Variance of Sequence Divergence

R. R. Hudson and G. B. Golding
National Institute of Environmental Health Sciences

Golding (1983) derives the variance of a quantity, \( D \), which is a population measure of divergence. He considers the divergence of two populations that have been isolated for \( t \) generations. Golding finds that the variance of \( D \), under a neutral model, is less than the binomial variance that is commonly assumed as the variance of divergence. We are concerned that Golding's results could be misinterpreted and misapplied to measures of divergence for which his results are not appropriate. In particular, his results should not be applied to a statistic, \( S \), based on a comparison of two sequences obtained as follows. Two gametes are randomly chosen, one from population 1 and one from population 2. Homologous genes are sequenced from each gamete. \( S \) is the proportion of sites which differ between the two sequences. \( S \) is an observable measure of divergence that is commonly calculated and has often been assumed to be binomially distributed (Holmquist 1972; Kimura and Ohta 1972; Kaplan and Risko 1982). The relationship of \( S \) to \( D \) is described below. We shall also show that, under the same neutral model that Golding considers, the variance of \( S \) is greater than binomial. We shall consider only the case where the mutation rate is the same at all sites, but the case where mutation rates vary from site to site can be analyzed in the same way.

Denote the frequency of allele \( j \) at site \( i \) in population 1 by \( p_j(i) \). Denote the corresponding frequency in population 2 by \( q_j(i) \). Given the frequencies \( p_j(i), q_j(i), j = 1, \ldots, k \), the probability that two randomly chosen gametes, one from each population, are identical at site \( i \) is \( x_i = \prod_j p_j(i) q_j(i) \). Following Nei (1972), Ewens (1979), and Takahata (1982), Golding defines \( D \) as

\[
\frac{1}{n} \sum_{i=1}^{n} (1 - x_i).
\]

\( S \) can be defined as

\[
\frac{1}{n} \sum_{i=1}^{n} z_i,
\]

where \( z_i \) is an indicator variable which is one if two sequences are different at site \( i \) and zero otherwise. Given \( x_i \), the probability that \( z_i = 1 \) is \( 1 - x_i \). It follows that \( D = \mathbb{E}(S|x_1, \ldots, x_n) \), and clearly the expectation of \( D \) equals the expectation of \( S \). In contrast to \( S \), \( D \) is a random variable that is not directly observable (unless the sequence of every individual in both populations could be obtained).

The variance of \( S \) is not in general equal to the variance of \( D \). The variances,

Address for correspondence and reprints: R. R. Hudson, NIEHS, P.O. Box 12233, Research Triangle Park, North Carolina 27709.

Permission to reprint a letter printed in this section may be obtained only from the authors.
of course, depend on how one envisions doing replications. Following Golding, we assume that replications would consist of completely rerunning the evolutionary history of the populations and, to obtain $S$, randomly drawing two gametes. Using elementary properties of conditional statistics, one can write the variance of $S$ as:

$$\text{var}(S) = E[\text{var}(S|X)] + \text{var}[E(S|X)],$$

where $X = (x_1, \ldots, x_n)$. But the expectation of $S$ conditional on $X$ is just $D$. So

$$\text{var}(S) = E[\text{var}(S|X)] + \text{var}(D).$$

The variance of $D$ for a single site is equal to the variance of Li and Nei’s (1975) genetic distance measure $J_{12}$. The variance of $S$ is always greater than or equal to the variance of $D$.

Under the neutral model, the variance of $S$ can be found as

$$\text{var}(S) = \left[ \sum_i \text{var}(z_i) + \sum_{i \neq j} \text{cov}(z_i, z_j) \right] / n^2.$$ 

$\text{Var}(z_i)$ and $\text{cov}(z_i, z_j)$ can be written in terms of the identity coefficients, $\Phi_1(t)$ and $\Phi_4(t)$, that Golding defines and for which he derives explicit expressions. $\Phi_1(t)$ is defined as the probability that two randomly chosen gametes (one from each population) are identical at site 1; $\Phi_4(t)$ is defined as the probability that two randomly chosen gametes (one from each population) are identical at both site 1 and site 2. $\text{Var}(z_i)$ is just $\Phi_1(t)[1 - \Phi_1(t)]$. $\text{cov}(z_i, z_j)$ is equal to $E(z_i z_j) - [1 - \Phi_1(t)]^2$. But $E(z_i z_j)$ is simply the probability that the two gametes differ at both sites, which is $1 - 2\Phi_1(t) + \Phi_4(t)$. So we can write the variance of $S$ as

$$\text{var}(S) = \frac{1}{n} \Phi_1(t)[1 - \Phi_1(t)] + \frac{n - 1}{n} [\Phi_4(t) - \Phi_1(t)^2].$$

The first term on the right-hand side is the binomial variance that is usually assumed for the variance of $S$. The second term is always positive for the equilibrium neutral model, since under this model $\Phi_4(t)$ is always greater than $\Phi_1(t)^2$. So the variance of $S$ is always greater than binomial. For most applications, the departure from the binomial variance is probably negligible, but for recently diverged populations that are highly polymorphic, the effect can be large.

The variance of $S$ is greater than binomial because of the polymorphism in the population at the time of the split. If the population is monomorphic at the time of the split, $[\Phi_1(0) = 1]$, then the variance is exactly binomial. This same effect, an increase in variance resulting from polymorphism, has been analyzed before in the context of the infinite-site model (Gillespie and Langley 1979). Note also that the population sizes after the split have no effect on the expectation or variance of $S$. This can be seen by examining the recursion relations for $\Phi_1(t)$ and $\Phi_4(t)$ in Golding’s appendix.

LITERATURE CITED


MASATUSHI NEI, reviewing editor

Received April 3, 1984; revision received April 30, 1984.
Interaction between gene duplication and natural selection in molecular evolution was investigated utilizing a phylogenetic tree constructed by the parsimony procedure from amino acid sequences of 50 calmodulin-family protein members. The 50 sequences, belonging to seven protein lineages related by gene duplication (calmodulin itself, troponin-C, alkali and regulatory light chains of myosin, parvalbumin, intestinal calcium-binding protein, and glial S-100 phenylalanine-rich protein), came from a wide range of eukaryotic taxa and yielded a denser tree (more branch points within each lineage) than in earlier studies. Evidence obtained from the reconstructed pattern of base substitutions and deletions in these ancestral loci suggests that, during the early history of the family, selection acted as a transforming force on expressed genes among the duplicates to encode molecular sites with new or modified functions. In later stages of descent, however, selection was a conserving force that preserved the structures of many coadapted functional sites. Each branch of the family was found to have a unique average tempo of evolutionary change, apparently regulated through functional constraints. Proteins whose functions dictate multiple interaction with several other macromolecules evolved more slowly than those which display fewer protein-protein and protein-ion interactions, e.g., calmodulin and next troponin-C evolved at the slowest average rates, whereas parvalbumin evolved at the fastest. The history of all lineages, however, appears to be characterized by rapid rates of evolutionary change in earlier periods, followed by slower rates in more recent periods. A particularly sharp contrast between such fast and slow rates is found in the evolution of calmodulin, whose rate of change in earlier eukaryotes was manifolds faster than the average rate over the past 1 billion years. In fact, the amino acid replacements in the nascent calmodulin lineage occurred at residue positions that in extant metazoans are largely invariable, lending further support to the Darwinian hypothesis that natural selection is both a creative and a conserving force in molecular evolution.

Introduction

Calmodulin, with its four calcium-binding domains produced by ancient gene duplications, its strongly conserved functions, and its ubiquitous distribution in eukaryotes, is the most generalized member of the family of low-molecular-weight proteins whose functions dictate multiple interaction with several other macromolecules evolved more slowly than those which display fewer protein-protein and protein-ion interactions, e.g., calmodulin and next troponin-C evolved at the slowest average rates, whereas parvalbumin evolved at the fastest. The history of all lineages, however, appears to be characterized by rapid rates of evolutionary change in earlier periods, followed by slower rates in more recent periods. A particularly sharp contrast between such fast and slow rates is found in the evolution of calmodulin, whose rate of change in earlier eukaryotes was manifolds faster than the average rate over the past 1 billion years. In fact, the amino acid replacements in the nascent calmodulin lineage occurred at residue positions that in extant metazoans are largely invariable, lending further support to the Darwinian hypothesis that natural selection is both a creative and a conserving force in molecular evolution.

Introduction

Calmodulin, with its four calcium-binding domains produced by ancient gene duplications, its strongly conserved functions, and its ubiquitous distribution in eukaryotes, is the most generalized member of the family of low-molecular-weight proteins whose functions dictate multiple interaction with several other macromolecules evolved more slowly than those which display fewer protein-protein and protein-ion interactions, e.g., calmodulin and next troponin-C evolved at the slowest average rates, whereas parvalbumin evolved at the fastest. The history of all lineages, however, appears to be characterized by rapid rates of evolutionary change in earlier periods, followed by slower rates in more recent periods. A particularly sharp contrast between such fast and slow rates is found in the evolution of calmodulin, whose rate of change in earlier eukaryotes was manifolds faster than the average rate over the past 1 billion years. In fact, the amino acid replacements in the nascent calmodulin lineage occurred at residue positions that in extant metazoans are largely invariable, lending further support to the Darwinian hypothesis that natural selection is both a creative and a conserving force in molecular evolution.

1. Key words: calmodulin evolution, Darwinian natural selection, maximum parsimony method, rates of molecular evolution.

Address for correspondence and reprints: Dr. Morris Goodman, Room 432, Medical Research Building, Wayne State University School of Medicine, Detroit, Michigan 48201.

© 1984 by The University of Chicago. All rights reserved.
0737-4038/84/0106-0001$02.00
intracellular proteins that mediate the bioregulatory actions of calcium (Goodman et al. 1979; Cheung 1980; Klee et al. 1980; Kretsinger 1980; Watterson et al. 1980). Utilizing many more amino acid sequences from calmodulin and other members of its family than were available for previous studies (Goodman and Pechère 1977; Goodman et al. 1979; Goodman 1980), we have continued to investigate by cladistic methods based on the parsimony principle (Goodman 1981) the evolutionary history of these proteins.

In seeking to understand the mechanisms responsible for diversification of protein branches in the calmodulin family, we agree with Ohno (1970) and others, such as Kimura (1983) and Li and Gojobori (1983), that gene duplications provided a means to tolerate previously "forbidden" mutations. However, rather than explain the emergence of new features by random drift of such mutations, we consider Darwinian selection for favorable mutants in expressed genes among duplicates to be the more important mechanism channeling evolution. Our findings indicate that calmodulin evolved in the stem of the Eucaryota at a fast rate resulting from the accumulation of many adaptive amino acid replacements. In later clades, however, the rate of calmodulin evolution was extremely slow. We suggest that the dramatic shift from fast to slow rates of change demonstrates both the transforming and conserving roles of natural selection.

Material and Methods
Sequences Analyzed


Phylogenetic Tree Construction

The evolutionary assumptions, mathematical foundations, and algorithms employed for finding, from the amino acid sequences, a minimum-length phylogenetic tree and most parsimonious set of ancestral sequences at the interior nodes of that tree have been thoroughly described elsewhere (Moore et al. 1973; Moore 1976;
Goodman et al. 1979; Goodman 1981). Our specific strategy for the tree constructed in the present study, involving aligned amino acid sequences of 50 members of the calmodulin protein family, followed closely that used in previous study of this family (Goodman et al. 1979; Goodman 1980).

The initial computer search for a parsimonious phylogenetic tree was carried out with intact sequences of the 50 proteins. We then aligned the four domain sequences of the seven protein lineages against one another on the basis of sequence identities and, with these 200 sequences (of which some were empty because of deleted domains in certain proteins), renewed the heuristic search for a most parsimonious dendrogram. Among the starting dendrograms were those in which each set of 50 domain sequences followed the same network previously found for the intact sequences. It proved most parsimonious when the four domain branches converged on two penultimate nodes of which one represented the common ancestor of domains I and III and the other the common ancestor of domains II and IV. This parsimony reconstruction using the separate domain sequences placed the root of the seven protein lineages and tallied the NRs between ancestral and descendant nodes.

The Time Scale Employed

The time scale in figure 1 is based on paleontological views on when speciation of the pertinent organisms occurred (Young 1962; Romer 1966; Schopf et al. 1973; Cloud 1974, 1976; Schopf and Oehler 1976; Schopf and Sovietov 1976; Lövtrup 1977; Valentine 1977; Cloud et al. 1980; Cloud and Glaessner 1982; Novacek 1982). Since the proteins from the seven major branches of the calmodulin family all show highly significant sequence similarity by the statistic of Goodman and Moore (1977) and have been found only in eukaryotic species, we assume that the gene duplication (apex of the tree in fig. 1) that produced the first four-domain protein occurred near the origin of eukaryotes 1,500 Myr ago. Nodes, including those depicting gene duplications, which fell between the nodes with dates assigned on the basis of the fossil record were dated by interpolation from the NR values of the internodal links of the path on which they fell. Wherever more than one path of links existed from an ancestral node with an assigned date to later descendant nodes with assigned dates, we always used the most ancient of these later descendant nodes to date the earlier nodes in the older parts of the tree. This procedure ensures that each ancestral node has a more ancient date than any of its descendant nodes.

Results

Phylogeny of Calmodulin Family Proteins

Figure 1 depicts phylogenetic relationships among the seven lineages of calmodulin family proteins established during more than 1 billion years of evolution.

2. Throughout this report, we use the symbol NR to represent nucleotide substitutions estimated via amino acid replacements.

3. That such “local clock” calculations can yield plausible dates is illustrated by those dates obtained utilizing this method—74 Myr and 21 Myr, respectively—in the RLC (regulatory light chain) branch for the ancestral splitting of Pectin marinus (scallop) from Patinopecten yessoensis (giant scallop) and Chlamys nipponensis (shellfish) and then the divergence of the latter two from each other. These dates are similar to those recently communicated to us by Dr. Thomas R. Waller (from his paleontological studies—70 MyrBP and 12–14 MyrBP, respectively). This method also yields implausible dates, however, as is evidenced by the RLC branch. If clock calculations were applied to the ancestral splitting of the RLC Oryctolagus (rabbit) and Gallus (chicken) lineages, the date generated would be about 6 Myr, rather than the 300 Myr date taken from paleontological evidence.
Confirming the results of earlier work (Goodman et al. 1979; Goodman 1980), this phylogenetic reconstruction shows all 50 proteins as originating from a one-domain polypeptide about 36-40 amino acid residues long with the central 12-residue Ca$^{2+}$-binding site followed by two tandem duplications to become first a two-domain protein and then a four-domain calmodulin-like protein in which each domain had the central Ca$^{2+}$-binding loop. As judged by present-day calmodulin and troponin-C, this central loop may have served as a triggering site for Ca$^{2+}$-dependent physiological processes. In the computer-generated ancestral sequence
Fig. 2.—Reconstructed four-domain ancestral sequences delimiting proto and early stages of calmodulin evolution. Row A: the primal four-domain ancestor produced by gene duplication (apex node of fig. 1) and before sequence divergence accumulated either between domains I and III or between domains II and IV. Row B: calmodulin-TNC ancestor of fig. 1 from which calmodulin stem and TNC stem descended. Row C: eukaryotic calmodulin ancestor of fig. 1 from which the ancestral lineage to *Tetrahymena* diverged from that to multicellular eukaryotes. X, Y, Z, \(-Y\), \(-X\), and \(-Z\) represent the six Ca\(^{2+}\)-coordinating ligands in the 12-residue central loop of each domain. Positions with amino acid differences between rows A and C are indicated by overlined residues in row A; positions with amino acid differences between rows B and C are indicated by underlined residues in row C. Of the 148 residue positions of calmodulin, domain I spans positions 6–41, domain II spans 42–78, domain III spans 79–115, and domain IV spans 116–148. Among sequenced extant calmodulins, amino acid differences are found in the 148 alignment at positions 3, 24, 60, 70, 71, 85, 90, 94, 96, 97, 99, 101, 129, 130, 135, 136, 143, 144, 146, and 147; all other positions are unvaried and have the residues shown in row C.
for this primal four-domain protein (fig. 2, row A), domains I and III (the protein's first and third quadrants) appear to have descended from the N-terminal half of the earlier two-domain protein, whereas domains II and IV descended from the C-terminal half. Subsequent duplications in the basal eukaryotes and later prevertebrate metazoans may then have produced the ancestral loci for major protein branches of the calmodulin family (fig. 1). The divergence of lineages within each protein branch conforms closely with phylogenetic data derived from paleontology and comparative morphology.

The most parsimonious branching arrangement for major protein lineages shown in figure 1 was found by requiring that the phylogenies of the taxa of each of the domains be identical with one another. This requirement is tantamount to the assumption that events which cause hybrid genes among duplicated loci, such as gene conversion or unequal homologous crossing-over, occur at a much lower frequency than point mutations. On giving lesser weight to genomic reshuffling events, we found alternative parsimony trees that depict hybrid fusions occurring early in the evolution of the duplicated loci. Nevertheless, these alternative trees still have their two penultimate groupings consisting of domains I + III sequences and domains II + IV sequences. Moreover, we calculated the NR distances between ancestral and descendant nodes on a representative alternative tree and found the same trends in evolutionary rates described below for the tree in figure 1.

Evolutionary Rates

Table 1 lists ancestral nodes from figure 1 with dates assigned from paleontological views and presents rates of evolutionary change along the lines of descent that separate these divergence events. Table 1 also compares average rates of change among the major calmodulin family lineages. These rates are recorded as the number of NRs per 100 codons per 10^8 y (NR%).

Average rates of evolution from the time of ancestral divergence to the present vary widely among the protein groups investigated. Calmodulin displays the slowest overall rate of amino acid replacement since its descent from the eukaryotic ancestor around 1.4 to 1.0 billion years ago (0.5–0.7 NR%). By comparison, the rate of change for TNC during the past 300 Myr is more rapid, averaging 2.7 NR%. The TNC rate is still quite conservative, however, when compared with the rates of evolution found in the alkali and regulatory light chains during the same time period (4.7 and 6.1 NR% for skeletal and cardiac ALC, respectively; 6.4 NR% for cardiac RLC).

Parvalbumins display the most rapid rates of change over the past 400 Myr (average rate of 9.3 NR% for α- and β-parvalbumins); the parvalbumin rate is nearly 40 times faster than the rate of calmodulin evolution for the same time period.

While each protein group has a unique tempo of change, all groups show a general tendency toward deceleration in the rate of amino acid replacement over time. The calmodulin rate decreased from 0.7 NR% in the metazoan stem lineage to the teleost-tetrapod ancestral node—at about 400 Myr—to only 0.2 NR% from this node to present, a 3½-fold decrease. Moreover, the calmodulin rate was 0 NR% from the eutherian ancestral node to present. Similarly, the parvalbumin rate of

4. We have recently become aware of the fact that chicken calmodulin has an amino acid sequence, as deduced from its cDNA nucleotide sequence (Putkey et al. 1983), which is identical to the amino acid sequence of human calmodulin. This further demonstrates that calmodulin evolution became extremely conservative in the higher vertebrates (i.e., in Aves and Mammalia).
Table 1: Variable Evolutionary Rates in Lineages of the Calmodulin Family

<table>
<thead>
<tr>
<th>Lineages</th>
<th>Time Myr (from-to)</th>
<th>Nucleotide Substitutions per 100 Codons per 10⁸ y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calmodulin:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic to mollusc-vertebrate ancestor</td>
<td>1,400/1,000*–680</td>
<td>0.7/1.5</td>
</tr>
<tr>
<td>Eukaryotic ancestor to present</td>
<td>1,400/1,000*–0</td>
<td>0.5/0.7 average</td>
</tr>
<tr>
<td>Mollusc-vertebrate to teleost-tetrapod ancestor</td>
<td>680–400</td>
<td>0.7</td>
</tr>
<tr>
<td>Teleost-tetrapod to eutherian* ancestor</td>
<td>400–85</td>
<td>0.2</td>
</tr>
<tr>
<td>Eutherian* ancestor to present</td>
<td>85–0</td>
<td>0 average</td>
</tr>
<tr>
<td>Troponin C:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetrapod to amniote SK ancestor</td>
<td>340–300</td>
<td>7.8</td>
</tr>
<tr>
<td>Tetrapod SK ancestor to present</td>
<td>340–0</td>
<td>3.2 average</td>
</tr>
<tr>
<td>Amniote to eutherian* SK ancestor</td>
<td>300–75</td>
<td>3.3</td>
</tr>
<tr>
<td>Eutherian* SK ancestor to present</td>
<td>75–0</td>
<td>0.4 average</td>
</tr>
<tr>
<td>Alkali light chain:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amniote SK ancestor to present</td>
<td>300–0</td>
<td>4.7 average</td>
</tr>
<tr>
<td>Amniote to eutherian* CARD ancestor</td>
<td>300–85</td>
<td>2.5</td>
</tr>
<tr>
<td>Amniote CARD ancestor to present</td>
<td>300–0</td>
<td>6.1 average</td>
</tr>
<tr>
<td>Eutherian* CARD ancestor to present</td>
<td>85–0</td>
<td>7.1 average</td>
</tr>
<tr>
<td>Regulatory light chain:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mollusc-vertebrate to amniote SK ancestor</td>
<td>680–300</td>
<td>16.2</td>
</tr>
<tr>
<td>Mollusc-vertebrate to amniote CARD ancestor</td>
<td>680–300</td>
<td>13.5</td>
</tr>
<tr>
<td>Mollusc-vertebrate ancestor to present</td>
<td>680–0</td>
<td>10.0 average</td>
</tr>
<tr>
<td>Amniote SK ancestor to present</td>
<td>300–0</td>
<td>2.0 average</td>
</tr>
<tr>
<td>Amniote to eutherian* CARD ancestor</td>
<td>300–85</td>
<td>6.9</td>
</tr>
<tr>
<td>Amniote CARD ancestor to present</td>
<td>300–0</td>
<td>6.4 average</td>
</tr>
<tr>
<td>Eutherian* CARD ancestor to present</td>
<td>85–0</td>
<td>11.9 average</td>
</tr>
<tr>
<td>Parvalbumin:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gnathostome to teleost-tetrapod β ancestor</td>
<td>425–400</td>
<td>103.6</td>
</tr>
<tr>
<td>Gnathostome β ancestor to present</td>
<td>425–0</td>
<td>14.1 average</td>
</tr>
<tr>
<td>Teleost-tetrapod β ancestor to present</td>
<td>400–0</td>
<td>9.1 average</td>
</tr>
<tr>
<td>Teleost-tetrapod to tetrapod α ancestor</td>
<td>400–340</td>
<td>32.3</td>
</tr>
<tr>
<td>Teleost-tetrapod α ancestor to present</td>
<td>400–0</td>
<td>9.5 average</td>
</tr>
<tr>
<td>Tetrapod α ancestor to present</td>
<td>340–0</td>
<td>7.4 average</td>
</tr>
<tr>
<td>Intestinal calcium-binding protein:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artiodactyl ancestor to present</td>
<td>55–0</td>
<td>15.1 average</td>
</tr>
</tbody>
</table>

* Cloud (1976), Schopf and Oechler (1976), and Schopf and Sovietov (1976) suggest that the divergence of Tetrahymena, Spinacea, and Metazoa occurred about 1–1.4 billion years ago. Dates for the other divergences are those employed previously (Goodman 1981); thus, the eutherian* (Oryctolagus-Homo) ancestral node, from cladistic evidence on a range of proteins, is depicted as a descendant of the eutherian* (Bos-Oryctolagus-Homo) ancestral node.

103.6 NR% in early jawed vertebrates (the fastest rate detected) became 9.1 NR% between 400 Myr and the present, an 11-fold decrease. Such steep decelerations are also apparent in TNC and RLC lineages on comparing rates before the amniote ancestral node at 300 Myr to those after. Such a pattern of rapid evolution during emergence of jawed vertebrates followed by deceleration in the descent of Aves and Mammalia has also been found for myoglobin, α hemoglobin, β hemoglobin, cytochrome c, and lens α-crystallin A (see table 11 in Goodman [1981]).

In the case of calmodulin evolution, we can readily deduce that its period of fastest rates occurred over a billion years ago before the separation of Protozoa from Animalia and Plantae. Suppose there was no such early period of rapid evolution and that, instead, calmodulin had evolved at a uniform rate from the time of its
ancestral divergence from troponin-C to the present. Then, calculations of divergence dates on the basis of the clock model of molecular evolution, with the mollusc-vertebrate divergence node used to set the calmodulin clock, places the last common ancestor of *Tetrahymena*, spinach, and metazoan calmodulins at over 2 billion years ago and the preceding troponin-C-calmodulin ancestral node at over 12 billion years ago. Since this latter date is billions of years earlier than the age of the Earth itself and thus impossibly ancient, we have to conclude that early calmodulin evolution was much more rapid than in lineages to extant eukaryotes. Using the more realistic dates for branch points presented in figure 1 yields a rate of evolution in the pre- and stem-calmodulin lineages which is about 40 times faster than the average rate between the eukaryotic calmodulin ancestor and its present-day descendants (27.1 NR% compared with <0.7 NR%).

Selected Amino Acid Substitutions in Calmodulin

Figure 2 compares the primal four-domain and calmodulin-TNC ancestral sequences to the eukaryotic calmodulin ancestral sequence for amino acid differences over the protein’s 148 residue positions and also identifies (see legend) the residue positions (only 21) that have been found to vary among extant calmodulins. These data show that a substantial proportion of amino acid replacements in the pre- and stem-calmodulin lineage occurred at those positions that are unvaried among *Tetrahymena*, *Spinacea*, and Animalia calmodulins. The proportions are 65 of 74 amino acid differences between the primal four-domain and eukaryotic calmodulin ancestral sequences (rows A and C, fig. 2) and (in the stem alone) 31 of 34 differences between the calmodulin-TNC and eukaryotic ancestral sequences (rows B and C, fig. 2). We conclude that, during the fast evolution of pre- and stem-calmodulin, most amino acid replacements were at positions that later became invariable and that a majority of such substitutions were adaptive and strongly selected. For example, the stretch of residue positions, 70–79, thought to be calmodulin’s interaction site with cyclic nucleotide phosphodiesterase (Kuznicki et al. 1981; Kilhofer et al. 1983), is so far found to be unvaried in the Metazoa, and most of this stretch (positions 72–79) is unvaried in Eucaryota. Yet, our reconstruction (fig. 2) has eight amino acid differences and a deletion between the primal four-domain and eukaryotic calmodulin ancestral sequences.

Discussion

The spectrum of evolutionary tempos displayed by members of the calmodulin family may mirror a variable spectrum of selective restraints related to the unique physiological role of each protein. The marked evolutionary conservatism of calmodulin may be explained by considering calmodulin’s extraordinary functional versatility in regulating the diverse activities of eukaryotic cells. To execute its regulatory role, calmodulin binds sequentially four Ca^{2+} ions to its domains, undergoes conformational changes, and interacts with a wide assortment of enzymes including phosphorylase kinase, cyclic nucleotide phosphodiesterase, adenylate cyclase, ATPase, myosin kinase, glycogen synthase kinase, phospholamban kinase, tryptophan/tyrosine hydroxylase kinase, protein-I kinase, and protein phosphatase-2B (Cohen 1982). Structural-functional constraints should thus be particularly severe—this protein conserves residues involved in the four ion-binding loops, in domain-domain connections, and in the surface configurations that interact with sites on target enzymes.

The protein displaying penultimate evolutionary conservatism within the calmodulin family is troponin-C, which, like calmodulin, is interlocked in a complex
set of macromolecular relationships. TNC is part of a trimer that mediates the interaction of actin and myosin in skeletal muscles. Within this complex, TNC is the key molecule regulating muscle contraction by calcium ions. Following a nerve impulse, TNC binds four Ca\(^{2+}\) ions causing a conformational change that permits interaction between actin and myosin and is ultimately responsible for muscle contraction. To fulfill its role, TNC must interact with both troponin-T and troponin-I in the trimer, as well as bind four Ca\(^{2+}\) ions. Although unlike calmodulin in its localization within skeletal muscle, troponin-C functioning does require multiple protein-protein and protein-ion interactions that may cause evolutionary stasis over a large proportion of the macromolecule's surface.

The most rapidly evolving calcium-modulated protein is parvalbumin, a skeletal muscle polypeptide whose physiological function is less well understood than those of calmodulin and troponin-C. While various functions have been proposed for parvalbumins (Kretsinger 1980; Celio and Heizmann 1982; Kilhoffer et al. 1983) and no single hypothesis has gained widespread acceptance, it seems that this protein is less important to the contractile mechanism than TNC (Kretsinger 1980). Furthermore, parvalbumin is free-floating in the cytosol and is involved in no known protein-membrane or protein-protein interactions. It is also significant that parvalbumin binds only two Ca\(^{2+}\) ions rather than four as in calmodulin and troponin-C. Thus parvalbumin may have fewer selective reins holding back its rate of mutational fixation. Nevertheless, the faster rate of parvalbumin evolution need not be attributed solely to random fixation of neutral mutations but may also reflect selection of adaptive substitutions, i.e., the shaping of less critical but still important new functional sites.

Although the functions of the light chains are not well understood, ALCs and RLCs appear to play a more direct structural role in muscle contraction than parvalbumin (Kretsinger 1980) yet apparently are not involved in as many protein-protein or protein-ion interactions as either calmodulin or troponin-C (e.g., ALCs have completely lost the ability to bind Ca\(^{2+}\), while RLCs bind only one ion). Thus the evidence available suggests that, within the spectrum of macromolecular interactions displayed by the calmodulin family proteins, ALC and RLC may have intermediate properties consistent with their moderate rates of evolutionary change.

The importance of natural selection in causing slow rates of molecular evolution has been recognized by neo-Darwinian selectionists and neutral-mutation theorists alike (Goodman 1976, 1981; Wilson et al. 1977; Kimura 1979). The spectrum of evolutionary tempos displayed by calmodulin family proteins provides further evidence that the strength of conserving selection is related to the density of coadapted functional sites among interacting molecules. The role of natural selection as a transforming agent in shaping the pattern of molecular evolution has been controversial, however. It has been argued that rapid or accelerated rates result from positive natural selection of advantageous mutations that bring about new and coadaptive functional sites within molecules (Goodman 1981; Czelusniak et al. 1982). It has also been argued that rapid change results primarily from relaxation of purifying selection, with selectively neutral amino acid changes occurring at the functionally less important residue positions (Li and Gojobori 1983). As we now discuss, these two views—selectionist and neutralist—are not necessarily conflicting.

In the billion-and-a-half-year-old history of calmodulin, early fast rates of evolution were followed by extremely slow rates. This history clearly fits a selectionist view of protein evolution (Goodman 1981). Nevertheless, calmodulin's history also can be interpreted by a hypothesis in which selectively neutral mutations played a
substantial role. The hypothesis which allows these neutral mutations, however, predicts decelerations in rate of molecular evolution and in that respect is closer to ideas proposed by Goodman (1961, 1963) and restated recently (Goodman et al. 1982; Goodman 1984) than those of Kimura (1968, 1979, 1983). The rate deceleration hypothesis (Goodman 1984) applied to a protein such as calmodulin holds that the rate of selectively neutral amino acid replacement was much greater in the protein's early history when far fewer of its surface molecular sites were engaged in functionally important protein-protein interactions than later in its history after a high density of functional sites had evolved. When the evolution toward higher density was occurring, at some stage in the process, the effects of purifying selection became so pervasive that decelerated rates ensued.

During protein evolution the most extensive accumulations of mutations by random drift probably occurred in earlier phylogenetic periods when the main branches of protein families originated from gene duplications. The redundant cistrons, to paraphrase Ohno (1970), were able to escape from the relentless pressure of natural selection, and, by escaping, they accumulated formerly forbidden mutations. We must reassert here, however, our conviction that Darwinian selection was eventually responsible for the emergence of new gene loci. Without the intervention of selection, mutations in the duplicated loci would sooner or later produce nonfunctional pseudogenes. Of course, replacement of defective portions of a pseudogene by corresponding portions of a functional gene could render the locus functional again, but then this reactivated gene would be much less likely to be neutral to natural selection than before the conversion. If it contained from its pseudogene portion only detrimental substitutions, it would be selected against. But if the substitutions helped encode a new protein with advantageous features, then positive selection would spread the duplicated gene throughout the species in which it occurred. Moreover, in all probability, some of the amino acid replacements encoded by this gene would be far from optimal in the new protein. Thus, most likely, there would be further rounds of positive selection for adaptive amino acid replacements that perfected the protein's functional performance.

Returning to the particular case of calmodulin's history, and to recapitulate, we have concluded that the high functional density of present-day calmodulins did not always exist but was shaped by adaptive amino acid replacements in the basal eukaryotes. Thus a fast rate of replacements in the nascent calmodulin lineage can be attributed, at least in part, to positive selection, whereas the subsequent extremely slow rates can be attributed to strong purifying selection. However, not all of calmodulin's functional sites currently being subjected to such conserving selection were necessarily brought about by positively selected amino acid replacements. For instance, some of the sites engaged in protein-protein interactions at first might have been molded by neutral amino acid replacements and later acquired their specific functional properties after adaptive replacements occurred in other proteins that evolved to interact with calmodulin. But then, the force of positive selection would tend to rapidly spread these later mutations, because random drift had already spread through the species the earlier mutations encoding potential protein-protein interaction sites. Nevertheless, it is apparent that during protein evolution such periods of positive selection tended to be briefer than ensuing periods of purifying selection. We wish to observe, however, that purifying selection can also help transform the course of phylogeny. For example, it seems possible to us that the basal eukaryotic species in which an adaptively advanced calmodulin first evolved had, because of this calmodulin, an advantage over all other eukaryotic
species of that time and thereby became the stem of a vast radiation of eukaryotic life.

Acknowledgments

The authors wish to thank Judith E. Beeber for technical assistance in assembly of data, Drs. D. Martin Watterson and Anthony R. Means for providing us with calmodulin sequences before publication, and Walter M. Fitch for stimulating discussion. This study was supported by National Science Foundation grant DEB 78-10717.

APPENDIX

Species List of Sequences in the Calmodulin Family Analyzed by the Maximum-Parsimony Method

<table>
<thead>
<tr>
<th>Protein and Species</th>
<th>Common Name</th>
<th>Protein and Species</th>
<th>Common Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troponin-C (TNC):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>Human (SK)</td>
<td>Pectin marinus</td>
<td>Scallop (EDTA)</td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Rabbit (SK)</td>
<td>Gallus gallus</td>
<td>Chicken gizzard</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Chicken (SK)</td>
<td>Oryctolagus</td>
<td></td>
</tr>
<tr>
<td>Rana esculenta</td>
<td>Frog (SK)</td>
<td>Gallus gallus</td>
<td></td>
</tr>
<tr>
<td>Bos taurus</td>
<td>Beef (CARD)</td>
<td>Homo sapiens</td>
<td></td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Rabbit (CARD)</td>
<td>Bos taurus</td>
<td>Bovine (CARD DTNB)</td>
</tr>
<tr>
<td>Calmodulin (CAL):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetrahymena</td>
<td>Tetrahymena</td>
<td>Gallus gallus</td>
<td>Chicken (CARD DTNB) L2B</td>
</tr>
<tr>
<td>Spinacea oleracea</td>
<td>Spinach</td>
<td>Gallus gallus</td>
<td>Chicken (CARD DTNB) L2A</td>
</tr>
<tr>
<td>Renilla reniformes</td>
<td>Renilla</td>
<td>Gallus gallus</td>
<td></td>
</tr>
<tr>
<td>Pectin marinus</td>
<td>Scallop</td>
<td>Parvalbumin:</td>
<td></td>
</tr>
<tr>
<td>Metridium senile</td>
<td>Sea anemone</td>
<td>Latimeria</td>
<td></td>
</tr>
<tr>
<td>Electrophorus</td>
<td></td>
<td>cchalamnae</td>
<td>Coelacanth α</td>
</tr>
<tr>
<td>Bovine</td>
<td>Eel</td>
<td>Esox lucius</td>
<td>Pike α</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>Human</td>
<td>Oryctolagus</td>
<td></td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Rabbit</td>
<td>cchuniculus</td>
<td>Rabbit α</td>
</tr>
<tr>
<td>Alkali light:</td>
<td></td>
<td>Rana esculenta</td>
<td>Frog α</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Chicken gizzard</td>
<td>Latimeria</td>
<td></td>
</tr>
<tr>
<td>Chains (ALC):</td>
<td></td>
<td>cchalamnae</td>
<td>Coelacanth β</td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Rabbit (SK) 1</td>
<td>Rana esculenta</td>
<td>Frog β</td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>Rabbit (SK) 2</td>
<td>Merlucius</td>
<td>Hake β</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Chicken (SK) 2</td>
<td>Gadus callarias</td>
<td>Cod β</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Chicken (SK) 1</td>
<td>Esox lucius</td>
<td>Pike β</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Chicken (CARD)</td>
<td>Gadus merlangus</td>
<td>Whiting β, chub β</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>Human (CARD)</td>
<td>Cyprinus carpio</td>
<td>Carp β</td>
</tr>
<tr>
<td>Bos taurus</td>
<td>Bovine (CARD)</td>
<td>Cyprinus carpio</td>
<td>Carp β</td>
</tr>
<tr>
<td>Regulatory light:</td>
<td></td>
<td>PAPI-B:</td>
<td>Bovine PAPI-B</td>
</tr>
<tr>
<td>Patinopecten yessoensis</td>
<td>Giant scallop (EDTA)</td>
<td>Bos taurus</td>
<td></td>
</tr>
<tr>
<td>Chlamys nipponensis</td>
<td>Shellfish (EDTA)</td>
<td>ICaBP:</td>
<td>Pig ICaBP</td>
</tr>
</tbody>
</table>

NOTE.—The sequences listed above are referenced in the text, in Goodman and Pechère (1977), Goodman et al. (1979), and Goodman (1980).
LITERATURE CITED


WALTER M. FITCH, reviewing editor

Received October 17, 1983; revision received April 19, 1984.
Nucleotide Sequence of the *Acinetobacter calcoaceticus trpGDC* Gene Cluster

Jeffrey B. Kaplan, Paul Goncharoff, Anita M. Seibold, and Brian P. Nichols

Department of Biological Sciences, University of Illinois at Chicago

A plasmid library of *Acinetobacter calcoaceticus* HindIII fragments was constructed, and clones that complemented an *Escherichia coli pabA* mutant were selected. Plasmids containing a 3.9-kb fragment of *A. calcoaceticus* DNA that also complemented *E. coli trpD* and *trpC−(trpF+)* mutants were obtained. We infer that complementation of *E. coli pabA* mutants was the result of the expression of the amphibolic anthranilate-synthase/p-aminobenzoate-synthase glutamine-amidotransferase gene and that the plasmid insert carried the entire *trpGDC* gene cluster. In *E. coli* minicells, the plasmid insert directed the synthesis of polypeptides of 44,000, 33,000, and 20,000 daltons, molecular masses that are consistent with the reported molecular masses of phosphoribosylanthranilate transferase, indoleglycerol-phosphate synthase, and anthranilate-synthase component II, respectively. A 3,105-bp nucleotide sequence was determined. Comparison of the *A. calcoaceticus trpGDC* sequences with other known *trp* gene sequences has allowed insight into (1) the evolution of the amphibolic *trpG* gene, (2) varied strategies for coordinate expression of *trp* genes, and (3) mechanisms of gene fusions in the *trp* operon.

Introduction

In all organisms investigated to date, ranging from gram-negative prokaryotes to lower-level eukaryotes, the five reactions that convert chorismate to tryptophan have not been found to differ, but the structure and organization of the genes involved in tryptophan biosynthesis vary considerably (Crawford 1975). *Acinetobacter calcoaceticus* contains seven separate genes that encode the enzymes of tryptophan biosynthesis (Twarog and Liggins 1970; Sawula and Crawford 1973), and these genes lie in three unlinked clusters (Sawula and Crawford 1972). The *Escherichia coli trp* operon, on the other hand, contains five contiguous genes encoding the tryptophan-biosynthetic enzymes (Yanofsky et al. 1981). Two gene fusions have occurred in the evolution of the *E. coli trp* operon, with the result that two of the *E. coli* genes encode bifunctional polypeptides (Creighton 1970; Greishaber and Bauerle 1972).

In certain organisms (*A. calcoaceticus* [Sawula and Crawford 1972], *Bacillus subtilis* [Kane et al. 1972], and *Pseudomonas acidovorans* [Buvinger et al. 1981]), tryptophan biosynthesis is “interlocked” with the biosynthesis of the vitamin dihydrofolate because the glutamine amidotransferase (GAT) subunit of anthranilate synthase (AS; EC 4.1.3.27) is amphibolic, that is, it also acts as the GAT subunit of

1. Key words: tryptophan genes, *Acinetobacter calcoaceticus*, gene rearrangements, gene fusions.

Address for correspondence and reprints: Brian P. Nichols, Department of Biological Sciences, University of Illinois at Chicago, Box 4348, Chicago, Illinois 60680.
Acinetobacter \textit{trpGDC} Sequence

A second enzyme, \textit{\textit{p}}-aminobenzoate synthase (PABS). AS and PABS use identical substrates, chorismate and glutamine, and produce very similar products—glutamate, pyruvate, and either anthranilate (\textit{o}-aminobenzoate) or \textit{\textit{p}}-aminobenzoate (PABA), the latter being incorporated into dihydrofolate.

We report here the nucleotide sequence of the \textit{\textit{A}}. \textit{calcoaceticus trpGDC} gene cluster. \textit{TrpG} encodes the amphibolic GAT subunit of AS and PABS. Since the nucleotide and amino acid sequences of genes encoding the pathway-specific GAT subunits—\textit{pabA}, \textit{trpG}, and \textit{trp(G)D}—in several enterobacterial species are known (Nichols et al. 1980; Tso et al. 1980; Kaplan and Nichols 1983), determination of the primary structure of an amphibolic subunit might assist in the elucidation of the functional and evolutionary origins of pathway-specific genes. \textit{TrpD} and \textit{trpC} encode anthranilate phosphoribosyltransferase (PRTase; EC 2.4.2.18) and indole-glycerol phosphate synthase (InGPS; EC 4.1.1.48), respectively. In several species of the Enterobacteriaceae, these activities lie on bifunctional polypeptides. In \textit{\textit{E}}. \textit{coli}, for example, PRTase and AS GAT activities are encoded by a single gene, \textit{\textit{trpD}} (Greishaber and Bauerle 1972; Miozzari and Yanofsky 1979). For clarity, we will follow the suggestion of Crawford (1975) and refer to the fused gene as \textit{\textit{trp(G)D}}. Similarly, in \textit{\textit{E}}. \textit{coli} InGPS lies on the same polypeptide with phosphoribosylanthranilate isomerase (PRAI) and is encoded by \textit{\textit{trpC}} (Creighton 1970). PRAI is encoded by \textit{\textit{trpF}} in \textit{\textit{A}}. \textit{calcoaceticus}, and we will refer to the fused \textit{\textit{E}}. \textit{coli} gene as \textit{\textit{trpC(F)}}. Comparisons of the separate and fused gene sequences can aid in the determination of the minimum amount of information necessary to encode each individual function as well as assist in the elucidation of events that may have resulted in fused genes.

Materials and Methods

Bacteriophages, Strains, and Strain Construction

The bacterial strains used in this work are listed in table 1. Bacteriophages \textit{\textit{M13mp8}} and \textit{\textit{M13mp9}} are those described by Messing and Vieira (1982) and were obtained from New England Biolabs (Beverly, Mass.). Bacteriophage \textit{\textit{P1kc}} was obtained from C. Yanofsky, and transductions were performed as described by Miller (1972). \textit{\textit{Escherichia coli BN100}} was constructed by transducing a \textit{\textit{P1kc}} lysate made from \textit{\textit{E}}. \textit{coli AB3292} (Huang and Pittard 1967) into \textit{\textit{E}}. \textit{coli W3110 \Delta\textit{trpLD102}} and selecting streptomycin-resistant colonies. Resistant colonies were then screened for PABA auxotrophy. \textit{\textit{Escherichia coli BN105 (\textit{\textit{trpR}}, \textit{\textit{minA}}, \textit{\textit{minB}}, \textit{\textit{leu}}, \textit{\textit{thi}}, \textit{\textit{lacY}}, \textit{\textit{ara}}, \textit{\textit{gal}}, \textit{\textit{malA}}, \textit{\textit{xyl}}, \textit{\textit{azi}}, \textit{\textit{tsx}}), tonA, and \textit{\textit{rpsL}})} was constructed by transduction of a \textit{\textit{P1kc}} lysate made from \textit{\textit{E}}. \textit{coli W3110 \Delta\textit{trpEA2 ina2}} into \textit{\textit{E}}. \textit{coli P678-54} and isolation of 5-methyltryptophan-resistant, threonine-independent colonies.

\textit{\textit{DNA manipulations.}}—Restriction endonucleases were either prepared in this laboratory according to published procedures or purchased from commercial distributors (New England Biolabs; P-L Biochemicals, Milwaukee; and Boehringer Mannheim Biochemicals, Indianapolis). Calf intestinal phosphatase was obtained from Boehringer Mannheim. Restriction endonuclease digests were performed according to the commercial supplier's recommendations. Ligations were performed at DNA concentrations of 30–60 \textit{\textmu}g/ml in 10 mM Tris–HCl (pH 7.4), 10 mM \textit{\textit{MgCl}}\textit{\textsubscript{2}}, 10 mM 2-mercaptoethanol, and 0.5 mM ATP for 1–4 h at 14 C. Transformation of \textit{\textit{E}}. \textit{coli} was carried out as described by Mandel and Higa (1970). \textit{Acinetobacter calcoaceticus} DNA was prepared according to Marmur (1966), except
Table 1
Bacterial Strains Cited

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Reference or Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acinetobacter calcoaceticus BD413</td>
<td>trpE27, ura</td>
<td>Crawford 1975</td>
</tr>
<tr>
<td>Escherichia coli AB3292</td>
<td>pabA1, rpsL704, proA2, argE3, lacY1, his-4, ilvC7, thi-1, galK2, xyl-5, met-1, tfi-3, tsx-358, supE44</td>
<td>Huang and Pittard 1967</td>
</tr>
<tr>
<td>E. coli BN100</td>
<td>pabA1, ΔtrpLD102, rpsL704</td>
<td>This work</td>
</tr>
<tr>
<td>E. coli BN101</td>
<td>pabA1, trpA33, rpsL704</td>
<td>Kaplan and Nichols 1983</td>
</tr>
<tr>
<td>E. coli p678-54</td>
<td>thr, ara, leu, azi, tonA, lacY, tsx, minA, minB, gal, malA, thi, xyl, rpsL</td>
<td>Matsumura et al. 1977</td>
</tr>
<tr>
<td>E. coli W3110</td>
<td>trpR, ΔtrpEA2, tna2</td>
<td>Yanofsky et al. 1981</td>
</tr>
<tr>
<td>E. coli BN105</td>
<td>trpR, ara, leu, azi, tonA, lacY, tsx, minA, minB, gal, malA, thi, xyl, rpsL</td>
<td>This work</td>
</tr>
<tr>
<td>E. coli W3110 trpC55</td>
<td>trpC, (trpF+)</td>
<td>Yanofsky et al. 1971</td>
</tr>
<tr>
<td>E. coli LE392</td>
<td>trpR55, metB1, hsdR514, galK2, galT22, lacY1, supE44, supF58</td>
<td>Enquist et al. 1979</td>
</tr>
<tr>
<td>E. coli JM103</td>
<td>Δ(lac-pro), thi, strA, supE, endA, sbc15, hsdR4/1, traD36, proAB, lacI*, ΔlacZM15</td>
<td>Messing et al. 1981</td>
</tr>
</tbody>
</table>

that DNA precipitated in ethanol was collected by centrifugation. Plasmid DNA was prepared by the method of Birnboim and Doly (1979).

Plasmid constructions.—Acinetobacter calcoaceticus DNA was prepared, and a plasmid library was constructed by ligating 20 µg of HindIII-digested A. calcoaceticus DNA with 10 µg of pBR322 treated with HindIII and phosphatase. A portion of the ligated DNA representing 3 µg of vector was transformed into E. coli LE392 (r^-m^+), and ampicillin-resistant colonies were selected. Approximately 7,000 ampicillin-resistant colonies were pooled, and the plasmid DNA was isolated. Two micrograms of the amplified plasmid library was transformed into E. coli BN100, and ampicillin-resistant, PABA-independent colonies were isolated.

Protein synthesis in E. coli minicells.—Plasmids containing A. calcoaceticus trp genes were transformed into E. coli BN105, and ampicillin-resistant colonies were isolated. Isolation of minicells and labeling of proteins with 35S-methionine (300 Ci/mm; Amersham, Arlington Heights, Ill.) were performed as described by Matsumura et al. (1977).

DNA-sequence analysis.—DNA sequences were determined by the procedure of Maxam and Gilbert (1980) or Sanger et al. (1977). DNA fragments were end-labeled either with T4-polynucleotide kinase (P-L) and (γ-32P)ATP (>2,000 Ci/mm; Amersham) or with E. coli DNA polymerase I (Klenow fragment;
Boehringer Mannheim) and (α-32P)dCTP (>800 Ci/mmol; Amersham) (Maxam and Gilbert 1980). Alternatively, restriction endonuclease–generated DNA fragments were cloned into M13mp8 or M13mp9 RF I. Recombinant phages were transformed into *E. coli* JM103, and single-stranded DNA was prepared as described by Messing and Vieira (1982). The polyacrylamide gel/urea electrophoresis system described by Sanger and Coulson (1978) was used for resolving DNA fragments. DNA sequences were analyzed in part by computer programs (Queen and Korn 1980; S. Weaver, B. Nichols, and C. Hutchison III, unpublished).

**Results**

**Construction and Characterization of Recombinant Plasmids containing the *Acinetobacter calcoaceticus trpGDC* Gene Cluster**

Plasmid libraries of *A. calcoaceticus* DNA were transformed into BN101 and selected for ampicillin resistance and PABA independence. Approximately 50 such colonies were obtained. Plasmid DNA prepared from six of the PABA-independent colonies obtained from the HindIII library was cleaved with HindIII. Agarose-gel electrophoresis showed that, in addition to the 4.3-kb pBR322-HindIII fragment, all of the plasmids contained a 3.9-kb HindIII fragment. Four of the six plasmids contained a small additional HindIII fragment. Retransformation of *Escherichia coli* BN101 with the two smaller plasmids resulted in a high proportion of PABA-independent colonies. One plasmid, designated pBN79, was selected for further characterization.

Plasmid pBN79 was mapped with several restriction endonucleases. The sites are indicated in figure 1. Since SalI cleaved the *A. calcoaceticus* DNA into approximately equal parts, the two SalI-HindIII DNA fragments were cloned separately into the SalI and HindIII sites of pBR322. Each was then transformed into *E. coli* BN101, and ampicillin-resistant colonies were tested for PABA independence. The plasmid containing the larger (2,350-bp) HindIII-SalI fragment (pBN78) was found to complement the *E. coli* pabA mutation, whereas the plasmid containing the smaller (1,750-bp) fragment (pMB466) did not.

It has been shown that the product of the *A. calcoaceticus trpG* gene, that is, AS Component II (AS Coll), does not complement an enterobacterial AS Component

![Fig. 1.—Physical and genetic maps of the *Acinetobacter calcoaceticus trpGDC* region including (a) genetic map of pBN79 illustrating the 3.9-kb HindIII fragment (light line) in pBR322 (heavy lines) with arrows indicating the position and direction of transcription of each gene, (b) restriction endonuclease map of the HindIII fragment in pBN79, (c) HindIII SalI fragments contained in plasmids pBN78 and pMB466, and (d) extent of the sequence analysis of the HindIII DNA fragment.](image)
I (AS CoI) (Sawula and Crawford 1973). Therefore, in order to confirm that transformation to PABA independence was the result of cloning the \textit{A. calcoaceticus} \textit{trpG} gene and not the result of anomalous expression of a normally cryptic \textit{pabA} gene, complementation tests were performed to test for the presence of other \textit{trp} genes known to be linked to \textit{A. calcoaceticus} \textit{trpG}.

Complementation tests for the presence of \textit{A. calcoaceticus} \textit{trpD} were performed by transforming the plasmids into \textit{E. coli} BN100. The tryptophan requirement of this strain cannot be circumvented by anthranilate unless the PRTase (\textit{trpD}) function is supplied by the plasmid. Only pBN79 gives rise to colonies on medium containing anthranilate and PABA. Since neither pBN78 nor pMB466 is \textit{trpD}+, separation of the \textit{HindIII} fragment at the \textit{SalI} site must either interrupt the \textit{A. calcoaceticus} \textit{trpD} gene or cause the loss of its expression in some other manner.

The test for complementation by \textit{A. calcoaceticus} \textit{trpC} was carried out in \textit{E. coli} W3110 \textit{trpC55}, a strain that contains a missense mutation in the first half of \textit{trpC}(F). This mutation causes the loss of InGPS activity without loss of PRAI activity. The strain harboring either pBN79 or pMB466 can grow on minimal medium, indicating the presence of \textit{A. calcoaceticus} \textit{trpC} on both of these plasmids. Plasmid pBN78 does not confer the ability to grow on minimal medium. The \textit{trpC} gene of the \textit{A. calcoaceticus} \textit{trpGDC} cluster must therefore be located to the right of the \textit{SalI} site (as indicated in fig. 1).

Growth-rate experiments indicate that \textit{A. calcoaceticus} \textit{trpG} and \textit{trpD} complement \textit{E. coli} \textit{pabA} and \textit{trpD} mutants as effectively as exogenous PABA or tryptophan (data not shown). \textit{Acinetobacter calcoaceticus} \textit{trpC}, on the other hand, complements the \textit{E. coli} \textit{trpC55} mutation slightly less effectively than exogenous tryptophan. These complementation data show that PABA independence is conferred on an \textit{E. coli} \textit{pabA} mutant by an \textit{A. calcoaceticus} \textit{trpG} gene that lies in or near the \textit{trpGDC} cluster on a 3.9-kb \textit{HindIII} DNA fragment. Since only three polypeptides are synthesized from this DNA fragment (see below), we conclude that PABA independence is conferred by \textit{trpG}. In addition, the data indicate that \textit{trpG} and \textit{trpC} are separated by a \textit{SalI} site that inactivates \textit{trpD}.

Analysis of Polypeptides Encoded by the \textit{A. calcoaceticus} DNA Fragment

The polypeptides encoded by the three plasmids were analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis of \textsuperscript{35}S methionine–labeled proteins produced by \textit{E. coli} minicells. Plasmids pBN79, pBN78, and pMB466 were transformed into \textit{E. coli} BN105, a minicell-producing strain containing a defective \textit{trp} aporepressor. The results are shown in figure 2. Plasmid pBN79 efficiently directs the synthesis of three major polypeptides in addition to the pBR322-encoded \textit{\beta}-lactamase. The molecular masses of 44,000, 33,000, and 20,000 daltons are consistent with the molecular masses of 37,543, 30,203, and 21,779 daltons predicted from DNA sequence analysis (see below). Molecular masses estimated by gel permeation chromatography for PRAI, InGPS, and \textit{AS CoII} were 42,500, 30,000, and 14,000, respectively (Twarog and Liggins 1970; Sawula and Crawford 1973). The difference in the estimates for \textit{AS CoII} may be due to the different analytical methods employed. The origin of the faint bands beneath the largest polypeptide is unknown but may be a result of artifactual translation initiation at internal methionine codons of \textit{trpD}.
Plasmids pBN78 and pMB466 each encode one major polypeptide in addition to β-lactamase, and the results are consistent with the complementation data described above. Plasmid pBN78 directs the synthesis of only the smallest polypeptide, that is, the trpG-gene product, and pMB466 directs the synthesis of the 30,000-dalton trpC-gene product. The largest polypeptide, the product of trpD, is absent from both the pBN78 and pMB466 lanes, confirming that cleavage at the SalI site inactivates the gene.

Nucleotide-Sequence Determination of the A. calcoaceticus trpGDC-Gene Cluster

The nucleotide sequence of 3,105 bp containing the A. calcoaceticus trpGDC gene cluster was determined by the methods of Sanger et al. (1977) and Maxam and Gilbert (1980). Restriction fragments from pBN78 and pMB466 were either end-labeled (Maxam and Gilbert procedure) or cloned into M13mp8 or M13mp9 (Sanger procedure). One restriction fragment from pBN79 was used to determine the sequence across the SalI site. All bases were determined on both strands of the DNA, and, with one exception, all restriction sites used for sequencing were overlapped. The exception was the HpaII site in trpG. We are confident that no information was lost, since this site occurs within a region of the DNA sequence that is highly conserved in several related GAT sequences. The nucleotide and the predicted amino acid sequences of the trpGDC gene cluster are presented in figure 3.

The three trp genes lie in proximity on the chromosome. Only three nucleotides (a termination codon) lie between trpG and trpD, and 13 nucleotides lie between trpD and trpC. The insufficient number of nucleotides for individual regulatory sequences between the genes suggests a polycistronic transcriptional unit, a suggestion that is in agreement with previous reports of the coordinate regulation of the trpGDC cluster (Cohn and Crawford 1976). The adjacent termination and initiation codons of A. calcoaceticus trpG and trpD are identical with those found between

---

**Fig. 2.** Autoradiogram of an SDS-polyacrylamide gel showing 35S methionine-labeled proteins synthesized from Escherichia coli minicell plasmids pBN79 (lane a), pBN78 (lane b), and pMB466 (lane c).
**FIG. 3.**—DNA and predicted amino acid sequences of the trpGDC gene cluster and flanking regions.

*E. coli* **thrB** and **thrC** (Cossart et al. 1981) and are reminiscent of the overlapping termination and initiation codons observed in the *E. coli* **trp** operon between **trpE** and **trp(G)D** (Nichols et al. 1980) and between **trpB** and **trpA** (Platt and Yanofsky 1975; Nichols and Yanofsky 1979). The overlapping termination and initiation codons in the *E. coli* **trp** operon have been implicated in the coordinate translation of the adjacent genes (Oppenheim and Yanofsky 1980; Aksoy et al. 1984), and it is possible that the adjacent termination and initiation codons play a similar role in the coordinate expression of *A. calcoaceticus trpG* and **trpD**. The three **trp** genes can be easily recognized by their similarity to both *E. coli* **trp(G)D** and the S-terminal portion of *E. coli trpC(F)*. Figure 4 shows the similarity between the aligned *A. calcoaceticus* and *E. coli* nucleotide sequences. The sequence similarity is evident along the entire length of the gene cluster and is strongest within the coding regions of the genes. Similarities are notably reduced around intercistronic regions and at the beginning and end of the *A. calcoaceticus* gene cluster. The intermittent nature of strong similarity suggests conservation of structurally and functionally critical amino acid sequences and divergence of expression-related signals and less important amino acid sequences.

**Amino Acid-Sequence Alignments**

Comparisons of the predicted amino acid sequences of *A. calcoaceticus* AS CoII, PRTase, and InGPS with the homologous sequences from other organisms are shown in figures 5, 6, and 7, respectively. Alignments are manual, and gaps have been inserted to increase similarity among the sequences. The three *A. calcoaceticus* amino acid sequences show 36%-39% similarity to the *E. coli* amino acid sequences (AS CoII, 37%; PRTase, 36%; InGPS, 39%), however, *A. calcoaceticus* AS CoII is much more similar to the *E. coli* PABS CoII (57%) than to AS CoII. The higher similarity may explain why *A. calcoaceticus trpG* functions in *E. coli* PABA synthesis but does not function in anthranilate synthesis.
FIG. 4.—Nucleotide-sequence similarity between *Acinetobacter calcoaceticus* trpGDC and the equivalent portion of the *Escherichia coli* trp operon. Each point represents the similarity between 40-nucleotide-long sequences. The positions of the *A. calcoaceticus* genes are indicated by the arrows along the horizontal axis.

**Discussion**

**Comparison of Amino Acid Sequences**

Alignment of the three AS GAT amino acid sequences with the *Escherichia coli* PABS GAT sequence underscores the common origin of the AS GAT and PABS GAT subunits (Kaplan and Nichols 1983). Strong similarity occurs in several regions throughout the sequences. One of these regions, at residues 74–83, contains a cysteine residue (position 79) that has been determined to be essential for catalytic activity. A glutamine analog covalently reacts with this cysteine residue in the *Pseudomonas putida* (Kawamura et al. 1978) and *Serratia marcescens* (Tso et al. 1980) AS GAT subunits. Another region, centered in the area of residue 103, is rather less conserved among the sequences but contains a lysine residue shown to be essential for activity (Bower and Zalkin 1982). Arginine replaces the essential lysine residue in the *Acinetobacter calcoaceticus* sequence. The specific roles of the other conserved sequences are not known, but their conservation across generic boundaries suggests critical roles in either substrate binding or catalysis.

The *A. calcoaceticus* trpG product is most closely related to the *P. putida* trpG product and least related to the *E. coli* trp(G) product. These relationships correlate...
Acinetobacter trpGDC Sequence

**Fig. 6.**—Alignment of the predicted amino acid sequences of *Acinetobacter calcoaceticus* anthranilate phosphoribosyltransferase (PRTase) (top row) and *Escherichia coli* PRTase (Horowitz et al. 1982) (bottom row). The *E. coli* sequence shown begins at residue 198 of the bifunctional glutamine amidotransferase PRTase polypeptide. Numbering refers to the *A. calcoaceticus* sequence.

with intragenic AS subunit exchange experiments. The *A. calcoaceticus* and *P. putida* GAT subunits can be exchanged without appreciable loss of function, but neither of the AS CoI subunits can function with an enterobacterial AS CoII (Sawula and Crawford 1973; Queener et al. 1973); however, at least two (*A. calcoaceticus* AS CoI and *E. coli* PABS CoII) of the three most similar sequences shown in figure 5 do function in PABA synthesis with *E. coli* PABS CoI in vivo. Because of the observed sequence similarity between *A. calcoaceticus* trpG and *P. putida* trpG,

**Fig. 7.**—Alignment of the predicted amino acid sequences of *Acinetobacter calcoaceticus* indoleglycerol phosphate synthase (InGPS) (top row) and *Escherichia coli* InGPS (Christie et al. 1978) (bottom row). Only the N-terminal portion of the fused *E. coli* InGPS—phosphoribosylanthranilate isomerase sequence is shown. Numbering refers to the *A. calcoaceticus* sequence.
it would not be surprising to find that *P. putida* AS CoII functioned similarly, although it has not yet been determined whether *P. putida* trpG encodes an amphibolic GAT subunit.

It is not obvious from inspection of the amino acid sequences why the *A. calcoaceticus* GAT subunit does not function amphibolically in *E. coli*. Evidence suggests that the N-terminal portion of *E. coli* AS CoII interacts with the AS CoI subunit (Greishaber and Bauerle 1972). Whether the same is true of the PABS subunit is unknown, however, if most of the remainder of the sequence is conserved for other functions, then by analogy with AS CoII, it is likely that the N-terminal region of the PABS GAT subunit also specifies interaction with the large PABS subunit. Since the *A. calcoaceticus* GAT cannot complement the *E. coli* AS CoI but behaves instead as a PABS-specific GAT subunit, it is likely that the amino acid differences observed in *E. coli* trp(G) in this region account in part for the functional differences in subunit specificity. Subunit-specificity determinants would also be expected to occur on the CoI subunits of AS and PABS, and according to observed complementation patterns, it is likely that in the subunit-interaction areas *A. calcoaceticus* AS CoI will show similarities to *A. calcoaceticus* PABS CoI that are not found between *E. coli* AS CoI and *E. coli* PABS CoI.

The predicted amino acid sequences of *A. calcoaceticus* and *E. coli* PRTase are compared in figure 6. The *E. coli* sequence shown begins within the fusion region between the trp(G) and trpD portions of the sequence. (The monofunctional PRTase-encoding sequence of *S. marcescens* begins at the position equivalent to *A. calcoaceticus* residue 3.) Similarity is quite strong in the central region of the polypeptide, whereas the N- and C-terminal portions show less similarity. The diversity at the N and C termini may be due to either a structural or a functional adaptation of the *E. coli* sequence to the fusion of trp(G) in conjunction with the evolution of the internal *E. coli* trpP, promoter (Morse and Yanofsky 1968; Horowitz et al. 1982; Horowitz and Platt 1982) or to an entirely different origin of this DNA region.

The amino acid sequence of the *A. calcoaceticus* trpC product aligns with the N-terminal portion of the fused *E. coli* trpC(F) product (Christie and Platt 1980). As in the PRTase sequences, the extent of similarity is strongest in the central portion of the sequence and is somewhat lower toward the termini. The product of the *E. coli* trpC(F) gene is 450 amino acids long and catalyzes both the InGPS and PRAI activities. Analysis of proteolytic fragments has determined that the two activities lie on independently folding domains and that the InGPS activity is confined to the 289 N-terminal residues (Kirschner et al. 1980). Comparison of the *E. coli* trpC(F) sequence with the monofunctional *A. calcoaceticus* trpC sequence suggests that the InGPS activity is encoded within the first 260 amino acid residues and that the domain containing PRAI activity lies in the remaining 190 amino acids. *Sacharomyces cerevisiae* TRP1 (equivalent to trpF) encodes InGPS on a separate polypeptide chain. The first recognizable similarity to the *E. coli* sequence occurs at positions 15–19 of TRP1, matching *E. coli* positions 266–271 as numbered in figure 7 (Tschumper and Carbon 1980). This latter sequence, KVCGLT, also overlaps the last three amino acids of the *A. calcoaceticus* InGPS sequence. While it has been shown that the *E. coli* trp(G)D fusion is the result of modification and translation through an ancestral intercistronic region (Miozzari and Yanofsky 1979), this is probably not the same mechanism that resulted in the fusion of trpC and trpF. We suggest instead that the trpC-trpF fusion occurred via a deletion between
separate but adjacent trpC and trpF genes with end points within each of the ancestral genes. This hypothesis is supported by the observation that the “fusion peptide” between the GAT and PRPase activities lies near the surface of the molecule and is accessible to proteases (Li et al. 1974), whereas the junction between the fused InGPS and PRAI activities is buried within the protein core (Kirschner et al. 1980).

Codon and Amino Acid Usage in the A. calcoaceticus trpGDC Gene Cluster

The choice of specific codons within a codon family differs between A. calcoaceticus trpGDC and the equivalent region of the E. coli trp operon (table 2). As seen in many prokaryotes, there is an avoidance of the use of such codons as ATA(Ile), CTA(Leu), CGR(Arg), and AGR(Arg) (Grantham et al. 1981). These preferences correlate with insignificant amounts of the cognate tRNAs in E. coli (Ikemura 1981), and it may be that the same is true in other prokaryotes. Other A. calcoaceticus codon usage trends that are similar to those of E. coli are preferred use of ATT (Ile), AAA (Lys), and GGY (Gly). Among the notable differences in A. calcoaceticus codon usage, however, are the following: (1) CTG (Leu) is not the preferred Leu codon; rather, TTA and TTG represent one-half of the Leu codons; (2) TCT and TCA (Ser) are preferred to TCG; (3) CCT (Pro) is preferred to CCG; and (4) in all of the two-codon families, the member ending in T or A is used from two to six times as often as is the alternative.

It has been observed that within the trp genes of the Enterobacteriaceae, codon usage is influenced by the genomic G + C content of the organism and that the third positions of the codons are the most influenced (Nichols et al. 1980, 1981). The same effect is seen in the A. calcoaceticus trpGDC gene cluster, albeit in the opposite direction: the G + C content of A. calcoaceticus is 42%, but in the third positions of the codons it is only 35%. As mentioned above, this is very noticeable in the two-codon families, but it is also obvious in some of the four-codon families (e.g., Ser and Pro).

In addition to the influence exerted on the third position of the codon, the genomic G + C content may also affect the choice of amino acids. That is, within a family of amino acids, those that are encoded by A/T-rich codons are used more frequently in A. calcoaceticus. For example, in the (aliphatic) hydrophobic class of amino acids, Ile (ATH) is used more frequently and Ala (GCX) used less frequently than in E. coli. A similar statement can be made concerning the choice between the basic amino acids Lys (AAR) and Arg (CGX + AGR). The true significance of this observation awaits further documentation of the amino acid composition of orthologous proteins from additional organisms with varied genomic G + C content.

Conclusion

The nucleotide sequence of the A. calcoaceticus trpGDC gene cluster emphasizes the fundamental similarity of the enzymes involved in tryptophan metabolism yet also illustrates the variety of genomic arrangements and regulatory differences coordinating their expression. The G-D-C gene order is common to gram-negative bacteria, but the amphibolic nature of trpG and the separateness of the trpGDC cluster are not. Transcription of the A. calcoaceticus trpGDC cluster is coordinated with the expression of trpE, but the trpFBA cluster is regulated differently (Cohn and Crawford 1976). In the absence of other results, we cannot identify potential
Table 2
Codon Usage in *Acinetobacter calcoaceticus* trpGDC and in the Equivalent Portion of the *Escherichia coli* trp Operon

<table>
<thead>
<tr>
<th>CODON</th>
<th>A. CALCOACETICUS TRP</th>
<th>E. COLI TRP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G</td>
<td>D</td>
</tr>
<tr>
<td>TTT (Phe)</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>TTC (Phe)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TTA (Leu)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>TTG (Leu)</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>CTT (Leu)</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>CTC (Leu)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CTA (Leu)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>CTG (Leu)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>ATT (Ile)</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>ATC (Ile)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>ATA (Ile)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ATG (Met)</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>GTT (Val)</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>GTC (Val)</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>GTA (Val)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>GTG (Val)</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>TCT (Ser)</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>TCC (Ser)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>TCA (Ser)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>TCG (Ser)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CCT (Pro)</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>CCC (Pro)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CCA (Pro)</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>CCG (Pro)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>ACT (Thr)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>ACC (Thr)</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>ACA (Thr)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>ACG (Thr)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>GCT (Ala)</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>GCC (Ala)</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>GCA (Ala)</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>GCG (Ala)</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>TAT (Tyr)</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>TAC (Tyr)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TAA (End)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TAG (End)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAT (His)</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>CAC (His)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CAA (Gln)</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>CAG (Gln)</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>AAT (Asn)</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>AAC (Asn)</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>AAA (Lys)</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>AAG (Lys)</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
transcriptional signals (either promoters or operators) by inspection of the 185-bp sequence flanking trpG. Similarly, we do not see evidence of secondary structures or leader peptide-coding regions that might reflect an attenuation mechanism of transcriptional regulation; however, the coupling of translation by the propinquity of termination and initiation signals of adjacent genes may rival gene fusion as a method of coordinating expression. We have pointed out above that the termination and initiation codons of trpG and trpD are adjacent to one another, and we have also noticed that the termination codon of trpD overlaps a nine-base sequence that closely resembles the Shine and Dalgarno (Escherichia coli) sequence (Shine and Dalgarno 1974). Although translational coupling has not been directly demonstrated in such arrangements, it is possible that termination of translation within or very near to an initiating environment influences the efficiency of the subsequent initiation event.

The evolution of an amphibolic trp/pab GAT subunit can be envisioned by several mechanisms, involving the duplication of one (CoI) or two (CoI + CoII) genes (with one of the duplicated CoII genes being deleted in the latter case). While the sequence data do not support one mechanism more than another, the similarity of the trp-linked amphibolic subunit to E. coli pabA rather than to E. coli trp(G) suggests that the pabA-like sequence may more closely resemble the ancestral GAT sequence. This is further supported by the fact that the trp-dedicated GAT sequences of lower eukaryotes also closely resemble the E. coli pabA sequence (Schectman and Yanofsky 1983; H. Zalkin, personal communication).

Gene fusions are commonly encountered in evolutionary comparisons of trp-gene arrangements. The fusion junctions that have thus far been characterized at
the nucleotide-sequence level by comparisons of fused and separate genes include enterobacterial \textit{trpG-trpD} (Miozzari and Yanofsky 1979; Nichols et al. 1980), \textit{Neurospora crassa} TRP1 (equivalent to \textit{trpg-trpC-trpF}) (Schechtman and Yanofsky 1983), and \textit{Sacharomyces cerevisiae} TRP5 (equivalent to \textit{trpA-trpB}) (Zalkin and Yanofsky 1981). In each case, the fusion regions are characterized by the presence of a "linking peptide" in the fused gene that did not previously exist in either of the separate genes (Yanofsky, personal communication). We report here the first complete nucleotide sequences of separate \textit{trpD} and \textit{trpC} genes. In conjunction with previous sequence data, we propose that an alternative fusion mechanism involving an in-phase deletion with end points within adjacent but separate genes may be responsible for joining \textit{trpC} and \textit{trpF}. The \textit{Bacillus subtilis} \textit{trp} cluster contains separate but adjacent \textit{trpC} and \textit{trpF} genes, and determination of the intercistronic region may provide further information regarding this hypothesis.

**Acknowledgments**

We thank the members of BioS 466 (Procaryotic Molecular Biology Laboratory)—G. Alberts, P. Balandyk, D. Baro, J.-D. Chen, S. Hahn, B. Howard, J.-W. Lee, P. Lysakowski, J. Malakooti, M. Sussman, H. Weidmann, and C. Wilke—for determining the sequence of the \textit{Acinetobacter calcoaceticus} DNA fragment contained in pMB466. W. K. Merkel for assisting with the complementation studies, I. Crawford for providing the \textit{A. calcoaceticus} strains and for helpful advice and discussions, and C. Yanofsky and H. Zalkin for communicating information prior to publication. This work was supported by U.S. Public Health Service grant AI 18639 and by the Department of Biological Sciences of the University of Illinois at Chicago.

**LITERATURE CITED**


NICHOLS, B. P., G. F. MIOZZARI, M. VAN CLEEMPUT, G. N. BENNETT, and C. YANOFSKY.


Walter M. Fitch, reviewing editor

Received January 16, 1984; revision received April 16, 1984.
One major parvalbumin each was isolated from the skeletal muscle of two reptiles, a boa snake, *Boa constrictor*, and a map turtle, *Graptemys geographica*, while two parvalbumins were isolated from an amphibian, the salamander *Amphiuma means*. The amino acid sequences of all four parvalbumins were determined from the sequences of their tryptic peptides, which were ordered partially by homology to other parvalbumins. Phylogenetic study of these and 16 other parvalbumin sequences revealed that the turtle parvalbumin belongs to \( \beta \) lineage, while the salamander sequences belong, one each, to the \( \alpha \) and \( \beta \) lineages defined by Goodman and Peche\'re (1977). Boa parvalbumin, however, while belonging to the \( \beta \) lineage, clusters within the fish in all reasonably parsimonious trees. The most parsimonious trees show many parallel or back mutations in the evolution of many parvalbumin residues, although the residues responsible for \( \text{Ca}^{2+} \) binding are very well conserved. These most parsimonious trees show an actinopterygian rather than a crossopterygian origin of the tetrapods in both the \( \alpha \) and \( \beta \) groups. One of two electric eel parvalbumins is evolving more than 10 times faster than its paralogous partner, suggesting it may be on its way to becoming a pseudogene. It is concluded that varying rates of amino acid replacement, much homoplasy, considerable gene duplication, plus complicated lineages make the set of parvalbumin sequences unsuitable for systematic study of the origin of the tetrapods and other higher-taxon divergence, although it may be suitable within a genus or family.

**Introduction**

Problems in evolution or taxonomy are classically studied by the analysis of fossil records and the morphology of current representatives. The fossils directly illuminating the origin of the tetrapods, particularly for the amphibians, are scarce, and the early evolution of the tetrapods has remained uncertain. Since Fitch and Margoliash (1967) demonstrated the use of orthologous amino acid sequences for the systematic study of evolutionary relationships, a variety of proteins and taxa has been examined with the method. It is important to determine the degree to
which the phylogenetic trees based on different proteins agree with each other as well as with the tree based on other biological data. For this, one must have several protein sets differing in function but covering the same organisms, but there has previously been insufficient effort to sequence proteins from species of taxonomic interest or which round out a set of proteins from a single taxon.

In an attempt to round out the amino acid sequences of orthologous proteins for the systematic study of the early evolution of tetrapods, we previously sequenced a myoglobin from the map turtle (Maeda and Fitch 1981a) and the lace monitor lizard (Maeda and Fitch 1981b) as well as a monomeric heme protein of bullfrog heart muscles (Maeda and Fitch 1982). We had difficulty, however, in isolating myoglobin from amphibian muscles, leaving the current myoglobin data inadequate for the resolution of tetrapod origins (Maeda and Fitch 1981b). Since it is important to compare several protein sets different in function but covering the same taxa as far as possible, we turned to parvalbumins.

Parvalbumins are small and acidic calcium-binding proteins especially abundant in the white muscle of fish and amphibians. The recent studies on the parvalbumins in muscles of rabbit (Lehky et al. 1974; Enfield et al. 1975), rat (Berchtold et al. 1982), and chicken (Heizmann and Strehler 1979), however, have revealed the presence of this protein in a wider range of vertebrates. Because of their relative ease of purification and small size, interest in this family of homologous proteins in systematic study has increased (Pechère et al. 1973; Goodman and Pechère 1977; Goodman et al. 1979).

A paper (Zhu et al. 1984) describes the isolation and amino acid sequence of β-type parvalbumins from muscle of the electric eel, Electrophorus electricus. In this paper we describe the isolation and sequence of two parvalbumins from a salamander, Amphiuma means, and one each from a boa snake, Boa constrictor, and a map turtle, Graptemys geographica, and discuss their evolutionary implications using a most-parsimonious-tree method (Fitch 1971; Fitch and Farris 1974).

Material and Methods

Animals

A salamander, Amphiuma means, and a snake, Boa constrictor, were purchased from Rand McNally and Company, Chicago, Illinois, and Marine Serpentarium Labs, Miami, Florida, respectively. Map turtles were a kind gift of Dr. J. Bull of the University of Wisconsin.

Protein Isolation and Amino Acid Sequence Determination

We shall use Apa, Bpa, and Tpa to designate Amphiuma, boa, and turtle parvalbumins, respectively. Isolation of proteins, disc gel electrophoresis, amino acid analysis, and fragmentation and separation of peptides are essentially the same as described previously (Maeda and Fitch 1981a; Zhu et al. 1984). As parvalbumins have a typical UV absorption spectrum, with maxima at 253, 259, 265, 269, and 276 nm (Pechère et al. 1973), elution and purification of protein were monitored by checking absorption spectra. Edman degradation of peptides (~0.1–0.2 μmol) was carried out manually, and phenylthiohydantoin derivatives were identified by thin-layer chromatography and by amino acid analysis after alkali or acid hydrolysis, as described previously (Maeda and Fitch 1981a).
Phylogenetic Reconstructions

Four parvalbumin sequences (this paper) and two of electric eels (Zhu et al. 1984) were compared with 14 other available sequences, and the most parsimonious tree was constructed by the method of Fitch (1971) and Fitch and Farris (1974).

Results

Isolation of *Amphiuma means* Parvalbumins, Apa-1 and Apa-2

From a homogenate of 300 g of *A. means* muscle, 300 mg of protein was recovered as a single fraction, Apa, by DE-52 ion exchange column chromatography, in 0.02 M Tris/HCl buffer, pH 8.4, containing 2 mM CaCl2, and NaCl increasing from 0 to 0.3 M. This fraction, though showing a single band of MW ~15,000 on SDS-disc gel electrophoresis, separated into three peaks on Sephadex G-50 gel filtration in 0.5 M acetic acid. The difference of amino acid compositions and of A260/A280 ratio of these peak components suggested that Apa contained at least two components. The relative amount of protein in the three Sephadex G-50 fractions differed depending on the Ca2+ ion concentration in the elution buffer. With 2 mM CaCl2 in the elution buffer, the third peak (Ve/V_T ~0.3) was the major peak, whereas without CaCl2, the second peak (Ve/V_T ~0.15) was the major peak, where Ve/V_T is the ratio of the elution volume of the peak to the total column volume. Electric eel parvalbumin-1 (Zhu et al. 1984) eluted at Ve/V_T ~0.3 under either condition.

Apa was separated into its components on a DE-52 column in 0.02 M sodium acetate buffer, pH 5.9, containing 1 mM CaCl2 with acetate buffer concentration increasing linearly to 0.2 M. Four fractions, Apa-1 through Apa-4, were obtained. Apa-1 gave a single band on gel electrophoresis at pH 9.5, whereas Apa-2, -3 and -4 each gave the same pattern of bands although with different intensities. Amino acid analysis of Apa-2, -3, and -4 showed no differences in their composition. They have two cysteine residues. Presumably, unstable polymerization occurred when Apa was desalted in 0.5 M acetic acid. Apa-1 and Apa-2 were used for amino acid sequence studies. The A260/A280 ratio of Apa-1 is 1.80, whereas it is 1.13 for Apa-2, reflecting the difference in tyrosine content (one and two residues per Apa-1 and Apa-2, respectively; table 1).

Isolation of Boa and Turtle Parvalbumin

From a homogenate of 350 g of boa muscle, 350 mg of major parvalbumin component (Bpa) was recovered from DE-52 column chromatography at pH 8.4. Rechromatography on DE-52 under the same condition or at pH 5.9, as well as on a Sephadex G-50 column, gave a single peak. Bpa was electrophoretically pure (SDS-disc gel electrophoresis and gel electrophoresis at pH 9.5). Two other fractions with UV absorption spectra typical of parvalbumins were recovered from muscle extract. One of them had a higher absorption at 180 nm than at 260 nm, suggesting the presence of a tryptophan. The amount was low (~2% of total parvalbumin fraction recovered), and we did not study it further.

Turtle muscle contains a relatively high amount of myoglobin (Maeda and Fitch 1981a). Parvalbumin co-eluted with myoglobin from a DE-52 column at pH 8.4. The parvalbumin (Tpa) was separated from myoglobin by DE-52 column chromatography in 0.02 M sodium acetate buffer, pH 5.9, containing 1 mM CaCl2, and rechromatographed under the same condition and gel filtered on Sephadex
G-50. Electrophoretically pure material (Tpa, 65 mg) was obtained from 200 g of muscle. As shown in table 1, Tpa composition shows somewhat less aspartic acid and somewhat more valine and lysine than determined from the sequence. This suggests that the preparation still contains a little myoglobin (Maeda and Fitch 1981a), but this did not cause any trouble in the sequence study.

Amino Acid Sequence Determination

Apa-1, which does not have any cysteine residues, was denatured at 90°C for 20 min in 0.05 M Tris/HCl buffer, pH 8.0 (11 mg/ml), prior to digestion with trypsin. Apa-2, Bpa, and Tpa were carboxymethylated with monoiodoacetic acid in 0.2 M Tris/HCl buffer, pH 8.5, containing 6 M guanidine-HCl, 2 mM EDTA, and 0.2 M dithiothreitol. Excess reagents were removed by gel filtration on a Sephadex G-75 column in the dark.

All four proteins have a blocked N-terminus. The N-terminal tryptic peptide of each was further digested with thermolysin and the N-terminal residue detected by thin-layer chromatography in n-butanol/acetic acid/H₂O (4:1:2, by volume) or in n-butanol/pyridine/H₂O (1:1:1, by volume) and by thin-layer electrophoresis at pH 4.8. For each of the four proteins, the mobility of this N-terminal residue is the same as a standard N-acetylalanine. Thus, the N-terminal amino group of each protein is presumably acetylated as in other parvalbumins (Pechère et al. 1973).

Table 1

<table>
<thead>
<tr>
<th>Amino Acid Composition of Parvalbumins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Asp</td>
</tr>
<tr>
<td>Thr</td>
</tr>
<tr>
<td>Ser</td>
</tr>
<tr>
<td>Glu</td>
</tr>
<tr>
<td>Pro</td>
</tr>
<tr>
<td>Gly</td>
</tr>
<tr>
<td>Ala</td>
</tr>
<tr>
<td>Cysb</td>
</tr>
<tr>
<td>Val</td>
</tr>
<tr>
<td>Met</td>
</tr>
<tr>
<td>Ilec</td>
</tr>
<tr>
<td>Leu</td>
</tr>
<tr>
<td>Tyr</td>
</tr>
<tr>
<td>Phe</td>
</tr>
<tr>
<td>His</td>
</tr>
<tr>
<td>Arg</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

A260/A280   1.80  1.13  1.81  1.60

* Relatively lower amount of aspartic acid and higher amounts of valine and lysine suggest that some myoglobin (Maeda and Fitch 1981a) remains in this preparation. The contamination caused no trouble in sequence analysis. The results are the average of three hydrolysates (24 h) and expressed in molar ratio with Gly + Ala set equal to 17, 19, 21, and 23 for Apa-1, Apa-2, Bpa, and Tpa, respectively. Values in parentheses give the numbers obtained from sequence analysis.

b As carboxymethylcysteine.

c Obtained from two 72-h hydrolysates.
Amino acid sequences were deduced mainly from those of tryptic peptides, as shown in figure 1. Some tryptic peptides were ordered by homology. Exceptions are as follows.

Cyanogen bromide cleavage of Apa-1 was carried out in 70% HCO₂H and in 75% trifluoroacetic acid (TFA). Neither condition produced a new terminal residue in good yield because the methionine is followed by threonine. Tpa has two methionines. One residue, followed by alanine, was cleaved by CNBr in 70% TFA, quantitatively producing peptide CNBrIV; the other, followed by threonine, was cleaved poorly.

Fourteen cycles of Edman degradation were carried out on CNBr-treated Apa-1, as shown in figure 1a.

Tryptic digestion of Apa-1 produced free arginine and lysine residues. The homology of the tryptic peptide sequences could not locate these residues. Clostridio peptidase B (Worthington Biochemical Co.) digestion of Apa-1 did not produce the digests specific to arginine residues in our experiment. The ε-amino groups of Apa-1 (3 mg) were blocked with 1,2,4-benzene-tricarboxylic anhydride (40 mg) as described previously (Maeda and Fitch 1981a). The tryptic digest was subjected to six cycles of manual Edman degradation without separating peptides. The mixture of PTII-amino acids produced at each step could be explained by the sequences of Th-3, Tk-d, and Td-2. This result located the arginine residue at position 38, prior to peptide Td-2, and the lysine residue, as a consequence, at position 28.

The ε-amino groups of Bpa were blocked in the same way, and the tryptic peptides were separated by SephadeX G-50 column chromatography. BTC-3 was isolated pure, which gave the alignment of TVS-5, TIIIf, TV-1, TVS-2, and TVS-3. BTC-2 was not free from BTC-1 contamination, but 13 cycles of Edman degradation could align TIIIc and TIIIf and explained the C-terminal part of the molecule.

Tryptic digestion of Tpa gave Tld and Tla in one experiment and peptides DE7180 and Tpa ppt in another experiment instead of the expected peptide for residues 55–75. This could have been caused by contaminating chymotryptic activities in the trypsin. It may also mean that the region is quite susceptible to enzymic digestion for some unknown reason.

Discussion

Figure 2 summarizes the 20 aligned parvalbumin sequences used to study the phylogeny. Gaps were introduced to improve the similarity. The rat sequence (Berchtold et al. 1982), which is very close to rabbit, was reported after this work was completed and so was not used. The most parsimonious tree found for the 20 sequences is shown in figure 3a. Parvalbumin sequences were divided into two major groups, α and β, by Pechère et al. (1973). Sequences 1–5 in figure 2 belong to the α group, and 6–20 belong to the β group. The ray is placed in this group because it is consistently associated with the β coelacanth when all 10,395 possible trees for eight taxa composed of sequences 6, 7, 8, 9, 10, and 11 and the ancestral sequences of the α group (1–5) and the fish β group (12–20) are examined.

All 20 sequences share 22 residues in common. They are Asp-12, Phe-26, Phe-31, Gly-36, Asp-53, Asp-55, Glu-61, Glu-64, Leu-65, Leu-69, Phe-72, Arg-77, Leu-79, Glu-83, Thr-84, Asp-92, Asp-94, Gly-95, Asp-96, Ile-99, Gly-100, and Glu-103. These residues locate and make two calcium-binding regions (Kretsinger and Nockolds 1973) except for Arg-77 and Glu-83, which are responsible for a salt
Fig. 1.—Sequence of parvalbumins of (a) Apa-1: *Amphiuma means; (b) Apa-2: *Amphiuma means; (c) Bpa: *Boa constrictor; and (d) Tpa: *Graptemys geographicus. —, residue determined by manual Edman degradation; X, residue determined by amino acid analysis of the remainder after several cycles of Edman degradation or of the fraction analyzed directly without hydrolysis. Sites of enzymic cleavage are indicated by 1, trypsin; 1, thermolysin; (T), Staphylococcus aureus V8; (l), nonspecific cleavage by trypsin.
FIG. 1 (Continued)
FIG. 2.—Comparison of amino acid sequences of parvalbumins. Sequences used for phylogenetic study were aligned with gaps (') to improve the similarity. Positions whose amino acid is identical in all 20 sequences are enclosed in boxes. The taxa are: pike = *Esox lucinus* (Frankenne et al. 1973; Gerday 1976); coelacanth = *Latimeria chalumnae* (Jauregui-Adell and Pechère 1978; Pechere et al. 1978); amphiuma = *Amphiuma means*; frog = *Rana esculenta* (Capony et al. 1975; Jauregui-Adell et al. 1982); rabbit = *Oryctolagus cuniculus* (Enfield et al. 1975); thornback ray = *Ruju clupeutu* (Thatcher and Pechère 1977); turtle = *Graptemys geographic*; boa = *Boa constrictor*; whiting = *Gudus merlingus* (Joassin and Gerday 1977); cod = *Gadus callarias* (Elsayed and Bennich 1975); hake = *Merluccius merluccius* (Capony et al. 1973); chub = *Leuciscus cephalus* (Gerday et al. 1978); electric eel = *Electrophorus electricus* (Zhu et al. 1984); and carp = *Cyprinus carpio* (Coffee and Bradshaw 1973; Coffee et al. 1974).
<table>
<thead>
<tr>
<th></th>
<th>Coelacanth</th>
<th>Pike 1</th>
<th>Pike 2</th>
<th>Amphluma 1</th>
<th>Amphluma 2</th>
<th>Frog</th>
<th>Amphluma 1</th>
<th>Frog</th>
<th>Thornback Ray</th>
<th>Thornback Ray</th>
<th>Coelacanth 2</th>
<th>Aoa</th>
<th>Whiting</th>
<th>Cod</th>
<th>Hake</th>
<th>Pike 2</th>
<th>Electric Eel 1</th>
<th>Electric Eel 2</th>
<th>Chubb</th>
<th>Carp 1</th>
<th>Carp 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Coelacanth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Electric Eel 1</td>
<td>Electric Eel 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Pike 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Amphluma 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Frog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Thornback Ray</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Coelacanth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Frog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Amphluma 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Coelacanth 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Aoa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Whiting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Cod</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Hake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Pike 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Carp 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Carp 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2 (Continued)
bridge, and Leu-79, which is responsible for hydrogen binding with Lys-66 (in carp). The function of Phe-72 and Thr-84 is not known. It is obvious that the residues in the C-terminal portion of the molecule are more conserved than those in the N-terminal part. Although there are two major groups, the only distinctive amino acid-sequence feature separating these two groups is position 34, where the α group sequences have hydrophobic residues and β group sequences have lysine residues in common.

Amphiuma Apa-1 belongs to the α group, whereas Apa-2 belongs to the β group, as does turtle Tpa. Boa Bpa, like other α sequences, has three residues following the nearly constant leu-val in positions 107–108. This differs from the β group, which has only two additional residues (except for whiting). Nonetheless, Bpa belongs to the β group.

Bpa, which, like Tpa and others, has cysteine at position 20, has a unique Cys at position 28. Cysteine-35 of Apa-2 is also present in frog α and whiting, while cysteine-73 of Apa-2 is a unique replacement. Other unique residues found in Apa-1 sequences are Ala-25 (Ser in all others), Phe-28 (His or Tyr is common), Arg-40 and Arg-56 (both Lys are common), and Asp-86 (the only other negative

---

**FIG. 3.**—The most parsimonious tree observed (a) and the biological tree (b) for 20 parvalbumins. The biological tree follows the views of Romer (1966) and Greenwood et al. (1966). Numbers on legs are numbers of nucleotide substitutions required to account for the descent from the ancestor to its immediate descendant. The total number of substitutions required is 579 and 603 for the most parsimonious tree and for the biological tree, respectively. Nodes with open diamonds denote gene duplications proven by the presence of the same species on each of the two branches descending from these nodes. The insets indicate how many codons had how many substitutions in the tree they are shown with (e.g., 1a shows that 27 codons were unsubstituted in the most parsimonious tree, while two codons were substituted 17 times).
charged side chain is Glu in Bpa). Unique residues found in Bpa are Gly-16 (Ala in all others), Asp-73, and Lys-75 (no other positive charge at this position), in addition to the Cys-28 and Glu-86 mentioned above. These positions are rather on the surface of the molecule and probably do not affect the function.

The biological tree (fig. 3b) is, among the fish (see table 2), generally consistent with the views of Romer (1966) and of Greenwood et al. (1966, fig. 1) in which the teleosts divide into three major groups represented here by the orders Salmoniformes (pike), Cypriniformes (carp, chub, and eel), and Gadiformes (cod, whiting, and hake). The order of their divergence is not apparent in these authors’ works, and the choice of making the Gadiformes come off first is based on the slender reed, in this case, of parsimony. The most parsimonious tree does violence to the cohesion of these three groups in that it takes the chub from the Cypriniformes and the whiting from the Salmoniformes and makes the two of them a sister group that together constitute the sister group of the cyprinids. Forcing the two carp sequences to be recently paralogous (i.e., the result of a gene duplication), we examined all 10,395 trees for these fish sequences (fig. 4); and the most parsimonious tree for that set is identical with their section of the most parsimonious tree. It would have cost four extra substitutions to make the chub the sister group of the carp and electric eel, leaving the whiting where it is.

That there are twice as many minimum base differences between whiting and cod as between whiting and carp (~60:30) suggests, since cod and whiting are classified in the same genus, that the cod and whiting sequences are paralogous.
Table 2  
Fish Relationships

<table>
<thead>
<tr>
<th>Chondrichthyes</th>
<th>Elasmobranchii</th>
<th>Batoidea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osteichthyes</td>
<td>Actinopterygii</td>
<td>Teleostei</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cypriniformes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Salmoniformes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gadiformes</td>
</tr>
<tr>
<td>Sarcopterygii</td>
<td>Crossopterygii</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** The levels shown are class, subclass, infraclass, and order as given by Romer (1966) and are consistent with those of Greenwood et al. (1966). The fish from this study are: batoid-ray; cyprinid-carp, chub, eel; salmonid-pike; gadid-cod, whiting, hake. See fig. 2 for genus and species names.

This is not unreasonable since there is ample evidence for gene duplication in the parvalbumins, the two pike, carp, and electric eel sequences clearly being the result of three separate gene-duplication events.

The other Osteichthys, the coelacanth, belongs to a separate subclass (Sarcopterygii) from the others (Actinopterygii) and is shown in the biological tree diverging

**Fig. 4.**—Distribution of the 10,395 possible unrooted trees for parvalbumins. The sequence data employed are ray, frog-$\beta$, amphiuma 2, turtle, boa, coelacanth, the ancestral sequence for five $\alpha$ sequences, and the ancestral sequence for nine fish $\beta$ sequences. The ancestral sequences were computed using the biological tree. Nucleotide substitutions ranged from 591 to 634. The topology of the two trees requiring only 591 substitutions is shown in the inset.
before the others but after the divergence of the ray, the only member of the Chondrichthyes. This set of higher-order relationships is preserved in the most parsimonious tree as well.

The biological tree shows the tetrapods as diverging from the crossopterygian lineage, but the sequence data demand that the frog, salamander, turtle, and rabbit sequences instead diverge early from the actinopterygian lineage. This is true in both the α and β sequence groups and is found by Baba et al. (1984), who examined some of these parvalbumins in the context of other calcium-binding proteins. To bring the tetrapods off the crossopterygian line would cost 17 extra nucleotide substitutions, 11 in the alphas and six in the betas. This casts doubt on the crossopterygian origin of the tetrapods.

The boa does not fall close to the other reptile, the turtle, in these data. Indeed, the boa sequence has 25% more unavoidable discordancies when compared with the turtle than it has when compared to the whiting (45 vs. 36). (Unavoidable discordancies are differences that must lead to extra homoplasious [= parallel + reverse] substitutions irrespective of tree structure; see Fitch (1977)). These unavoidable discordancies ultimately demand that the boa sequence be the sister sequence of the whiting; we have examined another set of 10,395 trees, trying without success to get the boa sequence to go elsewhere. This result could be explained by asserting that the two reptile sequences are paralogous (after all, there are only 25 unavoidable discordancies between the turtle and the salamander), but there is no other evidence to support paralogy. We have already suggested such an explanation involving the whiting and the chub. It should be noted that these two suggestions of paralogy are logically separated; one such gene duplication cannot be used to explain both observations.

Another conflict occurs within the tetrapods in that the divergence of the amniotes comes off different amphibian lineages in the two major parvalbumin groups, with the rabbit α being the sister sequence of the frog α and the turtle β being the sister sequence of the salamander sequence. The situation in the α group is consistent with the ideas of Jarvic (as discussed in Parsons and Williams (1963)). In neither group of sequences are the anurans and urodeles the sister sequences of each other. To make them such would cost five extra substitutions in the alphas plus three extra in the betas. Thus, for neither group are the amphibians monophyletic.

When the analysis of sequence data conflicts with established opinion, the reasons may be various, but the major dichotomy is between whether the sequence data or the established opinion is the problem. In the present case, the only instance where we argue that the sequence data give strong indications that the established phylogeny may be wrong is in their indication of the tetrapods arising from the actinopterygians. This is because both the α and β parvalbumins show the same result, and it is particularly strong in the α group.

The problems that see the whiting sequence associated with the chub and the boa with the whiting are just too radical to permit a suggestion that the phylogeny is wrong. That the two associations evidence two additional instances of paralogy is reasonable but should have additional support to rule out the possibility that grossly unequal rates of evolutionary change in different lineages are not the cause of the observations.

The problems within the tetrapods are best explained as the result of simple stochastic variability.
The relative rates of change in both the biological and most parsimonious trees show large differences. In the latter's case, the minimal number of substitutions in the two lines of descent since their common ancestor varies more than twofold for each of the pairs whiting:chub, cod:hake, and rabbit:frog. Moreover, it is more than tenfold for eel-1:eel-2, suggesting that eel-2 may no longer be under much selective constraint. This would be rather like a pseudogene if it were not for the fact that this gene is expressed. The degree of variability between these different lineages seems rather larger than that seen in other proteins and makes any phylogenetic conclusion drawn from the parvalbumins more hazardous than usual.

The number of codons showing substitutions is 89. The average number of nucleotide substitutions per codon is 6.5 (range 1–17; see insets, fig. 3a, b). This is rather large for only 20 taxa and indicates a potential source of problems in phylogeny reconstruction. Felsenstein (1979) has shown that high and unequal rates of character-state change in different lines of descent are precisely the conditions most apt to cause a parsimony procedure to give the incorrect historical relationships. Those conditions appear to be present in the parvalbumins and lead to the conclusion that parvalbumins are not currently suitable for the study of higher-level (families and above) vertebrate taxonomy. Indeed, they are unlikely ever to be so in the absence of work at the genomic level that sorts out all the paralogies and, possibly, increases the number of taxa as well, especially those in unrepresented orders.

Acknowledgments

We are grateful to Dr. R. L. Niece and D. Smith for technical help and discussion. This work was supported by the National Science Foundation under grants DEB 78-04191 and DEB 78-14137.

LITERATURE CITED


MASATOSHI NEI, reviewing editor

Received April 26, 1984; revision received May 17, 1984.
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 Cα1 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
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_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-1^2^ 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_k within *R. norvegicus* but must be understood in a context at least as broad as that consisting of closely related C_k genes within the genus *Rattus*. Recent studies on allelic forms of C_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_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, LOUVAINE (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_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_k probe (Sheppard and Gutman 1981b), washed, and autoradiographed to obtain a restriction-enzyme map of the C_k gene in its unrearranged genomic context and to determine the size of the C_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. 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-acrylylcystamine-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 M13 mp8 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-1a 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**

FIG. 1.—Restriction-enzyme maps of Cκ-containing genomic and cloned EcoRI fragments. The top two and bottom two maps represent different unrearranged Cκ 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-sequence 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 Cκ genes in these Australian rats must still be at least formally considered.)

**Cκ Genomic Clones**

Six independent, recombinant, phage-containing Cκ-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* Cκ 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_K 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_K Nucleotide Sequences**

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

![Nucleotide sequences of C_K 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_K-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*.](image)
of the 3' end of the J&x–C&x intervening region have been sequenced (the number of nucleotides varies for each species in this region, as discussed below). The C&x-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. cooktownensis 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&x 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&x.)

Maintenance of Repetitive Sequence

A number of short repetitive sequences have been identified in the Australian rat C&x sequences. These include the tetranucleotide GTET, 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&x 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&x 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&x-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&x 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&x genes reflects that predicted by serological data (Gutman and Baverstock 1980). The R. l. cooktownensis kappa chain cross-reacts quantitatively with Igk-la (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 C\(_k\)-Coding and Noncoding Regions

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Coding</th>
<th>Noncoding*</th>
<th>(P^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Rattus leucopus</em> cooktownensis vs. <em>Rattus</em> colletti</td>
<td>3.7</td>
<td>1.5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><em>R. l. cooktownensis</em> vs. LOU</td>
<td>2.8</td>
<td>2.0</td>
<td>NS</td>
</tr>
<tr>
<td><em>R. l. cooktownensis</em> vs. DA</td>
<td>5.0</td>
<td>1.9</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td><em>R. colletti</em> vs. LOU</td>
<td>5.6</td>
<td>1.6</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><em>R. colletti</em> vs. DA</td>
<td>6.2</td>
<td>1.9</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>LOU vs. DA</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.

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

*b* From \(x^2\) test for independence of differences between the coding and total noncoding regions. \(P > 0.05\) was considered to be not statistically significant.

* Sheppard and Gutman (1981b).

significant difference between coding and noncoding sequences). The *R. colletti* kappa chain cross-reacts not at all with Igk-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*</td>
<td>1.6</td>
<td>4.6</td>
<td>0.35</td>
</tr>
<tr>
<td><em>Rattus leucopus</em> cooktownensis vs. <em>R. colletti</em></td>
<td>3.9</td>
<td>5.9</td>
<td>0.7</td>
</tr>
<tr>
<td><em>R. l. cooktownensis</em> vs. LOU*</td>
<td>4.8</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td><em>R. l. cooktownensis</em> vs. DA*</td>
<td>3.2</td>
<td>5.9</td>
<td>0.5</td>
</tr>
<tr>
<td><em>R. l. cooktownensis</em> vs. Mouse*</td>
<td>19.5</td>
<td>10.6</td>
<td>1.8</td>
</tr>
<tr>
<td><em>R. l. cooktownensis</em> vs. Human*</td>
<td>76.0</td>
<td>29.9</td>
<td>2.5</td>
</tr>
<tr>
<td><em>R. colletti</em> vs. LOU*</td>
<td>8.4</td>
<td>6.7</td>
<td>1.3</td>
</tr>
<tr>
<td><em>R. colletti</em> vs. DA*</td>
<td>7.8</td>
<td>7.6</td>
<td>1.0</td>
</tr>
<tr>
<td><em>R. colletti</em> vs. Mouse*</td>
<td>26.5</td>
<td>8.7</td>
<td>3.1</td>
</tr>
<tr>
<td><em>R. colletti</em> vs. Human*</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 (\alpha)-globin*</td>
<td>83.0</td>
<td>8.4</td>
<td>9.9</td>
</tr>
<tr>
<td>Rat vs. Human insulin*</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.

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

*b* Sheppard and Gutman 1981b.

*c* Max et al. 1981.

*d* Heiter et al. 1980.

*e* 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 C_K-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 C_K genes whose proteins are known to be serologically closely related (see Gutman and Baverstock 1980 and discussion above). Thus, not only are C_K-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 C_K 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 C_K genes. This may indicate that, although they clearly exist (as discussed below), the selective forces acting to conserve C_K 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 C_K domain between the C_K 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 C_K 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-C_K model (kindly provided by P. Colman). The majority of C_K 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 C_K allelic differences in R. norvegicus. This pattern points to an appreciable degree of selective constraint on the C_K 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 C_K 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 C_K genes of at least three species within the genus Rattus. Coding regions
Table 3
Spatial Distribution of Amino Acid Replacements in Cκ Domains

<table>
<thead>
<tr>
<th>COMPARISON</th>
<th>C-terminal Bends</th>
<th>Remainder of Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rattus leucopus cooktownensis vs. Rattus colletti</td>
<td>6 (22)</td>
<td>5 (6)*</td>
</tr>
<tr>
<td>R. l. cooktownensis vs. LOU</td>
<td>3 (11)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>R. l. cooktownensis 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>

Note.—Amino acid positions were determined from a three-dimensional model of the Cκ 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.

* P < 0.05.
† P < 0.01.

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 R. colletti, respectively, but only 1:2 in that leading to R. l. cooktownensis; however, the total number of coding changes in the R. l. cooktownensis 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 Cκ genes in the genus 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 Cκ genes. Although these findings may be explained by hypothesizing as-yet-unrecognized constraints on noncoding DNA near Cκ genes from these closely related species (such as a required higher-order structure of either DNA or RNA [Sheppard and Gutman 1981b]), the fact that these unusual features do not hold for comparisons between more distantly related Cκ 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 $C_K$ 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 $C_K$ domains? Although $C_K$ domains bind covalently and noncovalently to the $C_H$ domain of heavy chains, there are no other specific functions assignable to the $C_K$ domain. As far as is known, $C_K$ 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 Nisino et al. 1975). Invoking the interaction of $C_K$ and $C_H$ domains as a primary focus of this diversifying selection is weakened by our finding that the differences between various Rattus species' $C_K$ 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 $C_K$ 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 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.
MOLECULAR BIOLOGY AND EVOLUTION extends its thanks to the following reviewers whose appreciable and appreciated efforts were so necessary to its success.

AIR, G. M.
ALLENDORF, F.
AQUADRO, C. F.
ARNHEIM, N.
AVISE, J. C.
BALL, L. A.
BARNES, W.
BELL, G. I.
BLAIR, W. F.
BRITTEN, R. J.
BROWN, A. II. D.
CAMPBELL, A. M.
CHAKRAVARTI, A.
CHAMBERS, G.
CHAO, L.
CRAWFORD, I. P.
CROW, J. F.
DIXON, G.
DOUGLAS, T. C.
EANES, W. F.
ELDREDGE, R. N.
FELSENSTEIN, J.
FITCH, W. M.
FOX, G.
GOJOBORI, T.
GOODMAN, M.
GRANTHAM, R.
GRAUR, D.
GRUNSTEIN, M.
HALL, B. G.
HARDIES, S. C.

HOLMQUIST, R.
HUANG, H.
HUDSON, R. R.
INGER, R. F.
IKEMURA, T.
JAMESON, D.
JOHNSON, M. J.
KANE, J.
KAPLAN, N.
KEDES, L.
KLEE, C. B.
KOEHN, R.
KOLODNER, R.
KREITMAN, M.
LANGLEY, C.
LENGYEL, P.
LEVIN, B.
LI, W.-H.
LIN, E. C. C.
LIPMAN, D.
MACINTOSH, L.
MCLACHLAN, A. D.
MAXSON, L.
MILKMAN, R.
MILLER, D. A.
MIYATA, T.
NEI, M.
NUTE, P. E.
OHTA, T.
PALESE, P.
PALMER, J. D.

PATIENT, R. K.
POWERS, D. A.
REZNIKOFF, W. S.
RILEY, M.
RUDIKOFF, S.
SCHOPF, T.
SCOTT, A. F.
SELANDER, R. K.
SMITH, G.
SMITH, T. F.
SOGIN, M.
SOMFRO, G.
STONEKING, M. A.
SUGIURA, M.
SYVANEN, M.
TAJIMA, F.
TAKAHATA, N.
TEMIN, H.
VARVIO-AHO, S.
WALSH, J. B.
WALSH, K. A.
WATERMAN, M.
WEBER, R.
WEISSMAN, S.
WEYSMAN, E.
WHTT, G. S.
WHITTAM, T. S.
WILBUR, W. J.
WOESE, C. R.
YOKOYAMA, S.

ZOUROS, E.
ZUCKOR, M.
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acinetobacter calcoaceticus</td>
<td>456</td>
</tr>
<tr>
<td>Adaptation 109; in species 1</td>
<td></td>
</tr>
<tr>
<td>Alleles: null 238; variant 84</td>
<td></td>
</tr>
<tr>
<td>Allendorf, Fred W., 183, 238, 438 (B)</td>
<td></td>
</tr>
<tr>
<td>Allozymes 67, 162</td>
<td></td>
</tr>
<tr>
<td>Amino acid replacement 473</td>
<td></td>
</tr>
<tr>
<td>Analysis, computer, 357</td>
<td></td>
</tr>
<tr>
<td>Antibodies 317; polyclonal 317</td>
<td></td>
</tr>
<tr>
<td>Antisera 317</td>
<td></td>
</tr>
<tr>
<td>Aquadro, Charles F., 38, 423</td>
<td></td>
</tr>
<tr>
<td>Arneheim, Norman, 29</td>
<td></td>
</tr>
<tr>
<td>Avise, John C., 38</td>
<td></td>
</tr>
<tr>
<td>Ayala, Francisco J., 287 (B)</td>
<td></td>
</tr>
<tr>
<td>Baba, Marietta L., 442</td>
<td></td>
</tr>
<tr>
<td>Baeverstock, P. R., 489</td>
<td></td>
</tr>
<tr>
<td>Berg, Claire M., 411 (R)</td>
<td></td>
</tr>
<tr>
<td>Berg, Douglas E., 411 (R)</td>
<td></td>
</tr>
<tr>
<td>Berger-Cohn, Janet, 442</td>
<td></td>
</tr>
<tr>
<td>Besta, R., 489</td>
<td></td>
</tr>
<tr>
<td>Biogeography 345</td>
<td></td>
</tr>
<tr>
<td>Boa constrictor 473</td>
<td></td>
</tr>
<tr>
<td>Boyer, Samuel H., 371</td>
<td></td>
</tr>
<tr>
<td>Boynton, John E., 317</td>
<td></td>
</tr>
<tr>
<td>Brownell, Elise, 29</td>
<td></td>
</tr>
<tr>
<td>Calhoun, David H., 109</td>
<td></td>
</tr>
<tr>
<td>Cellobiose 171</td>
<td></td>
</tr>
<tr>
<td>Chakravarti, Aravinda, 286 (B)</td>
<td></td>
</tr>
<tr>
<td>Chang, L.-Y. Edward, 371</td>
<td></td>
</tr>
<tr>
<td>Chlamydomonas 317</td>
<td></td>
</tr>
<tr>
<td>Chloroplast 291</td>
<td></td>
</tr>
<tr>
<td>Clegg, Michael T., 291 (R)</td>
<td></td>
</tr>
<tr>
<td>Clock, molecular, 249</td>
<td></td>
</tr>
<tr>
<td>Coding, overlapping frames, 260</td>
<td></td>
</tr>
<tr>
<td>Codon boundaries, 260</td>
<td></td>
</tr>
<tr>
<td>Conservation, sequence, 335</td>
<td></td>
</tr>
<tr>
<td>Conversion, DNA sequence, 371</td>
<td></td>
</tr>
<tr>
<td>Coss, Richard G., 249</td>
<td></td>
</tr>
<tr>
<td>Crossing-over event, 473</td>
<td></td>
</tr>
<tr>
<td>Crossingover, unequal, 195</td>
<td></td>
</tr>
<tr>
<td>Curtis, Stephanie E., 291 (R)</td>
<td></td>
</tr>
<tr>
<td>Czelusniak, John, 371</td>
<td></td>
</tr>
<tr>
<td>Daniel, Susan W., 38</td>
<td></td>
</tr>
<tr>
<td>De Framond, Jean, 162</td>
<td></td>
</tr>
<tr>
<td>Dehydrogenase: glucose-6-phosphate 162; lactate 238</td>
<td></td>
</tr>
<tr>
<td>Deletion(s) 269; base 442</td>
<td></td>
</tr>
<tr>
<td>Demaillé, Jacques G., 442</td>
<td></td>
</tr>
<tr>
<td>Development 183</td>
<td></td>
</tr>
<tr>
<td>Differentiation, mitochondrial DNA, 38</td>
<td></td>
</tr>
<tr>
<td>Distance: evolutionary 269; genetic 125</td>
<td></td>
</tr>
<tr>
<td>Distribution: negative binomial 125; gamma 125</td>
<td></td>
</tr>
<tr>
<td>Divergence 390; coding versus noncoding sequence 489; evolutionary 195; sequence 125</td>
<td></td>
</tr>
<tr>
<td>Diversity, antibody, 195</td>
<td></td>
</tr>
<tr>
<td>DNA: methylation 221; mitochondrial 38, 423; ribosomal 29</td>
<td></td>
</tr>
<tr>
<td>Drift, random genetic, 67</td>
<td></td>
</tr>
<tr>
<td>Duplications, internal, 335</td>
<td></td>
</tr>
<tr>
<td>Dykhuizen, Daniel E., 162</td>
<td></td>
</tr>
<tr>
<td>Electrophoresis 249</td>
<td></td>
</tr>
<tr>
<td>Enzymes, structural loci encoding, 183</td>
<td></td>
</tr>
<tr>
<td>Escherichia coli 171, 317</td>
<td></td>
</tr>
<tr>
<td>Evolution: adaptive 442; calmodulin 442; concerted 195; convergent 143; divergent 143; microbial 213; molecular 38, 291; primate 221; rates of molecular 442</td>
<td></td>
</tr>
<tr>
<td>Families, multigene, 302</td>
<td></td>
</tr>
<tr>
<td>Fitch, Walter M., 57, 473</td>
<td></td>
</tr>
<tr>
<td>Foulser, David E., 357</td>
<td></td>
</tr>
<tr>
<td>Frank, M. B., 489</td>
<td></td>
</tr>
<tr>
<td>Frazier, Marsha L., 335</td>
<td></td>
</tr>
<tr>
<td>Galas, David J., 260</td>
<td></td>
</tr>
<tr>
<td>Gene(s): aminocyclitol-3'-phosphotransferase, sequence, 57; ancestral, families, 390; β-like globin 390; cluster 390; conversion 195, 302, 371; cryptic 109, 213; duplicate 94, 238; duplication(s) 238, 335; evolution 371; expression 411; fused 143; fusions 456; globin 94, 269; glucagon 335; immunoglobulin C-kappa 489; insulin 269; interspecific, exchange, 411; intrachromosomal, conversion, 302; multiple, conversions, 371; nonfused 143; rearrangements 456; regulatory 183; ribosomal RNA 221; structural, rearrangements, 143; transfer 57; trpGDC, cluster, 456; variable region 195</td>
<td></td>
</tr>
<tr>
<td>Genetics, population, 84, 125</td>
<td></td>
</tr>
<tr>
<td>Genome rearrangements 411</td>
<td></td>
</tr>
</tbody>
</table>

505
Genus *Rattus* 489
Ghandour, Ghassan, 357
Gillham, Nicholas W., 317
Globin: $\beta$- 302; $\delta$- 302
Glucosides, $\beta$-, 171
Goujobori, Takashi, 94, 195
Golding, G. B., 125, 439 (L)
Goncharoff, Paul, 456
Goodman, Morris, 371, 442
Gorilla 371
Gray, Gary S., 57
Gutman, G. A., 489
Hall, Barry G., 109, 171
Hardison, Ross C., 302, 390, 435 (B)
Hartl, Daniel L., 143
Heath, Peter, 371
Homoplasy 473
Hudson, R. R., 439 (L)
Human(s) 371, 390
Identity block 357
Immunoglobulin 195
Immunology, albumin (microcomplement fixation), 345
Insertion 269
Kaplan, Jeffrey B., 456
Kaplan, Norman, 423
Karlin, Samuel, 357
Kimura, Motoo, 84
Knoller, Mechthilde, 221
Knudsen, Kathy L., 183
Korn, Laurence J., 357
Kricker, Maja, 171
Krystal, Mark, 29
Lansman, Robert A., 38
Leary, Robb F., 183
Li, Wen-Hsiung, 94, 213, 335
Linkage disequilibrium 67
Lopez, Linda C., 335
Luo, Chi-Cheng, 335
Macromolecules 442
Maeda, Nobuyo, 473
Maps, alignment, 357
Margot, Jean B., 302
Matsuda, Genji, 442
Maxson, Linda R., 288 (B), 345
Metabolism, tryptophan, 143
Method, maximum parsimony, 442
Model, mathematical, 423
Mutation(s): neutral 84; nonrandom 125; spontaneous 411; transition 94; transversion 94
Myers, Alan M., 317
Nei, Masatoshi, 195, 269
Neutrality 162; selective 162
Nichols, Brian P., 456
Nucleolus organizer regions 29
Ochman, Howard, 67
Operons, tryptophan, 143, 456
Orthology 390
Papillomavirus 357
Paralogy 473
Parsimony 473
Parvalbumin 473
Peptides, tryptic, 473
*Peromyscus* 38
Perutz, M. F., 1
Phosphoglucomutase 183
Phylogeny, *bufo*, 345
Plasmid insert 456
Polymorphism 238; protein 84
Population dynamics 213
Prass, William, 371
Proteins, chloroplast ribosomal, 317
Pseudogenes 29, 195, 302, 390
Rabbits 390
Rate(s): calibration 249; divergence 390; unequal substitution 269
Recognition, DNA sequence, 411
Recombination: homologous 411; nonreciprocal 390; substitutive 67
Regions, tryptophan regulatory, 143
Resistance: antibiotic 57, 411; Kanamycin 57
Restriction: digests 38; endonuclease 38
Risko, Kenneth J., 423
Ryman, Nils, 238
Salamander 473
Salmonids 238
Sasakawa, C., 411 (R)
Saunders, Grady F., 335
Schmickel, Roy D., 221
Schmidt, Robert J., 317
Scott, Alan F., 371
Seibold, Anita M., 456
Selander, Robert K., 67
Selection: Darwinian natural 442; periodic 67
Sequence(s): amino acid 473; evolution 423; homologous 143; insertion 260, 411; intervening 302; nucleotide 456

SHAPIRA, JOHN F., 38
Sites, coadapted functional, 442
SLIGHTOM, JERRY L., 371
SMITH, DAVID GLENN, 249
SMITH, TEMPLE F., 260
Speciation, mammalian, 38
Spermophiles 249

STAHN, GUNNAR, 238

Structure: hemoglobin 1; multilocus genetic 67

Substitution(s): base 442; nonsynonymous 94, 335, 371; nucleotide 84, 269, 291; rate of 94; silent nucleotide 489; synonymous 335

SZURA, L. LYNNE, 221

TADIMA, FUMIO, 269

Tetrapod origins 473

Theory, neutral, 84
Transcription, initiation site, 221
Transfection 57
Transitions 423
Transposition, mechanism of, 411
Transposon 57; Tn5 bacterial 411
Transversion 423
Trot, rainbow, 183
TRUSKO, STEPHEN, 371

Turtle, map, 473

Variability, intergenic genetic, 195
Variance, binomial, 125
Variation, geographic, 67

WHITTAM, THOMAS S., 67
WILSON, GOLDER N., 221

YANOFSKY, CHARLES, 143
YOKOYAMA, SHOZO, 109

ZHU, DEXU, 473
Information for Contributors

*Molecular Biology and Evolution* is a bimonthly journal devoted to the interdisciplinary science between molecular biology, and evolutionary biology. The journal emphasizes experimental papers, but theoretical papers are also published if they have a solid biological basis. Although this journal is primarily for original papers, review articles and book reviews normally written by solicited authors are also published. Brief discussion and comment on material published in this journal or on issues particularly relevant to readers of this journal will be published as "Letters to the Editor." Letters that refer to a paper handled by an Associate Editor should be sent to that editor or to the Editor in Chief.

To minimize publication delays, authors should follow the instructions given here and should also provide their telephone numbers.

**Submission of Manuscripts**

Send manuscripts (one original and two high-quality copies) to the Editor in Chief, Managing Editor, or any Associate Editor (addresses below). Any manuscript or any part of a manuscript which has been published or submitted for publication elsewhere cannot be accepted for publication. Anyone who wishes to write a review article should contact the Managing Editor. Correspondence about book reviews should also be addressed to the Managing Editor. Decision on acceptance of papers will be made as rapidly as possible. Papers that are not suitable to the journal will be returned immediately to authors without detailed review.

After a manuscript is accepted, its author will be requested to sign an agreement transferring copyright to the publisher. No published material may be reproduced or published elsewhere without the written permission of the copyright owner. The journal will not be responsible for the loss of manuscripts at any time.

Publication is taken to imply that the authors are prepared to make available to the public any unpublished sequences on which the paper is based and any clone of cells, DNA, or antibodies used in the experiments reported. This principle also applies to computer programs.

**Preparation of Manuscripts**

Papers must be written in English and organized in the sequence described below. Each section must be typed double-spaced on heavyweight, nonerasable bond; the page margins should be 1 1/2 inches wide to allow for corrections and manuscript editor's notes and queries. Special typefaces (e.g., italic or sans serif) should not be used, and right-hand margins should not be justified. Word-processing output on dot matrix printers is acceptable only if it is the quality of the standard typewriter. Handwritten items (e.g., Greek letters) must be identified in the margin. Non-English words must have correct diacritics. Although each major part of the paper (e.g., Literature Cited) must begin on a new page, the pages should be numbered consecutively throughout, beginning with the title page and continuing through the abstract, text, Appendix, Literature Cited, footnotes, tables, and ending with figure legends.

*Title page.*—This page should contain the paper's title, the names of all authors, the institution(s) at which the research was done, the current affiliations of all authors, the name and address for correspondence, and a footnote on nonstandard abbreviations used, if any (see below). Finally, the title page should also provide a running head (maximum of 50 characters and spaces).

*Abstract.*—The abstract should be a one-manuscript-page factual condensation of the entire paper, including a statement of purpose, a clear description of observations and findings, and a concise presentation of conclusions. It should not assert that the findings are discussed.

*Key words.*—A list of three to six words or phrases should be provided that will accurately index the subject matter of the article.

*Text.*—The text should comprise the following sections: (1) Introduction, (2) Material and Methods, (3) Results, (4) Discussion, and (5) Acknowledgments (if any). Papers should be concise but will not be restricted in length.
Information for Contributors

All organisms mentioned must be identified by their scientific binomens and underlined. Symbols for genetic loci must also be underlined and should follow the established rules of genetic nomenclature for the various organisms (consult M. Demerec et al., Genetics 54 [1966]:61-76). Include the formal IUB name and number of all enzymes mentioned.

Do not use abbreviations for words or phrases used less than five times. Abbreviations used by the Journal of Biological Chemistry will be regarded as standard; nonstandard abbreviations should be defined collectively in a footnote.

Mathematical equations must be carefully typewritten: spacing between characters should be correct as typed. It will be assumed that all characters in equations and their counterparts in the text will be set in italics unless the author specifies otherwise the first time a character appears. Equations should be numbered sequentially, in arabic numerals in parentheses, on the right-hand side of the page.

These and other guidelines can be found in the Council of Biology Editors Style Manual (4th ed., 1978). In general, all material should conform to the CBE format. See also recent issues of this journal.

Terminology.—A satisfactory interdisciplinary communication requires using words with careful attention to their precise meaning in both disciplines. Authors may use any word they choose provided only that its meaning is clear, consistent, and serves to increase the paper's comprehensibility. The following preferred usages are not prescriptive but will be assumed unless authors define them otherwise.

Where the alignments disagree, they are differences rather than changes since there may have been multiple changes to create a single difference. Differences or changes are replacements if the sequences are amino acids, substitutions if they are nucleotides. Mutations should be restricted to changes before selection has operated. Homology must be defined since it has two common meanings: (1) observed similarity and (2) inferred common ancestry. The term similarity is preferred for meaning 1 because sequences may have similarity acquired by convergence (analogy) rather than retained after divergence (homology). When homology arises via a gene duplication (all or part), it is properly called paralogy; when it arises via speciation, it is properly called orthology. Gaps are introduced into sequences to increase their similarity rather than to optimize similarity (homology), unless an algorithm is employed that guarantees an optimized result according to the way similarity (homology) is defined (e.g., as maximum matches—a third meaning of homology). Similarity should not be asserted to be significant unless patently obvious or accompanied by a probability statement and its method of determination ($\chi^2$, standard measure, binomial, etc.).

As recommended to the IUB, the preferred single letter code for nucleotide bases including ambiguity is: A = adenine, C = cytosine, G = guanine, T = thymine, U = uracil, R = A/G (purine), Y = C/T (pyrimidine), M = A/C, W = A/T, S = C/G, K = G/T, B = C/G/T (not A), D = A/G/T (not C), H = A/C/T (not G), V = A/C/G (not T), N = X = A/C/G/T (any or unknown). For ambiguous nucleotides, T and U are equivalent.

Literature Cited.—Literature in the text should be cited by author and year and, where citation is to a book, the relevant pages thereof. Text citations of two or more works at a time should be given in chronological order. When a paper written by three or more authors is cited, write the name of the first author plus et al. The Literature Cited section at the end of the paper should be arranged alphabetically and then chronologically and should contain only works specifically cited in the text. References to papers that have not yet been published will be as for articles (see below), except that “accepted” (along with the journal name) will replace the volume and page numbers. “In press” will not be used. (When such papers include authors of the submitted manuscript, copies of those papers must accompany the submitted manuscript.)

For the style of citations, please note the following examples:

Journal articles:
Information for Contributors

Books:


Book chapters:


The abbreviations of periodicals should be those used by the Council of Biology Editors. Periodical titles may also be written out.

Articles should include the name of the reviewing editor (the Editor or Associate Editor with whom author has corresponded) at the end of the Literature Cited section.

Footnotes.—Footnotes should be used sparingly. When necessary, they should be indicated in the text by superscript arabic numerals; the notes themselves should be typed on a page separate from the text. Footnotes to tables are referenced by superscript letters, except for significance levels, which use asterisks; table footnotes should be typed on the same page as the table to which they pertain.

Tables.—Each table must have a brief and self-explanatory title, be numbered with arabic numerals in order of its appearance in the text, and be typed on a separate page. Large, complex tables are discouraged. Guidelines for table format may be found in the CBE Style Manual and the Chicago Manual of Style or may be obtained by writing the editors of this journal.

Legends.—Figure legends should be typed on pages at the end of the manuscript, after tables. Each legend must be descriptive so that the illustration can be understood apart from the text and must define abbreviations used in the illustration.

Illustrations.—Each illustration (figure) should be an original, not a photocopy. Illustrations should be separate and have uniform lettering. They should be numbered consecutively, following the sequence in which they are mentioned in the text. The place where each illustration is to be inserted may be indicated by a circled note in the margin of the typescript. Names of authors, figure number, and an arrow indicating proper orientation should be written lightly in pencil on the back of each figure. Line drawings must be of high quality; typewritten or hand lettering is unacceptable. Photographs should be high-contrast, glossy prints. Magnifications may be indicated by a micron bar or in the legend. Photographs to be reproduced without further reduction must be so marked and may not exceed 4 3/4 inches wide by 7 1/4 inches long (122 by 194 mm) in order to fit the journal format. Please keep in mind that these dimensions are maxima. Because of the need for a figure legend, the illustration cannot be the maximum size in both dimensions.

Proofs and Reprints

Offprint order forms will be sent to the author (or in the case of multiple authors, to the senior author) with page proofs. There will be no page charge for publication.
Information for Contributors

Editor in Chief
Walter M. Fitch
Department of Physiological Chemistry
University of Wisconsin—Madison
1300 University Avenue
Madison, Wisconsin 53706

Roy J. Britten
Kerckhoff Marine Laboratory
California Institute of Technology
Corona del Mar, California 92625

Wesley M. Brown
Division of Biological Sciences
University of Michigan
Ann Arbor, Michigan 48109

Richard B. Flavell
Department of Cytogenetics
Plant Breeding Institute
Cambridge CB2 2LQ
England

Managing Editor
Masatoshi Nei
Center for Demographic and Population Genetics
University of Texas Health Science Center at Houston
P.O. Box 20334
Houston, Texas 77225

Associate Editors
Ken W. Jones
Department of Genetics
University of Edinburgh
West Mains Road
Edinburgh EH9, 3JN Scotland

Richard K. Koehn
Department of Ecology and Evolution
State University of New York at Stony Brook
Stony Brook, New York 11794

Robert K. Selander
Department of Biology
University of Rochester
Rochester, New York 14627
If the key word is genetics—

behavioral genetics  molecular genetics  biochemical genetics
clinical genetics and  population genetics  cytogenetics
  counseling                                       cancer genetics  mutagenesis/somatic
  immunogenetics                                  developmental genetics  cell genetics

—the key research is in

The American
Journal of Human
Genetics

Established in 1948, the AJHG is one of the leading journals addressing central issues in contemporary genetics. The Journal—which is the official publication of The American Society of Human Genetics—provides a record of research and review relating to heredity in humans and to the application of genetic principles in medicine, psychology, anthropology, and the social sciences as well as in related areas of molecular and cell biology.

Editor: Dr. David E. Comings.

Published bimonthly by The University of Chicago Press. 1-year rates: Individuals $75; Institutions $95. Outside of the U.S.A. postage surcharges vary according to destination. Visa and MasterCard accepted. Mail complete charge card information, payment, or purchase order to The University of Chicago Press, Journals Division, P.O. Box 37005, Chicago, IL 60637. Please note: Subscriptions entered on a calendar-year basis only. Students accepted as members of the society only. For student and all other membership rates, write to the Executive Office, American Society of Human Genetics, P.O. Box 6015, Rockville, MD 20850.

12/83

Subscribe to the AJHG.
Canadian Journal of
Genetics and Cytology

1984 Subscription Rates
(12 issues)
Canadian Journal of
Genetics and Cytology
Institutional
CANADIAN: $68.00
FOREIGN: $78.00
SINGLE COPY: $12.50
Personal
CANADIAN: $31.00
FOREIGN: $41.00
SINGLE COPY: $7.50

Canadian Journal of
Biochemistry and Cell Biology
Institutional
CANADIAN: $82.00
FOREIGN: $97.00
SINGLE COPY: $12.50
Personal
CANADIAN: $31.00
FOREIGN: $46.00
SINGLE COPY: $7.50

Payment must be enclosed with order
Cheques should be made payable to the Receiver General
for Canada, credit NRCC
Send to:
Distribution, R-88
(Can. J. Genet. Cytol.) or
National Research Council of
Canada, Ottawa, Ontario,
Canada K1A 0R6

EDITOR: Peter B. Moens, Department of Biology, York University,
4700 Keele Street, Downsview, Ontario, Canada M3J 1P3

The Canadian Journal of Genetics and Cytology publishes papers in
applied and basic genetics and cytology. Traditionally a substantial
number of the articles are of agricultural significance. In this respect
the Journal enjoys international contributions and a large readership.
Basic research articles are in the fields of molecular genetics,
population genetics, mutagenesis, and chromosome structure and
behaviour.

Indexed by Biological Abstracts, Current Contents, Science Citation Index, Chemical Abstracts, Biological and Agricultural Index, CAB Abstracts, and others

Canadian Journal of
Biochemistry and Cell Biology

EDITORS: Morris Kates and J. M. Neelin, Canadian Journal of
Biochemistry and Cell Biology, Faculty of Science and Engineering,
University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

The Canadian Journal of Biochemistry and Cell Biology, formerly
the Canadian Journal of Biochemistry, publishes papers in any field
of general biochemistry and in experimental cell biology. These
include full-length papers and rapid communications reporting
original work or invited reviews on topical subjects, as well as
certain symposia and special issues dedicated to a particular subject
or occasion. Recent special issues include papers devoted to "Gene
Replication and Expression: Organization and Function of Nucleus,
Chromosomes, and DNA" (March 1982), and "Mechanism of
Hormone Action" (July 1983); collections of symposia papers
include "Biochemical Evolution of the Translation Apparatus"
(April 1982), "Gene Regulation During Heat Shock" (June 1983),
"Basement Membrane" (late 1983), "Calmodulin" (late 1983), and
"Glycoprotein Biosynthesis" (late 1983). It is one of the leading
international journals on general biochemistry and cell biology, and
attracts subscriptions from more than 72 countries.

Indexed by Biological Abstracts, Current Contents, Science Citation Index, Chemical Abstracts, Excerpta Medica, Index Medicus, and others
That's how many pages we publish each year—original research, review articles, serial features, clinical briefs, medical perspectives, symposia and special articles—all devoted to increasing your understanding of infectious disease.

The Journal of Infectious Diseases is the foremost outlet for new research in infectious diseases—diseases that are responsible for more than 20% of hospital admissions and 50% of first office or clinic visits. Journal articles communicate information from clinical, epidemiologic, and laboratory experience. They provide a basis for understanding the causes, pathogenesis, diagnosis, and treatment of infections as well as the biology of the host response. Published monthly. Martha Yow, editor.

Reviews of Infectious Diseases, JID’s companion journal, fulfills a twofold purpose: first, to publish review articles offering comprehensive consideration of recent literature in a particular area or on a group of related topics; and second, to publish thematic studies focusing on central topics in infectious diseases. Often most or all of an issue is devoted to papers treating a single theme. Published bimonthly. Frequent supplements. Edward H. Kass, editor.

Sponsored by the Infectious Diseases Society of America.

One-year subscriptions: JID/RID combination: Institutions $225; Individuals $105. Special rates for students, interns, residents, and fellows as well as some society members. For further information, write to Circulation Manager, University of Chicago Press, P.O. Box 37005, Chicago, IL 60637 USA.
this publication is available in microform

Please send me additional information.

Name ______________________________________
Institution ______________________________________
Street ______________________________________
City ______________________________________
State _______ Zip _______

University Microfilms International
300 North Zeeb Road
Dept. P.R.
Ann Arbor, MI 48106
U.S.A.

18 Bedford Row
Dept. P.R.
London, WC1R 4EJ
England
In 1867, four renegades founded a journal.

Of the four men who left the Museum of Comparative Zoology in Cambridge to found The American Naturalist, two also participated in the founding, sixteen years later, of the American Society of Naturalists. The Society and the journal have worked closely together ever since, furthering the advancement and diffusion of knowledge about organic evolution and other broad biological principles.

In 1984, AN is still at the forefront.

One of the oldest scientific periodicals in the United States, The American Naturalist is also one of the foremost publications in the world for research in ecology, evolution, and population biology. Drawing upon examinations of all biotic kingdoms, the Naturalist emphasizes sophisticated analyses, innovative theoretical syntheses, and empirical discoveries with broad general implications.

Published monthly. George W. Salt, Editor.

The November 1983 special issue is now a book

Ecology and Evolutionary Biology
The papers in this collection outline the major positions in the current debate on approaches to community ecology. Considered together, these discussions contribute to the formulation of a new conceptual base for research in ecology and evolutionary biology.
Transcription and translation
a practical approach

J.B. Gurdon, in his introduction to the book, states
"... During recent years, experimental systems have greatly improved both in the
range and efficiency of the gene expression steps which they carry out. Further-
more, there has been a great proliferation in the types and sources of systems which
can be usefully applied to a particular problem. I therefore believe that the present
volume will be very widely welcomed. The chapters have been contributed by those
who have extensive experience of the procedures involved, and who, in many cases,
have been directly involved in their development..."

Contents
Introduction: J.B.Gurdon
- Expression of exogenous DNA in mammalian cells.
  D.A.Spandidos and N.M. Wilkie
- Expression of exogenous DNA in Xenopus Oocytes. A.Colman
- Transcription of eukaryotic genes in a whole-cell extract.
  J.L.Manley
- Transcription of RNA in isolated nuclei. W.F.Marzluff and
  R.C.C.Huang
- Transcription of chromatin. R.S.Gilmour
- In vivo gene expression systems in prokaryotes. N.G.Stoker, J.M.Pratt and I.B.Holland
- Coupled transcription-translation in prokaryotic cell-free systems. J.M.Pratt
- Purification of eukaryotic messenger RNA. M.J.Clemens
- Translation of eukaryotic messenger RNA in cell-free extracts. M.J.Clemens
- Translation of eukaryotic messenger RNA in Xenopus Oocytes. A.Colman

Appendices: I Nucleic acid and polypeptide molecular weight markers. S.Minter and P.Sealey
II List of suppliers

April 1984; 360pp; 0 904147 52 5 (soft)
£12.00/US$24.00

Centrifugation (2nd Edition)
a practical approach

A revised and completely updated edition which provides detailed experimental
protocols for all types of centrifugal separations from macromolecules to whole
cells. Extensive coverage is also given to the applications of centrifuges ranging
from simple bench machines to analytical centrifuges.

Contents
The theory and practice of centrifugation. D.Rickwood
- Choice of conditions for density gradient centrifugation. B.D.Hames
- Centrifugal methods for characterising macromolecules and their interactions. D.Rickwood and
  J.A.A.Chambers
- Measurements of sedimentation coefficients and computer
  simulation of rate-zonal separations. B.D.Young
- Isolation of subcellular
  organelles and membranes. J.Graham
- Centrifugal separations of mammalian
  cells. A.Brouwer, R.J.Berelds and D.Knook
- Separations in zonal rotors. J.Graham
- Analytical ultracentrifugation. R.Eason

Appendices: I Nomogram and equation for computing relative centrifugal force.
II Chemical resistance chart for tubes and zonal rotors. III Specifications of
ultracentrifuge rotors. IV Equations relating the refractive index to the density of
solutions. V Marker enzymes and chemical assays for the analysis of subcellular
fractions. J.Graham and T.C.Ford. VI Names and addresses of suppliers of
centrifuges and ancillary equipment. VII Glossary of terms.

April 1984; 250pp; 0 904147 55 X (soft)
£10.00/US$20.00
Forthcoming

Codon Usage and tRNA Content in Unicellular and Multicellular Organisms
Toshimichi Ikemura

Alignment Maps and Homology Analysis of the J-C Intron in Human, Mouse, and Rabbit Immunoglobulin Kappa Gene
Samuel Karlin and Ghassan Ghandour

DNA Sequence Comparisons of the Human, Mouse, and Rabbit Immunoglobulin Kappa Gene
Samuel Karlin, Ghassan Ghandour, and David E. Foulser

The Tetracycline Repressor of pSC101
Mary Ann D. Brow, Richard Pesin, and J. Gregor Sutcliffe

Methods for Computing the Standard Errors of Branching Points in an Evolutionary Tree and Their Application to Molecular Data from Humans and Apes
Masatoshi Nei, J. Claiborne Stephens, and Naruya Saitou

Biochemical Pathways in Prokaryotes Can Be Traced backward through Evolutionary Time
Roy A. Jensen
Molecular Biology and Evolution
Volume 1, Number 6, November 1984

439  Variance of Sequence Divergence
R. R. Hudson and G. B. Golding
(Letter to the Editor)

442  The Early Adaptive Evolution of Calmodulin
Marietta L. Baba, Morris Goodman, Janet Berger-Cohn, Jacques G. Demaille, and Genji Matsuda

456  Nucleotide Sequences of the Acinetobacter calcoaceticus trpGDC Gene Cluster
Jeffrey B. Kaplan, Paul Goncharoff, Anita M. Seibold, and Brian P. Nichols

473  Amino Acid Sequences of Lower Vertebrate Parvalbumins and Their Evolution:
Parvalbumins of Boa, Turtle, and Salamander
Nobuyo Maeda, Dexu Zhu, and Walter M. Fitch

489  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

503  MBE Reviewer Index

505  MBE Author-Subject Index