Discontinuous translation and mRNA structure of the coding region

Mitsuo Zama
National Institute of Radiological Sciences, 9-1 Anagawa-4-chome, Inage-ku, Chiba 263-8555, Japan

ABSTRACT
An analysis of the published data of the electrophoresis of the products of silk fibroin exhibiting discontinuities or pauses in the translational process showed that about 60 times of pauses occur during the translation of the mRNA from silk fibroin. In addition, we estimated that silk fibroin gene as a whole (15kb) is composed of about 60 alternate, crystalline–noncrystalline, arrays of nucleotide sequence elements each connected by the boundary sequence, on the basis of the nucleotide sequence of the partial cDNA clone of the heavy chain gene of Bombyx mori available in the literature. The coincidence of the total number of pauses with that of the alternate arrays suggests that one translational pausing site exists in each of the unit (alternate array + boundary ) nucleotide sequence in the silk fibroin mRNA template.

INTRODUCTION
It was shown earlier (1) by studies of translation of silk fibroin in a cell free system and in intact silk gland cells that, in addition to full-size product, a large number of smaller polypeptides are detectable upon analysis by sodium dodecyl sulfate/polyacrylamide gel electrophoresis. Evidence was presented that these smaller polypeptides are growing fibroin chains that transiently accumulate as discrete size classes due to discontinuities in the translation process. It was pointed out that these discontinuities or pauses occur at specific sites in the fibroin mRNA template. We have shown, on the other hand, that the coding region of silk fibroin gene is composed of a repeating bipartite unit sequence, by sequencing a short fragment of Bombyx mori silk fibroin heavy chain cDNA(2). From the distinctly different codon usage patterns in the two regions of the unit sequence, it has been suggested that selection has operated on the codon usage to optimize the secondary structure characteristic of the fibroin mRNA and correlation between the mRNA structure and discontinuities in the translation process of fibroin coding region has been implicated (2). In the present study, we pursued further the correlation by analyzing the data of electrophoresis showing discontinuities of translation (1) and of the nucleotide sequences of a number of silk fibroin heavy chain cDNAs (3).

RESULTS AND DISCUSSION
As a step to identify the specific sites in the fibroin mRNA template where translational pauses occur, we first tried to obtain the total number of pauses occurring on the mRNA template. Translational products of fibroin mRNA exhibit the ladder of electrophoretic bands smaller than full-size fibroin (1). The electrophoretic band pattern of translational products from intact gland cells is nearly identical to that obtained from in vitro translation of purified fibroin mRNA. The intensities of electrophoretic bands of smaller size fibroin are relatively weak and the distribution of the bands of larger size fibroin polypeptides is too dense to discriminate between neighboring bands. Our inspection of the electrophoresis data (see Fig.1 in Ref.1) revealed the existence of about 16 electrophoretic bands between size markers of molecular weights 116,000 and 212,000. This led us to estimate at about 60 electrophoretic bands for the total number of the bands appearing in the gel. That is, it was estimated that about 60 times of pauses occur during the translation of silk fibroin mRNA. The calculation is based on the following assumptions:
the length of total coding region is 15 kb, and the average molecular weight per amino acid of fibroin is 72. The latter value was obtained from the average molecular weight per amino acid for the 216-base unit sequence from fibroin (2). This is reasonable because the amino acid composition of the 216-base unit sequence is in excellent agreement with that from fibroin (2).

The highly repetitive structure and its organization of the silk fibroin gene has recently been shown, by sequencing a number of cDNAs representing the Bombyx mori silk fibroin heavy chain transcript (3). The fibroin gene is composed of alternate arrays of the crystalline element and the noncrystalline element. The crystalline element is characterized by repeats of a highly conserved 18-base sequence coding perfect repeats of the unit peptide Gly-Ala-Gly-Ala-Gly-Ser. The noncrystalline element is composed of repeats of a less-conserved 30-bp sequence which codes a peptide similar to that in the crystalline element except that Ser is replaced by Tyr and that there are irregular substitutions of Ala to Val or Tyr. Heterogeneities in numbers of repeats are observed for both the crystalline and the noncrystalline elements within an alternate array. The boundary sequence, appearing in boundary from noncrystalline element to crystalline element, is identified by a quite homogeneous 18-base sequence.

Our closer examination of the nucleotide and predicted amino acid sequences of the 3'-end cDNA clone of the silk fibroin heavy chain gene of Bombyx mori (3) revealed the existence of 6 alternate arrays within a block of 1.5 kb partitioned both ends by a homogeneous nonrepetitive amorphous domain around 100 bp in length. In addition, it has been suggested that the gene structure similar to that seen in the 3'-end cDNA clone repeats throughout the rest of the silk fibroin gene (3). On the basis of these data we estimated that fibroin gene as a whole (15 kb) is composed of about 60 alternate arrays of nucleotide sequence elements each connected by the boundary sequences.

The coincidence of the total number of pauses with that of the alternate arrays suggests that one translational pausing site exists in each of the unit (alternate array + boundary) nucleotide sequence in the silk fibroin mRNA template. Nonrepeating and conserved nucleotide sequences would be preferable as the translational pausing site. Although the crystalline elements are highly conserved in the nucleotide sequence but repeats of the element as well as heterogeneities in numbers of repeats of the element are observed. The nucleotide sequences of the noncrystalline element is, on the other hand, less conserved. Thus, it appears that neither the crystalline nor the noncrystalline element is suitable for the locus where a pausing site resides. In this respect, the 18-base boundary sequence is noted since the sequence is nonrepeating by appearing only once per crystalline-noncrystalline array, and moreover it is a quite homogeneous sequence. Thus, it appears probable that the site of the translational pause is located in the boundary sequence.

Discontinuous translation has also been observed for other proteins. In a series of studies for such proteins, type I collagen (4), colicin A (5), chloroplast photosystem II reaction center protein D1 (6), globin (7) and spider fibroin (8) we have so far presented evidence to support the view that the discontinuities or pauses may be attributable to the mRNA secondary structure of the protein-coding region. In the case of silk fibroin, the distinctly different codon usage patterns have been shown for each of the crystalline, noncrystalline and boundary sequences, implying mRNA secondary structures characteristic of each nucleotide sequence (2). Elucidation of site specific mRNA structures responsible for the pauses during translational elongation for silk fibroin gene as well as for other protein genes remains unsolved for future studies.

REFERENCES