The encapsidation signal of hepatitis B virus facilitates preC AUG recognition resulting in inefficient translation of the downstream genes

Wen-Ling Hwang† and Tsung-Sheng Su1,2,3

1,2 Institutes of Microbiology & Immunology†, and Genetics2, National Yang-Ming University, Taipei, Taiwan 112, Republic of China
3 Department of Medical Research and Education, Veterans General Hospital-Taipei, Taipei, Taiwan 112, Republic of China

Introduction

Hepatitis B virus (HBV) is a small, partially double-stranded, circular DNA virus. Replication of HBV is carried out by the polymerase (P) protein through reverse transcription of the pregenomic RNA. Virus encapsidation is mediated by a specific cis-acting sequence, the ε stem–loop structure, present at the 5′ end of the pregenomic RNA. The P protein binds to this structure to initiate the encapsidation reaction and DNA priming (Hirsch et al., 1990; Junker-Niepmann et al., 1990; Bartenschlager & Schaller, 1992; Pollack & Ganem, 1993; Wang & Seeger, 1993; Tavis et al., 1994; Nassal & Rieger, 1996; Rieger & Nassal, 1996). In particular, the P protein prefers to bind in a cis-assembly fashion to RNA templates from which it is translated (Bartenschlager et al., 1990; Hirsch et al., 1990).

The P protein is translated from the bicistronic pregenomic RNA. The P open reading frame (ORF) is located downstream to an overlapping out-of-frame core (C) ORF. Translation of the HBV P protein is initiated by a leaky scanning mechanism (Lin & Lo, 1992; Fouillot et al., 1993; Hwang & Su, 1998). The majority of the initiation is carried out by ribosomes that terminate the translation of an upstream minicistron within the C region (Fouillot et al., 1993; Hwang & Su, 1998). Although the initiator codon of the P gene is in a suboptimal context, the efficiency of the translation initiation of the P gene can reach one-tenth of that of the C gene when using lacZ as a reporter in a translation efficiency assay (Hwang & Su, 1998). On the other hand, other studies have shown that few P and C proteins are expressed from another genomic RNA, the preC RNA (Ou et al., 1990; Lin & Lo, 1992; Fouillot & Rossignol, 1996). Instead, the preC RNA directs synthesis of the precore protein which is the precursor of a secreted e antigen, termed...
HBeAg (Ou et al., 1986; Schlicht et al., 1987; Standring et al., 1988).

The structure of preC RNA differs from that of pregenomic RNA by the presence of some 30 nt at its 5’ end that encompass the initiator codon of the preC region (Yaginuma et al., 1987; Junker-Niepmann et al., 1990). Translation initiation of the C and P genes from the preC mRNA, when it occurs, probably involves ribosomal leaky scanning. The unfavourable expression of the C and P proteins from the preC RNA suggests that few 40S ribosomal subunits scan past the preC AUG codon. By sequence inspection, the context of the preC AUG (ACCAUGG) is not fully compatible with the consensus sequence of an initiator codon [(A/G)CCAUGG] (Kozak, 1989). On the other hand, Kozak (1990) has demonstrated that a downstream secondary structure of moderate strength can facilitate the recognition of an initiator codon in a suboptimal context. Such a secondary structure would slow down ribosomal scanning, thereby providing more time for the 40S ribosomal subunits to properly recognize an initiator codon. The ε stem–loop structure is situated 30 nt downstream of the initiator codon of the preC ORF (Fig. 1). Conceivably, its presence may enhance recognition of the preC AUG codon, resulting in the translation of the downstream genes.

Another characteristic of the preC RNA is that it directs synthesis of a secretory protein, the HBeAg. The protein translated from the preC ORF is a p25 protein. The first 19 amino acids of p25, constituting a signal sequence, are removed during targeting to the endoplasmic reticulum (ER), generating a p22ε derivative in the process. The p22ε protein is further cleaved in a post-ER compartment to produce HBeAg (Ou et al., 1986; Schlicht et al., 1987; Standring et al., 1988). In mammalian cells, secretory proteins are co-translationally translocated across the ER membrane in a signal recognition particle (SRP)-dependent manner (reviewed by Walter & Johnson, 1994). When a signal peptide emerges from the large ribosomal subunit, it is associated with an SRP to initiate elongation arrest of the nascent chain. The SRP escorts the complex to the ER membrane where it binds to the SRP receptor. Translation is then able to resume only when the ribosome binds to a membrane translocation complex and the signal sequence is inserted into a membrane channel. In the process of elongation arrest, movement of the 40S ribosomal subunits to the downstream cistron is blocked. It is conceivable that during translation of the precore protein, elongation arrest, when it occurs, can retard the movement of the scanning ribosomal subunits to the C and P cistrons located downstream leading to reduced synthesis of the C and P proteins from the preC RNA.

The preC RNA is some 30 nt longer at its 5’ end than the pregenomic RNA and yet this RNA is used exclusively for precore protein production. To investigate why the C and P proteins are expressed unfavourably from the preC RNA, a genetic approach was taken in this study. Primarily, we examined the possible role of the ε structure and the secretory nature of the precore protein in expression of the downstream P gene using lacZ as a reporter. In addition, a radioimmuno-precipitation assay using anti-HBc antisera was performed to study the effect of stability of the ε structure on expression of the precore and C proteins.

Methods

Plasmid construction and mutagenesis. Plasmid pPreC-CP-Z (Fig. 2 a) was constructed by inserting a 653 bp HBV ayw subtype DNA fragment (nt 3091–561, the A residue of the C gene initiator codon being nt 1), which included the preC region (nt 3096–3182), the C ORF (nt 1–552) and 155 nt of the N terminus of the P ORF (nt 407–561), into the N terminus of the lacZ ORF of plasmid pCH110 (Pharmacia). This construct carries the P ORF fused in-frame to the N terminus of the lacZ ORF (P-Z fusion) and is under the transcriptional control of the simian virus 40 (SV40) early promoter. pPreC-CP-Z(εm), pPreC-CP-Z(εm#), and pPreC-CP-Z(εm*) are derivatives of pPreC-CP-Z with mutations introduced into the stem region of the ε structure to increase the free energy of the RNA secondary structure from −26.5 kcal/mol in the WT to −18.3, −13 and −9.7 kcal/mol in εm, εm# and εm*, respectively. RNA folding and the energy rules are based on the programs of Zuker (1989) and Walter et al. (1994). Construct pPreC-CP-Z(εm*)–CP-Z, a derivative of pPreC-CP-Z, carries a mutated signal sequence. This was achieved by inserting a G residue at the beginning of signal sequence to create a frame-shift mutation, then deleting a downstream T residue to restore the reading frame. Several proline residues have been introduced to impair the z-helix structure of the signal peptide. This sequence is no longer recognized as a signal sequence based on the computer-assisted prediction for a signal peptide (Nielsen et al., 1997). The nucleotide and the amino acid sequences between the initiator codons of the preC and C ORFs, which include the signal sequence and the ε sequence of the WT HBV genome.

Fig. 1. Structural features of HBV genomic RNA. (a) Transcription start sites of the preC RNA and pregenomic RNA are indicated by arrows with the preC RNA starting at nt 3068 and 3074. The numbering system is such that nt 1 starts at the A residue of the C gene initiator codon. Locations of the preC, C and P ORFs are indicated by boxes with their ends numbered. The filled-in box within the preC region is where the coding sequence for the signal peptide is located. The encapsidation signal is marked (c). (b) Folded RNA structure of the ε sequence. Initiator codons of the C and preC ORFs are indicated by arrows.
and its mutant derivatives, are listed in Fig. 3. Site-directed mutagenesis was performed as described by Kunkel (1985) or by the Promega Altered Sites II in vitro Mutagenesis system. A summary of the characteristics of the WT construct pPreC-CP-Z and its mutant derivatives, with emphasis on the nature of the various ε structures, the start site mutations and nature of signal sequence of each construct are indicated. WT, Wild-type; M, mutant.

<table>
<thead>
<tr>
<th>construct</th>
<th>abbreviation</th>
<th>preC AUG</th>
<th>signal seq</th>
<th>ε</th>
<th>C AUG</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPreC-CP-Z</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>−26.5</td>
<td>WT</td>
</tr>
<tr>
<td>pPreC&lt;sup&gt;m&lt;/sup&gt;-CP-Z</td>
<td>preC&lt;sup&gt;m&lt;/sup&gt;</td>
<td>AGU</td>
<td>WT</td>
<td>−26.5</td>
<td>WT</td>
</tr>
<tr>
<td>pPreC-C&lt;sup&gt;m&lt;/sup&gt;-P-Z</td>
<td>C&lt;sup&gt;m&lt;/sup&gt;</td>
<td>WT</td>
<td>WT</td>
<td>−20.1</td>
<td>CCG</td>
</tr>
<tr>
<td>pPreC-CP-Z&lt;sup&gt;(ε&lt;sup&gt;1&lt;/sup&gt;)&lt;/sup&gt;</td>
<td>ε&lt;sup&gt;1&lt;/sup&gt;</td>
<td>WT</td>
<td>WT</td>
<td>−18.3</td>
<td>WT</td>
</tr>
<tr>
<td>pPreC-CP-Z&lt;sup&gt;(ε&lt;sup&gt;2&lt;/sup&gt;)&lt;/sup&gt;</td>
<td>ε&lt;sup&gt;2&lt;/sup&gt;</td>
<td>WT</td>
<td>WT</td>
<td>−13.0</td>
<td>WT</td>
</tr>
<tr>
<td>pPreC-CP-Z&lt;sup&gt;(ε&lt;sup&gt;3&lt;/sup&gt;)&lt;/sup&gt;</td>
<td>ε&lt;sup&gt;3&lt;/sup&gt;</td>
<td>WT</td>
<td>WT</td>
<td>−9.7</td>
<td>WT</td>
</tr>
<tr>
<td>pPreC&lt;sup&gt;m&lt;/sup&gt;-CP-Z&lt;sup&gt;(ε&lt;sup&gt;2&lt;/sup&gt;)&lt;/sup&gt;</td>
<td>preC&lt;sup&gt;m&lt;/sup&gt;-ε&lt;sup&gt;2&lt;/sup&gt;</td>
<td>AGU</td>
<td>WT</td>
<td>−13.0</td>
<td>WT</td>
</tr>
<tr>
<td>pPreC&lt;sup&gt;m&lt;/sup&gt;-CP-Z&lt;sup&gt;(ε&lt;sup&gt;3&lt;/sup&gt;)&lt;/sup&gt;</td>
<td>preC&lt;sup&gt;m&lt;/sup&gt;-ε&lt;sup&gt;3&lt;/sup&gt;</td>
<td>AGU</td>
<td>WT</td>
<td>−9.7</td>
<td>WT</td>
</tr>
<tr>
<td>pPreC&lt;sup&gt;ε&lt;sup&gt;i&lt;/sup&gt;&lt;/sup&gt;-CP-Z</td>
<td>sig&lt;sup&gt;−&lt;/sup&gt;</td>
<td>WT</td>
<td>M</td>
<td>−26.5</td>
<td>WT</td>
</tr>
</tbody>
</table>

Fig. 2. Structure of pPreC-CP-Z and its mutant derivatives. (a) Schematic diagram of plasmid pPreC-CP-Z. This construct contains the P ORF fused in-frame to the lacZ gene. Transcription is under the control of the SV40 early promoter and the translation start site for each protein is indicated by an arrow. The regions of signal sequence (sig) and encapsidation signal (ε) are indicated. (b) Characteristics of constructs derived from pPreC-CP-Z. The abbreviated name, free energy of ε structure (ΔG in kcal/mol), status of preC and C initiator codons and nature of signal sequence of each construct are indicated. WT, Wild-type; M, mutant.

Results

Translation initiation at the preC initiator codon is required to exert the negative effect on translation of the downstream genes

To study why the downstream C and P genes are translated unfavourably from the preC RNA, an in vivo expression system using transfected genes was employed in this study. To facilitate analysis, the lacZ gene was fused in-frame to the P ORF (Fig. 2a), so that the β-galactosidase level could reflect the efficiency of the P gene initiator. One difference between the preC RNA and the pregenomic RNA is the presence of an extra AUG, namely the preC AUG, found in the 5′ region of the preC RNA. When this AUG codon was inactivated, there was an 8-fold enhancement in β-galactosidase activity (Fig. 4a; compare the activity of WT and preC<sup>m</sup>). To rule out the possibility that the increase in activity merely reflects the leaky scanning mode of translation initiation, β-galactosidase activity was assayed when the C gene initiator was inactivated (Fig. 4a; C<sup>m</sup>); only a 2.5-fold increase in activity was observed. The result supports not only a ribosomal leaky scanning mode of P gene translation from preC RNA but it is also evidence of the importance of initiation at the preC AUG codon on translation of the downstream gene. However, it is possible that initiation at the 5′-proximal AUG codon mostly affects leaky scanning (Kozak, 1995). In this regard, we observed only a 1.5-fold increase in P gene translation in the same assay system when the 5′-proximal C gene AUG was inactivated in a construct where the C ORF preceded the P-Z fusion gene (Hwang & Su, 1998). The β-galactosidase values presented in this study were from four independent experiments and have been adjusted by the mRNA level of β-galactosidase to normalize the transfection efficiency (Fig. 4c).

Efficiency of translation initiation in the preC AUG sequence context is about 40% of that of the C gene AUG

If translation initiation from the preC AUG codon is responsible for suppression of P gene translation, and if the ribosomal leaky scanning mechanism is involved in P initiation from preC RNA, one would expect that the preC initiator codon is an effective initiation site where the majority of the initiation complex is formed. One of the important determinants for the strength of an initiator codon is its context. The optimal context proposed by Kozak (1989a) is (A/G) CCAUGG, in which nucleotides at positions −3 (A or G) and +4 (G) relative to the AUG codon (underlined) are highly conserved. The presence of nucleotide C instead of the
preferred G in position + 4 of the preC initiator codon (ACC\text{\textasciitilde}AU\textit{G})C suggests that it is a rather weak initiation site. In this regard, the nucleotides at \(-3\) and \(+4\) of the C gene AUG, GG\textit{CA\textasciitilde}UG\textit{G}, match that of the consensus sequence. To verify that the preC AUG codon is in a suboptimal context and to study how the context affects the strength of the preC and C initiator codons, the local context of the particular initiator codon was introduced into plasmid pP-Z which carries a P-Z fusion gene under the transcriptional control of the SV40 promoter. \(\beta\)-Galactosidase activity from transfected cells was then assayed. If the activity from the construct carrying the C gene AUG context was set as 100, the activity of that carrying the C AUG context appeared to be 39% of that of the C AUG codon, indicating an efficiency of translation initiation of the preC AUG context is about 40% of that of the C AUG codon, indicating an unfavourable primary sequence context.

The \(\epsilon\) RNA structure facilitates favourable recognition of the preC AUG codon

If the preC initiator codon is in a suboptimal context, there may be other elements to facilitate its recognition by the ribosomes. An RNA secondary structure, the \(\epsilon\) stem–loop structure, residing 30 nt downstream of the initiator codon of the preC ORF may be involved. One would predict that a reduction in the stability of this structure would result in an enhancement in translation efficiency at the C and P genes since the 40S ribosomal subunits would then bypass the preC AUG to continue the scan to initiate at the downstream AUGs. At the same time, expression of the precore protein should decrease to reflect a reduction of initiation at the preC AUG. To test this hypothesis, protein levels of the preC, C and P gene products from construct pPreC-CP-Z (Fig. 2a) and its derivatives were examined in a transient transfection system. The precore and C proteins were analysed by a radioimmunoprecipitation assay using anti-HBc antisera (Dako), whereas the activity of the P gene initiator codon was measured by the activity of the \(\text{lacZ}\) reporter gene. Derivatives of pPreC-CP-Z, namely \(\epsilon^{m1}\), \(\epsilon^{m2}\) and \(\epsilon^{m3}\), with increased \(\epsilon\) structure free energies (from \(-26.5\) to \(-18.3\), \(-13\) and \(-9.7\) kcal/mol, respectively) were constructed and tested. It is noteworthy that no alteration in the sequences of the C ORF or its initiation context was introduced into these mutants (Fig. 3).

If the \(\epsilon\) structure does play a role in recognition of the preC AUG, one would expect a reduction of the precore product when the free energy of the \(\epsilon\) structure was increased. The precore protein is a 25 kDa protein (p25) from which the signal sequence is cleaved to generate p22e, which is further processed to produce the secretory HBV \(\epsilon\) antigen of p17, whereas the C protein is a 21 kDa cytoplasmic protein (p21). The p25 and p22e precocore products and p21 C protein were analysed by immunoprecipitation using anti-HBc antisera from cells labelled with \(^{35}\text{S}\)methionine. As shown in Fig. 4(b), when transfected with the WT construct (pPreC-CP-Z), both p25 and p22e could be detected, whereas p21, the C gene product, was undetectable. When the preC AUG codon was inactivated, the p21 protein, instead of p25 and p22e, appeared (Fig. 4b), preC\textsuperscript{m1}, preC\textsuperscript{m2} and preC\textsuperscript{m3} were indeed the products initiated from the preC AUG codon. Data also confirm the previous conclusion that initiation at the preC AUG codon is responsible for the suppression of translation of the downstream genes from preC RNA. When transfected with a construct carrying a less stable \(\epsilon\) structure, i.e. \(\epsilon^{m1}\), the intensity of both the p25 and p22e bands decreased to about 65% of that of the WT level. A cross-reacting protein of 18.5 kDa (Fig. 4b; labelled \(\star\)) that was also visible in the control transfection with salmon sperm DNA (Fig. 4b, mock lane) was used to normalize the signal intensity between lanes of the WT and the \(\epsilon^{m1}\) samples. It is noteworthy that transfection efficiency, as measured by the intensity of the RNA capable of hybridizing to the \(\text{lacZ}\) promoter.
sequence (Fig. 4c), was the same between these samples. The result shows that a less stable \( \varepsilon \) structure is associated with reduced precore protein synthesis. When transfected with a construct carrying the \( \varepsilon^{m2} \) structure, the p21 protein that initiated from the C AUG codon could clearly be observed (Fig. 4b, \( \varepsilon^{m2} \)). This protein was not detected in cells transfected with the construct carrying \( \varepsilon^{m1} \) (Fig. 4b, \( \varepsilon^{m1} \)). Nevertheless, in an enzyme immunoassay of proteins from crude extract that reacted with anti-HBe, a 1.6- and 7.1-fold increase in activity in constructs carrying \( \varepsilon^{m1} \) and \( \varepsilon^{m2} \), respectively, was observed relative to the WT construct. It is possible that the radio-immunoprecipitation assay used here is not sensitive enough to detect a small amount of C protein expressed in the construct carrying \( \varepsilon^{m1} \). On the other hand, when carrying the \( \varepsilon^{m2} \) mutation (Fig. 4b, \( \varepsilon^{m2} \)), both p25 and p22e showed a slightly faster gel mobility, which is probably due to some amino acid alterations in the precore region. In addition, the ratio of p25 to p22e differed from that in the WT sample. This observation could not have resulted from inefficient signal peptide cleavage since no mutations had been introduced into the signal sequence during the generation of the \( \varepsilon^{m2} \) construct (Fig. 3). The stability of the precore protein was most likely affected when the amino acid sequence in the \( \varepsilon \) region was altered. It is thus not possible to compare the precore protein produced from \( \varepsilon^{m2} \) and the WT. However, no such concern appeared in the \( \varepsilon^{m1} \) mutation since none of the amino acid sequences were affected due to codon degeneracy (Fig. 3).

The \( \varepsilon \) structure is within the coding region and overlaps the signal sequence of the precore protein (Fig. 1). Thus, any in vivo study addressing effects of the \( \varepsilon \) RNA structure on the expression of precore protein may inevitably be complicated by the fact that protein stability or signal sequence recognition may be altered when modifying the \( \varepsilon \) RNA secondary structure. To circumvent this difficulty, the activity of the preC initiator codon was assayed indirectly by the expression level of the P gene. In particular, the \( \beta \)-galactosidase activity assay was sensitive and quantitative enough when using lacZ as a reporter to reflect the activity of the P gene initiator codon. In agreement with the prediction, an increase in \( \beta \)-galactosidase activity of 2.5- and 5-fold in constructs carrying the \( \varepsilon^{m1} \) and \( \varepsilon^{m2} \) mutations, respectively, was observed (Fig. 4a, \( \varepsilon^{m1} \) and \( \varepsilon^{m2} \)). The enhancing effect observed was probably maximal since \( \beta \)-galactosidase activity no longer increased even when the stability was further reduced to \(-9.7 \, \text{ kcal/mol in } \varepsilon^{m3} \). We did not test the construct when the free energy of the \( \varepsilon \) structure increased further since the \( \varepsilon \) sequence in such a construct would pair with sequences from other regions to form unpredictable structure, thus affecting translation. Similar levels of mRNA capable of hybridizing to the lacZ sequence were observed among cells transfected with various constructs indicating that levels of \( \beta \)-galactosidase activity were indeed a reflection of regulation at the translational level. Taken together, our data show a reduction in precore expression when preC RNA carries a less stable \( \varepsilon \) structure, whereas such an alteration results in an increase in the translation efficiency of the downstream C and P genes suggesting that the encapsidation signal facilitates favourable recognition of the

### Table

<table>
<thead>
<tr>
<th>Construct</th>
<th>( \Delta G ) (kcal/mol)</th>
<th>( \beta )-galactosidase (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-26.5</td>
<td>1</td>
</tr>
<tr>
<td>preCm</td>
<td>-26.5</td>
<td>8.16 ± 0.63</td>
</tr>
<tr>
<td>Cm</td>
<td>-20.1</td>
<td>2.51 ± 0.09</td>
</tr>
<tr>
<td>( \varepsilon^{m1} )</td>
<td>-18.3</td>
<td>2.52 ± 0.08</td>
</tr>
<tr>
<td>( \varepsilon^{m2} )</td>
<td>-13.0</td>
<td>5.02 ± 0.53</td>
</tr>
<tr>
<td>( \varepsilon^{m3} )</td>
<td>-9.7</td>
<td>4.88 ± 0.45</td>
</tr>
<tr>
<td>preCm-( \varepsilon^{m2} )</td>
<td>-13.0</td>
<td>12.35 ± 0.45</td>
</tr>
<tr>
<td>preCm-( \varepsilon^{m3} )</td>
<td>-9.7</td>
<td>12.30 ± 0.67</td>
</tr>
<tr>
<td>sig-</td>
<td>-26.5</td>
<td>1.81 ± 0.20</td>
</tr>
</tbody>
</table>

---

Fig. 4. Effects of the initiator codon, \( \varepsilon \) structure and signal sequence on preC RNA translation. (a) Effect on P translation. All constructs are derived from WT pPreC-CP-Z. Plasmid DNA of individual constructs was transfected to HuH-7 cells. Sixty hours after transfection, cells were removed for \( \beta \)-galactosidase assay and for total RNA isolation. The level of \( \beta \)-galactosidase activity in cells transfected with WT construct was set at 1. Data are adjusted by the mRNA level of \( \beta \)-galactosidase and represent the mean±SD from four independent experiments. (b) Effect on preC and C translation. At 28 h post-transfection, HuH-7 cells were starved of methionine for 1 h, then labelled with \( {^{[35]}S} \)methionine for 12.5 min before harvest. \( {^{[35]}S} \)Methionine-labelled proteins were examined by SDS–PAGE after immunoprecipitation with anti-HBe antisera (Dako). Positions of p25, p22e and p21 are indicated. A non-specific, cross-reacting protein of 5 kDa, also visible in the control transfection with salmon sperm DNA (mock lane), is marked (\( \oplus \)). The intensity of this protein band, quantified by PhosphorImager analysis, was used to normalize for the efficiency of \( {^{[35]}S} \)methionine incorporation. (c) A representative Northern blot using lacZ sequence as a probe. The plasmid transfected is indicated at the top of each lane. The RNA detected is about 4.4 kb in size.
preC initiator codon. The data also suggest that this facilitation is mainly through the secondary structure of the encapsidation signal instead of the primary sequence.

**The ε RNA structure has a negative effect on translation initiation from pregenomic RNA**

Besides the positive effect described above, the secondary structure of mRNA may exert a negative effect on translation when the structure resides upstream to an initiator codon (Kozak, 1989b). In this regard, the initiator codon of the C gene is found within the lower stem of the ε structure (Fig. 1). To examine the possibility of a negative effect and to eliminate the possible effect originating from the ε structure on the preC initiator codon, experiments were performed using construct preC<sup>m</sup> with an inactivated preC AUG. Thus, the mRNA of preC<sup>m</sup> has an identical number of AUG codons to pregenomic RNA. The β-galactosidase activity was found to increase by about 1.5-fold when carrying a less stable ε structure (Fig. 4a; compare the activities of preC<sup>m</sup> to that of preC<sup>m-ε<sup>m2</sup></sup> or preC<sup>m-ε<sup>m3</sup></sup>). Similarly, C gene translation was also affected, as shown by the increase in the amount of p21 in a construct carrying ε<sup>m2</sup> (Fig. 4b; compare the p21 intensity of preC<sup>m</sup> and preC<sup>m-ε<sup>m2</sup></sup>). Thus, the ε structure acts to reduce expression of both the C and P genes.

**Elongation arrest during targeting of the nascent peptide chains into ER can, to a small extent, reduce translation efficiency of the downstream genes**

When a bicistronic mRNA encodes a secretory protein in the first cistron, the scanning ribosomes initiating at the downstream cistron can be impeded when elongation arrest occurs during targeting of the nascent peptide chains bearing a signal sequence into the ER. When this occurs, one would expect a mild increase in P gene initiation from the preC RNA when the signal sequence is inactivated. Indeed, a 1.8-fold increase in β-galactosidase activity was observed (Fig. 4a; compare the activity between WT and sig<sup>−</sup>). This difference is statistically significant at the 5% level using a t-test. It is noted that the mutation affects the recognition of the signal sequence since only the p25 protein but not p22e was detected in cells transfected with a construct that had a mutated signal sequence as examined by immunoprecipitation using anti-HBc antisera (Fig. 4b, sig<sup>−</sup>). In addition, no secreted HBeAg was detected in the culture media of cells transfected with this construct by enzyme immunoassay.

**Discussion**

In this study, we have described evidence to indicate that preferential expression of the HBV P protein from pregenomic RNA instead of from preC RNA is as a result of several factors. Firstly, in the preC RNA of the HBV ayw subtype (Galibert et al., 1979), the P gene is preceded by six AUG codons instead of five in the pregenomic RNA. When the 40S ribosomal subunits attach to the 5′ end of the preC RNA and scan for a suitable AUG codon, an additional AUG codon needs to be accounted for before the 40S ribosomal subunits reach the P gene initiator codon. Therefore, fewer initiations at the P gene initiator are expected compared with that from the pregenomic RNA. Secondly, although the preC AUG codon is not in an optimal context for translation initiation, the downstream ε structure can facilitate its recognition so that the number of 40S ribosomal subunits available to reach the P gene initiator codon is greatly reduced. This appears to be the major mechanism responsible for inefficient expression of the P protein from preC RNA. Thirdly, during translation of the precore protein, translocation of the growing polypeptide chains into the endoplasmic reticulum is the first step in targeting proteins for secretion. The elongation arrest occurring during the targeting step can interfere or retard movement of the 40S ribosomal subunits. Therefore, initiation at the downstream AUG codon is reduced. The extent of reduction depends on the structural nature of the signal peptide since the sequence can affect the steps of SRP recognition and membrane-embedded translocation (Hatzawa et al., 1997). Thus, duration of elongation arrest can be a mechanism used by the cells to regulate expression of the secretory protein and also proteins that are expressed from some downstream cistrons. As a result of the combined action of the factors described, the P protein is translated inefficiently from preC RNA. Likewise, C protein expression from preC RNA is similarly affected. The data confirm studies showing that the HBV P gene is translated from the pregenomic RNA by a leaky scanning mechanism (Lin & Lo, 1992; Fouillot et al., 1993; Hwang & Su, 1998) and also suggest that translation initiations at the C and P genes from the preC mRNA, if any, may occur by a similar mechanism. However, this conclusion may not be applicable to other members of the hepadnaviruses, especially as the study by Chang et al. (1990) has shown that translation of the P gene in duck HBV depends on a mechanism other than scanning to allow internal entry of ribosomes to the region of the P initiator.

This study reveals that the encapsidation signal can facilitate recognition of the upstream preC initiator codon so that most scanning ribosomes would start at this site. Kozak (1990) has shown that this effect depends on the position of the downstream hairpin structure. The strongest facilitation is seen when the hairpin is separated from the preceding AUG codon by 14 nt, and the facilitated initiator recognition would diminish when this distance increases to 32 nt. The distance between the ε stem–loop structure and the preC AUG codon is an unfavourable 30 nt in length. It is possible that the stability of the hairpin structure may also be important. The free energy of the ε structure is −26.5 kcal/mol whereas that of the hairpin structure used in the study of Kozak (1990) is −19 kcal/mol, based on the energy rules of Tinoco et al. (1973). This value corresponds to −12 kcal/mol by the energy rules of Walter et al. (1994), which is the algorithm
employed to calculate the free energy of the ε structure. It is conceivable that a relatively stable ε structure compensates for the unfavourable distance effect. On the other hand, mRNA secondary structure can have a negative effect on translation when such secondary structure resides upstream to an initiator codon (Kozak, 1989b). In this respect, the initiator codon of the C gene is within the lower stem of the ε structure (Fig. 1).

This stem–loop structure could somehow reduce the translation efficiency of the C and P genes from the pregenomic RNA (Fig. 4). It is noteworthy that this inhibition occurs in the absence of the P protein. In the presence of the P protein, formation of a pre-assembly complex by binding of the P protein to the ε structure would result in a complete block of 40S ribosomal scanning (Nassal et al., 1990). However, such an effect may not occur when using preC RNA as a template since the 80S elongating ribosomes can disrupt duplex structure during translation of the precore protein (Kozak, 1989b; Nassal et al., 1990). Therefore, the negative effect of the ε structure on C and P protein translation from the preC RNA is mainly due to its facilitating effect on the precore translation rather than a blockage of the scanning ribosomes initiating at the C and P initiator codons. According to Nassal et al. (1990), inactivation of the ε structure by the 80S translating ribosomes is responsible for the preC RNA exclusion from encapsidation. Our finding that the ε structure enhances preC translation from the preC RNA while reducing C gene expression from the pregenomic RNA is consistent with the results of Ou et al. (1990) in that the in vitro translation efficiency of the preC RNA for precore protein synthesis was 7-fold higher than that of the core protein translating from pregenomic RNA. It is also noteworthy that translation initiation with the context of preC initiator codon is only 40% efficient in comparison to that of the C gene. Nevertheless, one would not expect these two proteins to differ significantly in amount since the preC RNA is only a minor species. In addition, the experiments carried out by Ou et al. (1990) were done presumably in the presence of the P protein.

The ε stem–loop structure at the 5’ end of the hepadnavirus pregenomic RNA has been assigned the roles of an RNA encapsidation signal (Hirsch et al., 1990; Junker-Niepmann et al., 1990; Pollack & Ganem, 1993) and as an origin of reverse transcription (Wang & Seeger, 1993; Tavis et al., 1994; Nassal & Rieger, 1996; Rieger & Nassal, 1996). Recently, Tavis et al. (1998) have suggested that the ε stem–loop structure is an essential cofactor for the maturation of the polymerase to an enzymatically active form. Results from the present study assign to the ε stem–loop structure a fourth function as a downstream secondary structure that facilitates recognition of the preC initiator codon. One can imagine that expression of the P protein from preC RNA is undesirable since this RNA is not packagable (Ender et al., 1987; Will et al., 1987; Junker-Niepmann et al., 1990; Nassal et al., 1990). It is also possible that instead of regulating expression of the P protein, a control via the ε stem–loop structure may ensure that enough precore gene products are expressed from a minor species of mRNA with a suboptimal initiation context.

Although the precore gene is not required for virus replication, it is important in establishing a persistent infection after neonatal infection (reviewed by Ou, 1997). In addition, the precore product p22e is shown to be important in the regulation of HBV replication (Lamberts et al., 1993; Scaglioni et al., 1997). It is of interest to ask why HBV does not accomplish a favourable precore expression instead with the preC initiator codon in an optimal context. Our preliminary data suggest that it may not be feasible to have an optimal AUG context and yet not to compensate for signal peptide processing (W.-L. Hwang & T.-S. Su, unpublished data). HBV may rather employ the ε structure to enhance the efficiency of its preC AUG codon. In fact, among members of hepadnavirus, none has a G at the +4 position of the preC AUG codon (Galibert et al., 1982; Seeger et al., 1984; Sprengel et al., 1985, 1988). However, for a virus with a compact genome organization, such an arrangement may occur fortuitously. Many lessons remain to be learned from this virus.

We thank K.-B. Choo and S. J. Lo for helpful discussion and critical reading of the manuscript. This work was supported in part by grants NSC-83-0419-B075-003-MH and NSC-84-2331-B075-007-MH from the National Science Council and by a grant from the Veterans General Hospital-Taipei, Taiwan, Republic of China.

References


