

## Evidence for quasispecies distributions in the human hepatitis A virus genome

Glòria Sánchez,<sup>a</sup> Albert Bosch,<sup>a,\*</sup> Gema Gómez-Mariano,<sup>b</sup> Esteban Domingo,<sup>b</sup> and Rosa M. Pintó<sup>a</sup>

<sup>a</sup> Grup Virus Entèrics, Department of Microbiology, University of Barcelona, 08028 Barcelona, Spain

<sup>b</sup> Centro de Biología Molecular "Severo Ochoa," CSIC-UAM, Cantoblanco, 28049 Madrid, Spain

Received 7 April 2003; returned to author for revision 20 May 2003; accepted 12 June 2003

### Abstract

Nucleotide sequence analysis of multiple molecular clones of the hepatitis A virus (HAV), generated by reverse transcription-PCR of two capsid-coding regions, revealed a degree of heterogeneity compatible with a quasispecies structure in three clinical samples. Passage of plaque-purified reference strain HAV pHM175 43c in FRhK-4 cells documented the generation of a mutant distribution of HAV genomes. The mutant spectra showed mutation frequencies in the range of  $1 \times 10^{-3}$  to  $1 \times 10^{-4}$  substitutions per nucleotide, with a dominance of transition over transversion mutations. While in the VP3-coding region, nonsynonymous mutations were predominant; in the VP1-coding region they were uncommon. Around 50% of the amino acid replacements involved residues located at or near antigenic sites. Most of the detected mutations occurred at or in the vicinity of rare codons, suggesting a dynamics of mutation-selection, predominantly at and around rare codons. The results indicate that despite antigenic conservation, HAV replicates as a complex distribution of mutants, a feature of viral quasispecies.

© 2003 Elsevier Inc. All rights reserved.

**Keywords:** HAV; Quasispecies; Antigenic sites; Codon-usage; Rare codon; Normal codon

### Introduction

Hepatitis A virus (HAV), classified as the type species of the genus *Hepatitisvirus* within the *Picornaviridae* family (van Regenmortel et al., 2000), is a hepatotropic virus which represents a significant problem for human health (Battegay and Feinstone, 1997; Hollinger and Emerson, 2001). The virion capsid is composed of the structural proteins VP1, VP2, VP3, and possibly VP4, encoded in the P1 region of the genome (Hollinger and Emerson, 2001; Racaniello, 2001).

Some degree of nucleotide sequence heterogeneity of the P1 genomic region has been observed among independent HAV isolates from different regions of the world (Lemon et

al., 1987; Robertson et al., 1992; Taylor, 1997; Arauz-Ruiz et al., 2001; Costa-Mattioli et al., 2001). However, this variability at the nucleotide level is not reflected in an equivalent degree of variation at the amino acid level (Lemon and Robertson, 1993; Hollinger and Emerson, 2001; Sánchez et al., 2003). The high degree of conservation of the amino acid sequences of the capsid proteins of HAV entails a low antigenic diversity, and therefore, only a single serotype of human HAV has been recognized (Hollinger and Emerson, 2001). This suggests the operation of severe structural constraints in the HAV capsid (Sánchez et al., 2003). HAV shows a high codon usage bias, with the repeated occurrence of 22 rare codons for 14 amino acids (Sánchez et al., 2003). Most of the carboxy-terminal regions of  $\beta$ -barrels and  $\alpha$ -helices, that are predicted in the capsid proteins (Luo et al., 1988), include residues encoded by conserved rare codons, suggesting a potential function of such codons in a decrease of the rate of translation to facilitate the proper folding of the capsid proteins (Sánchez et al., 2003).

The molecular basis of the genetic variability of RNA

\* Corresponding author. Department of Microbiology, School of Biology, University of Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain. Fax: +34-934034629.

E-mail address: [abosch@ub.edu](mailto:abosch@ub.edu) (A. Bosch).

viruses has been associated with the absence of a 3' → 5' exonuclease proofreading activity in viral RNA-dependent RNA polymerases and reverse transcriptases, together with lack of postreplicative repair mechanisms that can act on DNA but not on RNA (Holland et al., 1992; Steinhauer et al., 1992; Williams and Loeb, 1992; Domingo and Holland, 1997). Mutation rates for a variety of RNA viruses range between  $10^{-4}$  and  $10^{-5}$  substitutions per nucleotide copied (Batschelet et al., 1976; Drake, 1993; Drake and Holland, 1999). As a consequence, RNA viruses replicate as complex dynamic mutant distributions, termed viral quasispecies (Eigen and Biebricher, 1988; Holland et al., 1992; Domingo and Holland, 1997; Domingo et al., 2001). The open reading frame of the putative HAV 3D (the RNA-dependent RNA polymerase) does not provide any evidence for the presence of a proofreading or error-correcting activity known to be associated with several DNA-dependent DNA polymerases (Kunkel, 1988; Zimmern, 1988; Friedberg et al., 1995). Therefore, mutation rates and frequencies for HAV are not expected to differ significantly from those of other picornaviruses and RNA viruses in general (Drake and Holland, 1999). In this view, the 90% or higher amino acid sequence identities among independent strains and isolates of HAV (Hollinger and Emerson, 2001) would be the result of negative selection on many newly arising mutants and convergence of consensus or average sequences (Eigen and Biebricher, 1988; Holland et al., 1992; Domingo et al., 2001). Yet if mutational pressure originated a number of mutants hidden in a mutant spectrum, such mutants would provide evidence of quasispecies dynamics, implying the presence of a variant reservoir for HAV adaptation. Despite the biological significance of this population structure, no such analyses of HAV mutant spectra have been reported.

In the present work, results suggesting the occurrence of distributions of related genomes in HAV are presented, both with clinical isolates of HAV and in populations evolved in cell culture from a biological clone of HAV. The nature of the mutations detected in the mutant spectra supports the operation of structural constraints in the capsid of HAV. Furthermore, the results of HAV evolution in cell culture revealed the presence of antigenic variants that were generated in the absence of immune selection—as previously observed with other viruses (Domingo et al., 1993).

## Results

### *A mutant spectrum in clinical isolates of HAV*

To test whether the limited amino acid variation among consensus sequences of independent HAV isolates was paralleled by a homogeneous mutant spectrum within a HAV population, a clonal analysis of three clinical samples of HAV was performed. The samples were Val9 (stool), Val10 (stool), and Val12 (serum) (Sánchez et al., 2002), whose origin is described under Materials and methods. A mean of

25 molecular clones from a single reverse transcription (RT)-PCR amplification of a fragment of the VP3 (from nucleotides 1470 to 1839, encoding amino acids 1 to 123) and a fragment of the VP1 (from nucleotides 2459 to 2943, encoding amino acids 85 to 245), which span sequences encoding the main antigenic sites of HAV (Nainan et al., 1992; Ping and Lemon, 1992; Bosch et al., 1998), were analyzed for each sample. Minimum mutation frequencies (counting repeated mutations only once) ranged between  $1.2 \times 10^{-3}$  and  $1.4 \times 10^{-4}$  substitutions per nucleotide (Table 1). In some analyses, the maximum mutation frequency (counting all mutations relative to the consensus) increased over the minimal mutation frequency. The values of Shannon entropy (a measure of the proportion of different sequences in the set analyzed) indicated heterogeneity of the mutant spectra. Control experiments (described in detail under Materials and methods) indicated that the mutation frequencies observed cannot be the result of misincorporations during RT-PCR amplification of HAV RNA. Statistical analysis (nonparametric Mann–Whitney *U* test and ANOVA) revealed significant differences between the mutation frequencies of the patient samples and the error rate of the RT-PCR system, except in the Val10 sample. However, despite this lack of statistical significance, the localization of the mutations (see below) make it extremely unlikely that they could be just due to the occurrence of random misincorporations. The mutation-frequencies and Shannon entropies of the mutant spectra are in the range observed in other viral quasispecies (see Discussion).

### *A mutant spectrum in serially passaged HAV pHM175 43c*

The sequence comparisons indicated the occurrence in the HAV isolates of dominant genomic sequences together with several variants. These results suggest generation of mutants in the course of replication of each individual HAV. However, a definitive proof of mutant generation requires a demonstration that multiple variants are produced *de novo* upon replication of a single HAV genome (Domingo et al., 1978; Domingo, 1996). To this effect, the reference HAV strain pHM175 43c, adapted to FRhK-4 cells, was subjected to three successive plaque isolations to produce pHM175 43c P0, and then this clonal population was subjected to 26 serial passages in the same cells (about  $10^6$  PFU infecting  $10^6$  cells per passage), as detailed under Materials and methods. The resulting population is termed pHM175 43c P26. No mutations were detected in the consensus sequences of the VP1- and VP3-coding regions analyzed, at passages 1, 5, 11, 16, and 26. However, a clonal analysis revealed 23 mutations in 100 molecular clones from passage 26 (Table 2). The minimum mutation frequencies were  $7.0 \times 10^{-4}$  and  $3.3 \times 10^{-4}$  substitutions per nucleotide for the VP3- and VP1-coding regions, respectively. Only for the VP3-coding region the maximum mutation frequency was slightly higher than the minimum mutation frequency.

Table 1  
Characterization of the mutant spectrum of HAV in clinical samples

Genomic region	Mutations <sup>a</sup> nucleotides sequenced	$T_s^b$	$T_v^b$	$N_{syn}^b$	Syn <sup>b</sup>	Indel or stop <sup>b</sup>	Nucleotide mutation frequency <sup>c</sup>		Maximum amino acid mutation frequency <sup>c</sup>	$S_N^d$
							Minimum	Maximum		
VP3 Val9	6/8773	3	3	4	2	0	$6.8 \times 10^{-4}$	$6.8 \times 10^{-4}$	$1.3 \times 10^{-3}$	0.31
VP3 Val10	1/6960	1	0	0	1	0	$1.4 \times 10^{-4}$	$1.4 \times 10^{-4}$	0	0.05
VP3 Val12	5/9913	4	1	4	1	0	$4.0 \times 10^{-4}$	$5.0 \times 10^{-4}$	$1.2 \times 10^{-3}$	0.22
VP1 Val9	40/13,035	38	2	0	40	0	$1.2 \times 10^{-3}$	$3.1 \times 10^{-3}$	0	0.50
VP1 Val10	2/9962	2	0	0	2	0	$2.0 \times 10^{-4}$	$2.0 \times 10^{-4}$	0	0.12
VP1 Val12	19/12,248	19	0	0	19	0	$6.5 \times 10^{-4}$	$1.5 \times 10^{-3}$	0	0.26

<sup>a</sup> Mutant residues are those that vary relative to the corresponding consensus sequences

<sup>b</sup> Detected mutations were classified into transitions ( $T_s$ ) versus transversions ( $T_v$ ) nonsynonymous (inducing an amino acid substitution) versus synonymous or silent (not inducing an amino acid substitution) and insertion/deletion mutations (inducing an erroneous reading frame)

<sup>c</sup> The minimum nucleotide mutation frequency is the number of different mutations found divided by the total number of nucleotides sequenced. The maximum nucleotide mutation frequency is the total number of mutations found divided by the total number of nucleotides sequenced. The maximum amino acid mutation frequency is the total number of nonsynonymous mutations divided by the number of amino acids encoded in the sequence analyzed. Mutation frequencies are expressed as substitutions per nucleotide or amino acid substitutions per amino acid.

<sup>d</sup> The normalized Shannon entropy was calculated as  $S_n = [-\sum_i (p_i \times \ln p_i)]/1nN$ , where  $p_i$  is the frequency of each sequence and  $N$  is the total number of sequences in the spectrum of mutants.

Statistical analysis (nonparametric Mann–Whitney  $U$  test and ANOVA) revealed significant differences between these mutation frequencies and the experimental error rate. The values of Shannon entropy indicated heterogeneity of the mutant spectra. The consensus and the dominant sequences were identical. Thus, the results document constancy of the average or consensus sequences, and the generation of a mutant spectrum upon replication of a biological clone of HAV. These are features of quasispecies dynamics (Eigen and Biebricher, 1988). Experiments are in progress to elucidate whether the virus mutant distribution varies with cell passage.

#### Types and location of mutations: mutation clustering

A dominance of transitions over transversions was observed in all HAV mutant spectra analyzed (Tables 1 and 2). In contrast, the ratio of synonymous-to-nonsynonymous

mutations was 33 for the VP1-coding region and 0.5 for the VP3-coding region (average for all mutations detected in the four mutant spectra analyzed, Tables 1 and 2). This suggests higher constraints for variation at the amino acid level for VP1 than for VP3. As expected, no bias was observed in the few mutations recorded in the control amplification of an RNA transcript (basal RT-PCR error, described under Materials and methods).

The positions of the mutations and amino acid substitutions indicate the occurrence of several clusters of mutations in the mutant spectra of both clinical samples and in the clonal population pHM175 43c P26 (Fig. 1). Clusters of substitutions at amino acid positions 37, 40, 41 and at 82, 83, 84 of VP3 in the mutant spectra of pHM175 43c P26, and at 179, 180, and 181 of VP1 in the mutant spectra of the clinical samples were remarkable. Some substitutions were observed in independent mutant spectra (i.e., at positions 93, 104, 116, 131, 140, 181, and 217 of VP1), suggesting

Table 2  
Characterization of the mutant spectrum of the pHM175 43c P26

Genomic region	Mutations <sup>a</sup> nucleotides sequenced	$T_s^b$	$T_v^b$	$N_{syn}^b$	Syn <sup>b</sup>	Indel or stop <sup>b</sup>	Nucleotide mutation frequency <sup>c</sup>		Maximum amino acid mutation frequency <sup>c</sup>	$S_N^d$
							Minimum	Maximum		
VP3	15/18,500	12	3	10	5	0	$7.0 \times 10^{-4}$	$8.1 \times 10^{-4}$	$1.6 \times 10^{-3}$	0.31
VP1	8/24,150	7	1	2	5	1 Stop	$3.3 \times 10^{-4}$	$3.3 \times 10^{-4}$	$2.5 \times 10^{-4}$	0.20

<sup>a</sup> Mutant residues are those that vary in comparison with the consensus sequence at P26.

<sup>b</sup> Detected mutations were classified as indicated in footnote b of Table 1.

<sup>c</sup> Mutation frequencies are as defined in footnote c of Table 1.

<sup>d</sup> The normalized Shannon entropy was calculated as indicated in footnote d of Table 1.

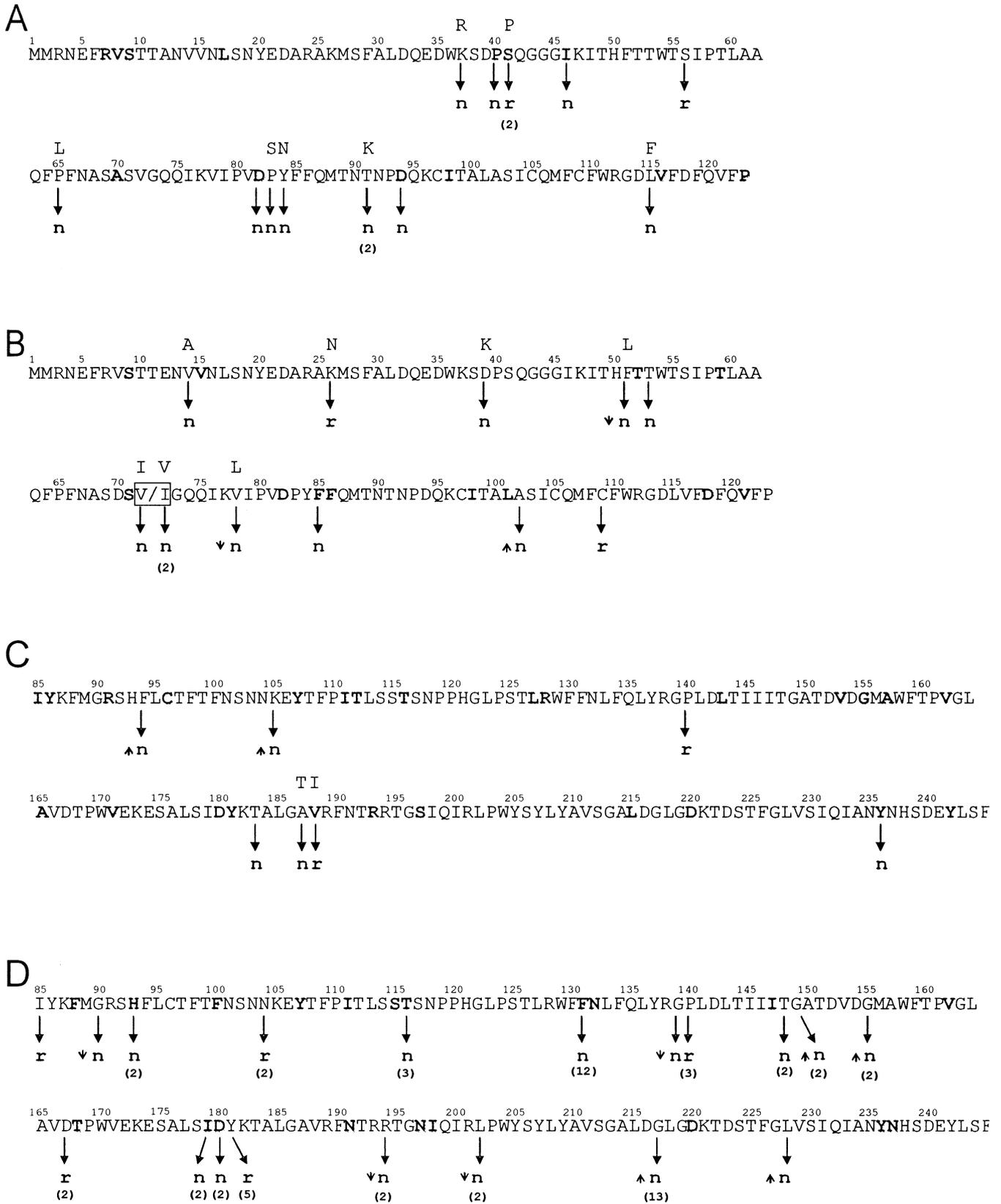


Fig. 1. Amino acid sequence of the VP3 and VP1 fragments analyzed in the present study. (A) VP3 sequence of the pHM175 43c strain (B) VP3 sequence of the clinical samples. (C) VP1 sequence of the pHM175 43c strain. (D) VP1 sequence of the clinical samples. In bold are depicted those amino acids encoded by rare codons. Arrows indicate those positions in which a mutation occurred. Regarding the codon usage (Sánchez et al., 2003), the letter n designates a normal (frequent) new codon; ↑ n, a significantly more normal (frequent) new codon than the old one; ↓ n, a significantly less normal (frequent) new codon than the old one; and r, a rare new codon. Numbers in parentheses indicate the number of clones with a given substitution. At those positions in which a nonsynonymous substitution occurred, the new amino acid is depicted over the sequence row. At position 72 in (B) either a valine or an isoleucine may be present depending on the clinical sample.

again the operation of constraints that limit the number of positions that tolerate replacements, giving the appearance of mutational hot spots (see Discussion). Seven of 15 amino acid substitutions were either at residues previously identified as belonging to antigenic sites (i.e., 65, 72, and 115 of VP3) or at residues in the neighborhood of antigenic sites (i.e., 78, 83, 84, and 91 of VP3) (Ping and Lemon, 1992; Bosch et al., 1998; Hollinger and Emerson, 2001).

A large proportion of mutations in the mutant spectra analyzed involved either rare codons, defined as those codons used at low frequencies (Gavrillin et al., 2000) or normal codons located in the neighborhood of rare codons (at a maximum distance of three triplets from a rare codon) (Fig. 1). The majority of the mutations involving rare codons (90%) resulted in the change to a normal codon. However, most of the resulting normal codons (78%) were contiguous to a rare codon (Fig. 1). In contrast, most mutations involving normal codons (74%) induced changes which significantly affected the normality of the codon, and a high percentage of them (72%) were again closely located to rare codons (Fig. 1). Therefore, both the types of mutations and the types of codons affected suggest variation constraints in HAV.

## Discussion

Despite the invariance of consensus sequences, at the VP3- and VP1-coding regions examined, upon serial passage of a biological clone of HAV, each of the populations analyzed consisted of a mutant spectrum in agreement with a quasispecies population structure for this important human pathogen (Tables 1 and 2). From the minimum mutation frequencies and since the HAV genome is about 7500 nucleotides in size [specifically, the genome of the original HM175 wild-type HAV is 7493 nucleotides long (Cohen et al., 1987) and the genome of the cell culture-adapted strain pHM175 43c is 7503 nucleotides long (Lemon et al., 1991)], the average number of mutations per genome (relative to the consensus sequence) for Val9, Val10, Val12, and the clonal population pHM175 43c P26 is 7.0, 1.2, 3.9, and 3.8, respectively. This calculation assumes that the mutation frequencies for the analyzed genomic regions reflect those of other genomic regions, a point that would require sequencing of other genomic regions from components of mutant spectra. However, comparison of nucleotide sequences from different HAV isolates suggests a considerable similarity in the degree of conservation of different genomic regions (Hollinger and Emerson, 2001; Ching et al., 2002; Sánchez et al., 2003). In these comparisons, the most variable genomic regions are those encoding 2B, 2C, and 3B (representing 25% of the genome), with an average of 1.1-fold the genetic distance calculated for the analyzed VP3- and VP1-coding regions, the most conserved genomic regions are the highly conserved noncoding regions and those encoding 3C and 3D (representing 39% of the ge-

nome), with an average of 0.8-fold the genetic distance calculated for the VP3- and VP1-coding regions. Since in other viruses conservation among independent isolates often parallels conservation within mutant spectra (Arias et al., 2001; Domingo et al., 2001), the heterogeneity quantitated on the basis of the VP3- and VP1-coding regions analyzed is unlikely to differ substantially from the heterogeneity for the entire HAV genome. With this assumption, the proportion of genomes with no mutations (the dominant sequence), calculated from the Poisson distribution, would be 0.09, 30, 2.0, and 2.2% for Val9, Val10, Val12, and pHM175 43c P26 respectively. These values are comparable to those estimated for clonal populations of other RNA viruses such as bacteriophage Q $\beta$  (Domingo et al., 1978) or FMDV (Sobrino et al., 1983; Arias et al., 2001; review in Domingo et al., 2001). Thus, in these HAV populations the nonmutated class of genomes were a minority of the total.

Of the analyzed mutant spectra, only Val10 had a level of heterogeneity that could be influenced by misincorporations during RT-PCR amplification. Several arguments support the fact that the great majority of the mutations found, even in Val10, was present in the RNA populations under study and were not the result of misincorporations during the RT-PCR procedure employed. First, a basal error rate for the RT-PCR procedures was experimentally determined for the same VP3- and VP1-coding regions under study. This control experiment (detailed under Materials and methods) used recombinant clones to prepare VP3 and VP1 transcripts with T7 DNA-dependent RNA polymerase. The transcripts were then subjected to the same RT-PCR amplification used with HAV RNA and 50 cDNA clones were sequenced. The final error rate attributable to the system is  $7.9 \times 10^{-5}$  by  $8.1 \times 10^{-5}$  substitutions per nucleotide for the VP3- and VP1-coding regions, respectively. Moreover, the mutations that occurred during the RNA synthesis and the RT-PCR amplification (three transitions and two transversions, three nonsynonymous and two synonymous mutations) did not show any of the mutational type bias or mutation clustering observed in the HAV populations. For obvious statistical reasons, the probability of mutation clustering as an RT-PCR artifact is negligible. The mutations in the mutant spectrum of Val10 were located in the vicinity of the sites where mutations for the other samples have been mapped. Therefore, the vast majority of mutations scored in the mutant spectra must have been present in the HAV RNA populations examined. Although it cannot be excluded that some mutation could have been generated during the *in vitro* amplification procedure, its exclusion would not significantly modify the quantifications and conclusions on mutant spectrum complexity of HAV. The narrower mutant spectrum in Val10 could not be the result of a limitation in the number of RNA template molecules in the sample, since in all cases a dilution of at least 1:100 of the preparation of the RNAs used as template produced a positive amplification, which excludes a bias in the sequence repertoire (Airaksinen et al., 2003). The narrower mutant spectrum of

this isolate could result from a shorter time from a clonal origin of the infection, a limited HAV replication, or higher sequence constraints than for the other isolates, among other possibilities (Domingo et al., 2001).

A complex mutant spectrum for HAV isolates poses a problem regarding the concept of strain for this virus. For hepatitis C virus, a genomic distance higher than 5%, in a genomic fragment encompassing the E2-NS2 junction, is required for two HCVs to be considered a different strain (Cabot et al., 2000), and different HAV and poliovirus genotypes are those with higher than 15% nucleotide sequence divergence in the VP1-2A-coding region (Rico-Hesse et al., 1987; Hollinger and Emerson, 2001). As more consensus sequences and data on mutant spectra of HAV become available, it may be necessary to redefine limits for strain and genotype classifications.

The mean ratio of transition to transversion mutations in the RNA of the analyzed populations was 8.6, probably reflecting the misincorporations tendencies of RNA dependent RNA polymerases (Domingo et al., 1978; Kuge et al., 1989; Schnerder and Roossinck, 2000). Also, the proportion of transitions tends to decrease with the divergence of the genes compared (Villanueva et al., 1983; Nei, 1987), supporting a recent occurrence of the mutations observed in these HAV samples. Examination of the sequence context in which mutations are found indicates that 25% of the mutations in clusters are located at the ends of short oligo (A) or oligo (U) stretches, suggesting a possible contribution of polymerase slippage in the generation of some of these mutations (Ripley, 1990; Arias et al., 2001). However not all homopolymeric tracts were associated with mutations only 2 of 14 (2/14) A3 tracts; 3/33 U3; 1/7 A4; 1/10 U4; 2/3 A5; 4/8 U5.

Several amino acid substitutions in components of the mutant spectra are located at or near recognized antigenic sites of HAV (Fig. 1). In particular in population pHM175 43c P26 passaged in cell culture in the absence of immune selection, substitutions at positions 65 (Pro → Leu), 83 (Pro → Ser), and 84 (Tyr → Asn) of VP3 are related to the immunodominant site of HAV (Nainan et al., 1992; Luo et al., 1988), and 115 (Leu → Phe) and 91 (Thr → Lys) affect a continuous epitope of VP3 (Bosch et al., 1998; M. Luo, personal communication). Antigenic variation in the absence of immune selection has been observed in several other viruses (reviewed in Domingo et al., 1993) and has been attributed to higher tolerance to replacements of surface residues which are relatively free of structural constraints (Domingo et al., 1993; Haydon and Woolhouse, 1998). Replacements at or in the neighborhood of antigenic sites were also found in the mutant spectra of Va19 [substitution Val→Ile at position 72 of VP3, which is involved in the epitope defined by the monoclonal antibody K34C8 (Sánchez et al., 2002)] or Val12 [substitutions Ile→Val at position 72 of VP3, and Val→Leu at position 78 of VP3]. These replacements are in the vicinity of the immunodominant site (Luo et al., 1988).

One of the unusual features of HAV is the abundance of rare codons at some capsid sites, compatible with a modulating effect on HAV translation (Sánchez et al., 2003). The fact that the mutations observed tended to maintain a minimal frequency of rare codons suggests a constraint in the dynamics of mutation-selection to preserve the previously postulated balance between rate of translation and capsid protein folding (Sánchez et al., 2003). In HAV, this may be an important evolutionary constraint, additional to those generally accepted to limit evolutionary rates of RNA viruses despite high mutation rates (Simmonds and Smith, 1999).

A quasispecies dynamics for HAV may be of consequence for the natural history of this pathogen and its control in the human population. Pathogenic manifestations of HAV are quite variable (reviewed in Hollinger and Emerson, 2001). Patients may be asymptomatic despite active viral replication, but relapsing and fulminant forms of hepatitis, as well as extrahepatic manifestations such as encephalopathy, have been described. Although a strong host component undoubtedly must influence disease outcome, a possible participation of virus variants to respond to host defense mechanisms or to favor replication in the face of physiological alterations cannot be excluded. In this context, it is remarkable that the pattern of mutations found in the analyzed VP3 and VP1 regions differed between the pHM175 and the clinical isolates, likely due to different selective pressures. The existence of variant reservoirs in HAV populations should also be taken into consideration in the design of preventive and therapeutic treatments, despite antigenic conservation of the virus.

## Material and methods

### *Cells, viruses, and infections*

The cytopathogenic pHM175 43c strain of HAV (courtesy of T. Cromeans, Centers for Disease Control, Atlanta, GA) was three times plaque-purified in FRhK-4 cells, as previously described (Cromeans et al., 1987), and a biological clone (pHM175 43c P0) was serially passaged 26 times in the same cell line, as previously described (Bosch et al., 1998) to yield population pHM175 43c P26.

### *Clinical samples*

Three HAV strains were isolated from three different patients of an outbreak of acute hepatitis A associated with the consumption of coquina clams in Valencia, Spain, during autumn–winter 1999 (Sánchez et al., 2002). One virus strain (Va19) was isolated from 60  $\mu$ l of serum, and two additional strains (Va110 and Va112) were isolated from 60  $\mu$ l of extracted feces as previously described (Sánchez et al., 2002). All strains belonged to genotype IB and presented an overall nucleotide homology of the consensus sequences of

99.7% in a fragment of the 5'NCR and 99.3% in the region encoding the VP1-2A junction (Sánchez et al., 2002), which indicates an epidemic relationship among them.

### Molecular cloning and sequencing

RNA extracted from pHM175 43c P26 or from the clinical samples was retrotranscribed to a cDNA with the M-MLV reverse transcriptase (Promega), and the cDNA was copied and amplified by the thermostable *Pwo* pol from *Pyrococcus woesei*, which has proofreading activity (error rate of  $3.2 \times 10^{-6}$  substitutions/nucleotide) (Mullan et al., 2001). Two genomic regions coding for capsid proteins were amplified. A fragment of the VP3-coding region (nucleotides 1470 to 1839, corresponding to amino acids 1 to 123) and a fragment of the VP1-coding region (nucleotides 2459 to 2943, corresponding to amino acids 85 to 245), which include most of the epitopes so far described in HAV (Nainan et al., 1992; Ping and Lemon, 1992; Bosch et al., 1998). The primers used to copy and amplify the VP3-coding region are NH<sub>2</sub>-VP3 (5' TCTACCTGAATGATATTTGG 3') for the cDNA synthesis, and NH<sub>2</sub>-VP3 and VP3-1431B (5' CTTGGATCCCCTCAATGTTT-TAGCTAGA 3') for PCR amplification. The primers used to copy and amplify the VP1-coding regions are VP1-2965 (5' TCTGTGACAGACAGACAAATAACAAC 3') for the cDNA synthesis, and VP1-2965 and VP1-2428 (5' GAGG-GATCCGACATACATCAGATCATATGTC 3') for PCR amplification. The synthesis of cDNAs was performed in a final volume of 25  $\mu$ l containing 8 units of M-MLV RT, 0.2 mM of each nucleotide, 0.5  $\mu$ M primer, 50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl<sub>2</sub>, and 10 mM DTT. Ten microliters of HAV RNA was denatured at 99°C for 5-min and incubated at 45°C for 1 h. DNA amplification was performed following the manufacturer's specifications in a final volume of 50  $\mu$ l containing 0.5 U of the *Pwo* pol, 0.2 mM of each nucleotide 0.5  $\mu$ M of each primer, 10 mM Tris-HCl, 25 mM KCl, 5 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, and 10  $\mu$ l of the RT product, and using an annealing temperature of 50°C in both reactions. Since the DNA fragments produced by *Pwo* pol are blunt-ended, and primers VP3-1431B and VP1-2428 were designed to include a *Bam*HI restriction site at their 5' end, the amplification products were cloned into pGEM-3Zf(+). PCR products were digested with *Bam*HI and the plasmid vector with both *Bam*HI and *Hinc*II. Digested DNAs were purified with the High Pure PCR Product Purification Kit (Roche) following the directions of the manufacturer. DNA ligations were performed overnight at 16°C using T4 DNA ligase. Ligation products were transformed in *Escherichia coli* DH5 $\alpha$ , and transformant clones were screened first by the standard white/blue  $\beta$ -galactosidase colorimetric reaction and then confirmed by colony hybridization with specific digoxigenin-labeled probes. Plasmid DNA from each clone was purified by using the Wizard Plus SV Minipreps Kit (Promega). Nucleotide sequencing was carried out in an ABI

PRISM 377 automated DNA sequencer, with the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and using vector-derived primers, as described elsewhere (Arias et al., 2001). All mutations were confirmed by sequencing both strands of DNA.

### Error rate of the RT-PCR system

To ensure that the observed heterogeneity was due to the HAV polymerase and not to artifactual misincorporations introduced by the RT or *Pwo* polymerases during the amplification procedure, a control experiment to determine the error rate of the system was carried out. One recombinant clone of the VP3 fragment and one recombinant clone of the VP1 fragment, obtained by copying RNA from pHM175 43c, were in vitro transcribed with the T7 polymerase. The RNA transcripts, diluted 1/1000, were subjected to the original RT-PCR procedure and the amplified products were subcloned. The sequence of 50 clones from each VP3- and VP1-coding region yielded an error rate for the amplification system of  $1.1 \times 10^{-4}$  and  $1.7 \times 10^{-4}$  mutations per nucleotide, respectively. Since the error rate of the T7 RNA polymerase is about  $2.9 \times 10^{-5}$  (Remington et al., 1998), the final error rate of the system was calculated to be  $8.1 \times 10^{-5}$  and  $1.4 \times 10^{-4}$  for the VP3- and VP1-coding regions, respectively. The ANOVA and the nonparametric Mann-Whitney *U* tests were employed to compare viral mutation frequencies and the system error rates, revealing statistical significance in all but one case (Va110). These data, together with the types and location of the mutations scored (described under Results and Discussion) indicate that the vast majority of mutations detected must have been present in the mutant spectra of the HAV samples analyzed.

### Sequence analysis

The quasispecies complexity was analyzed by calculating the mutation frequencies and the Shannon entropy. Minimum and maximum mutation frequencies were determined as previously described (Arias et al., 2001). Normalized Shannon entropies were calculated following the formula  $S_n = -[\sum_i (p_i \times \ln p_i)]/\ln N$ , where  $p_i$  is the frequency of each sequence and  $N$  is the total number of sequences in the spectrum of mutants (Airaksinen et al., 2003).  $S_n$  ranges from 0 (no diversity) to 1 (maximum diversity). The codon usage table of HAV defined previously (Sánchez et al., 2003) was used in the analyses of codon abundances.

### Acknowledgments

We acknowledge the technical expertise of the Serveis Científic-Tècnics of the University of Barcelona. Work in Barcelona was supported in part by Grants ERB3514PL973098, QLRT-1999-0634, and QLRT-1999-

0594 from the European Union, BIO2002-02625 from the CICYT, Spain, and 2001/SGR/00098 from the Generalitat de Catalunya. Work in Madrid was supported by Grants ERB3514PL973098 from EU, BMC2001-1823-C02-01 from CICYT, and by Fundación Ramón Areces.

## References

- Airaksinen A., Pariente, N., Menéndez-Arias, L., Domingo, E., 2003. Curing of foot-and-mouth disease virus from persistently infected cells by ribavirin involves enhanced mutagenesis. *Virology*, 311, 334–349.
- Arauz-Ruiz, P., Sundqvist, L., Garcia, Z., Taylor, L., Visona, K., Norder, H., Magnius, L.O., 2001. Presumed common source outbreaks of hepatitis A in an endemic area confirmed by limited sequencing within the VP1 region. *J. Med. Virol.* 65, 449–456.
- Arias, A., Lázaro, E., Escarmis, C., Domingo, E., 2001. Molecular intermediates of fitness gain of an RNA virus: characterization of a mutant spectrum by biological and molecular cloning. *J. Gen. Virol.* 82, 1049–1060.
- Batschelet, E., Domingo, E., Weissmann, C., 1976. The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* 1, 27–32.
- Battegay, M., Feinstone, S.M., 1997. Hepatitis A virus, in: Willson, R.A. (Ed.), *Viral Hepatitis*, Marcel Dekker, New York, pp. 35–84.
- Bosch, A., Gonzalez-Dankaart, J.F., Haro, I., Gajardo, R., Perez, J.A., Pinto, R.M., 1998. A new continuous epitope of hepatitis A virus. *J. Med. Virol.* 54, 95–102.
- Cabot, B., Martell, M., Esteban, J.I., Sauleda, S., Otero, T., Esteban, R., Guardia, J., Gomez, J., 2000. Nucleotide and amino acid complexity of hepatitis C virus quasispecies in serum and liver. *J. Virol.* 74, 805–811.
- Ching, K.Z., Nakano, T., Chapman, L.E., Demby, A., Robertson, B.H., 2002. Genetic characterization of wild-type genotype VII hepatitis A virus. *J. Gen. Virol.* 83, 53–60.
- Cohen, J.I., Ticehurst, J.R., Purcell, R.H., Buckler-White, A., Baroudy, B.M., 1987. Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses. *J. Virol.* 61, 50–59.
- Costa-Mattioli, M., Ferre, V., Monpoeho, S., Garcia, L., Colina, R., Billaudel, S., Vega, I., Perez-Bercoff, R., Cristina, J., 2001. Genetic variability of hepatitis A virus in South America reveals heterogeneity and co-circulation during epidemic outbreaks. *J. Gen. Virol.* 82, 2647–2652.
- Cromeans, T., Sobsey, M.D., Fields, H.A., 1987. Development of a plaque assay for a cytopathic, rapidly replicating isolate of hepatitis A virus. *J. Med. Virol.* 22, 45–56.
- Domingo, E., 1996. Biological significance of viral quasispecies. *Viral Hepatitis Rev.* 2, 247–261.
- Domingo, E., Biebricher, C., Eigen, M., Holland, J.J., 2001. Quasispecies and RNA Virus Evolution: Principles and Consequences. Landes Bioscience, Austin.
- Domingo, E., Diez, J., Martínez, M.A., Hernández, J., Holguin, A., Borrego, B., Mateu, M.G., 1993. New observations on antigenic diversification of RNA viruses. Antigenic variation is not dependent on immune selection. *J. Gen. Virol.* 74, 2039–2045.
- Domingo, E., Holland, J.J., 1997. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51, 151–178.
- Domingo, E., Sabo, D., Taniguchi, T., Weissmann, C., 1978. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13, 735–744.
- Drake, J.W., 1993. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. USA* 90, 4171–4175.
- Drake, J.W., Holland, J.J., 1999. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* 96, 13910–13913.
- Eigen, M., Biebricher, C.K., 1988. Sequence space and quasispecies distribution, in: Domingo, E., Ahlquist, P., Holland, J.J. (Eds.), *RNA Genetics*, Vol. 3. CRC Press, Boca Raton, FL, pp. 211–245.
- Friedberg, E.C., Walker, G.C., Siede, W., 1995. *DNA Repair and Mutagenesis*. American Society for Microbiology, Washington, DC.
- Gavrilin, G.V., Cherkasova, E.A., Lipskaya, G.Y., Kew, O.M., Agol, V.I., 2000. Evolution of circulating wild poliovirus and of vaccine-derived poliovirus in an immunodeficient patient: a unifying model. *J. Virol.* 74, 7381–7390.
- Haydon, D.T., Woolhouse, M.E., 1998. Immune avoidance strategies in RNA viruses: fitness continuums arising from trade-offs between immunogenicity and antigenic variability. *J. Theor. Biol.* 193, 601–612.
- Holland, J.J., de La Torre, J.C., Steinhauer, D.A., 1992. RNA virus populations as quasispecies. *Curr. Top. Microbiol. Immunol.* 176, 1–20.
- Hollinger F.B., Emerson (2001). Hepatitis A virus, in: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, Lippincott Williams and Wilkins, Philadelphia PA, pp. 779–840.
- Kuge, S., Kawamura, N., Nomoto, A., 1989. Strong inclination toward transition mutation in nucleotide substitutions by poliovirus replicase. *J. Mol. Biol.* 207, 175–182.
- Kunkel, T.A., 1988. Exonucleolytic proofreading. *Cell* 53, 837–840.
- Lemon, S.M., Chao, S.F., Jansen, R.W., Binn, L.N., LeDuc, J.W., 1987. Genomic heterogeneity among human and nonhuman strains of hepatitis A virus. *J. Virol.* 61, 735–742.
- Lemon, S.M., Murphy, P.C., Shields, P.A., Ping, L.H., Feinstone, S.M., Cromeans, T., Jansen, R.W., 1991. Antigenic and genetic variation in cytopathic hepatitis A virus variants arising during persistent infection: evidence for genetic recombination. *J. Virol.* 65, 2056–2065.
- Lemon, S.M., Robertson, B.H., 1993. Current perspectives in the virology and molecular biology of hepatitis A virus. *Semin. Virol.* 4, 285–295.
- Luo, M., Rossmann, M.G., Palmenberg, A.C., 1988. Prediction of three-dimensional models for foot-and-mouth disease virus and hepatitis A virus. *Virology* 166, 503–514.
- Mullan, B., Kenny-Walsh, E., Collins, J.K., Shanahan, F., Fanning, L. J., 2001. Inferred hepatitis C virus quasispecies diversity is influenced by choice of DNA polymerase in reverse transcriptase-polymerase chain reactions. *Anal. Biochem.* 289, 137–146.
- Nainan, O.V., Brinton, M.A., Margolis, H.S., 1992. Identification of amino acids located in the antibody binding sites of human hepatitis A virus. *Virology* 191, 984–987.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University, New York.
- Ping, L.H., Lemon, S.M., 1992. Antigenic structure of human hepatitis A virus defined by analysis of escape mutants selected against murine monoclonal antibodies. *J. Virol.* 66, 2208–2216.
- Racaniello, V.R., 2001. Picornaviridae: The viruses and their replication, in: Knipe, D.M., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B., Straus, S.E. (Eds.), *Fields Virology*, Lippincott Williams & Wilkins, Philadelphia, pp. 685–722.
- Remington, K.M., Bennett, S.E., Harris, C.M., Harris, T.M., Bebenek, K., 1998. Highly mutagenic bypass synthesis by T7 RNA polymerase of site-specific benzo[a]pyrene diol epoxide-adducted template DNA. *J. Biol. Chem.* 273, 13170–13176.
- Rico-Hesse, R., Pallansch, M.A., Nottay, B.K., Kew, O.M., 1987. Geographic distribution of wild poliovirus type 1 genotypes. *Virology* 160, 311–322.
- Ripley, L.S., 1990. Frameshift mutation: determinants of specificity. *Annu. Rev. Genet.* 24, 189–213.
- Robertson, B.H., Jansen, R.W., Khanna, B., Totsuka, A., Nainan, O.V., Siegl, G., Widell, A., Margolis, H.S., Isomura, S., Ito, K., Ishiku, T., Moritsugu, Y., Lemon, S.M., 1992. Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. *J. Gen. Virol.* 73, 1365–1377.

- Sánchez, G., Bosch, A., Pinto, R.M., 2003. Genome variability and capsid structural constraints of hepatitis A virus. *J. Virol.* 77, 452–459.
- Sánchez, G., Pinto, R.M., Vanaclocha, H., Bosch, A., 2002. Molecular characterization of hepatitis A virus isolates from a transcontinental shellfish-borne outbreak. *J. Clin. Microbiol.* 40, 4148–4155.
- Schneider, W.L., Roossinck, M.J., 2000. Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. *J. Virol.* 74, 3130–3134.
- Simmonds, P., Smith, D.B., 1999. Structural constraints on RNA virus evolution. *J. Virol.* 73, 5787–5794.
- Sobrinho, F., Dávila, M., Ortin, J., Domingo, E., 1983. Multiple genetic variants arise in the course of replication of foot-and-mouth disease virus in cell culture. *Virology* 128, 310–318.
- Steinhauer, D.A., Domingo, E., Holland, J.J., 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 122, 281–288.
- Taylor, M.B., 1997. Molecular epidemiology of South African strains of hepatitis A virus: 1982–1996. *J. Med. Virol.* 51, 273–279.
- van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., Mc Geoch, D.J., Pringle, C.R., Wickner, R.B., 2000. *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses.* Academic Press, San Diego.
- Villanueva, N., Dávila, M., Ortin, J., Domingo, E., 1983. Molecular cloning of cDNA from foot-and-mouth disease virus C1-Santa Pau (C-S8). Sequence of protein-VP1-coding segment. *Gene* 23, 185–194.
- Williams, K.J., Loeb, L.A., 1992. Retroviral reverse transcriptases: error frequencies and mutagenesis. *Curr. Top. Microbiol. Immunol.* 176, 165–180.
- Zimmern, D., 1988. Evolution of RNA viruses, in: Domingo, E., Holland, J.J., Ahlquist, P. (Eds.), *RNA Genetics*, Vol. 2, CRC Press, FL, pp. 211–240.