Functional Evolution in Orthologous Cell-encoded RNA-dependent RNA Polymerases*

Received for publication, August 18, 2015, and in revised form, February 4, 2016. Published, JBC Papers in Press, February 23, 2016, DOI 10.1074/jbc.M115.685933

Xinlei Qian†, Fursham M. Hamid‡, Abbas El Sahili‡, Dina Amallia Darwis‡, Yee Hwa Wong‡, Shashi Bhushan‡, Eugene V. Makeyev§, and Julien Lescar¶

From the †Division of Structural Biology and Biochemistry, School of Biological Sciences, Nanyang Technological University, 138673 Singapore, Singapore, the ‡Medical Research Council Centre for Developmental Neurobiology, King’s College, London SE1 1UL, United Kingdom, and the ¶UPMC UMRS CR7-CNRS ERL 8255-INSEM U1135 Centre d’Immunologie et des Maladies Infectieuses, Faculté de Médecine Pierre et Marie Curie, Centre Hospitalier Universitaire Pitié-Salpêtrière, 75031 Paris, France

Many eukaryotic organisms encode more than one RNA-dependent RNA polymerase (RdRP) that probably emerged as a result of gene duplication. Such RdRP paralogs often participate in distinct RNA silencing pathways and show characteristic repertoires of enzymatic activities in vitro. However, to what extent members of individual paralogous groups can undergo functional changes during speciation remains an open question. We show that orthologs of QDE-1, an RdRP component of the quelling pathway in Neurospora crassa, have rapidly diverged in evolution at the amino acid sequence level. Analyses of purified QDE-1 polymerases from N. crassa (QDE-1Ncr) and related fungi, Thielavia terrestris (QDE-1Tte) and Myceliophthora thermophila (QDE-1Mth), show that all three enzymes can synthesize RNA, but the precise modes of their action differ considerably. Unlike their QDE-1Ncr counterpart favoring processive RNA synthesis, QDE-1Tte and QDE-1Mth produce predominantly short RNA copies via primer-independent initiation. Surprisingly, a 3.19 Å resolution crystal structure of QDE-1Tte reveals a quasisymmetric dimer similar to QDE-1Ncr. Further electron microscopy analyses confirm that QDE-1Tte occurs as a dimer in solution and retains this status upon interaction with a template. We conclude that divergence of orthologous RdRPs can result in functional innovation while retaining overall protein fold and quaternary structure.

Eukaryotic cells widely use small RNA (sRNA)* guides to limit proliferation of viruses and transposable elements, maintain proper chromosomal structure, and control endogenous gene expression in a sequence-specific manner (1–10). sRNA thwart improper chromosomal structure, and control endogenous limit proliferation of viruses and transposable elements, mainly because of the presence of (11–13). In fungi, protozoans, and some metazoans, sRNA production often depends on template-dependent RNA synthesis catalyzed by cell-encoded RNA-dependent RNA polymerases (RdRPs) (14, 15).

Many RdRPs contribute to maintenance and amplification of gene silencing signals initiated by primary sRNAs originating from viral genomes, endogenous sequences, or dsRNAs experimentally delivered into a cell (16–23). In such cases, long single-stranded RNA (ssRNA) targets of an initial round of silencing become templates for RdRP-catalyzed RNA synthesis, ultimately giving rise to secondary sRNAs. Moreover, some RdRPs may trigger a silencing response with no apparent need for preexisting sRNAs (22–27). This may involve RdRP recruitment to transcripts with aberrant or unusual molecular features generated by other RNA polymerases (23, 24, 28, 29).

In the filamentous fungus Neurospora crassa, silencing, or “quelling,” of transgenic arrays relies on RdRP QDE-1, PIWI/PAZ protein QDE-2, DNA helicase QDE-3, and several other components, including the ssDNA-binding protein RPA (2). Interestingly, QDE-1 may trigger quelling by producing long aberrant ssRNA copies (aRNAs) of ssDNA intermediates that frequently arise in tandem-duplicated genomic sequences (30–32). This reaction depends on QDE-3 and RPA and is stimulated by DNA damage. QDE-1 can subsequently use its aRNA product as a template to produce dsRNA intermediates. These are subsequently converted into sRNAs that associate with QDE-2 and target complementary sequences. Thus, at least some RdRPs may initiate gene silencing de novo.

In line with their diverse biological functions, individual RdRPs and their protein complexes isolated from various sources show different enzymatic properties in vitro. These include polymerase-specific preferences between ssRNA versus ssDNA templates and primer-dependent versus primer-independent initiation of RNA synthesis (33–37). Some RdRPs can also function as template-dependent terminal transferases (34, 35). Notably, two distinct template-dependent polymerization modes have been described for previously studied RdRPs: (a) processive synthesis of long double-stranded products, which is typically initiated at or close to the 3′-end of a single-stranded template using either primer-independent or so-called “back-priming” mechanisms, and (b) non-processive...
synthesis of sRNA copies initiated in a primer-independent manner at internal positions of a single-stranded template (33).

Different RdRPs appear to utilize the two modes with markedly different efficiencies. For example, Rdr1 from Schizosaccharomyces pombe or RDR6 from Arabidopsis efficiently synthesizes long products (34, 38), whereas RRF-1 from Caenorhabditis elegans specializes in production of sRNAs (39). These biochemical differences are consistent with in vivo evidence; long dsRNAs generated by Rdr1, RDR2, and RDR6 must be processed by Dicer/RNase III-like endoribonucleases to generate functional small interfering RNAs (siRNAs), whereas sRNA products of RRF-1 apparently do not require further processing for their secondary siRNA function (16, 19, 20).

Purified QDE-1 from N. crassa can use both primer-independent and “back-priming” mechanisms in vitro (33, 35). The crystal structure of QDE-1 catalytic fragment, thus far the only known structure of a cell-encoded RdRP, suggests that this enzyme is a homodimer with the two subunits adopting either “closed” or “open” conformation (40). It has been proposed that the two structurally distinct conformations may help this remarkably versatile enzyme choose between different activities (40). However, in the absence of structural information for corresponding enzyme-substrate complexes, whether QDE-1 in fact remains a dimer upon template binding is unknown. Moreover, it remains to be seen whether other RdRPs can form homodimers, an important question, given that at least some RdRPs behave as monomers in solution (36).

On a more fundamental level, how new functional properties evolve in the RdRPs (and in other protein families for that matter) is poorly understood. Gene duplication followed by paralog divergence is a major driving force in protein evolution (41, 42), and it clearly contributed to RdRP diversification. Indeed, many species encode more than one distinct RdRP, with three paralogous genes present in N. crassa (QDE-1, SAD-1, and RRP-3), four in C. elegans, and six in Arabidopsis (14). The last eukaryotic common ancestor might have contained three functionally distinct RdRPs giving rise to the α, β, and γ branches of the RdRP genealogy, an arrangement that was further modified by lineage-specific gene duplications and losses (43).

Species-specific members of individual paralogous groups, referred to as orthologs, are typically assumed to have similar biological activities (44). However, it has been alternatively proposed that divergence of orthologous sequences might frequently result in acquisition of novel functional properties (45). Until recently, it has been difficult to investigate these possibilities experimentally because genomes of just a few distantly related model organisms have been sequenced completely. Here we took advantage of the increasing number of whole-genome sequences available for fungal species and examined evolutionary trends in QDE-1 orthologs using phylogenetic, biochemical, and structural approaches.

**Experimental Procedures**

**Phylogenetic Analyses**—Amino acid sequences of fungal polymerases were downloaded from OrthoDB (46) and aligned using MUSCLE (47). Phylogenetic trees were constructed in MEGA6 (48) by computing evolutionary distances using Poisson correction and inferring evolutionary history by the neighbor-joining method. Tree topology was tested using bootstrapping. Amino acid sequence conservation profiles were plotted in EMBOSS/plotcon using a 50-amino acid sliding window (49). Protein structures were color-coded according to interspecies conservation using Chimera (50). Chimera was also used to predict ligand positions in N. crassa QDE-1 aponzyme (Protein Data Bank code 2J7N) (40) based on the known structure of the polymerase II elongation complex (Protein Data Bank code 1R9T) (51).

**Protein Expression and Purification**—Recombinant proteins were expressed and purified as described elsewhere (52, 53). Briefly, synthetic open reading frames (ORFs) encoding catalytic fragments of QDE-1^Tte^ and QDE-1^Mth^ were obtained from Genscript, and the sequence-encoding catalytic fragment of QDE-1^Ncr^ (QDE-1AN) was amplified from pEM55 (33). Catalytically inactive QDE-1^Mth^ D607A mutant with the DYDDG motif substituted by AYDGD was prepared by QuikChange mutagenesis (Agilent). The ORFs were amplified using primers shown in Table 1, cloned into the pFB-LIC-Bse (a gift from Opher Gileadi; Addgene plasmid 26108) using ligation-independent cloning (54), and subsequently transformed into DH10Bac (Life Technologies) to produce the recombinant bacmids. Viral stocks generated by introducing the bacmids into Sf9 insect cells were further amplified and used to infect Sf9 cells for large scale protein expression. Virus-infected cells were harvested by centrifugation at 4,000 × g for 15 min at 4 °C. Cell pellets were resuspended in 20 mM HEPES-NaOH, pH 8.0, 5 mM imidazole, 300 mM NaCl, 5% (v/v) glycerol with complete EDTA-free protease inhibitors (Roche Applied Science) and subjected to sonication. Soluble fractions were isolated by centrifugation at 50,000 × g for 30 min at 4 °C and incubated with His tag purification resin (Roche Applied Science) for 1 h at room temperature. Non-specifically bound proteins were eluted by 20 mM HEPES-NaOH, pH 8.0, 15 mM imidazole, 300 mM NaCl, 5% (v/v) glycerol, and 0.5 mM tris-(2-carboxyethyl) phosphine. His tag-containing proteins were eluted with 20 mM HEPES-NaOH, pH 8.0, 500 mM imidazole, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM tris(2-carboxyethyl)phosphine. Fractions containing the protein were concentrated using a 100 kDa cutoff concentrator (Sartorius) and further purified by size exclusion chromatography using Superdex 200 (GE Healthcare) pre-equilibrated with 20 mM HEPES-NaOH, pH 8.0, and 0.5 mM tris(2-carboxyethyl)phosphine additionally containing 150 mM NaCl and 5% (v/v) glycerol (QDE-1^Tte^), 300 mM NaCl and 5% (v/v) glycerol (QDE-1^Ncr^), or 500 mM NaCl and 10% (v/v) glycerol (QDE-1^Mth^ and QDE-1^Mth^ (D607A)). Eluted proteins were concentrated and stored at −80 °C. Chemicals were from Sigma-Aldrich unless stated otherwise.

### Table 1: Primers used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’→3’)</th>
<th>Name</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QDE1_Ncr_forward</td>
<td>GACAAGTTGTACAGGTTTCAGCTCCAGAGGGCG</td>
<td>QDE1_Ncr_reverse</td>
<td>TATCCACCTTTACTGTCAATAATCGCCATTCCCTGAA</td>
</tr>
<tr>
<td>QDE1_Mth_forward</td>
<td>TATCCACCTTTACGTGTTACTACCTGCTACAGGACGC</td>
<td>QDE1_Mth_reverse</td>
<td>TATCCACCTTTACTGTGTAATTACTCGTCACCTGAATCGC</td>
</tr>
<tr>
<td>QDE1_Tte_forward</td>
<td>TATCCACCTTTACTGTGTAATTACTCGTCACCTGAATCGC</td>
<td>QDE1_Tte_reverse</td>
<td>TATCCACCTTTACTGTCAATAATCGCCATTCCCTGAA</td>
</tr>
<tr>
<td>D607A_forward</td>
<td>GACAAGTTGTACAGGTTTCAGCTCCAGAGGGCG</td>
<td>D607A_reverse</td>
<td>TATCCACCTTTACTGTCAATAATCGCCATTCCCTGAA</td>
</tr>
</tbody>
</table>
Protein Thermostability Assay—Protein thermostability was determined by monitoring temperature-induced fluorescence changes, as described elsewhere (55). Purified proteins were incubated at 1 mg/ml in the gel filtration buffer with 1,000-fold diluted SYPRO Orange stock (Life Technologies) in 96-well PCR plates (Bio-Rad) sealed with optical sealing tape (Bio-Rad). Fluorescence was measured using an iCycler iQ5 real-time PCR detection system (Bio-Rad) with excitation and emission wavelengths set at 490 and 575 nm, respectively. The temperature was increased from 20 to 90 °C with 1 °C increments, and the mixture was incubated for 12 s at each temperature. Protein melting temperatures were calculated using the iQ5 Optical System software, version 2.1 (Bio-Rad).

RNA Polymerase Assays—A synthetic single-stranded DNA (ssDNA) template (5’-CTGACTGCTTCTGTGTGTCTTCT-3’), containing the fragment corresponding to the sTTE catalytic fragment (Protein Data Bank code 2J7N) (40) as a search probe. The model was built interactively using Coot (57), and the structure was refined using REFMAC from the CCP4 package, with tight non-crystallographic symmetry restraints between the four independent monomers (each monomer was considered as a group) with individual atom isotropic temperature factors and TLS refinement (58). Each monomer of the QDE-1 catalytic fragment used for crystallization contains 1,034 residues, including the His tag and the tobacco etch virus protease cleavage site. Of these, 924/ TABLE 2

<table>
<thead>
<tr>
<th>Data collection and refinement statistics for QDE-1Tth apoenzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection statistics</strong></td>
</tr>
<tr>
<td>Detector type</td>
</tr>
<tr>
<td>Synchrotron</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Unit cell parameters</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
</tr>
<tr>
<td>α, β, γ (degrees)</td>
</tr>
<tr>
<td>Measured reflections*</td>
</tr>
<tr>
<td>Unique reflections*</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>I/σ(I)</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Rmerge*²</td>
</tr>
</tbody>
</table>

**Refinement statistics**

| Reflections used for refinement | 75,503 |
| Non-hydrogen atoms | 28,130 |
| Rwork (%) | 21.30 (25.06) |
| Rfree (%) | 25.09 (26.20) |
| Root mean square deviations |
| Bond lengths (Å) | 0.015 |
| Bond angles (degrees) | 1.85 |
| Mean B-factor (Å²) | 60.58 (56.62, 73.99, 64.8, 77.5) |
| Ramachandran plot |
| Most favored (%) | 83.2 |
| Allowed (%) | 13.8 |
| Outliers (%) | 1.4 |
| Protein Data Bank code | 5FSW |

* The values for the highest resolution shell are shown in parenthesis.
* Residues missing from the model are as follows: 213–225, 566–567, 813–816, 867–875, 895–906, 923–925, 967–972, and 1007–1034.
* The B-factor values are given for all of the protein atoms and for each of the four polypeptide chains in the asymmetric unit.

Filter Binding Assay—Purified QDE-1∶Ncr, QDE-1∶Tte, or bovine serum albumin (BSA) control was incubated at 1 μM in the presence of binding buffer (50 mM HEPES-NaOH, pH 7.8, 0.1 mM EDTA, 2% (v/v) Triton X-100, 100 mM NH₄OAc, and 2 mM MgCl₂) and 0.03–10 μM ATP (PerkinElmer Life Sciences) and T4 polynucleotide kinase (New England Biolabs). Following a 10-min incubation at 20 °C, mixtures were passed through nitrocellulose membrane prewashed with the binding buffer using a Bio-Rad slot blot apparatus. The labeled RdRP-ssDNA complex retained on the membrane was quantified using phosphorimaging (Typhoon Trio, GE Healthcare).

Protein Crystallization, Diffraction Data Collection, and Structure Determination—Crystals of the QDE-1∶Tte catalytic fragment were obtained at 30 °C by mixing 2 μl of the protein solution at 2 mg/ml with 1 μl of crystallization solution (100 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, and 10 mM MgCl₂, 0.1 units/μl recombinant RNasin (Promega), 2 mM spermine). Crystals were transferred to a drop containing the crystallization solution supplemented with 10% (v/v) glycerol and incubated at 12 °C overnight. Before freezing, crystals were dehydrated at room temperature in two steps of 20 and 30% glycerol. X-ray diffraction data were collected to 3.19 Å resolution at 100 K at the PXIII beamline of the Swiss Light Source (Villingen, Switzerland) using a Pilatus 6M detector (Dectris). The crystal belonged to the P2₁ space group with the following unit cell dimensions: a = 84.23 Å, b = 165.84 Å, c = 173.83 Å, and β = 90.10°. The data collection and structure refinement parameters are listed in Table 2. The structure was determined by molecular replacement using the known structure of the QDE-1∶Ncr catalytic fragment (Protein Data Bank code 2J7N) (40) as a search probe. The model was built interactively using Coot (57), and the structure was refined using REFMAC from the CCP4 package, with tight non-crystallographic symmetry restraints between the four independent monomers (each monomer was considered as a group) with individual atom isotropic temperature factors and TLS refinement (58). Each monomer of the QDE-1∶Tte fragment used for crystallization contains 1,034 residues, including the His tag and the tobacco etch virus protease cleavage site. Of these, 924/
292 could be built in dimer A/B and 922/922 in dimer C/D in the two non-crystallographic dimers of the asymmetric unit (Table 2). Missing residues in the model belong to the N and C termini and flexible loops (Table 2). Missing residues in the model belong to the N and C termini and flexible loops (Table 2).

**Electron Microscopy (EM)—Purified QDE-1**

QDE-1 was diluted to 10 μg/ml with 100 mM Tris–HCl, pH 7.5, 75 mM NaCl, 5 mM MgCl₂, and 5% (v/v) glycerol. A volume of 4 μl of protein sample was applied to a glow-discharged carbon-coated transmission EM grid and stained with 2% (v/v) uranyl acetate. RdRP-ssDNA substrate complex was formed by incubating 1.3 μM QDE-1 with 1.7 μM (i.e. ~10-fold higher than the apparent $K_a$ determined in the filter binding assay) DNA oligonucleotide 5′-CTTATATTGTATAG-3′ before transferring acutely diluted mixture to a grid and subsequently staining with uranyl acetate as above. Images were recorded at a monochromator energy of 66.350 using a FEI T12 transmission electron microscope equipped with a 4K CCD camera (FEI) under low dose conditions.

**Results**

Orthologs of *N. crassa* QDE-1 Belong to a Rapidly Evolving Protein Group—To gain insights into RdRP evolution, we analyzed corresponding genes from 40 taxonomically diverse fungi with completely sequenced genomes (Table 3 and Fig. 1). Most species encoded RdRPs proteins clustering with *N. crassa* QDE-1, Sad-1, or RRP-3, except for three species from the Eurotiales and Hypocreales orders that had QDE-1 and Sad-1 but not RRP-3. *S. pombe* and *S. cerevisiae* had no RdRPs, as expected. Within each of the three groups, RdRPs clustered according to their taxonomic origin. This topology suggested that the last common ancestor of fungi might have contained at least three RdRP paralogs that evolved as orthologously related lineages or were occasionally lost during speciation.
1. To identify regions accounting for their apparently accelerated evolution, we examined a QDE-1 conservation plot (Fig. 2). The non-catalytic N-terminal part showed extremely low conservation scores, as pointed out previously (33). However, even within the generally less divergent C-terminal part, a prominent peak of sequence conservation was detected only in the vicinity of the catalytic DYDGD motif (14, 33). This contrasted with the SAD-1 and RRP-3 plots that contained substantially broader regions of relatively high conservation. Conservation was even more uniform for two polymerase II subunits, RPB1 and RPB2, distantly related to cell-encoded RdRPs (40) (Fig. 2A). Quantitative analyses of amino acid substitution rates showed that QDE-1 was significantly more divergent than SAD-1, RRP-3, RPB1, and RPB2 (Fig. 2B).

Interestingly, conserved amino acid residues showed prominent clustering around the active center of the known crystal structure of the QDE-1Ncr apoenzyme interacting with one of the two catalytic Mg$^{2+}$/H$_{11001}$ ions (Mg$^{2+}$/H$_{11001}$A) (Protein Data Bank code 2J7N) (40) (Figs. 2C and 3). We modeled positions of other molecules participating in RNA polymerization, including an incoming ATP monomer, the second Mg$^{2+}$/H$_{11001}$B, the template, and the RNA product based on the structure of the polymerase II elongation complex (Protein Data Bank code 1R9T) (51). This placed the ATP and Mg$^{2+}$/H$_{11001}$B near the conserved surface of the nucleotide pore and the nascent 3' end of the RNA product along with the corresponding template nucleotides near the conserved DYDGD loop and its Mg$^{2+}$/H$_{11001}$A ligand (Figs. 2C and 3). On the other hand, a more distal segment of the template-product duplex egressing from the active center was surrounded by substantially more divergent QDE-1 surfaces (Figs. 2C and 3). Notably, all contacts made by the template and the nascent RNA with the RPB1 and RPB2 subunits of the polymerase II complex appear highly conserved in evolution (Fig. 2C). We concluded that the QDE-1 group is generally more divergent than its RdRP paralogs and DdRP relatives and that it shows unusually strong sequence variability outside of the active center and the NTP-interacting surfaces.

QDE-1 Orthologs Generate Markedly Different Combinations of Long and Short RNA Products—To examine whether divergent QDE-1 proteins had distinct functional properties, we purified catalytic fragments of QDE-1Ncr and its ortholog from Thielavia terrestris (QDE-1Tte) from the Chaetomiaceae family distantely related to N. crassa (Sordariaceae family) using a standardized protocol (see “Experimental Procedures”) and analyzed the RNA polymerase activity of these two proteins (Figs. 4 and 5). QDE-1Ncr is known to accept either ssRNA or ssDNA templates and generate their continuous end-to-end RNA copi-
FIGURE 2. **QDE-1 orthologs rapidly diverge in evolution.** A, similarity plots for the QDE-1, SAD-1, and RRP-3 orthologs as well as for RPB1 and RPB2 (bottom) subunits of polymerase II DdRP from 37 fungal species containing more than one paralogous RdRP. Red arrowheads mark positions of the catalytic D(Y/F/L)DGD motif in the RdRP and RPB1 active centers (RPB2 lacks this sequence). The gray dimension line indicates the C-terminal part of QDE-1 previously shown to be enzymatically active (33). Note that the DYDGD-adjacent region in the QDE-1 orthologs is conserved noticeably better than the rest of the sequence. This contrasts with the rest of the proteins containing wider areas of relatively strong conservation. B, box and whisker plot comparison of amino acid substitution scores showing significantly higher divergence rate in QDE-1 proteins compared with their SAD-1, RRP-3, RPB1, and RPB2 counterparts. Corresponding Wilcoxon test \( p \) values are indicated at the top. C, surfaces proximal (≤4 Å) to the template (dark gray) and the nascent RNA product (black) in the “closed” subunit of *N. crassa* QDE-1 (QDE-1\text{Ncr}) (Protein Data Bank code 2J7N) (40) and *S. cerevisiae* polymerase II (Protein Data Bank code 1R9T) (51). Incoming ATP monomer and the two catalytic ions, Mg\(^{2+}\)\(_A\) and Mg\(^{2+}\)\(_B\), are shown in black and green, respectively. Protein surfaces are colored using cyan for low, white for intermediate, and maroon for high conservation. In the case of QDE-1\text{Ncr}, Mg\(^{2+}\)\(_A\) position is determined experimentally, whereas template, RNA product, incoming ATP, and Mg\(^{2+}\)\(_B\) are modeled based on the polymerase II elongation structure on the right. Error bars represent the standard errors.
FIGURE 3. **Spatial clustering of evolutionarily conserved amino acid sequences in QDE-1Ncr structure.** Surface representations of the QDE-1Ncr protein homodimer show the "closed" A subunit colored according to interspecies conservation and the "open" B subunit colored in beige. The known position of Mg\(^{2+}\)/H11001 and predicted positions of the second catalytic Mg\(^{2+}\) ion (Mg\(^{2+}\)/H11001), an incoming ATP, a template, and an RNA product are also indicated. Note that highly conserved amino acid residues cluster in the vicinity of the catalytic center and the NTP pore.

FIGURE 4. **QDE-1 orthologs have distinct enzymatic properties.** A and B, RNA polymerase activities of purified catalytic fragments of QDE-1Ncr and QDE-1Tte were assayed at 30–60 °C in the presence of an ssRNA (A) or an ssDNA template (B). The reaction products were separated by native agarose gel electrophoresis and visualized using ethidium bromide staining or \(^{32}\)P phosphorimaging, as indicated. Note that QDE-1Tte is substantially more efficient than QDE-1Ncr in generating sRNA products that migrate either at a low molecular weight position or in a template base-paired form. C, RNA polymerase activities of QDE-1Tte and QDE-1Ncr were assayed at 45 °C in the presence of a ssRNA template and analyzed by denaturing agarose gel electrophoresis. D, RNA products from C were incubated with increasing concentrations of RNase ONE or RNase ONE reaction buffer, as specified under "Experimental Procedures." Positions of the 1× full template-length products of processive end-to-end polymerization initiated in a primer-independent manner and sRNA products of non-processive polymerization are indicated on the right. Also shown is an expected position of "back-primed" 2× template-length products, which QDE-1Ncr can generate for some but not all ssRNA templates (33).
The newly analyzed QDE-1 ortholog, the newly analyzed QDE-1 ortholog, was a highly active RNA polymerase (Fig. 4A). However, it showed a striking bias toward generating short RNA products rather than full-length copies (Fig. 4A). This activity produced readily detectable amounts of template-product duplexes migrating slower than the ssRNA template but faster than the corresponding full-length dsRNA on EtBr-stained gels (Fig. 4A). Gel autoradiography additionally revealed a prominent low molecular weight band probably corresponding to short RNA products migrating on native gels as full-length dsRNAs and partially dsRNA species, respectively (Fig. 4A).

Similar to its N. crassa ortholog, the newly analyzed QDE-1 Tte was a highly active RNA polymerase (Fig. 4A). However, it showed a striking bias toward generating short RNA products rather than full-length copies (Fig. 4A). This activity produced readily detectable amounts of template-product duplexes migrating slower than the ssRNA template but faster than the corresponding full-length dsRNA on EtBr-stained gels (Fig. 4A). Gel autoradiography additionally revealed a prominent low molecular weight band probably corresponding to short RNA products released from the template (Fig. 4A). A similarly migrating band of short RNAs was also a major reaction product in the ssDNA-programmed reactions containing QDE-1 Tte but not QDE-1 Ncr (Fig. 4B). On the other hand, both enzymes gave rise to full-length DNA-RNA template-product duplexes with comparable efficiencies (Fig. 4B).

To better understand the nature of RNA products, we analyzed ssRNA-programmed reactions by denaturing agarose gel electrophoresis (Fig. 4C). Under these conditions, short RNA copies quantitatively dissociated from the template and migrated at the expected low molecular weight position. This analysis additionally revealed RNA products of 1× template length, thus suggesting that, under conditions used in our RdRP assays, both QDE-1 Ncr and QDE-1 Tte can initiate end-to-end RNA synthesis in a predominantly primer-independent manner (33, 35) (Fig. 4C). Consistent with the native gel analyses, the ratio between short and long RNA products was noticeably higher for QDE-1 Tte than for QDE-1 Ncr (Fig. 4C). Importantly, the bias for short RNA products was consistently detected for several independently purified batches of QDE-1 Tte and observed over a wide temperature range (Fig. 4, A and B). This ruled out the trivial explanation that the enzymatic differences between QDE-1 Tte and QDE-1 Ncr were related to a higher growth temperature limit of T. terrestris compared with N. crassa (60–62). In fact, the thermal stability of the QDE-1 Tte protein exceeded that of QDE-1 Ncr by only 3.7 °C ($T_m = 54.0 ± 0.0 °C$ versus $T_m = 50.3 ± 0.6 °C$; Fig. 5B). The two polymerases also had comparable single-stranded template binding properties (Fig. 5, C and D). Thus, distinct QDE-1 orthologs may have markedly different RNA polymerization properties.

Evolutionary Divergence between QDE-1 Family Members Modulates Their Product Preferences—To ensure that functional differences between QDE-1 Ncr and QDE-1 Tte resulted from evolutionary innovation, we purified a catalytic fragment of QDE-1 from Myceliophthora thermophila, a Chaetomiaceae fungus more closely related to T. terrestris than to N. crassa (Figs. 1 and 5). The $T_m$ of QDE-1 Mth was comparable with that of QDE-1 Ncr and QDE-1 Tte ($T_m = 52.3 ± 0.6 °C$ versus $T_m = 50.3 ± 0.6 °C$ and $T_m = 54.0 ± 0.0 °C$, respectively; Fig. 5).
To compare enzymatic properties of QDE-1<sup>Ncr</sup>, QDE-1<sup>Tte</sup>, and QDE-1<sup>Mth</sup>, we incubated the three polymerases with either ssRNA or ssDNA template for 1 h at 45 °C and separated the reaction products by urea-containing PAGE, affording simultaneous detection of long RNAs migrating at the top of the lane and sRNA products visualized at single-nucleotide resolution (Fig. 6). In ssRNA-programmed reactions, QDE-1<sup>Mth</sup> polymerase was clearly more efficient in producing short RNA products than QDE-1<sup>Ncr</sup>, albeit not to the same extent as QDE-1<sup>Tte</sup> (Fig. 6A). Moreover, QDE-1<sup>Ncr</sup> synthesized detectable amounts of 23-nt and 28–31-nt products but virtually no 24–27-nt sRNAs (Fig. 6A). On the other hand, QDE-1<sup>Tte</sup> and QDE-1<sup>Mth</sup> efficiently produced 24–27-nt sRNA in addition to shorter products (Fig. 6A). Interestingly, each of the three polymerases generated a unique blend of sRNA products in ssDNA-programmed reactions (Fig. 6B). As an additional control, we purified a D607A mutant QDE-1<sup>Mth</sup>, where the first Asp residue of the DYDGD motif was mutated to Ala. As expected, the mutant protein lacked detectable nucleotidyltransferase activity (Fig. 7). We concluded that sequence divergence between QDE-1 orthologs appears to underlie differences in their activities.

QDE-1<sup>Tte</sup> Is Structurally Similar to QDE-1<sup>Ncr</sup>—We next wondered whether distinct functional properties of QDE-1 orthologs might be due to major differences in spatial structures of these enzymes. This appeared plausible, given the degree of amino acid sequence divergence outside of the active center (see Fig. 2). To this end, we determined QDE-1<sup>Tte</sup> three-dimensional structure using x-ray crystallography. A 3.19 Å resolution trace of the QDE-1<sup>Tte</sup> polypeptide chain revealed two closely similar QDE-1<sup>Ncr</sup>-like homodimers in the asymmetric unit (root mean square deviation of 1.06 Å between dimer A/B and C/D for 917 carbon atoms). In each dimer, the “head” (residues 312–413 and 535–639) located at the base of each subunit (Fig. 8, A and B). The “head” (residues

FIGURE 6. QDE-1 orthologs synthesize different blends of sRNA products. A and B, purified QDE-1<sup>Ncr</sup> and QDE-1<sup>Tte</sup> and QDE-1<sup>Mth</sup> were assayed at 45 °C with an ssRNA (A) or an ssDNA (B) template and analyzed by urea-containing 15% polyacrylamide gel electrophoresis. Note that in the presence of a long ssRNA, QDE-1<sup>Tte</sup> and QDE-1<sup>Mth</sup> generate a relatively higher fraction of sRNAs than QDE-1<sup>Ncr</sup>, especially those of 12 nt and 24–27 nt. On the other hand, QDE-1<sup>Ncr</sup> synthesizes a relatively larger amount of long RNA products migrating close to the top of the lane. Each of the three polymerases produces a unique combination of sRNA products from the ssDNA. In the lane scans provided at the bottom of each panel, maximal intensity of the 18-nt marker band is set to 1.
457–507 and 822–1006 of the catalytic fragment) and the “neck” domains (residues 428–532 and 784–821) of the two subunits adopted a slightly tilted back-to-back position (Fig. 8A), and the “slab” domain (residues 1–244) of each subunit protruded from the catalytic domain toward the corresponding head domain. The five domains formed an extensive groove, probably accommodating the template and the nascent RNA product, with the catalytic site located at the bottom (Fig. 8, A and B). In general, Ca atoms in the QDE-1Ncr and QDE-1Tte dimers superimposed very well with a root mean square deviation score of 1.34 Å for 917 α-carbon atoms (Fig. 8C).

As found previously for QDE-1Ncr, one of the two QDE-1Tte subunits adopted a slightly more closed conformation than the other. Superposition of 922 Ca atoms between the two monomers (dimer A/B) using the SSM server returned a root mean square deviation of 2.15 Å (Fig. 8D). The largest conformational

FIGURE 7. Lack of RNA polymerase activity in the QDE-1Mth (D607A) mutant. The QDE-1Mth (D607A) mutant containing the AYGDG sequence instead of the wild-type catalytic DYDGD motif shows no RNA polymerase activity in the presence of either ssRNA or ssDNA template and over a wide range of reaction temperatures.

FIGURE 8. QDE-1Tte crystal structure. A, QDE-1Tte is an asymmetric dimer with each subunit containing a DPBB1 (shown in cyan for subunit A), a DPBB2 (green), a “head” (magenta), a “neck” (blue), and a “slab” domain (red). The entire B subunit is colored in gray. B, magnified view of the DPBB1 and DPBB2 domains with the three catalytic aspartate side chains shown as balls and sticks. C, structural alignment between QDE-1Tte and previously solved QDE-1Ncr structure showing a considerable overlap between the two proteins. D, structural alignment between subunits A and B QDE-1Tte demonstrating that subunit A adopts a more “closed” conformation than subunit B. The two subunits are color-coded as in A.
differences between the two monomers occurred closer to the C-terminal end of the polypeptide chain. The homodimer interface is stabilized by a total of 119 and 115 residues of monomer A and B, respectively (63). Most residues at the interface originate from the upper “neck” and “head” domains, where the subunits display the largest structural differences (Fig. 8C). These data suggest that QDE-1Ncr and QDE-1Tte have remarkably similar tertiary and quaternary structures despite their primary sequence differences.

**QDE-1Tte Can Function as a Dimer in Solution**—Extensive contacts between the A and B subunits observed in the QDE-1Ncr and QDE-1Tte crystals suggested that these proteins may form catalytically active dimers in solution with the two monomers oscillating between the closed and open conformations. To test this possibility directly, we used EM to compare negatively stained images of the QDE-1Tte apoenzyme and its complex with a ssDNA template (Fig. 9), using reference-free two-dimensional class averaging (Fig. 9B). These initial EM reconstructions carried out using 3,000 apoenzyme and 3,000 ssDNA bound particles revealed largely similar pyramidal-shaped homodimer structures, each containing two quasi-symmetric grooves and showing at this resolution a generally good fit with the QDE-1Tte crystal structure obtained without bound DNA (Fig. 9C). As in the crystal structure, one of the two template-product grooves in the QDE-1Tte dimer adopted a more open conformation in the EM reconstructions (Fig. 9C). To obtain a better resolved solution structure of QDE-1Tte in the presence of ssDNA, we extended our EM data analysis to include a total of 11,992 particles. This allowed us to visualize this complex to a resolution of 20 Å (Fig. 10, A and B). As
expected, the structural fit between the EM and crystal structures of QDE-1\textsuperscript{Tte} further improved when the EM map was prepared using the crystal structure as a reference (Fig. 10B). Of note, even this improved resolution was not sufficiently high to unambiguously assign density for the ssDNA. Likewise, understanding of the subtle conformational changes in the QDE-1\textsuperscript{Tte} dimer upon template binding will require a higher resolution structure of the RdRP-ssDNA complex, a challenge that will be addressed in the future. Nonetheless, our present data strongly suggest that QDE-1 orthologs function as dimers in solution.

Discussion

This study suggests that fungal QDE-1 orthologs underwent an unusually rapid sequence divergence outside of the catalytic center and the nucleotide-binding site. This corresponds to a significantly higher amino acid substitution rate in this group compared with its RdRP paralogs and distant DdRP relatives (Fig. 2). Consistent with their stronger divergence from QDE-1\textsuperscript{Ncr} than from each other (Fig. 1), QDE-1\textsuperscript{Tte} and QDE-1\textsuperscript{Mth} differ from QDE-1\textsuperscript{Ncr} in their enhanced ability to generate sRNA copies and reduced ability to synthesize long RNA products (Fig. 4). Because QDE-1\textsuperscript{Ncr}, QDE-1\textsuperscript{Tte}, and QDE-1\textsuperscript{Mth} originate from a single taxonomic order (Sordariales), these data argue that protein functions can undergo substantial changes over relatively short periods of evolutionary history. Notably, this effect was not due to a recent gene duplication event followed by functional specialization of the newly emerged paralogs because our BLAST searches did not reveal any additional RdRP gene in
the *N. crassa*, *T. terrestris*, and *M. thermophile* genomes besides QDE-1, SAD-1, and RRP-3.

What could be the biological significance of such accelerated functional evolution? One possible answer relates to the role of RNA silencing in cellular defense against viruses and transposable elements (3, 5, 6). Different species encounter distinct sets of pathogens, which necessitate corresponding changes in cellular defense mechanisms (64). Interestingly, *N. crassa* is a mesophilic fungus colonizing burnt vegetation and occasionally living plants (60, 61), whereas *T. terrestris* and *M. thermophila* typically inhabit self-heating composts (62). Exposure to diverse environments and correspondingly different sets of biohazards might have exerted substantial evolutionary pressure on cellular immunity factors, including QDE-1. An interesting direction for future studies would be to test whether this also rewired RNA silencing pathways in a more fundamental manner (e.g., incorporating sRNA products of QDE-1<sup>Tte</sup> and QDE-1<sup>Mth</sup> into signal amplification loops similar to the secondary siRNA pathway in *C. elegans* (16, 19–21, 39)).

The striking structural similarity between QDE-1<sup>Tte</sup> and QDE-1<sup>Mcr</sup> uncovered in this work (Fig. 8) suggests that evolution in this group probably proceeded through alteration of functionally important surface residues rather than major changes in the protein fold. In other words, QDE-1 orthologs show a combination of structural robustness and functional innovability (65). Although further work will be required to identify specific structural elements underlying functional divergence in the QDE-1 group, these will probably include amino acid residues directly interacting with the template-product duplex or affecting protein flexibility. Indeed, structural alignment of QDE-1<sup>Ncr</sup> or QDE-1<sup>Tte</sup> with the yeast polymerase II elongation complex (Protein Data Bank code 1R9T) (51) using structural similarity between the QDE-1 DPBB2 domain and the only DPBB domain of the polymerase II RPB1 subunit suggests that the QDE-1 enzyme may have to undergo major conformational changes to allow egress of template-product duplexes longer than ~10 base pairs.

The EM reconstructions presented here provide an unprecedented insight into solution structure of an RdRP enzyme (Figs. 9 and 10). In line with the earlier prediction (40), QDE-1 assembles into a pyramid-shaped homodimer with each of the two subunits containing a groove suitable for template-product binding. One of the two subunits in the EM density maps adopts a more "closed" conformation than the other. This supports the "two-stroke motor" model of QDE-1 activity that was proposed earlier based on the x-ray structure of QDE-1<sup>Ncr</sup> apoenzyme (40). Importantly, our data indicate that QDE-1 retains its dimeric form in the presence of a single-stranded DNA template (Figs. 9 and 10).

We finally note that the efficient synthesis of sRNA products by the newly isolated QDE-1 enzymes might facilitate a range of research and diagnostic applications. QDE-1<sup>Tte</sup> appears to generate more than one copy of a given template sequence (Fig. 4, A and B). Therefore, this enzyme might be especially useful for amplifying either an entire nucleic acid target or its parts accessible to the polymerase. Combined with deep sequencing technology, this may open up new possibilities in high-throughput analyses of the transcriptome composition, RNA conformation and ribonucleoprotein complex structure. Enzymatic properties of QDE-1<sup>Tte</sup> could be further improved by knowledge-based mutagenesis of its evolutionarily variable parts or by "shuffling" corresponding sequences with their counterparts from other QDE-1 orthologs.

In conclusion, our study argues that acquisition of novel enzymatic properties through divergence of orthologous sequences could be a more common evolutionary scenario than anticipated previously. This work also improves our understanding of molecular mechanisms underlying RdRP functions and expands the existing molecular biology toolkit. We predict that further comparative analyses of this remarkably diverse class of enzymes will be a rewarding experience for evolutionary biologists and biochemists alike.

**Author Contributions**—X. Q. and D. A. D. expressed the proteins (QDE-1<sup>Ncr</sup>, QDE-1<sup>Mth</sup>, QDE-1<sup>Tte</sup>, and QDE-1<sup>Mth</sup> (D607A)). X. Q. purified the proteins, conducted the thermofluor assay, crystallized QDE-1<sup>Tte</sup>, and solved its structure. X. Q. and F. M. H. conducted the polymerase activity assay. Y. H. W. generated the mutated polymerase II elongation complex (Protein Data Bank code 1R9T), and solved its structure. X. Q. and J. L. performed all EM and refinement studies. J. L. collected the diffraction data, with A. E. S., D. D. M., and J. L. wrote the paper.

**Acknowledgments**—We thank Andrew Wong for technical help with EM and Bernard Henriassat and Snezhka Oliferenko for helpful discussions.

**References**


Functional Evolution of Fungal RdRPs


Functional Evolution in Orthologous Cell-encoded RNA-dependent RNA Polymerases
Xinlei Qian, Fursham M. Hamid, Abbas El Sahili, Dina Amallia Darwis, Yee Hwa Wong, Shashi Bhushan, Eugene V. Makeyev and Julien Lescar

doi: 10.1074/jbc.M115.685933 originally published online February 23, 2016

Access the most updated version of this article at doi: 10.1074/jbc.M115.685933

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 65 references, 18 of which can be accessed free at http://www.jbc.org/content/291/17/9295.full.html#ref-list-1