Genomic organization and linkage via a bidirectional promoter of the AP-3 (adaptor protein-3) μ 3A and AK (adenosine kinase) genes: deletion mutants of AK in Chinese hamster cells extend into the AP-3 μ 3A gene

Bhag SINGH and Radhey S. GUPTA¹

Department of Biochemistry, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

The cDNA and genomic DNA for the μ 3A subunit of the AP-3 (adaptor protein-3) complex were cloned from Chinese hamster cells. The AP-3 μ 3A genes in Chinese hamster, human and mouse each comprise nine exons and eight introns, with all introns located in identical positions in the species studied. The AP-3 μ 3A genes in these species are linked in a head-to-head fashion with the gene for the purine salvage pathway enzyme AK (adenosine kinase). These genes share the first exon, and a 512 bp fragment covering the intervening untranslated sequence has the characteristic of a CpG island promoter, and it effectively carried out transcription in both directions. Deletion studies indicate that this region contains both positive and negative regulatory elements affecting transcription of these genes. In comparison with the AP-3 μ 3A gene (27 kb), the AK gene in human is very large (558 kb), with average exon and intron lengths of approx. 100 bp and 55.7 kb respectively. The ratio of non-coding to cod-

ing sequence in the human AK gene is > 550, which is the highest reported for any gene. We also present evidence that a number of AK[−] mutants of Chinese hamster ovary cells contain large deletions that affect both of these genes. In addition to lacking part of the AK gene, two of these mutants also lacked all of the exons and introns corresponding to the AP-3 μ 3A gene. These mutants should prove useful in elucidating the role of AP-3 μ 3A in vesicle-mediated protein sorting – a process that is altered in Hermansky–Pudlak syndrome. Detailed phylogenetic analysis of the μ family of proteins presented here also provides insight into how different AP complexes are related and may have evolved.

Key words: adaptor protein μ chain, adenosine kinase, deletion mutant, Hermansky–Pudlak syndrome, intron–exon arrangement, phylogenetic analysis.

INTRODUCTION

The AP (adaptor protein) complexes are essential components of vesicles that mediate transport of proteins between different cellular compartments [1,2]. These complexes are responsible for selecting vesicle cargo by recognizing the sorting motifs in proteins, and also for the attachment of vesicles to the membrane. In mammalian cells, four different AP complexes (AP-1 to AP-4) have been identified. Each of these complexes has a similar heterotetrameric structure consisting of two large subunits ($\gamma/\alpha/\delta/\epsilon$ and $\beta 1$ – $\beta 4$), one medium-sized subunit of 47–50 kDa ($\mu 1$ – $\mu 4$) and one smaller subunit ($\sigma 1$ – $\sigma 4$) of ~20 kDa. A mutation in the $\beta 3A$ subunit of the AP-3 complex is responsible for type 2 Hermansky–Pudlak syndrome [1,2], a human genetic disorder characterized by defects in various lysosome-related organelles, such as melanosomes and platelet dense granules [3,4]. Mutations in other AP components (AP-3 $\delta 3$ and AP-3 $\mu 3$) also cause pigmentation defects associated with abnormal vesicular transport in animal models [1–4].

AK (adenosine kinase) is a major purine salvage pathway enzyme which catalyses the conversion of adenosine into AMP and plays a central role in maintaining intracellular and interstitial levels of adenosine [5,6]. Because of the potent pharmacological activity of adenosine in the cardiovascular, nervous, respiratory and immune systems, there is much interest in understanding how AK activity is regulated *in vivo* and can be modulated. The AK cDNA has been cloned from several species and important

information concerning its structure–activity relationship has been obtained [7–12]. However, our understanding of AK gene structure in mammalian species is still incomplete. In an earlier study, we reported that the AK gene in human, mouse and Chinese hamster is composed of 11 exons [13]. However, the exact lengths of the introns, which were indicated to be very large, were not determined. An unexpected observation made during that work was that the AP-3 μ 3A gene is located immediately upstream of the AK gene [14,15]. The two genes are linked in a head-to-head fashion and their transcribed regions are separated by less than 300 bp [13]. This observation strongly suggested that these genes share a common or overlapping promoter. Another interesting observation made during this work was that many of the AK[−] mutants of CHO (Chinese hamster ovary) cells harboured large deletions leading to the loss of several of introns and exons [7,13]. Because of the close linkage seen between the AP-3 μ 3A and AK genes, it is of interest to determine whether any of these deletions also affect the AP-3 μ 3A gene.

In the present work, we have cloned the cDNA and genomic DNA for AP-3 μ 3A from CHO cells and have compared its intron–exon organization in different species. We also demonstrate that the sequence region that lies between the AP-3 μ 3A and AK genes can effectively carry out transcription in both orientations and that it serves as a bidirectional promoter. We also report that the AK gene in humans is very large (558 kb), with average intron length of > 50 kb. A number of AK[−] mutants of CHO cells are shown to harbour large deletions that also cause

Abbreviations used: AK, adenosine kinase; AP, adaptor protein; CHO, Chinese hamster ovary; Tub^R, tubercidin-resistant; WT, wild-type.

¹ To whom correspondence should be addressed (e-mail gupta@mcmaster.ca).

The nucleotide sequence data reported will appear in GenBank[®], DDBJ, EMBL and GSDN Nucleotide Sequence Databases under accession numbers AY346106–AY346109.

deletion of the neighbouring AP-3 μ 3A gene. Additionally, we present phylogenetic analysis of the μ family of protein sequences that provides insight into their evolution.

MATERIALS AND METHODS

Cell lines and culture conditions

The sources of various cell lines used in the present studies, i.e. WT (wild-type) CHO cells, human diploid HSC-172 fibroblasts and mouse Ltk- cells, have been described in earlier work [16]. Single-step mutants of CHO cells resistant to the adenosine analogue tubercidin (Tub^R cells) were selected as described previously [17]. These mutants exhibit a high degree of resistance (> 100-fold) to various adenosine analogues and contain no detectable AK activity (AK⁻ mutants) [17,18]. All cells were grown at 37 °C in α -minimal essential medium supplemented with 5% (v/v) fetal bovine serum in 5% CO₂/95% air atmosphere.

Screening of genomic libraries and Southern blot hybridization

A genomic library of CHO cells in λ EMBL-3 vector (cat. no. JL1003j) was purchased from Clontech. Approx. 2×10^5 phage particles (plaque-forming units) of this library were plated on *Escherichia coli* L392 cells and the plaques were transferred on to duplicate nitrocellulose filters. The filters were hybridized under stringent conditions to a ³²P-labelled DNA probe made for CHO μ 3A cDNA. The positive clones were purified and analysed for the presence of different exons and introns of the μ 3A gene. For Southern blot analysis, 10 μ g of genomic DNA from CHO WT and mutant cell lines was digested to completion with the indicated restriction enzymes. After electrophoresis in 0.8% (w/v) agarose gels, the DNA fragments were transferred to a nylon membrane and hybridized to labelled AK or μ 3A cDNA probes under stringent conditions [7,13].

Reporter constructs

The sequence upstream of the translation initiation codons for the μ 3A and AK genes was PCR amplified from human HSC-172 cells, mouse Ltk- cells and WT CHO DNA using specific primers containing *Bgl*II and *Kpn*I restriction sites at the ends of the two primers. The clones in the reverse orientation, as well as various deletions for this region, were also made by means of PCR using appropriate specific primers containing the *Bgl*II and *Kpn*I restriction sites. The amplified fragments were initially cloned in the Pdrive vector (Qiagen). After digestion of the vector with *Bgl*II and *Kpn*I, the fragments obtained were subcloned into the pGL3-basic reporter vector (Promega) upstream of the luciferase gene lacking any promoter or enhancer sequences. Cos-1 cells were transfected with purified DNA from various constructs using lipofectAMINETM reagent (Invitrogen) using the manufacturer's protocol. As an internal control for transfection efficiency, DNA samples from all control and test inserts were each mixed with an equivalent amount of DNA from another construct expressing the β -galactosidase gene. Positive and negative control vectors with either simian virus 40 promoter and enhancer sequences or simply a basic vector with no additional sequence were also included in all experiments. All transfections were performed in triplicate to minimize sample variability. Luciferase activity in the cell extracts was determined using a luciferase assay system from Promega and read on an EG&G Berthold luminometer. Luciferase counts obtained were normalized for both β -galactosidase expression and total protein.

Primer extension

Primer extension was carried out with radiolabelled primers. The primers used are as follows: AK, 5'-TCCACCTTCAGCTTTTGTG-3'; μ 3A, 5'-TGGGCGACTCAGTGGTTCGG-3'. Total RNA was purified by Trizol Reagent (Life Technology). The oligonucleotides were end-labelled using [γ -³²P]ATP and T4 polynucleotide kinase. Primer extensions were performed by annealing labelled oligonucleotides to total RNA as follows. The oligonucleotides and RNA mixture was heated to 65 °C for 5 min and cooled slowly to room temperature. Primers were extended with Moloney murine leukaemia virus reverse transcriptase by incubating the reaction mixture at 45 °C. The reactions were terminated by the addition of loading buffer. The extension products were resolved on a sequencing gel along with the sequence reaction of a DNA template performed with the same primers.

Phylogenetic analysis

Phylogenetic analysis of protein sequences was carried out as described in our earlier work [19,20]. A multiple sequence alignment of the μ (or Mu) family of APs from various species was created using the ALIGN program. All sequence regions where the alignment was unreliable, or where large inserts or deletions were present in certain homologues, were excluded from phylogenetic analysis. A neighbour-joining distance tree showing the branch lengths was constructed using the PROTDIST, NEIGHBOR and DRAWTREE programs. The aligned sequences were also bootstrapped 100 times using the SEQBOOT program, and a consensus tree based on the bootstrapped sequences was obtained using the PROTDIST, NEIGHBOR and CONSENSE programs. All phylogenetic programs used are part of the PHYLIP 3.5 software [21]. Pairwise similarity between different protein sequences was determined using the PALIGN program from the PCGENE software package. The GenBank accession numbers of the various μ sequences that were used in these studies are as follows: human μ 3A, 6912240; mouse μ 3A, 20531985; rat μ 3A, 19424296; human μ 3B, 5803000; mouse μ 3B, 20073160; rat μ 3B, 18959246; *Discophyga ommata* μ 3, 1351947; *Drosophila melanogaster* μ 3, 28571412; *Anopheles gambiae* μ 3, 31242415; *Caenorhabditis elegans* μ 3, 17568099; *Dictyostelium discoideum* μ 3, 12000359; *Arabidopsis thaliana* μ 3, 25404135; *Trypanosoma brucei* μ 3, 22347748; human μ 1A, 14210504; mouse μ 1A, 6671557; human μ 1B, 9506797; mouse μ 1B, 13277588; *Anopheles gambiae* μ 1, 31227815; *Drosophila melanogaster* μ 1, 24645407; *Caenorhabditis elegans* μ 1, 1078840; *Neurospora crassa* μ 1, 28949965; *Schizosaccharomyces pombe* μ 1, 19112573; *Dictyostelium discoideum* μ 1, 12000357; *Arabidopsis thaliana* μ 1, 15220202; *Plasmodium falciparum* μ 1, 23619052; human μ 2, 6753074; *Drosophila melanogaster* μ 2, 24649014; *Dictyostelium discoideum* μ 2, 28828638; *Giardia lamblia* μ 2, 19110270; human μ 4, 5442366; mouse μ 4, 10946736.

RESULTS

Cloning of the Chinese hamster μ 3A gene and intron-exon arrangement within the gene

In our earlier work on the CHO AK gene, a number of phage clones were isolated which contained the first exon of the AK gene along with the 5'-end untranslated regions of both the AK and μ 3A gene transcripts [13,14]. Furthermore, screening of the CHO genomic library with a DNA probe for the Chinese hamster μ 3A upstream region and a cDNA probe for the mouse μ 3A gene led to the isolation of several genomic clones containing

Table 1 Intron–exon organization of the Chinese hamster μ 3A gene

The first and last exons also contain non-coding sequences, whose lengths vary in the reported clones. The exon and intron sequences are shown in upper and lower case respectively.

Exon no.	Length (bp)	5'-Exon (end)	Donor	Intron length (bp)	Acceptor	3'-Exon (beginning)
1	264	GTGTCGGG (non-coding)	gtgagcgctg...	6400	...gttttttag	AAA ATG ATC M I
2	276	ACT TTT CAG T F Q	gtttgtattc...	669	...attttgttag	GAC TAC TTT D Y F
3	172	TCT ATT ACA G S I T	gtatgaaagc...	3500	...taattcttag	GC AGT AGT G S S
4	138	GAT AAA TCA G D K S	gtaattgtgt...	866	...ctttcttag	GA TCT ACA G S T
5	86	TCT TTC ATG S F M	gtaagtttg...	857	...ttctgacag	AAC CCA AGG N P R
6	134	TCA CAG AA S S N	gtaaattgat...	937	...ctttgttag	T CTT GTG L V
7	208	GTA ACC AAG V T K	gtacagccgc...	1458	...tcttttaag	GTA CTA GCA V L A
8	145	ATT TCA G I S	gtaaagatag...	422	...tttgacacag	GC TTA AAA G L K
9	105	GGG AAG TTC CAA GTG AGG ACA TGA G K G Q V R T -	(3' end non-coding sequence)

Figure 1 Intron–exon organization of the μ 3 genes in different eukaryotic organisms

The exons (Ex) sequences are drawn to scale. For the introns whose lengths vary in different species, only their lengths and positions are indicated. For the human, the gene structures for both the μ 3A and μ 3B genes are shown.

the entire Chinese hamster μ 3A gene sequence. Based on this information, a cDNA clone for the Chinese hamster μ 3A gene was also isolated by means of reverse transcription–PCR. The sequence data for these clones have been deposited in GenBank under accession numbers AY346106–AY346109. The μ 3A protein in Chinese hamster consists of 418 amino acids, and it differs from its mouse and human homologues in only a single amino acid residue. However, the rat homologue differs from these species at four positions (results not shown), which is unexpected. From a comparison of the cDNA and genomic sequences, we have deduced the intron–exon arrangement within the μ 3A gene (Table 1). The μ 3A gene in Chinese hamster consists of nine exons and eight introns. The length of the exons in the μ 3A

gene varies from 86 bp to 276 bp, whereas introns range in size from 0.42 kb to approx. 6.4 kb. The largest of the μ 3A introns lies just upstream of the initiation codon. Hence all of the upstream untranslated sequence for the μ 3A transcript is present in the first exon (0.56 kb), which also contains part of the AK coding sequence. The intron–exon arrangement of the μ 3A gene in other mammalian species is identical with that in Chinese hamster (Figure 1). All of the intron–exon boundaries, as well as the lengths of various exons, are the same in human, mouse and Chinese hamster. However, the lengths of the intervening introns vary among species, with the human introns generally being longer than those in rodents, resulting in an overall length of the μ 3A gene in these species of between 13 and 27 kb. In

Figure 2 Genomic arrangement of the μ 3A and AK genes in mammalian species

The first exon is common to both of these genes, and its complete sequence for Chinese hamster is shown. This exon contains all except 3 bp of the non-coding sequence of the μ 3A transcript, as well as the entire upstream sequence of the AK-long transcript. This exon also contains the sequence information for the first 21 amino acids that are present only in the AK-long form of the protein. The single transcription sites for the μ 3A gene transcript as well as the multiple transcription initiation sites for AK that have been identified in the present work are marked. The positions of various sequence motifs for transcription factors identified by computer analysis are indicated by the lines above the sequence.

contrast with the mammalian species, intron–exon boundaries for the μ 3 (Mu3) gene are not conserved in other species, such as *C. elegans* and *D. melanogaster* (Figure 1).

Mammalian species contain two different homologues of the μ 3 protein [1,14,15]. The μ 3A and μ 3B homologues from human, rat and mouse exhibit approx. 84–85% sequence identity at the amino acid sequence level [15]. We have also examined the intron–exon arrangement of the human μ 3B gene, sequence information for which is now available (Figure 1) [22]. Except for the first intron, which in this case contains only part of the 5'-untranslated sequence, all of the other introns and exons in the μ 3B gene are located in the same positions as in the μ 3A gene, providing strong evidence that these genes are closely related.

Promoter activity of the sequence region linking the AK and μ 3A genes

The AK and μ 3A genes in Chinese hamster are linked in a head-to-head fashion, with the transcribed regions of the two genes being separated by less than 300 bp [13]. The genetic organization of these two genes and the sequence of the first exon depicting the intervening region is shown in Figure 2. Since the introns are very large in comparison with the exons, they are not drawn to the scale and only their positions are indicated. A similar close arrangement of these genes is observed in the human and mouse genomes [13] (results not shown), indicating that a head-to-head orientation of these two genes is a general characteristic of mammalian species. The first exon, whose length varies between 560 and 570 bp in mouse, human and Chinese hamster, is common to both genes, and contains nearly all of the upstream untranslated sequence for the μ 3A gene as well as all of the 5'-upstream sequence for the AK gene and the first 62 bp of its coding sequence (Figure 2). Based on the observed arrangement, it is expected that the intervening

512 bp sequence between these genes (Figure 2) should contain the promoter sequences for both of them.

Computer analysis of the first exon using the EMBOSS (EMBL-EBI) Searcher program identified a 400 bp region located approx. 50 bp upstream of the initiator ATG in the AK sequence as a CpG island, which is a common characteristic of many eukaryotic promoters [23–25]. This was the only CpG island detected by the program in a large gene sequence (> 5 kb) surrounding this exon. We also carried out experiments to determine the transcription start sites for the two genes in CHO cells by primer extension analysis. These studies identified a single strong transcription initiation site for the μ 3A gene in CHO cells which is located 280 bp upstream of the initiator codon (see Figure 2). In contrast with the μ 3A gene, primer extension analysis indicated the presence of multiple closely spaced transcription start sites for the AK gene, located 40–60 bp upstream of the initiator codon (Figure 2). Computer analysis also detected binding sites for a number of transcription factors (i.e. Sp1, activator protein-1, activating transcription factor, nuclear factor- κ B and androgen receptor) in the intervening sequence, but no well demarcated TATA or CAAT boxes were identified. The absence of a TATA box, the presence of multiple transcription start sites, and location within a GC-rich region are characteristics common to many eukaryotic housekeeping gene promoters, a number of which exhibit bidirectional activity [26–30].

The ability of the intervening sequence to serve as a promoter in either direction was assayed directly using a reporter system from Promega. For these studies, 512–530 bp fragments that comprised nearly all of the untranslated sequence that lies between the initiator codons for these genes were amplified from CHO, mouse and human cells, and inserted upstream of a reporter luciferase gene in the pGL3-basic vector (Promega) that lacked any promoter or enhancer sequences. The fragments were inserted in both

orientations in order to measure the potential promoter activity of the sequence in both directions. Column-purified DNA from different clones was used to transfect monkey COS-1 cells and the activity of the reporter luciferase gene was assayed. The luciferase counts obtained were normalized for both control β -galactosidase expression and total protein. The results of these experiments indicated that transfection of cells with the vector constructs containing the intervening sequence (in both orientations) from all three species resulted in >200-fold stimulation of luciferase activity in comparison with controls transfected with DNA from only the basic PGL-3 vector (results not shown). Furthermore, for all three of these mammalian species, the promoter activity in the direction of the μ 3A gene was 4–6-fold higher than that towards the AK orientation. These results provided evidence that the intervening sequence between the AK and μ 3A genes is able to serve as a bidirectional promoter to drive transcription in both directions.

In order to delineate the transcription regulatory elements that may be present in the intervening region, a number of deletions were made in both orientations. For these studies, different portions of the intervening sequence were PCR-amplified and cloned in both orientations upstream of the reporter luciferase gene. The luciferase activity of the various constructs was measured as before, and the results of these studies are presented in Figure 3(A). For promoter activity in the AK orientation, deletion of the first 200 bp that lies immediately upstream of the AK initiator ATG, as in deletion #15, led to a 3-fold enhancement in promoter activity over that seen with the entire 512 bp intervening fragment. This suggests that this region contains certain negatively regulatory elements for AK gene expression. However, deletions that extended beyond this point resulted in a severe reduction in reporter gene activity, indicating that the sequence lying beyond this point is important for promoter activity in the AK orientation. This inference is also supported by the activities of the deletions from the other end of the sequence. In this case, maximum promoter activity in the AK orientation was seen with a fragment (corresponding to deletion #3N) that contained the sequence from bp 50 to 400. The promoter activity of this fragment was more than 3-fold that seen with the entire 512 bp intervening fragment. These results indicate that the first 50 bp of the intervening sequence (from the μ 3A end) also exerts a negative regulatory effect on AK gene expression. All constructs that involved deletion of more than 100 bp from the μ 3A end resulted in a drastic loss of reporter gene activity in the AK orientation. These studies indicate that the sequence region from 50 to 315 bp (from μ 3A end) contains most of the promoter elements necessary for optimal transcription of the AK gene.

Similar deletion studies were also performed to investigate promoter activity in the opposite orientation (Figure 3B). As indicated above, the promoter activity of the intervening sequence in the μ 3A orientation was about 4–6-fold higher than that in the AK direction. None of the deletions in this case showed a significant increase in reporter gene activity over that seen with the entire 512 bp intervening fragment. Similar to the promoter activity in the AK orientation, 200 bp could be deleted from the AK end (deletion #28/2) without any loss of reporter gene activity. Further deletion from that end did lead to a loss of promoter activity. On the other hand, removal of first 100 bp from the μ 3A end caused a drastic reduction in promoter activity. The results of these studies provide evidence that the sequence between 50 and 350 bp from the μ 3A end is acting as a bidirectional promoter, and that all important elements required for optimal transcription of both the AK and μ 3A genes are present within this region. It should be mentioned that although computer analyses indicated that the intervening region contained a number of sequence motifs

Figure 3 Promoter activity of various fragments of the intervening sequence between the AK and μ 3A genes in both orientations

Different fragments of the intervening sequence between the AK and μ 3A genes were PCR-amplified, and these fragments were fused upstream of the luciferase (LUC) gene in either the AK orientation (A) or the μ 3A orientation (B). The sequence information that is present in different deletion clones, as well as their orientations with respect to the luciferase gene whose transcription was studied, are shown in the right-hand panels. The clones with the 1–512 bp fragment contain the entire intervening sequence. After transfection of Cos-1 cells with these DNA fragments, the relative luciferase activity in cell extracts was determined and normalized (left panels). Control experiments with DNA from the basic vector lacking any promoter or insert were included in all experiments.

for transcription factors (see Figure 2), deletion studies indicated that none of these played an essential role in the transcription of these genes.

The AK gene in humans is unusually large and contains very large introns

Our earlier work on the AK gene in human and Chinese hamster has indicated that some of the introns within this gene are very large (>15–20 kb) [13]. However, the exact lengths of the introns or the size of the AK gene had not been determined. The availability of the human genome sequence has now enabled us to clarify this issue [22]. The AK gene in humans, as in other mammalian species (mouse, Chinese hamster and rat), consists of 11 exons. The sizes of various exons and introns that are found within this gene are given in Table 2. The exons in the AK gene range in length from 36 to 765 nt. In contrast, the introns in the

Table 2 Intron–exon organization of the human AK gene

The first and last exons also contain non-coding sequences, whose lengths vary in the reported clones. Exon and intron sequences are shown in upper and lower case respectively. The AK gene is located on human chromosome 10 at position 10 cen-q24 on contig Hs10_8740 (accession no. NT_008583.15).

Exon no.	Length (bp)	Location on chrom. 10 contig	5'-Exon (end)			Donor	Intron length (bp)	Acceptor	3'-Exon (beginning)		
1	115	24 463 259–24 463 373	GCG A	CTG L	AG	gtgagcgtg...	48 321	...gttttttag	A	GAA E	AAT N
2	75	24 511 694–24 511 768	CTT L	GAT D	AA	gtaagtatta...	23 700	...tgttatacag	G	TAT Y	TCT S
3	54	24 535 468–24 535 521	AAG K	GAA E	CT	gtaagtgcatt...	90 076	...ttccttatag	G	TTT F	GAT D
4	79	24 625 597–24 625 657	GCT A	CAG Q		gttggttaat...	79 414	...tgttttacag		TGG W	ATG M
5	173	24 705 071–24 705 243	GAC D	AAC N	AG	gtcagtgtaa...	4158	...gttggttttag	G	TCC S	CTC L
6	109	24 709 401–24 709 509	ATA I	GCA A		gtaagtactt...	126 677	...cttcacccag		GGC G	TTT F
7	171	25 836 186–24 836 356	GAG E	ACA T		gtgagtacc...	63 856	...tttatttcag		GAA E	GCT A
8	36	24 900 212–24 900 247	TTT F	GAG E		gtgagttaac...	11 062	...ctattaatag		ACT T	AAA K
9	115	24 911 309–24 911 423	GCT A	ACA T	G	gtacatgtgg...	69 689	...ctaattgcag	AA	AGT S	GAA E
10	87	24 981 112–24 981 199	GTT V	GGA G	G	gtacagacta...	38 052	...tgtgttctag	GT	TTT F	CTG L
11	122	25 019 251–	CCA P	GAC D	TTC F	CAC H	TGA (+ 3' end non-coding sequence)				

AK gene vary in length from the smallest of 4.2 kb to the largest of 127 kb. The average length of the 10 introns in this gene is 55.7 kb, giving rise to a total length of the AK gene in the human of 558 kb. Since the protein coding sequence for AK is only 1083 bp, the ratio of non-coding to coding sequence in this gene is > 550, which is one of the highest observed in mammalian species [31,32].

Deletion mutants affecting the AK gene extend into the μ 3A gene

AK, which catalyses the conversion of adenosine into AMP, is also responsible for converting a wide variety of adenosine analogues into their toxic phosphorylated forms [33–36]. In our earlier work, mutants resistant to a large number of adenosine analogues (such as tubercidin, toyocamycin, formycin A, pyrazofurin, etc.) were selected and characterized [17,18,36,37]. The most common type of resistant mutants (referred to as class A mutants), which exhibit high degree of resistance to various adenosine analogues (> 500-fold), are devoid of any detectable AK activity in cell extracts. Such mutants occur at an unusually high spontaneous frequency ($\sim 10^{-3}$ – 10^{-4}) in CHO cells [18,37]. Two of the class A mutants in CHO cells were shown previously to contain very large deletions within the AK gene. One of these mutants involved loss of all DNA between exons 5 and 8, whereas in the other the deletion extended from exon 2 to exon 8 [13].

Since the AK gene (558 kb) is very large in comparison with the μ 3A gene (13–27 kb), the possibility that deletions in some of the AK mutants may also extend into the μ 3A gene was investigated. To investigate this, Southern blot analysis was performed on DNA from a number of independent Tub^R mutants of CHO cells that we have isolated. The blots were hybridized to ³²P-labelled DNA probes made for either the first three exons of AK cDNA or the full-length μ 3A cDNA. The results of these studies for three of the Tub^R AK[−] mutants and WT CHO cells are presented in Figure 4(A). As seen from Figure 4(A), WT CHO cells and the Tub^R-16 mutant showed similar hybridization patterns. The cDNA probe in both of these cell lines hybridized

to two different fragments of the genomic DNA. In contrast, for two of the other Tub^R mutants (Tub^R-3 and Tub^R-5) there was no hybridization observed for the AK cDNA probe, indicating that these mutants lack this region (i.e. the first three exons) of the AK gene. After stripping of the labelled probe, the same blot was subsequently hybridized to a cDNA probe for the μ 3A cDNA. The WT CHO cells and the Tub^R-16 mutant again showed similar hybridization patterns, indicating that the μ 3A gene was not affected in this mutant. However, in the Tub^R-3 and Tub^R-5 mutant lines, which lacked the first three exons of the AK gene, no hybridization was observed to the μ 3A cDNA probe, indicating that in these mutants the entire μ 3A gene was deleted.

To confirm this inference, PCR amplification was carried out on the genomic DNA from different cell lines using specific primers for different exons within the AK (exons 1, 5 and 7) and μ 3A (exons 2, 3 and 8) genes. The results of these studies are shown in Figure 4(B). With WT cells and the Tub^R-16 mutant, all of the exon-specific primers amplified the expected sizes of fragments for both the AK and μ 3A genes, indicating that this mutant does not involve any large deletion affecting these genes. However, for the Tub^R-3 and Tub^R-5 mutants, no amplification was observed with primers for AK exon 1, but fragments of expected sizes were amplified for the other two exons (5 and 7). The DNA from the Tub^R-3 and Tub^R-5 mutants also failed to amplify any of the expected fragments of the μ 3A gene using PCR primers specific for exons 2, 3 and 8. These results confirm that the Tub^R-3 and Tub^R-5 mutants contain large deletions covering the first three exons of AK as well as the entire μ 3A gene.

Evolutionary relationships within the μ (Mu) family of proteins

Four different types of AP complexes have been identified in eukaryotic organisms [1,2,15]. However, the relationship among these complexes and how they have evolved is presently unclear. Because homologues of the μ chain are present in all of these complexes, they provide an opportunity to gain insight in this regard. Table 3 presents a pair-wise similarity matrix of

Figure 4 Deletion of the μ 3A gene in AK⁻ mutants of CHO cells

(A) Southern blot analysis of *Eco*RI-digested genomic DNA from WT CHO cells and three AK⁻ mutants (Tub^R-3, Tub^R-5 and Tub^R-16). The blot was initially hybridized to a DNA probe covering the first three exons of the AK coding sequence from CHO cells (left panel). After stripping of the labelled probe, the same blot was subsequently hybridized to a DNA probe for the full-length μ 3A cDNA from CHO cells. (B) PCR amplification with DNA from WT cells and AK⁻ mutants using specific primers for different AK and μ 3A exons. The primers used for AK were for regions that covered exons 1, 5 and 7, and they amplified (as expected) fragments of 400, 168 and 86 bp respectively. The PCR primers for the μ 3A gene were for regions that flanked exons 2, 3 and 8, and the expected sizes of the amplified fragments in this case were 379, 449 and 574 bp respectively.

representative members of the μ family of protein sequences from different AP complexes. The μ homologues from AP-1, AP-2 and AP-3 complexes have been found in a broad range of eukaryotic species [1], whereas the AP-4 μ homologues have thus far been identified only in mammalian species. From the pair-wise identity scores for different sequences, it is clear that for the AP-1, AP-2 and AP-3 families of proteins the highest identity/similarity was seen for other members of the same family as compared with those belonging to the other families. Further, the μ homologues from AP-1 and AP-2 complexes exhibited greater similarity to each other (in the range 39–42%) than seen for either the AP-3 or AP-4 family members (range 26–32%). Of the different μ protein homologues, those belonging to the AP-4 family were found to be highly divergent, exhibiting only 23–30% identity with other μ homologues. The mammalian species contain two different homologues for the μ 1 and μ 3 proteins. The μ 1A (or μ 1B) homologues from human, rat and mouse are much more similar to each other than to the μ 1B (or μ 1A) homologue from the same species [15]. The same is true for the μ 3A and μ 3B homologues.

To gain further insight into their evolutionary relationships, a phylogenetic tree was constructed from a multiple alignment of the μ family of protein sequences. Figure 5 shows a neighbour-joining distance tree indicating how different μ homologues are related to each other. The numbers on different nodes (or branch points) denote percentage bootstrap scores, which provides a measure of the reliability of the observed relationships [38]. Thus a bootstrap value of 100 at the nodes leading to the AP-3 family indicate that, 100% of the time, the sequences belonging to this family group together rather than branching with any of the other sequences. Bootstrap scores of >75% are considered quite reliable. As seen from Figure 5, the μ homologues from AP-1, AP-2 and AP-3 complexes form well-defined clades in the tree whose distinctness is strongly supported by various criteria (e.g. branch lengths, bootstrap scores). Furthermore, the AP-1 and AP-2 families of homologues are much more closely related to each other in the tree than they are to the AP-3 family of protein sequences. The μ homologues from the AP-4 family are characterized by very long branch lengths, indicating their highly

Table 3 Amino acid identity matrix for the μ family of adaptin protein sequences

Values are percentage identity between pairs of protein sequences. These were calculated using the program PALIGN from the POGENE software package using the structure–gene matrix and unit gap and open gap costs of 2 and 7 respectively. The genus abbreviations used in the species names are: *D.*, *Dictyostelium*; *S.*, *Schizosaccharomyces*; *A.*, *Arabidopsis*; *G.*, *Giardia*. GenBank accession numbers of the various protein sequences used are as follows: (A) 6 912 240; (B) 5 803 000; (C) 28 571 412; (D) 17 568 099; (E) 12 000 359; (F) 25 404 135 (G) 14 210 504; (H) 9 506 797; (I) 12 000 357; (J) 19 112 573; (K) 6 753 074; (L) 24 649 014; (M) 28 828 638; (N) 19 110 270; (O) 5 442 366.

		Identity (%)														
Species/homologue		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Human AP3 μ 3A	A	100	84.2	69.4	68.7	51.7	43.9	28.7	31.3	29.7	27.8	26.8	26.8	27.8	25.8	23.0
Human AP3 μ 3B	B		100	69.9	67.2	50.5	43.6	26.8	30.1	29.7	27.0	26.6	27.8	27.8	23.0	24.9
<i>Drosophila</i> AP3 μ 3	C			100	67.5	47.5	42.2	29.9	31.1	27.2	30.1	29.6	27.5	30.4	23.4	25.1
<i>C. elegans</i> AP3 μ 3	D				100	52.4	41.3	31.9	31.6	33.1	30.7	30.4	30.4	29.2	25.6	22.6
<i>D. discoideum</i> AP3 μ 3	E					100	43.2	30.6	32.3	30.6	30.7	28.5	27.6	29.7	25.2	24.5
<i>A. thaliana</i> AP3 μ 3	F						100	28.1	26.9	28.5	26.1	25.7	26.1	27.3	26.4	24.9
Human AP1 μ 1A	G							100	79.7	69.5	63.6	40.2	41.1	42.3	30.0	29.6
Human AP-1 μ 1B	H								100	66.4	60.0	39.2	38.8	40.0	32.6	30.3
<i>D. discoideum</i> AP-1 μ 1	I									100	59.4	41.4	40.7	42.8	31.5	27.3
<i>S. pombe</i> AP-1 μ 1	J										100	39.7	39.9	41.1	29.6	29.3
Human AP-2 μ 2	K											100	87.6	55.9	33.0	28.5
<i>Drosophila</i> AP-2 μ 2	L												100	52.6	32.5	27.5
<i>D. discoideum</i> AP-2 μ 2	M													100	34.3	28.4
<i>G. lamblia</i> AP	N														100	25.6
Human AP-4 μ 4	O															100

Figure 5 Phylogenetic tree based on the μ family of protein sequences

The tree shown is a neighbour-joining distance tree constructed using the PHYLIP program (see the Materials and methods section). The sequences were bootstrapped 100 times, and bootstrap scores for some of the important nodes are indicated. GenBank accession numbers are given in the Materials and methods section.

divergent nature. The AP-2 family of proteins is indicated to be most closely related to the AP-4 family. The phylogenetic tree also clearly illustrates that the $\mu 3A$ homologues from different mammalian species are much more closely related to each other than they are to the $\mu 3B$ homologues from the same species. Different $\mu 3$ homologues are not found in earlier branching species such as insects, nematodes and plants. These observations provide evidence that the $\mu 3A$ and $\mu 3B$ family of sequences has resulted from a gene duplication event that probably occurred just prior to the evolution of the mammals. In a similar manner, the $\mu 1A$ and $\mu 1B$ family of sequences has also resulted from another independent gene duplication event, probably occurring at a similar time.

DISCUSSION

The present paper sheds light on several aspects of the mammalian AK and $\mu 3A$ genes, which are closely linked in mammalian species. In human, rat, mouse and Chinese hamster, these genes are arranged in a head-to-head fashion, sharing the first exon and with their transcription start sites separated by less than 200 bp. We have presented evidence that a 315 bp fragment between these two genes that partially overlaps with the transcribed region of the $\mu 3A$ gene has the ability effectively to carry out transcription in both directions. Any deletion within this region affects transcription in both orientations, indicating strongly that it serves as a bidirectional promoter for both genes. This sequence region that functions as a bidirectional promoter lies within a CpG island, which are commonly present in the promoter regions of many

housekeeping genes [23,39]. Similar to other CpG promoters, this region lacks a TATA box and initiates transcription at a number of closely related sites (for the AK gene) [28–30]. The arrangement of genes in a bidirectional divergent fashion, as seen here for the AK and $\mu 3A$ genes, appears to be a common organizational motif in mammalian genomes. In the human genome a substantial fraction (15–30%) of the housekeeping genes, controlled by CpG island promoters, are arranged in this manner [26]. One possible utility of a shared promoter is to regulate the expression of the two genes co-ordinately. However, the possible physiological significance of this for the AK and $\mu 3A$ genes, which carry out apparently unrelated functions, is presently unclear. Nevertheless, it is of interest to determine whether expression of these two genes is co-ordinately regulated under different conditions.

The present work also clarifies the gene structures, including intron–exon organization, of the $\mu 3A$ and AK genes. The $\mu 3A$ gene in various mammalian species is shown to be made up of nine exons and eight introns, with the introns located in the same positions in all species studied. The mean lengths of the exons and introns in the human $\mu 3A$ gene were 170 bp and 3215 bp respectively, close to the average values seen for the human genome [32,40]. The human $\mu 3B$ gene has a very similar intron–exon organization to $\mu 3A$, and the only difference is in the position of the first intron. In contrast with mammalian species, the numbers and positions of the introns in the $\mu 3$ gene in other eukaryotic species are not conserved. In comparison with the $\mu 3A$ gene (27 kb), the AK gene in humans is very large (558 kb). While the exons are of similar lengths in the two genes, the introns within the AK gene are unusually large, with an average length

of 55.7 kb. Dystrophin is one of the largest genes found within the human genome, with a total length of 2400 kb [31,32]. The average length of introns in the dystrophin gene is approx. 30 kb, and the ratio of intron sequence to protein coding sequence (3685 amino acids; 11055 bp) is approx. 215. The AK gene, in comparison, has a protein coding sequence of only 1083 nt (361 amino acids). This translates into a ratio of non-coding to coding sequence of > 550, which to our knowledge is the highest reported for any gene [31,32].

McNally et al. [9] reported the existence of two different forms of AK gene transcripts in human cells and tissues. The protein products encoded by the two transcripts (known as AK-long and AK-short) differ only at their N-termini. The long form of AK is similar to that seen in other species (including Chinese hamster, mouse and rat), and it differs from AK-short in the first 21 amino acids at the N-terminal end. In AK-short, the first 21 amino acids are replaced with 4 amino acids not found in the long form. The two transcripts also differ in terms of their upstream non-translated sequences, but their encoded sequences beyond the first 4 amino acids in the short form are identical [9]. The two forms of AK exhibit similar catalytic activity, which is supported by the results of deletion studies where removal of the first 21 amino acids from the long form of AK had minimal effect on its catalytic activity [12]. McNally et al. [9] suggested that the two AK gene transcripts arise by differential splicing of a single transcript. However, at that time no information was available on the AK gene or its promoter sequence. The promoter region for AK that we have now identified corresponds to the AK-long transcripts. The first exon in the identified AK gene, in addition to containing the promoter sequence, contains all of the upstream non-coding sequence found in the AK-long form, as well as coding sequence for the first 21 amino acids that are distinctive for this form (Figure 2). The observed characteristics of this exon indicate that it cannot account for or transcribe the AK-short form. To understand the possible origin of the AK-short transcript, we have performed BLAST searches of the human genome with the upstream sequence found in the AK-short form. The results of these studies reveal that a 200 bp fragment containing the upstream non-coding sequence as well as the coding sequence for the first 4 amino acids of the AK-short form matches perfectly with a sequence in human chromosome 10 located between positions 24487616 and 24487815 (results not shown). This sequence lies within the first intron described in the present work (see Table 2). This observation strongly suggests the existence of a separate promoter for the AK gene, which is responsible for the synthesis of the AK-short transcript. The two AK gene transcripts found in human cells thus are likely to result from splicing of the independently transcribed first exons with the remainder of the AK exons, which are shared by both transcripts. The nature of the second AK promoter, and whether it is a unique characteristic of human cells, is presently unclear and will be examined in future studies. Earlier biochemical studies have reported large variations (38–56 kDa) in the estimated molecular mass of AK [41–43]. Although the exact basis of this variation remains to be determined, molecular sequence data indicate that much of this variation is likely have been due to technical reasons.

The unusually large introns within the AK gene presumably are responsible for the high mutation frequency at this gene locus in mammalian cells. In CHO cells, stable mutants of AK are obtained at a spontaneous frequency of between 10^{-3} and 10^{-4} , which is 100–1000-fold higher than that for other gene loci in CHO and other mammalian cells [17,18,44]. Many of these mutants contain large deletions within the AK gene, leading to loss of several of its introns and exons [13]. We have now presented evidence that, in a number of these AK⁻ mutants, the deletions that begin within

the AK gene extend upstream and lead to the loss of all exons and introns corresponding to the μ 3A gene. AK and μ 3A are neighbouring genes on chromosome 10 and they are flanked by the VCL (vinculin) gene on the μ 3A side and by MORF (monocytic leukaemia Zn²⁺ finger protein-related factor) on the AK side [22]. It will be of interest to determine whether the deletions in some of the AK⁻ mutants also extend into other neighbouring genes.

The AP-3 complex, of which μ 3A is a part, is involved in the vesicle-mediated transport of proteins from the *trans*-Golgi network to lysosomes and related organelles [1–4]. Much interest in AP-3 complex function has been generated by the finding that a defect in the β 3A subunit of the AP-3 complex is responsible for type 2 Hermansky–Pudlak syndrome [3,4,14]. This group of genetic disorders is characterized by the abnormal formation and functioning of lysosome-related vesicles, such as melanosomes and platelet dense granules, leading to a number of problems, including hypopigmentation, prolonged bleeding and pulmonary fibrosis [3,4]. Two mouse models of Hermansky–Pudlak syndrome (mocha and pearl) contain mutations in the δ and β 3A subunits of the AP-3 complex and exhibit similar phenotypic abnormalities. In *Drosophila*, a number of different eye colour mutants are affected in the genes encoding different subunits of the AP-3 complex [1,2,4]. Of these, the *carmine* mutants involve defects in the μ 3 subunit. Mammalian cells that are deficient in the β 3A subunit show increased trafficking of certain lysosomal membrane proteins via the plasma membrane [14]. In yeast, mutants lacking AP-3 subunits show altered sorting of certain vacuolar proteins (alkaline phosphatase and Vam3p), without an effect on that of carboxypeptidase Y [1,2]. All of these observations support the view that the AP-3 complex and its component subunits are involved in the sorting of proteins from the *trans*-Golgi network to lysosomes and related organelles. As indicated above, two different μ 3 protein homologues are found in mammalian cells. Of these, μ 3A (known previously as p47A) is expressed ubiquitously, while μ 3B (p47B) expression is seen only in neurons and neuroendocrine cells [15]. Thus far, no mammalian cells containing mutant μ 3A have been described. In this context, the mutants described in the present work which involve deletion of the μ 3A subunit are of interest. These mutants grow much more slowly in comparison with the WT CHO cells, but their biochemical or cell biological phenotype has not yet been examined. Future studies with these mutants should prove helpful in elucidating the function of this subunit in protein sorting and cell function.

Phylogenetic analysis of μ protein homologues provides strong evidence that the four adaptor families (AP-1, AP-2, AP-3 and AP-4), although related, are phylogenetically distinct from each other. Therefore they are likely to have diverged from each other by gene duplications at early stages in eukaryotic cell evolution. Of these, the AP-1 and AP-2 families of homologues are most closely related, whereas those belonging to the AP-4 family are most divergent. Although, based on our analysis, it is difficult to confidently infer which of these complexes evolved first, certain observations are suggestive that the AP-2 family may have evolved before the others. In *G. lamblia*, which is one of the earliest branching eukaryotic species [45,46], the only μ homologue found corresponds to the μ 2 family, suggesting that the AP-2 adaptor family evolved very early in evolution. If one roots the phylogenetic tree shown in Figure 5 at the midpoint, then the root again lies close to the AP-2 or AP-4 family of proteins. Of these, the AP-4 complexes show a very narrow distribution in eukaryotic organisms, indicating that they appeared at a later stage in evolution. Thus it is likely that the root of the tree lies in the AP-2 family of proteins, and that the other μ protein families (and by inference the AP complexes) have evolved from it by means

of gene duplications and divergence. Later gene duplications within the AP-1 and AP-3 families are indicated to have given rise to the $\mu 1A$ and $\mu 1B$ (or $\mu 3A$ and $\mu 3B$) homologues. These latter gene duplications are likely to have occurred before the evolution of the mammals, as different $\mu 1$ and $\mu 3$ homologues are not found in earlier branching eukaryotic species. One of these gene duplication events presumably also resulted in the head-to-head linkage of the $\mu 3A$ and AK genes in mammalian species.

This work was supported by research grant T-4688 from the Heart and Stroke Foundation of Ontario.

REFERENCES

- Robinson, M. S. and Bonifacino, J. S. (2001) Adaptor-related proteins. *Curr. Opin. Cell Biol.* **13**, 444–453
- Odorizzi, G., Cowles, C. R. and Emr, S. D. (1998) The AP-3 complex: a coat of many colours. *Trends Cell Biol.* **8**, 282–288
- Shotelersuk, V. and Gahl, W. A. (1998) Hermansky–Pudlak syndrome: models for intracellular vesicle formation. *Mol. Genet. Metab.* **65**, 85–96
- Starcevic, M., Nazarian, R. and Dell'Angelica, E. C. (2002) The molecular machinery for the biogenesis of lysosome-related organelles: lessons from Hermansky–Pudlak syndrome. *Semin. Cell Dev. Biol.* **13**, 271–278
- Berne, R. M. (1993) Adenosine – a cardioprotective and therapeutic agent. *Cardiovasc. Res.* **27**, 2
- Manfredi, J. P. and Holmes, E. W. (1985) Purine salvage pathways in myocardium. *Annu. Rev. Physiol.* **47**, 691–705
- Singh, B., Hao, W., Wu, Z., Eigl, B. and Gupta, R. S. (1996) Cloning and characterization of cDNA for adenosine kinase from mammalian (Chinese hamster, mouse, human and rat) species. High frequency mutants of Chinese hamster ovary cells involve structural alterations in the gene. *Eur. J. Biochem.* **241**, 564–571
- Spychala, J., Datta, N. S., Takabayashi, K., Datta, M., Fox, I. H., Gribbin, T. and Mitchell, B. S. (1996) Cloning of human adenosine kinase cDNA: sequence similarity to microbial ribokinases and fructokinases. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1232–1237
- McNally, T., Helfrich, R. J., Cowart, M., Dorwin, S. A., Meuth, J. L., Idler, K. B., Klute, K. A., Simmer, R. L., Kowaluk, E. A. and Halbert, D. N. (1997) Cloning and expression of the adenosine kinase gene from rat and human tissues. *Biochem. Biophys. Res. Commun.* **231**, 645–650
- Moffatt, B. A., Wang, L., Allen, M. S., Stevens, Y. Y., Qin, W., Snider, J. and von Schwarzenberg, K. (2000) Adenosine kinase of Arabidopsis. Kinetic properties and gene expression. *Plant Physiol.* **124**, 1775–1785
- Maj, M. C., Singh, B. and Gupta, R. S. (2002) Pentavalent ions dependency is a conserved property of adenosine kinase from diverse sources: identification of a novel motif implicated in phosphate and magnesium ion binding and substrate inhibition. *Biochemistry* **41**, 4059–4069
- Maj, M. C., Singh, B. and Gupta, R. S. (2000) Structure-activity studies on mammalian adenosine kinase. *Biochem. Biophys. Res. Commun.* **275**, 386–393
- Singh, B., Lin, A., Wu, Z. C. and Gupta, R. S. (2001) Gene structure for adenosine kinase in Chinese hamster and human: high-frequency mutants of CHO cells involve deletions of several introns and exons. *DNA Cell Biol.* **20**, 53–65
- Dell'Angelica, E. C., Shotelersuk, V., Aguilar, R. C., Gahl, W. A. and Bonifacino, J. S. (1999) Altered trafficking of lysosomal proteins in Hermansky–Pudlak syndrome due to mutations in the beta 3A subunit of the AP-3 adaptor. *Mol. Cell* **3**, 11–21
- Pevsner, J., Volkandt, W., Wong, B. R. and Scheller, R. H. (1994) Two rat homologs of clathrin-associated adaptor proteins. *Gene* **146**, 279–283
- Gupta, R. S. and Dudani, A. K. (1987) Species-specific differences in the toxicity of rhodamine 123 towards cultured mammalian cells. *J. Cell. Physiol.* **130**, 321–327
- Gupta, R. S. and Siminovich, L. (1978) Genetic and biochemical studies with the adenosine analogs toycamycin and tubercidin: mutation at the adenosine kinase locus in Chinese hamster cells. *Somatic Cell Genet.* **4**, 715–735
- Gupta, R. S. and Singh, B. (1983) Quantitative mutagenesis at the adenosine kinase locus in Chinese hamster ovary cells. Development and characteristics of the selection system. *Mutat. Res.* **113**, 441–454
- Gupta, R. S. and Singh, B. (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Curr. Biol.* **4**, 1104–1114
- Gupta, R. S. (1995) Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. *Mol. Biol. Evol.* **12**, 1063–1073
- Felsenstein, J. (1994) PHYLIP, version 3.5, University of Washington, Seattle, WA
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome. *Nature (London)* **409**, 860–921
- Antequera, F. and Bird, A. (1999) CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr. Biol.* **9**, R661–R667
- Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11995–11999
- Takai, D. and Jones, P. A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3740–3745
- Adachi, N. and Lieber, M. R. (2002) Bidirectional gene organization: a common architectural feature of the human genome. *Cell* **109**, 807–809
- Brenner, V., Nyakatura, G., Rosenthal, A. and Platzer, M. (1997) Genomic organization of two novel genes on human Xq28: compact head to head arrangement of IDH gamma and TRAP delta is conserved in rat and mouse. *Genomics* **44**, 8–14
- Phan Dinh, T. F., Porteu, A., Kahn, A. and Skala, H. (2001) Bidirectional activity and orientation-dependent specificity of the rat aldolase C promoter in transgenic mice. *FEBS Lett.* **499**, 143–146
- Liao, W. C., Ash, J. and Johnson, L. F. (1994) Bidirectional promoter of the mouse thymidylate synthase gene. *Nucleic Acids Res.* **22**, 4044–4049
- Ame, J. C., Schreiber, V., Fraulob, V., Dolle, P., de Murcia, G. and Niedergang, C. P. (2001) A bidirectional promoter connects the poly(ADP-ribose) polymerase 2 (PARP-2) gene to the gene for RNase P RNA. Structure and expression of the mouse PARP-2 gene. *J. Biol. Chem.* **276**, 11092–11099
- Tennyson, C. N., Klamut, H. J. and Worton, R. G. (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* **9**, 184–190
- Strachan, T. and Read, A. P. (1999) Human Molecular Genetics, Wiley-Liss, New York
- Lindberg, B., Klenow, H. and Hansen, K. (1967) Some properties of partially purified mammalian adenosine kinase. *J. Biol. Chem.* **242**, 350–356
- Miller, R. L., Adamczyk, D. L., Miller, W. H., Koszalka, G. W., Rideout, J. L., Beacham, III, L. M., Chao, E. Y., Haggerty, J. J., Krenitsky, T. A. and Elion, G. B. (1979) Adenosine kinase from rabbit liver. II. Substrate and inhibitor specificity. *J. Biol. Chem.* **254**, 2346–2352
- Bennett, Jr, L. L., Schnebli, H. P., Vail, M. H., Allan, P. W. and Montgomery, J. A. (1966) Purine ribonucleoside kinase activity and resistance to some analogs of adenosine. *Mol. Pharmacol.* **2**, 432–443
- Gupta, R. S. (1989) Purine nucleoside analogs. In *Drug Resistance in Mammalian Cells*, vol. 1: Antimetabolite and Cytotoxic Analogs (Gupta, R. S., ed.), pp. 89–110, CRC Press, Boca Raton
- Gupta, R. S. and Mehta, K. D. (1986) Genetic and biochemical characteristics of three different types of mutants of mammalian cells affected in adenosine kinase. *Adv. Exp. Med. Biol.* **195**, 595–603
- Felsenstein, J. (1985) Confidence limits in phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791
- Cross, S. H. and Bird, A. P. (1995) CpG islands and genes. *Curr. Opin. Genet. Dev.* **5**, 309–314
- Deutsch, M. and Long, M. (1999) Intron–exon structures of eukaryotic model organisms. *Nucleic Acids Res.* **27**, 3219–3228
- Juranka, P. and Chan, V. L. (1985) Analysis of adenosine kinase mutants of baby hamster kidney cells using affinity-purified antibody. *J. Biol. Chem.* **260**, 7738–7743
- Chang, C. H., Brockman, R. W. and Bennett, Jr, L. L. (1980) Adenosine kinase from L1210 cells. Purification and some properties of the enzyme. *J. Biol. Chem.* **255**, 2366–2371
- Andres, C. M. and Fox, I. H. (1979) Purification and properties of human placental adenosine kinase. *J. Biol. Chem.* **254**, 11388–11393
- Simon, A. E., Taylor, M. W., Bradley, W. E. and Thompson, L. H. (1982) Model involving gene inactivation in the generation of autosomal recessive mutants in mammalian cells in culture. *Mol. Cell. Biol.* **2**, 1126–1133
- Gupta, R. S., Aitken, K., Falah, M. and Singh, B. (1994) Cloning of Giardia lamblia heat shock protein HSP70 homologs: implications regarding origin of eukaryotic cells and of endoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2895–2899
- Kabnick, K. S. and Peattie, D. A. (1991) Giardia: A missing link between prokaryotes and eukaryotes. *Am. Sci.* **79**, 34–43

Received 11 August 2003/7 October 2003; accepted 24 October 2003

Published as BJ Immediate Publication 24 October 2003, DOI 10.1042/BJ20031219