Kappa-Chain Constant-Region Gene Sequences in Genus *Rattus*: Coding Regions Are Diverging More Rapidly Than Noncoding Regions

M. B. Frank,* R. M. Besta,* P. R. Baverstock,† and G. A. Gutman*

*Department of Microbiology and Molecular Genetics, University of California, Irvine; and †South Australian Museum, Adelaide

We have determined the nucleotide sequence of a 1,200-base pair (bp) genomic fragment that includes the kappa-chain constant-region gene (Ck) from two species of native Australian rodents, *Rattus leucopus cooktownensis* and *Rattus colletti*. Comparison of these sequences with each other and with other rodent Ck genes shows three surprising features. First, the coding regions are diverging at a rate severalfold higher than that of the nearby noncoding regions. Second, replacement changes within the coding region are accumulating at a rate at least as great as that of silent changes. Third, most of the amino acid replacements are localized in one region of the Ck domain—namely, the carboxy-terminal "bends" in the alpha-carbon backbone. These three features have previously been described from comparisons of the two allelic forms of Ck genes in *R. norvegicus*. These data imply the existence of considerable evolutionary constraints on the noncoding regions (based on as yet undetermined functions) or powerful positive selection to diversify a portion of the constant-region domain (whose physiological significance is not known). These surprising features of Ck evolution appear to be characteristic only of closely related Ck genes, since comparison of rodent with human sequences shows the expected greater conservation of coding regions, as well as a predominance of silent nucleotide substitutions within the coding regions.

Introduction

Immunoglobulin light chains consist of two globular domains. The N-terminal V-domain is involved (together with the corresponding V-domain of the heavy chain) in the formation of an antigen-specific combining site; the C-terminal constant domain has no identified biological function but interacts by covalent and noncovalent bonds with the C\textsubscript{H}1 domain of the heavy chain to stabilize the polymeric immunoglobulin molecule.

Contrary to the expectation that coding regions should be more highly conserved than noncoding regions, Sheppard and Gutman (1981b) found that the coding regions of the allelic forms of constant-region genes of kappa light chains (Ck) in *Rattus norvegicus* are more than three times as different from one another as are adjacent noncoding regions. In addition, substitutions within the coding regions that result in amino acid replacements outnumber silent changes. These

---

1. Key words: immunoglobulin C-kappa genes, genus *Rattus*, coding/noncoding sequence divergence.

Address for correspondence: M. B. Frank, Department of Microbiology and Molecular Genetics, California College of Medicine, University of California, Irvine, California 92717.
results have been interpreted as implying the existence of considerable evolutionary constraint at the nucleotide sequence level.

The present work was undertaken to determine whether this pattern of conservation is an unusual or unique feature of the two allelic forms of kappa chains within *Rattus norvegicus* or a pattern generally found within closely related kappa-chain genes. We have cloned C\textsubscript{k} genes from two species of native Australian *Rattus* and determined their nucleotide sequence. A comparison of these genes with each other and with those from *R. norvegicus* shows the same unusual features previously seen with the Igk-1\textsuperscript{2} alleles: coding regions are more diverse than noncoding regions, and silent changes within the coding region do not outnumber replacement changes. This pattern is therefore not a unique feature of the allelic forms of C\textsubscript{k} within *R. norvegicus* but must be understood in a context at least as broad as that consisting of closely related C\textsubscript{k} genes within the genus *Rattus*. Recent studies on allelic forms of C\textsubscript{k} genes in rabbits (Pavirani et al. 1983) reveal a greater degree of divergence between coding regions than that between 3' untranslated regions of cDNA. Thus, this pattern may be generally true of closely related C\textsubscript{k} genes within mammals.

**Material and Methods**

**Animals**

Australian *Rattus* species used in these studies were collected in 1979. *Rattus leucopus cooktownensis* was trapped at Kuranda, Queensland, and *R. colletti* was trapped at Cannon Hill, Northern Territory. Animals were maintained in the laboratory and sacrificed to obtain plasma and liver. Livers were either used immediately or stored at -70°C until used.

These rats represent two of some 14 endemic forms of *Rattus* that have evolved in Australia and New Guinea over the past few Myr. They have been extensively studied with respect to their taxonomic relationships, karyotypes, and biochemical genetics (Taylor and Horner 1973; Baverstock et al. 1977, 1979, 1983). The members of this monophyletic group of *Rattus* species are more closely related to one another than they are to Asian *Rattus* species, whose taxonomy is complex and controversial. The two other types of rats to which we refer in this work—namely, LOUVAIN (LOU) and DA—are two inbred strains of *R. norvegicus*, and genetic differences between them therefore represent polymorphic loci in this species.

**Identification and Cloning of the C\textsubscript{k} Genes**

The DNA was extracted from the livers of one *R. leucopus cooktownensis* and one *R. colletti* according to the method of Blin and Stafford (1976), digested with a variety of restriction endonucleases, and electrophoresed in a 0.8% agarose gel prior to Southern blotting (Southern 1975). The nitrocellulose filter was hybridized with an *R. norvegicus* nick-translated C\textsubscript{k} probe (Sheppard and Gutman 1981b), washed, and autoradiographed to obtain a restriction-enzyme map of the C\textsubscript{k} gene in its unrearranged genomic context and to determine the size of the C\textsubscript{k}-containing EcoRI fragments for subsequent cloning.

Additional DNA was digested with EcoRI and was size-selected using NaCl zonal gradient centrifugation. Fractions with DNA in the range of 6–7 kilobase pairs (kbp) for *R. l. cooktownensis* and 4–5 kbp for *R. colletti* were separately pooled for each species and ligated to an EcoRI-digested λgtWES vector (Enquist et al. 1989).

2. The standard nomenclature for rat immunoglobulin loci is used (Gutman et al. 1983).
1976) after removal of the \(\lambda B\) insert. The resulting DNA was packaged in vitro (Hohn and Murray 1977), plated on \(E. coli\) strain LE-392, and screened by in situ hybridization (Benton and Davis 1977) with the probe described above. Following plaque purification, the \(C_K\)-containing \(EcoRI\) fragments were isolated and either used directly or subcloned into the \(EcoRI\) site of pBR325.

Restriction enzyme mapping of the \(C_K\)-containing \(EcoRI\) fragments from each clone revealed that the \(C_K\) gene of both species is located in an approximately 1,200-bp \(BspRI\) fragment, as is the case in \(R. norvegicus\) (Sheppard and Gutman 1981b). This fragment was isolated from each species using a preparative 5% \(N, N'\)-bis-acrylyl cystamine-cross-linked polyacrylamide gel (Bio-Rad Laboratories, Richmond, Calif.) following reduction and DEAE-cellulose ion-exchange chromatography.

Sequencing the \(C_K\) Genes

Subfragments of these 1,200-bp regions were ligated into the M13mp8 vector (Messing and Vieira 1982) following the method provided by Bethesda Research Laboratories (Gaithersburg, Md.) and with the addition of spermidine to a final concentration of 0.14 mM (Sanger et al. 1980). The ligated DNA was transformed in \(E. coli\) strain JM103, and the resulting clear colonies were screened for the presence of the \(C_K\) gene by in situ hybridization with a nick-translated \(C_K\) probe (see previous subsection). The latter step proved particularly useful in the case of blunt end-ligated inserts, which had a high (30%-50%) rate of false positives when screened solely on the basis of these clones' inability to produce \(\beta\)-galactosidase. The difficulty in successfully cloning blunt-end inserts was directly related to the insert size; in one series of experiments, the number of positive clones per microgram of insert used for ligation was 53 for inserts <200 bp, 16 for inserts having 200-400 bp, and seven for the 1,200-bp \(BspRI\) fragment.

Single-stranded DNA was prepared from positive clones by the method of Sanger et al. (1980), hybridized to a 17-base single-stranded primer (Collaborative Research, Lexington, Mass.), and sequenced by the dideoxy method of Sanger et al. (1977). For each clone, an additional reaction and gel lane was included using deoxyinosine 5'-triphosphate and 2',3'-dideoxy GTP to identify regions of “compression” of bands in the “G” lanes (W. Barnes, personal communication). If there were any ambiguities in the reading of the gel, they were resolved by sequencing either the complementary strand or an overlapping fragment of the same strand.

Serological Typing

Serological typing for Igk-1 was carried out using a solid-phase radioimmunoassay (Gutman 1977). The level of serological cross-reactivity with the Igk-1a allotype of \(R. norvegicus\) (Gutman et al. 1983) was 94% ± 0% and 2% ± 1% for the \(R. l. cooktownensis\) and \(R. colletti\) animals studied, respectively. As is the case with all non-\(R. norvegicus\) animals (Gutman 1981), no serological cross-reaction was detected with the Igk-1b allotype.

Data Analysis

Computer searches for sequence similarity and alignment of overlapping sequences were carried out using computer programs developed by Staden (1980), modifications of these programs kindly provided by Isono (1982), and the Molgen programs (Sumex Project, Stanford University).
Nucleotide substitutions in coding regions from different species were analyzed by a computer program written (in the SIMULA language) to determine the proportion of replacement and silent changes according to the method described by Perler et al. (1980). These calculations adjust the percentage of difference between two coding regions on the basis of two factors: first, random nucleotide changes in codons are expected to produce more replacement than silent changes (≈3:1 in our sequences); second, multiple nucleotide changes may occur in the same codon over time. The first adjustment is dealt with by producing a catalog (for each pair of sequences to be compared) of the results of all possible nucleotide substitutions within each codon (replacement vs. silent) and considering the actual substitutions as a proportion of the potential number of replacement or silent ones. The second adjustment, based on the assumption that substitutions occur randomly with respect to one another, involves a Poisson correction. Following these adjustments, random fixation of unselected base changes should yield a ratio of silent to replacement changes of 1:1. (These calculations have been criticized [Gojobori et al. 1982] for their underlying assumption of equal rates of transitions and transversions, but this assumption is problematic mainly in comparisons of genes considerably more distantly related than those herein studied.)

Three-Dimensional Structure of Ck Domains

Although no X-ray crystallographic studies have yet been carried out on rat immunoglobulins, a great deal is known about the three-dimensional structure of mouse kappa chains (see Beale and Feinstein 1976). The Ck domain consists mainly of two antiparallel beta-pleated sheets covalently and noncovalently bound to each other. The remainder comprises six regions where the alpha-carbon backbone turns back on itself to join one beta-pleated-sheet strand with another. These six regions of surface loops we refer to as bends; three of them are in the carboxyl half (C-terminal bends) of the domain, which connects with the heavy chain via a disulfide bridge, and the other three are in the amino terminal half, which connects with the V-domain.

We used an alpha-carbon-backbone model of the mouse Ck domain that was kindly provided by P. Colman (Sydney) to locate the sites of amino acid differences between different forms of Ck. With the help of R. Feldman (NIH, Bethesda, Md.), we have shown that the rat amino acid sequence is readily accommodated by this same backbone structure (see Gutman 1981).

Results

Multiple Genomic Contexts of Ck Genes in Rattus colletti and R. l. cooktownensis

Restriction-enzyme mapping of liver DNA from R. l. cooktownensis and R. colletti revealed that the unrearranged Ck gene is present in each species in two distinct contexts (fig. 1). Although serological studies have indicated the existence of polymorphisms among Ck genes of certain species of Australian Rattus (Gutman and Baverstock 1980), in neither of the animals studied here did serological data predict the presence of allelic genes. In the case of R. l. cooktownensis, no serological Ck polymorphism has been found. Although at least two levels of cross-reactivity with Igk-la have been described in R. colletti, the animal of this species that we studied was typed serologically as a homozygote for the allele of least cross-reactivity (see Material and Methods). Therefore, the two different Ck genomic contexts that
Rattus leucopus cooktownensis

![Map of Rattus leucopus cooktownensis](image)

Rattus colletti

![Map of Rattus colletti](image)

FIG. 1.—Restriction-enzyme maps of Ck-containing genomic and cloned EcoRI fragments. The top two and bottom two maps represent different unrearranged Ck configurations of Rattus leucopus cooktownensis and Rattus colletti, respectively, as determined by Southern blotting of liver DNA. The center two maps are expanded views of cloned EcoRI fragments from each of these two species. B = BamHI; R = EcoRI; X = XbaI; Hd = HindIII; Ps = PstI; Hc = HincII; Pv = PvuII; Bs = BspRI; Hp = HpaI; and S = SalI. Unmarked vertical bars indicate restriction-enzyme sites identical to those found on the map directly above. Only two BspRI sites are shown, representing the distal ends of the nucleotide-sequenced region.

...we found in both of these species must be serologically indistinguishable, either because of serologically undetectable changes within the coding region or because of nucleotide changes outside the coding region. As described below, the latter situation holds for the two R. l. cooktownensis genes. The nature of the polymorphism in R. colletti remains to be explained, since only one of the two genes has been cloned and sequenced. (Although it has essentially been ruled out in R. norvegicus [Sheppard and Gutman 1981a], the possibility of duplicated Ck genes in these Australian rats must still be at least formally considered.)

Ck Genomic Clones

Six independent, recombinant, phage-containing Ck-nucleotide sequences were isolated from a size-selected EcoRI library of R. l. cooktownensis liver DNA, and one such sequence was isolated from an R. colletti library. Each of the R. l. cooktownensis clones contained a single hybridizing 6.5-kb EcoRI fragment, whereas the R. colletti clone contained a 4.4-kb EcoRI hybridizing fragment (fig. 1).

Analysis of restriction enzyme digests of the six R. l. cooktownensis Ck clones revealed that one clone showed a set of restriction enzyme fragments different from those of the others. These two types of clones presumably correspond to the two genomic contexts shown in figure 1, although in the absence of mapped differences...
within the genomic EcoRI fragments, we cannot determine which clone is derived from which context. In the case of the *R. colletti* C<sub>K</sub> clone, the presence of the BamHI site within the EcoRI fragment unambiguously identifies the genomic context from which it was derived (see fig. 1).

There is a high degree of similarity between the genomic restriction maps of *R. colletti* and *R. l. cooktownensis*, as well as between these Australian *Rattus* species and *R. norvegicus*. The maps of *R. colletti* and *R. l. cooktownensis* differ from each other only in the positions of the distal EcoRI sites and in one HincII site, a site which, in *R. colletti* but not *R. l. cooktownensis*, is also recognized by SalI. Only two of the approximately 14 BspRI sites are shown in figure 1; the 1.2-kb fragment defined by these sites is conserved in all *Rattus* examined thus far, and it was this nucleotide sequence that was determined for each species.

**C<sub>K</sub> Nucleotide Sequences**

The nucleotide sequences of the C<sub>K</sub> genes from *R. l. cooktownensis* and *R. colletti* have been aligned by eye to maximize identities with the LOU C<sub>K</sub> sequence (fig. 2). Beginning with the 5' BspRI site, approximately 440 nucleotides

---

**Fig. 2.**—Nucleotide sequences of C<sub>K</sub> genes of three *Rattus* species. Sequences have been aligned by eye to maximize identity, and the hyphens indicate identity with the topmost (LOU) sequence (Sheppard and Gutman 1981b); size differences are discussed in Results. The C<sub>K</sub>-coding regions are shown in brackets, preceded by the 5' flanking region. The presumed site of polyadenylation is shown by a vertical bar (demarcating the 3' untranslated and 3' flanking regions) 19 bases downstream from the underlined polyadenylation recognition sequence AATAAA (between nucleotides 984 and 985). The first two G and the last two C residues are inferred from the BspRI recognition sequence. The LOU designation is the same as that given in the note to table 1. R.I.c. = *Rattus leucopus cooktownensis*, and R.c. = *Rattus colletti*. 
of the 3' end of the \( J_K - C_K \) intervening region have been sequenced (the number of nucleotides varies for each species in this region, as discussed below). The \( C_K \)-coding regions are located between nucleotides 464 and 784 and are followed by a 199-base 3'-untranslated region. An additional 209 nucleotides of the 3' flanking region are also shown, which end in the 3' BspRI site. Only one of the two \( R. \ l. \) cooke\textit{townensis} sequences is shown; the second clone differs from this at only two positions in the 3' flanking region—namely, position 1017 (A → G) and position 1091 (A → C). These two genes therefore code for identical polypeptides and (as discussed above) represent serologically indistinguishable alleles (or duplicate genes).

Comparison of the three \( Rattus \ C_K \) sequences reveals several size differences in the noncoding regions. One such difference (an 11-bp duplication of nucleotides 72–82 following position 51) distinguishes the \( R. \ colletti \) sequence from the other two. Other size differences distinguish \( R. \ norvegicus \) from both Australian \( Rattus \) species—namely, an 8-bp and an 11-bp tandem duplication in the region between positions 358 and 396 of \( R. \ norvegicus \) and an 8-bp and 1-bp size difference following position 409 and at 442, respectively. (There are no size differences between the LOU and DA alleles of \( C_K \).)

Maintenance of Repetitive Sequence

A number of short repetitive sequences have been identified in the Australian rat \( C_K \) sequences. These include the tetranucleotide GTTGT, which is repeated four times in the 5' intron between nucleotides 111 and 151, another four times between nucleotides 346 and 377, and four more times in the coding region between nucleotides 536 and 627. The tetranucleotide TCCT occurs four times between bases 421 and 503—a region overlapping the 3' end of the 5' intervening sequence and the beginning of the coding region—as well as seven times in the 3' untranslated region. Similarly, the tetranucleotide TTTG is present four times between nucleotides 929 and 990 at the end of the 3' untranslated region. There is also an imperfect 34-base palindromic sequence in the 5' flanking region between bases 239 and 274. Although many of these repeated sequences have been reported in human, rat, and mouse \( C_K \) genes (Hieter et al. 1980; Max et al. 1981; Sheppard and Gutman 1981b), their significance remains unknown.

Increased Sequence Divergence in Coding versus Noncoding Regions

The differences between the \( C_K \) genes of the two Australian \( Rattus \) and \( R. \ norvegicus \) are shown in table 1, using the alignment shown in figure 2. Although the noncoding sequences may be divided into three regions (i.e., 5' intervening sequence, 3' untranslated, and 3' flanking), the results are not markedly different for the three, and they have been joined under the general term “noncoding” for analytical purposes. The \( C_K \)-coding sequences from different \( Rattus \) species have diverged approximately three times as much as have the noncoding regions, a result similar to that previously shown for the allelic forms of \( C_K \) genes in both \( R. \ norvegicus \) (Sheppard and Gutman 1981b) and rabbits (Pavirani et al. 1983).

The degree of similarity between the coding-region sequences of \( Rattus \ C_K \) genes reflects that predicted by serological data (Gutman and Baverstock 1980). The \( R. \ l. \) cooke\textit{townensis} kappa chain cross-reacts quantitatively with Igk-1a (LOU), and the nucleotide-sequence difference between these two coding regions is the lowest shown in table 1 (and this is the only comparison that fails to show a statistically
Table 1
Differences Between Cκ-Coding and Noncoding Regions

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Coding</th>
<th>Noncoding</th>
<th>P&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rattus leucopus cooktownensis vs. Rattus colletti</td>
<td>3.7</td>
<td>1.5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>R. l. cooktownensis vs. LOU</td>
<td>2.8</td>
<td>2.0</td>
<td>NS</td>
</tr>
<tr>
<td>R. l. cooktownensis vs. DA</td>
<td>5.0</td>
<td>1.9</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>R. colletti vs. LOU</td>
<td>5.6</td>
<td>1.6</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>R. colletti vs. DA</td>
<td>6.2</td>
<td>1.9</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>LOU vs. DA&lt;sup&gt;;c&lt;/sup&gt;</td>
<td>3.8</td>
<td>1.1</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

NOTE.—LOU and DA are inbred strains of Rattus norvegicus. NS = not significant.

<sup>a</sup> Each size difference in the noncoding region was counted as a single difference, since each is presumably the result of a single event.

<sup>b</sup> From x<sup>2</sup> test for independence of differences between the coding and total noncoding regions. P > 0.05 was considered to be not statistically significant.

<sup>c</sup> Sheppard and Gutman (1981b).

A significant difference between coding and noncoding sequences. The R. colletti kappa chain cross-reacts not at all with Igκ-1a, and table 1 shows it to be more different from both LOU and DA than is R. l. cooktownensis.

We have further analyzed the differences between coding regions to determine the rate of change at replacement sites (which result in amino acid differences in the protein) vs. that at silent sites. Application of the calculations described by Perler et al. (1980) yields the divergence values shown in table 2. If conservation of

Table 2
Divergence of Silent and Replacement Positions in the Coding Region

<table>
<thead>
<tr>
<th>COMPARISON</th>
<th>Silent (S)</th>
<th>Replacement (R)</th>
<th>Ratio (S/R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-kappa genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DA vs. LOU&lt;sup&gt;;b&lt;/sup&gt;</td>
<td>1.6</td>
<td>4.6</td>
<td>0.35</td>
</tr>
<tr>
<td>Rattus leucopus cooktownensis vs. R. colletti</td>
<td>3.9</td>
<td>5.9</td>
<td>0.7</td>
</tr>
<tr>
<td>R. l. cooktownensis vs. LOU&lt;sup&gt;;b&lt;/sup&gt;</td>
<td>4.8</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>R. l. cooktownensis vs. DA&lt;sup&gt;;b&lt;/sup&gt;</td>
<td>3.2</td>
<td>5.9</td>
<td>0.5</td>
</tr>
<tr>
<td>R. l. cooktownensis vs. Mouse&lt;sup&gt;;c&lt;/sup&gt;</td>
<td>19.5</td>
<td>10.6</td>
<td>1.8</td>
</tr>
<tr>
<td>R. l. cooktownensis vs. Human&lt;sup&gt;;d&lt;/sup&gt;</td>
<td>76.0</td>
<td>29.9</td>
<td>2.5</td>
</tr>
<tr>
<td>Rattus colletti vs. LOU&lt;sup&gt;;b&lt;/sup&gt;</td>
<td>8.4</td>
<td>6.7</td>
<td>1.3</td>
</tr>
<tr>
<td>R. colletti vs. DA&lt;sup&gt;;b&lt;/sup&gt;</td>
<td>7.8</td>
<td>7.6</td>
<td>1.0</td>
</tr>
<tr>
<td>R. colletti vs. Mouse&lt;sup&gt;;c&lt;/sup&gt;</td>
<td>26.5</td>
<td>8.7</td>
<td>3.1</td>
</tr>
<tr>
<td>R. colletti vs. Human&lt;sup&gt;;d&lt;/sup&gt;</td>
<td>90.4</td>
<td>29.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Other genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse vs. Human α-globin&lt;sup&gt;;e&lt;/sup&gt;</td>
<td>83.0</td>
<td>8.4</td>
<td>9.9</td>
</tr>
<tr>
<td>Rat vs. Human insulin&lt;sup&gt;;e&lt;/sup&gt;</td>
<td>76.0</td>
<td>5.2</td>
<td>14.6</td>
</tr>
</tbody>
</table>

NOTE.—Designations are the same as those listed in the note to table 1.

<sup>a</sup> Estimated number of nucleotide substitutions per 100 nucleotides, based on the method of Perler et al. (1980) described in Material and Methods.

<sup>b</sup> Sheppard and Gutman 1981b.

<sup>c</sup> Max et al. 1981.

<sup>d</sup> Heiter et al. 1980.

<sup*e</sup> Perler et al. 1980.
Protein structure is the major evolutionary constraint on nucleotide substitutions in coding regions, one would expect to observe an abundance of silent vs. replacement changes. However, our comparisons of Ck-coding regions from different Rattus species fail to show a predominance of silent changes. As seen in table 2, only two of the six silent:replacement ratios within Rattus are above unity. The highest ratio (2.0 for R. l. cooktownensis vs. LOU) is for a pair of Ck genes whose proteins are known to be serologically closely related (see Gutman and Baverstock 1980 and discussion above). Thus, not only are Ck-coding regions diverging more rapidly within Rattus than are noncoding regions (as discussed above), but amino acid-replacement changes within coding regions are accumulating at least as rapidly as are silent changes.

Interestingly, the ratio of silent to replacement changes for Ck genes from more distantly related species (e.g., rodent and human) is approximately three to five times less than that for other genes analyzed by this method (table 2). Although the general increase in the ratio of silent to replacement changes is related to the time of divergence between species, the high rate of replacement changes in the most distant comparison appears to be a distinctive feature of Ck genes. This may indicate that, although they clearly exist (as discussed below), the selective forces acting to conserve Ck amino acid sequences are not as strong as those acting to conserve globins and insulins.

Location of Amino Acid-Sequence Substitutions

We have previously noted that the amino acid differences in the Ck domain between the Ck of LOU and DA strains of R. norvegicus are not randomly distributed over the three-dimensional structure of the domain: eight of the 11 differences reside on the three surface loops (bends) in the alpha-carbon backbone that occur at the carboxyl half of the domain (Gutman 1981). The Ck amino acid sequences of R. l. cooktownensis and R. colletti have been determined by translating their nucleotide sequences, and the positions of amino acid substitutions between species have been localized on a three-dimensional rat-Ck model (kindly provided by P. Colman). The majority of Ck amino acid substitutions between these Rattus species (table 3) are also localized in the carboxy-terminal bends of the domain, which is in agreement with the above findings for the Ck allelic differences in R. norvegicus. This pattern points to an appreciable degree of selective constraint on the Ck domain (as discussed below).

Discussion

If the selective forces operating on diverging genes act primarily to conserve protein structure and function, two expectations follow: first, noncoding regions should show a higher rate of divergence than do coding regions; second, within coding regions, silent changes should accumulate at a higher rate than do replacement changes. Although these expectations have been borne out by analysis of globin, ovalbumin, insulin, and metallothionein genes from distantly related species (see Perler et al. 1980; Jeffreys 1981; Griffith et al. 1983), we have described precisely the opposite in the case of two allelic forms of Ck genes in the laboratory rat R. norvegicus (Sheppard and Gutman 1981b).

Tables 1 and 2 and figure 2 show that the surprising features of Igk-1 allelic divergence are not unique to these alleles but are also seen when comparison is made of Ck genes of at least three species within the genus Rattus. Coding regions
Table 3

Spatial Distribution of Amino Acid Replacements in C\textsubscript{K} Domains

<table>
<thead>
<tr>
<th>COMPARISON</th>
<th>C-terminal Bends</th>
<th>Remainder of Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rattus leucopus cooke	townensis vs. Rattus colletti</td>
<td>6 (22)</td>
<td>5 (6)*</td>
</tr>
<tr>
<td>R. l. cooke	townensis vs. LOU</td>
<td>3 (11)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>R. l. cooke	townensis vs. DA</td>
<td>8 (30)</td>
<td>5 (6)†</td>
</tr>
<tr>
<td>R. colletti vs. LOU</td>
<td>7 (26)</td>
<td>7 (9)</td>
</tr>
<tr>
<td>R. colletti vs. DA</td>
<td>8 (30)</td>
<td>7 (9)*</td>
</tr>
<tr>
<td>LOU vs. DA</td>
<td>8 (30)</td>
<td>3 (4)†</td>
</tr>
</tbody>
</table>

\begin{itemize}
\item \textit{NOTE.}—Amino acid positions were determined from a three-dimensional model of the C\textsubscript{K} domain (see Gutman 1981). C-terminal bend residues (see Materials and Methods) consist of amino acids 121-129, 150-158, and 182-190—or 27 residues out of a total of 106 in the constant region (the numbering is that of Gutman et al. [1975] and refers to the entire kappa chain beginning at the N-terminus of the V-region). Designations are the same as those listed in the note to table 1.
\item \(\ast P < 0.05\).
\item \(\dagger P < 0.01\).
\end{itemize}

are diverging at an overall rate two to three times higher than that of noncoding regions, and replacement changes within coding regions are accumulating as fast as (or faster than) silent changes. Construction of a tree (see Fitch 1980) indicates that the coding to noncoding ratios are 5:1 and 8:1 in the lines leading to LOU and \textit{R. colletti}, respectively, but only 1:2 in that leading to \textit{R. l. cooke	townensis}; however, the total number of coding changes in the \textit{R. l. cooke	townensis} line is too small for statistical confidence regarding these differences. Thus, whatever caused the high rate of divergence in these coding regions and replacement sites within these coding regions, it is operating in at least as broad a context as are the C\textsubscript{K} genes in the genus \textit{Rattus}. Recent comparisons of partial cDNA sequences of allelic forms of rabbit kappa chains have similarly shown a higher degree of conservation of 3' untranslated sequences than of the coding regions (Pavirani et al. 1983); this may therefore be a general characteristic of closely related mammalian C\textsubscript{K} genes. Although these findings may be explained by hypothesizing as-yet-unrecognized constraints on noncoding DNA near C\textsubscript{K} genes from these closely related species (such as a required higher-order structure of either DNA or RNA [Sheppard and Gutman 1981\textit{b}]), the fact that these unusual features do not hold for comparisons between more distantly related C\textsubscript{K} genes (e.g., rodent:human [table 2 and Gutman and Sheppard 1983]) remains a puzzle. We have found that the overall rate of divergence of coding regions is two to three times greater than that of noncoding regions and that the rate of divergence within the three C-terminal bends is higher still. If the amino acid–sequence differences we see in these localized regions are all selectively neutral, then considerable negative selection must be invoked to explain the slower rate of evolution in noncoding regions. On the other hand, if these amino acid substitutions are the result of positive selection, there must be important functional differences between the different forms of C-kappa proteins. However, any positive selection invoked, must operate within the context of the considerable conservative influences readily apparent in C-kappa evolution; the three-dimensional structure of the C-kappa domain is highly conserved between species as different as the mouse and the human (Beale and Feinstein 1976; Poljak 1978).
Three broad questions remain to be answered. First, what functions might be maintained by conservation of noncoding nucleotide sequences adjacent to the \( C_\kappa \) gene? The high level of similarity of these noncoding regions between different species of rat as well as between rat and mouse strongly implicates the existence of as yet undefined functions of this region. Known enhancer sequences, for instance, can result in considerable conservation of noncoding regions; however, the only such sequences identified in kappa-chain genes (Picard and Schaffner 1984) are located outside the region herein studied. Our current knowledge of the regulation of immunoglobulin light-chain gene expression at the level of nucleotide sequences as well as of nucleic-acid housekeeping functions limits the development of a testable hypothesis at this time.

Second, what functions might be the object of having diversifying selection operating at the level of the protein for \( C_\kappa \) domains? Although \( C_\kappa \) domains bind covalently and noncovalently to the \( C_\text{H}1 \) domain of heavy chains, there are no other specific functions assignable to the \( C_\kappa \) domain. As far as is known, \( C_\kappa \) domains do not interact with complement components, are not involved in antigen binding, and do not bind to any of the receptors with which immunoglobulins can interact (cytophilic receptors, secretory component, yolk-sac receptors, etc.) (see Nisnoff et al. 1975). Invoking the interaction of \( C_\kappa \) and \( C_\text{H}1 \) domains as a primary focus of this diversifying selection is weakened by our finding that the differences between various \textit{Rattus} species' \( C_\kappa \) domains are mostly localized elsewhere than the sites of interaction of these two domains (which involves the four-strand \( \beta \)-pleated sheet). In the absence of a clearly defined \( C_\kappa \) function other than that of stabilization of heavy-light chain interactions, it is difficult to test the possibility that the various forms of kappa chains we have identified in \textit{Rattus} are physiologically distinguishable.

Finally, the question of whether these unexpected patterns of nucleotide substitutions may be a consequence of saltatory events in early rodent evolution (which may make them unique to the Muridae) or whether they may be features of short term evolutionary events in general (which have not yet been extensively studied) is currently being pursued in our laboratory by the analysis of nucleotide sequences of other recently divergent genes.

Acknowledgments

This work was supported in part by grants from the U.S. Public Health Service (AI 14774) and from the KROC Foundation. G.A.G. was the recipient of a Research Career Development Award (AI 00286), and M.B.F. was supported in part by an Individual National Research Service Award (AI 06763) from the U.S. Public Health Service. We thank Dr. H. W. Sheppard for valuable help in the initial stages of this work.

LITERATURE CITED


WALTER M. FITCH, reviewing editor

Received April 23, 1984; revision received July 30, 1984.