Hepadnavirus evolution and molecular strategy of adaptation in a new host

Claudio Argentini, Valentina La Sorsa, Roberto Bruni, Emilio D’Ugo, Roberto Giuseppetti and Maria Rapicetta

Laboratory of Virology, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

In order to elucidate the mechanisms of hepadnavirus evolution in vivo and to trace the fate of known quasispecies in a single animal during the acute phase of infection, a woodchuck (Marmota monax) was infected with the hepadnavirus woodchuck hepatitis B virus (WHV). Woodchuck 197 (W197) was injected intravenously with pooled sera collected from a chronic carrier that had been infected originally with a molecular clone of known genome sequence (WHV7). Viral genome variants from both the inoculum and the follow-up sera from W197 were characterized for the presence of quasispecies related to the WHV7 sequence. Interestingly, WHV7-related genomes were predominant 6 weeks post-infection (p.i.), whereas a highly heterogeneous virus population was present in the first viraemic serum (4 weeks p.i.). Using WHV7 as the prototype, the variability of the Pol and PreS/S regions in the first 11 weeks p.i. has been calculated. The sequence population in serum collected 6 weeks p.i. was highly homogeneous, with a mean variability of 0–36% in the region analysed. Mean variability values ranging from 0.82% to 1.61% were found in quasispecies from the other sera. The presence of possible selective pressure was analysed by means of the non-synonymous versus synonymous variation ratio (dn/ds). We found that the dn/ds values were stable for the S ORF (ranging from 2.6 to 3.0), whereas a wider range was observed for the Pol ORF (from 1.4 to 3.0). Furthermore, from the analysis of the variability of the codon positions for the two overlapping ORFs it was found that, in most cases, non-synonymous mutations at position 1 of the Pol ORF (position 3 of the S ORF) corresponded to synonymous variation in the S (Pol) ORF, indicating independent evolution of the encoded proteins.

Introduction

Virus populations exist in vivo as a heterogeneous mixture of variants, termed quasispecies (Domingo & Holland, 1994). The host represents an ‘environment’ and selective forces such as immune surveillance, virus cell-tropism and therapies may affect virus diversity. Studies on virus quasispecies and their evolution may help to clarify several pathogenetic aspects of infection, such as adaptation to a new host, compartmentalization in different tissues and organs, pathologic features, prognosis of infection and resistance to antivirals (Hughes et al., 1997; Julias et al., 1997; Lane et al., 1995; McDonald et al., 1997; Meyer & Southern, 1997; Wong et al., 1997). Recently, it has been shown in two different virus models that quasispecies composition is a major determinant in the establishment of persistent infection and disease progression (Meyer & Southern, 1997; Rowe et al., 1997).

Nucleotide mutations in the viral genome are introduced by the error-prone activity of the enzymes involved in virus replication, such as RNA polymerases and reverse transcriptases. Variations may, under particular conditions, overcome the ‘error threshold’, which defines the minimum level of survival for a virus (Eigen & Biebricher, 1988); Lee et al., 1997). Furthermore, a high quasispecies complexity leads to unpredictable virus evolution (Martinez et al., 1997).

Woodchuck hepatitis B virus (WHV) is a member of the family Hepadnaviridae. Members of this family infect humans, chimpanzees, gibbons, woolly monkeys, ground squirrels,
Fig. 1. Phylogenetic trees of WHV nucleotide sequences obtained from the inoculum and from W197 follow-up sera. The trees are based on the neighbour-joining method. The panels refer to the different sera studied. Numbers following underscores specify the recombinant clones containing cDNAs that represent WHV variants. WHV7 is the prototype sequence. Bar, 0.02 substitutions per nucleotide position.

arctic ground squirrels, herons and ducks (Norder et al., 1996; Seeger et al., 1984; Schodel et al., 1991; Sprengel et al., 1988; Testut et al., 1996). The partially double-stranded DNA genomes exhibit many common features, including an extremely compact organization. The polymerase, Core, X and PreS/S polypeptides are encoded by overlapping reading frames, thus increasing the utilization of the small genome (Schodel et al., 1991; Sprengel et al., 1988).

Hepatitis B virus (HBV) and WHV are made up of heterogeneous virus populations (Li et al., 1996; Pult et al., 1997). The possible association of variants (PreCore, Core, S and X mutants) with particular clinical manifestations has not been definitively established. The complexity of the virus population has recently been reported to influence the fate of infection (Sterneck et al., 1996). However, few data are available on the distribution of hepadnavirus mutants in quasispecies and on the dynamics of evolution in HBV-infected patients (Sterneck et al., 1997, 1998; Gunther et al., 1998). In the present study, an experimental infection model has been designed to elucidate the mechanisms of hepadnavirus evolution in vivo. Woodchuck 197 (W197) was infected with a virus strain, WHV7, that has been characterized previously (Cohen et al., 1988; Miller et al., 1990).

All variants found in the inoculum and in the W197 follow-up sera were highly related to the prototype virus, WHV7, and the productive infection led to virus persistence. The confirmation of an 'ancestor' genetic signature allowed multifacial analysis of the virus quasispecies in vivo. Overlapping frames
could mutate independently under the selective pressures of the newly infected host, by means of the peculiar genome organization.

Methods

■ Study design. Pooled ‘chronic’ sera (Miller et al., 1990) from a chronically infected woodchuck (W7) that has been characterized previously (Cohen et al., 1986; Miller et al., 1990) (kindly provided by A. Ponzetto, Osp. Molinette, Turin, Italy) was used as inoculum to infect woodchuck W197. The W7 infection originated from a molecular transfected (Cohen et al., 1988). Sera were collected from week 4 to week 11 post-infection (p.i.) and 41 clones, 7–10 from each serum sample, were analysed for WHV DNA. The 7–10 clones were considered to represent the molecular species most abundant in the quasispecies. The amplified region (see Fig. 1) was 1253 nucleotides long, more than one-third of the entire WHV genome.

■ Woodchuck W197 experimental infection. W197, an adult wild woodchuck trapped in the Middle Atlantic area of North America and purchased from International Animal Exchange (Fernandale, MI, USA), was maintained in the animal facility of the Istituto Superiore di Sanità, Rome, and received human care in compliance with the guidelines of the Italian Ministry of Health. Experimental infection of W197 (WHV+) was performed by intravenous inoculation with diluted serum to give a final virus load of 0.4 × 10^6 infectious units. Foetal calf serum, previously inactivated by incubation at 56 °C for 30 min, was used for dilution. Blood samples were drawn weekly during the follow-up from week 4 to week 11 p.i. and sera were stored at −20 °C. The animal died of pulmonary infection 10 months after inoculation and was viraemic (WHV DNA−) until its death.

■ Detection of WHV DNA. The development of WHV infection was monitored by standard procedures (Faune et al., 1995, 1997). Briefly, WHV DNA was extracted from a serum sample (20 µl) by mixing with 30 µl 1 M NaOH and incubating at 65 °C for 10 min. After incubation, the mixture was chilled on ice for 10 min and 50 µl of a chilled solution containing 1:2 M Tris–HCl (pH 7.5) and 3 M NaCl was added. The samples were spotted onto nitrocellulose. Hybridization was performed with a ^32P-labelled, full-length WHV genome probe (a gift from J. Summers, Chase Cancer Center, Philadelphia, USA). A random priming DNA labelling kit (Boehringer Mannheim) was used according to the manufacturer’s instructions. WHV DNA was quantified by comparison with the signals obtained from serial dilutions of a cloned WHV DNA spotted in parallel to the samples. Autoradiograph images were scanned with a Pharmacia LKB-Ultroscan XL. Serum DNA levels ranged from 5 to 200 ng/ml. Additionally, WHV DNA was detected directly from serum by PCR using the following reaction mixture: 1 × PCR buffer II (Perkin Elmer), 2.5 mM MgCl₂, 0.5 mM primers, 0.2 mM dNTPs and 2.5 U Taq polymerase (Perkin Elmer). The selected region (nucleotides 2490–419), which encodes a portion of P, PreS and the amino-terminal part of the S protein, was amplified from 0.25 µl serum, quantified as 10^7–10^9 copies depending on the sample, with primers 02490 (5’ CTTCTAGGTCCCCAGAAGACGCACTCCC) and A419 (5’ CCCCTGGAAAAACTGAGAGAATGTCACACC) PCR was performed in a Thermal Cycler (Gene Amp PCR System 9600, Perkin Elmer) with a preliminary denaturation at 92 °C for 3 min followed by 35 cycles of three steps: denaturation at 94 °C for 15 s, annealing at 65 °C for 15 s and polymerization at 72 °C for 90 s.

■ Cloning of the WHV DNA fragment. PCR products, length 1253 bp, were cloned prior to DNA sequencing. The amplified fragment was cloned into vector PCR II by a ligase reaction by using the Original TA cloning kit (version C, Invitrogen) according to the manufacturer’s instructions. The ligation products were transfected into Escherichia coli XL1 Blue by the heat-shock protocol (Sambrook et al., 1989). Clones bearing the insert were identified first by selection with ampicillin, IPTG and X-Gal and then by screening with the cracking procedure. Briefly, bacteria were resuspended in 50 µl lysis buffer (0.2 M NaOH, 0.5% SDS

Fig. 2. Schematic representation of the WHV genome region studied (1253 nucleotides) (a) and a nucleotide variability table (b). The scheme (a) shows the overlapping Core, S and PreS/S ORFs; the region has been divided into the five subregions corresponding to the overlapping ORFs. Percentages (b) represent mean nucleotide variability of intra-serum variants as compared with the WHV7 sequence. Total percentages were calculated for the entire region.
and 20% sucrose) and incubated at 70 °C for 5 min and then 4 μl 4 M KCl was added to each sample. After centrifugation, the supernatant was analysed for the presence and size of plasmid DNA by electrophoresis on an agarose gel (0-7%) (Sambruk et al., 1989). Purified DNA for sequencing was obtained by using the Qiagen purification kit according to the manufacturer’s instructions.

DNA sequencing and analysis. DNA was sequenced by using the Taq Dye-Primer and ABI Prism Dye-Terminator cycle sequencing kits (Perkin Elmer). Sequencing reactions were carried out on a Gene Amp PCR system 9600 (Perkin Elmer) and analysed on an Applied Biosystems 373A automated sequencer. Sequences were named as the serum number (In, 4w, 6w, 7w and 11w) separated by an underscore from the recombinant clone number assigned during the bacteria screening. The sequences of WHV variants in the inoculum and in W197 sera were analysed to evaluate the genetic relatedness. All sequences were aligned by using the program PILEUP from the GCG package and then edited by using the Genetic Data Environment (GDE) software (Smith et al., 1994).

Neighbour joining, as implemented in the phylogeny inference package (PHYLIP; Felsenstein, 1993), was used to construct the phylogenetic trees with a two-parameter distance model (modified Kimura model) with a ratio of transitions to transversions of 2:0. DNA sequence divergence was calculated by using the neighbour-joining method in PHYLIP. Nucleotide sequences were translated by using the GDE package. Amino acid variability for the polymerase polypeptide was calculated by splitting the region into two separate segments: amino acids 1–167 and 168–417. Segment 168–417 overlaps the PreS/S polypeptide (PreS1, 1–149; PreS2, 150–209; and S, 210–250). With regard to the analysis of intra-serum quasispecies complexity, the sequences with genetic distances lower than 0.2 were considered as predominant. The presence of mutations in each position within codons of the polymerase or PreS ORFs was calculated considering the complete overlapping region (nucleotides 2992–419). Multiple mutations (two or three) in a single ORFs was calculated considering the complete overlapping region.

The introduction of single base substitutions and frameshift errors by Taq DNA polymerase was considered to be negligible, based on the known error-rate of this enzyme (Eckert & Kunkel, 1990).

Terminology. Following the results of phylogenetic analysis, WHV7 was termed the prototype and the sequenced clones were defined as variants, whereas variants presenting mutations in the predicted amino acid sequence of the S region were considered mutants (M). Mutants with a stop codon in position 13, 17, 21 or 28 of the S polypeptide were designated as termination mutants (TM).

Results

Phylogenetic analysis of WHV7 and WHV197 quasispecies and genome variability

The sequences of WHV variants (nucleotides 2490–419, according to the numeration of the WHV7 sequence as reported in the EMBL database with accession no. M18752) from the inoculum and from the W197 follow-up sera were analysed to evaluate genetic relatedness. Fig. 1 shows the phylogenetic trees obtained by the neighbour-joining method. The sequence of variant In_53 was identical to the WHV7 sequence. Identities between the WHV7 sequence and the sequences of the inoculum variants ranged from 97-97% to 100% (data not shown). The sequence of WHV7 in the database (Cohen et al., 1988; Miller et al., 1990) was used as the prototype in the following analysis, performed for all amplified clones (from inoculum and from sera collected during the first 11 weeks of infection of W197: 4w, 6w, 7w and 11w).

Identities, calculated against the WHV7 sequence, ranged from 95-53% to 100%. These findings confirm both the validity of the analysis based on the relatedness of all clones to the prototype WHV7 and the common origin of the W7 and W197 infections, as shown by the central clustering of WHV7 in all the phylogenetic trees.

The region of the WHV genome studied is shown in Fig. 2(a). Fig. 2(b) reports the mean variability for each fragment in five subregions, selected on the basis of their encoded ORFs. The first fragment, Core/Pol (nucleotides 1–98), encodes both the carboxy terminus of the core protein and the amino terminus of the polymerase; the second fragment, Pol (nucleotides 147–417), encodes the polymerase without other overlapping ORFs; the third fragment, PreS1/Pol (nucleotides 503–949), contains the polymerase and the PreS1 ORFs; the fourth, PreS2/Pol (nucleotides 950–1129), encodes polymerase and PreS2; and the fifth, S/Pol (nucleotides 1130–1253), encodes the polymerase and S regions. The mean value of variability for the inoculum clones, when compared with WHV7, was 0.82%. The highest values were found in the
Hepadnavirus evolution in a new host

Fig. 4. Sequential analysis of amino acid variability in Pol and PreS/S ORFs. The partial Pol sequence was divided into two segments corresponding to (a) the non-overlapping region (segment 1–167) and (b) the region overlapping the PreS/S genes (segment 168–417).

PreS1/Pol and PreS2/Pol fragments. Similar results were observed in sera 6w and 11w for all regions analysed, whereas a higher variability was present in serum 4w for PreS1/Pol and Pol/S. Notably, the Core/Pol region displayed few nucleotide variations, indicating that the rate of introduction of new mutations by Taq polymerase was low. The variants in serum 4w were the most divergent compared with WHV7, whereas those in serum 6w were highly related to the prototype.

Fig. 3 shows an amino acid sequence alignment of the amino terminus of the S protein, derived from the translation of nucleotide sequences into amino acid sequences by computer-assisted prediction. Three ‘hot-spot’ residues, 17, 21 and 28, were identified, with changes in 19 variants out of 24. Six of these mutants coded for S proteins truncated at position 28. The stop codons were associated with mutations either in position 17 (7w_16) or 21 (In_61, 4w_7, 4w_13 and 11w_4) or in both the residues (4w_12).

The amino acid sequence variability of the WHV variants in comparison with the WHV7 prototype is reported in Fig. 4. In Fig. 4 (a), the variability of the polymerase ORF in the Core/Pol plus Pol fragments (segment 1–167) and in the Pol/PreS/S fragments (segment 168–417) is shown. Serum 4w was characterized by the presence of variants with higher mean variability in both segments (2.7% and 4.0%, respectively); the lowest values (0.3% and 1.2%) were observed for serum 6w, in which the Pol variants were more similar to WHV7 Pol, particularly in the first segment (1–167). Similar variabilities were observed for the inoculum and serum 11w.

The amino acid sequence variability of the PreS/S ORFs are shown in Fig. 4(b). Serum 6w variants were highly related to WHV7. Serum 4w clones diverged from WHV7, especially in the PreS1 and S regions, where the variability reached 5–90% and 4–18%, respectively. Similarly, serum 7w variants showed high variability in the PreS2 and S segments.

Nucleotide variability in Pol and S ORFs according to codon position

The extremely compact hepadnavirus genomes have a common gene organization. Fig. 5(a) shows the way the Pol and PreS/S codons overlap each other. Position 3 of the codons of the Pol ORF corresponds to position 2 of the S codons and position 3 of the S ORF is the first position of the Pol frame. This peculiar organization might allow various selection pressures to affect the variability of the two frames differently. To test whether this operates in vivo, the rate of introduction of mutations in each codon position for fragment 168–417 of the Pol ORF (Fig. 5b) and for the PreS/S ORFs (Fig. 5c) was calculated. The significance of differences was
evaluated by a \( \chi^2 \) test. This analysis showed recurrent variations of position 1 in the Pol ORF, with the exception of serum 7w. In this position, we found 9, 16, 5, 10 and 12 variations for the inoculum and sera 4w, 6w, 7w and 11w, respectively (46.8% of total mutations). Some variations were significantly higher in comparison with the total analysed data \((P < 0.05\) or \(P < 0.01\); see Fig. 5). Conversely, in the PreS/S analysis (Fig. 5c), the higher occurrence of second position mutations in serum 7w variants was significant (83 variations; 59.7% of total mutations). With the aim of studying the effect of non-synonymous variations in each ORF and the existence of selective pressures, the analysis was extended to the evaluation of the \( d_n \) / \( d_s \) ratio of position 3. Furthermore the \( d_n \) / \( d_s \) ratios for all three positions were calculated for Pol and PreS/S ORFs (Table 1). The \( d_n \) / \( d_s \) values were homogeneous in the PreS/S ORFs (range 2.6–3.2), whereas they were more variable in the Pol ORF (lower values for 7w and inoculum; 1.4 and 1.9, respectively) with a high value in serum 4w (3.0), comparable to that found for the PreS/S ORFs (3.2). In serum 7w, the synonymous/non-synonymous mutation ratio in the third position of the Pol ORF was high (19.0), in contrast with the value obtained in the corresponding position 2 of the PreS/S ORF (6.5), indicating a high rate of non-synonymous variations in this latter ORF. The range of values for the \( d_n \) / \( d_s \) ratio was similar in the two ORFs if serum 7w was not considered (2.8–6.3 for Pol and 2.5–6.5 for PreS/S). Interestingly, in serum 6w the ratios were equal in the different frames, and in the inoculum the ratios were quite high for the Pol ORF. None of the ratios was significantly different if compared with the total data.

**Analysis of the virus population composition**

The distribution of genetic distances among the 41 clones characterized was calculated in order to analyse the quasi-

---

**Fig. 5.** Variability analysis applied to the three different codon positions. (a) Respective positions in Pol and PreS/S codons. (b)–(c) Percentage variability calculated in Pol (b) or in PreS/S (c), in single positions (1, 2 or 3) or in multiple positions (1 and 2; 1 and 3; 2 and 3; or 1, 2 and 3, reported as T) in the codons. Significant differences calculated by the \( \chi^2 \) test are indicated by one asterisk \((P < 0.05)\) or two asterisks \((P < 0.01)\).
Table 1. Analysis of synonymous and non-synonymous mutations

The value d_n/d_s is the ratio of the mean number of non-synonymous substitutions per non-synonymous site to the mean number of synonymous substitutions per synonymous site.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Pol ORF</th>
<th>PreS/S ORFs</th>
<th>d_n/d_s</th>
<th>Pol ORF</th>
<th>PreS/S ORFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inoculum (In)</td>
<td>6.3</td>
<td>2.7</td>
<td>1.9</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>4w</td>
<td>3.0</td>
<td>5.0</td>
<td>3.0</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>6w</td>
<td>2.5</td>
<td>2.5</td>
<td>2.2</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>7w</td>
<td>19.0</td>
<td>6.5</td>
<td>1.4</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>11w</td>
<td>2.8</td>
<td>3.8</td>
<td>2.9</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4.3</td>
<td>3.5</td>
<td>2.3</td>
<td>2.9</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Quasispecies composition and homogeneity. (a) Composition of the virus populations. The percentage representing the predominant form is indicated. (b) MGD calculated for the variations in each serum. (c) Percentages of the virus population corresponding to mutants in positions 17, 21 and 28 (M, dotted line) or to termination mutants (TM, solid line).

species in W197, because of the high relatedness of the variants (data not shown). Most distances were equal to or less than 0.02%. This value corresponds to fewer than 25 variations in the 1192 nucleotides evaluated. On this basis, the variants with genetic distances lower than 0.02 were clustered as the predominant species of the virus population in each serum. The quasispecies composition and its homogeneity were analysed by considering the mean genetic distance (MGD), which describes the genetic relatedness quantitatively, as well as the percentage of M or TM variants in the S ORF. The data are summarized in Fig. 6.

A marked heterogeneity amongst the virus population was observed for serum 4w. This serum gave the highest MGD, of 0.0358 (Fig. 6b), and no sequences corresponding to the predominant sequence of the inoculum were identified (Fig. 6a). In this serum, the virus population was largely composed of TM and M variants (42.8% and 28.6%) (Fig. 6c). Conversely, in the serum 6w population, all nine variants analysed could be considered as belonging to the predominant species (MGD 0.0057); no TMs were observed and only 28.6% of the population was mutated in the S region (Fig. 6c). The quasispecies of the inoculum and serum 11w gave similar MGD values and comparable presence of mutants but different homogeneity, with the predominant sequence making up only 50% of the variants analysed in serum 11w versus 78% in the inoculum. The composition of serum 7w was similar to that of...
Discussion

The present study evaluated the evolution of hepadnaviruses in a new host. In the experimental transmission described, a predominant species of WHV was present in the inoculum. The phylogenetic analysis reported in Fig. 1 demonstrates that WHV7 (Cohen et al., 1988; Miller et al., 1990) is the ‘mainframe’ genome. WHV7 DNA has never been used in our laboratory, thus excluding the possibility of PCR contamination. Nevertheless, the sequence of variant In_53 was identical to WHV7. When this virus population was inoculated into W197, the WHV7 genome was found repeatedly (clones 4w–16 and 6w–11). The selective advantage of WHV7-like genomes was demonstrated by the low variability and homogeneity of the virus quasispecies observed in serum 6w and in the sera collected subsequently (Figs 2 and 4), although it has to be remarked that the analysed sequences represent only the most abundant variants and not all quasispecies.

The sera that constituted the inoculum were derived from an in vivo transfection (Cohen et al., 1988). It has been shown previously (Cohen et al., 1988; Miller et al., 1990) that the transmission of WHV7 led to the predominance of this genome; however, the composition of populations and the evolution of quasispecies were not studied. In this work we have described the evolution of the best-represented variants of the quasispecies in the follow-up sera of the newly infected animal. The first viraemic serum, 4w, was extremely heterogeneous and the variants were much less closely related to WHV7 than in the other samples (Figs 1–3 and 6b). These variants do not contain, even as a small percentage, any predominant species (Fig. 6a). In fact, this was the only serum in which WHV7-related sequences did not predominate. The second serum (6w), on the other hand, presented high homogeneity and genetic relatedness, both among the variants and with the prototype. After 7 weeks p.i. (serum 7w) the quasispecies were much more similar to the inoculum.

These results illustrate the adaptation of the virus, as shown by the radiation of variants in the first viraemic serum and by the following phases, characterized by the predominance of one group of these variants highly related to WHV7.

A new ‘hot-spot’ mutation was described at the amino terminus of the S protein and a number of variants showed a protein sequence truncated at position 28 (Fig. 3). Two other amino acid variations, at residues 17 and 21, were also associated with premature termination. The presence and the quantity of mutants (Fig. 6c) might be related to the characteristics of the environment, and selective pressures might have influenced the quasispecies distribution soon after 1 month p.i. At this stage, a large number of TM variants were present, which later disappeared. The existence of these mutants could be a feature of W197 infection and, to our knowledge, no other data on the region of WHV studied are available. In other models, such as mouse hepatitis virus (MHV) and lymphocyte choriomeningitis virus (Meyer & Southern, 1997; Rowe et al., 1997), several mutants were observed in the quasispecies during the early phase of virus replication in vivo. In the case of the spike protein of MHV, the hypothesis of an association of deletion mutants and ‘antigenic oscillation’, to avoid virus clearance, was put forward (Rowe et al., 1997). Further experimental data could clarify the mechanisms involved in the appearance and disappearance of S mutants in our model.

One part of the present study has considered the peculiar feature of the WHV genome whereby, because of its small size (3 kb), it has developed the strategy of multiple overlapping genes encoding different proteins. As shown in Fig. 5, the same nucleotide is in position 1 of the reading frame of the pol gene and in position 3 of the PreS/S codons; similarly, position 2 corresponds to position 1 and position 3 to position 2. We have shown that one of the two ORFs can mutate without the involvement of the other. From the mutation frequencies in each position, we observed higher values of variability for position 1 in the Pol ORF and for position 2 in the S ORF in serum 7w (P < 0.05). Furthermore, analysis of the sequences performed by the evaluation of synonymous/non-synonymous mutation ratios in position 3 has shown that changes in the S ORF did not affect the Pol ORF (ratio in serum 7w, 19:0). Therefore, single nucleotide mutations affected only one of the overlapping ORFs and could produce a neutral (i.e. synonymous) mutation in the other. These natural mutations can provide a reservoir of nucleotide variability in both frames which may allow rapid adaptability or may drive the evolution of virus proteins, thereby ‘subverting essentiality’ (Martinez et al., 1997). This has been recently shown for foot-and-mouth disease virus (Martinez et al., 1997). Furthermore, the mutations were not concentrated in the same codon position in the Pol and S ORFs.

Finally, we have calculated the d<sub>n</sub>/d<sub>s</sub> index. Recent studies have supported the theory that a high number of non-synonymous variations and the related increase of the index may indicate a predominance of positive pressures on the virus population (Zhang et al., 1997; Salvatori et al., 1997). Using this index, we analysed the changes that occurred as a consequence of transmission to another host and virus adaptation during the follow-up. As shown in Table 1, increasing pressures in the Pol ORF were found in serum 4w and 11w (3:0 and 2:9) with a strong decrease in serum 7w (1:4). Pressure on S genes was evident, especially in sera 4w, 7w and 11w (3:2, 2:8 and 2:8, respectively).

These data suggest that the early phases of the infection of W197 might have been driven by evolution of both structural and non-structural proteins (serum 4w). We have not studied the nature of the pressures involved; however, it has been
established for human HBV infections that cytotoxic T cells recognize multiple epitopes of the core, envelope and polymerase polypeptides, producing a vigorous response in acute self-limiting hepatitis (Bertoni et al., 1997). The precise onset and the link with the CD4+ response and B-cell activity is not known (Koziel, 1996; Chisari, 1997). These findings have been partially confirmed in the woodchuck model (Menne et al., 1997, 1998). In woodchucks that were infected acutely, the T-cell response against the envelope proteins began 3 weeks after infection, followed by responses to the immunodominant T-cell epitope of WHcAg (Menne et al., 1998).

Other possible selective pressures acting on hepadnavirus quasispecies are those involved in the functions of the S protein and of the polymerase (reviewed by Lau & Wright, 1993), such as binding to the receptor and replicase activity. Therefore, the absence of or the limited pressures on the pol gene observed in serum 7w may be explained in terms of the successful adaptation of the predominant sequence of serum 6w or in terms of the absence of immunological attack on this region. Conversely, the increased variability in serum 11w might indicate a change of cell targets near the conclusion of the acute phase, which can be characterized by an increase of the cytotoxic T-cell response.

In conclusion, this study attempted to trace the fate of an experimental infection in a new host. A homogeneous virus population was propagated and the resulting quasispecies were able to infect a new host persistently. Selective forces in the new host shaped the infecting population but the high fitness of WHV7 gave rise to the predominance of variants highly related to the prototype genome. Moreover, the dynamics operating on virus progeny may now be described more thoroughly. The successful molecular organization of hepadnavirus genome that controlled the variability in overlapping frames was characterized. Further studies will evaluate infection in more than one woodchuck, to clarify the functions of the TM and M variants and to trace the variations of quasispecies linked to the passage from the acute to the chronic phases of infection.

We would like to thank Armando Cesolini in the animal facility for technical support and Cinzia Bisegra for editorial and secretarial assistance. This work was supported in part by the Viral Hepatitis Project, Istituto Superiore di Sanità, and by EC grant no. BIO4-CT98-0002 (DG 12 – SSMI).

References


Received 4 August 1998; Accepted 21 October 1998