. (1983) Purification and characterization of a type II casein kinase from *Drosophila melanogaster*. *J. Biol. Chem.* **258**, 3258–3265.
- Maroni, G. (1993) An Atlas of *Drosophila* Genes: Sequences and Molecular Features, Oxford Univ. Press, New York.
- Burke, T. W., and Kadonaga, J. T. (1996) *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10**, 711–724.

28. Bidwai, A. P., Reed, J. C., and Glover, C. V. (1995) Cloning and disruption of CKB1, the gene encoding the 38-kDa beta subunit of *Saccharomyces cerevisiae* casein kinase II (CKII). Deletion of CKII regulatory subunits elicits a salt-sensitive phenotype. *J. Biol. Chem.* **270**, 10395–10404.
29. Mount, S. M. (1993) in *An Atlas of Drosophila Genes: Sequences and Molecular Features*, pp. 333–358. Oxford Univ. Press, New York.
30. Logsdon, J. (1998) The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* **8**, 637–648.
31. Blaxter, M. (1998) *Caenorhabditis elegans* is a nematode. *Science* **282**, 2041–6.
32. de Souza, S. J., Long, M., Klein, R. J., Roy, S., Lin, S., and Gilbert, W. (1998) Toward a resolution of the introns early/late debate: Only phase zero introns are correlated with the structure of ancient proteins. *Proc. Natl. Acad. Sci. USA* **95**, 5094–5099.