

PALINDROMIC-NUCLEOTIDE SUBSTITUTIONS (PNS) OF HEPATITIS C VIRUS GENOTYPES 1 AND 5a FROM SOUTH AFRICA

^{1,2*}N. Prabdial-Sing, ³M. Giangaspero, ^{1,2}A. J. Puren, ⁴J. Mahlangu, ⁴P. Barrow and ^{5,}
⁶S.M. Bowyer

¹Specialized Molecular Diagnostics, Hepatitis Unit, National Institute for Communicable Diseases (NICD), of the National Health Laboratory Services (NHLS), Private bag 4, Sandringham, 2131 and ²Division of Virology and Communicable Diseases Surveillance, School of Pathology, University of Witwatersrand, Johannesburg, South Africa, ³Veterinary Microbiology, School of Veterinary Medicine, Faculty of Agriculture, Iwate University, Morioka, Japan, ⁴Charlotte Maxeke Hospital, Johannesburg, South Africa, ^{5a}Department of Medical Virology, University of Pretoria, Pretoria, South Africa, ⁶Tshwane Academic Division, National Health Laboratory Services (NHLS), Pretoria, South Africa

*Corresponding author. Tel.: +27 11 3866347; fax. +27 11 3866411.
Email address: niship@nicd.ac.za

Summary

The HCV stem-loop subdomains III-a, -b and -c have been shown to reflect the characteristics of the virus and identify isolates by genus, genotype and subtype. The aim of this study was to investigate the genotype-specific PNS within the 5'UTR of prevalent HCV genotypes (1 and 5a) found in South Africa. The genotype 5a (N=35) and genotype 1 sequences (N=20) were from patients presenting with liver disease or haemophilia, respectively. PNS HCV typing characteristics, defined previously, were observed. The PNS method differentiated subtypes 1a and 1c from subtype 1b by the base change at nucleotide position 243. A lack of structural data from the variable loci V1 of the 5'UTR did not allow us to further differentiate the subtypes of 1. A nucleotide change from a thymine (T) to a cytosine (C) at position 183 was found among genotype 5a sequences. This mutation changed the stable U-AA bond to a Y AA bulge at base-pair position 32. There was an insertion of a single adenine (A) at position 207. At present PNS analysis is

labour intensive but, with development of further software to aid the computer analysis, it has the potential to provide a rapid, reliable alternative to phylogenetic analysis.

Key words: palindrome, HCV, stem-loop, genotype 5a

Hepatitis C virus (HCV) translation initiation is mediated via an RNA structure called the internal ribosome entry site (IRES) which encompasses four highly conserved secondary structure domains in the 5' untranslated region (UTR) of the virus (Collier et al., 2002). Highly conserved palindrome-like binding patterns (PNS) form stable stem regions which maintain the basic 3D secondary structure (Harasawa and Giangaspero, 1998). These patterns have been found in the stem-loop structures of pestiviruses (Harasawa and Giangaspero, 1998) and HCV (Giangaspero et al., 2008) and are the result of viral evolution over time. PNS patterns provide information on the base-pair signatures to which the primary sequence conforms in the stem-loop. The stem-loop structures of domains III-a, -b and -c (Piron et al., 2005) reflects the characteristics of the HCV isolate, by genus, genotype and subtype (Giangaspero et al., 2008). RNA replication and infectivity are affected when mutations are introduced into the stem, but not the loop, regions of the four 5'UTR secondary structure domains (Friebe and Bartenschlager, 2009), indicating that stem regions are required for IRES activity and, therefore, viral replication.

Numbering the base-pairs from the start of the secondary structure, the genotype and subtype PNS characteristics of genotype 1 (Figure 1 and 2; Giangaspero et al., 2008) are

as follows: *Genotype characteristic*: A-U in position 28 of domain III; GA bulge in position 4 of the variable region 1 (V1). *Subtype characteristic*: A (subtype 1a or 1c) or G (subtype 1b) at position 10 of domain III; AA (subtype 1a or 1c) or AG (subtype 1b) bulge at position 1, U*G (1b or 1c) or U-A (1a) in position 2 and GG (1c) or GA (1a or 1b) bulge in position 7 of V1. From previous alignments, genotype 5a specimens lack the initial 59-95 nucleotides of the 5'UTR and do not have a V1 structure. Two PNS patterns in domain III are specific for this genotype. *Genotype characteristic*: The presence of both U*G in position 24 and G-C in position 28 of domain III.

The aim of this study was to investigate these genotype specific palindromic substitutions within the 5'UTR of the most prevalent HCV genotypes (1 and 5a) in South Africa. The study was retrospective and approved by the ethics committee of the University of the Witwatersrand, Johannesburg, South Africa (WITS HREC M051114), and was therefore performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. The study described the secondary structure in the 5'UTR sequences of specimens previously genotyped as 1 and 5a in two patient cohorts attending a Johannesburg hospital (Prabdi-Sing et al., 2008). Sixty-one sequences (genotype 1, N=26; genotype 5, N=35) were formatted and aligned using ClustalW in BioEdit, version 7.0 (Hall, 1997) over the informative regions (179 nucleotides) of the 5'UTR as described by Giangaspero et al., 2008. Reference strains used were: M62321 (1a), D50480 (1b), D14853 (1c), D17763 (3a), D49374 (3b), Y12083 (6a), D50409 (2c), D10988 (2b), D00944 (2a), U33430 (5a), AM502711, AM502710, DQ164748- DQ164751 (5a), AY033769 and L28057 (5a). Sequences were then analysed for PNS patterns in the predicted secondary

structures according to the algorithm of Zuker and Stiegler (1981) using the Genetyx-mac Software program (version 10.1 Software Development Co., Ltd., Tokyo, Japan) as described previously (Giangaspero et al., 2008).

The PNS method was 100% concordant with the sequencing method (Table 1a and b) at the genotype level. The characteristic genotype PNS patterns identified by Giangaspero et al., 2008 for genotype 1 and genotype 5a were observed in this study. Tables 1a and b indicate relevant sections of the primary nucleotide alignment of the genotype 1(141-252, -201 to -89, according to Choo et al., 1991) and 5a sequences (102-227, -240 to -114, according to Choo et al., 1991), respectively. The shaded regions show important base pair genotype signatures in the stem-loop sequences. Row 1 of the table refers to the PNS pattern number (Fig. 1b) while row 2 gives the actual nucleotide number in the primary sequence data. The RNA folding is a result of base-pairing of upstream nucleotides with complementary nucleotides further downstream. For example, PNS pattern 28 pairs nucleotides 179 and 220 (Table 1a).

Figure 1 shows the typical pattern of folding in the 5'UTR region and shows the position of the structural regions V1 and domain III described in this study. Figure 2 shows the predicted PNS patterns for genotype 1 (left) and genotype 5 (right) determined from the alignments in this study. Stem-loop III is formed from the folding of nucleotides 141-252 to form 47 PNS sites (Fig. 2). The numbering of the nucleotides and the base pair positions are according to Giangaspero et al., 2008. Table 1 aligns sequence data of this study, together with appropriate reference sequences, over relevant sections of domains II

and III to show all variation found. Highlighted regions in the tables are definitive PNS signatures. The arrows indicate sites of variation not reported previously but observed in this study.

Nucleotide 243 does not base pair and forms a bulge at PNS position 10 and is an identified PNS characteristic for subtyping genotype 1 into subtypes 1a or 1c (Giangaspero et al., 2008). The 5'UTR sequencing results corroborate that the subtype change from A to guanosine (G) correlates with a switch from genotype 1a or 1c to 1b in all but one specimen (1108/H). Sequencing in the NS5B region, considered the gold standard for subtyping, called specimen 1485/H as subtype 1a contrary to the two 5'UTR analyses.

There was one variable site in the genotype 1 data (Table 1a) at nucleotide 204 where a nucleotide base change from C to A was seen in specimens 420/LD1, 1958/H, 1956/H, 1108/H, 1485/H and 2763/H. Alternatively, there was a C to T change as seen in the genotype 1c reference sequence, D14853, and 6641/H.

The specific combination of PNS positions to define genotype 5a can be seen in the alignment (Table 1b) at position 24 (U*G) and position 28 (G-C; Giangaspero et al., 2008). There was also a T to C mutation at position 183 in one of isolates studied, 2577/LD1 as well as three European reference specimens, two from France (AM502710, AM502711) and the other from Belgium (DQ164750, Table 1b). In three of the specimens (3065/LD1, 2014/LD1 and 918/LD1) and the Brazilian isolate (AY033769)

there was a single A insertion at position 207 (Table 1b). The genotype 5a specimen sequences with these changes did not cluster significantly differently in the phylogeny, although 2577/LD1 was an outlier to the main genotype 5a clade in the 5'UTR region (Prabdi-Sing et al., 2008).

The data partitioned into genotypes 1 and 5a identically by PNS analysis based on changes in the secondary structure when compared to both 5'UTR and NS5B phylogenetic analysis of the primary structure. Specimens infected with these two genotypes were chosen for the PNS analysis as they are the predominant genotypes found in HCV-infected individuals in SA. The phylogenetic analyses were published previously (Prabdi-Sing et al., 2008). Three more variable sites were found, at positions 106, 183 and 207, in genotype 5a sequences. The nucleotide change at position 106 lies in a region of the 5'UTR which is unpaired in genotype 5a whereas the change at position 183 lies in the stem region between the loops of domain IIIa and IIIb. The T183C change does not affect its complementary nucleotides AA at position 215 and 216. However, the stable U-AA bond is predicted to change to a Y: AA bulge at position 32 (Fig. 1b). The insertion at position 207, seen in some of the genotype 5a sequences, was predicted to form part of the domain IIIb loop. The insertion of a single A at position 207 and CA at position 197 at base pair 41 and 47 in the structural alignment, respectively, were among the unique PNS characteristics described for subtypes 6a and 6b (Giangaspero et al., 2008). They are predicted to form a higher secondary loop structure but do not change the stem. This same insertion at position 41 in the genotype 5a sequences was not predicted to elongate the loop, but should be noted as specific changes are important for primer design.

Functionally, since this occurs in the loop structure, little or no effect on viral viability and replication is expected.

The PNS method differentiated subtypes 1a and 1c from subtype 1b by the base pair change at nucleotide position 243, observed in PNS position 10. However, since A is also definitive for subtype 1c, subtype 1a and 1c could not be discerned. The availability of sequence data from the nucleotides 1 to 80 at the amino terminal would have cleared this ambiguity as the PNS characteristic for genotype 1c includes a unique site (GG bulge at position 7) in the V1 structure (Giangaspero et al., 2008). PCR amplification of the amino terminal of the 5'UTR was attempted using the rapid amplification of cDNA ends (RACE) method. Although high viral load specimens were used (range 769161-4664180 IU/ml), the RACE method failed as the initial concentration of extracted viral RNA was too low (0.2-1.6ng) for the process of d(A) tailing and amplification, where 0.2µg of RNA is required. Better design of primers (Giangaspero et al., 2008) and optimisation of a PCR to sequence the first 60-80 nucleotides of genotype 1 is necessary for effective subtyping. All specimens without the subtype 1b specific G at 243 were subtyped previously as 1a, but two others typed by PNS as subtype 1b were typed as subtype 1a by phylogenetic analysis. Interestingly, both of these have an A at position 204. This might be expected to affect the phylogenetic analysis in the conserved 5'UTR but 1485/H was typed in the NS5B region. The nucleotide change seen among the genotype 1 specimens at position 204 was recognised by Giangaspero et al., 2008 and made no structural change to the IIIb loop.

The PNS method partitioned the genotype 1 and 5a specimens, successfully, and provided the secondary structure associated with primary sequence data from SA. The latter could be used to design targets for siRNA therapies (Prabhu et al., 2006) and reverse genetics (Friebe and Bartenschlager, 2009).

For clearer subtyping of genotype 1, the first 80 nucleotides of the 5'UTR should be sequenced for the full application of the PNS method. At present PNS analysis is labour intensive, but the method would have potential were easily accessible computer programmes available. When fully automated, it could provide a reliable alternative for rapid typing and subtyping of HCV in the conserved 5'UTR which is a very robust region for PCR of the less characterised genotypes. At the same time, because of the importance of the 5'UTR 3D structure, structural changes are more reliable than primary sequence changes.

References:

Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G., Houghton, M., 1991. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A* 88, 2451-5.

Collier, A.J., Gallego, J., Roscoe, K., Cole, P.T., Harris, S.J., Harrison, G.P., Aboul-ela, F., Varani, G., Walker, S., 2002. A conserved RNA structure within the HCV IRES Eif3-binding site. *Nat Structural Biol* 9: 375-380.

Fraser, C.S., Doudna, J.A., 2007. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat Rev Microbiol.* 5, 29-38.

Friebe, P., Bartenschlager, R., 2009. Role of RNA structures in genome terminal sequences of the hepatitis C virus for replication and assembly. *J Virol.* 83, 11989-5.

Giagaspero, M., Harasawa, R., Zanetti, A., 2008. Taxonomy of genus Hepacivirus. Application of palindromic nucleotide substitutions for the determination of genotypes of human hepatitis C virus species. *J Virol Methods.* 153, 280-99.

Hall, T., 1997. "BioEdit." <http://www.mbio.ncsu.edu>. Last accessed 2009/02/22, 14:41PM.

Harasawa, R., Giangaspero M., 1998. A novel method for pestivirus genotyping based on palindromic nucleotide substitutions in the 5'-untranslated region. *J Virol Methods*. 70, 225-30.

Piron, M., Beguiristain, N., Nadal, A., Martinez-Salas, E., Gomez, J., 2005. Characterizing the function and structural organization of the 5' tRNA-like motif within the hepatitis C virus quasispecies. *Nucleic Acids Res*. 33, 1487-502.

Prabdial-Sing, N., Puren, A.J., Mahlangu, J., Barrow, P., Bowyer, S.M., 2008. Hepatitis C virus genotypes in two different patient cohorts in Johannesburg, South Africa. *Arch Virol*. 153, 2049-58.

Prabhu, R., Garry, R.F., Dash, S., 2006. Small interfering RNA targeted to stem-loop II of the 5' untranslated region effectively inhibits expression of six HCV genotypes. *Virology*. 3, 100.

Zuker, M., Stiegler, P., 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9, 133-48.

Table 1a. Nucleotide alignment (nucleotides 141-252) of sequences in this study with HCV genotype I reference sequences from GenBank, M62321 (1a), D50480 (1b) and D14853 (1c). The shaded regions (nucleotides 179 plus 220 and 243) highlight structural information at base-pair 28 and 10, respectively, which are specific PNS signatures for genotype 1. The arrow marks the previously unreported variation at nucleotide 204 which was observed in this study.

↓

Genotype	1 2 3 25 26 27 28 29 30 31 39 40 41 42 41 40 39 38 37 28 27 26 25 24 23 22 21 13 12 11 10 9 8 7 6 5 4 3 2 1																												PNS vs sequencing																
	141	142	143	176	177	178	179	180	181	182	201	202	203	204	205	206	207	208	209	220	221	222	223	224	225	226	227	228	229	242	243	244	245	246	247	248	249	250	251	252	5'UTR	NS5B	PNS		
M62321(1a)	la	T	A	G	G	C	C	A	G	G	A	G	A	T	C	A	A	C	C	C	T	G	G	A	G	A	T	T	C	G	C	A	A	G	A	C	T	G	C	T	A	reference 1a			
420/LD1	la	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	la or lc	la	la or lc
1958/H	la	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	la or lc	la	la or lc
1956/H	la	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	la or lc	la	la or lc
307/H	la	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	la or lc	la	la or lc
9049/LD1	la	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	la or lc	la	la or lc
D50480(1b)	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	reference 1b		
1108/H	? 1b	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	outlier la or lc	ND	1b
1485/H	? 1a	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	la	1b
2763/H	1b	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
4434/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
967/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
2771/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
2585/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
5307/LD2	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
3894/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
6373/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
1229/LD1	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	1b	1b
2205/LD1	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	1b	1b
356/LD1	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	1b	1b
1109/H	1b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	1b	1b
6641/H	1b	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1b	ND	1b
D14853(1c)	1c	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	reference 1c		

Fig.1. Secondary structure of the 5'UTR of the HCV genome (Fraser and Doudna, 2007). The variable region (V1) used by Giangaspero et al. 2008 to confirm some of the subtypes is boxed. Genotype 5a specimens start between nucleotides 59—95 of domain II. The definitive double AA at nucleotides 106 and 107 which define genotype 5 in the primary sequence would replace the UG, marked by the arrows, in domain II. The numbering is according to Giangaspero et al., 2008. Row 1 indicates the base pairings numbered from the bottom of the secondary structure and row 2 the nucleotide position as described in the text. The right hand 3 columns compare the results of PNS typing with 5'UTR and NS5B sequencing. Inconsistencies are highlighted in grey.

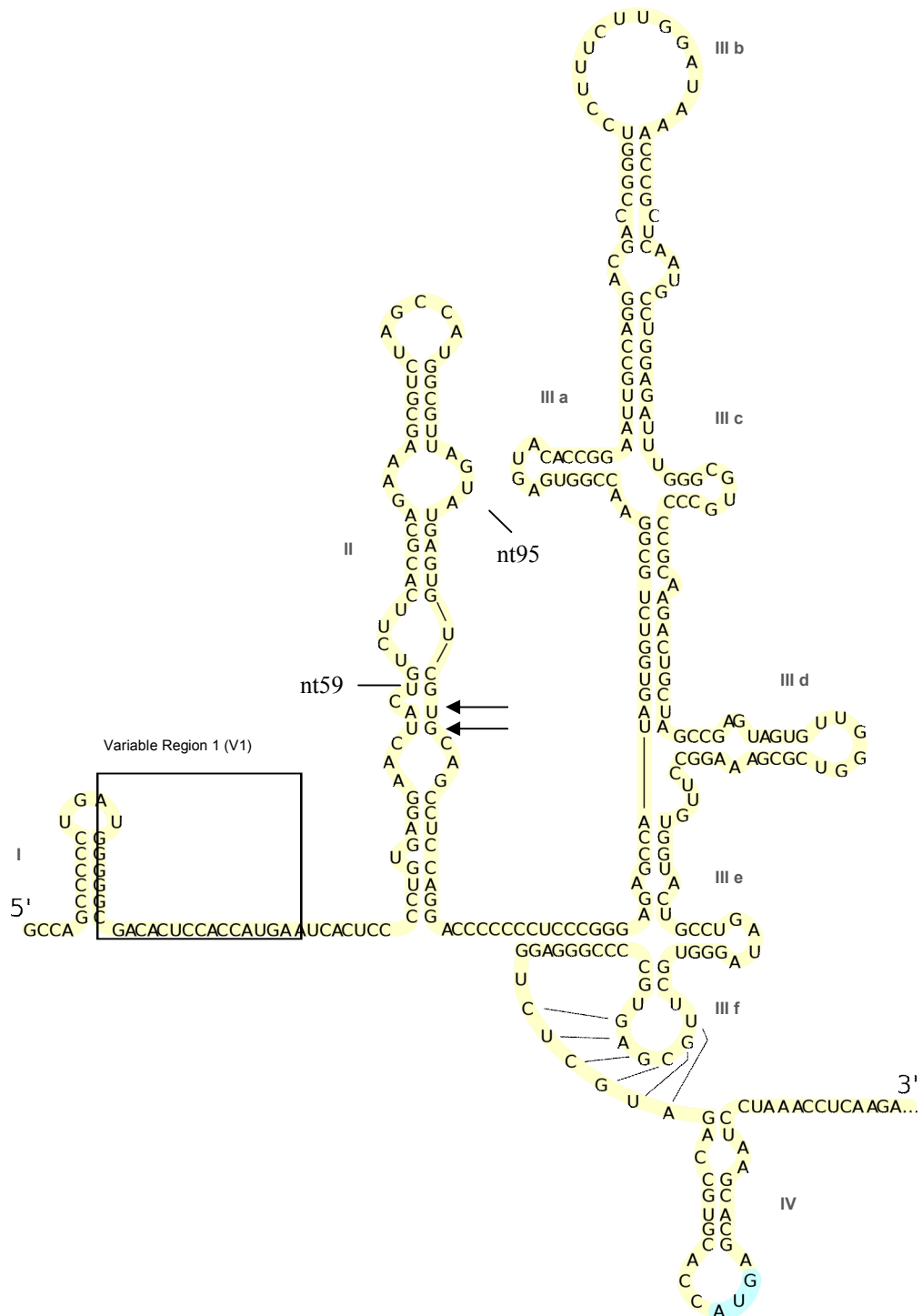


Fig.2. Secondary structure of domain IIIa-c predicted for genotype 1 (left) and 5a (right) specimens sequenced in this study. The 3D structures show the variable loci definitive of the genus *Hepacivirus* and the two predominant genotypes in South Africa and their subtypes. Base pairings characteristic to the genus (PNS genus specific) are shown in bold; those distinguishing the HCV genotypes (PNS genotype specific) are represented in bold and underlined, while subtype-specific PNS are shown in bold and italic. Watson–Crick base pairings are indicated by a dash (–); tolerated pairings in secondary structure are indicated by an asterisk (*); interchangeable base pairings are indicated by a colon (:). R =A or G; W=A or U; Y = C or U; B= C or G or U; H=A or C or U.

