

Highly similar structural frames link the template tunnel and NTP entry tunnel to the exterior surface in RNA-dependent RNA polymerases

Dorothy M. Lang^{1,*}, A. T. Zemla^{2,*} and C. L. Ecale Zhou²

¹Physical and Life Sciences Division and ²Global Security Computing Applications Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

Received July 18, 2012; Revised October 31, 2012; Accepted November 1, 2012

ABSTRACT

RNA-dependent RNA polymerase (RdRp) is essential to viral replication and is therefore one of the primary targets of countermeasures against these dangerous infectious agents. Development of broad-spectrum therapeutics targeting polymerases has been hampered by the extreme sequence variability of these sequences. RdRps range in length from 400–800 residues, yet contain only ~20 residues that are conserved in most species. In this study, we made structure-based comparisons that are independent of sequence composition using a recently developed algorithm. We identified residue-to-residue correspondences of multiple protein structures and created (two-dimensional) structure-based alignment maps of 37 polymerase structures that provide both sequence and structure details. Using these maps, we determined that ~75% of each polymerase species consists of seven protein segments, each of which has high structural similarity to segments in other species, though they are widely divergent in sequence composition and order. We define each of these segments as a ‘homomorph’, and each includes (though most are much larger than) the well-known conserved polymerase motifs. All homomorphs contact the template tunnel or nucleoside triphosphate (NTP) entry tunnel and the exterior of the protein, suggesting they constitute a structural and functional skeleton common among the polymerases.

INTRODUCTION

The polymerase protein family has been studied extensively for >40 years. This interest has been motivated by

their unique function—to replicate all forms of life, and confounded by their sequence diversity. As more tertiary structures of polymerase were solved, it became apparent that widely diverse sequences form highly similar structures. There has not, until recently, been a time-effective computational method to make detailed comparisons of these observations. The objective of this study was to clarify the relationship between structure and sequence in a group of RNA-dependent RNA polymerases (RdRps) that replicate many of the viruses that represent significant threats to life throughout the world. We selected well-studied species in order to maximize the amount of experimental data that could be used to evaluate the association of functional residues and structure (Table 1). We used the StralSV algorithm (1) to perform structure comparisons between all of the selected species. We created maps of residue-to-residue (R2R) correspondence from which we determined the boundaries of structurally similar segments—which we named ‘homomorphs’. In contrast to the relatively short lengths of previously described motifs, we found that most homomorphs are long, and each provides a structural connection between the template tunnel or NTP entry tunnel and the exterior of the protein.

The tertiary structure of the replicative unit of most RdRps is highly conserved (2). It resembles that of a right-handed palm, with finger-like folds curved inward to form a tunnel that encircles the template that is being processed (3). Most single-unit polymerases are 400–800 residues in length. Early polymerase studies found that ~22 residues are highly conserved in all polymerases, and in most species they are in the same sequential order (4). Some are clustered, with two to four highly conserved residues within a segment ~10 residues in length. The sequence segment that includes each of the highly conserved residues or clusters has been described as a motif. The motifs are arranged in most species in the order: G-F1-F2-F3-A-B-C-D-E. The birnaviruses (IBDV and IPNV) differ from this scheme due to a

*To whom correspondence should be addressed. Tel: +1 209 839 1880; Fax: +1 925 423 6437; Email: dorothylang@gmail.com
Correspondence may also be addressed to A. T. Zemla. Tel: +1 925 423 5571; Fax: +1 925 423 6437; Email: zemla1@llnl.gov

Table 1. Viral species and PDB structures used as queries in this study

Queries	Abbreviations	PDB	Length of structure
Picornaviridae, RdRp, +ssRNA			
Poliovirus	PV	1RA6 (3)	461
Poliovirus (open)	PV	3OL6 (9)	471
Poliovirus (closed)	PV	3OL7 (9)	471
Coxsackie virus	COXV	3DDK (11)	462
Human rhinovirus	HRV	1XR6 (12)	460
Human rhinovirus	HRV	1XR7 (12)	460
Foot and mouth disease virus	FMDV	1U09 (2)	476
Caliciviridae, RdRp, +ssRNA			
Norwalk virus	NV	1SH0 (13)	510
Rabbit hemorrhagic disease virus	RHDV	1KHV (14)	516
Rabbit hemorrhagic disease virus	RHDV	1KHW (14)	516
Sapporo virus	SAPV	2CKW (15)	515
Flaviviridae, RdRp, +ssRNA			
Hepatitis C virus	HCV	1NB4 (16)	570
Hepatitis C virus	HCV	1NB7 (16)	570
Bovine viral diarrhea virus	BVDV	1S48 (17)	609
West Nile (Kunjin) virus	WNV	2HCS (18)	595
Dengue virus	DENV	2J7U (19)	635
Cystoviridae, RdRp, dsRNA			
Bacteriophage Phi6	PHI6	1HHS (20)	664
Bacteriophage Phi6	PHI6	1HI0 (20)	664
Reoviridae, RdRp, dsRNA			
Reovirus lambda 3	REOV	1MUK (21)	1267
Reovirus lambda 3	REOV	1N35 (21)	1267
Rotavirus	ROTAV	2R7X (22)	1095
Birnaviridae, RdRp, dsRNA			
Infectious bursal disease, birnavirus	IBDV	2PGG (5)	774
Infectious pancreatic necrosis	IPNV	2YI8 (23)	799
Infectious pancreatic necrosis	IPNV	2YI9 (23)	799
Infectious pancreatic necrosis	IPNV	2YIA (23)	799
RdDp			
HIV1	HIV1	1RT1 (24)	560
HIV1 (closed)	HIV1	1RTD (25)	554
HIV1 (open)	HIV1	2HMI (26)	558
Tribolium castaneum	TERT	3DU6 (27)	596
DdRp			
Bacteriophage T7 (closed)	T7 RNAP	1S77 (28)	883
Bacteriophage T7 (closed)	T7 RNAP	2AJQ (29)	704
Bacteriophage T7 (open)	T7 RNAP	1MSW (30)	883
Bacteriophage N4	N4	3C2P (31)	1117
DdDp			
Thermus aquaticus (open)	TAQ	2KTQ (32)	538
Thermus aquaticus (closed)	TAQ	3KTQ (32)	540
Thermus aquaticus (open)	TAQ	4KTQ (32)	539
Bacteriophage T7	T7 DNAP	1T7P (33)	698

Queries of both open and closed structures were used to assess the impact of this feature on R2R correspondence. Enterobacteria bacteriophage T7 [NCBI:NC_001604], which has a genome of 39937bp and 60 proteins and infects *Escherichia coli*, contains both RNA polymerase (T7 RNAP) and DNA polymerase (T7 DNAP). The T7 RNAP, typified by PDB:1S77 replicates multiple sequences of RNA initiated from transcription sites throughout the genome. The T7 DNAP, typified by PDB:1T7P, ‘...fills DNA gaps that arise during DNA repair, recombination and replication...’—a Family A polymerase (NCBI:NC_001604).

transversion involving the C Motif (C-A-B) (5). The references for each of the motifs and species that have been studied are listed in Table 2; these references were selected because they included either alignments of one or more motifs with several species, or alignments for a particular motif not found elsewhere. Apart from these conserved motifs, the RdRp sequences are highly variable. An extensive study of picornaviruses by Koonin *et al.*(6) illustrates this variability. The basis of Koonin *et al.*'s study was an alignment of 64 species using the algorithm multiple sequence

comparison by log expectation (MUSCLE) (7) and a manual adjustment. The alignment included four species that are also in our sample group, and these had a total of 32 conserved residues within sequences ~550 residues in length.

The analysis that we present in the following pages demonstrates that highly similar structures can be formed by very different sequences. Structure comparison, rather than sequence comparison, enabled us to readily recognize functionally significant segments of similarity and difference between sequences.

Table 2. Summary of reference publications that identify multiple motifs in RdRps included in the query set

Motifs	Conserved SxGresidues		KxE; KxR; Q;K DYxxxD		SGxxxTxxxN		GDD		GLTxxxDK		LKR, E	
	Motif G	Motif F	Motif A	Motif B	Motif C	Motif D	Motif E	Motif F	Motif G	Motif H	Motif I	Motif J
PV	Xu <i>et al.</i> (34)	Ferrer-Orta <i>et al.</i> (2)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Poch <i>et al.</i> (4)	Ferrer-Orta <i>et al.</i> (2)
HRV	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)
FMDV	Gorbalenya <i>et al.</i> (35)	Ferrer-Orta <i>et al.</i> (2)	Poch <i>et al.</i> (4)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)
NV	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)	Ferrer-Orta <i>et al.</i> (2)
RHDV	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Ferrer-Orta <i>et al.</i> (2)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Love <i>et al.</i> (12)	Ferrer-Orta <i>et al.</i> (2)
HCV	Pan <i>et al.</i> (5)	Choi <i>et al.</i> (17)	Bressanelli (36)	Pan <i>et al.</i> (5)	Bressanelli <i>et al.</i> (36)	Bressanelli <i>et al.</i> (36)	Bressanelli <i>et al.</i> (36)	Bressanelli <i>et al.</i> (36)	Bressanelli <i>et al.</i> (36)	Bressanelli <i>et al.</i> (36)	Bressanelli <i>et al.</i> (36)	Ferrer-Orta <i>et al.</i> (2)
BVDV	Pan <i>et al.</i> (5)	Choi <i>et al.</i> (17)	Pan <i>et al.</i> (5)	Choi <i>et al.</i> (17)	Choi <i>et al.</i> (17)	Choi <i>et al.</i> (17)	Choi <i>et al.</i> (17)	Choi <i>et al.</i> (17)	Choi <i>et al.</i> (17)	Choi <i>et al.</i> (17)	Choi <i>et al.</i> (17)	Bressanelli <i>et al.</i> (36)
HIV	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Choi <i>et al.</i> (17)
PH16	Pan <i>et al.</i> (5)	Butcher <i>et al.</i> (20)	Butcher <i>et al.</i> (20)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Butcher <i>et al.</i> (20)
REOV	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Pan <i>et al.</i> (5)	Butcher <i>et al.</i> (20)

MATERIALS AND METHODS

Sample selection and processing

In total, 18 well-studied viral species with solved RNA polymerase structures and four viral species with solved DNA polymerase structures were selected for analysis. We used the StralSV algorithm (<http://proteinmodel.org/>), described in detail previously (1), to perform the analyses. StralSV compares the R2R (structural) correspondence of each sequence in a set of reference sequences to a specified query sequence, beginning at the start of the sequence and continuing to the end, by evaluating successive overlapping segments of a user-selected length. Each of the selected structures was used as a query to all structures available in the Protein Data Bank (PDB release 2011_01_25; number of chains 176365) (8). The results were filtered for structural segments of at least 55% LGA_S structure similarity (1) to at least one query segment of 90 amino acids in length (size cutoff for the structural context) from which R2R correspondences were extracted from local tightly superimposed spans: continuous segments of the minimum length of five amino acids. These parameters together contributed to the identification of common regions of structure similarity, which were used to distinguish regions of conservation (structure matches) from regions in which structure deviates (non-matches). The StralSV comparison of each initial query to all structures in PDB resulted in the identification of a final set of 37 PDB structures that were used as a reference polymerase structure set in this study. One representative of each species within the set of 37 PDB structures was used to create an all-against-all structure comparison using StralSV. The species and PDB identities of this reference set are summarized in Table 1.

Creation of structure maps

The output from the all-against-all structure comparisons was parsed to extract R2R correspondences for each query/template pair. In each comparison the full query sequence was represented. At some positions in the template sequences, gaps occurred either due to structure deviation exceeding the alignment cutoff (5 Å) or because the template contained additional residues (e.g. a loop) without correspondence in the query structure.

A structure map was created for each set of R2R correspondences derived from each query/template-set StralSV comparison by combining the data for each binary alignment in an Excel spreadsheet. In this article, we report the structure map using poliovirus RdRp as the primary query (Figures 2–8), although in most cases we include structure maps with other species as query (Figures 2 and 4–8). Alternative query species were used when structure similarity of some of the templates was not identified with poliovirus as query (e.g. Motif G in WNV and DENV were identified based on DENV as the query). The query that contributes to non-poliovirus matches is indicated on each alignment ('q' following the species abbreviation).

The structure maps were used as the basis for the structure alignments described in this article. On all structure maps, we identified the Motifs A–F as described by Gong and Peersen (9) by coloring the background of the columns matching the residues of the motif orange and the background of the columns matching the highly conserved residues yellow. A similar coloring scheme was used for Motif G, except that it depicts the residues identified by Pan *et al.* (5) because Motif G was not specified in the Gong and Peersen study (9). On all structure maps, we colored the residues of the picornaviruses blue; the caliciviruses green; HCV and BVDV (flaviviruses) black; WN and DENV (flaviviruses) red; PHI6 black; REOV brown; ROTAV, IBDV, IPNV black; HIV purple; TERT, T7RNAP, N4 black; TAQ turquoise and T7DNAP black.

The segment of conserved structure adjacent to each motif was determined from the StralSV maps. For each of the queries and each of the motifs, the location at which structure conservation of most species became discontinuous was noted. We defined the boundaries of a homomorph as the position at which the structural segment shared by all representatives in a set became discontinuous in more than two species. In all structure maps, the conserved segments that we identified based on StralSV R2R correspondences are colored light blue. We defined the homomorph of each motif as the segment consisting of the conserved motif plus the adjacent structurally conserved segments. The length of each homomorph varied somewhat depending on the query. For each query species, the start and end of the homomorphic segment of each motif were recorded, and a 20 × 20 matrix for each motif was generated to compile the data from each query (data not shown). This matrix was used to identify the minimum start and maximum end of each homomorph, and these values are summarized in Supplementary Table S1. These values are plotted in Figure 1, which illustrates the maximal expanse of the homomorphic segments that include each of the polymerase motifs. All of the tertiary structures were illustrated using the Cn3D program (10).

RESULTS

Overview

Structural examination of the sequence motif regions yielded extended regions of structural conservation. We named each of these regions a ‘homomorph’, defined as a sequence segment that shares a highly similar tertiary structure with other species, independent of the sequence composition. We found that most of the homomorphs were at least twice as long as the corresponding sequence motif. The extent of this expansion is illustrated in Figure 1. The length of each homomorph was determined separately for each species, using a single structure for each species as the query in a StralSV analysis. The identity of the start and end of each homomorph depends on the structural similarity to a given query. Therefore, there is some minor query-specific variability of the location of the ends of homomorphs that

can be observed in Figure 1. Within the homomorphs, non-matching residues can be used to identify minor differences between species. In most single-stranded RdRp (ss-RdRp) species (PV, COXS, HRV, FMDV, NV, RHDV, SAPV, HCV, BVDV, WN and DENV), the homomorphs of Motif G are the largest (median of 53 residues), followed by A (49), B (46), E (37), F3 (28), D (23), C (17), F1 (10) and F2 (8). In double-stranded RdRps (ds-RdRps) (PHI6, REOV, ROTAV, IBDV and IPNV), the homomorphs of Motif G are relatively short (median of 12 residues). Several other homomorphs are also shorter in ds-RdRps: B (48), A (41), E (31), F3 (28), D (16), C (12), F1 (11) and F2 (2). The lengths and occurrences of homomorphs of polymerases that are associated with DNA (HIV, TERT, TAQ, T7 DNAP, T7 RNAP and N4) are variable and will be discussed in the sections describing each motif.

The homomorphs of all species are similarly distributed over the length of the polymerase (Figure 1). The ss-RdRps (PV, COXS, HRV, FMDV, NV, RHDV, SAPV, HCV, BVDV, KUNJ and DENV) are most similar to each other. The spacing between homomorphs is more variable in the ds-RdRps (PHI6, REOV, ROTAV, IBDV and IPNV), and in general larger than in the ss-RdRps. The homomorph of Motif C (hmC) is identified in the birnaviruses despite a sequence inversion that places it before Motif A (5). Relatively large segments between homomorphs occur in PHI6 between F1 and F3, and in KUNJ, DENV and PHI6 between B and C. The spacing between motifs is notably reduced in HIV and TERT (RdDps). In birnaviruses (IBDV, IPNV), the homomorphs of C and A are only three residues apart, and the distance between the homomorphs of Motif F3 and C is greater than the typical F3-B distance. Most homomorphs are separated from each other by a segment that contains a turn (secondary structure), or there is a turn at the beginning or end of the homomorph.

Within all RdRps, all motifs occur within a length of 375 residues. In T7-DdRp (T7 RNAP) and N4, the motifs are spread out over approximately 600 residues. The amount of R2R correspondence for most of the RdRps, determined from the minimum and maximum values of all homomorphs, is ~75% over the span from Motif G through Motif E (Supplementary Table S1).

Homomorph of Motif G results

The structurally aligned sequences that comprise hmG are summarized in Figure 2A.

The R2R correspondences of WN and DENV could not be evaluated for the Motif G region (approximately PV 101–121), as the structural configuration of the segments of these viruses that would be expected to match the homomorph of Motif G (hmG) segment had not been determined. Within the homomorph, most of the ss-RdRps were highly similar (Figure 2A, top). BVDV is similar to the other ss-RdRps in the N-terminal segment, but no longer matches them at the C-terminal segment. Only Motif G, and not a homomorph, was identified in PHI6, IBDV and IPNV. In the region of Motif G, StralSV did not identify R2R correspondences between any of the

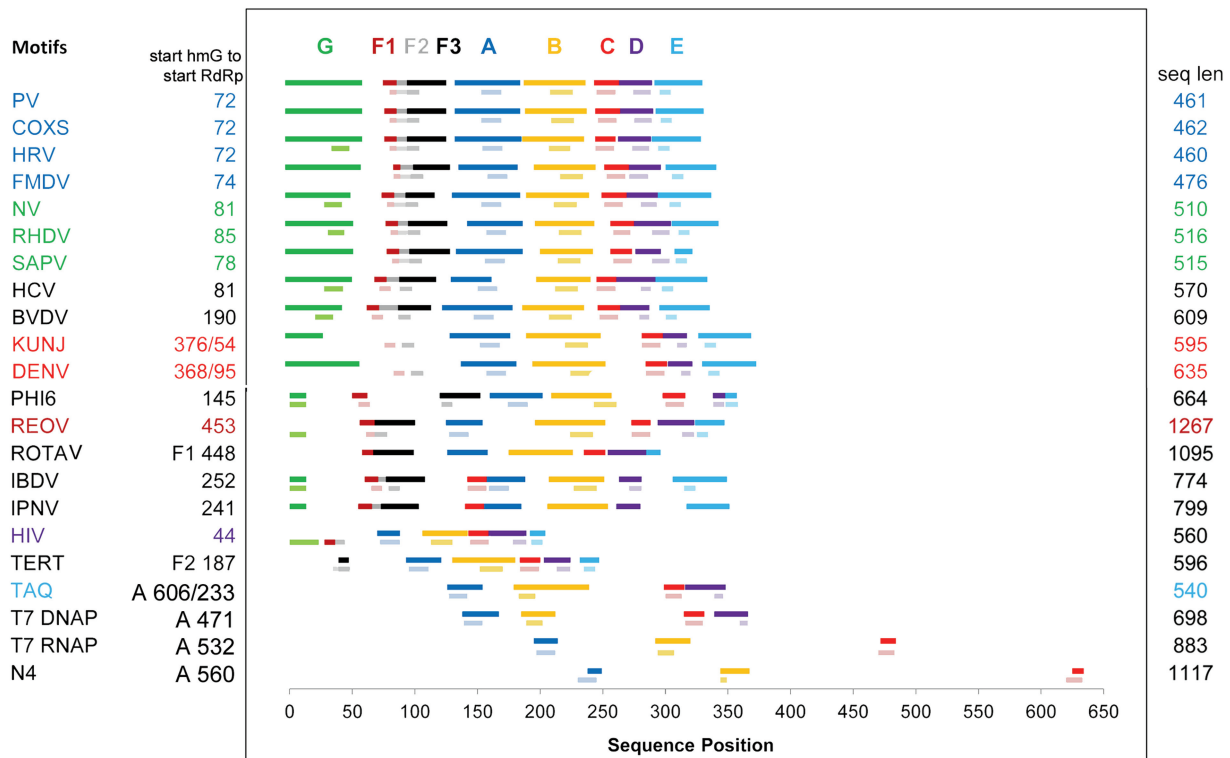


Figure 1. The distributions and lengths of the homomorphs and conserved motifs of viral polymerases in this study. The homomorphs of each species are illustrated from the start of Motif G, sequence position = 0, through the end of Motif E. The homomorphs and motifs are colored as follows: G (green), F1–F2–F3 (maroon–gray–black), A (blue), B (gold), C (red), D (purple), E (aqua). The darker bars show the sequence positions of the homomorphs, and the lighter bars show the sequence positions of the motifs as described by Gong and Peersen (9) or Pan *et al.* (5). The number of residues from the start of the polymerase structure to the start of the first homomorph is identified for each species at the left of the chart. The PDB structures, and consequently, sequence position numbers for KUNJ, DENV and TAQ, which are used throughout this article, do not begin at the polymerase; therefore, for this figure, the distance from the start of the polymerase is shown after the slash. For species lacking Motif G, the first identified homomorph is indicated at the left of the start position. The length of the polymerase of each species is listed at the right of the chart.

RNA polymerases and REOV, ROTAV, HIV, TERT or DNA-dependent polymerases (TAQ, T7 DNAP, T7 RNAP and N4).

There were structural discontinuities within the homomorph (noted by x in Figure 2A) and similar discontinuities within the motif. These minor discontinuities identify species-specific differences within a segment that is otherwise highly continuous in several species. For example, FMDV has 3 AA between S87 and T90 (PV numbering) and therefore does not match the structure of PV 88-LD-89, which has only 2 AA. In contrast, WN and DENV have the same number of residues for the gaps from A388–R392 and G385–R389, respectively, but these segments were not structurally aligned by StralSV within the parameters used in this study. The segment PV-Y102 to A109 is a β -hairpin unique to picornavirus RdRps(11). The numbering on NV and HCV clarifies that these regions are continuous in these species (and the other caliciviruses, RHDV and SAPV, though not numbered).

Figure 2B and C illustrates the tertiary structure of the homomorph using a poliovirus structure (PDB:1RA6). Most of the N-terminal segment is a single helix that extends over nearly half of the surface of the protein.

Both ends of the homomorph terminate at the exterior surface of the protein.

The distance between the homomorphs of Motifs G and F1 was 20–37 residues in the ss-RdRps (in all species where both were present) and longer in the ds-RdRps (median 47 residues) (Figure 1).

Homomorph of Motif F results

Three components of Motif F have been recognized: F1, F2 and F3 (2,5). In some species there are sequence segments between these motifs. In all the species in our sample set except PHI6, in those species that have R2R correspondence within Motif F, the three F motifs are continuous; therefore, we have combined them, and the adjacent structurally aligned segments, into a single homomorph. The structurally aligned sequences that comprised homomorph of Motif F (hmF) for RdRps and HIV are summarized in Figure 3A. HmF extended five residues upstream from the N-terminal edge of Motif F1 [as defined by Gong and Peersen (9)] and ~20 residues downstream from Motif F3 [as defined by Gong and Peersen (9)]. HmF1 was found in all RdRp species except WN and DENV; it was not possible to evaluate R2R correspondence for this segment of WN and

A Structure map of the homomorph of Motif F (hmF)

QUERY	start	Motif F1	end	start	Motif F2	end	start	Motif F3	end
PV, Gong [18]	157- x x x x x	Y V K D E L	-162	163- R S K T	K V	-168	171- G K S R L I E		-177
NV, Pan [11]							180- K K R L L W G S D		-188
HRV, Pan [11]							172- K T R V I E A S S		-180
PV	152- L P L V T	Y V K D E L	-162	163- R S K T	K V E Q	-170	171- G K S R L I E	A S S L N D S V A M R M A F G N L Y A A F	-198
COXS	152- L P M V T	Y V K D E L	-162	163- R S K T	K V A K	-170	171- G K S R L I E	A S S L N D S V A M R Q T F G N L Y K T F	-198
HRV	152- L P M I T	F L K D E L	-162	163- R K K T	K I A S A	-170	171- G K T R V I E	A S S I N D T I L F R T V G N L F S K F	-198
FMDV	157- F A C Q T	F L K D E L	-167	168- R P M E	K V R A	-175	178- T R I V D	L P V E H I L Y T R M M I G R F C A Q M	-203
NV	159- P V Y T G	A L K D E L	-169	170- V K T D	K I Y G K	-178	179- I K K R L L W	G S D L A T M I R C A R A F G G L M D E L	-206
RHDV	167- x I Y A C	G L K D E L	-176	177- R P L D	K V K E	-184	185- G K K R L L W	G C D V G V A V C A A A V F H N I C Y K L	-212
SAPV	161- x A Y K L	A L K D E L	-170	171- R P I E	K N K A	-178	179- G K K R L L W	G C D A A T T L I A T A A F K A V A T R L Q	-207
HCV	135- x D T T I	M A K S E V	-144	145- F C v q	p e k g r	-154	155- K P A R L I V	x P D L G V R V C E K M A L Y D V V S	-180
BVDV	257- x Y E T A	I P K N E K	-266	267- r d	vs d d w q a g d l v v e	-281	282- K R P R V I Q	Y P E A K T R L A I T K V x Y N	-304
WN							473- x S R A I W	F M W L G A R F L E F E A L G F L N E D H	-498
DENV							470- x S R A I W	F M W L G A R Y L E F E A L G F L N E D	-494
PHI6	197- G A Y Y V	V Y R A	-205	266- c			267- E R R R T A M	G G P F A L N A P I M A V A Q P V R N K I Y	-295
REOV	511- A D T S M	G L R N Q V	-521	522- q			523- R R P R S I M	P L N V P Q Q Q V S A P H T L T A D Y I N Y	-551
ROTO	448- x x x	P L G R R D V F	-455	456- g			457- R R T R I I E	I L P Y E Y F I A Q H A V V E K M L I Y A K	-485
IBDV	314- S C G L L	F P K A E R	-324	325- Y D K S T W L		-331	333- x K T R N I W	S A P S P T H L M I S M I T W P V M S N	-358
IPNV	298- S C G L L	F P K G E R	-308	309- Y T Q V D W D		-315	316- K K T R N I W	S A P Y P T H L L S M V S T P V M N E	-342
HIV	58- x x	T P V F A I	-63	67- d s		-68	69- T K W R K L W	D F R E L N K R T Q D F W	-88

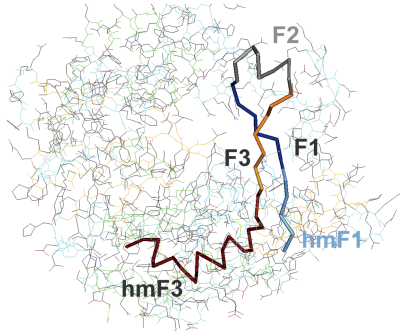
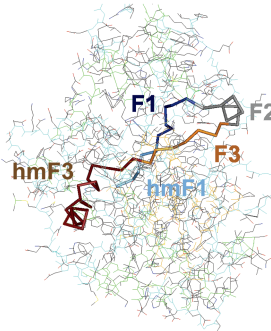
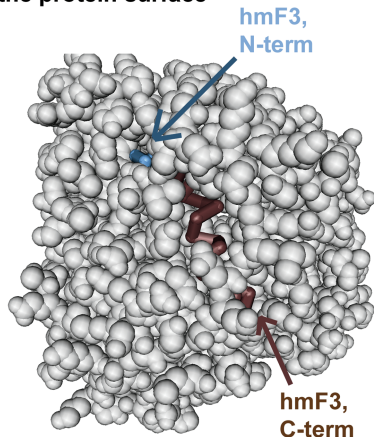
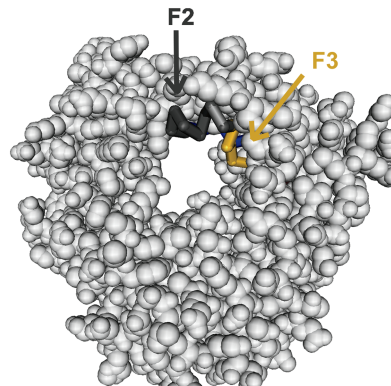
B. HmF in poliovirus [PDB:1RA6]**C. HmF transects the protein****D. Terminal residues of hmF at the protein surface****E. Motif F2 is at the apex of the hmF fold**

Figure 3. (A) The residues of Motifs F1 and F2, according to Gong and Peersen (9), are indicated by an orange background and the highly conserved residues by a yellow background. HmF includes the residues with light blue background, and Motifs F1, F2 and F3. Lower case residues indicate that they are present but do not match the structure. (B) HmF1 (light blue), illustrated using the poliovirus structure 1RA6, begins about five amino acids upstream from Motif F1 (dark blue) and ends at hmF3 (dark brown), about 20 amino acids downstream from Motif F3 (orange). Highly conserved residues from Motifs F1, F2 and F3 line the template tunnel (empty space within the crystal near the intersection of F1–F2–F3). Motif F2, which varies in length and composition between species, is at the exterior of the protein. (C) Most of hmF is folded back on itself, and Motif F2 (gray) is at the apex of this fold at the exterior surface of the protein. (D) The terminal residues of hmF (blue) are located at the surface of the protein. In addition to the N-terminal (blue) and the C-terminal (brown) of hmF, several residues of the C-terminal segment (brown) 186-AMRMA-190 are also at the surface of the protein. (E) The length of Motif F2 (black) is highly variable in different species. Motif F3 (gold) is the start of a highly conserved segment of the homomorph that transects the protein and terminates at a nearly opposite exterior surface.

and IPNV (ds-RdRps) have fewer aligned residues. REOV (ds-RdRp) and HIV and TERT (RdDps) do not have R2R correspondence with the ss-RdRps. The DdRps (T7 RNAP and N4) and DdDps (TAQ and T7DNAP) share a homomorphic structure within the N-terminal segment, but it is substantially different from the RdRp structure and therefore is not included in the homomorph or Figure 4A. Within the motif, HIV corresponds only to NV and SAPV (only found using an HIV query),

indicating a significant structural difference from other species; HIV also lacks R2R correspondence beyond the motif and therefore is not included in the homomorph. At the C-terminal segment of the homomorph, most species in the sample set, except HIV, have a homologous structure. At some sequence positions within hmA, a particular residue composition is conserved throughout a viral family (e.g. picornavirus), and a different residue composition is conserved in another viral family at the same

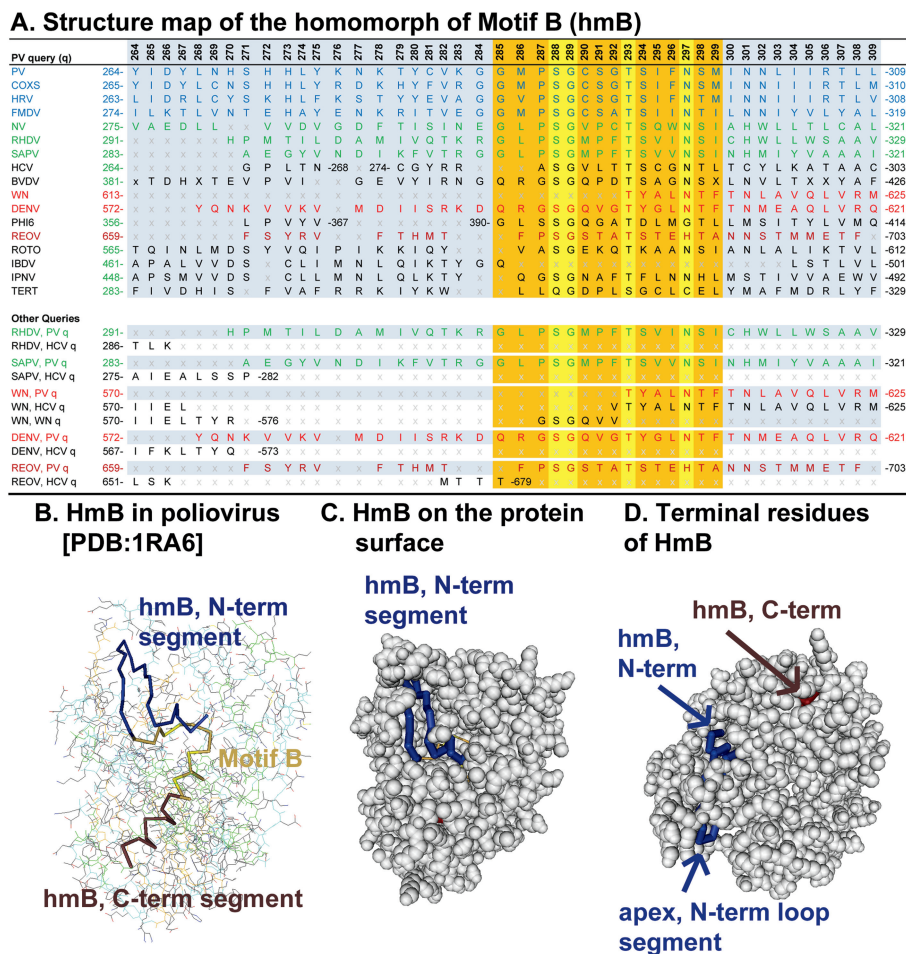


Figure 5. (A) A StralSV alignment based on the PV query is located at the top section of this figure. The middle section compares some of the R2R matches using other queries, to those found by poliovirus. Residues common to both alignments indicate extremely close matches, and those differently represented are closer to the respective query. The species TAQ-N4-T7 DNAP-T7 RNAP-T7 RNAP align with each other over about 20 residues, but do not align structurally with the other species. (B) HmB [illustrated using poliovirus (PDB:1RA6)] is the largest homomorph in the RdRps. The N-terminal segment of HmB (blue) begins at the exterior surface of the protein, folds back on itself to form a classical β -hairpin (at the apex, PV 275-YKN-278), and then continues to Motif B (gold), which is located at the surface of the template tunnel. The C-terminal segment of the homomorph (brown) continues as a single chain to the exterior surface of the protein. (C) The N-terminal segment of HmB is at the surface of the protein. (D) Each of the terminal residues of HmB, and the apex of the N-terminal loop are on the exterior surface of the protein.

these species matched a poliovirus query, indicating there is greater structural similarity than in other homomorphs and motifs. The N-terminal segment of the homomorph contains some discontinuities that are resolved by using R2R matches for alternative queries (Figure 5A, lower section). The C-terminal segment of the homomorph is well represented in all RNA polymerases and TERT. No R2R correspondence was found between the residues comprising hmB in the RNA polymerases and residues in the DNA-dependent polymerases (T7 RNAP, N4, TAQ and T7 DNAP).

The lower section of Figure 5A illustrates the dependence of the R2R correspondence on the query sequence. These differences make it possible to identify fine details between structures. Our definition of each of the homomorphs, however, is based on the inclusion of all R2R alignments using all queries in the sample set.

The position of the hmB within the tertiary structure of PV is illustrated in Figure 5B. The N-terminal residue is at

the exterior surface of the protein. The N-terminal segment is a classical β -hairpin protein structure that is folded back on itself and is almost entirely exposed on a surface nearly perpendicular to the face of the protein that contains the N-terminal residue (Figure 5C). The base of the loop transitions to Motif B at the template tunnel. The C-terminal side of the homomorph extends from the tunnel to the exterior surface of the protein (Figure 5D).

The distance between the homomorphs of Motifs B and C (hmC) (Figure 1) is <6–17 in all RdRps except KUNJ and DENV, which are 36 and 35 residues, respectively. In the DNA-dependent polymerases, this distance is between 98 (TAQ) and 258 (N4) residues. In IBDV and IPNV, the segments between the homomorphs of Motifs B and D are 16 and 11 residues, respectively.

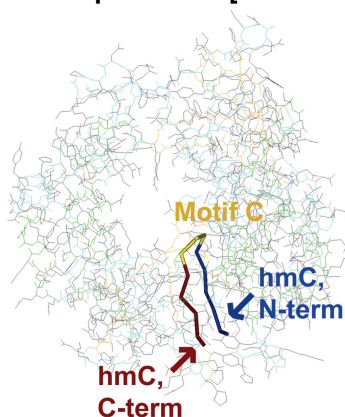
HmC results

The structurally aligned sequences that comprise hmC are shown in Figure 6A. Motif C is the only RdRp motif that is not a component of a larger homomorphous structure.

A. Structure map of the homomorph of Motif C (hmC)

PV query (q)	
PV	322- K M I A Y G D D V I A S Y -334
COXS	323- R M I A Y G D D V I A S Y -335
HRN	321- K I I A Y G D D V I F S Y -333
FMDV	332- T M I S Y G D D I V V A S -344
NV	337- L F S F Y G D D E I V S T -349
RHDV	348- P F Y T Y G D D G V Y x x -358
SAPV	341- T I H T Y G D D C M Y x x -324
HCV	312- T M L V N G D D L V V I C -324
BVDV	442- R I H V C G D D G x x x x -450
WN	663- x M A V S G D D C V V x x -672
DENV	658- x M A I S G D D C V V x x -667
PHI6	448- x Q I S K S D D A M L G W -459
REOV	729- x Y V C Q G D D G L x x x -737
ROTAV	626- x I R V D G D D N x x x x -633
IBDV	398- x x L V Y A D N I Y I x x -406
IPNV	384- x x M I Y A D N I Y I x x -392
HIV	179- V I Y Q Y M D D L Y V G S -191
IPNV q	
IBDV	396- K A L V Y A D N I Y I V H -408
IPNV	382- L V M I Y A D N I Y I L Q -394
TERT q	
TERT	337- F I H R T V D D Y F F C S -349
T7 DNAP q	
HCV	312- T M L V x x x L V V I C E -325
ROTAV	624- K I I x x x x N Y A V L Q -638
HIV	181- Y Q Y M x x x D D L Y V G -190
TERT	342- x x x x V x x D D Y F F C -348
N4	948- x x K I F x x D G M N I G -946
TAQ	780- L L Q V H x x D E L V L E -790
T7 DNAP	649- M A W V H x x D E I Q V G -659
T7 RNAP	809- x x L I H x x D S F G T I -817

B. HmC in poliovirus [PDB:1RA6]



C. Terminal residues of hmC

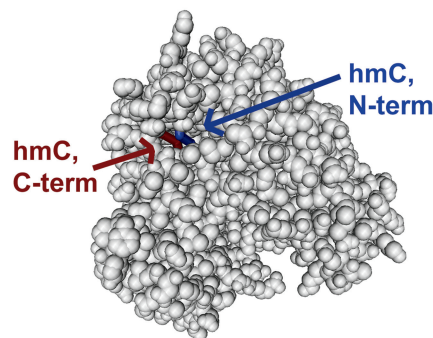


Figure 6. (A) The high number of species that align to PV indicates that the structure of Motif C is highly conserved. Although a T7 DNAP query was required to identify the matches for the N4-TAQ-T7 RNAP species, it was achievable. HmC is the only homomorph for which there is R2R correspondence in all species of the study group. (B) Motif C (gold) is the only motif in the RdRps that is not a component of a larger structure. Motif C [illustrated using poliovirus (PDB:1RA6)] is tightly folded upon itself in a manner that places the highly conserved residues (yellow) at the tunnel wall, whereas the N-terminal segment of the motif (blue) and C-terminal segment of the motif (brown) are parallel to each other and penetrate the protein. (C) The terminal residues of both the N- and C-terminal segments are at the surface of the protein.

The segments immediately adjacent to both flanks of Motif C do not even cluster into subgroups. Motif C is short—12 residues in most RdRps and folds sharply back on itself (Figure 6B). The highly conserved residues (labeled Motif C) are at the surface of the template tunnel and both the N-terminal and C-terminal residues are at the exterior surface of the protein (Figure 6C).

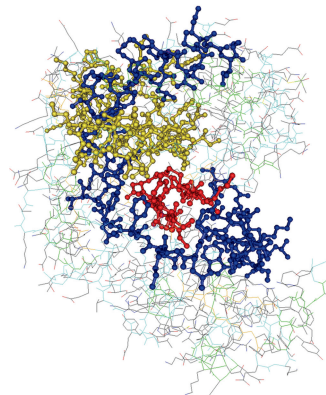
In the birnaviruses IBDV and IPNV, there is a sequence inversion that results in the relocation of Motif C to a position immediately preceding Motif A. Figure 7A shows an alignment that documents this inversion. The top and bottom segments of Figure 7A illustrate that all species are well aligned upstream of Motif C (IPNV positions 365–372) and within Motif A (IPNV positions

399–409). RHDV, SAPV and BVDV are not well aligned within Motif C using the IPNV query, and therefore are missing from the middle section of Figure 7A (IPNV positions 382–393). The numbering of IPNV and IBDV is sequential, indicating that Motif C precedes Motif A in these species. The numberings of NV and HCV indicate there are R2R matches with IPNV at Motif C, but that over this segment the match is not in sequential order. Using a PV query, however, all of these species have R2R matches over this segment (shown in Figure 6A). The IPNV query indicates that the structure of Motif C of the birnaviruses more closely matches NV and HCV than the others in the sample set. The difference in linear order that results from the sequence inversion is

A. Structure map of Motif C using a birnavirus query

IPNV seq pos	IPNV	NV	RHDV	SAPV	HCV	BVDV	IBDV
365	M	G.225	V.232	M.226	V.201	F.326	L.382
366	D	P.226	D.233	Q.227	E.202	D.327	N.383
367	R	I.227	V.234	V.228	F.203	K.328	R.384
368	I	I.228	I.235	M.229	L.204	V.329	I.385
369	M	F.229	I.236	N.230	V.205	R.330	V.386
370	T	E.230	N.237	D.231	N.206	K.331	E.387
371	I	R.231	N.238	S.232	T.207	E.332	W.388
372	I	H.232	L.239	L.233	W.208	W.333	I.389
373	R		T.240		K.209	D.334	L.390
374	D		S.241		S.210	S.335	
375	S		K.242				
376	L						
377	D						
378	N						
379	D						
380	E						
381	D				C.311		
Motif C - 382	L	L.337			T.312		K.396
383	V	F.338			M.313		A.397
384	M	S.339			L.314		L.398
385	I	F.340			V.315		V.399
386	Y	Y.341			N.316		Y.400
387	A	G.342			G.317		A.401
388	D	D.343			D.318		D.402
389	N	D.344			D.319		N.403
390	I	E.345			L.320		I.404
391	Y	I.346			V.321		Y.405
392	I	V.347			V.322		I.406
393	L				I.323		V.407
394	Q						H.408
395	D						S.409
396	N						N.410
397	T				V.340		T.411
398	W				A.341		W.412
Motif A - 399	Y		Y.239	F.217	V.342	Y.413	
400	S	D.240	C.240	S.218	S.343	S.414	
401	I	A.241	L.241	Y.219	F.344	I.415	
402	D	D.242	D.242	D.220	D.345	D.416	
403	L	Y.243	Y.243	T.221	T.346	L.417	
404	E	S.244	S.244	R.222	K.347	E.418	
405	K	R.245	W.254	K.245	C.223	A.348	K.419
406	G	W.246	D.255	W.246	F.224	W.349	G.420
407	E	D.247	S.256	D.247	D.225	D.350	E.421
408	A	S.248	T.257	S.248	S.226	T.351	A.422
409	N	T.249	M.258	T.249	T.227	Q.352	N.423

B. Homomorphs A-B-C in poliovirus [PDB:1RA6]



C. Homomorphs C-A-B in IBDV [PDB:2PGG]

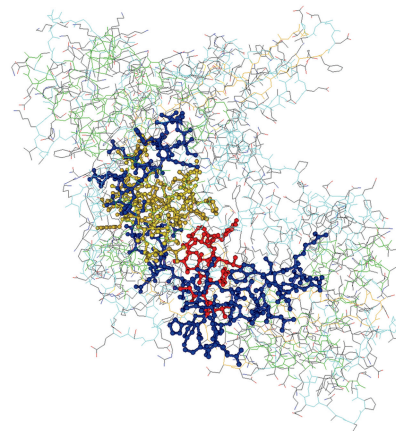


Figure 7. (A) The sequence of the IPNV query is listed vertically in this table, with the sequence position number of each IPNV residue at the left in the first column. The R2R corresponding residues for each species are listed in the columns to the right. In birnaviruses (IBDV and IPNV), unlike other RdRps, Motif C precedes Motif A. The top segment of this figure (IPNV positions 365–372) shows that all species (that align to IPNV) are well aligned prior to Motif C. IPNV, NV, HCV and IBDV are aligned in Motif C (IPNV positions 382–393), but there are no R2R matches for RHDV, SAPV and BVDV. In Motif A, at the bottom segment of the figure (IPNV segments 399–409) all species are again, well aligned (though fewer residues are aligned in RHDV). (B) In poliovirus (illustrated here with PDB:1RA6), the sequence order of homomorphs is: A (blue), B (gold) and C (red). (C) In IBDV, the difference in sequence order is compensated by a modified structure that maintains the motifs within a tertiary position that is similar to all other RdRps. The sequence order of the IBDV homomorphs is: C (red), A (blue) and B (gold).

compensated by a modified structure that maintains the motifs within a tertiary position that is similar to all other RdRps (Figure 7B and C).

The distance between the hmC and hmD (homomorph of Motif D) is <10 residues in the RdRps. It is relatively large in PHI6 (25 residues) and is indeterminate in the DdDps, as neither Motif D nor its homomorph is within the PDB structures included in this study.

HmD results

The structurally aligned sequences that comprise the hmD are shown in Figure 8A. The homomorph is 21 residues long and consists of a 10-residue extension from the N-terminal edge of the motif plus the motif itself. The structure of the N-terminal segment is more highly conserved (i.e. has more R2R matches) than the motif. Various query sequences were tested with the expectation that they would capture additional alignments. The

middle section of Figure 8A illustrates that this produced some improvement. For example, using an HCV query, there are R2R matches to TERT, TAQ and T7 DNAP. The C-terminal edge of the motif has some R2R correspondence, suggesting that the structure of the motif is moderately conserved. Using T7 DNAP as a query (lowest segment of the figure), only a small portion of the C-terminal edge of Motif D and a few species have similar structures. There is no alignment of PHI6 within the N-terminal segment of the homomorph, because in this region PHI6 consists of a 24-residue loop between the end of Motif C and the start of Motif D.

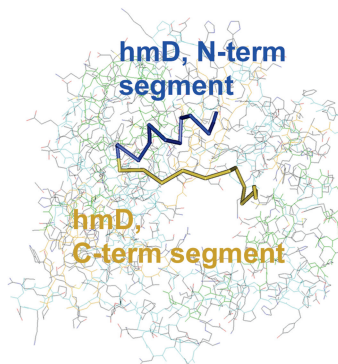
The tertiary structure of the hmD is illustrated in Figure 8B and C. This homomorph lies mostly at the exterior surface of the protein. The motif lines the wall of the polymerase tunnel.

The segment between the hmD and homomorph of Motif E (hmE) is <15 residues in all structures in the

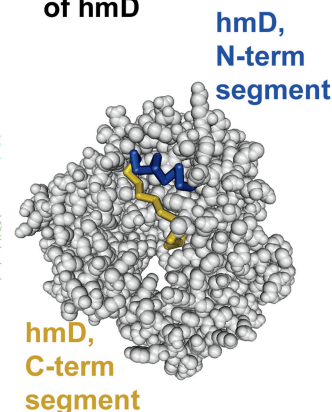
A. Structure map of the homomorph of Motif D (hmD)

Motif D, PV query (q)	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361		
PV	340-	A	S	L	L	A	Q	S	G	K	D	Y	G	L	T	M	T	P	A	D	K	S	A	-361
COXS	341-	A	S	L	L	A	E	A	G	K	G	Y	G	L	I	M	T	P	A	D	K	G	E	-362
HRV	339-	M	E	A	I	A	N	E	G	K _o	K	Y	G	L	T	I	T	P	A	D	K	S	T _s	-360
FMDV	350-	F	E	A	L	K	P	H	F	K	S	L	G	Q	T	I	T	P	A	D	K	S	D	-371
NV	355-	P	E	K	L	T	A	K	L	K	E	Y	G	L	K	P	T	R	P	D	K	T	E	-376
RHDV	368-	L	P	A	I	I	E	N	L	R	D	Y	G	L	S	P	T	A	x	D	K	T	E	-389
SAPV	363-	x	x	T	V	L	A	N	L	T	S	Y	G	L	K	x	x	x	x	x	x	x	x	-374
HCV	336-	L	R	A	F	T	E	A	M	T	R	Y	S	A	P	P	x	x	x	x	x	x	x	-350
BVDV	463-	A	N	K	G	x	Q	I	L	H	E	A	G	K	P	Q	x	x	x	x	x	x	x	-477
WN	684-	x	x	x	x	x	H	F	L	N	A	M	S	x	x	x	x	x	x	x	x	x	x	-690
DENV	679-	x	x	x	x	x	L	A	L	N	D	M	G	x	x	x	x	x	x	x	x	x	x	-685
PHI6	487-	x	x	x	x	x	x	x	x	x	x	x	x	K	I	S	Y	E	x	x	x	x	x	-491
REOV	757-	L	E	L	I	S	K	Y	G	E	E	F	G	W	K	Y	D	x	x	x	x	x	x	-772
ROTA	652-	S	N	D	V	R	E	T	Y	A	R	M	N	A	K	V	K	A	x	x	x	x	x	-668
IBDV	518-	x	x	x	x	x	K	S	I	E	D	x	x	I	N	F	K	I	x	x	x	x	x	-530
IPNV	504-	x	x	x	F	M	D	L	E	E	K	T	G	I	N	F	K	I	E	x	x	x	x	-518
HIV	202-	I	E	E	L	R	Q	H	L	L	R	W	G	L	T	T	P	D	x	x	x	x	x	-218
HCV, Gong [12]	347-												S	A	P	P	G	D	x	x	x	x	x	-351
HCV, HCV q	336-	L	R	A	F	T	E	A	M	T	R	Y	S	A	P	P	G	D	D	P	Q	P	E	-357
BVDV, HCV q	463-	A	N	K	G	x	Q	I	L	H	E	A	G	K	P	Q	K	I	x	x	K	V	A	-488
WN, HCV q	684-	x	x	x	x	x	H	F	L	N	A	M	S	K	V	R	x	x	x	x	x	x	W	-705
DENV, HCV q	679-	x	x	x	x	x	L	A	L	N	D	M	G	K	V	R	x	x	x	x	K	G	W	-700
PHI6, PHI6 q	485-	x	x	x	x	x	x	x	x	x	x	x	Y	M	K	I	S	Y	E	H	x	x	x	-492
ROTA, HCV q	652-	S	N	D	V	R	E	T	Y	A	R	M	N	A	K	V	x	x	A	L	V	S	-671	
HIV, HCV q	202-	I	E	E	L	R	Q	H	L	L	R	W	G	L	T	T	P	D	K	K	x	x	x	-220
TERT, HIV q	360-	V	Y	D	F	E	L	L	I	K	G	x	x	x	x	x	x	x	x	x	x	x	x	-365
TAQ, HCV q	800-	A	R	A	L	K	E	V	M	E	G	V	x	x	x	x	x	x	V	D	I	D	-824	
T7 DNAP, HCV q	670-	I	E	Y	A	Q	E	A	M	R	W	V	G	x	x	x	x	x	G	K	M	G	-697	
T7 DNAP q																								
HCV	351-	x	x	P	Q	P	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	-356
ROTA	667-	K	A	L	V	S	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	-671
TAQ	819-	V	E	V	G	I	G	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	-824
T7 DNAP	692-	T	E	G	K	M	G	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	-697

B. HmD in poliovirus [PDB:1RA6]



C. Surface exposure of hmD



D. Terminal residues of hmD

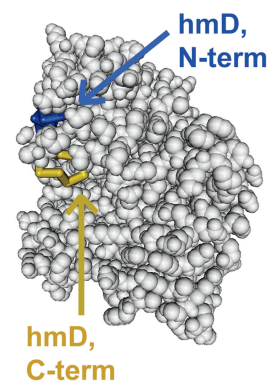


Figure 8. (A) The structure of the N-terminal segment of hmD (blue background) is more conserved than that of Motif D (orange background). There is a high amount of conservation of some residues (pink background) within the motif. Residues consistent with the motif described by Gong and Peersen (9) are bold. In the middle section of the figure, different queries were used to identify more of the R2R correspondences within the motif. In the lowest section of the figure, the alignment is based on a T7 DNAP query. There is an off-by-one alignment of ROTAV (compare lines labeled 'ROTA, HCV q' and 'ROTA, T7 DNAP q', both italicized), suggesting an unusual structural conformation. (B) The homomorph of Motif D [illustrated using poliovirus (PDB:1RA6)] includes the motif itself (gold) and an adjacent upstream segment (blue). The motif lines the wall of the template tunnel. (C) Most HmD residues are located on the surface of the protein. (D) The terminal residues of hmD are located at the exterior surface of the protein, on a different face than the main section of the homomorph.

sample set, except in IBDV and IPNV in which it is 28 and 40 residues, respectively.

HmE results

The structurally aligned sequences that comprise hmE are summarized in Figure 9A. HmE is large and in most of the ss-RdRps (PV, COXS, HRV, FMDV, NV, RHDV,

SAPV, HCV, BVDV, WN and DENV) it is highly conserved. The motif is near the N-terminal edge and a loop region is located near the middle of the homomorph. The sequences vary in length due to the loop region. The length of hmE in the caliciviruses (30–34 residues) is shorter than those in the picornaviruses (36–37 residues); HCV and BVDV loops are 37 and 35 residues, respectively, and the loops of WN and DENV are the longest at

A. Structure map of the homomorph of Motif E (hmE)

PV query (q)	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404			
PV	368-	W	E	N	V	T	F	L	K	R	R	F	F	R	A	D	E	K	Y	P	F	L	I	H	P	V	M	P	M	K	E	I	H	E	S	I	R	W	T	-404
COXS	369-	W	T	N	V	T	F	L	K	R	Y	F	R	A	D	E	Q	Y	P	F	L	V	H	P	V	M	P	M	K	D	I	H	E	S	I	R	W	T	-405	
HRV	367-	Y	N ₀	N	V	T	F	L	K	R	Q	F	R	Q	D	E _d	K	H _y	T _k	F	L	I	H	P	T	F	P	V	E	E	I	Y	E	S	I	R	W	T	-403	
FMDV	380-	I	T	D	V	T	F	L	K	R	H	F	H	M	D	Y	G	T	-396	x	399-	Y	K	P	V	M	A	S	K	T	L	E	A	I	L	S	F	A	-415	
NV	385-	L	N	G	L	T	F	L	R	R	T	V	T	R	D	-398	x	x	x	402-	W	F	G	K	L	E	Q	S	S	I	L	R	Q	M	Y	W	T	-418		
RHDV	396-	L	N	K	I	S	F	L	K	R	T	F	E	L	-408	x	x	x	x	412-	G	W	V	S	K	L	D	K	S	S	I	L	R	Q	L	E	W	S	-429	
SAPV	391-	x	x	x	x	V	F	L	K	R	T	F	T	Q	T	-400	x	x	x	404-	I	R	A	L	L	D	I	T	S	I	T	R	Q	F	Y	W	L	-420		
HCV	360-	L	E	L	I	T	S	C	S	S	N	V	S	V	A	-373	x	x	x	381-	V	Y	L	T	x	D	P	T	T	P	L	A	R	A	A	x	x	-396		
BVDV	492-	x	E	D	I	E	F	C	S	H	T	P	V	V	R	-505	x	x	x	512-	S	H	X	A	G	x	D	T	A	V	I	L	S	K	x	x	-526			
WN	708-	W	Q	Q	V	P	F	C	S	H	H	F	T	E	L	I	-722	x	x	x	x	x	x	x	x	735-	G	Q	D	E	L	V	G	R	A	R	I	x	-745	
DENV	703-	W	Q	Q	V	P	F	C	S	H	H	F	H	E	L	I	M	-718	x	724-	L	V	V	P	C	x	P	Q	D	E	L	I	G	R	A	R	I	S	-741	
PH16	495-	x	x	x	x	A	F	L	G	D	I	x	x	x	x	x	x	x	x	x	x	x	x	V	G	N	I	N	S	M	L	N	N	Q	F	x	-526			
REOV	779-	x	x	x	x	A	E	Y	L	K	L	Y	x	x	x	x	x	x	x	x	x	x	x	x	x	P	N	L	S	R	H	x	x	x	x	x	-798			
ROTAV, NV q	675-	x	x	x	x	A	E	I	A	K	R	Y	I	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	-682			
IBDV	571-	x	x	x	x	L	G	W	S	A	T	Y	S	K	D	L	G	-581	x	x	387	V	L	D	K	E	R	L	F	C	S	A	A	Y	x	-599				
IPNV	560-	x	x	x	x	L	L	G	W	S	A	I	Y	S	R	Q	M	E	-572	x	x	377	V	L	E	N	E	R	L	I	A	S	A	A	x	-588				

Only TERT-HIV match
TERT q
HIV, TERT q 225- X P P F L W M G Y E L H X -235
TERT, TERT q 385- X D E I P Y C G K I F N L -396

PV	368-	W	E	N	V	T	F	L	K	R	R	F	F	R	A	D	E	K	Y	P	F	L	I	H	P	V	M	P	M	K	E	I	H	E	S	I	R	W	T	-404
DENV	703-	W	Q	Q	V	P	F	C	S	H	H	F	H	E	L	I	M	x	x	x	L	V	V	P	C	x	P	Q	D	E	L	I	G	R	A	R	I	S	-741	

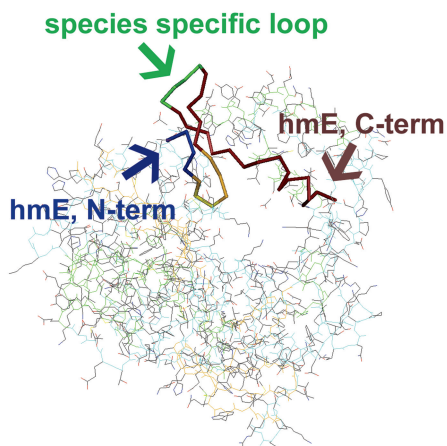
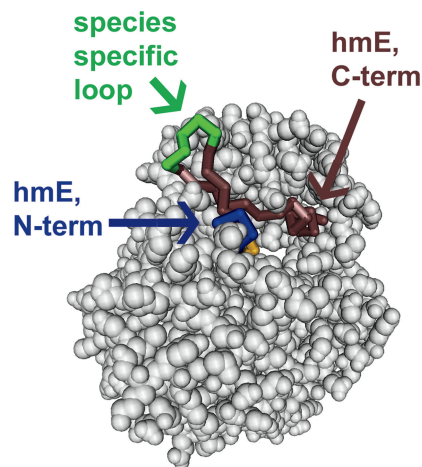
B. HmE in poliovirus [PDB:1RA6]**C. Terminal residues of hmE**

Figure 9. (A) Motif E is a component of a large, well-defined homomorph. There is considerable sequence similarity between PV and DENV in this homomorph, illustrated at the bottom of the figure. TERT and HIV have R2R correspondence over the motif segment and three amino acids on each side of it. HIV does not have any R2R correspondence with any species in the sample set in this region. (B) Motif E (gold), [illustrated using poliovirus (PDB:1RA6)] forms part of the NTP entry tunnel. Both the N-terminal segment of the homomorph (blue) and the C-terminal segment (brown) extend to the surface of the protein. (C) From the C-terminal edge of the motif, the homomorph extends to the surface of the protein to expose a species specific loop (of variable length) (green) at the surface, then turns, then transects the protein and extends to an opposite surface.

38 and 39 residues, respectively. There is strain-specific amino acid variability in this segment of HRV. HmE is well represented by all RdRps. No R2R correspondence was found with HIV or TERT. These species, however, are structurally matched to each other (Figure 9A, middle section). There is considerable sequence similarity between PV and DENV within this homomorph; this is illustrated in the bottom section of Figure 9A by the shaded conserved residues. DdRps and DdRps are not included in the analysis of this region because the region is missing from the structures in our sample group.

The tertiary structure of the hmE is illustrated in Figure 9B and C. Most of the homomorph is at the exterior of the protein near the NTP entry tunnel. Although it has extensive surface exposure, each terminus of the homomorph appears to be anchored by residues that are not part of the homomorph; as a result, the terminal residue at each end of the homomorph is

exposed as a single residue at the exterior surface of the protein. Motif E is located near the N-terminal edge of the homomorph and contacts the surface of the NTP entry tunnel (2). The C-terminal segment of the homomorph is folded back on itself in a manner that places the species-specific loop at the surface of the protein (Figure 8C). The homomorph forms a double strand through PV_M392, at which point the remainder of the homomorph is a single-stranded helix that emerges at the exterior surface of the protein. In PV, the C-terminal of hmE (R402) is exposed at the surface the protein and surrounded by the segment 28-SAFHYVFEG-36.

DISCUSSION

Structure-based sequence alignment using the StralSV algorithm (1) enabled us to identify seven distinct homologous structures in most of the polymerases in our

collection of 22 species. In the RdRps, the combined regions of structural homology represent ~75% of the sequence from the start of homomorph of Motif G (hmG) through the end of hmE in each species (~375 residues). There is <10% conservation of sequence composition among these species.

Each of the homomorphs includes a sequence motif consisting of characteristic highly conserved functional residues that are essential to replication. The tertiary position of each of the homomorphs includes at least one residue (and sometimes more) in contact with the exterior surface of the protein and one or more highly conserved functional residues located within or at the wall of the template tunnel.

We defined the boundaries of a homomorph as the position where the structural segment shared by all representatives in a set became discontinuous in more than two species. For many queries, this position could be confidently identified. However, these positions sometimes varied by one or two residues, depending on the query sequence. Query-dependent differences in R2R matches were also observed within the motifs themselves, where minor differences in structure resulted in a lack of R2R matches for short segments of some queries. Our approach was to set the boundary at the position where most queries were in agreement, but to keep in mind that these edges might vary by one or two residues.

Poliovirus had R2R correspondence with other species in the sample set more often than did any other structure. In almost all instances, we were able to map functional features of other proteins to a structurally similar segment of poliovirus. This property of centrality makes it a useful template for polymerase structure properties.

HmG discussion

HmG is shared by picornaviruses, caliciviruses and flaviviruses, although the structures of each of these groups begin to diverge within the C-terminal segment of Motif G. Motif G is characterized by the conserved motif [T/Sx₁₋₂G], which is located near the outer edge of the template tunnel. The motif may enforce the correct orientation of essential residues and a primer (35). Each flank of the homomorph contains amino acid residues that significantly affect the life cycle of the species. In PV, mutations at the N-terminal residue of the homomorph (D71A/E72A) are lethal (37). Mutations located outside the N-terminal edge of the motif (PV D105A/E108A) result in small plaques (37). Downstream from the C-terminal edge of the motif, there is a nuclear localization signal (NLS) in the picornaviruses and caliciviruses. The NLS is located two residues from the C-terminus of the homomorph and mutations in the NLS (K125A/K126A/K127A and K127A/R128A/D129A) are lethal to PV (37).

HmF discussion

Previous research found that Motif F occurs in all RdRps (38), that it recognizes the incoming NTP (39), serves as the primary fidelity checkpoint for RdRp and reorients the proper triphosphate into a position for efficient catalysis

(40). HmF is an extensive structure with surface exposure at both ends and near its mid-section at Motif F2 (Figure 3A–D). Motif F2 (Figure 3E) is analogous to the loop in hmG that varies in composition and length; it is upstream of a highly conserved motif and is species-specific. The large size of this homomorph and its positioning that transects the protein while maintaining contact with the template tunnel is consistent with its established role in transcription, which requires both fine-scale stability and large-scale mobility. Motif F3 consists of mostly basic amino acid residues and forms the roof of the NTP entry tunnel (41); the characteristic conserved *arg* residue is essential to nucleotide binding (38). The required orientation of the F motifs would be stabilized by the loop formed by hmF and the double-stranded segment formed by the extension of the homomorph beyond the motifs. Both the N-terminal and C-terminal residues of the homomorph are exposed at the exterior surface of the protein. In PV, mutations of residues adjacent to the N-terminal are lethal: G149-i-I150 (42) and H149A/K150A (37).

HmA discussion

The conserved residues of Motif A (in PV, D233 and D238) control the function of the metal ions at the active site (41,43), which perform the phosphotransfer essential to polymerase activity (41). D233 is ligand to the metal (44). D238 is essential to NTP binding (3). Similar functions for the residues of Motif A have been identified for HCV (45), HRV (45) and FMDV (2). Motif A is centered within a spring-like homomorph (Figure 4B–D). Each end terminates at the exterior surface of the protein, and the beginning and end of the homomorph terminate nearly opposite each other. Mutations in the N-terminal segment of the homomorph (in PV at E226A/E227A) result in small plaques (37), suggesting that these residues influence the rate of catalysis. This is the region where species-specific structures protrude from the homomorph (PV L224–L229). This position, relative to the conserved motif, is analogous to a similar structure in hmG and Motif F2. All these structures contain a segment that varies in length and composition by species and is located upstream from a highly conserved motif, essential to replication.

An insertion at the C-terminal edge of Motif A (L241-i-S242) is lethal (Figure 4A) (42). The structure of the homomorph is highly conserved in this region, suggesting that the structural consequences of an insertion are not tolerated. The position of this lethal insertion is similar to the position of lethal mutations in Motif G, although the major effect in hmG may be the loss of the nuclear localization signal. Mutations near the C-terminal residue of the Motif A homomorph (PV G257) affect function: E254A/K255A is lethal (37), and the insertion I256-ile-G257 results in temperature sensitivity (46).

HmA provides a structural connection between the functional residues at the template tunnel and the exterior surface of the protein. Sequence residues immediately adjacent to the motif have a high degree of

functionality. It is possible that the orientation of the conserved segments that comprise the homomorph would be affected by changes in the orientation of residues at the edges of the motif. N-terminal and C-terminal segments of the homomorph are helices, which are likely to be relatively rigid.

HmB discussion

Motif B is near the center of a very large homomorph that contacts the exterior surface at nearly opposite positions. As stated by Bruenn (38), Motif B forms the base of the template-entry channel and may function in guiding the template entry into the active site. Choi *et al.* (17) observed that the highly conserved *asn* (N414 in BVDV) is conserved in all picornavirus. Hansen *et al.* (47) found that in HRV, N297 is involved in positioning NTP for recognition. Ferrer-Orta *et al.* (2) determined that the equivalent FMDV-N307 and D245 (Motif A) together are involved in ribonucleoside triphosphate (rNTP) selection. Tao *et al.* (21) and Butcher *et al.* (20) proposed that Motif B interacts with the 2'-OH group on the incoming nucleotide. Korneeva and Cameron (48) determined that FMDV-N307 interacts with the C-terminal-OH in the uridylylation complex, but with the 2'-OH in the elongation complex. The role of Motif B in the mechanisms of active site closure has recently been described in detail by Gong and Peersen (9). These experiments document the role of the highly conserved *asn* in the motif in multiple species and suggest that structural alignment may be useful for the identification of potential functionally equivalent residues in structures that have R2R correspondences.

StralSV structure analysis indicates that the structure of Motif B is highly conserved in all RdRps, unlike some of the other motifs that have unmatched R2R correspondences. This highly conserved structure is consistent with its role in NTP recognition. An insertion in Motif B at PV C290-s-S291 is lethal (42). Within the N-terminal segment of the hmB, the mutation in PV-K276L results in small plaques (49). The structural position of this mutation (within the homomorph and upstream from the motif) is similar to that of rate-affecting mutations in the homomorphs of Motifs G and A. At the C-terminal end of the homomorph, in BVDV (BVDV-F426), mutation of residues C427, S428 and R447 to *ala* reduces primer-dependent RNA elongation and abolishes *de novo* synthesis (17).

HmC discussion

Motif C is not a component of a larger structurally conserved segment, but has the same key features of the other homomorphs. It is folded in a manner that places the apex of the fold at the wall of the template tunnel, and both the N-terminus and C-terminus at the exterior surface of the protein (Figure 6B and C). Therefore, Motif C as defined in the literature comprises the homomorph. The absence of R2R correspondence adjacent to hmC indicates that the structures of the adjacent sequence segments are highly specific to each species. HmC is highly conserved in the RdRps and

highly similar to the DNA-dependent polymerases. Although there is a sequence inversion in the birnaviruses (Motif C precedes Motif A), Figure 7B and C illustrates that despite the difference in sequence order, the homomorphs occupy a similar tertiary position. StralSV analysis indicates that the structure of Motif C is highly conserved in the RNA-dependent polymerases, though slightly different in the DNA-dependent polymerases (Figure 6A).

Motif C is part of the classic 'RRM-fold' that forms the core of the palm domain of all these polymerases (together with that part of Motif A that forms a β -sheet with Motif C). Experimental studies have demonstrated that several residues within Motif C are sensitive to the position and composition of mutants. The highly conserved residues, GDD, occur near the center of the motif. The primary function of these residues is to coordinate the metal ions associated with the incoming rNTP(45,43). In PV, mutation of D to E in either or both positions (D328 or D329) is lethal (50). In HCV, mutation of G317A is also lethal (51). However, in birnaviruses the highly conserved residues are ADN, rather than GDD, and mutation to GDD increases RNA synthesis activity (5). Certain mutations immediately upstream from the GDD motif are lethal in PV: Y326[CHIMS] (50). This is similar to the effect of the L241-i-S242 at the downstream edge of Motif A in PV. Near the N-terminal end of Motif C in HCV (HCV T312), the mutation D311A characterizes chronic hepatitis (52). Mutation at the edge of a highly conserved structure seems to have a substantial effect on the viral life cycle. The R2R comparisons summarized in these structure maps identify the types of sequence variability that can occur while maintaining the same spatial structure (Figure 6A) and demonstrates and identifies the variations in composition that can be tolerated even within a key functional motif. The R2R correspondence of other RdRps with birnaviruses in Motif C (Figure 7A), despite the sequence inversion in birnaviruses (CAB) supports the premise that conservation of structure is a significant, if not dominant, factor in evolution.

HmD discussion

HmD is different from the other homomorphs in that it lies mostly on the surface of the protein (Figure 8C). Like the others, however, its terminal residues are located at a distinctive surface (Figure 8D). In the case of hmD, they come from an opposite surface rather than the interior of the protein. The N-terminal segment of hmD is more conserved than the motif itself, which forms the C-terminal segment. The motion of Motif D in the active state has not been captured by the existing structures in PDB (40). Therefore, the lack of R2R correspondence in the motif may be a reflection of the limitations of the available structures.

Residues within hmD perform varied functions. In PV, e.g. polymerases form an extensive lattice system by polymerase-polymerase interactions; L342 and D349, located within hmD, contribute to interface I of this lattice system (53). The most highly conserved residue within the homomorph is a *gly* (PV G351) at the N-terminal edge

of the motif and central to the homomorph; *gly* in this position would facilitate the folding of the homomorph, and is consistent with Cameron *et al.*'s (40) hypothesis that Motif D may be the most dynamic structural element of RdRps and RTs. Another conserved residue is a *lys* near the C-terminal edge (PV-K359). Residues equivalent to PV-K359 supply a proton to the nucleotidyl transfer reaction that increases the rate constant for nucleotide addition by 50- to 1000-fold (40). In PV, within the motif, the insertion T353-t-M354 results in small plaques, likely due to delayed RNA synthesis (54). In other homomorphs, mutations that affect the rate of synthesis occur more commonly outside of the motifs. Immediately downstream from the homomorph, PV-T362I is an attenuating mutation for the Sabin vaccine (40).

HmE discussion

HmE is ~36 amino acids in length (Figure 9A), and well represented by all RNA-dependent members of the sample set, except that no correspondence was found with HIV or TERT. These two species, however, are structurally matched to each other. There is considerable sequence similarity between PV and DENV within this homomorph, shown in the bottom segment of Figure 9A. Appleby *et al.* (55) determined that Motif E is unique to RNA polymerases.

HmE forms part of the NTP entry tunnel and has a considerable amount of exposure on the surface of the protein. The species-specific loop within hmE is at the outermost edge of the protein, a feature found in other homomorphs (G, F2 and A; Figures 2A, 3A and 4A, respectively). Huang *et al.* (25) found that the Motif E loop region acts as a pivot point for thumb subdomain movement upon template–primer binding. Motif E may also function in the proper positioning of the thumb relative to the palm (5). The turn of the loop projects into the active site cavity where it has been implicated in helping to position the C-terminal end of the primer strand for attachment to the α -phosphate of the NTP during phosphoryl transfer (56). Motif E in HCV plays a role in binding the priming nucleotide (not the incoming nucleotides) (38); HCV has a longer loop (Figure 9A), possibly related to this function. In PV, the C-terminal of hmE (R402) emerges from the protein into the segment 28-SAFHYVFEG-36. This segment contains residues F30 and F34, which interact with W403 to maintain the polymerase structure (39).

SUMMARY

Comparisons of the tertiary structures of the RdRps of 18 viral species indicated that most of the highly conserved residues essential to polymerase function are embedded in large sequence segments that are highly conserved structurally, yet disparate in composition. We have named these conserved segments 'homomorphs' and have identified the composition and length of each homomorph that includes previously recognized polymerase motifs (Table 2). We have demonstrated that the RNA

polymerases have structural skeletons (frames) that are highly conserved, with flexible segments between them, and that extensive segments of structure similarity can be identified by the methods we have described. These methods are applicable to the studies of other groups of proteins, and we anticipate that by accessing structure similarity independent of sequence composition, skeletal frameworks will be found in other groups of proteins. Additionally, after structure similarity is identified, differences between members of the group become readily apparent.

All of the homomorphs included residues that connect the template tunnel or the NTP entry tunnel with the outer surface of the protein. Although some of the surface residues within these homomorphs have specific functional roles, as reported in the literature (see citations in previous paragraphs), we anticipate that they may all be important for polymerase function; the consistent occurrence of homomorphs embedding motifs—even when a defined sequence motif is small in size—suggests a structure–function relationship between the motif and its structurally conserved flanking regions. It would be interesting to explore the possibility that interactions at the surface of the protein (e.g. protein–protein contact at surface homomorph residues) may subtly affect function buried deep beneath, within the tunnel. Furthermore, each homomorph is either divided by or is separated from another homomorph by a flexible secondary structure. Identification of the span of each homomorph and the terminal residue enables us to identify specific residues on the surface that would not, in many cases, be otherwise noticed. By comparing experimental data with the surface location of the ends of the homomorphs, we have found that these are often the sites of key functional interactions of the protein. A paper describing these sites is in preparation.

We have compared the effects of currently recognized mutations within the motifs and within the homomorphs. Most mutations within the motifs are function-specific, related to either a change in charge or size, and in most cases the mutations are lethal [mA (42), mB (42), mC (50),(51)]. Mutations outside of the motif (but within the homomorph) are more often rate related and located in a segment that bulges from the homomorph by an amount that varies by species (Figures 2A, 3A and 4A). These differences support the hypothesis that residues actively involved in template processing are essential to viability, and most of them are components of a consistent, stable structure that places and/or maintains them in their appropriate functional position. However, the practice of mutating residues to *ala* has resulted in a somewhat 'all or nothing' perspective of mutations. StralSV analysis can facilitate informed selection of alternative residues of various compositions, which could possibly affect replication rates to different extents. Experiments involving this type of testing would enhance predictive models and may provide new insights for the design and development of medical countermeasures.

The extension of all homomorphs from the template tunnel to the exterior of the protein was an unexpected finding. Its universality in the polymerase family suggests

a functional significance. Residues within the homomorphs that were localized to the surface often had species-specific loops. The most likely reason these features have not been identified previously is due to the limitations of existing sequence and structure comparison tools—in particular, the ability to perform multi-species comparisons of structures, using overlapping windows of a size determined by the user, and the ability to select the criteria for R2R matches. The homomorphs as defined in this work add structural clarity and context to sequence-based functional motifs previously observed by numerous authors performing comparative studies among polymerases. The structure maps created from the R2R correspondences identified by the StralSV algorithm provided a unique and informative perspective of structure and function in RdRps. They readily identify unique regions of each species and those shared by proteins within a family. These are features that would be useful for studies of any protein family. Based on the results of this study, it may be possible to define characteristic homomorphs for many other protein families, despite considerable sequence variation. It may be feasible to classify homomorphs in a manner analogous to the SCOP database, and in doing so provide new insight into protein evolution.

The StralSV algorithm simultaneously, rapidly and quantitatively identifies the similarities and differences of the structural components of multiple species and provides an output that facilitates the comparison of three-dimensional structure information. StralSV enabled us to cluster protein segments that have the same tertiary structure, independent of sequence variability. In a sense, it is an analog of Blast, although based on structure rather than sequence. The precision of StralSV makes it easy to identify small differences between and within species. The ability to process multiple species at the same time can rapidly accelerate our understanding of differences between them. The identified structural associations may also facilitate the transfer of structure-related functional information among proteins.

The traditional perspective of the relationship between the amino acid sequence of a protein and its tertiary structure has been that sequence determines structure. Under this premise, sequence-based evolutionary studies and phylogeny would inherently incorporate structure. In this study of RdRps, we demonstrated that structure accommodates substantial sequence variability, and that highly diverse sequences can generate highly similar tertiary structures. Structure-based phylogeny may provide new perspectives of protein evolution.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2.

ACKNOWLEDGEMENTS

A.Z. designed and developed the StralSV algorithm and performed calculations for all viral species in the study.

C.Z. and D.L. wrote codes and developed methods for post-processing of StralSV results, and performed literature searches for interpretation of biological significance of various residue positions. D.L. performed the greater part of the detailed sequence comparisons and prepared the manuscript, with contributions from C.Z. and A.Z. All authors participated in the discussions and shaped the ideas that led to the experimental design and results of this work. All authors read and approved the manuscript.

FUNDING

Lawrence Livermore National Laboratory (DOE Contract DE-AC52-07NA27344); UC-LLNS Fees grant (PI: CEZ). Funding for open access charge: University of California and Lawrence Livermore National Security Fees Grant (PI: CEZ).

Conflict of interest statement. None declared.

REFERENCES

- Zemla, A.T., Lang, D.M., Kostova, T., Andino, R. and Ecale Zhou, C.L. (2011) StralSV: assessment of sequence variability within similar 3D structures and application to polio RNA-dependent RNA polymerase. *BMC Bioinformatics*, **12**, e226.
- Ferrer-Orta, C., Arias, A., Perez-Luque, E., Escarmis, C., Domingo, E. and Verdagué, N. (2004) Structure of Foot-and-Mouth Disease Virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. *J. Biol. Chem.*, **279**, 47212–47221.
- Thompson, A.A. and Peersen, O.B. (2004) Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J.*, **23**, 3462–3471.
- Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.*, **8**, 3867–3874.
- Pan, J., Vakharia, V.N. and Tao, Y.J. (2007) The structure of a birnavirus polymerase reveals a distinct active site topology. *Proc. Natl Acad. Sci. USA*, **104**, 7385–7390.
- Koonin, E.V., Wolf, Y.I., Nagasaki, K. and Dolja, V.V. (2008) The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat. Rev. Microbiol.*, **6**, 925–939.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
- Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S. *et al.* (2002) The Protein Data Bank. *Acta Crystallogr., D: Biol. Crystallogr.*, **58**, 899–907.
- Gong, P. and Peersen, O.B. (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl Acad. Sci. USA*, **107**, 22505–22510.
- Wang, Y., Geer, L.Y., Chappay, C., Kans, J.A. and Bryant, S.H. (2000) Cn3D: sequence and structure views for Entrez. *Trends Biochem. Sci.*, **25**, 300–302.
- Campognola, G., Weygandt, M.H., Scoggin, K.E. and Peersen, O.B. (2008) Crystal structure of coxsackievirus B3 3Dpol highlights the functional importance of residue 5 in picornavirus polymerases. *J. Virol.*, **82**, 9458–9464.
- Love, R.A., Maegley, K.A., Yu, X., Ferre, R.A., Lingardo, L.K., Diehl, W., Parge, H.E., Dragovich, P.S. and Fuhrman, S.A. (2004) The crystal structure of the RNA-dependent RNA polymerase from human rhinovirus: a dual function target for common cold antiviral therapy. *Structure*, **12**, 1533–1544.
- Ng, K.K., Pendas-Franco, N., Rojo, J., Boga, J.A., Machin, A., Alonso, J.M. and Parra, F. (2004) Crystal structure of norwalk

- virus polymerase reveals the carboxyl terminus in the active site cleft. *J. Biol. Chem.*, **279**, 16638–16645.
14. Ng, K.K., Cherney, M.M., Vazquez, A.L., Machin, A., Alonso, J.M., Parra, F. and James, M.N. (2002) Crystal structures of active and inactive conformations of a caliciviral RNA-dependent RNA polymerase. *J. Biol. Chem.*, **277**, 1381–1387.
 15. Fullerton, S.W.B., Blaschke, M., Coutard, B., Gebhardt, J., Gorbalenya, A., Canard, B., Tucker, P.A. and Rohayem, J. (2007) The 2.3 Å resolution structure of the saporin virus RNA dependant RNA polymerase. *J. Virol.*, **81**, 1858–1871.
 16. O'Farrell, D., Trowbridge, R., Rowlands, D. and Jager, J. (2003) Substrate complexes of hepatitis C virus RNA polymerase (HC-J4): structural evidence for nucleotide import and De-novo initiation. *J. Mol. Biol.*, **326**, 1025–1035.
 17. Choi, K.H., Groarke, J.M., Young, D.C., Kuhn, R.J., Smith, J.L., Pevar, D.C. and Rossmann, M.G. (2004) The structure of the RNA-dependent RNA polymerase from bovine viral diarrhoea virus establishes the role of GTP in de novo initiation. *Proc. Natl Acad. Sci. USA*, **101**, 4425–4430.
 18. Malet, H., Egloff, M.P., Selisko, B., Butcher, R.E., Wright, P.H., Roberts, M., Gruez, A., Sulzenbacher, G., Vonnrhein, C., Bricogne, G. et al. (2007) Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. *J. Biol. Chem.*, **282**, 10678–10698.
 19. Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G. and Lescar, J. (2007) Crystal structure of the Dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J. Virol.*, **81**, 4753–4765.
 20. Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H. and Stuart, D.I. (2001) A mechanism for initiating RNA-dependent RNA polymerization. *Nature*, **410**, 235–240.
 21. Tao, Y., Faretta, D.L., Nibert, M.L. and Harrison, S.C. (2002) RNA synthesis in a cage—structural studies of reovirus polymerase lambda3. *Cell*, **111**, 733–745.
 22. Lu, X., McDonald, S.M., Tortorici, M.A., Tao, Y.J., Vasquez-Del Carpio, R., Nibert, M.L., Patton, J.T. and Harrison, T.C. (2008) Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. *Structure*, **16**, 1678–1688.
 23. Graham, R.L., Sarin, L.P., Bahar, M.W., Myers, R.A., Stuart, D.H. and Grimes, J.M. (2011) The N-terminus of the RNA polymerase from infectious pancreatic necrosis virus is the determinant of genome attachment. *PLoS Pathog.*, **7**, e1002085.
 24. Hopkins, A.L., Ren, J., Esnouf, R.M., Willcox, B.E., Jones, E.Y., Ross, C., Miyasaka, T., Walker, R.T., Tanaka, H., Stammers, D.K. et al. (1996) Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.*, **39**, 1589–1600.
 25. Huang, H., Chopra, R., Verdine, G.L. and Harrison, S.C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669–1675.
 26. Ding, J., Das, K., Hsiou, Y., Sarafianos, S.G., Clark, A.D. Jr, Jacobo-Molina, A., Tantillo, C., Hughes, S.H. and Arnold, E. (1998) Structure and functional Implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.*, **284**, 1095–1111.
 27. Gillis, A.J., Schuller, A.P. and Skordalakes, E. (2008) Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature*, **455**, 633–637.
 28. Yin, Y.W. and Steitz, T.A. (2001) The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, **116**, 393–404.
 29. Brieba, L. and Ellenberger, T. (2005) Structure of replicative DNA polymerase provides insights into the mechanisms for processivity, frameshifting and editing, doi: 10.2210/pdb2ajq/pdb.
 30. Yin, Y.W. and Steitz, T.A. (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA Polymerase. *Science*, **298**, 1387–1395.
 31. Gleghorn, M.L., Davydova, E.K., Rothman-Denes, L.B. and Murakami, K.S. (2008) Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol. Cell*, **32**, 707–717.
 32. Li, Y., Korolev, S. and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.*, **17**, 7514–7525.
 33. Doublet, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2Å resolution. *Nature*, **391**, 251–258.
 34. Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S.G. and Ding, J. (2003) Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res.*, **31**, 7117–7120.
 35. Gorbalenya, A.E., Pringle, D.M., Zeddiam, J.L., Luke, B.T., Cameron, C.E., Kalkmakoff, J., Hanzlik, T.N., Gordon, K.H.J. and Ward, V.K. (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J. Mol. Biol.*, **324**, 47–62.
 36. Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., DeFrancesco, R. and Rey, F.A. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 13034–13039.
 37. Diamond, S.E. and Kirkegaard, K. (1994) Clustered charged-to-alanine mutagenesis of poliovirus RNA dependent RNA polymerase yields multiple temperature-sensitive mutants defective in RNA synthesis. *J. Virol.*, **68**, 863–876.
 38. Bruenn, J.A. (2003) A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.*, **31**, 1821–1829.
 39. Thompson, A.A., Albertini, R.A. and Peersen, O.B. (2007) Stabilization of poliovirus polymerase by NTP binding and fingers–thumb interactions. *J. Mol. Biol.*, **366**, 1459–1474.
 40. Cameron, C.E., Mostafa, I.M. and Arnold, J.J. (2009) Dynamics: the missing link between structure and function of the viral RNA-dependent RNA polymerase? *Curr. Opin. Struct. Biol.*, **19**, 768–774.
 41. Ng, K.K., Arnold, J.J. and Cameron, C.E. (2008) Structure-function relationships among RNA-dependent RNA polymerases. *Curr. Top. Microbiol. Immunol.*, **320**, 137–156.
 42. Burns, C.C., Lawson, M.A., Semler, B.L. and Ehrenfeld, E. (1989) Effects of mutations in poliovirus 3Dpol on RNA polymerase activity and on polyprotein cleavage. *J. Virol.*, **63**, 4866–4874.
 43. Steitz, T.A. (1998) A mechanism for all polymerases. *Nature*, **391**, 231–232.
 44. Arnold, J.J., Vignuzzi, M., Stone, J.K., Andino, R. and Cameron, C.E. (2005) Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA Polymerase. *J. Biol. Chem.*, **280**, 25706–25716.
 45. Beese, L.S. and Steitz, T.A. (1991) Structural basis for the C-terminal-N-terminal exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25–33.
 46. Charini, W.A., Burns, C.C., Ehrenfeld, E. and Semler, B.L. (1991) trans rescue of a mutant poliovirus RNA polymerase function. *J. Virol.*, **65**, 2655–2665.
 47. Hansen, J.L., Long, A.M. and Schultz, S.C. (1997) Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*, **5**, 1109–1122.
 48. Korneeva, V.S. and Cameron, C.E. (2007) Structure-function relationships of the viral RNA-dependent RNA polymerase: fidelity, replication speed, and initiation mechanism determined by a residue in the ribose-binding pocket. *J. Biol. Chem.*, **282**, 16135–16145.
 49. Richards, O.C., Spagnolo, J.F., Lyle, J.M., Vleck, S.E., Kuchta, R.D. and Kirkegaard, K. (1996) Intramolecular and intermolecular uridylation by poliovirus RNA-dependent RNA polymerase. *J. Virol.*, **80**, 7405–7415.
 50. Jablonski, S.A. and Morrow, C.D. (1995) Mutation of the aspartic acid residues of the GDD sequence motif of poliovirus RNA-dependent RNA polymerase results in enzymes with altered metal ion requirements for activity. *J. Virol.*, **69**, 1532–1539.

51. Lohmann,V., Korner,F., Herian,U. and Bartenschlager,R. (1997) Biochemical properties of Hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J. Virol.*, **71**, 8416–8428.
52. Okura,I., Horiike,N., Michitaka,K. and Onji,M. (2004) Effect of mutation in the hepatitis C virus nonstructural 5B region on HCV replication. *J. Gastroenterol.*, **39**, 449–454.
53. Hobson,S.D., Rosenblum,E.S., Richards,O.C., Richmond,K., Kirkegaard,K. and Schultz,S.C. (2001) Oligomeric structures of poliovirus polymerase are important for function. *EMBO J.*, **20**, 1153–1163.
54. Bernstein,H.D., Sarnow,P. and Baltimore,D. (1996) Genetic complementation among poliovirus mutants derived from an infectious cDNA clone. *J. Virol.*, **60**, 1040–1049.
55. Appleby,T.C., Luecke,H., Shim,J.H., Wu,J.Z., Cheney,I.W., Zhong,W., Vogeley,L., Hong,Z. and Yao,N. (2005) Crystal structure of complete rhinovirus RNA polymerase suggests front loading of protein primer. *J. Virol.*, **79**, 277–288.
56. Jacobo-Molina,A., Ding,J., Nanni,R.G., Clark,A.D. Jr, Lu,X., Tantillo,C., Williams,R.L., Kamer,G., Ferris,A.L., Clark,P. *et al.* (1992) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.