Dual Evolutionary Modes in the Bovine Globin Locus†

Amy M. Brunner, John C. Schimenti,‡ and Craig H. Duncan*

Division of Basic Science, Children's Hospital Research Foundation, Cincinnati, Ohio 45229

Received January 29, 1986; Revised Manuscript Received March 19, 1986

ABSTRACT: Five bovine globin pseudogenes were subjected to sequence analysis. These genes include the three pseudogenes in the β-type globin gene cluster as well as two allelic forms. Comparison of the sequences with those of the adult and fetal bovine globin genes shows that together they form a multigene family that was created by large-scale duplication. The structures are explained by invoking sequence exchange mediated by gene conversion. After their creation these genes evolved in a concerted fashion, exchanging sequence freely by intrachromosomal gene conversion. Subsequently, one by one, the genes were uncoupled from this exchange. This was accomplished by the creation of nonhomologies that formed barriers to gene conversion. These nonhomologies were several hundred bases in length and were formed by either deletion or by insertion of short repetitive sequences within the gene structures. In this way the genes made the transition from a rapid, coupled mode to a slow, solitary mode of evolution. Allelic gene polymorphisms were distributed inhomogeneously in the bovine globin family. It is proposed that this was due to interruption of interchromosomal gene conversion by a recent pseudogene duplication in the fetal globin gene cluster.

The recent surge of structural information about genes has presented an opportunity to understand the complex relations between changes in DNA structure and fitness of the organism. These studies have shown that point by point mutation of DNA sequence is not the only source of evolutionary flux. Two larger scale processes of genomic alteration also play central roles. These mechanisms operate on the point by point variation by establishing multiple targets on which it can act and by combining variations that originally arose in separate genes. The result is that spontaneous variations are channeled into the creation of novel genes.

One of these processes is DNA duplication. Although the mechanism is unknown, large stretches of DNA sequence, up to 40 kb in length (Townes et al., 1984), give rise to copies of themselves. Repeated copying generates families of genes lying in a tandem arrangement on the chromosome. The present-day arrangements of globin genes (Goodman et al., 1984; Hardison, 1984; Hardies et al., 1984), histocompatability genes (Steinmetz et al., 1982), and immunoglobulin genes (Honjo, 1983; Seidman et al., 1978) all arose in this manner.

Genes situated in such families can exchange sequence information by virtue of a second process, intrachromosomal gene conversion (Baltimore, 1981; Kourilsky, 1983). Again, the mechanism is obscure, but a single conversion event can transmit an extended sequence of DNA from one gene to another member of the same family (Schulze et al., 1983; Weiss et al., 1983; Ollo & Rougeon, 1983; Clarke & Rudikoff, 1984). In effect, rare spontaneous variations are united in novel combinations that would not have occurred in the absence of gene conversion. Among known mammalian gene families, histocompatability genes are the epitome of this process. Essentially all of the diversity in these highly polymorphic loci arises through intrachromosomal information exchange among the 30-odd members of this family (Coligan, 1984; Lalanne et al., 1982; Pease et al., 1983).

Globin genes also exchange information via gene conversion but in a more limited way. In primate β globin gene loci, the duplicated homologous fetal globin genes undergo frequent conversion with each other but not with adjacent adult globin genes or pseudogenes (Slightom et al., 1980; Scott et al., 1984; Slightom et al., 1985). Primate α globin gene loci have also undergone gene conversion (Liebhaber et al., 1980; Liebhaber & Begley, 1983). In ruminant globin gene loci, some pairs of genes interact via gene conversion (Schimenti & Duncan, 1985a; Schon et al., 1982), while other pairs do not participate (Schimenti & Duncan, 1984). In order to better understand these differences in recombinational behavior, we have analyzed the DNA sequence of all the bovine (*Bos taurus*) globin pseudogenes (ψ) and compared them to the adult (β) and fetal (γ) globin genes of this species. The results suggest some principles of general interest to the evolution of gene structure.

Materials and Methods

DNA. The pseudogenes described in this paper were isolated from bovine cosmid and bacteriophage libraries constructed in this laboratory. The methods of construction, clone isolation, and mapping are presented in a previous report (Schimenti & Duncan, 1985b). Pseudogenes ψ^1 and ψ^2 were isolated from cosmid clone 19, ψ^{1A} from phage clone 11, ψ^{2A} from cosmid clone 10, and ψ^3 from phage clone 7. DNAs to be sequenced were cut with appropriate restriction enzymes, subcloned in the M13 vectors mp18 and mp19, and propagated on the host strain JM105 (Norrander et al., 1983). Both strands of the DNA were determined, as well as overlapping sequences at restriction enzyme joints. The sequences were analyzed with the computer programs developed by Fristensky et al. (1982). Dot matrix analyses were done with the DNA Inspector II programs developed by Textco Inc., West Lebanon, NH (Gross, 1986).

A modification (Duncan, 1985) of the dideoxy chain termination method (Sanger et al., 1977) was used to generate the sequence ladders. ssDNA template (100–200 ng), 20 μCi of [α-^{32}P]dATP (3000 Ci/mmol), 50 pmol of dGTP,[1] and 50

† This work was supported by the following grants: March of Dimes Birth Defects Foundation Grant 1-961 (to C.H.D.) and National Institutes of Health Grant HL15996-11 to the Comprehensive Sickle Cell Center of Cincinnati.

* Correspondence should be addressed to this author.

‡ Present address: Department of Molecular Biology, Princeton University, Princeton, NJ.

0006-2960/86/0425-5028$01.50/0

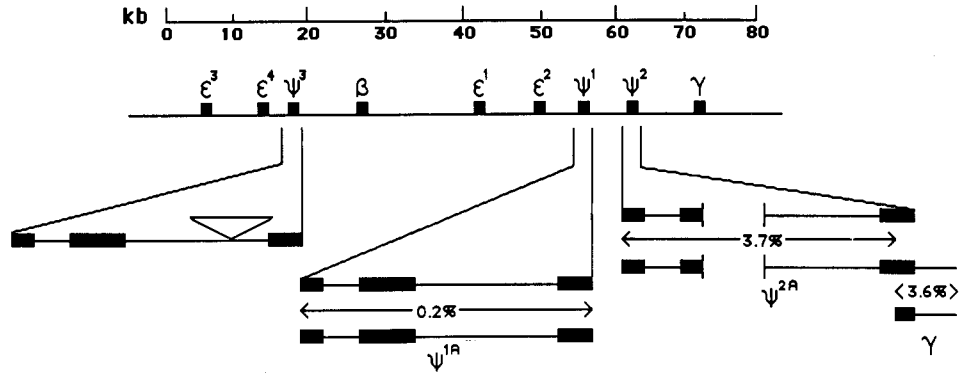

FIGURE 1: Linkage map of the bovine β globin locus, displaying structures of the pseudogenes. Percent divergence between pseudogene alleles and the 3′ regions of ψ^{2A} and the γ gene is presented. Regions homologous to coding sequences are indicated by thicker lines.

pmol of dTTP were lyophilized to dryness and dissolved in 2.5 μL of RT buffer containing 1 pmol of a synthetic primer d(GTTTTCCCAGTCACGAC). RT buffer was 50 mM NaCl, 34 mM Tris (pH 8.3, 42 °C), and 6 mM $MgCl_2$. After 10-min incubation at 42 °C, 0.5 μL of RT buffer containing 0.6 unit of reverse transcriptase (diluted on the day of use from a preparation of 20 units/μL sold by Life Sciences, Inc., St. Petersburg, FL) was added and the reaction incubated at 42 °C for 5 min. An additional 2 μL (2.4 units) of the diluted reverse transcriptase was added, and 1 μL of the resulting mixture was dispensed to each of four tubes containing 1.5 μL of N quasi buffers, where N can be either G, A, T, or C. The N quasi buffers were RT buffer containing 250 μM dGTP, 250 μM dATP, 250 μM dTTP, 250 μM dCTP, and 12.5 μM ddNTP. After 10-min incubation at 42 °C, the reactions were stopped by adding 6 μL of 90% deionized formamide, 10 mM EDTA, 0.05% bromphenol blue, and 0.1% xylene cyanol to each tube. The reaction mixes were then heated at 90 °C for 2 min and chilled on ice, and 3 μL was loaded onto each of two 100 cm long 5% polyacrylamide DNA sequencing gels.

The gels were made and electrophoresis was performed basically as described by Smith and Calvo (1980), with equipment purchased from American BioNuclear, Emeryville, CA. The first gel was run at 2300 V for 6–7 h, while the second gel was run at 2300 V for 36 h. The gels were transferred in two pieces to Whatman 3MM filter paper and dried under vacuum for 15 min at 80 °C on a 35 × 44 cm slab gel dryer (Hoefer Scientific Instruments). The dried gels were autoradiographed in 35 × 43 cm film cassettes with a sheet of XAR-2 film (Eastman Kodak) at room temperature for 24–48 h. DNA sequence was routinely read in the region from 20 to 600 bases adjacent to the primer site.

RESULTS

Structure of Pseudogenes. The cow's β-type globin locus contains three pseudogenes (Figure 1). These genes were derived from two duplication events—one large duplication, which copied a four gene set, followed by a localized duplication, which created a second pseudogene in the fetal cluster (Schimenti & Duncan, 1985b). The sequences (Figures 2–4) all contain defects that render them nonfunctional. ψ^1 has point mutations in splicing consensus sequences of both introns. ψ^2 has a portion of its second exon deleted. ψ^3, as previously discussed (Schimenti & Duncan, 1985b), has a frameshift mutation in its first exon. In this paper, the words intron, exon, and coding sequence will be used to denote regions of the pseudogene sequences that are homologous to introns, exons, and coding sequences in functional globin genes.

Table I: Percent Divergence Comparisons of Bovine Globin Gene and Pseudogene Sequences[a]

	β	γ	ψ^1	ψ^{1A}	ψ^2	ψ^{2A}	ψ^3
β		10.6	18.6	18.7	21.7	20.9	24.5
γ	7.5		20.5	20.3	24.0	21.4	23.1
ψ^1	15.5	16.2		0.2	21.8	22.7	24.3
ψ^{1A}	15.5	16.2	0.0		21.9	22.9	24.3
ψ^2	–	–	–	–		3.7[b]	15.3
ψ^{2A}	9.6	3.6	17.3	17.3	–		16.1
ψ^3	–	–	–	–	18.1	–	

[a] Percent divergence for structural gene (upper right) and 3′ region (lower left) comparisons are presented. The 3′ region includes 75 bp of the third coding region and extends into the flanking region for a total of 415 bp, except in the case of ψ^2 and ψ^3 where only 282 bp of the 3′ region have been sequenced. Insertions in β, γ, and ψ^3 were removed before comparison. Gaps were introduced to provide maximum overlap and were counted as one difference. Dashes indicate greater than 30% divergence. [b] A 75-bp region at the 3′ end was highly divergent relative to the rest of the gene (see text) and was excluded from the comparison. When this region was included, divergence was 6.1%.

These pseudogenes share extensive homology with each other and with the cow β globin and γ globin DNA sequences (Table I). The homology is not confined to coding sequences but extends also over introns and flanking sequences. To illustrate the sequence homologies between these genes, we performed dot matrix analyses on all pairwise combinations of the cow globin genes and also between the cow genes and the human δ globin gene. Every comparison led to similar conclusions, which are illustrated by the examples shown (Figure 5). When the cow ψ^1 gene was compared with its human δ globin counterpart, the dot matrix analysis shows that only the coding regions are conserved—no homology is retained in the introns. In contrast, comparison of the cow ψ^1 with the cow β gene sequence shows homology throughout the introns and coding sequences, which is interrupted in a discontinuous fashion by a block of nonhomology in the 3′ region of the second intron (Figure 5). This result is typical of pairwise comparisons between the cow pseudogenes, the β gene, and the γ gene. A survey of such analyses shows (Figure 6) that ψ^2 has suffered a deletion in the 5′ region of the gene, while ψ^3 has received an insertion in the 3′ region. The goat (*Capra hircus*) orthologue of ψ^3, $\Psi\beta^Z$, does not contain this insertion (Cleary et al., 1981). This insertion is flanked by 11-bp direct repeats and is a repetitive element that appears in at least seven

[1] Abbreviations: dGTP, 2′-deoxyguanosine 5′-triphosphate; dTTP, thymidine 5′-triphosphate; dATP, 2′-deoxyadenosine 5′-triphosphate; dCTP, 2′-deoxycytidine 5′-triphosphate; ddNTP, dideoxynucleotide triphosphate; ddGTP, 2′,3′-dideoxyguanosine 5′-triphosphate; ddATP, 2′,3′-dideoxyadenosine 5′-triphosphate; ddTTP, 2′,3′-dideoxythymidine 5′-triphosphate; ddCTP, 2′,3′-dideoxycytidine 5′-triphosphate; EDTA, ethylenediaminetetraacetic acid, pH 8.0, with NaOH; Tris, tris(hydroxymethyl)aminomethane; bp, base pair.

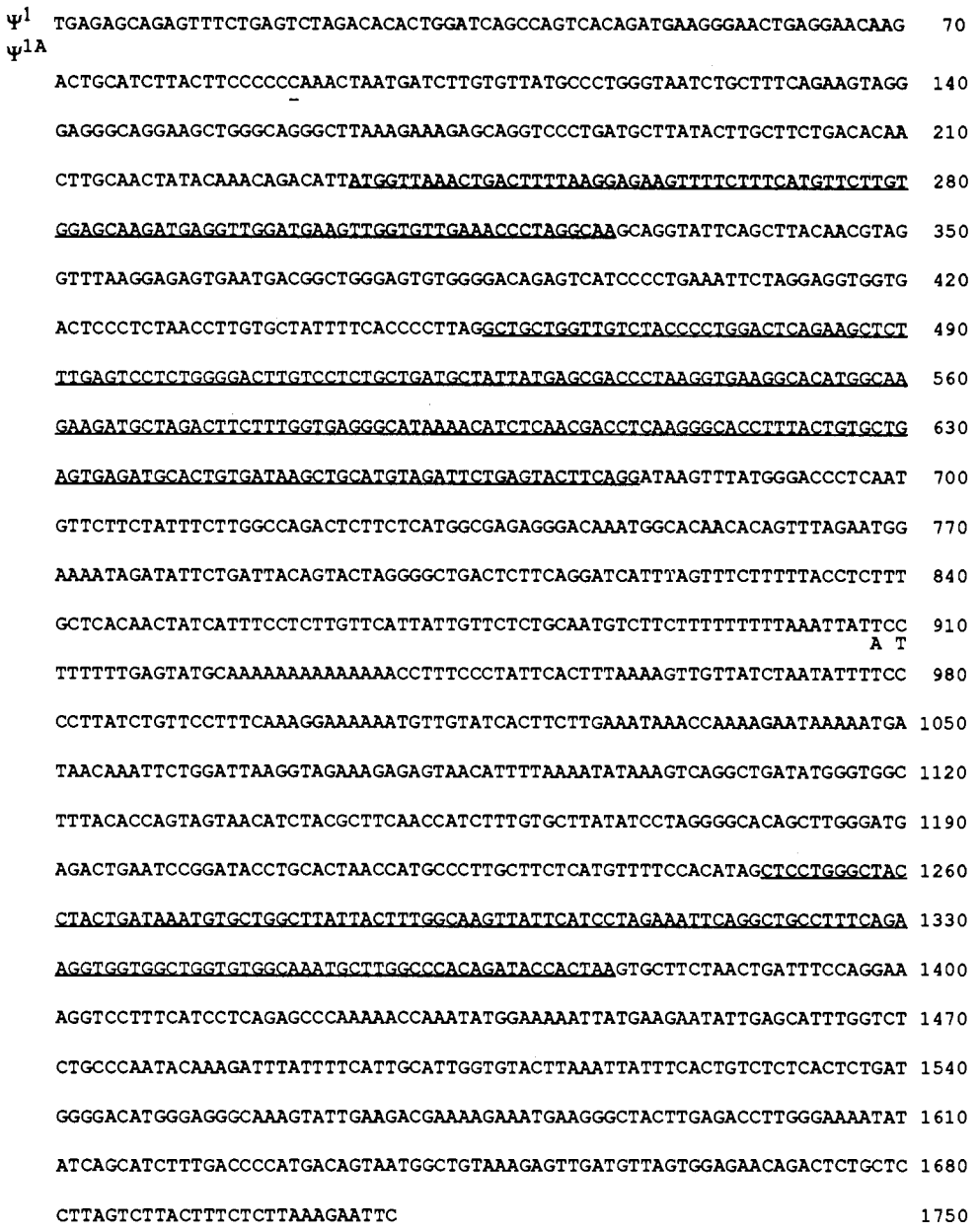

FIGURE 2: Nucleotide sequence comparison of alleles ψ^1 and ψ^{1A}. The sequence of ψ^1 is displayed. Positions where ψ^{1A} differs from ψ^1 are indicated. Regions homologous to coding sequences are underlined.

other locations in the cow globin locus. It is distinct from previously described ruminant *Alu*-type DNA (Watanabe et al., 1982; Schon et al., 1981; Schimenti & Duncan, 1984; Spence et al., 1985) and will be described further in a future report.

Our previous work (Schimenti & Duncan, 1985b) indicated that there was a concentration of allelic restriction fragment length polymorphisms in the ψ^2 region as compared to the ψ^1 region, implying an uneven distribution of sequence divergence in this region. Comparison of the allelic pseudogene sequences (Figures 2 and 3) supports this conclusion. ψ^1 and ψ^{1A} are only 0.2% divergent, compared to 3.7% divergence for ψ^2 and ψ^{2A}. Differences between the sequences of ψ^2 and ψ^{2A} are evenly distributed until the middle of the third coding region (Figure 3), but the last 75 bp of this region are highly divergent. In addition, the 3′ untranslated and 3′ flanking regions of these alleles are not homologous, while the sequences 3′ to ψ^1 and ψ^{1A} are identical. Pairwise comparisons (Table I) between the 3′ regions of the bovine globin genes showed high homology between ψ^{2A} and the γ gene. The 3′ sequences of ψ^{2A} and the γ gene are 6-fold more homologous than the remainder of these genes. At the point where homology between ψ^2 and ψ^{2A} begins to decrease, homology between ψ^{2A} and the γ gene increases. This can be explained by presuming an intrachromosomal gene conversion that substituted a >200-bp patch of γ gene sequence in the 3′ region of the ψ^{2A} gene. Similar events that transferred several hundred base-pair regions of sequence have been reported in the mouse histocompatability locus (Devlin et al., 1985) and in the silk moth chorion gene family (Eickbush & Burke, 1985).

Only a single form of the ψ^3 sequence is presented, because our attempts to find the other allele of ψ^3 failed. We obtained the allelic forms of the ψ^1 and ψ^2 genes from the DNA of one heterozygous animal and so tried to obtain an allelic form of the ψ^3 sequence from this same animal. Accordingly, we tested five independent clones containing ψ^3 with a battery of restriction enzymes that recognize four base-pair sequences but were unable to detect any restriction site polymorphisms. Either all these clones represented only one allele or they included both alleles, but the nucleotide sequence differences

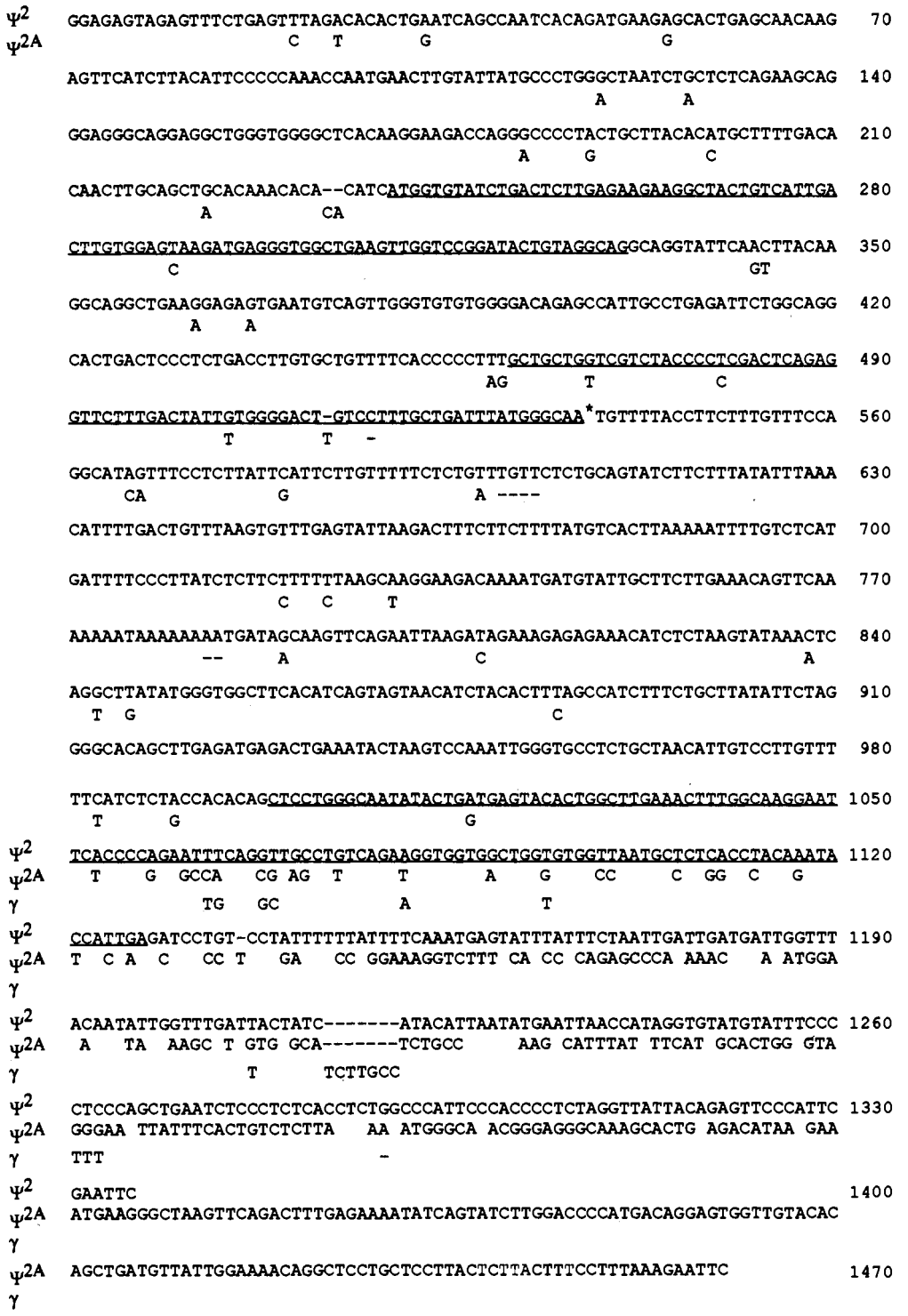

FIGURE 3: Nucleotide sequence comparison of alleles ψ^2 and ψ^{2A}. The sequence of ψ^2 is displayed. Positions where ψ^{2A} differs from ψ^2 are indicated. Regions homologous to coding sequences are underlined. An asterisk marks the site of a 292-bp deletion relative to the canonical globin gene. Differences between the 3′ regions of ψ^{2A} and the γ gene are presented for comparison.

were too few to cause a site polymorphism for the restriction enzymes we used. We believe that the latter explanation is right, because genomic blots performed with three additional six-base specific restriction enzymes also failed to show any restriction fragment length polymorphisms anywhere in the adult globin gene cluster (data not shown). It is likely that the ψ^3 gene, like the ψ^1 gene, has a low level of sequence variation between alleles.

Discussion

Insertions and Deletions in Ruminant Globin Genes Reveal an Evolutionary Process. Evolution is the sum of several processes that alter DNA sequence. A central component is sequence exchange between nonallelic genes. Our current concepts stem from diverse sources (Edelman & Gally, 1970; Hood et al., 1975; Slightom et al., 1980; Baltimore, 1981; Nagylaki & Petes, 1982; Ohta, 1982). Seidman et al. (1978) and Kourilsky (1983) recognized that these sequence exchanges create a fundamental dichotomy. They classified genes into two categories, which represent different modes of evolutionary change. One type is characterized by genes that are conserved over long periods of evolutionary time and maintain stable structures. Such genes are found once per genome or as highly divergent members of an ancient gene

```
Ψ3 GGAGAATAAAGTTTCTGAGTCTAGACACACTGGATCAGCCAATCACAGATGAAGGGCACTGAGGAACAGG   70
   AGTGCATCTTACATTCCCCCAAACCAATGAACTTGTATTATGCCCTGGGCTAATCTGCTCAGAGCAGAGA  140
   GGGCAGGGGGCTGGGTGGGGCTCACAAGCAAGACCAGGGCCCCTACTGCTTACACTTGCTTCTAACACAA  210
   CTTGCAACTGCACAAACACACATCATGGTGCATCTGACTCTTGAGGGGAAGGCTACTTGTCACTGCCCTG  280
   CAGACGAAAATGAGGGTGGCTGAAGTTGGTGTTGAAACCTTAGGCAGGCAGGTATTCAGCTTACAAGGCA  350
   AGGAGAGTGAATGTCAGCTGGGTGTGTGGGGACAGAGCCATTGCCTGGGATTCTGGCAGGCATTGACTCC  420
   CTCCTTCCTTATGCTGTTTTCACCCTGTAGGCTTCTGGTTGTCTACCCCTGGACTCAGAGGTTCTTTGAG  490
   TCCTTTGGGAACTTGCCCTCTGCTGATCTATTATGGGCAACCTAAGGTGAAGGCCCATGACAAGAAGGTG  560
   CTAGACTCCTTTACCAAAGGCCTGAAGCATGTTGACCACCTCAAGGGTGTCTTTGCTTTGCTAAGTGAGT  630
   TGCACTGTAAGAATCTGCATGTCAGTCCTGAGAACATCAGTGTGAGTCTACGGGATGCTTAATATTCTCC  700
   ATCTATTTTTTTTCTTCTGGTGGTTAAGTTCCTATCATGAGGAGAGAGTTAAGCAGCAGGATACAGTTCA  770
   GAATGGAAAAGAGATATTCTGGTTACATCACTATGGATTCCTCAGGAACATTTAGTTTCCTTTACTTTCT  840
   TTGTTCCCAGCCATCATTTCCTCTTACCCAATCTTGTTTTTTTCTGTTTGTTCTTTACAGTATCTTCTTT  910
   TTATTCAAACATTTTGAGTATTTAAAAACACTTTTATATTTTAAGTCACTTAAAATTTTATCTCATATTT  980
   TCCCCTTACCTCTTCCTTTCAAAGCAAGGGAGACAAAATGATGCATTGTGTCTTGAAATGGTTCAAAAGA 1050
   ATAAAAAATGATAACAGGCTATGGACTAAGACAGAAAGGCAGAAACATTTCTAAGAACAAGTTCAGGCTG 1120
   CTAT*CAATTCAGTTCAGTCACTCAGTCGTGTCCAACTCCTTTGCAACCCCATGAATCATAGTGCGCCAA 1190
   GCCTCCCTGTCCATCACCAACTCCCGACTCTAGAGGATCCCCGGAGTTCACTCAGACTCACGTCCATTGA 1260
   GTCAGTGATGCCATCCAGCCATCTCATCCTCTGTCATCCCCTTCTCCTCCTGCCCCCAATCCCTCCCAGC 1330
   ATCAGAGTCTTTTCCAATGAGTCAACTCTTCGCATGAGGTGGCCAAAGTATTGGAGTTTCAGCTTTAGCA 1400
   TCATTCCTTCCAAAGAAATCCCAGGGCTGATCTCCTTCAGAATGGACTGGCTGATAT*GGGTGGGTTCAT 1470
   ATTAGAAGTAACATCTATACTTCAGCCATCTTTCTACTTATATTCTAGGGGCACAGCTTGGGATGAGACT 1540
   GAAATACTCTTTAGTCTGAATTGGGTGCCTCTGCTAACCATGTCCTTGTTTTTTTATTCCTTCCACACAG 1610
   CTCCTTGGCAACATACTGGTAATTACACTGGCTCAAAACTTTGGCAAGGAATTCACCCCGGAGTTCCTGG 1680
   CTGCCTATCAGAAGGTGGTGGCTGGTGTGCTAATGCCCTCACCTAAAAATACCACTGGGATCCTGGCCAT 1750
   TTTCTTTAAAAAAGGAGAAAATTTATTTTTAATTGATTGATGATTGGTTTACAATACTGGTTTGATCACT 1820
   ATCACATATTAACAGGAATTAATCATAGGTGTACATATATCCCCTCCCATTTGAATCTCCCTCCCATCTC 1890
   CCACCAATTCACACTCCTCTAGGTTATTACAGAGCCCTATTTGAATTC                        1960
```

FIGURE 4: Nucleotide sequence of ψ^3. Regions homologous to coding sequences are underlined. Asterisks mark the boundaries of an insertion, whose sequence is shown in italics. Oligonucleotide repeats flanking the insertion are italicized and underlined.

family. Examples are the insulin gene, the serum albumin–α-fetoprotein pair, or the chicken ovalbumin gene family.

The other style of evolution is seen in large multigene families, whose members retain high sequence homology. In these cases, the emphasis is placed not on the variation of any individual gene but rather on the interactions between members of the gene family. The entire set of genes is evolving "in concert". Variations occurring in one member of the family are thought to be transmitted to other members at high frequency by the process of intrachromosomal gene conversion although, in mammals, rigorous evidence for the interaction of two genes on the same chromosome has been provided only in a few cases (Slightom et al., 1980; Liskay & Stachelek, 1983; Liebhaber et al., 1984). The multiple genes form many targets to accumulate sequence variation. These sequence variations can then be recombined via gene conversion to yield novel gene structures over short periods of evolutionary time. One prediction of this model is that gene duplication should be followed by accelerated divergence. This phenomenon has been reported for goat and sheep globin genes (Li & Gojobori, 1983).

This theory explains otherwise puzzling features of the cow globin locus. One of these is the remarkable homology between the introns of the pseudogenes and the β or γ genes (see Results). This homology is not the result of restricted divergence caused by recent creation of these pseudogenes. The precursor of these pseudogenes predates the radiation of placental mammals. Among its descendants are the δ globin gene in humans (Goodman et al., 1984), the $\psi\beta 2$ gene in rabbits (Hardison, 1984), and the $\psi\beta$h2–$\psi\beta$h3 pair in mice (Hardies

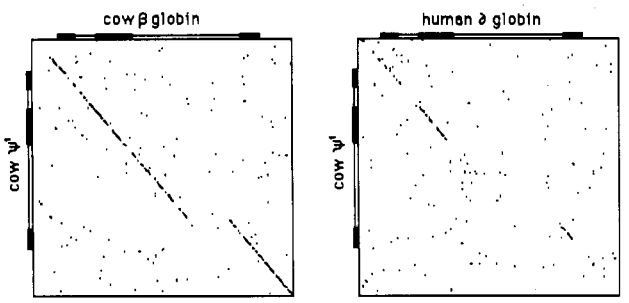

FIGURE 5: Dot matrix comparisons of globin genes. The cow ψ^1 sequence was compared to the cow β globin (Schimenti & Duncan, 1984) and human δ globin (Spritz et al., 1980) sequences. The gene structures are represented schematically on the top and left-hand boundaries of the figures with dark bars for coding sequences and hollow bars for introns. The matching criteria in this analysis were 10-base strings with a maximum of one mismatch per string.

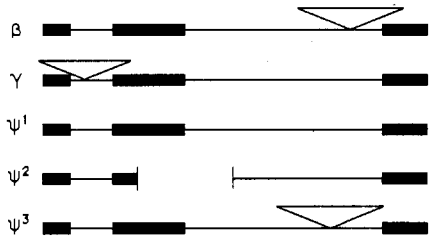

FIGURE 6: Structures of bovine globin genes and pseudogenes. Regions homologous to coding sequences are indicated by thicker lines. Locations of insertions and deletions are shown.

et al., 1984). The cow pseudogenes retain homology with these orthologous genes only within coding regions (see Results and Figure 5).

These findings can be accounted for by the rapid evolutionary mode. After the large duplication event, the cow globin genes and pseudogenes would be locked in the collective mode of evolution. Intrachromosomal gene conversion would impose sequence uniformity to create a homologous sequence throughout the introns. Because intron sequences are not under stringent selection, this shared intron sequence would be free to vary from that of globin genes in other mammalian orders. Protein coding sequences would also be subject to gene conversion, but in these regions the force of natural selection would maintain the homology between mammalian orders.

Alu-type transposable DNA elements have also played a role in the evolution of cow globin genes. It has been proposed that these elements form local blocks to intrachromosomal gene conversion (Hess et al., 1983; Michaelson & Orkin, 1983; Kourilsky, 1983). Data to support this proposal have come from studies of globin genes and intergenic DNA (Hess et al., 1983; Schimenti & Duncan, 1984) as well as from studies of γ-crystallin cDNAs (Bhat & Spector, 1984). In essence, random transposition of *Alu*-type DNA into one member of a duplicated DNA sequence creates a region of nonhomology between the duplicated regions. This area of nonhomology is thought to interfere with intrachromosomal gene conversion in its neighborhood by inhibiting heteroduplex formation. In this way, an *Alu*-type insertion could uncouple a gene from exchanging sequence with others in its gene family. In the general sense, *Alu*-type repeats could be considered as genomic catalysts, promoting transitions from the rapid, collective mode of evolution to the stable solitary mode.

The structures of the cow globin genes and pseudogenes provide a vivid example of these concepts. These genes are completing the transition from the collective evolutionary mode to the stable mode. We can reconstruct their evolutionary history as follows. After duplication of an original four-gene set (Schimenti & Duncan, 1985b), the duplicated β-type genes and the pseudogenes evolved in concert as a multigene family. The first gene to assume a differentiated structure was the present-day adult gene. We suggested (Schimenti & Duncan, 1984) that this gene's structure was then stabilized by an insertion. This event occurred before the divergence of cows and goats 20 million years ago (Schimenti & Duncan, 1984). Rapid evolution in the collective mode continued for the incipient fetal gene and the pseudogenes. As the incipient fetal globin gene developed a structure suited to its role, it became evolutionarily advantageous to uncouple its variation from that of the pseudogenes. Insertions and deletions have accumulated in such a way as to bring an end to the intrachromosomal recombinations of these genes. The result is the present-day structures of the cow globin genes and pseudogenes (Figure 6).

Sequence Polymorphism in the Duplicated Pseudogene Region. Allelic forms of a gene are also substrates for gene conversion events. In this case, the recombination occurs between sister chromosomes and it is called interchromosomal gene conversion. This process has been extensively analyzed in fungal systems, where it is possible to examine all of the products of a single meiosis (Szostak et al., 1983). It also occurs in the higher eukaryotes, where it is much more difficult to study. From an evolutionary standpoint, interchromosomal gene conversion is important because it can transmit and recombine variation among the gene pool of a species (Dover, 1982; Walsh, 1983; Lamb, 1985). Its function is analogous to that of intrachromosomal gene conversion in coupling the sequences of a multigene family.

The effects of interchromosomal gene conversion can be studied by structural comparison of allelic genes (Antonarkis et al., 1984). In the bovine globin locus, sequence polymorphisms are not randomly distributed. In our initial studies, this was indicated by the appearance of polymorphic restriction enzyme sites (Schimenti & Duncan, 1985b). Comparison of allelic DNA sequences (Table I) shows an 18-fold difference in polymorphism frequency between two adjacent pseudogenes; ψ^2 and ψ^{2A} are 3.7% divergent, compared to 0.2% divergence for the allelic neighbors ψ^1 and ψ^{1A}. Being pseudogenes, there is no obvious selective pressure for such a discrepancy.

Several other examples of clustered allelic sequence polymorphisms have been reported. Maeda et al. (1983) compared four alleles of a human globin intergenic region, as well as the corresponding segment of chimpanzee DNA. Sequence polymorphisms were clustered in a manner very unlikely to have arisen by chance. Clustered polymorphisms have also been observed in both human (Cohen et al., 1984) and mouse (Steinmetz et al., 1984) histocompatability gene loci. None of these clusters was confined to regions of DNA coding for proteins, and in no case was there an obvious explanation for these events.

In the case of the cow globin genes, there is a clue to the cause of the high-divergence region. The pseudogenes ψ^1 and ψ^2 were created by a localized DNA duplication (Schimenti & Duncan, 1985b). Because this duplication did not involve any functional gene, it was almost certainly a neutral event in the genome. Thus, the pseudogene duplication could have been carried for long evolutionary times heterozygous with the single pseudogene from which it was derived. During this time the duplicated element formed a region of nonhomology that might well have inhibited interchromosomal gene conversion in this region of the chromosome. This suppression could have caused accumulation of polymorphisms between allelic forms. Recently, during the genetic fixation events caused by human

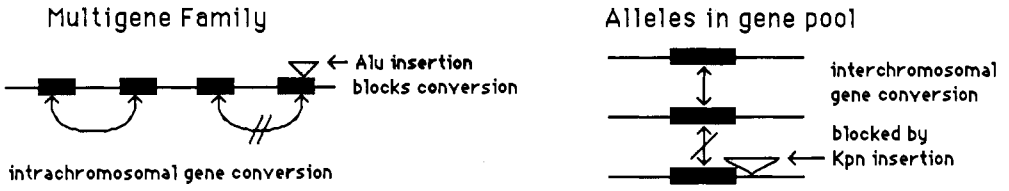

FIGURE 7: Proposed role of transposable DNA elements in modulating gene conversion.

efforts at breeding cows, two such divergent alleles may have been fixed to lead to the structures we have observed.

Just as it can be adaptive to block intrachromosomal gene conversion in a multigene family, sometimes it can surely be adaptive to shield certain alleles from interchromosomal gene conversion. Such a block would be beneficial to an allele that is undergoing a structural shift corresponding to a speciation event. What agent could promote this uncoupling? As for multigene families, an insertion could provide a region of nonhomology that would suppress the formation of heteroduplex DNA. Because they block intrachromosomal gene conversion, it is reasonable to suppose that *Alu*-type elements could also perform this function for interchromosomal gene conversion. Recent studies indicate otherwise. When two genes with overlapping deletions were introduced into cultured cells (Kucherlapati et al., 1984; Small & Scangos, 1983; Rubintz & Subramani, 1985; de Saint Vincent & Wahl, 1983) or incubated with extracts from cultured cells (Darby & Blattner, 1984; Kucherlapati et al., 1985), recombinants were generated in a process thought to represent interchromosomal gene conversion in vitro. A region of nonhomology the size of *Alu* elements did not inhibit recombination in these studies.

The idea that *Alu*-type repeats may block intrachromosomal but not interchromosomal gene conversion implies that these two processes have different mechanisms. Recent work indicates that this is indeed the case in yeast genetics (Klein, 1984; Klar & Strathern, 1984; Fink & Petes, 1984). If this is also true for mammalian systems, one can speculate that a larger region of nonhomology than an *Alu* element is required to block interchromosomal gene conversion. There would be several ways to generate such a large nonhomology. One is via a DNA duplication at least several kilobases in length. The duplicated region would then be out of register with its solo copy in a heterozygous animal and thus be unable to exchange sequence via interchromosomal gene conversion. The duplicated cow pseudogenes are an example.

Long transposable elements might also inhibit interchromosomal gene conversion. There exists just such a sequence, the *Kpn* family of transposable elements (Potter, 1984; Martin et al., 1984; Shafit-Zagardo et al., 1982). This family is composed of long (up to 6-kb) DNA sequences that show a high degree of sequence conservation among orders of mammals. There are about 20000 copies per mammalian genome of this sequence. According to this scenario *Kpn* sequences would perform a function for allelic genes analogous to the function of *Alu* sequences in gene families. Thus, the two major classes of interspersed repetitive DNA in mammalian genomes would play corresponding roles in directing variation into the evolution of new forms (Figure 7).

Gene conversion plays a profound role in evolution. It promotes the flow of DNA sequences, both among the members of a multigene family and among the individuals of a species. As novel variations arise, they are transmitted and recombined along this two-dimensional network. Specific mechanisms exist to uncouple a gene from this interaction. A fuller understanding of these processes may lead to the solution of three outstanding problems in evolutionary theory: the question of how and at what level natural selection acts on DNA, the question of how isolating mechanisms work at the level of the gene, and the question of how gene variation is related to punctuated equilibria during evolution (Eldredge & Gould, 1972).

ACKNOWLEDGMENTS

The technical assistance of S. Sirkin and the secretarial skills of M. Loescher were greatly appreciated.

REFERENCES

Antonarkis, S. E., Boehm, C. D., Serjeant, G. R., Theisen, C. E., Dover, G. J., & Kazazian, H. K., Jr. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 853–856.

Baltimore, D. (1981) *Cell (Cambridge, Mass.) 24*, 592–594.

Bhat, S. P., & Spector, A. (1984) *DNA 3*, 287–295.

Clarke, S. H., & Rudikoff, S. (1984) *J. Exp. Med. 159*, 773–782.

Cleary, M. L., Schon, E. A., & Lingrel, J. B. (1981) *Cell (Cambridge, Mass.) 26*, 181–190.

Cohen, D., Le Gall, I., Marcadet, A., Font, M. P., Lalouel, J.-M., & Dausset, J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7870–7874.

Coligan, J. E. (1984) *Surv. Immunol. Res. 3*, 176–178.

Darby, V., & Blattner, F. (1984) *Science (Washington, D.C.) 226*, 1213–1215.

de Saint Vincent, B. R., & Wahl, G. M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2002–2006.

Devlin, J. J., Weiss, E. H., Paulson, M., & Flavell, R. A. (1985) *EMBO J. 4*, 3203–3207.

Dover, G. (1982) *Nature (London) 299*, 111–117.

Duncan, C. H. (1985) *N. Engl. Nuclear Prod. News 4(3)*, 6–7.

Edelman, G. M., & Gally, J. A. (1970) in *Neuroscience; Second Study Program*, pp 962–972, Rockefeller University, New York.

Eickbush, T. H., & Burke, W. D. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2814–2818.

Eldredge, N., & Gould, S. J. (1972) in *Models in Paleobiology* (Schopf, T. J. M., Ed.) pp 82–115, W. H. Freeman, San Francisco.

Fink, G. R., & Petes, T. D. (1984) *Nature (London) 310*, 728–729.

Fristensky, B., Lis, J., & Wu, R. (1982) *Nucleic Acids Res. 10*, 6451–6463.

Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L., & Slightom, J. L. (1984) *J. Mol. Biol. 180*, 803–823.

Gross, R. H. (1986) *Nucleic Acids Res. 14*, 591–596.

Hardies, S. C., Edgell, M. H., & Hutchinson, C. A., III (1984) *J. Biol. Chem. 259*, 3748–3756.

Hardison, R. C. (1984) *Mol. Biol. Evol. 1*, 390–410.

Hess, J. F., Fox, M., Schmid, C., & Shen, C.-K. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5970–5974.

Honjo, T. (1983) *Annu. Rev. Immunol. 1*, 499–528.

Hood, L., Campbell, J. H., & Elgin, S. C. R. (1975) *Annu. Rev. Genet. 9*, 305–353.

Klar, A. J. S., & Strathern, J. N. (1984) *Nature (London) 310*, 744–747.

Klein, H. L. (1984) *Nature (London) 310*, 748–750.
Kourilsky, P. (1983) *Biochimie 65*, 85–93.
Kucherlapati, R. S., Eves, E. M., Song, K.-Y., Morse, B. S., & Smithies, O. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3153–3157.
Kucherlapati, R. S., Spencer, J., & Moore, P. D. (1985) *Mol. Cell. Biol. 5*, 714–720.
Lalanne, J. L., Bregegere, F., Delarbre, C., Abastado, J. P., Gachelin, G., & Kourilsky, P. (1982) *Nucleic Acids Res. 10*, 1039–1049.
Lamb, B. C. (1985) *Heredity 53*, 113–138.
Li, W.-H., & Gojobori, T. (1983) *Mol. Biol. Evol. 1*, 94–108.
Liebhaber, S. A., & Begley, K. A. (1983) *Nucleic Acids Res. 11*, 8915–8930.
Liebhaber, S. A., Goossens, M., & Kan, Y. W. (1980) *Nature (London) 290*, 26–29.
Liebhaber, S. A., Rappaport, E. F., Cash, F. E., Ballas, S. K., Schwartz, E., & Surrey, S. (1984) *Science (Washington, D.C.) 226*, 1449–1451.
Liskay, R. M., & Stachelek, J. L. (1983) *Cell (Cambridge, Mass.) 35*, 157–165.
Maeda, N., Bliska, J. B., & Smithies, O. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5012–5016.
Martin, S. L., Voliva, C. F., Burton, F. H., Edgell, M. H., & Hutchison, C. A., III (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2308–2312.
Michaelson, A. M., & Orkin, S. H. (1983) *J. Biol. Chem. 258*, 15245–15254.
Nagylaki, T., & Petes, T. D. (1982) *Genetics 100*, 315–337.
Norrander, J., Kempe, T., & Messing, J. (1983) *Gene 26*, 101–106.
Ohta, T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1940–1944.
Ollo, R., & Rougeon, F. (1983) *Cell (Cambridge, Mass.) 32*, 515–523.
Pease, L. R., Schulze, D. H., Pfaffenbach, G. M., & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 242–246.
Potter, S. S. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1012–1016.
Rubintz, J., & Subramani, S. (1985) *Mol. Cell. Biol. 5*, 529–537.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schimenti, J. C., & Duncan, C. H. (1984) *Nucleic Acids Res. 12*, 1641–1655.
Schimenti, J. C., & Duncan, C. H. (1985a) *Mol. Biol. Evol. 2*, 505–513.
Schimenti, J. C., & Duncan, C. H. (1985b) *Mol. Biol. Evol. 2*, 514–525.
Schon, E. A., Cleary, M. L., Haynes, J. R., & Lingrel, J. B. (1981) *Cell (Cambridge, Mass.) 27*, 359–369.
Schon, E. A., Wernke, S. M., & Lingrel, J. B. (1982) *J. Biol. Chem. 257*, 6825–6835.
Schulze, D. H., Pease, L. R., Geier, S. S., Reyes, A. A., Sarmiento, L. A., Wallace, R. B., & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2007–2011.
Scott, A. F., Heath, P., Trusko, S., Boyer, S. H., Prass, W., Goodman, M., Czelusniak, J., Chang, E., & Slightom, J. L. (1984) *Mol. Biol. Evol. 1*, 371–389.
Seidman, J. G., Leder, A., Nau, M., Norman, B., & Leder, P. (1978) *Science (Washington, D.C.) 202*, 11–17.
Shafit-Zagardo, B., Maio, J. J., & Brown, F. L. (1982) *Nucleic Acids Res. 10*, 3175–3193.
Slightom, J. L., Blechl, A. E., & Smithies, O. (1980) *Cell (Cambridge, Mass.) 21*, 627–638.
Slightom, J. L., Chang, L. Y. E., Koop, B. F., & Goodman, M. (1985) *Mol. Biol. Evol. 2*, 370–389.
Small, J., & Scangos, G. (1983) *Science (Washington, D.C.) 219*, 174–176.
Smith, D. R., & Calvo, J. M. (1980) *Nucleic Acids Res. 8*, 2255–2274.
Spence, S. E., Young, R. M., Garner, K. J., & Lingrel, J. B. (1985) *Nucleic Acids Res. 13*, 2171–2186.
Spritz, R. A., DeRiel, J. K., Forget, B. G., & Weissman, S. M. (1980) *Cell (Cambridge, Mass.) 21*, 639–646.
Steinmetz, M., Winoto, A., Minard, K., & Hood, L. (1982) *Cell (Cambridge, Mass.) 28*, 489–498.
Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R. A., Dastoornikoo, G. R., Stephan, D., Gibb, E., & Romaniuk, R. (1984) *EMBO J. 3*, 2995–3003.
Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., & Stahl, F. (1983) *Cell (Cambridge, Mass.) 33*, 25–36.
Townes, T. M., Fitzgerald, M. C., & Lingrel, J. B. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6589–6593.
Walsh, J. B. (1983) *Genetics 105*, 461–468.
Watanabe, Y., Tsukada, T., Notake, M., Nakanishi, S., & Numa, S. (1982) *Nucleic Acids Res. 10*, 1459–1468.
Weiss, E. H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J., & Flavell, R. A. (1983) *Nature (London) 310*, 671–674.