Bioinformatic prediction of polymerase elements in the rotavirus VP1 protein

RODRIGO VÁSQUEZ-DEL CARPIO*, JAIME L MORALES, MARIO BARRO*, ALBA RICARDO, EUGENIO SPENCER*

Laboratorio de Virología, Facultad de Química y Biología, Universidad de Santiago de Chile, Casilla 40 Correo 33, Santiago, Chile.

* Present address: Laboratory of Infectious diseases, NIAID, National Institutes of Health, 50 South Drive MSC 8026, Room 6314, Bethesda, MD 20892.

SUMMARY

Rotaviruses are the major cause of acute gastroenteritis in infants world-wide. The genome consists of eleven double stranded RNA segments. The major segment encodes the structural protein VP1, the viral RNA-dependent RNA polymerase (RdRp), which is a minor component of the viral inner core. This study is a detailed bioinformatic assessment of the VP1 sequence. Using various methods we have identified canonical motifs within the VP1 sequence which correspond to motifs previously identified within RdRps of other positive strand, double-strand RNA viruses. The study also predicts an overall structural conservation in the middle region that may correspond to the palm subdomain and part of the fingers and thumb subdomains, which comprise the polymerase core of the protein. Based on this analysis, we suggest that the rotavirus replicase has the minimal elements to function as an RNA-dependent RNA polymerase. VP1, besides having common RdRp features, also contains large unique regions that might be responsible for characteristic features observed in the *Reoviridae* family.

Key terms: Rotavirus, genome replication, RNA polymerase, bioinformatic analysis.

INTRODUCTION

Rotaviruses, members of the *Reoviridae* family, are the major world-wide cause of acute gastroenteritis in infants and young children (Parashar et al., 2003). The rotavirus virion is a non-enveloped icosahedron, consisting of three concentric protein layers and a viral genome composed of eleven double stranded RNA (dsRNA) segments (Kapikian et al., 2001). These segments encode six structural and six non-structural proteins (Estes, 2001). Segment one encodes the structural protein VP1, the rotavirus putative RNA-dependent RNA polymerase (RdRp) (Eiden & Hirshon, 1993, Gallegos & Patton, 1989, Valenzuela et al., 1991). This enzyme is proposed to

possess transcriptase and replicase functions. These activities catalyze the formation of the viral mRNA (plus strand synthesis) and the dsRNA genome (minus strand synthesis), respectively (Eiden & Hirshon, 1993, Estes, 2001, Patton et al., 2003). VP1 together with VP3, the capping enzyme (Liu et al., 1992, Pizarro et al., 1991), are the minor components of the viral core (a single copy of each per five-fold lattice) (Prasad et al., 1996).

VP1 specifically recognizes multiple signals contained in the last 60 nucleotides of the 3' end of the viral mRNA of gene 8 (Patton, 1996, Tortorici et al., 2003). To date, the region(s) of VP1 responsible for this specific recognition of the 3' end of the plus strand has not been identified. This is

The abbreviations used are: RdRp, RNA-dependent RNA polymerase; dsRNA, double stranded RNA; mRNA, messenger RNA; aa, amino acid(s); kDa, kilo Dalton.

*Corresponding author. Mailing address: Laboratorio de Virología, Facultad de Química y Biología, Universidad de Santiago de Chile, Casilla 40 Correo 33, Santiago, Chile. Phone: (562) 681-0185. Email: espencher@lauca.usach.cl

Received: April 3, 2006. Accepted: May 23, 2006

not unexpected due to the lack of information about the structural-functional features of this viral protein, i.e. the location of the nucleic acid-protein and the protein-protein interacting regions in VP1 are unknown. This can be attributed in part to its large size (1,088 aa, 125 kDa), to the lack of a solved structure and to the requirement of VP1 for the core lattice protein (VP2) to form a competent replication complex (Patton, 1996, Tortorici et al., 2003). To date, the VP1 sequence has been poorly studied; specifically, few amino acids have been suggested to play an important role in protein function. Previous studies include the identification of conserved residues by full sequence alignments of VP1 from group A, B and C rotaviruses (Bremont et al., 1992, Eiden & Hirshon, 1993, Mitchell & Both, 1990), and by short alignments using other viral replicases of positive strand RNA viruses (Cohen et al., 1989, Mitchell & Both, 1990). The function of these conserved amino acids has been described in other viral RNA polymerases (i.e. phage ϕ 6, HCV, Poliovirus RdRp), using biochemical (site-directed mutagenesis) and/or biophysical methods (crystallography and X-ray diffraction) (Bressanelli et al., 1999, Butcher et al., 2001, Ribas & Wickner, 1992). In such enzymes, these residues are organized in typical motifs and are situated in the catalytic core of the polymerases, playing an important role in catalysis. In rotavirus, there is no experimental data to show that these amino acids form part of the motifs and could be involved in polymerization or interaction with the template. Some early studies have suggested the presence of these motifs in the rotavirus VP1 protein by sequence comparison of viral RdRps (Bruenn, 1991, Almanza et al., 1994). However, these studies did not clearly describe the classical motifs containing the conserved amino acids, or analyze the VP1 sequence for features characteristic for other viral RNA polymerases.

The vast information available about viral RNA polymerases has allowed us to perform a detailed study of the rotavirus VP1 sequence using a bioinformatic

approach. This study is based on the observed conservation of the secondary structure of the palm subdomain in other replicases of positive strand and double strand RNA viruses. This conservation helped us first, to visualize and describe the possible motifs and amino acids that may be implicated in polymerization and sugar selection; and second, to delimit the polymerase region in the VP1 sequence guided by the predicted structural conservation in the middle part of the protein. Finally, the prediction of a polymerase region in VP1 enabled us to observe, by a conserved domain search, the possible tertiary structure of the central region of the protein.

METHODS

Cell culture, virus propagation and RNA purification

Rhesus monkey fetal kidney cells (MA104) were maintained in minimal essential media (MEM) supplemented with 10% fetal bovine serum and grown at 37 °C. Simian Rotavirus strain SA11 was propagated and titrated in these cells. The dsRNA was purified from viral particles as previously described (Spencer & Arias, 1981). The gene 1 was purified from the viral dsRNA genome by gel electrophoresis in 0.8% low melt point (LMP) agarose following previously described procedures (Sambrook et al., 1989); approximately 50 ng of purified gene one dsRNA was used for each RT-PCR amplification reaction.

cDNA synthesis, cloning and sequence determination

The complete open reading frame of gene one, which corresponds to VP1, was amplified using a two step RT-PCR reaction: AMV-RT (Promega) and Elongase (Invitrogen) were used for the cDNA synthesis and amplification steps, respectively. Using combinations of primers that contain Sal I or Hind III restriction sites, the amplified products were cloned in pFastBac Hta (Invitrogen)

and transformed into *Escherichia coli* DH5 α following general protocols described elsewhere (Sambrook et al., 1989). A combination of the primers: (5'-TAGCGTCGACGAATGGGGAAGTACAA TCTAATC-3'; 5' GGGAAGCTTCTATGG TTTATCAACATTCACTGG-3'), (5'-TAG CGTCGACGACCAGTGAATGTTGATAAA CCA-3'; 5'-GGGAAGCTTCTATCTCTTT TCATTATTAAGTAG-3') and (5'-TAGCG TCGACGACTACTTAATAATGAAAAGAGA-3'; 5'-GGGAAGCTTCTAATCTTGAAAG AAGTTCGCGTT-3'), was used to amplify the nucleic acid sequence corresponding to the N-termini, middle and C-termini region of VP1, respectively. Transformants were selected on LB agar plates supplemented with ampicillin (100 mg/ml). Plasmid purification was realized by the common alkaline lysis procedure (Sambrook et al., 1989). The presence of the correct insert was confirmed by restriction enzyme digestion and PCR. The DNA was visualized by electrophoresis on agarose gels [0.9%] stained with ethidium bromide [0.5 μ g/ml]. The insert was sequenced at least three times by automated sequencing with an ABI PRISM 3100 genetic analyzer (PE Applied Biosystems), the full sequence was determined by contiguous assembly.

Sequence manipulation and analysis

The obtained VP1 sequence was compared with various RdRp sequences available in the Genbank database (<http://www.ncbi.nih.gov/Genbank/index.html>). The alignments were made using *ClustalW* 3.0 (<http://www2.ebi.ac.uk/clustalw/>) and edited in *BOXSHADE* 3.21 (http://www.ch.embnet.org/software/BOX_form.html). The secondary structure prediction of complete amino acid sequences was made using the *PSIPRED* protein structure prediction server (<http://insulin.brunel.ac.uk/psipred/>). The visualization of the possible tertiary structure of the palm subdomain was initially performed using the conserved domain (CD) search option in the Blast page (<http://www.ncbi.nlm.nih.gov/BLAST/>), where the initial output is an alignment with various polymerases

including among them a protein with a solved structure. The alignment between the query protein and the protein with a solved structure allows the modeling program *Cn3D 4.0* (NCBI/NIH), to show the regions of similarity on the crystal structure of the solved protein. The motif searcher program *MEME* (GCG/Wisconsin package) was used to find and confirm the previous motifs identified by visual determination. The input to the program consisted of an appropriate pool of sequences (i.e. short sequences containing the described motifs). The sequences were of 457 aa in length and corresponded to the middle region of the proteins including, for instance, the RdRp canonical motifs.

Sequence accession numbers

The amino acid sequences of RdRps from various viruses selected for this study were: brome mosaic virus (BMV) (822 aa) accession No. CAA41362, tobacco mosaic virus (TMV) (1,616 aa) No. AAD44327.1, tomato bushy stunt virus (TBSV) (818 aa) No. CAB56480, poliovirus (462 aa) No. NP_041277, encephalomyocarditis virus (EMCV) (459 aa) No. GNNYEB, foot and mouth disease virus (FMDV) (470 aa) No. AAG35699, hepatitis A virus (HAV) (415 aa) No. CAA33490, hepatitis C virus (HCV) (591 aa) No. CAB10747, bacteriophage Q β (589 aa) No. CAA32872, vesicular stomatitis virus (VSV) (2,109 aa) No. NP_041716, infectious haematopoietic necrotic virus (IHNV) (1,986 aa) No. CAA61500, measles virus (MV) (2,183 aa) No. AAD29097, avian rotavirus P013 strain (1,088 aa) No. BAA24146, human rotavirus IDIR strain (1,159 aa) No. A44280, porcine rotavirus Cowden strain (1,082 aa) No. P1XRPC, grass carp reovirus VP2 (1,274 aa) No. AAG10436, blue tongue virus (BTV) (1,302 aa) No. RRXRBT, *Saccharomyces cerevisiae* L-A virus (731 aa) No. AAA50508, P2 bacteriophage ϕ 6 (665 aa) No. AAA32355 and P2 bacteriophage ϕ 13 (713 aa) No. AAG00444. The accession number of the rotavirus VP1 sequence originated in this study is DQ457016.

RESULTS

Comparison of VP1 sequences available in the database.

The complete VP1 nucleic and amino acid sequence reported in this study was used as a template to search for other related rotavirus VP1 sequences in the database (<http://www.ncbi.nlm.nih.gov/BLAST/>). The VP1 sequences were then aligned using the program *ClustalW*. Interestingly, when the amino acid sequence is aligned with VP1 sequences from the same group (A) (i.e. bovine UK and RF, porcine Gottfried, avian) or with other rotavirus groups (IDIR-B, Cowden-C), the main differences were found in the amino- and carboxi-termini of the protein and not in the central region (data not shown). This suggests that the middle section of the protein represents a region of high homology with respect to the amino acid variations that are seen for the rest of the protein and for instance this zone could have an important role in enzyme function, for example, polymerization.

Localization of canonical motifs displayed by other viral RdRps in the rotavirus VP1 sequence.

In different families of polymerases, or even within the same family there are only a few conserved amino acids, and these are essential for the catalytic function of the enzyme. These amino acids are in a particular and strictly structural context in the catalytic core of the enzyme (O'Reilly & Kao, 1998). Some of these amino acids are involved in the coordination of the bivalent cations needed for the nucleophilic attack mediated by a two-metal-ion mechanism (Steitz, 1998), while other residues are important for sugar selection (Brautigam & Steitz, 1998, O'Reilly & Kao, 1998). The motifs that contain these conserved amino acids are well described for different groups of polymerases, in particular the motif involved in the coordination of the bivalent ions (motif C). The nature of these motifs is diverse, they vary depending on the sugar selection of the enzyme and according to the class of polymerase that they belong to, for

example, reverse transcriptases (RTs), multimeric DNA-dependent RNA polymerases (RNA pols), RdRps (Joyce & Steitz, 1995). Several works previously described the nature of the motifs shared by viral RdRps (Koonin, 1991, Koonin et al., 1989, O'Reilly & Kao, 1998, Poch et al., 1989). This information was used to search the rotavirus VP1 sequence for the classical motifs described in other RdRps. In order to accomplish this, our initial approach was to search for motifs in VP1 guided by the defined structural context in which these are situated. The most notorious of these is the ubiquitous motif C, which is strictly located between two short beta strands usually present in the middle portion of the protein (O'Reilly & Kao, 1998). Koonin *et al.*, had initially described eight motifs in positive strand RNA viruses, where three of them are present in almost every RdRp, with very few exceptions (Shwed et al., 2002). We were able to find, guided by a particular secondary structure context, these three motifs (A, B and C, or often called motifs IV, V and VI), as well as two of the other five motifs initially described in positive strand RNA viruses, motifs D and F (Fig. 1) (Johnson et al., 2001, Koonin et al., 1989, O'Reilly & Kao, 1998). All of the motifs that were identified are in the same classical arrangement as is seen for other positive strand, double-strand RNA viruses (O'Reilly & Kao, 1998). In the case of motif E, which is involved in a hydrophobic interaction with the thumb subdomain (O'Reilly & Kao, 1998), we were unable to define the motif precisely. This was due to the fact that the motif could not be located in a correct structural boundary, which is usually described to have serine residues and to be in a rich beta strand region following the motif D (Johnson et al., 2001) (Fig. 1). Once the motifs were defined, short sequence alignments with other positive, and negative strand, double-strand RNA viruses were made (Fig. 2). The conserved amino acids that we found in VP1 are in a very defined structural context as was described for other viral RdRps. The results also show that the motifs belonging to positive strand and dsRNA viruses are much more similar between them than when compared with the

representatives of the negative strand RNA viruses. This could give us an idea of the phylogeny of the RNA viruses, where possibly the minus strand RNA viruses could have had an early divergence or, alternatively, have a different origin. This suggestion is in agreement with a phylogenetic tree made in the *MEGA* program (Version 2.1) (data not shown) and with more extended studies (Zanotto et al., 1996).

In a second approach, with the aim to find novel motifs in the rotavirus sequence, we performed a motif search using the MEME program (GCG/Wisconsin package) (Bailey & Elkan, 1994). The results of the program show the presence of the canonical motifs in the VP1 sequence as previously identified in this study (Fig. 1) and no novel motifs. The results obtained using this program serve to define similarities with the motifs detected earlier by visual analysis of the rotavirus VP1 sequence.

In order to find other amino acids of interest in the rotavirus replicase, we performed a PSI-Blast protein search (non-redundant database), using small VP1 sequence segments. Each segment corresponds to 128 aa of the full sequence with an overlap of 64 aa between them. This approach was made with the aim to get only short nearly exact matches and for instance, to avoid the polymerase region bias during

the search. The search was performed until the iterations converged. Only one site of interest with a value sufficiently higher than the threshold background level was found. It is located towards the C-terminus, between the amino acids 802 and 856. This site presents similarity with mitochondrial ATPases/ATP synthases, and could be a novel ATP binding site in VP1, different to the more common NTP binding loop used for polymerization, which is composed of amino acids corresponding to the fingers and palm subdomains (motif F) (Butcher et al., 2001).

Prediction of a polymerase region in VP1 based on the RdRp structural conservation of other positive strand and dsRNA viruses

Despite their great structural heterogeneity, distinct types of polymerases share common features that are responsible for their catalytic function of polymerization of nucleic acids (Steitz, 1998, Steitz, 1999). Their domains have a similar overall architecture, with a conformation resembling a right-hand (Brautigam & Steitz, 1998, Cramer, 2002). The polymerases can be classified by their catalytic function, for example, according to the nature of the template that is recognized and to the type of substrate (rNTPs or dNTPs) that is used for the polymerization, and/or to the structural

Figure 1: Identification of the canonical motifs present in other viral RdRps in rotavirus VP1. The motifs in the VP1 protein are in a particular structural context and in a specific arrangement, as is the case for other viral RdRps. The identified motifs are shown by underlines of different colors. In addition a tentative location for motif E is shown. Alpha helix and beta strand predicted secondary structures are represented as an arrow and a tubular structure, respectively. Numbers denote the amino acid at which the sequence shown starts or ends.

Figure 2: Alignments of various viral RdRp sequences from positive, double and negative stranded RNA viruses. The three canonical motifs located in a precise structural context in the RdRps are shown. Grey highlighting denotes amino acid similarity. Black highlighting denotes identity between the sequences belonging to the same group. The amino acids highlighted with different colors (yellow, cyan, red or blue) also represent identity, and are the conserved amino acids described in several studies that comprise the active site and are critical for the polymerase activity. Numbers at the left of each sequence show the position at which the residue starts in the full sequence. Graphics in the upper part of the alignment show the representative and conserved secondary structure predicted for the motifs. The blue line shown in the double stranded group indicates rotavirus sequences.

architecture of their subdomains (fingers, palm and thumb). The RdRps, as is the case for other classes of polymerases, share a particular overall structural conservation of their basic domains (fingers, palm and thumb) within the family (O'Reilly & Kao, 1998). With the objective to determine if the rotavirus polymerase adhered to this structural conservation, we performed the secondary structure prediction of the VP1 sequence together with RdRps from positive strand and dsRNA viruses. The program *PSIPRED* was used to predict the secondary structures of the different viral sequences. Results of the predictions suggest a structural conservation of the polymerase core region in the RdRps secondary structure, as was previously described (Fig.

3) (O'Reilly & Kao, 1998). This conservation is extended to the predicted rotavirus VP1 secondary structure, where the conserved part is in its middle region (Fig. 3). This region, similar to the other viral RNA polymerases, is predicted to be located the palm subdomain and part of the fingers subdomain. The predicted conservation in the middle part of the protein is in concordance with the distribution of the motifs and amino acids implicated in the polymerization reaction (located mostly within the palm subdomain) (Brautigam & Steitz, 1998). The observed degree of conservation decreases towards the amino- and carboxy-termini, zones in which the main part of the fingers (the distal region to the palm) and the complete thumb

subdomains are probably located. Finally, close to the N- and C-termini of the protein, the predicted structural similarity completely breaks down. The polymerase region does not cover these zones of the protein, and is predicted to end approximately 320 aa from the N-termini and 220 from the C-termini of the VP1 sequence, thus comprising almost half of the total protein sequence length (Fig. 3). The N- and C-termini of the protein vary greatly from one polymerase to another, these unique regions may be involved in the specific recognition of template by the polymerase, or alternatively, may give some special catalytic properties to the protein (eg. a proofreading domain in the Klenow fragment, or a RNase H domain in the RT of HIV) (O'Reilly & Kao, 1998).

Visualization of a possible three-dimensional structure of the VP1 middle region based on the partial sequence similarity with the rabbit hemorrhagic disease virus polymerase

The possible three-dimensional structure of the rotavirus polymerase middle region was observed by a conserved domain alignment (CD alignment) (Fig. 4). This was possible using the CD search option in the Blast

page (see methodology) (Marchler-Bauer et al., 2003). This program compares a protein sequence against the conserved domain database (Smart and Pfam) using the RPS-BLAST program. This allows the prediction of known functional and structural domains in protein query sequences. Enough similarity was found to make a partial alignment of the VP1 sequence with the sequence of a crystallized protein, in this case the RdRp of the Rabbit Hemorrhagic Disease Virus (RHDV), a positive strand RNA virus belonging to the *Caliciviridae* family (Ng et al., 2002). The program was able to perform an alignment of residues 458-633 of the VP1 sequence and 186-356 of the RHDV RdRp sequence (where most of the motifs of both proteins are located), and to show this region of VP1 as a three-dimensional structure based on the structure of this region from the crystallized RHDV protein (Fig. 4). Thus, this approach was possible using the sequence corresponding to the middle region of the protein, from residues 438 to 776 of VP1, and not the rest of the protein due to the lack of general sequence similarity and structural conservation that occurs outside of the middle region of the replicases. The regions of similarity obtained by the alignment and

Figure 3: Secondary structure conservation at the polymerase region of the rotavirus replicase. The superimposition of the predicted secondary structures belonging to plus and double stranded viral RdRp sequences is shown. Motif C (GDD) of the palm subdomain was used as a signature to start the superimposition of the secondary structures towards N- and C-termini of the protein. The dashed lines in some structures denote a gap in the superimposed predicted structure. As was seen previously, the continuity of the palm subdomain is stopped by part of the fingers subdomains (proximal to the palm region) in these types of polymerase (RdRps or RTs). Some of the predicted secondary structures do not have complete structural equivalence compared to the crystallized form of the protein. The predicted secondary structures were chosen for the structural superimposition in order to have a consistent bias.

observed in the crystal structure of the RHDV RdRp corresponds to the predicted palm and part of the fingers subdomain in VP1, where the conserved residues (identical or similar) are those that lie in the active site of the protein and are mainly located in the palm subdomain. The rest of the VP1 sequence in the alignment does not have any similarity with RHDV RdRp and could correspond to part of the putative fingers and part of the thumb subdomain, where the similarity decreases. The results obtained in this section provide us a basis to postulate that the central core of the rotavirus replicase probably has a typical

right handed conformation described for other RdRps of positive strand and dsRNA viruses.

DISCUSSION

The gene one of the rotavirus genome encodes for the structural protein VP1. This viral protein forms part of the rotavirus replication machinery that also includes the core lattice protein VP2 and the capping enzyme VP3. VP1 needs the presence of VP2 to form a competent replication complex and display replicase activity. VP1

Figure 4: Region of similarity between the central core of the rotavirus VP1 and the middle region of the crystallized RHDV polymerase. **A**, RHDV RdRp X-ray crystal structure (PDB: *IKHV*), determined by Ng *et al.* The pink tubular and the cyan arrow-shaped structures denote alpha helix and beta strand structures, respectively. The worm-shaped structure denotes the α -backbone of the protein. **B**, region of the RHDV replicase that has sequence similarity with the rotavirus VP1 protein, shown in the same orientation as figure A. This region corresponds to the middle region of the RHDV RdRp where most of the palm and part of the fingers subdomains are located, this region matches with the middle region of VP1. The residues forming part of the region with similarity are shown in blue, while identical residues are shown in red. **C**, the structure in figure B was rotated 90° towards the viewer in the y-axis from top to front, to show the conserved amino acids of the different motifs (yellow). **D**, sequence alignment of the CD search results, the numbers denotes the residue positions in the RHDV RdRp sequence. The colors have been maintained with respect to figures B and C. The corresponding motifs are shown under the sequence alignment. Figures were prepared using the *Cn3D* modeling program (NCBI/NIH).

has been assigned as the viral RdRp but little is known about its functional domains. This protein is not well studied, i.e. there has been no information reported about which region of VP1 is interacting with the viral mRNA or dsRNA templates prior to polymerization, or with the structural proteins VP2 and VP3, or the non-structural protein NSP2. Also, there have been no reports of which amino acids of the rotavirus replicase have an important role in the catalytic function of the protein. With the aim to provide a better understanding of this viral protein, we performed a detailed analysis of its amino acid sequence.

Bioinformatic programs available on the world wide web (internet), were used to predict motifs shared by viral RdRps in the rotavirus VP1 sequence. We were able to predict five of the eight motifs initially described in plus stranded virus polymerases using the structural conservation in the middle part of the viral polymerases as a guide. In addition, we identified conserved amino acids that make up these motifs, which could be involved in polymerization and sugar selection during nucleic acid synthesis. A predicted structural conservation of the putative palm subdomain in rotavirus VP1 by comparison with the predicted secondary structures of other viral RNA polymerases was shown. Finally, this predicted conservation in the middle part of the protein allowed us to perform a conserved domain search, and further, observe the possible three-dimensional structure of the VP1 middle part based on the RHDV polymerase. Based on this approach, we suggest that the rotavirus replicase contains features common to other positive strand, dsRNA viral polymerases, namely motifs and conserved amino acids, including an overall structural conservation of the middle portion (proximal part of the predicted fingers and palm subdomains).

As was expected, the predicted structural conservation was reduced towards the distal fingers (fingertips) and thumb putative subdomains, and was absent towards the N- and C-termini, based on other viral RdRps. Since the palm subdomain harbors nearly all the described

motifs (only one known motif is present in the fingers subdomain, motif F, and there are no classical motifs described for the thumb), it is reasonable to suggest that this domain is structurally conserved and that this structural conservation decreases closer to the other subdomains, where there is almost no motif that comprises the catalytic nuclei of the enzyme. We also suggest that in the case of a large size protein, such as the rotavirus replicase (1,088 aa, 125 kDa), the polymerase region corresponds to only one half of the total protein, where the other half of the VP1 protein (distributed in the N- and C-termini of the protein) could be unique regions. This data is in concordance with the size observed for other RdRps polymerase domains (eg. in phage $\phi 6$ is approximately 600 aa) (Butcher et al., 2001). This is even more evident in the solved structure of the reovirus $\lambda 3$ RdRp (Tao et al., 2002), where the polymerase domain corresponds to only 509 of a total of 1,267 aa. This protein has two accessory domains, an N-terminal domain and a C-terminal (bracelet) domain, that give it a three-dimensional cage-like structure with other characteristic features. It is presumable that related viruses belonging to the same *Reoviridae* family, like rotavirus, will have the similar overall domain dispositions and arrangements, including some unique characteristics. It is postulated that every polymerase must have two types of interactions with its template: one specific, the interaction that is made prior to the formation of the pre-initiation complex (binding); and another unspecific, the interaction that the polymerase has while bound to the template during elongation (Steitz, 1998). It will be interesting to see if the rotavirus RdRp has a defined domain for the specific recognition of its template, and where this domain is located. This type of bioinformatic approach enabled us to suggest unique regions in the protein, where one can look for novel protein functions. For example, a novel site in the rotavirus VP1 sequence was found (residues 802 to 856) that could be related to a possible ATPase activity and be an

attractive target for mutagenesis. Such an activity could be useful for a helicase activity during the transcriptase mode of VP1, when the enzyme utilizes dsRNA as a template, and be related to the high *in vitro* polymerase activity observed for the rotavirus double-layered particles (Spencer & Arias, 1981). We are aware that these are predictions and they will require further experimental studies to corroborate the data, particularly using biophysical methods such as crystallography/X-ray diffraction or NMR techniques and functional methods such as mutagenesis and biochemical studies. Nonetheless using this kind of approach we have strengthened the notion that VP1 is the rotavirus RNA-dependent RNA polymerase.

ACKNOWLEDGMENTS

We are grateful to Dr. John T. Patton (NIAID/NIH) for the resources kindly provided during the VP1 SA11 sequence determination and to Dr. F.D. Gonzalez-Nilo (U. de Talca) for the computational resources and expert advice. We also want to thank Dr. J. Chnaiderman for critical review of the manuscript. R.V.D was a fellowship recipient from DAAD, Germany and from CONICYT, Chile. This work was partially supported by FONDECYT grant #1050002 to ES.

REFERENCES

ALMANZA L, ARIAS CF, LÓPEZ S (1994) Amino acid sequence of the porcine rotavirus YM VP1 protein. *Res Virol* 145:313-317

BAILEY TL, ELKAN C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2: 28-36

BRAUTIGAM CA, STEITZ TA (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr Opin Struct Biol* 8: 54-63

BREMONT M, JUSTE-LESAGE P, CHABANNE-VAUTHEROT D, CHARPILLENNE A & COHEN J (1992) Sequences of the four larger proteins of a porcine group C rotavirus and comparison with the equivalent group A rotavirus proteins. *Virology* 18: 684-92

BRESSANELLI S, TOMEI L, ROUSSEL A, INCITTI I,

VITALE RL, MATHIEU M, DE FRANCESCO R, REY FA (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc Natl Acad Sci U S A* 96: 13034-9

BRUENN JA (1991) Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucl Acids Res* 19:217-226

BUTCHER SJ, GRIMES JM, MAKEYEV EV, BAMFORD DH, STUART DI (2001). A mechanism for initiating RNA-dependent RNA polymerization. *Nature* 410: 235-40

COHEN J, CHARPILLENNE A, CHILMONCZYK S, ESTES MK (1989) Nucleotide sequence of bovine rotavirus gene 1 and expression of the gene product in baculovirus. *Virology* 171: 131-40

CRAMER P (2002) Common structural features of nucleic acid polymerases. *Bioessays* 24: 724-9

EIDEN JJ, HIRSHON C (1993) Sequence analysis of group B rotavirus gene 1 and definition of a rotavirus-specific sequence motif within the RNA polymerase gene. *Virology* 192: 154-60

ESTES MK (2001) Rotavirus and their Replication. In *Fields Fundamental Virology*, pp 1747-1785. Edited by DM Knipe, PM Howley. Philadelphia: Lippincott Williams & Wilkins Press

GALLEGOS CO, PATTON JT (1989) Characterization of rotavirus replication intermediates: a model for the assembly of single-shelled particles. *Virology* 172: 616-27

JOHNSON KN, JOHNSON KL, DASGUPTA R, GRATSCHEV T, BALL LA (2001) Comparisons among the larger genome segments of six nodaviruses and their encoded RNA replicases. *J Gen Virol* 82: 1855-66

JOYCE CM, STEITZ TA (1995) Polymerase structures and function: variations on a theme? *J Bacteriol* 177: 6321-9

KAPIKIAN AZ, HOSHINO Y, CHANOCK RM (2001) Rotaviruses. In *Fields Fundamental Virology*, 4th Edition edn, pp 1787-1833. Edited by DM Knipe, PM Howley. Philadelphia: Lippincott Williams & Wilkins Press

KOONIN EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72: 2197-206

KOONIN EV, GORBALENYA AE, CHUMAKOV KM (1989) Tentative identification of RNA-dependent RNA polymerases of dsRNA viruses and their relationship to positive strand RNA viral polymerases. *FEBS Lett* 252: 42-6

LIU M, MATTION NM, ESTES MK (1992) Rotavirus VP3 expressed in insect cells possesses guanylyltransferase activity. *Virology* 188: 77-84

MARCHLER-BAUER A, ANDERSON JB, DEWEESE-SCOTT C, FEDOROVA ND, GEER LY, HE S, HURWITZ DI, JACKSON JD, JACOBS AR, LANCZYCKI CJ, LIEBERT CA, LIU C, MADEJ T, MARCHLER GH, MAZUMDER R, NIKOLSKAYA AN, PANCHENKO AR, RAO BS, SHOEMAKER BA, SIMONYAN V, SONG JS, THIESSEN PA, VASUDEVAN S, WANG Y, YAMASHITA RA, YIN JJ, BRYANT SH (2003) CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* 31: 383-7

MITCHELL DB, BOTH GW (1990) Completion of the genomic sequence of the simian rotavirus SA11: nucleotide sequences of segments 1, 2, and 3. *Virology* 177: 324-31.

NG KK, CHERNEY MM, VÁZQUEZ AL, MACHIN A, ALONSO JM, PARRA F, JAMES MN (2002) Crystal

- structures of active and inactive conformations of a caliciviral RNA-dependent RNA polymerase. *J Biol Chem* 277: 1381-7
- O'REILLY EK, KAO CC (1998) Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology* 252: 287-303
- PARASHAR UD, HUMMELMAN EG, BRESEE JS, MILLER MA, GLASS RI (2003) Global illness and deaths caused by rotavirus disease in children. *Emerg Infect Dis* 9: 565-72
- PATTON JT (1996) Rotavirus VP1 alone specifically binds to the 3' end of viral mRNA, but the interaction is not sufficient to initiate minus-strand synthesis. *J Virol* 70: 7940-7
- PATTON JT, KEARNEY K, TARAPOREWALA ZF (2003) Rotavirus genome replication: role of RNA-binding proteins. In *Perspectives in Medical Virology: Viral Gastroenteritis*, pp 165-183. Edited by J Gray, U Dusselberger. Amsterdam: Elsevier Science B.V.
- PIZARRO JL, SANDINO AM, PIZARRO JM, FERNANDEZ J, SPENCER E (1991) Characterization of rotavirus guanylyltransferase activity associated with polypeptide VP3. *J Gen Virol* 72: 325-32
- POCH O, SAUVAGET I, DELARUE M, TORDO N (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *Embo J* 8: 3867-74
- PRASAD BV, ROTHNAGEL R, ZENG CQ, JAKANA J, LAWTON JA, CHIU W, ESTES MK (1996) Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature* 382: 471-3
- RIBAS JC, WICKNER RB (1992) RNA-dependent RNA polymerase consensus sequence of the L-A double-stranded RNA virus: definition of essential domains. *Proc Natl Acad Sci U S A* 89: 2185-9
- SAMBROOK J, FRITSCH EF, MANIATIS T (1989) *Molecular Cloning: A Laboratory Manual*, 2 edn. New York: Cold Spring Harbor Laboratory Press
- SHWED PS, DOBOS P, CAMERON LA, VAKHARIA VN, DUNCAN R (2002) Birnavirus VP1 proteins form a distinct subgroup of RNA-dependent RNA polymerases lacking a GDD motif. *Virology* 296: 241-50
- SPENCER E & ARIAS ML (1981) In vitro transcription catalyzed by heat-treated human rotavirus. *J Virol* 40: 1-10
- STEITZ TA (1998) A mechanism for all polymerases. *Nature* 391: 231-2
- STEITZ TA (1999) DNA polymerases: structural diversity and common mechanisms. *J Biol Chem* 274, 17395-8
- TAO Y, FARSETTA DL, NIBERT ML, HARRISON SC (2002) RNA synthesis in a cage—structural studies of reovirus polymerase lambda3. *Cell* 111: 733-45
- TORTORICI MA, BROERING TJ, NIBERT ML, PATTON JT (2003) Template recognition and formation of initiation complexes by the replicase of a segmented double-stranded RNA virus. *J Biol Chem* 278: 32673-82
- VALENZUELA S, PIZARRO J, SANDINO AM, VÁSQUEZ M, FERNÁNDEZ J, HERNÁNDEZ O, PATTON J, SPENCER E (1991) Photoaffinity labeling of rotavirus VP1 with 8-azido-ATP: identification of the viral RNA polymerase. *J Virol* 65: 3964-7
- ZANOTTO PM, GIBBS MJ, GOULD EA, HOLMES EC (1996) A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *J Virol* 70: 6083-96