Repeated Nucleotide Sequence Arrays in Balbiani Ring 1 of *Chironomus tentans* Contain Internally Nonrepeating and Subrepeating Elements*

(Received for publication, November 22, 1982)

Steven T. Case‡ and Michael R. Byer§

From the Department of Biochemistry, The University of Mississippi Medical Center, Jackson, Mississippi 39216

Balbiani rings in *Chironomus* are large puffs on salivary gland polytene chromosomes that contain functionally related, but nonidentical genes that code for tissue-specific secretory polypeptides. *In situ* hybridization was used to select a recombinant plasmid (pCtBR1-1) that contained an insert of *Chironomus tentans* genomic DNA that originated from Balbiani ring 1. Mapping with restriction endonucleases demonstrated that the insert was 385 bp *(base pair)* and it contained duplicate clusters of certain cleavage sites about 250 bp apart. This repeat was shown to be part of tandem sequence arrays in the genome by hybridization of radioactive pCtBR1-1 to nitrocellulose blots containing limit and partial restriction endonuclease digests of nuclear DNA. Subsequent sequence analysis of the cloned DNA confirmed the presence of one complete copy of a 246-bp repeat comprised of a 114-bp internally nonrepeating segment and a 132-bp segment containing four 33-bp subrepeats. The subrepeats apparently evolved from a simple 9-bp sequence encoding a consensus tripeptide (Lys-Pro-Ser) in which the first two codons (AAA-CCA) were highly conserved at the nucleotide level. Comparisons between intragenic and interspecific (BRb in *Chironomus thummi*) copies of corresponding sequences revealed that, during the evolution of these tandemly repeated protein-coding sequences, internally nonrepeating segments were highly conserved and most likely became interspersed by variable segments containing subrepeats that arose from reduplication and divergence of 9-bp repeats.

Two types of sequence repetitions have been described that occur within protein-coding genes. In one group, a short nucleotide sequence is tandemly duplicated within the coding portion of a gene and is reflected by a repeating amino acid sequence of functional significance (1, 2). An alternative repetition pattern is found among members of multigene families (3, 4) where short intergenic sequence homologies exist between domains of certain functionally related but nonidentical genes (5–9). Nucleotide and amino acid sequence data reveal that an underlying evolutionary feature of these genes is that they appear to have evolved by reduplication, divergence, and even translocation of a small ancestral sequence that ultimately yields discrete genes that may possess overlapping phenotypic functions.

A group of functionally related, but nonidentical genes code for discrete polypeptides that make up saliva which is secreted from the salivary glands of *Chironomus* larvae (10, 11). Cytologically, two of these genes can be observed as large chrosomal puffs, known as BR1 and BR2 in *Chironomus tentans*, on salivary gland polytene chromosomes. BR1 and BR2 transcripts appear similar in that they are 75 S RNA molecules (12, 13) with internal sequence repetition (14–16). However, *in situ* hybridization of BR1 versus BR2 RNAs detect little, if any, sequence homology (14), and their rates and patterns of reassociation to genomic DNA are distinctly different (16). Both transcripts emerge from the nucleus as 37-kb cytoplasmic RNAs (17) that can direct *in vitro* synthesis of incomplete secretory polypeptides (18). Secretory polypeptides are unusually large (19, 20), yet they have a simple amino acid composition (21) and tryptic peptide pattern (19), indicative of short, repeated amino acid sequences. These repeated sequences presumably permit overlapping alignment of individual polypeptides during polymerization of a long, highly elastic, salivary fiber that is used by larvae to spin a protective cylindrical tube (11).

Because BR genes afford an opportunity to examine the evolution of tandemly repeated protein coding sequences at the intragenic, intergenic, and interspecific levels, it is desirable to obtain nucleotide sequence data for their basic repeating units. It is becoming increasingly apparent that the organization of repeated nucleotide sequences in this system will have to be derived from a series of small cloned DNA fragments since larger fragments are remarkably unstable (22–24). Nucleotide sequence data is currently available for BR1 (24, 25) and BR2 (26) in *C. tentans* as well as their homologous genes, BRb (27) and BRc (23, 28) in *Chironomus thummi*, and BR2 in *Chironomus pallidivittatus* (29). In each instance, the existence of tandem arrays within a BR gene has been indicated by hybridizing cloned DNA to genomic Southern blots (30). The emerging pattern of BR gene structure is one of tandem arrays which contain an alternating pattern of internally repeating and nonrepeating nucleotide sequences (23, 25–29).

We report here the nucleotide sequence of a 348-bp genomic fragment of BR1 in *C. tentans*. The cloned fragment contained

---

1. This work was supported by the National Institutes of Health Research Grant GM 26362 from the Institute of General Medical Sciences and a PROPHET computer terminal from the Biotechnology Resources Program, Division of Research Resources. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom all correspondence should be sent.

§ Recipient of a Dean's Medical Student Research Stipend.
a complete copy of a 246-bp repeated sequence that is apparently organized as tandem arrays in the genome. The 246-bp repeat was divided into a conserved internally nonrepeating segment and a somewhat more divergent segment that contained subrepeats which may have evolved from simple 9-bp repeats. Oryzaleria for BR1 support a recent model (26) for the evolution of 37-kb BR2 genes from a 110 to 120-bp primordial sequence.

EXPERIMENTAL PROCEDURES

Construction, Selection, and Purification of pCtBR1-1—Details have been published regarding the construction of a partial genomic library from which pCtBR1-1 was selected (22). Briefly, BR1 and BR2 sequences were found to reside on 32 to 40-kb fragments after limited digestion of C. tentans nuclear DNA with EcoRI. Such fragments were enriched by gradient sedimentation, randomly sheared, tailed with oligo(dC), and annealed to oligo(dG)-tailed PstI-cleaved pBR322. Recombinants were screened by colony hybridization with radioactive salivary gland 75 S RNA that simultaneously detected BR1 + BR2 sequences. Plasmid DNAs used in this study were purified from detergent lysates of host bacterial cells by two rounds of bound precipitated centrifugation in CsCl gradients containing ethidium bromide (31). Recombinant DNA experiments were conducted under containment conditions prescribed by National Institutes of Health “Guidelines for Research Involving Recombinant DNA Molecules.”

Radiolabeling of Nucleic Acids—Purified (17) salivary gland 75 S RNA was partially hydrolyzed and labeled with [γ-32P]ATP using polynucleotide kinase (32). Purified plasmid DNAs used for hybridization probes were labeled by the nick translation reaction (33, 34) using either [3H]dTTP or [α-32P]dCTP. Phage or plasmid DNAs used as molecular weight markers on gels were labeled by nick translation without the addition of exogenous DNase. DNA fragments cleaved by restriction endonucleases were end-labeled with [γ-32P]ATP in an exchange reaction (35). Radioisotopes were purchased from New England Nuclear.

In Situ Hybridization—Squashed preparations (36) of salivary gland polytene chromosomes were used for in situ hybridization of H-labeled plasmid DNAs. Hybridizations were done in 4 × SET (1 X SET is 0.15 M NaCl, 30 mM Tris-HCl, pH 8.0, and 2 mM EDTA), 30% formamide at 37 °C. Posthybridization treatment included several rinses in excess hybridization media at 37 °C and two rinses in 1 X SET at 65 °C. In some instances, a stringent rinse was performed in 0.1 X SET at 65 °C. Hybridized chromosome preparations were subjected to autoradiography and stained with Giemsa prior to being photographed (37).

Restriction Endonuclease Mapping and Gel Blotting—Restriction endonucleases were purchased from New England Biolabs, Beverly, MA, or Bethesda Research Laboratories. Endonuclease cleavage sites within the insert were initially identified by comparing the cleavage patterns of pCtBR1-1 and pBR322 by electrophoresis in 1% agarose gels (38). Subsequent mapping was done by comparing autoradiograms made from polycrylamide gels (39) containing appropriate single and double enzymatic digests of randomly labeled (nick-translated) versus end-labeled (kinased) samples of the 385-bp insert (excised by XhoI) or an internal 250-bp MboII fragment (see Fig. 2A). These data were confirmed by partial digests of end-labeled fragments, according to published procedures (40). Limit or partial endonuclease digests of C. tentans nuclear DNA were fractionated on agarose gels and transferred to nitrocellulose blots (30). 32P-labeled pCtBR1-1 was hybridized to blots in 4 × SET, 30% formamide, 10 × Denhardt’s solution, 40 mM Tris-HCl, pH 7.4, 50 mM NaCl, and 0.1% sodium pyrophosphate at 37 °C. Posthybridization washes were the same as described for in situ hybridization. All gels contained internal radioactive molecular weight markers made by cleaving λ DNA with HindIII (42) or pBR322 DNA with HinFI, HaeIII, or HpaII (43).

DNA Sequencing—The pCtBR1-1 insert and appropriate fragments were subcloned into the replicative form of bacteriophage M13mp9. The complete insert was subcloned as a PstI fragment. The leftward, central, and rightward Rsal fragments (and adjacent vector sequences, refer to Fig. 2A) were separated by polycrylamide gel electrophoresis and repurified (44), and synthetic EcoRI linkers were added by blunt end ligation (45). The central Rsal fragment was then cleaved with EcoRI to trim linker oligomers and cloned into the EcoRI site of M13mp9 DNA. The leftward and rightward fragments were cleaved with EcoRI + PstI and cloned into phage DNA cleaved by the same enzymes. Recombinant phage containing complementary strands of inserted restriction fragments were sorted by DNA hybridization of phase lattices (46). Transcribed versus nontranscribed strands were identified by their ability to form RNase-resistant hybrids with 32P-labeled 75 S RNA. DNA sequencing was performed by using recombinant phage as templates for DNA synthesis in the presence of a primer (47) and dideoxyribonucleoside triphosphate chain terminators (48). Short repeated nucleotide sequence homologies were detected by using Align and Makedotgraphic procedures which are based upon a previously published procedure (49) and available on the PROPHET computer network.

RESULTS AND DISCUSSION

Balbiani Ring Origin of pCtBR1-1—pCtBR1-1 is comprised of a fragment of C. tentans nuclear DNA inserted into the PstI recognition site of pBR322 via oligo(dG):oligo(dC) homopolymeric tails. The recombinant was originally selected as a potential BR clone by colony hybridization using radioactive 75 S RNA that contained both BR1 and BR2 sequences. To determine which BR the insert originated from, purified pCtBR1-1 DNA was radiolabeled in vitro with [3H]dTTP by nick translation and hybridized in situ to squashed preparations of salivary gland polynucle chromosomes. Photomicrographs showed that autoradiographic exposure of silver grains occurred over chromosome IV, especially in the region of BR1 (Fig. 1A). Peripheral grains were often seen over BR2, similar to BR1.
repeats are indicated by a previous observation (24), but control experiments indicated that these could be greatly reduced by a stringent posthybridization rinse (Fig. 1B). In parallel experiments it was shown that ³²P-labeled pBR322 did not hybridize to polytene chromosomes. We, therefore, conclude that the genetic insert contained within pCtBR1-1 originates from BR1. Hybridization to BR2 was subsequently explained when a partial sequence homology was identified between this recombinant and a cloned BR2 sequence (26).

Endonuclease Cleavage Map and Genomic Organization of the pCtBR1-1 Insert—Restriction endonuclease cleavage mapping of the recombinant (Fig. 2A) indicated that the size of the insert was 385 bp. The central portion of the insert contained four, approximately evenly spaced Sau3A (MboI) sites flanked by two identical clusters of AluI, MboII, HinfI, and RsaI recognition sites about 250 bp apart. These results first suggested that the pCtBR1-1 insert contained a repeated nucleotide sequence.

To determine the organization of the pCtBR1-1 insert within the C. tentans genome, Southern blots (30) containing restriction endonuclease-cleaved fragments of nuclear DNA were probed with ³²P-labeled pCtBR1-1. Autoradiograms of blots made from limit MboII digests revealed a single band of hybridization at 250 bp (Fig. 3A), while partial MboII digests showed a ladder of bands with a 250-bp interval ranging from 250 bp to at least 2.5 kb (Fig. 3B). Identical results were obtained with HinfI and RsaI (data not shown). These results indicated that the 250-bp MboII fragment of the pCtBR1-1 insert might be organized within BR1 as tandem arrays consisting of at least 10 copies.

In an attempt to estimate the maximum number of tandem sequence arrays in BR1, blots containing nuclear DNA cleaved with TaqI or HaeIII were hybridized with ³²P-labeled pCtBR1-1. Neither enzyme cleaved the pCtBR1-1 insert, yet each enzyme cleaved most of C. tentans DNA to fragments of less than 4 kb. TaqI digests mainly showed hybridization with 19- and 22-kb bands, while HaeIII digests had a prominent 30-kb band (Figs. 3, C and D). These same bands hybridized to ³²P-labeled 75 S RNA (data not shown) which is highly enriched in BR1 + BR2 sequences, although we could not tell if they contained any nontranscribed flanking sequences. Nonetheless, these results raised the possibility that a substantial portion of a 37-kb BR1 gene may be comprised of tandem arrays of the 250-bp repeated sequence contained within the recombinant. We subsequently found that the faint 28-kb bands obtained in both digests could be eliminated by the same stringent rinse used after in situ hybridization. Thus, we conclude that these bands also result from a sequence homology with BR2.

Nucleotide and Amino Acid Sequence—The DNA sequencing strategy employed in this study involved subcloning the entire 385-bp insert and appropriate subfragments (Fig. 2A) into the replicative form (double-stranded DNA) of bacteriophage M13mp9. This procedure allowed us to unambiguously identify recombinants containing transcribed versus nontranscribed strands by filter hybridization with ³²P-labeled 75 S RNA (Fig. 2B). All segments were sequenced at least twice and 95% of the sequence was confirmed by independently sequencing complementary strands of the original insert. Due to the presence of a long homopolymeric sequence at the 3’-end of the insert (Fig. 2A), it was not possible to obtain confirmation of the 10-bp sequence between the nearest RsaI and AluI sites on the nontranscribed strand (Fig. 2B).

The nucleotide and encoded amino acid sequence of the nontranscribed strand of the pCtBR1-1 insert is presented in Fig. 4. The insert contains 348 bp of C. tentans DNA and homopolymeric stretches of (dG)₉ and (dC)₆ located at the transcriptional 5’- and 3’-ends, respectively. Direct evidence was found for the repeated sequence previously suggested by endonuclease cleavage mapping (Fig. 2A) and genomic blotting experiments (Fig. 3). The repeat length was 246 bp, beginning at nucleotide position 10 and extending through to position 256. The sequence beginning at position 256 corresponds to the start of an incomplete (93/246 bp) copy of what

---

**Fig. 2. DNA sequence organization within the pCtBR1-1 insert.** A, restriction endonuclease cleavage map with sites indicated for EcoRI (E), AluI (A), MboII (M), HinfI (H), RsaI (R), and Sau3A (S). The thick black line represents genomic DNA, while wavy lines at both ends represent homopolymeric sequences of (dG)₉(dC). B, the 5’ to 3’ transcriptional polarity is shown relative to the endonuclease cleavage map. Below that, the strategy for DNA sequencing is shown. The short dotted line indicates a small region of the nontranscribed strand that did not yield clear sequence data. C, the location of internally nonrepeating segments and a segment containing subrepeats (SR1, SR2, SR3, and SR4).

---

**Fig. 3. Autoradiograms showing hybridization of ³²P-labeled pCtBR1-1 to nitrocellulose blots made from agarose gels that contained C. tentans nuclear DNA digested with various restriction endonucleases.** A, complete MboII digest fractionated on a 2% agarose gel. B, partial digestion with MboII for 10 (1), 20 (2), 40 (3), or 60 (4) min and fractionated on a 1% agarose gel. C and D, complete digests with TaqI or HaeIII, respectively, fractionated on 0.3% agarose gels. Numbers to the side of each lane indicate the size of DNA fragments that hybridized the probe.
would have been the next tandem array. The 246-bp repeat had 42% G + C content, in agreement with the base composition of BR1 RNA (50). The amino acid sequence was derived from the only open translational reading frame found in the transcribed strand. The next two reading frames were interrupted by two and 24 termination codons, respectively. About 64% of the encoded polypeptide consisted of only Lys, Pro, Ser, and Arg residues (Table I). These same four amino acids have been identified as the most abundant residues in hydrolysates of the putative BR1 polypeptide. Ser codons frequently occurred within sequences that were compatible with reported phosphorylation sites in other proteins (51). We have found that greater than 80% of \( ^{32} \)P incorporated into the putative BR1 product in vivo can be recovered as phosphate.

The 246-bp Repeat Contains an Internally Nonrepeating Segment and a Segment with Subrepeats—While this manuscript was in preparation, models of BR gene structure were published which predicted that segments of highly repeated sequences alternated with nonrepeated sequence segments (23, 25–29). The 246-bp repeat within the pCtBR1-1 insert could similarly be divided into two regions consisting of distinctive nucleotide sequence patterns. These regions were designated as INR and a segment containing SR based upon examination of the complete nucleotide sequence. The INR and SR segments alternate with one another and that two SR/INR junctions occurred within the pCtBR1-1 insert (Figs. 2C and 4).

The INR1 segment (Fig. 2C) contained 114 bp extending from nucleotide position 10 through 123, inclusive (Fig. 4). No obvious patterns of internal sequence repetition were found within INR1, and the clustered AluI, MboII, HinfI, and Rsal recognition sites were located within this segment (compare Figs. 2A and C). A second, partial INR segment (INR2) begins at nucleotide position 256 and extends 93 bp toward the 3'-end of the insert. These INR segments exhibited remarkable sequence homology with only a single base difference occurring between them. The nucleotide position at 54 within INR1 was Ado, while the corresponding location within INR2 (position 282) had Thd. This variation invoked a silent mutation in the wobble position of an Ala codon; thus, the amino acid sequence is perfectly conserved throughout the homologous regions of INR1 and INR2.

The 132-bp SR segment was characterized by four direct repeats (designated SR1, SR2, SR3, and SR4) of a 33-bp sequence (Figs. 2C and 4). Although these direct repeats had considerable sequence homology, more sequence heterogeneity exists between them than between INR segments. SR2 and SR4 had complete homology in their DNA sequence and contained the apparent consensus nucleotide and amino acid sequence for subrepeats within pCtBR1-1. SR3 was 97% (32/33 bp) homologous with the consensus nucleotide sequence. At position 215, Ado was substituted for Guo resulting in a Lys codon instead of Arg. SR1 exhibited the greatest divergence from the subrepeat consensus sequence with 91% (30/33 bp) homology. The first three codons contained nucleotide substitutions: Guo instead of Ado at position 124, Thd instead of Ado at position 129, and Cyt instead of Thd at position 132. Only the first substitution resulted in a different amino acid (Glu replaced Lys as the first codon), while the latter two were silent mutations. Finally, the sequence from nucleotide positions 1 through 9 was completely homologous with the last 9 bp of the subrepeat consensus sequence. This implied that it represented the distal three codons at the 3'-end of the last subrepeat preceding INR1 (Fig. 2C).

*Distribution of Amino Acids and Codons between INR and SR Segments—*Certain amino acids and codons were not randomly distributed between INR and SR segments of the pCtBR1-1 insert. The INR1 segment was characterized by containing 70% (7/10) of the Arg residues found within the complete (INR1 + SR) 246-bp repeat (Table I). Less frequently occurring amino acids (Cys, Asn, Met, Phe, and Thr), that individually accounted for less than 7% of the residues,
Codon selection in INR and SR segments of a cloned 246-bp repeat from Balbiani ring 1 in C. tentans

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Per cent composition</th>
<th>Possible codons</th>
<th>Frequency of occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INR1</td>
<td>SR</td>
<td>Total</td>
</tr>
<tr>
<td>Lys</td>
<td>4.9</td>
<td>14.6</td>
<td>19.5</td>
</tr>
<tr>
<td>Pro</td>
<td>1.2</td>
<td>14.6</td>
<td>15.9</td>
</tr>
<tr>
<td>Ser</td>
<td>6.1</td>
<td>9.8</td>
<td>15.9</td>
</tr>
<tr>
<td>Arg</td>
<td>8.5</td>
<td>3.7</td>
<td>12.2</td>
</tr>
<tr>
<td>Gly</td>
<td>3.6</td>
<td>4.9</td>
<td>8.5</td>
</tr>
<tr>
<td>Glu</td>
<td>2.4</td>
<td>6.1</td>
<td>8.5</td>
</tr>
<tr>
<td>Ala</td>
<td>7.3</td>
<td>0</td>
<td>7.3</td>
</tr>
<tr>
<td>Cys</td>
<td>4.9</td>
<td>0</td>
<td>4.9</td>
</tr>
<tr>
<td>Asn</td>
<td>3.7</td>
<td>0</td>
<td>3.7</td>
</tr>
<tr>
<td>Met</td>
<td>1.2</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>Thr</td>
<td>1.2</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

were also sequestered in INR1. The 4 Cys residues were preceded by either Arg (nucleotide positions 10 and 88) or Lys (positions 46 and 82). Three of these dipeptides were spaced 10 residues apart from each other and the fourth dipeptide immediately followed the third. Although Gly residues were nearly evenly distributed between INR and SR segments, GGU codons were restricted to INR1 while GGA codons were restricted to the SR segment (Table I). The SR segment contained the majority of Pro (12/13), Lys (12/16), Glu (5/7), and Ser (6/13) residues coded for by the 246-bp repeat (Table I). As a rule, Pro residues were always preceded by a charged amino acid and most frequently found within the tripeptides Lys-Pro-Ser or Arg-Pro-Glu. We noticed that the only Pro residue within INR1 (Table I) was also part of a Lys-Pro-Ser tripeptide located very near SR1 (see Fig. 4, nucleotide positions 106 through 114).

Subrepeats Contain Conserved Tripeptides with a 6-bp Consensus Sequence—The redundant pattern of Lys-Pro-Ser and Arg-Pro-Glu tripeptides implied that the 33-bp subrepeats may have evolved by duplication and divergence of a smaller repeated nucleotide sequence. A computer-assisted alignment revealed a series of tandem 9-bp homologies beginning 18 bp prior to the INR1/SR1 junction and encompassing the adjacent SR segment (nucleotide positions 106 through 255). A 3-bp insertion was required, between the eighth and ninth codon within each subrepeat, that presumably reflects deletion of a codon during the evolution of these repeats (Fig. 5). This essentially converted each subrepeat into a block of four 9-bp repeats. The 9-bp repeats exhibited polarity with regard to conservation in nucleotide selection at each position (Fig. 5). That is, bases were more conserved at the 5' end of each 9-bp repeat than the 3' end. In fact, no consensus nucleotide sequence was found for the last three nucleotides of the 9-bp repeat. However, 71% (10/14) of the 3' terminal trinucleotides aligned at this position were synonymous Ser codons. So while the nucleotide sequence was conserved for only the first 6 bp, Lys-Pro-Ser prevailed as the consensus amino acid sequence.

Intragenic and Interspecific Conservation of Nonrepeating Segments and Subrepeats of BR1—There has been strong intragenic conservation among five independent INR segments from BR1 in C. tentans. A 150-bp BR1 sequence published previously (24) contained a 114-bp overlap (designated here as INR3; Fig. 6) with our INR1 segment. In addition, the recently published (25) nucleotide sequence of a cDNA clone from BR1 contained two nonrepeated segments (designated here as INR4 and INR5; Fig. 6). A minimum of

**TABLE I**

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>106</td>
</tr>
<tr>
<td>Pro</td>
<td>106</td>
</tr>
<tr>
<td>Ser</td>
<td>106</td>
</tr>
<tr>
<td>Arg</td>
<td>106</td>
</tr>
<tr>
<td>Gly</td>
<td>106</td>
</tr>
<tr>
<td>Glu</td>
<td>106</td>
</tr>
<tr>
<td>Ala</td>
<td>106</td>
</tr>
<tr>
<td>Cys</td>
<td>106</td>
</tr>
<tr>
<td>Asn</td>
<td>106</td>
</tr>
<tr>
<td>Met</td>
<td>106</td>
</tr>
<tr>
<td>Thr</td>
<td>106</td>
</tr>
</tbody>
</table>

Consensus Lys Pro Ser

---

**Fig. 5.** Derivation of a consensus tripeptide and conserved hexanucleotide sequence by aligning 9-bp repeats within pCtBR1-1. Numbers in parentheses indicate nucleotide positions within the complete sequence (Fig. 4). The sequence reads from left to right, top to bottom. The proposed locations of deleted codons (Δ) are indicated. Each row is designated by its origin in INR or SR segments of the insert (Fig. 2A), then consecutively numbered and lettered as indicated. The consensus nucleotide for each of the first 6 columns is shown at the bottom along with its frequency of occurrence. Divergent nucleotides are circled. For the last three columns, related codons are shown along with their frequency of occurrence.
that the ancestral sequence for BR1 in C. tentans. The BR1 sequence organization found in this study in BR2 in sequence that remained conserved while it became interspersed the sequence AAA-CCA. This sequence occurred not only as of BR1 subrepeats (Fig. 5), but also as part of the consensus that they were most frequently encoded (11/16 dipeptides) by (ie. 

Lys-Pro and Arg-Pro dipeptides in this region (codons 13 and 14 in INR1) was the most diverged portion of the BRb INR was near the 3'-end of the overlapped region; however, the characteristic Lys-Cys dipeptide in this region (codons 13 and 14 in INR1) was conserved despite a base substitution. BR1 subrepeats also exhibited intragenic and interspecific sequence conservation. The published (24) sequence surrounding INR3 (in Fig. 6) directly corresponds to subrepeats in BR1. The 20 nucleotides preceding INR3 were identical with the consensus sequence for a BR1 subrepeat obtained in this report. The 16 nucleotides following INR3 were also identical with the 5'-end of SR1 and duplicated an INR1/SR1 junction. Similarly, subrepeats were found (25) on both sides of INR5 whose sequence varied by not more than three nucleotides from the 132-bp SR segment reported here. Most of the published BRb sequence (27) was comprised of six inexact copies of 8 to 11-codon subrepeats that contained the consensus tripeptide Lys-Pro-Ser-Lys. Remnants of the consensus sequence for a BR1 subrepeat were also present in BRb. For example, the 11 codons preceding the BRb/INR1 overlap (nucleotides 160 to 192 in Ref. 27) exhibited 79% nucleotide and 100% amino acid sequence homology with the SR3 sequence described here (Fig. 4). Lys-Pro and Arg-Pro dipeptides were abundant throughout BRb subrepeats. We noted that they were most frequently encoded (11/16 dipeptides) by the sequence AAA-CCA. This sequence occurred not only as the conserved hexanucleotide within the consensus tripeptide of BR1 subrepeats (Fig. 5), but also as part of the consensus sequence AGC-AAA-CA found for subrepeats in BR2 (26).

It was recently postulated (26) that 215-bp tandem arrays in BR2 in C. tentans evolved from an ancestral 106-bp sequence that remained conserved while it became interspersed as a result of reduplication of a terminal 9-bp sequence that eventually resulted in six adjacent and divergent 18-bp repeats. The BR1 sequence organization found in this study (i.e. 246-bp arrays comprised of INR segments flanked by tandem 9-bp repeats) supported such a model and suggested that the ancestral sequence for BR1 in C. tentans was also about 105 to 114 bp in length and may be represented by segment INR1. We have noted that INR1 shares about 70% sequence homology with a similar segment in BR2. A detailed intergenic (BR1 versus BR2) comparison of INR segments will be made in conjunction with the description of a variant BR2 sequence (56).

The fundamental functioning unit within secretory polypeptides in Chironomus may be reflected by 200 to 300-bp repeats within BR genes. In turn, conserved and variable domains within these units may be represented by alternating INR and SR segments, respectively. Variable domains have tolerated limited divergence in amino composition and length due to variations in short tandemly repeated sequences. However, certain key characteristics were retained, reflected by the occurrence of prevalent di- and tripeptides. Homologous but unequal crossing over between tandemly repeated sequences can lead to variations in nucleotide sequence length (52) that can presumably lead to expansion or contraction of gene size (53). To better understand the evolution of BR genes it is desirable to identify major sequence variants that might occur, as well as obtain sequence data from other functionally related genes that code for salivary polypeptides (54, 55).

Acknowledgments—Mary Jane Bevill and Virginia Hill provided diligent technical assistance throughout much of this study. Joachim Messing kindly provided us with bacteriophage M13mp9 and detailed protocols regarding its use. We also wish to thank Ronnie Brown for her patience in typing the manuscript and figures containing sequence data.

REFERENCES
Repeated Nucleotide Sequences in Balbiani Ring 1

Repeated nucleotide sequence arrays in Balbiani ring 1 of Chironomus tentans contain internally nonrepeating and subrepeating elements.

S T Case and M R Byers


Access the most updated version of this article at http://www.jbc.org/content/258/12/7793

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/258/12/7793.full.html#ref-list-1