Molecular characterization analysis of the outer protein layer (VP7) from human rotavirus A genotype G1 isolate identified in Iran: implications for vaccine development

Zabihollah Shoja[1], Maria Tagliamonte[2], Somayeh Jalilvand[1], Abed-Ali Ziaee[3], Talat Mokhtari-Azad[1], Rasool Hamkar[1], Shohreh Shahmahmoodi[1], Farhad Rezaei[1], MariaLina Tornesello[2], Franco M. Buonaguro[2], Rakhshandeh Nategh[1], Luigi Buonaguro[2]

[1]School of Public Health, Tehran University of Medical Sciences, Tehran, Iran;
[2]Molecular Biology and Viral Oncogenesis Unit, Istituto Nazionale Tumori "Fond. G. Pascale", Naples, Italy;
[3]Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

SUMMARY

The full open reading frame of the outer protein layer VP7 from an isolate of human rotavirus identified in 2010 in an Iranian child admitted to hospital with gastroenteritis was amplified from a clinical stool specimen and subjected to molecular characterization. Genetic and phylogenetic analyses indicated that the analyzed gene falls into the G1 genotype forming a sub-cluster with sequences recently identified in Iran and geographically distant countries. Such results were confirmed by protein sequence alignment, showing a highly conserved "G1-like" amino acid sequence pattern within the known three main immunodominant regions.
These results are extremely relevant in a perspective of vaccine development. Indeed, the present study confirms that the A group G1 genotype is the most prevalent Rotavirus circulating in Iran and supports the development of G1 genotype-based rotavirus vaccine for this country.

KEY WORDS: Rotavirus, VP7, Iran, G1 genotype, Phylogenetic analysis.

Received April 05, 2012 *Accepted* July 20, 2012

INTRODUCTION

Rotaviruses represent one genus of the Reoviridae family (Bishop *et al.,* 1973; Bishop *et al.,* 1974; Flewett *et al.*, 1973; Kapikian *et al.,* 1974). Rotavirus particles are non-enveloped icosahedral virions, composed of triple layer capsids enclosing 11 segments of double-stranded RNA. Each RNA segment encodes for single structural (VP1 to VP7) or nonstructural proteins (NSP1 to NSP4) with the exception of segment 11, which encodes for two proteins (NSP5 and NSP6). The inner nucleocapsid layer is made by the VP2 core protein associated with VP1, VP3 and the viral RNA segments. The nucleocapsid is surrounded by a protein capsid layer made of VP6 proteins, forming the so-called double layer subviral particles. A third and most external protein capsid layer is made of VP7 with protruding VP4 proteins, resulting in the triple-layered mature virion (Bishop, 1996; Estes and Cohen, 1989).
According to the serological cross-reactivity of the middle protein VP6, seven serogroups (A-G) have been firmly established (Bishop, 1996), with group A being the major causative agent of severe diarrhea in infants and young children worldwide (Estes and Kapikian, 2007). Within the serogroup A, VP7 and VP4 outer capsid proteins are relevant for a binary classification system, similar to Influenza virus, based on the glycoprotein VP7 (G) and the protease-cleaved pro-

Corresponding author
Luigi Buonaguro
Lab. of Molecular Biology and Viral Oncogenesis
Istituto Nazionale Tumori "Fond. G. Pascale"
20131 Naples, Italy
E-mail: irccsvir@unina.it

tein VP4 (P) types (Estes and Kapikian, 2007). Indeed, 27 G genotypes (corresponding to 15 G serotypes) and 35 P genotypes (corresponding to 14 P serotypes) have been reported so far (Estes and Kapikian, 2007; Matthijnssens *et al.,* 2011; Matthijnssens *et al.,* 2008; Schumann *et al.,* 2009; Solberg *et al.,* 2009; Trojnar *et al.,* 2009). Global molecular epidemiologic surveys have identified G1P[8], G2P[4], G3P[8], G4P[8], and G9P[8] as the most common G-P genotypes causing diarrhea in humans worldwide (Estes and Kapikian, 2007; Glass *et al.,* 2006).

Management of Rotavirus infection is considered a high priority for the healthcare system in Iran (Glass *et al.,* 2008), where seroepidemiological studies have shown that rotavirus was responsible for 15.3%-67.6% of acute diarrhea occurring in <5 years old children (Emamghorashi *et al.,* 2008; Hamkar *et al.,* 2010; Kordidarian *et al.,* 2007; Malek *et al.,* 2010; Modarres *et al.,* 2008; Sadeghian *et al.,* 2010; Samarbafzadeh *et al.,* 2005). The most commonly detected genotype in Iran was the G1P[8] (Farahtaj *et al.,* 2007; Khalili *et al.,* 2004; Modaress *et al.*, 2011), although in a single study it was reported a prevalence of the G4 genotype (Esteghamati *et al.*, 2009).

The outer protein layer VP7, together with the VP4, represents the major neutralization antigen of rotavirus (Greenberg *et al.,* 1983; Hoshino *et al.,* 1988; Offit *et al.,* 1986; Taniguchi *et al.,* 1988), inducing homotypic as well as heterotypic protection in animal studies (Bishop *et al.*, 1986; Hoshino *et al.,* 1988; Offit *et al.,* 1986; Wyatt *et al.,* 1979). Therefore, the last generation of experimental rotavirus vaccines, based on recombinant or subunit proteins (including Virus-Like Particles), are focused on VP7 and VP4 because of their importance in inducing both passive and active cross-protective immunity in animal models (Crawford *et al.,* 1999; McNeal *et al.,* 1998; Parez *et al.,* 2006).

In this framework, molecular studies on genetic and phenetic evolution of VP7 proteins in clinical isolates are extremely relevant not only for molecular epidemiological reasons but also for vaccine development.

In the present study, we report the full genetic, phylogenetic and phenetic characterization of human rotavirus VP7 identified in a clinical stool specimen from an Iranian child admitted to hospital in 2010 with gastroenteritis.

MATERIALS AND METHODS

Patient and specimens

Children under five years of age with gastroenteritis were admitted to Mofid Pediatric Hospital in Tehran City and, after confirmation of rotavirus infection by ELISA, 10 stool specimens were collected and transferred to the Virology Department in Tehran University of Medical Sciences for molecular analysis.

Specimen preparation and RNA extraction

Stool specimens were suspended in phosphate-buffered saline (PBS), centrifuged at 1500 g for 20 min (WHO, 2009) and double stranded RNA was extracted from the clarified supernatant by QIAamp Viral RNA Mini Kit (Qiagen).

Detection and confirmation of G1 genotype by RT-PCR

Amplification and G genotyping of the VP7 gene was performed using RT-PCR and a nested multiplex PCR with type-specific primers according to the standard rotavirus detection and characterization method (Gouvea *et al.,* 1990; WHO, 2009). Rotavirus genomic double-stranded RNA was denatured by heating at 97°C for 5 min and quickly quenched in a dry ice-ethanol bath. RT-PCR was performed with aliquot of the dsRNA (15-20 ng), 10 µM each of specific primers (Table 1), 2X Reaction Mix (a buffer containing 0.4 mM of each dNTP, 2.4 mM $MgSO_4$) (invitrogen) and 1U SuperScript™ III RT/ Platinum® Taq High Fidelity Enzyme Mix (Invitrogen) to a final volume 50 microliters. The full-length (1062bp) cDNA synthesis was obtained by one cycle at 50°C for 30 minutes and 94°C for 2 minutes, followed by PCR amplification with 40 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 1 minute and 1 cycle at 68°C for 5 minutes. The second round was performed on 5 µl of the first amplification at the same experimental conditions as for the first round, except for the annealing temperature at 45°C, using a mixture of type-specific primers (Table 1).

Amplification and cloning of rotavirus gene coding for VP7

The complete open reading frame (ORF) of VP7 of rotavirus G1 genotype (981bp) was amplified

TABLE 1 - *Specific primers used in the RT-PCR and subsequent nested multiplex PCR for genotyping of rotavirus VP7 ORF or for amplification of full-length VP7 ORF.*

Primer	*Sequence 5'to 3'*	*Sense*	*Position*	*Result when paired to End 9*
End 9	GGT CAC ATC ATA CAA TTC TAA TCTAAG	-	1062-1036	
Beg 9	GGC TTT AAA AGA GAG AAT TTC CGT CTG G	+	1-28	RT and 1st round
aBT1	CAA GTA CTC AAA TCA ATG ATG G	+	314-335	G1 genotyping in the 2nd round
aCT1	CAA TGA TAT TAA CAC ATT TTC TGT G	+	411-435	G2 genotyping in the 2nd round
G3-Aust	ACG AAC TCA ACA CGA GAR G	+	250-269	G3 genotyping in the 2nd round
aDT4	CGT TTC TGG TGA GGA GTT G	+	480-498	G4 genotyping in the 2nd round
aAT8	GTC ACA CCA TTT GTA AAT TCG	+	178-198	G8 genotyping in the 2nd round
mG9	CTT GAT GTG ACT AYA AAT AC	+	757-776	G9 genotyping in the 2nd round
Primer	*Sequence 5'to 3'*	*Sense*	*Position*	*Expected result*
VP7-F VP7-R	ATG TAT GGT ATT GAA TAT ACC CTA TAC TCT ATA ATA AAA AGCT	+ -	49-69 1008-1029	Amplification of Full VP7 ORF

by RT-PCR using the primer pair VP7-F and VP7-R (Table 1), using the same experimental conditions described above. The full ORF VP7 was cloned by blunt-end ligation into the Klenow filled-in *Nco*I site of the pBiEx-3 dual-host expression vector (Novagen). The recombinant pBiEx-3 was transformed into competent *E. coli* JM83 strain, ampicillin-resistant colonies were screened by colony hybridization and cloning of the whole VP7 ORF was confirmed by restriction enzyme digestion.

Genetic and phylogenetic analysis

The nucleotide sequence of the cloned VP7 ORF was obtained by sequencing with ABI Prism 310 Automatic Sequencer (Applied Biosystem). Our sequence was aligned with reference VP7 sequences of different G genotypes available in GeneBank. Multiple sequence alignment was performed with BioEdit software using the Clustal W method (Thompson *et al.*, 1994) and identity matrices for both nucleotide and amino acid sequences were generated. The phylogenetic analysis was constructed using Maximum Likelihood method (MEGA 5.05) (Jukes and Cantor, 1969; Tamura *et al.*, 2011). The tree was measured by the bootstrap method with 1000 replicates.

The accession numbers of reference VP7 sequences used in this study were as follows: GU598246- GQ996881- GQ996876- GQ996880- GQ996879- GQ996878- GQ996877- AY631052- AY631050- GU598247- AY631049- AY631053- AY631051- AY098670- AF260946- AY635048- AF260944- EF088837- GU979204- AF501579- FJ152132- AF260945- EF088838- EF088836- EF088835- GU979202- HQ738582- FJ811903- GQ996894- AF508733- GQ996893- GQ996895- GQ996892- GU598245- AF260956- GU979198- GQ996896- FJ747619- AF260957- AF260958- U04350- GU393006- EF088832- EF088831- AB527008- HQ738583- GU598248- GQ229048- AF254139- EU708602- AF450294- AF161820- FJ598040- FJ598041- AF170837- GU979200- DQ015685- DQ015686- EF077484- DQ515961- EF159576- EF159575- EF672588- AF207063- AJ487832- DQ122400- AF532202- EF672609- AF207062- AF421183- AJ487833- AF143689- HM136995- FJ861661- GQ414545- AF143690- AF529871- AF529872- GQ869843- AF529869- GQ869842- AF529870- AB527030- AB527031- AB527032- AB527033- AB527035 GQ996923- AY629560- GQ869841- AF260959- AB527034- AB527029- AB527018- GQ996921- AF501580- DQ207390- AY630924- AJ605303- GQ996924- GQ996922- GQ996920- HQ445973- HQ445974-

FJ152110- FJ747630- FJ152121- FJ747618- GQ996935- GQ996932- AY206861- AF508734- GQ996934- GQ996933- GQ996931

Evaluation of silent/ non-silent nucleotide changes

The VP7 nucleotide sequence obtained in the present study was compared with indicated reference sequences for silent and non-silent mutations by the Highlighter tool and the Synonymous Non-synonymous Analysis Program (SNAP) [www.hiv.lanl.gov] (Korber, 2000). Highlighter software enables codon-aligned nucleotide sequences to be compared with a query master sequence and highlights silent and non-silent mutations with color-code rule. Moreover, it generates a phylogenetic tree based on the pattern of mutations. In parallel, the SNAP software calculates synonymous and non-synonymous substitution rates based on a set of codon-aligned nucleotide sequences. Synonymous and non-synonymous analyses were also carried out using the Maximum Likelihood method and HyPhy software package (Pond *et al.*, 2005). Evolutionary analyses were conducted in MEGA5 (Tamura *et al.,* 2011).

Protein structure analysis

Protein structures were analyzed with the ICM Browser Pro by Molsoft using as prototype the VP7 structure revealed by the X-ray structure (Protein DataBase code 3FMG) (Aoki *et al.*, 2009).

Nucleotide sequence accession numbers

The Iranian VP7 gene sequence described in this study was deposited in the GenBank database and given accession number JF719063.

RESULTS

Amplification and cloning of VP7 ORF

Stool specimens (n. 10) positive for rotavirus infection, as confirmed by ELISA, were screened for genotype classification using RT-PCR and a nested multiplex PCR with type-specific primers, according to the standard rotavirus detection and characterization method (Gouvea *et al.,* 1990; WHO, 2009). Among the 5 samples identified as G1 genotype (Figure 1A), the full VP7 ORF was obtained from one sample (designated *Iran/2010*) using a specific primer pair designed by us and

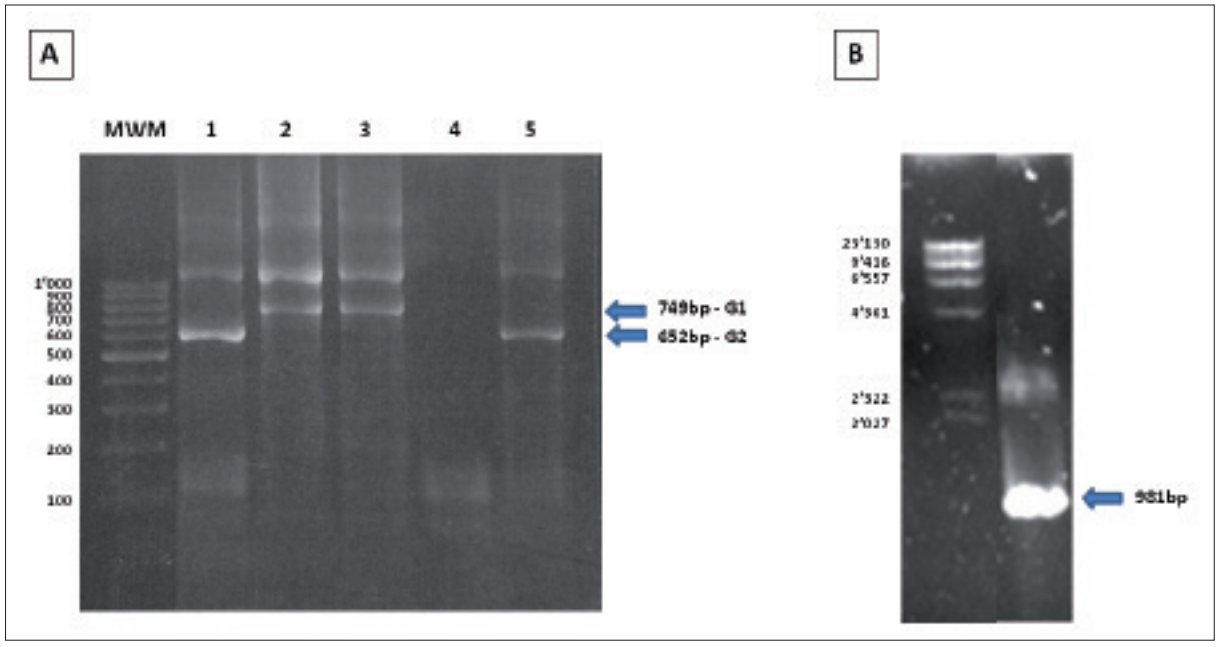

FIGURE 1 - *Genotype screening by multiplex PCR. (A) Rotavirus was genotyped by nested multiplex PCR specific for VP7 gene. Second round amplification identified a band of 749bp in G1 genotypes (lanes 2 and 3) and of 652bp in G2 genotypes (lanes 1 and 5). (B) Amplification of VP7 ORF (981bp) was obtained in one of the G1 genotypes identified by nested multiplex PCR.*

designated VP7-F and VP7-R (Table 1). The expected amplified fragment of 981bp (Figure 1B) was subsequently cloned by blunt-end ligation into the Klenow filled-in *Nco*I site of the pBiEx-3 vector.

Genetic analysis of VP7 ORF

The nucleotide sequence of full VP7 ORF (981 bp) identified in the specimen from the *Iran/2010* sample was obtained and aligned with VP7 reference sequences of G1 to G6, G8, G9 and G12 rotavirus genotypes available in the GeneBank. The similarity matrix obtained upon multiple alignment, showed that the *Iran/2010* VP7 sequence has an average 93.2% of similarity versus reference sequences of the G1 genotype and an average 71.9-75.6% of similarity versus all other genotypes ($p<0.001$) (Figure 2A). Moreover, within the G1 genotype, it showed a similarity >97% with three reference sequences identified in Germany (accession number GU979202), Russia (accession number HQ738582) and Iran (accession number FJ811903) (Figure 2B). Such results confirmed that the *Iran/2010* rotavirus isolate belongs to the G1 genotype.

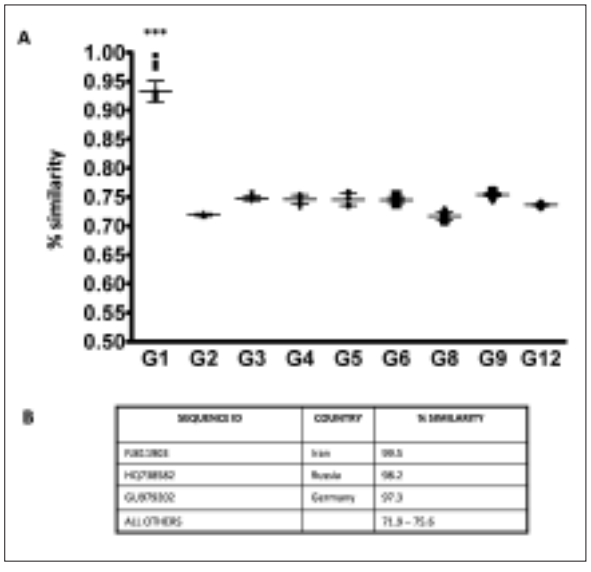

FIGURE 2 - *Analysis of nucleotide divergence. (A) The nucleotide sequence of VP7 ORF from the Iran/2010 sample was aligned to sequences of different genotypes. Percentage of similarity versus each reference sequence has been plotted according to genotype, showing average and standard deviation. (B) Percentage of similarity versus highest similar G1 sequences are shown.*

Phylogenetic analysis

The sequences' multiple alignment obtained for genetic analysis was used for phylogenetic analysis performed by MEGA 5.05 software, and the evolutionary history was inferred using the Maximum Likelihood method. The phylogenetic analyses for VP7 showed that the VP7 sequence identified in the *Iran/2010* isolate (Accession no. JF719063) belongs to the G1 cluster statistically supported by 93% bootstrap value (Figure 3). Moreover, it was confirmed that, within the G1 cluster, the 2010 Iranian sequence formed a minor cluster together with the three reference sequences with highest nucleotide homology described above. Such sub-cluster was separated by the other G1 reference sequences with a 99% bootstrap value (Figure 3). In particular, the 4 VP7 sequences including the *Iran/2010*, formed a monophyletic cluster strongly suggesting a common ancestor, although no epidemiological data to support such speculation were available to us. In addition, the phylogenetic analysis confirmed

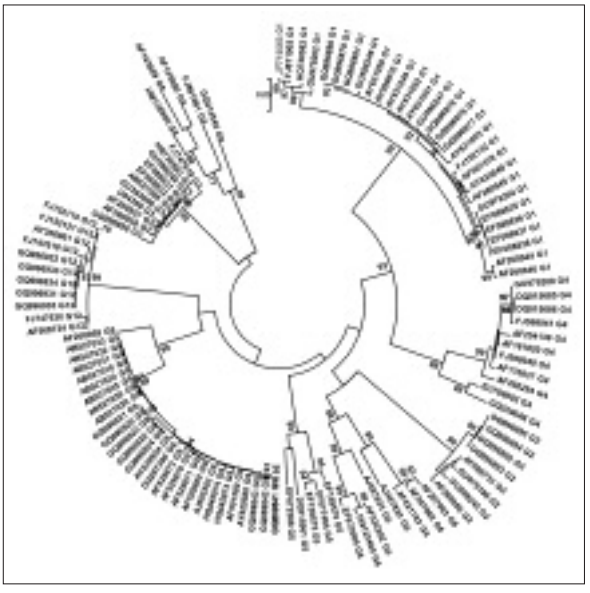

FIGURE 3 - *Maximum Likelihood phylogenetic analysis of VP7 gene. The alignment of the VP7 sequence (891 bp) from the Iran/2010 sample (Accession #JF719063) with all sequences of different genotypes was used to generate a phylogenetic tree. The clusters of different genotypes are indicated. Branch lengths were measured in the number of substitutions per site. The tree was drawn to scale with the bar at the top indicating 0.05 nucleotide substitutions per site. Significant bootstrap values are indicated. The Iran/2010 sample (Accession #JF719063) is marked in green.*

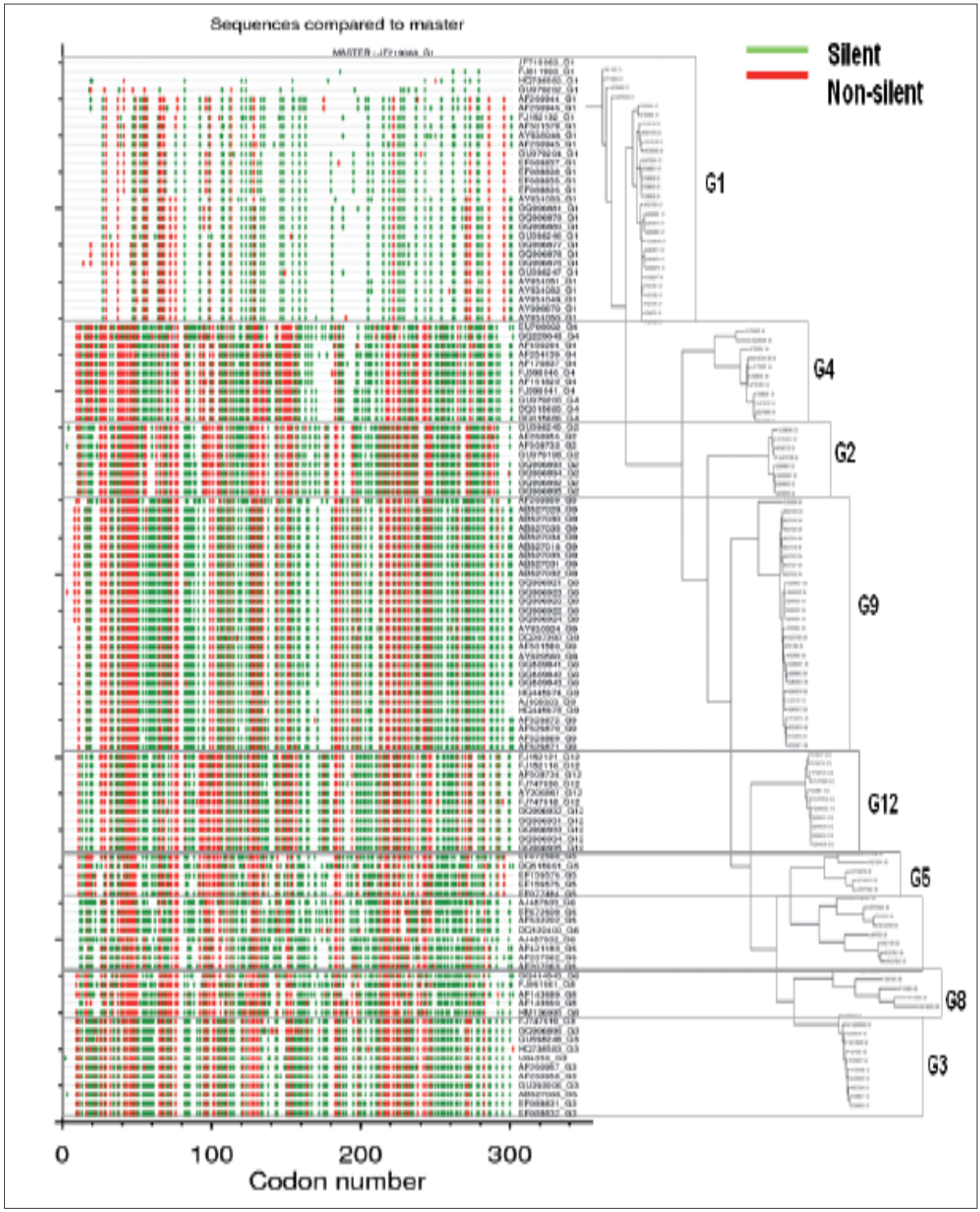

FIGURE 4 - *Analysis of silent and non-silent mutations. Codon-aligned VP7 sequences were analyzed with the Highlighter tool [www.hiv.lanl.gov] for identification of silent (green dots) and non-silent (red dots) mutations. The VP7 from the Iran/2010 sample (Accession #JF719063) has been used as master for comparing all other sequences. On the right side of the figure, a phylogenetic tree based on the pattern of silent vs non-silent mutations is shown.*

the distant relationship between our sequence and those of other G genotypes (Figure 3).

Synonymous and non-synonymous changes analysis

The genetic characterization of the *Iran/2010* sequence was further evaluated by analysis of synonymous and non-synonymous substitutions, to verify whether VP7 glycoprotein nucleotide sequences of different genotypes can also be classified according to a qualitative pattern of substitutions. For this purpose, codon-aligned sequences were compared using the Highlighter and SNAP tools [www.hiv.lanl.gov] (Korber, 2000) taking our sequence as reference master.

The results showed that sequences from the same genotype do cluster according to similar mutation patterns (Figure 4). Moreover, regardless the genotype compared to the *Iran/2010* sequence, the number of synonymous substitutions (green dots) were significantly higher than non-synonymous ones (red dots), suggesting an overall phenotypic stability of the VP7 glycoprotein sequence across genotypes (Figure 4). Such observation was further supported by the SNAP analysis, indicating that the ratio between synonymous (ds) and non-synonymous (dn) substitutions was well above 1 (dS/dN>>1). Maximum Likelihood computations of dN and dS were conducted using HyPhy software. In this case, non-synonymous rate was less than synonymous rate and confirmed dS/dN>1 that concluded using SNAP tools.

Such analysis confirmed the very high homology between the *Iran/2010* sequence and the three clustering G1 sequences described above. Indeed, the *Iran/2010* sequence showed only 4 synonymous mutations with the reference FJ811903 Iranian sequence, 13 synonymous and 2 non-synonymous mutations with the reference HQ738582 Russian sequence, 18 synonymous and 6 non-synonymous mutations with the reference GU979202 German sequence (Figure 4).

The overall picture across the genotype VP7 glycoprotein identified regions (mainly the C-terminal) where synonymous substitutions were concentrated, indicating a strong stability of the amino sequence, and regions (mainly the N-terminal) where non-synonymous substitutions were prevalent, indicating a significant phenotypic variability (Figure 4). In particular, looking more closely to the three immunodominant re-

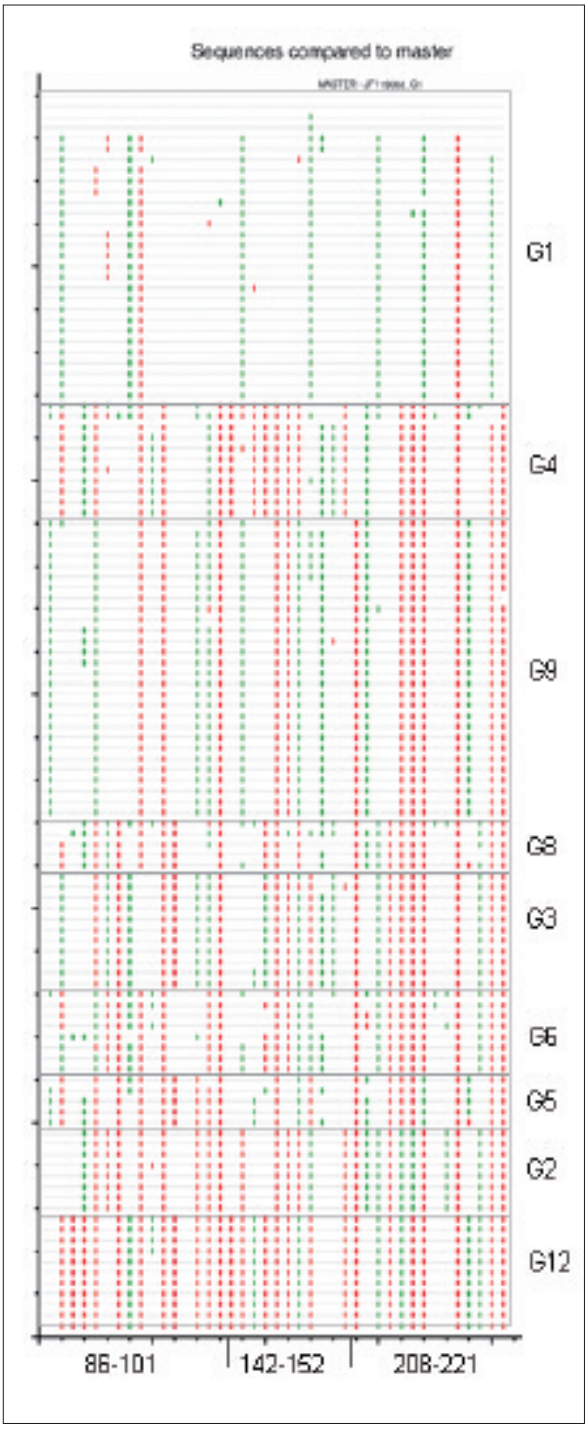

FIGURE 5 - *Analysis of silent and non-silent mutations in the immunodominant regions of VP7. A zoom-in of the silent vs non-silent mutation analysis, focused on the three immunodominant regions of the VP7 glycoprotein, is shown. The VP7 from the Iran/2010 sample (Accession #JF719063) was used as master for comparing all other sequences.*

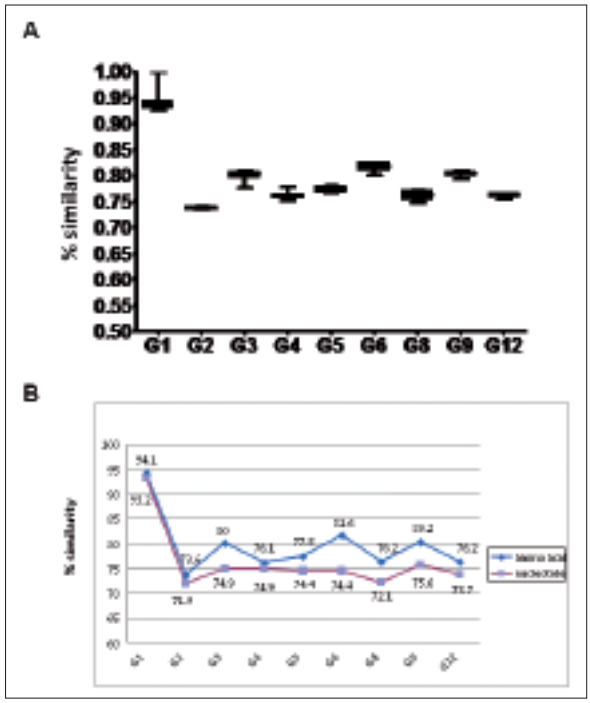

FIGURE 6 - *Analysis of amino acid divergence. A) The amino acid sequence of VP7 from the Iran/2010 sample was aligned to sequences of different genotypes. Percentage of similarity versus each reference sequence has been plotted according to genotype, showing average and standard deviation. (B) Side-by-side comparison of percentage of similarities in nucleotide and amino acid sequences, between our VP7 sequence and sequences of different genotypes.*

gions recognized in the VP7 glycoprotein (Dyall-Smith *et al.*, 1986), it was interesting to observe that, when compared to the *Iran/2010* sequence, very few mainly synonymous mutations were identified in the G1 reference sequences, while non-synonymous mutations were predominant in other G genotypes (Figure 5). These results indicated that the *Iran/2010* VP7 glycoprotein sequence was very conserved within the G1 genotype also in such immunodominant regions. Moreover, the complete absence of non-synonymous substitutions compared to the three clus-

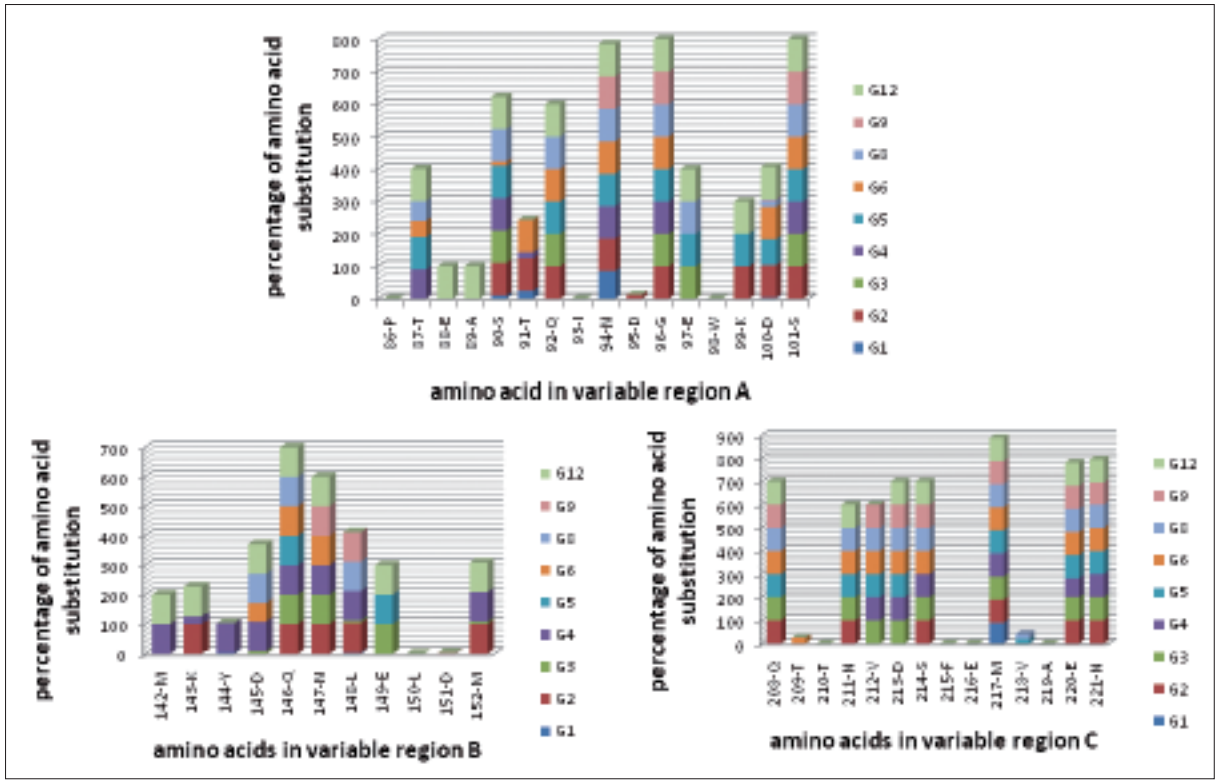

FIGURE 7 - *Amino acid substitutions in three immunodominant regions A, B and C. The amino acid sequence of the three immunodominant regions of VP7 from the Iran/2010 sample was aligned to sequences of different genotypes. The percentage of substitution for each amino acid residue in the different genotypes in shown.*

TABLE 2 - *Amino acid residue for each position in the three immunodominant VP7 regions compared to the Iran/2010 sequence. Numbers indicate the percentage of the indicated residue in sequences of the genotype. Asterisks (*) indicate 100% conservation of the aa residue in all the sequences of the genotype.*

Region A

	86	*87*	*88*	*89*	*90*	*91*	*92*	*93*	*94*	*95*	*96*	*97*	*98*	*99*	*100*	*101*
VP7 Iran/2010	*P*	*T*	*E*	*A*	*S*	*T*	*Q*	*I*	*N*	*D*	*G*	*E*	*W*	*K*	*D*	*S*
G1	*	*	*	*	$S_{89.3}$ $N_{10.7}$	T_{75} $N_{17.8}$,$A_{7.2}$	*	*	$N_{14.3}$ $S_{85.7}$	*	*	*	*	*	$D_{96.5}$ $N_{3.5}$	*
G2	*	*	*	*	K_{100}	N_{100}	E_{100}	*	S_{100}	$D_{87.5}$ $H_{12.5}$	N_{100}	*	*	E_{100}	N_{100}	T_{100}
G3	*	*	*	*	A_{100}	*	E_{100}	*	*	*	N_{91},D_9	S_{100}				T_{100}
G4	*	S_{82} P_9, T_9	*	*	P_{82},Q_{18}	T_{82} P_9,I_9	*	*	S100	*	T_{82},N_{18}	*	*	*	*	T_{91},A_9
G5	*	N_{100}	*	*	A_{100}	*	E_{100}	*	A_{100}	*	N_{60}, D_{40} T_{100}	K_{100}	*	T_{100}	E_{80} D_{20}	T_{100}
G6	*	V_{50} T_{50}	*	*	$S_{87.5}$ $A_{12.5}$	N_{100}	E_{100}	*	A_{100}	*		*	*	*	N_{100}	T_{100}
G8	*	A_{60} T_{40}	*	*	E_{100}	*	E_{100}	*	A_{100}	*	S_{100}	S_{100}	*	*	E_{20} D_{80}	T_{100}
G9	*	*	*	*	*	*	*	*	G_{100}	*	T_{100}	*	*	*	$D_{96.5}$ $N_{3.5}$	T_{100}
G12	*	S_{100}	S_{100}	V_{100}	T_{100}	*	E_{100}	*	T_{100}	*	P_{100}	D_{100}	*	T_{100}	N_{82},S_{18}	T_{100}

Region B

	142	*143*	*144*	*145*	*146*	*147*	*148*	*149*	*150*	*151*	*152*
VP7 Iran/2010	*M*	*K*	*Y*	*D*	*Q*	*N*	*L*	*E*	*L*	*D*	*M*
G1	*	*	$Y_{96.5}$ $S_{3.5}$	*	*	*	$L_{96.5}$ $F_{3.5}$	*	*	*	*
G2	*	R_{100}	*	*	N_{100}	T_{100}	S_{100}	*	*	*	A_{100}
G3	*	*	*	D_{91} V_9	A_{100}	T_{91},R_9	L_{91} V_9	Q_{100}	*	*	M_{91} R_9
G4	I_{100}	K_{73} R_{27}	F_{100}	A_{64},T_{36}	S_{100}	G_{100}	E_{100}	*	*	*	I_{100}
G5	*	*	*	*	A_{60},I_{20},V_{20}	*	*	Q_{100}	*	*	*
G6	*	*	*	N_{50},$A_{12.5}$ $D_{37.5}$	S_{100}	$T_{62.5}$,$A_{37.5}$	*	*	*	*	*
G8	*	*	*	N_{80},S_{20}	A_{100}	*	S_{100}	*	*	*	*
G9	*	*	*	*	S_{100}	T_{100}	*	*	*	$D_{96.5}$ $G_{3.5}$	*
G12	I_{100}	Q_{100}	*	Q_{100}	N_{100}	S_{100}	*	A_{100}	*	*	V_{100}

Region C

	208	*209*	*210*	*211*	*212*	*213*	*214*	*215*	*216*	*217*	*218*	*219*	*220*	*221*
VP7 Iran/2010	*Q*	*T*	*T*	*N*	*V*	*D*	*S*	*F*	*E*	*M*	*V*	*A*	*E*	*N*
G1	*	*	*	*	*	*	*	*	*	T_{89} M_{11}	*	*	*	*
G2	K_{100}	*	*	D_{100}	*	*	T_{100}	*	*	$I_{87.5}$, $N_{12.5}$	*	*	S_{100}	S_{100}
G3	L_{100}	*	*	D_{100}	T_{100}	N_{100}	T_{100}	*	*	E_{100}	*	*	T_{100}	A_{100}
G4	*	*	*	*	T_{100}	A_{82},N_{18}	T_{100}	*	*	T_{100}	*	*	D_{82} E_{18}	S_{100}
G5	L_{100}	*	*	D_{100}	T_{100}	N_{60},S_{40}	*	*	*	I_{60},T_{40}	V_{80} I_{20}	*	N_{60},S_{40}	A_{80},T_{20}
G6	$L_{87.5}$,$S_{12.5}$	T_{75} I_{25}	*	D_{100}	P_{100}	$N_{62.5}$,$S_{37.5}$	T_{100}	*	*	T_{100}	*	*	T_{100}	T_{100}
G8	L_{100}	*	*	D_{100}	T_{100}	T_{100}	T_{100}	*	*	E_{100}	V_{80} I_{20}	*	T_{100}	A_{100}
G9	I_{96},T_4	*	*	*	T_{96},P_4	A_{100}	T_{100}	*	*	E_{100}	*	*	T_{93},A_7	S_{96} N_4
G12	T_{100}	*	*	D_{100}	*	T_{91},A_9	T_{100}	*	*	E_{100}	*	*	N_{100}	A_{100}

tering G1 sequences (Figure 5) may explain the formation of such sub-cluster in the phylogenetic analysis.

Amino acid analysis

To confirm the data obtained with the genomic analysis, the amino acid sequence of the *Iran/2010* VP7 glycoprotein identified in the present study was deduced from the ORF of 981 nucleotides which extends from the initiation codon at nucleotides 49-51 to the termination codon at nucleotides 1024-1026 (numeration referring to segment 9 of viral genomic RNA) (Richardson *et al.,* 1984). The resulting amino acid sequence of 297 residues was compared to the same reference sequences of different genotypes mentioned above. Similar to the nucleotide sequence analysis, the amino acid sequence homology also showed that the *Iran/2010* VP7 glycoprotein belongs to the G1 genotype (92.5%-100% homology) with a protein homology >97.9% to the three clustering G1 variants mentioned above (Figure 6A). In particular, our sequence showed perfect similarity (100%) with the reference FJ811903 Iranian sequence, further supporting their close molecular relationship. By contrast, the *Iran/2010* sequence showed a significantly lower ($p<0.001$) average homology to the other G genotypes, with values ranging from 73.6 to 81.6% (Figure 6A). The overall percentage of similarity for the amino acid sequences was significantly higher than for the nucleotide sequences, further confirming the higher rate of synonymous mutations and, therefore, the lower rate of amino acid substitutions (Figure 6B).

In particular, looking more closely at the three VP7 immunodominant regions, it was confirmed that the *Iran/2010* sequence presents a very small number of mutations compared to G1 reference genotypes (Figure 7). Moreover, it is extremely interesting to identify residues with different pattern across genotypes. Indeed, some of them were perfectly conserved (i.e. 86P, 150L), some were extremely variable (i.e. 96G, 217M), while some were variable only in one or few specific genotypes (i.e. 88E, 149E) (Figure 7). Such data strongly suggested that in the three immunodominant regions of VP7 glycoprotein there were residues across genotypes under a significantly different pressure by the immune system. Moreover, in each amino acid position, when a substitution occurs, the substituting residue was

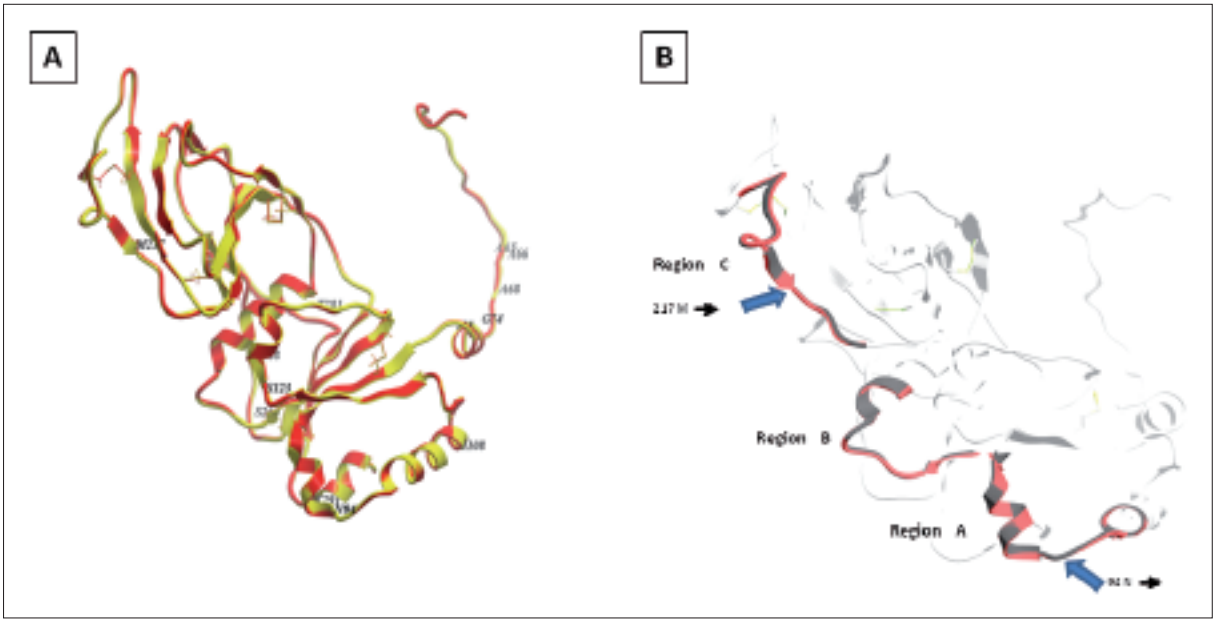

FIGURE 8 - *Protein structure analysis. The protein structure of VP7 from the Iran/2010 sample (red) has been superimposed to the VP7 consensus G1 protein sequence (yellow). (A) Mutations between Iran/2010 and G1 consensus sequence throughout the sequences are shown. (B) The three immunodominant regions are highlighted on the background of the whole VP7 proteins. The two mutations in such regions between Iran/2010 and G1 consensus sequence are shown.*

mostly genotype-specific, giving rise to a genotype-specific immune pattern (Table 2).

Protein structure analysis

In order to evaluate the effect of amino acid mutations on the VP7 glycoprotein structure and possibly protein-antibody binding, we performed a bio-informatic structural analysis using the *Iran/2010* sequence and a consensus of the G1 genotype. The analysis showed that the two VP7 sequences were perfectly superimposable, indicating that the few amino acid substitutions were all conservative and did not affect the molecule conformation (Figure 8A). Moreover, looking more closely at the VP7 immunodominant regions, the structure of all three regions (A, B and C) was conserved even in the presence of the two substitutions (N94S and M217T) (Figure 8B), confirming that the *Iran/2010* sequence may represent an important immunological prototype for the G1 genotype.

DISCUSSION

Epidemiological studies have shown that rotavirus genotypes G1P[8], G3P[8], G4P[8], G2P[4], and G9P[8] are the most common types detected in <5-year-old children with acute gastroenteritis in most countries during the last two decades. In particular, G1P[8] strains have been consistently present and often predominant in most locations (Gentsch *et al.,* 2005; Kirkwood, 2010; Santos and Hoshino, 2005), including Iran (Farahtaj *et al.,* 2007; Khalili *et al.,* 2004; Modaress *et al.,* 2011).

The present study describes the molecular characterization of rotavirus VP7 glycoprotein of G1 genotype identified in Iran in 2010 (*Iran/2010* isolate). As expected, the average nucleotide similarity of the *Iran/2010* sequence versus rotavirus sequences of different genotypes is much higher than G1 genotypes (average 93.2%) compared to other genotypes (average 71.9-75.6%). Moreover, the *Iran/2010* isolate shows the highest similarity with three G1 isolates from Iran, Russia and Germany. In particular, the extremely high homology (99.5%) to an Iranian sequence identified 2 years earlier suggests the persistence of genetic rotavirus variants circulating in Iran during the last couple of years.

The phylogenetic analysis confirms the prediction of genotype classification, showing that the *Iran/2010* isolate clusters with G1 genotype sequences, forming a separated monophyletic subcluster together with the three high homologous isolates to suggest a common ancestor, although such speculation is not supported by available epidemiological data. The only exception is represented by the G1 and G4 genotypes which show the closest phylogenetic relatedness, suggesting a more recent evolutionary divergence compared to other genotypes.

The analysis of synonymous and non-synonymous substitutions showed that regardless the genotype compared to the *Iran/2010* sequence the number of synonymous substitutions are significantly higher than non-synonymous ones. These results suggest that the VP7 glycoprotein, although under the pressure of the immune system, shows an overall phenotypic stability across genotypes with few highly variable regions. Considering the three immunodominant regions, the *Iran/2010* VP7 sequence shows very few mainly synonymous mutations compared to G1 reference sequences, while non-synonymous mutations are predominant compared to other genotypes. These results are confirmed by comparison of amino acid sequences, indicating that the *Iran/2010* VP7 glycoprotein sequence is very conserved within the G1 genotype also in such immunodominant regions. Moreover, the complete absence of non-synonymous substitutions and amino acid changes compared to the three phylogenetically related sequences identified in the present study may explain the sub-clustering within the G1 genotype and further supports their origin from a common ancestor.

Structural analysis shows that the amino acid mutations in the immunodominant regions of the *Iran/2010* VP7 sequence are conservative and do not affect the molecule conformation when compared to the consensus of the G1 genotype.

The results from the molecular characterization of G1 genotype VP7 from the *Iran/2010* isolate show a very limited intra-genotype genetic and antigenic variation. Such results indicate a limited antigenic variation within a single genotype also in the immunodominant regions of VP7 which needs to be validated by further and continuous molecular epidemiological studies. Indeed, the present data suggest the possible iden-

tification of candidate molecules to be used as antigen for eliciting genotype-specific protective immunity.

ACKNOWLEDGEMENTS

We thank Mofid Pediatric Hospital in Tehran for providing clinical samples. The VP7 nucleotide sequence of the Iran/2010 isolate has been deposited in the GenBank database and received the accession number JF719063. This study was funded and supported by Tehran University of Medical Sciences (TUMS); Grant n. 9429. It has also been part of a PhD thesis supported by Tehran University of Medical Sciences; Grant no. 240.4303.

REFERENCES

Aoki S.T., Settembre E.C., Trask S.D., Greenberg H.B., Harrison S.C, Dormitzer P.R. (2009). Structure of rotavirus outer-layer protein VP7 bound with a neutralizing. *Fab. Science.* **324**, 1444-1447.

Bishop R.F. (1996). Natural history of human rotavirus infection. *Arch. Virol. Suppl.* **12**, 119-128.

Bishop R.F., Davidson G.P., Holmes I.H, Ruck B.J. (1973). Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *Lancet.* **2**, 1281-1283.

Bishop R.F., Davidson G.P., Holmes I.H, Ruck B.J. (1974). Detection of a new virus by electron microscopy of faecal extracts from children with acute gastroenteritis. *Lancet.* **1**, 149-151.

Bishop R.F., Tzipori S.R., Coulson B.S., Unicomb L.E., Albert M.J, Barnes G.L. (1986). Heterologous protection against rotavirus-induced disease in gnotobiotic piglets. *J. Clin. Microbiol.* **24**, 1023-1028.

Crawford S.E., Estes M.K., Ciarlet M., Barone C., O'Neal C.M., Cohen J., et al. (1999). Heterotypic protection and induction of a broad heterotypic neutralization response by rotavirus-like particles. *J. Virol.* **73**, 4813-4822.

Dyall-Smith M.L., Lazdins I., Tregear G.W, Holmes I.H. (1986). Location of the major antigenic sites involved in rotavirus serotype-specific neutralization. *Proc. Natl. Acad. Sci.* **83**, 3465-3468.

Emamghorashi F., Rajabi S.H., Shadmand A., Zendegani N., Rabiee S., et al. (2008). Frequency of Rotavirus Infection in Children with Acute Gastroenteritis in Jahrom, South of Iran. *Iran. J. Med. Sci.* **33**, 84-87.

Esteghamati A., Gouya M., Keshtkar A., Najafi L., Zali M.R., Sanaei M., et al. (2009). Sentinel hospital-based surveillance of rotavirus diarrhea in Iran. *J. Infect. Dis.* Suppl 1, S244-S247.

Estes M.K., Cohen J. (1989). Rotavirus gene structure and function. *Microbiol. Rev.* **53**, 410-449.

Estes M.K., Kapikian A.Z. (2007). Rotavirus. Page 1917 in Fields Virology. Knipe D.M., Howly P.M., ed. Lippincott Williams & Wilkins, Philadelphia.

Farahtaj F., Gallimore C.I., Iturriza-Gomara M., Taremi M., Zali M.R., Edalatkhah H., et al. (2007). Rotavirus VP7, VP4 and VP6 genotypes co-circulating in Tehran, Iran, between 2003 and 2004. *Epidemiol. Infect.* **135**, 834-838.

Flewett T.H., Bryden A.S, Davies H. (1973). Letter: Virus particles in gastroenteritis. *Lancet.* **2**, 1497.

Gentsch J.R., Laird A.R., Bielfelt B., Griffin D.D., Banyai K., Ramachandran M., et al. (2005). Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. *J. Infect. Dis.* **192** (Suppl. 1), S146-S159.

Glass R.I., Parashar U.D., Bresee J.S., Turcios R., Fischer T.K., Widdowson M.A., et al. (2006). Rotavirus vaccines: current prospects and future challenges. *Lancet.* **368**, 323-332.

Glass, R.I., Wecker J., De Quadros C., Flem E., Toscano C., Parashar U.D., et al. (2008). Proceedings from the 8th International Rotavirus Symposium.

Gouvea V., Glass R.I., Woods P., Taniguchi K., Clark H.F., Forrester B., et al. (1990). Polymerase chain reaction amplification and typing of rotavirus nucleic acid from stool specimens. *J. Clin. Microbiol.* **28**, 276-282.

Greenberg H.B., Valdesuso J., van W.K., Midthun K., Walsh M., McAuliffe V., et al. (1983). Production and preliminary characterization of monoclonal antibodies directed at two surface proteins of rhesus rotavirus. *J. Virol.* **47**, 267-275.

Hamkar R., Yahyapour Y., Noroozi M., Nourijelyani K., Jalilvand S., Adibi L., et al. (2010). Prevalence of Rotavirus, Adenovirus and Astrovirus infections among patients with acute gastroenteritisin Northern Iran. Iranian. *J. Publ. Health.* **39**, 45-51.

Hoshino Y., Saif L.J., Sereno M.M., Chanock R.M, Kapikian A.Z. (1988). Infection immunity of piglets to either VP3 or VP7 outer capsid protein confers resistance to challenge with a virulent rotavirus bearing the corresponding antigen. *J. Virol.* **62**, 744-748.

Jukes T.H., Cantor C.R. (1969). Evolution of protein molecules. In Munro HN, editor, Mammalian Protein Metabolism, pp. 21-132, Academic Press, New York.

Kapikian A.Z., Kim H.W., Wyatt R.G., Rodriguez W.J., Ross S., Cline W.L., et al. (1974). Reoviruslike agent in stools: association with infantile diarrhea and development of serologic tests. *Science.* **185**, 1049-1053.

Khalili B., Cuevas L.E., Reisi N., Dove W., Cunliffe N.A, Hart C.A. (2004). Epidemiology of rotavirus diarrhoea in Iranian children. *J. Med. Virol.* **73**, 309-312.

Kirkwood C.D. (2010). Genetic and antigenic diversity of human rotaviruses: potential impact on vac-

cination programs. *J. Infect. Dis.* **202** (Suppl.), S43-S48.

KORBER B. (2000). HIV signature and sequence variation analysis. Page 55 in Computational analysis of HIV molecular sequences. G. R. Allen, H. L. Gerald, and eds, eds. Kluwer Academic Publishers, Dordrecht, Netherlands.

KORDIDARIAN R., KELISHADI R, ARJMANDFAR Y. (2007). Nosocomial infection due to rotavirus in infants in Alzahra Hospital, Isfahan, Iran. *J. Health Popul. Nutr.* **25**, 231-235.

MALEK M A., TELEB N., ABU-ELYAZEED R., RIDDLE M.S., SHERIF M.E., STEELE A.D., ET AL. (2010). The epidemiology of rotavirus diarrhea in countries in the Eastern Mediterranean Region. *J. Infect. Dis.* **202** (Suppl.), S12-S22.

MATTHIJNSSENS J., CIARLET M., HEIMAN E., ARIJS I., DELBEKE T., MCDONALD S.M., ET AL. (2008). Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J. Virol.* **82**, 3204-3219.

MATTHIJNSSENS J., CIARLET M., MCDONALD S.M., ATTOUI H., BANYAI K., BRISTER J.R., ET AL. (2011). Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch. Virol.* **156**, 1397-1413.

MCNEAL M.M., RAE M.N., CONNER M.E., WARD R.L. (1998). Stimulation of local immunity and protection in mice by intramuscular immunization with triple- or double-layered rotavirus particles and QS-21. *Virology.* **243**, 158-166.

MODARESS S., RAHBARIMANESH A.A., EDALAT R., SOHRABI A., MODARRES S., GOMARI H., ET AL. (2011). Human rotavirus genotypes detection among hospitalized children, a study in Tehran, Iran. *Arch. Iran.Med.* **14**, 39-45.

MODARRES S., RAHBARIMANESH A.A., KARIMI M., MODARRES S., MOTAMEDI-RAD M., SOHRABI A, ET AL. (2008). Electrophoretic RNA genomic profiles of rotavirus strains prevailing among hospitalized children with acute gastroenteritis in tehran, Iran. *Arch. Iran. Med.* **11**, 526-531.

OFFIT P.A., CLARK H.F., BLAVAT G., GREENBERG HB. (1986). Reassortant rotaviruses containing structural proteins vp3 and vp7 from different parents induce antibodies protective against each parental serotype. *J. Virol.* **60**, 491-496.

PAREZ N., FOURGEUX C., MOHAMED A., DUBUQUOY C., PILLOT M., DEHEE A., ET AL. (2006). Rectal immunization with rotavirus virus-like particles induces systemic and mucosal humoral immune responses and protects mice against rotavirus infection. *J. Virol.* **80**, 1752-1761.

RICHARDSON M.A., IWAMOTO A., IKEGAMI N., NOMOTO A, FURUICHI Y. (1984). Nucleotide sequence of the gene encoding the serotype-specific antigen of human (Wa) rotavirus: comparison with the homologous genes from simian SA11 and UK bovine rotaviruses. *J. Virol.* **51**, 860-862.

SADEGHIAN A., HAMEDI A., SADEGHIAN M., SADEGHIAN H. (2010). Incidence of rotavirus diarrhea in children under 6 years referred to the pediatric emergency and clinic of Ghaem Hospital, Mashhad, Iran. *Acta. Med. Iran.* **48**, 263-265.

SAMARBAFZADEH A., TEHRANI E.M., MAKVANDI M., TAREMI M. (2005). Epidemiological aspects of rotavirus infection in Ahwaz, Iran. *J. Health Popul. Nutr.* **23**, 245-249.

SANTOS N., HOSHINO Y. (2005). Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. *Rev. Med. Virol.* **15**, 29-56.

SCHUMANN T., HOTZEL H., OTTO P, JOHNE R. (2009). Evidence of interspecies transmission and reassortment among avian group A rotaviruses. *Virology.* 386, 334-343.

POND S.L., FROST S.D., MUSE S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* **21**, 676-679.

SOLBERG O.D., HASING M.E., TRUEBA G, EISENBERG J.N. (2009). Characterization of novel VP7, VP4, and VP6 genotypes of a previously untypeable group A rotavirus. *Virology.* **385**, 58-67.

TAMURA K., PETERSON D., PETERSON N., STECHER G., NEI M., KUMAR S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol* (in Press).

TANIGUCHI K., HOSHINO Y., NISHIKAWA K., GREEN K.Y., MALOY W.L., MORITA Y., ET AL. (1988). Cross-reactive and serotype-specific neutralization epitopes on VP7 of human rotavirus: nucleotide sequence analysis of antigenic mutants selected with monoclonal antibodies. *J. Virol.* **62**, 1870-1874.

THOMPSON J.D., HIGGINS D.G., GIBSON T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic. Acids. Res.* **22**, 4673-4680.

TROJNAR E., OTTO P., JOHNE R. (2009). The first complete genome sequence of a chicken group A rotavirus indicates independent evolution of mammalian and avian strains. *Virology.* **386**, 325-333.

WHO. (2009). Manual of rotavirus detection and characterization methods. Geneva.

WYATT R.G., MEBUS C.A., YOLKEN R.H., KALICA A.R., JAMES H.D. JR., KAPIKIAN A.Z, ET AL. (1979). Rotaviral immunity in gnotobiotic calves: heterologous resistance to human virus induced by bovine virus. *Science.* **203**, 548-550.