Translational modulation in hepatitis B virus preS-S open reading frame expression

A. Gallina,1 A. De Koning,1 F. Rossi,1 R. Calogero,2 R. Manservigi3 and G. Milanesi1*

1Istituto di Genetica Biochimica ed Evoluzionistica, Consiglio Nazionale delle Ricerche, Via Abbiategrasso 207, 1-27100 Pavia, 2SORIN Biomedica, Strada Crescentino, I-13040 Saluggia (VC) and 3Istituto di Microbiologia, Facoltà di Medicina, Università di Ferrara, Via Luigi Borsari 46, I-44140 Ferrara, Italy

A series of hepatitis B virus open reading frame (ORF) preS-S variants, including mutants in which the relative order of the in-frame start codons (AUG1, AUG2 and AUG3) and nearby sequences had been altered, was expressed both in vivo (in HepG2 hepatoblastoma cells) and in vitro (by T7 promoter-driven transcription followed by translation in a rabbit reticulocyte lysate). The ratio of the synthesis of the large, middle (M) and major (S) proteins or their chimeric counterparts was analysed to study the translational regulation of ORF expression. As expected on the basis of the ribosome scanning model, the AUG sequence context was found to be a prominent factor in determining the different translational behaviour of the two preS-S-specific mRNAs of 2-4 kb (predominantly translated from AUG1) and 2.1 kb (which includes AUG2 and/or AUG3 and can be translated from either). Results from both experimental systems suggested that initiation at internal AUGs in the 2-4 kb RNA is possible. In experiments in vitro, preS-S mutants bearing lesions in a region 5' to AUG2, which has been implicated in AUG2/AUG3 cis repression, showed no increase in the utilization of internal AUGs. In addition, the chimeric envelope polypeptides produced in transfected HepG2 cells in this study were informative with respect to preS-mediated endoplasmic retention: replacement of the preS2 N terminus with that from preS1 generated a chimeric M protein that was glycosylated within the putative preS1 retention sequence and was not secreted. Thus, the preS1 retention sequence most likely acts inside the lumen of endoplasmic reticulum and its function is insensitive to glycosylation. A similar element might be active at the N terminus of M protein.

Introduction

Hepatitis B virus (HBV) is an enveloped, partially double-stranded circular DNA virus which causes acute and chronic liver disease in humans (Tiollais et al., 1985; Ganem & Varmus, 1987; Blum et al., 1989). Its envelope (env) gene, the open reading frame (ORF) preS-S (Valenzuela et al., 1979; Galibert et al., 1979; Pasek et al., 1979; Ono et al., 1983; Kobayashi & Koike, 1984), encodes a set of related glycoproteins which are expressed by initiation at three in-frame start codons, AUG1, AUG2 and AUG3. Initiation at AUG3 generates the major (S) protein (p24 and its glycosylated form, gp21); initiation at AUG2 adds 55 amino acids (the so-called preS2 region) to the N terminus, generating the middle (M) protein (p30 and its mono- and diglycosylated forms, gp33 and gp36, respectively); initiation at AUG1 adds a further 108 to 119 amino acid extension (the preS1 region) to the N terminus of the M protein, generating the large (L) protein (p39 and its monoglycosylated form, gp41) (Heermann et al., 1984) (Fig. 1a). The rates of synthesis of L protein, and M and S proteins appear to be regulated separately at the transcriptional level. A hepato-specific promoter, SPI, distally located upstream from ATG1 (Chang & Ting, 1989; Chang et al., 1989; Raney et al., 1990), directs transcription of a 2-4 kb mRNA species which encodes the complete ORF and is believed to be translated predominantly into L protein (Cheng et al., 1986; Standring et al., 1986). A proximal, simian virus 40 (SV40)-like promoter, SPII, located within the preS1 coding region (Malpiece et al., 1983), directs transcription of 2-1 kb mRNAs starting near (either upstream or downstream of) AUG2 (Cataneo et al., 1983; Standring et al., 1984) (Fig. 1a). The 2-1 kb subspecies including AUG2 is believed to be translated into both M and S proteins (Cheng et al., 1986; Standring et al., 1986; McLachlan et al., 1987; Ou & Rutter, 1987) (Fig. 1a). The independent expression of L, and M and S proteins might reflect the functional specialization of the former. Indeed, L protein has been.

0001-0466 © 1992 SGM
found to be necessary, like S and in contrast to M protein, for HBV virion assembly and secretion (Bruss & Ganem, 1991). Moreover, it is retained within the rough endoplasmic reticulum (RER) and is able to trans-inhibit hamper virion secretion (Bruss & Ganem, 1991).

The constitutive secretion of the shortest envelope forms (Cheng et al., 1986; Chisari et al., 1986; Persing et al., 1986; Standring et al., 1986; McLachlan et al., 1987; Ou & Rutter, 1987). A short amino acid stretch at the N terminus of the preS1 region is involved in this retention process (Kuroki et al., 1989). This property is not directly related to L protein function in virogensis, because deletion of the L protein retention sequence does not hamper virion secretion (Bruss & Ganem, 1991).

The transcripational programme described is accompanied at the translational level by differential control of start codon usage. Both 2-4 kb mRNA and the 2-1 kb subspecies covering AUG2 offer ribosomes multiple start sites, but ribosomes apparently prefer AUG1 in the 2-4 kb mRNA, whereas they are capable of initiating at high frequency also at AUG3 in AUG2-containing 2-1 kb templates. This behaviour can be explained in terms of AUG sequence context, according to Kozak’s ‘modified scanning’ model (Kozak, 1988a, 1989a). The sequence around AUG1 [(A/G)GCAUGG] closely resembles the consensus sequence for an efficiently recognized start site [(A/G)CCAUGG, with a purine at position -3 and guanine at position +4 being of particular importance] (Kozak, 1987a). In contrast, AUG2 lies in a relatively ‘weak’ context (GCAUGC), which could allow for ‘leaky scanning’ in favour of AUG3 (which is within a strong context, A(A/U)CAUGG). On the other hand, a recent report (Masuda et al., 1990) suggested that a cis-acting sequence located upstream from AUG2 might prevent initiation at 2-4 kb mRNA internal AUG codons, in a manner independent of the context of AUG1.

The present report deals with these translational aspects of ORF preS-S regulation. A collection of deleted or rearranged ORF preS-S variants was compared with wild-type (wt) variants for the relative usage of initiation codons in two distinct experimental systems, cell transfection with an episomal vector and in vitro translation. As a by-product of this study, observation of the secretion and glycosylation phenotypes of chimeric envelope proteins expressed in transfected cells allowed us to gain some insight into the endoplasmic retention phenomenon.

Methods

Construction of plasmids. Vector pRPRSV has been described previously (Manservigi et al., 1990). It contains a defective BK virus genome which expresses large T antigen and enables the vector to replicate as a high copy number episome in human cells (Milanesi et al., 1984). In its expression cassette, a unique BanHI cloning site is preceded by the Rous sarcoma virus (RSV) long terminal repeat (LTR), which promotes transcription initiation (Stoltzfs, 1988), 44 nucleotides upstream from the cloning site. LTR-driven mRNAs thus bear a short heterologous 5' untranslated region (UTR), the first 34 residues of which correspond to nucleotides 1 to 34 of the RSV mRNA leader [hence excluding RSV 5' UTR micrinisceptors (Hensel et al., 1989)].

pRPRSV derivatives carrying various HBV genome fragments in the correct orientation for surface protein expression were constructed by standard cloning techniques (Sambrook et al., 1989) (Fig. 1b, c). pRPAUG3 bears the 2 kb BanHI (map position 30)-BglII (map position 1986) fragment of HBV ad2 DNA (map position zero is the unique EcoRI site in HBV DNA) including only the S protein coding region. For construction of pRPAUG2.3, the 2 kb EcoRI (map position 0/3221)-BglII (map position 1986) fragment from HBV ad2 DNA, the EcoRI extremity of which is encoded by five nucleotides downstream from AUG2, was ligated into the homologous sites of plasmid pKSV-10 (Pharmacia). The product was recloned at the EcoRI site, at which a head-to-head dimer of a synthetic BglII-BglRII adaptor reproducing the HBV sequence from 10 bases upstream of AUG2 to the EcoRI site (‘...AGGCTTGAGCCCAAGG’). The complete preS-S coding sequence was then extracted as a 2 kb BglII fragment and cloned into pRPRSV to generate the construct of interest. pRPAUG1.2.3 ayw and pRPAUG1.3 ayw, respectively, the 2-3 kb BglII fragment (map position 2839)-BglII (map position 1986) HBV ayw genomic segment and the 2-3 kb Akt (map position 2847)-BanHI (map position 1986) HBV ad2 genomic segment, spanning the entire preS-S coding region of either serotype.

Plasmids pRPAUG2.2.3, pRPAUG1.3 ayw and pRPAUG1.3 adw2 express rearranged versions of ORF preS-S. In RPAUG2.2.3, the 23 bp BglII-BanHI fragment from pKSV/preS-Slink, covering the AUG2 site and surrounding sequences, and the first 12 preS2 codons, replaces the 5' terminus of the preS1 coding region to the preS2-S portion in both HBV ayw and adw2 serotypes. The preS2-S portion in both HBV ayw and adw2 serotypes. In all the plasmids described above, the viral sequences downstream of the enhancer [enh (Shaal et al., 1985)], ORF X and the polyadenylation (pA) signal (Simonsen & Levinson, 1983). Preliminary experiments had shown that the inclusion of both enh and ORF X augmented LTR activity between two- and five-fold compared to that of constructs lacking either element or both, most probably by virtue of the trans-activating effect of the ORF X product on both enh and the LTR (Spandau & Lee, 1988).

In all the plasmids described above, the viral sequences downstream from ORF preS-S constantly covered the HBV enhancer [enh (Shaal et al., 1985)], ORF X and the polyadenylation (pA) signal (Simonsen & Levinson, 1983). Preliminary experiments had shown that the inclusion of both enh and ORF X augmented LTR activity between two- and five-fold compared to that of constructs lacking either element or both, most probably by virtue of the trans-activating effect of the ORF X product on both enh and the LTR (Spandau & Lee, 1988).

The same set of ORF preS-S variants was expressed in vitro by cloning them downstream from the T7 promoter into the plasmid BlueScribe (BS; Stratagene) polylinker (Fig. 1b, c). In this case, inserts spanned HBV sequences downstream from ORF preS-S to the SplI (map position 1265) site.

For construction of deletions within the preS1 coding region, 5 µg of plasmid BSAUG1.3 adw2 was cleaved within the preS2 coding region with BanHI and then treated with nuclease Bal 31 (Bethesda Research Laboratories) for between 30 s and 5 min at 25°C in the
appropriate buffer (20 mM-Tris-HCl pH 8.1, 12 mM-MgCl2, 12 mM-CaCl2, 200 mM-NaCl, 1 mM-Na2EDTA). After phenol extraction and ethanol precipitation, DNA was resuspended in 25 μl Klenow buffer (50 mM-Tris–HCl pH 7.2, 10 mM-MgSO4, 0.1 mM-DTT, 0.5 mM each dNTP, 5 μg/ml acetylated bovine serum albumin) and exposed to Klenow polymerase (2 units; Bethesda Research Laboratories) for 30 min at 22 °C, to blunt staggered DNA extremities. After another melting point agarose gel (Seaplaque; FMC) and ligated 'in gel' to the Linear plasmid containing unidirectionally eroded preS1 coding BamHI nuclease. The ligation products bore ORF preS-S mutants with deletions starting 10 nucleotides upstream from AUG2 and extending into the preS1 coding region for various distances. After amplification (Nunc) in MEM with 10% foetal calf serum (FCS). One day before transfection, they were plated at 2 x 10^5 cells/flask. Transfection was onto activated nylon membranes (GeneScreen plus; DuPont) and by the calcium phosphate method (Gorman, 1985), using 20 μg of the indicated plasmid, for 12 h. After a 20 min shock with 10% dimethylsulphoxide, fresh MEM and 10% FCS were added and cells were incubated for 48 h at 37 °C in a 5% CO2 atmosphere.

RNA analysis. Cellular RNA was harvested according to Chomczynski & Sacchi (1987). For Northern blot analysis, 20 μg RNA/transfectant was resolved through a 1.4% agarose/formaldehyde gel, blotted onto activated nylon membranes (GeneScreen plus; DuPont) and probed with the 1.4 kb BamHI fragment (map position 1403) of HBV adw2 DNA, labelled to >10^9 c.p.m.μg by the random priming method. After autoradiography, blots were stripped of HBV probe and rehybridized with a probe for human cytomegalovirus (hCMV) (Amersham) DNA. Autoradiographic bands from both hybridizations were quantified by densitometry; normalization for transfection efficiency was obtained by calculating the HBV/hCMV mRNA signal ratio for each lane.

Immune precipitations. Two days post-transfection, cell monolayers were washed four times with PBS and then incubated in 800 μl methionine-free MEM for 90 min at 37 °C, after which [35S]methionine (100 μCi, >1000 Ci/mmol; Amersham) was added. For pulse experiments, incubation was continued for another 3 h, then cells were harvested. For pulse-chase experiments, after 2 h of labelling, cells were washed again and incubated in non-radioactive MEM and 10% FCS for 18 h, then processed.

To prepare cytoplasmic extracts, cells were washed in PBS, resuspended in 800 μl lysis buffer (10 mM-Tris–HCl pH 8, 150 mM-NaCl, 1% Triton X-100, 10 mM-Na2EDTA), subjected to a double cycle of freezing (5 min in liquid nitrogen) and thawing (5 min at 37 °C) and then centrifuged in an Eppendorf minifuge to pellet cell debris. Cytoplasmic supernatants, in parallel with cleared culture media, were mixed with 10 μl of sheep anti-subviral HBV particles (HBSAg) IgG–Sepharose conjugate slurry (20 μg antibody per immune precipitation) and shaken overnight at 4 °C. Sepharose–antibody–surface protein immune complexes were washed four times in lysis buffer plus 0.2% Tween 20 and twice in 10 mM-Tris–HCl pH 8, 150 mM-NaCl, and resuspended in 50 mM-Tris–HCl pH 6.8. An equal volume of 2 x sample buffer (25 mM-Tris–HCl pH 6.8, 2% SDS, 6% v/v 2-mercaptoethanol, 0.02% bromphenol blue) was added and the suspension was heated at 95 °C for 5 min. After discarding Sepharose beads, samples were electrophoresed on 12% polyacrylamide gels containing 0.1% SDS. Fixed gels were soaked in En3Hance (New England Nuclear), dried and exposed for 24 to 48 h to X-ray film before development. Protein bands were quantified by densitometry.

In vitro transcription and translation of BS–ORF preS-S constructs. BlueScribe derivatives were linearized with HindIII, cutting into the plasmid polylinker just downstream from the HBV insert. Run-off transcription with phage T7 RNA polymerase (Promega) and translation of m'GpppG-capped synthetic mRNAs in rabbit reticulocyte lysates (Promega) were performed exactly as described by Kozak (1989b). Salt conditions in the translation mixture were adjusted to 2.2 mM-Mg2+, 45 mM-KCl, 90 mM-potassium acetate to ensure that the results would be comparable to in vivo data (Kozak, 1990a).

Results

The strategy adopted in this study is summarized in Fig. 1. We reasoned that to analyse the effects of sequence context and relative position on ORF preS-S AUG codon usage, a simple approach would be to change the relative order of the AUG codons. Inspection of the ORF preS-S nucleotide sequence revealed the presence of compatible restriction sites, positioned equally with respect to the reading frame, downstream from both AUG1 and AUG2 (Fig. 1b). These were exploited to construct recombinant variants of ORF preS-S (see Methods). In variant pRPAUG2.2.3 (Fig. 1c), a segment starting upstream of AUG2 to codon 12 of the preS2 coding region (thus including all of the AUG2 codon and surrounding sequences) was fused in-phase to codon 27 of the preS1 coding region (adw2), substituting the (theoretically) ‘weak’ AUG2 for the ‘strong’ AUG1 codon at the beginning of ORF preS-S. In variants pRPAUG1.3 ayw and pRPAUG1.3 adw2 (Fig. 1c), on the other hand, deletion around AUG2 fused preS1 codons 19 or 26, respectively, in phase to preS2 codon 13, putting AUG1 ad AUG3 in direct succession. Comparison of the two subtype versions is of interest owing to the presence of an 11 codon extension in subtype adw2 which generates a serial duplication of AUG1 and surrounding sequences. The internal duplication of AUG1 (herein referred to as AUG1-bis) in subtype adw2 lies in a ‘strong’ sequence context and has been shown to be utilized, at low efficiency, for initiation of 2.4 kb mRNA. Recombinant variants were expected to synthesize, from their distal start codons, chimeric counterparts of natural envelope proteins: a large chimera (c-L) bearing a duplication of preS2 N-terminal amino acids at its N terminus in place of the preS1 N terminus, translated from the pRPAUG2.2.3 variant; and proteins of the size of M protein (c-M, ayw or adw2 versions) substituting preS1 N-terminal amino acids for the N-terminal tip of M protein, translated from pRPAUG1.3 variants.
The translational behaviour of the re-assembled ORF preS-S variants was compared with that of variants pRPAUG3, pRPAUG2.3, pRPAUG1.2.3 ayw and pRPAUG1.2.3 adw2 (Fig. 1c) spanning, respectively, the S region (subtype adw2), the preS2-S region (subtype adw2) and the entire ORF preS-S of either subtype adw2 or ayw. All of them could express only normal envelope proteins.

**Translation of ORF preS-S variants transfected into HepG2 cells**

Variants were cloned into the expression cassette of vector pRPRSV (Manservigi et al., 1990), downstream from the RSV LTR (Fig. 1b, c). Cloned sequences spanned the HBV genome from ORF preS-S through the HBV pA site (Simonsen & Levinson, 1983). Thus, LTR-directed mRNAs were expected to share the following characteristics: (i) to span ORF preS-S variants completely, (ii) to bear a short heterologous 5' UTR upstream of HBV sequences and (iii) to terminate at HBV pA. Thus, the LTR was expected to drive the synthesis of mRNAs mimicking either the 2-1 kb mRNA (constructs pRPAUG2.3, pRPAUG1.3 ayw, pRPAUG1.3 adw2 and pRPAUG3) or 2.4 kb mRNA (constructs pRPAUG1.2.3 ayw, pRPAUG1.2.3 adw2 and pRPAUG2.3) in length, and regularly initiating upstream from the distal AUG, allowing direct evaluation of the relative use of AUG codons from the abundance of expressed proteins. A potential difficulty in our experiments was the presence in the latter group of constructs of HBV promoter SPII, embedded within the preS1 coding sequence (Fig. 1a, b, c), which could enable transcription of some natural 2-1 kb mRNA, contributing to the synthesis of M and S proteins and interfering with the analysis of translational events.

pRPRSV derivatives were transfected into HepG2 cells, a differentiated human hepatoblastoma cell line (Knowles et al., 1980) permissive for HBV replication (Sureau et al., 1986; Sells et al., 1988), offering a cellular milieu closely resembling that of hepatocytes. Total RNA was extracted 48 h later. Alternatively, cells were labelled with [35S]methionine and cell lysates prepared for immunoprecipitation.

The structure of plasmid-directed mRNAs was analysed in Northern blots. All of the transfecants expressed mRNAs whose size was consistent with initiation from the LTR and polyadenylation at the HBV pA site (Fig. 2). That the cap site was at the +1 nucleotide of the RSV mRNA leader was verified by RNase protection experiments (data not shown). In addition, the three constructs harbouring SPII produced natural 2-1 kb mRNA (Fig. 2, lanes 1 to 3). It was noteworthy that with these constructs both the overall
Hepatitis B virus surface gene translation

Fig. 2. (a) Transcription in vivo of ORF preS-S mRNAs. Total RNA was extracted from HepG2 cells transfected with different plasmids 48 h post-transfection and subjected to Northern analysis using an S RNA-specific probe. To the left, the approximate size (in kbp) of transcripts is shown; 2.4 kb mRNA bands detected in cells transfected with pRPAUG1.2.3 adw2 (lane 3), pRPAUG1.2.3 ayw (lane 2) or pRPAUG2.2.3 (lane 1) correspond to LTR-directed transcripts, as does the 2.1 kb transcript of pRPAUG2.3 (lane 4). In contrast, minor 2.1 kb transcripts in lanes 1 to 3 represent HBV SPII-directed mRNAs. To the right, the position of ribosomal RNAs is shown. Reprobing for hDHFR (b) enabled comparison of the amount of ORF preS-S-specific mRNAs between samples.

level of synthesis and the ratio of recombinant 2.4 kb mRNA to natural 2.1 kb mRNA (approximately 5:1) were very similar. Not unexpectedly, the ratio of 2.1 kb mRNA subspecies initiating on either side of AUG2 was also constant (two upstream : one downstream, as determined by RNase protection). Thus, constructs pRPAUG1.2.3 ayw, pRPAUG1.2.3 adw2 and pRPAUG2.2.3 were equivalent transcriptionally.

Cytoplasmic extracts and culture supernatants from pulse-labelled cells were immunoprecipitated with antisera to HBsAg. A series of polypeptide bands of the sizes expected for unglycosylated and glycosylated forms of L, M and S proteins or their chimeric counterparts was observed (Fig. 3) and quantified by densitometry (Table 1).

Construct pRPAUG3 synthesized only S polypeptides, p24 and its monoglycosylated form gp27. Both

Table 1. Expression ratio* of HBV surface protein variants in transfected HepG2 cells

<table>
<thead>
<tr>
<th>Construct</th>
<th>L</th>
<th>M</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRPAUG3</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>pRPAUG2.3</td>
<td>-</td>
<td>54.5</td>
<td>45.5</td>
</tr>
<tr>
<td>pRPAUG1.2.3 adw2</td>
<td>65.5†</td>
<td>24.1</td>
<td>10.4‡</td>
</tr>
<tr>
<td>pRPAUG1.2.3 ayw</td>
<td>60.1</td>
<td>28.4</td>
<td>11.5</td>
</tr>
<tr>
<td>pRPAUG2.2.3</td>
<td>37.3‡</td>
<td>49.4</td>
<td>13.3</td>
</tr>
<tr>
<td>pRPAUG1.3 adw2</td>
<td>-</td>
<td>91.2‡</td>
<td>8.8</td>
</tr>
<tr>
<td>pRPAUG1.3 ayw</td>
<td>-</td>
<td>88-0‡</td>
<td>12.0</td>
</tr>
</tbody>
</table>

* Ratios (percent) were calculated from densitometric profiles by summing peak areas of each of the S, M and L proteins or their chimeric counterparts (non-glycosylated plus glycosylated forms) and dividing the result by the cumulative lane area.

† This value is the total of initiation events at AUG1 and AUG1-bis.
‡ Chimeric polypeptide.

were abundantly secreted into the culture medium during labelling. (Fig. 3a, b).

pRPAUG2.3 synthesized equivalent amounts of S (p24 and gp27) and M proteins (p30, and its mono-
diglycosylated forms, gp33 and gp36, the latter bearing the second N-linked oligosaccharide at asparagine 4 in the preS2 region) (Fig. 3a). The envelope forms were again released into the medium, but at an appreciably lower rate; only glycosylated forms of M protein (which showed the characteristic 34K to 36K smeared band resulting from Golgi oligosaccharide trimming) were secreted (Fig. 3b).

Chimeric 2.1 kb mRNAs transcribed from pRPAUG1.3 ayw and pRPAUG1.3 adw2 were translated predominately as c-M polypeptides, with minor amounts of p24 and gp27 being synthesized (Fig. 3a). All envelope forms were absent from the medium (Fig. 3b). Of note, in this case, was the presence in both c-M subtype variants of a diglycosylated form. [This interpretation of the slowest migrating band was supported by the observation that it shifted with the intermediate band to the fastest migrating c-M band after endoglycosidase F or H treatment. In experiments to be reported elsewhere (A. Gallina et al., unpublished results) we have now demonstrated by site-directed mutagenesis that the actual site of the second glycosylation is at asparagine 4 in the ayw subtype, or 15 in the adw2 subtype, within the recombinant preS1 N terminus of c-M.] Furthermore, adw2 c-M, besides the non-, and mono- and diglycosylated forms translated from AUG1 (expectedly with a slightly lower electrophoretic mobility than ayw forms), showed less abundant polypeptides initiated at AUG1-bis (comigrating with ayw forms).

Transcripts from pRPAUG1.2.3 ayw, pRPAUG1.2.3 adw2 and pRPAUG2.2.3 expressed L, M and S polypeptides (Fig. 3a). No secretion was observed in any case (Fig. 3b). S protein forms were always the least abundant. L protein (p39 and gp41, glycosylated at the only S protein site) prevailed over M protein in transfecteds expressing wt ORF preS-S; utilization of AUG1-bis in the adw2 variant was not apparent. In contrast, in pRPAUG2.2.3 transfecteds M protein was over-expressed with respect to chimeric L protein (chimeric p39 and gp41). Interestingly, the preS2 N terminus, transplanted ahead of preS1 sequences, apparently failed to be glycosylated.

Control pulse–chase experiments (not shown) demonstrated that all envelope proteins, including the chimeric forms, were stable during labelling and after, so that distinct half-lives cannot be used to explain differences in their abundance. Thus, such differences should be attributable to translational effects. Our results suggest, in particular, that the distal AUG sequence context is always a determinant in deciding the synthesis ratio (Table 1). LTR-directed 2.1 kb transcripts, including AUG2 and surrounding sequences, and mimicking natural 2.1 kb mRNA, were translated as both M and S proteins, as previously reported. This was attributable to the 'weakness' of AUG2, as its substitution with the 'strong' AUG1 decreased initiation at AUG3.

The context of AUG1 determined synthesis of L protein from 2.4 kb templates; its substitution with AUG2 and surrounding sequences derepressed the use of proximal AUGs (Table 1). This interpretation is not affected by overlapping transcription from an internal HBV promoter, because, owing to the homogeneous activity of LTR and SPII in all constructs, no variation can be attributed to transcriptional factors.

Table 2. Expression ratio* of HBV surface protein variants expressed in vitro

<table>
<thead>
<tr>
<th>Construct</th>
<th>Surface variant synthesized (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
</tr>
<tr>
<td>pRPAUG3</td>
<td></td>
</tr>
<tr>
<td>pRPAUG2.3</td>
<td></td>
</tr>
<tr>
<td>pRPAUG1.2.3 adw2</td>
<td>62±4</td>
</tr>
<tr>
<td>pRPAUG1.2.3 ayw</td>
<td>70±3</td>
</tr>
<tr>
<td>pRPAUG2.2.3</td>
<td>36±2±2</td>
</tr>
<tr>
<td>pRPAUG1.3 adw2</td>
<td></td>
</tr>
<tr>
<td>pRPAUG1.3 ayw</td>
<td></td>
</tr>
</tbody>
</table>

* Ratios (percent) were calculated from densitometric profiles by summing peak areas of each of the S, M and L proteins or their chimeric counterparts and dividing the result by the cumulative lane area.
†,‡ See footnotes to Table 1.

Fig. 4. Synthesis in vitro of surface protein variants. (a) T7 transcripts produced from BSAUG3, -2.3, -1.2.3 ayw, -1.2.3 adw2, -2.2.3, -1.3 ayw or -1.3 adw2 (lanes 1 to 7) were translated in vitro in rabbit reticulocyte lysates. The [35S]methionine-labelled polypeptides generated were analysed by SDS–PAGE and fluorography. (b) Effect of in-frame deletions upstream from AUG2 on the translation pattern of ORF preS-S in vitro. Plasmids BSpreS1A18 (lane 3) and BSpreS1Δ8 (lane 2), bearing an ORF preS-S insert with deletions of nucleotides –186/–11 and –83/–11, respectively, upstream from AUG2 were processed as described for (a); for comparison, plasmid BSAUG1.2.3 adw2 (lane 1), bearing the wt ORF, was analysed in parallel. M and S proteins bands, clearly observed on the original autoradiograph, are only faintly reproduced on this photograph. In (a) and (b) the position of wt surface proteins is shown to the left.
effects could be avoided. After cloning downstream from a phage T7 promoter, ORF preS-S variants (Fig. 1b, c) were transcribed in vitro and translated in a rabbit reticulocyte lysate (Fig. 4a), under buffer conditions known to reflect the in vivo initiation fidelity and context requirements (Kozak, 1989b, 1990a). The data, shown in Table 2, were in agreement with in vivo results. On templates mimicking 2.1 kb mRNA, AUG2 allowed effects could be avoided. After cloning downstream from a phage T7 promoter, ORF preS-S variants (Fig. 1b, c) were transcribed in vitro and translated in a rabbit reticulocyte lysate (Fig. 4a), under buffer conditions known to reflect the in vivo initiation fidelity and context requirements (Kozak, 1989b, 1990a). The data, shown in Table 2, were in agreement with in vivo results. On templates mimicking 2.1 kb mRNA, AUG2 allowed for 'leaky scanning' in favour of AUG3 in little more than half of the scanning events. By comparison, AUG1 ayw and AUG1 + AUG1-bis captured more than 80% of initiation events.

Similarly, on transcripts mimicking the 2.4 kb mRNA AUG arrangement, both AUG1 and AUG1 + AUG1-bis captured most, but not all, initiation events (Table 2), which unambiguously proves, in this system, the imperfect dominance of the distal start codon and any possible cis-repressing element over internal AUGs.

The possibility that cis-regulatory sequences lie upstream of AUG2 was addressed directly by analysing the translational performance of two ORF preS-S mutants bearing in-phase deletions within the preS1 coding sequence in vitro. Both Δ−83/−11 and Δ−186/−11 (numbering from AUG2) mutants (plasmids BSpreS1A8 and BSpreSA18, respectively), deleting a 3′ portion or all, respectively, of the putative −102/−38 inhibitory sequence identified by Masuda et al. (1990), did not show an appreciable increase in internal AUG usage with respect to wt virus (Fig. 4b).

Discussion

The mechanism whereby eukaryotic ribosomes choose a start codon on a capped mRNA molecule is currently accounted for by the 'modified scanning' model (Kozak, 1988a, 1989a), in which 40S ribosomal subunits bind to the 5′ end of the mRNA and scan linearly toward the 3′ end until they encounter an AUG within a sequence context acceptable for initiation (Kozak, 1987a). Aspects further influencing the use of a given AUG codon include its distance from the 5′ cap (Johanson et al., 1984; Kozak, 1986, 1987b, 1988b; Sedman et al., 1990), mRNA secondary structure (Pavlakis et al., 1980; Pelletier & Sonenberg, 1985; Lawson et al., 1986; Kozak, 1988b, 1990b; Fu et al., 1991) and the presence of cis-acting elements, which can increase or decrease translation of downstream regions in cooperation with trans-acting proteins (Hentzel et al., 1987; Aziz & Munro, 1987; Parkin et al., 1988; Rosen, 1991). The combination of these factors determines which of the potential cistrons on a translational template are actually expressed, and to what extent. Most cellular and viral mRNAs are translated from the AUG, in a good context, closest to the 5′ end. In a minority of transcripts both 5′ and internal AUG(s) are exploited, with more than one protein being expressed (Kozak, 1988a).

These alternatives coexist for HBV env gene expression strategy: 2.4 kb mRNA is translated mainly from the distal AUG (AUG1), whereas in 2.1 kb mRNA, both the first and second AUG (AUG2 and AUG3, respectively) are utilized. The results presented here indicate that, in both cases, AUG sequence context is a determining factor for such behaviour. AUG1 context, in concordance with the consensus for a strong AUG, reduced initiation at downstream start codons not only when at its natural position, but also when transplanted directly upstream from AUG3. By contrast, the AUG2 'weak' sequence context, when substituted for AUG1 ahead of ORF preS-S, derepressed the use of downstream AUGs (Fig. 3 and 4; Tables 1 and 2). This effect of the context of the distal AUG on the overall expression ratio gives an experimental demonstration that the scanning process on 2.4 and 2.1 kb templates starts upstream from the first possible AUG, that is, at or near the 5′ end, as predicted by Kozak's general model. Thus, our data render unlikely the possibility of internal entry of ribosomes into the preS coding region for expression from internal AUGs. This alternative mechanism of ribosome binding, well established for translation of uncapped picornavirus mRNAs (Jackson et al., 1990; Sonenberg, 1991), most likely acts in the expression of HBV ORF pol (Jean-Jean et al., 1989) and of the homologous gene of duck hepatitis B virus (Chang et al., 1990).

Our results suggest, on the other hand, that the preclusion from using internal AUGs of the 2.4 kb mRNA was not as absolute in our experiments as previously reported for the Xenopus (Standring et al., 1986) and the vaccinia virus (Cheng et al., 1986) systems. Direct evidence came from in vitro experiments, in which internal initiation of transcripts spanning the entire ORF was observed, and approximately equal amounts of protein were translated from AUG2 and AUG3 (Fig. 4; Table 2). In HepG2 cells transfected with constructs pRPAUG1.2.3 adw2 and pRPAUG1.2.3 ayw, M protein was expressed to a relatively high extent, clearly prevailing over S protein (approximately 6 L : 3 M : 1 S) (Fig. 3; Table 1). These cells harboured both LTR-driven 2.4 kb mRNA, covering the entire ORF, and minor amounts of SPII-driven 2.1 kb mRNA, the larger (and more abundant) subspecies of which includes AUG2. Whether M protein is translated predominantly from natural transcripts (the translational efficiency of which, given the relative levels of transcription of 2.4 kb and 2.1 kb mRNA species, would then be postulated to be significantly higher), from LTR-driven transcripts (by internal initiation), or from both, one has to conclude
that, in this case, AUG2 is capable of intercepting much more than 50% of ribosome passages, contradicting previous observations (Cheng et al., 1986; Standing et al., 1986; McLachlan et al., 1987; Ou & Rutter, 1987) and our observations (this work) on recombinant 2.1 kb mRNA translation. Interestingly, in repeated transfections of 293 cells [a human embryonic kidney cell line expressing adenovirus E1A/E1B proteins (Graham et al., 1977)] with pRPAUG1.2.3-adw2 and pRPAUG1.2.3-ayw, even higher levels of M protein, close to those of L protein, were observed (A. Gallina, A. De Koning & G. Milanesi, unpublished results). Thus, a degree of modulation in the usage of initiation codons might exist in addition to the rigid effect of AUG sequence context, as revealed in cells differing in their tissue origin/differentiation state. Whether this plays a role in natural infection, in addition to transcriptional regulation, in adjusting the relative rate of envelope protein synthesis, can not be determined on the basis of our experiments. As a line of evidence, not only the L:S protein but also the M:S protein ratio, in the presence of ongoing viral replication, is markedly increased and is reflected in virion envelope composition (Tiollais et al., 1985).

An example of a defined cis-regulating sequence which might participate in translational modulation has recently been reported by Masuda et al. (1990). Analysing the expression of ORF preS-S deletion variants in a hamster cell line, these authors identified a cis-repressing element, mapping in the region −102 to −38 nucleotides upstream from AUG2, which inhibits initiation at AUG2 and AUG3. Within this region, a potential hairpin structure of moderate stability is predicted. Data presented here suggest that, both in HepG2 cells and in rabbit reticulocyte lysates, the effect of such a sequence is relatively unimportant compared to the context of AUG1 (Fig. 3). Moreover, two mutants bearing in-phase deletions within the preS1 region that damage or eliminate the putative cis-acting sequence were analysed by in vitro translation, and no significant increase in internal AUG utilization was observed in relation to the intact gene (Fig. 4b). A possible explanation of this discrepancy might be the absence from rabbit reticulocyte lysates of a factor which is essential in mediating the downstream repression effect in rodent cells. Alternatively, in hamster cells the 40S ribosomal subunits could be more sensitive to secondary structures in mRNA than in rabbit reticulocyte lysates. This issue deserves further investigation.

Although primarily intended for the study of translational modulation, our experiments gave additional information about the mechanisms involved in HBV envelope protein RER retention. The ability of L protein to retain itself and the shorter envelope forms within a RER or early Golgi compartment, verified here in HepG2 cells (Fig. 3), has been observed previously in various experimental systems, including transfected mammalian cells (Persing et al., 1986; McLachlan et al., 1987; Ou & Rutter, 1987), Xenopus oocytes (Persing et al., 1986; Standing et al., 1986) and transgenic mice (Chisari et al., 1986). A recent report from Kuroki et al. (1989) demonstrated that a specific sequence [endoplasmic retention sequence (ERS)], mapping between amino acids 5 and 19 of the L protein of subtype adw2 (corresponding to amino acids 1 to 8 of subtypes ayw and adw) is necessary and sufficient for RER retention. The non-secretory characteristics of the two c-M mutants described in the present work (Fig. 3) agree with these findings. These chimeric proteins, bearing amino acids 1 to 19 or 1 to 26 of the N terminus of L protein (subtype ayw or adw2, respectively) fused to amino acid 13 of the preS2 region, include the ERS. It is noteworthy that the N terminus of c-M proteins is N-glycosylated in both subtype variants at the asparagine of a target Asn-Leu-Ser sequence within the putative ERS (A. Gallina et al., unpublished results). This potential glycosylation site is normally unused in L protein (see Fig. 1a), but constitutes an absolutely conserved sequence among all serotypes. As the attachment of an N-linked oligosaccharide follows translocation into the intracisternal space of the RER, our data suggest that (i) the ERS of c-M proteins is transported into the lumen of the RER, where it mediates retention, and (ii) glycosylation at a position located centrally within ERS does not interfere with its function. These conclusions point to an interaction of ERS with a soluble or anchored component inside the RER, and counter the hypothesis that it can act as a cytoplasmic anchor.

A puzzling aspect of our experiments was that the complex of surface proteins, including a chimeric L protein form devoid of the N-terminal ERS, synthesized by pAFAUG2.2.3 transfectants also was retained within cells (Fig. 3). The idea that the preS2 N terminus replacing the natural preS1 N terminus introduces some gross misfolding in c-L impairing transportability, is contradicted by the absence of an absolute preclusion of c-L oligomerization and secretion, because high-level co-expression of S protein by cotransfection of pRPAUG2.2.3 and pRPAUG3 allowed secretion of all envelope forms (data not shown). An attractive alternative is that the preS2 domain, similarly to preS1, harbours within amino acids 1 to 12 an ERS, possibly also active when transplanted to the N terminus of L protein. We note, in fact, that the most abundantly expressed envelope form in these transfectants was the M protein, with a ninefold molar excess of preS2 N termini compared to S protein. We are currently testing our hypothesis by constructing site-specific and deletion mutants of the preS2 coding region.
This work was supported in part by Progetto Finalizzato ‘Biotecnologie e Biostrumentazione’ and by Progetto Finalizzato ‘Ingenieria Genetica’ of Consiglio Nazionale delle Ricerche. A.G. was the recipient of a fellowship from Fondazione Camillo Golgi, Brescia. A.D.K. was the recipient of a fellowship from the Commission of European Communities.

References


(Received 1 July 1991; Accepted 2 September 1991)