Two novel thermally resistant endolysins encoded by pseudo T-even bacteriophages RB43 and RB49

Galina V. Mikoulinskaia,1,* Sergei V. Chernyshov,1 Maria S. Shavrina,1 Nikolai V. Molochkov,2 Valentina Ya. Lysanskaya3 and Andrei A. Zimin3

Abstract

Identification and cloning of genes as well as biochemical characterization of the gene products were carried out for two novel endolysins of pseudo T-even lytic bacteriophages RB43 and RB49, which represent different myovirus groups of the subfamily Tevenvirinae. Genes RB43ORF159c and RB49p102 were cloned in E. coli cells, and their products were purified to electrophoretic homogeneity with an up to 80 % yield of total activity. In respect to substrate specificity, both enzymes were found to be lytic L-alanoyl-D-glutamate peptidases belonging to the M15 family. The pH optimum functioning of both endolysins was within the range 7.0–9.0, whereas the optimal values of ionic strength were different for the two proteins (25 mM vs 100 mM for the RB43 and RB49 endolysins respectively). Both peptidases were thermally resistant, with the RB43 endolysin being more stable (it restored 81 % of enzyme activity and 96 % of secondary structure after a 10 min heating at 90 °C) than its RB49 counterpart (27% and 77% respectively). The possible origin of genes of lytic L-alanoyl-D-glutamate peptidases of myoviruses as a result of horizontal transfer in the variable parts of genomes between unrelated phages having a common host is discussed.

INTRODUCTION

Peptidoglycan-degrading enzymes (PDEs) break down peptidoglycan of bacterial cell walls, which results in the osmotic lysis of the bacterial cell. In the last 15 years, they have drawn a lot of attention among scientists, mainly as potential antibacterial agents, which could provide an alternative to conventional antibiotic therapy [1–4]. It is especially important in regard to the spread of antibiotic-resistant strains of bacterial pathogens, which are difficult to treat using traditional methods. Along with their ability to affect antibiotic-resistant strains, other advantages of PDEs include high activity, low immunogenicity, ability to synergistically interact with other antibacterial agents and, in many cases, species- and genus-specificity allowing the risk of dysbiosis to be considerably reduced [5].

Because of their multiplicity, a definite host range and peculiarities of life cycle, bacteriophages provide an inexhaustible source of new bacteria-specific PDEs. To lyse bacterial cells and ensure release of their progeny, lytic phages belonging to the families Podoviridae, Myoviridae and Siphoviridae of the order Caudovirales code for a system of lytic proteins, the main components of which are the pore-forming protein holin and the PDE endolysin [5]. Phages of Gram-negative hosts also carry auxiliary lytic proteins necessary for the destruction of the outer bacterial membrane: i-spanin and o-spanin, which form a complex [6].

By the specificity of the bond disrupted in peptidoglycan, endolysins are divided into five classes: (1) lysozyme-like muramidases hydrolysing the glycoside bond between N-acetylmuramic acid and N-acetyl glucosamine; (2) lytic transglycosylases, which act on the same bonds as muramidases but further catalyse the intramolecular transfer of O-muramyl residue to the own C6 hydroxy; (3) N-acetyl-β-D-glucose aminidases hydrolysing the glycoside bond between N-acetylglucosamine and N-acetylmuramic acid; (4) N-acetylmuramyl-L-alanine amidases hydrolysing the bonds between N-acetylmuramic acid and L-alanine; and (5) peptidases hydrolysing definite peptide bonds. Some
spanins of bacteriophages Considering lytic proteins in general, holins and one of the virinae (T4, RB43 and RB49) are shown in Fig. 1(b).

The objective of this work was to identify, clone and express genes, as well as to analyse activity and specificity of the gene products, of peptidoglycan hydrolases of pseudo T-even bacteriophages RB43 and RB49, which belong to different myovirus genera of the Tevenvirinae subfamily. Fundamentally, comparative study of endolysins from the bacteriophages of the Caudovirales order is important to elucidate the evolution of the lytic module and the role of horizontal gene transfer in the formation of phage genomes. Investigation of these enzymes could help to determine the evolutionary relationships between the lytic systems of unrelated phages having a common host, including T-even and pseudo T-even phages and bacteriophage T5.

**RESULTS**

**Identification, cloning and expression of genes of RB43 and RB49 endolysins**

In the genomes of myobacteriophages RB43 and RB49, we found ORFs coded by the genes RB43ORF159c and RB49p102 respectively (GenBank accession nos. YP_239135.1 and NP_891673.1). These hypothetical proteins are annotated in the genomes as having a d-alanyl-d-alanine carboxypeptidase region; in the first case, the protein is indicated as a potential member of the VanY peptidase (M15_2) family; in the second case, as a member of the M15_4 peptidase family.

The proteins are orthologous to the zinc-containing l-alanoyl-d-glutamate peptidase of the siphovirus T5 (EC 3.4.24.*) characterized by us earlier [8]. The product of the RB49p102 gene has a 35% identity to the peptidase; the product of the RB43ORF159c gene has a 44% identity. Both amino acid sequences contain two histidines and two asparagines constituting the conservative zinc-binding site of peptidases of the M15_3 family: H62, D69, D124, H127 - ORF159c, H62, D69, D114, H117 - ORFp102 (Fig. 1a). It should be mentioned that endolysin of bacteriophage T4, the well-studied relative of those pseudo T-even phages, is a glcosyl hydrolase – muramidase (lysozyme). The relative gene locations of lytic proteins (holins, endolysins and spanins) in the genomes of three representatives of the subfamily Tevenvirinae (T4, RB43 and RB49) are shown in Fig. 1(b).

Considering lytic proteins in general, holins and one of the spanins of bacteriophages T4, RB43 and RB49 have evident signs of similarity, in particular, there is 49% identity between holins and the product of gene t of phage T4 and 38–48% identity between o-spanins and the product of gene pseT.2 of bacteriophage T4. i-spanin, an analogue of the product of gene pseT.3, also has a homologous sequence in the RB49 genome (34% identity), whereas i-spanin of RB43 is much shorter and not homologous to them. Genes of lytic proteins of bacteriophages T4, RB43 and RB49 have a similar transcriptional organization: (i) they are located in different regions of the genome, (ii) spanins form a cluster and overlap by three nucleotides, (iii) spanins and endolysin are early genes and (iv) holin is located at the end of the genome within a late transcript (see Fig. 1b). (For comparison, in the genome of phage T5 all genes of lytic proteins are located sequentially.)

Genes RB43ORF159c and RB49p102 were cloned into expression plasmid pET30b under the bacteriophage T7 promoter; the constructions were called pEndoRB43 and pEndoRB49 respectively. The RB49p102 product (129 amino acid residues; 14.75 kDa) was not toxic for the producer strain cells and was effectively expressed in the E. coli strain BL21(DE3). The RB43ORF159c product (131 amino acid residues; 14.81 kDa) was toxic and caused lysis of cells of that strain. We managed to express this gene in another E. coli strain, C41(DE3), which was specially selected for toxic proteins [9].

The products of both genes lysed chloroform-permeabilized cells of E. coli, i.e. they both proved to be PDEs. Hereafter, these enzymes are referred to as EndoRB43 and EndoRB49 (from endolysin).

**Purification of EndoRB43 and EndoRB49**

EndoRB43 and EndoRB49 were purified to electrophoretic homogeneity using chromatographic techniques. The results of the purification procedure are given in Table 1. It is interesting that EndoRB43 showed an anomalous electrophoretic mobility as it run through the gel much slower than EndoRB49, although their molecular masses are almost the same. The rate of EndoRB43 movement was even lower than that of a larger enzyme from the bacteriophage T5 (EndoT5; 15.266 kDa) (Fig. 2). The most probable explanation of abnormal electrophoretic mobility involves an anomalous binding of SDS with a certain amino acid sequence. As was shown earlier, the replacement of a single amino acid residue could change mobility of a protein in the denaturing gel in such a way that its apparent molecular mass would grow by 2 kDa [10].

**Stability of enzyme preparations**

The preparations of EndoRB43 and EndoRB49 were stored at −20 °C for half a year in a storage buffer (25 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA), showing no loss in activity. At +4 °C and above, the proteins were subjected to autocatalytic cleavage, which was especially characteristic of EndoRB43. The addition of Zn2+ to the storage medium increased storage time at +4 °C 10-fold, apparently due to inhibition of autoproteolysis.
Activity of enzyme preparations

EndoRB43 and EndoRB49 showed different levels of enzymatic activity. Table 1 gives values of specific activity for the two proteins measured in the standard buffer, which was the same for both enzymes (see Methods). However, the maximal specific activity (measured under optimal, for each of the enzymes, conditions) was an order of magnitude higher in the case of EndoRB43: 20555.1±112.7 U mg⁻¹ versus 1708.70±109.09 U mg⁻¹ for EndoRB49.

Optimal conditions for enzyme operation

EndoRB43 and EndoRB49 were shown to be active in the pH range 7.0–9.0, with the optimal activity lying around pH 8.0 (Fig. 3a). There were, however, substantial differences in

Table 1. Purification of recombinant EndoRB49 and EndoRB43

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Fraction</th>
<th>Fraction volume, ml</th>
<th>Protein concentration, mg ml⁻¹</th>
<th>Specific activity*, U mg⁻¹</th>
<th>Total activity, U</th>
<th>Purification factor (fold)</th>
<th>Yield, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>EndoRB49</td>
<td>Crude extract</td>
<td>8.6</td>
<td>13.60</td>
<td>422.37±13.97</td>
<td>49400.12</td>
<td>1.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>Chromatography on Toyopearl 650M</td>
<td>25.0</td>
<td>4.05</td>
<td>471.42±46.74</td>
<td>47731.67</td>
<td>1.12</td>
<td>96.62</td>
</tr>
<tr>
<td></td>
<td>Chromatography on phosphocellulose</td>
<td>19.2</td>
<td>0.87</td>
<td>1286.42±108.16</td>
<td>21488.32</td>
<td>3.05</td>
<td>43.50</td>
</tr>
<tr>
<td>EndoRB43</td>
<td>Crude extract</td>
<td>9.0</td>
<td>18.750</td>
<td>438.78±23.87</td>
<td>74044.80</td>
<td>1.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>Chromatography on Toyopearl 650M</td>
<td>27.0</td>
<td>2.907</td>
<td>925.32±36.87</td>
<td>72627.30</td>
<td>2.11</td>
<td>98.09</td>
</tr>
<tr>
<td></td>
<td>Chromatography on phosphocellulose</td>
<td>22.3</td>
<td>0.433</td>
<td>6074.36±611.23</td>
<td>58653.46</td>
<td>13.84</td>
<td>79.21</td>
</tr>
</tbody>
</table>

*These values represent the mean±SD (n≥3).
EndoRB43 and EndoRB49 were found to retain a considerable part of their enzymatic activity even after a 30 min heating at 90°C (Fig. 4a, b; Table S1, available in the online version of this article). EndoRB43 was more resistant retaining two-thirds of its initial activity, whereas EndoRB49 retained only one-fourth of it. Examination of the secondary structure of the enzymes on the basis of circular dichroism (CD) spectra [before, during and after heating at 90°C; Fig. 5(a, b)] showed that the proteins refolded a significant part of their secondary structure after thermal denaturation and subsequent cooling. The calculation results of CD spectra changes at chosen wavelengths are represented in Fig. 5(c, d), as denaturation curves. At the separate wavelengths (210, 215, 220 nm) we observed sigmoidal profiles of thermal denaturation. The experimental profiles of denaturation fit the two-state transition model, and were used to calculate the half transition temperature ($T_{1/2}$). Within the framework of the two-state model, $T_{1/2}$ for both EndoRB43 and EndoRB49 was found to be 43±2°C. After the final cooling of a sample, EndoRB43 and EndoRB49 would renature 82±3% and 65±4% of its native structure respectively.

Calculation of the percentage of alpha-spiral segments in the enzyme molecules before and after heating indicated that EndoRB43 and EndoRB49 restored 96 and 77% of their alpha structure respectively, thus confirming their high conformational stability.

**Determination of the hydrolysis site**

To determine the site of hydrolysis in peptidoglycan, apart from EndoRB43 and EndoRB49, two other available peptidoglycan hydrolases with known substrate specificity were used: commercial preparation of crystalline egg lysozyme (muramidase) and a preparation of L-alanoyl-D-glutamate peptidase from bacteriophage T5 (EndoT5), obtained earlier in our laboratory. It was established that hydrolysis of NaBH$_4$-reduced peptidoglycan with EndoRB43 and EndoRB49 did not yield free glycosyl hydroxyls, whereas hydrolysis of acetylated peptidoglycan resulted in the formation of free amino groups (Table S2). Thus, the enzymes were peptidases.

In order to determine the exact location of the hydrolysis site, amino groups of peptidoglycan hydrolysates, which were released in the cleavage reactions catalysed by the above-mentioned four enzymes, were labelled with dinitrophenyl (DNP) derivative [1]. As shown in Fig. 6, the DNP-derivative formed in the hydrolysates of EndoT5, EndoRB43 and EndoRB49 was glutamic acid. Amino acid analysis of the remaining EndoRB43 and EndoRB49 peptidoglycan hydrolysates showed a decreased content of glutamic acid comparatively to the control (Table S3). Thus, both these enzymes are L-alanoyl-D-glutamate peptidases, like endolysin of bacteriophage T5. On the basis of their amino acid sequence (Fig. 1a), they can be attributed to peptidases of the subfamily C of the M15 family (according to the

The preferences of the two enzymes for ionic strength as measurements of activity depending on the buffer concentration revealed that EndoRB49 was most active in 100 mM buffer, whereas EndoRB43 had the highest activity in 25 mM buffer (Fig. 3b).

Estimation of the effects of various inhibitors, first of all, metallochelators, showed that none of the typical chelators of Ca$^{2+}$ or Mg$^{2+}$ (EDTA, EGTA, BAPTA; 1 mM) inhibited the enzymatic activity of EndoRB43 and EndoRB49, which makes the two enzymes different from their orthologue, EndoT5 (Fig. 3c). The activity of EndoRB49 was also unaffected by 1 mM 1,10-phenanthroline (Zn$^{2+}$ chelator), CaCl$_2$, MgCl$_2$ and MnCl$_2$. The activity of EndoRB43, however, was partially inhibited by those agents, with the enzyme retaining 20–70% of its initial activity. The higher lability of EndoRB43 might result from its structural peculiarities, e.g. its 'loose' structure. Both EndoRB43 and EndoRB49 were quantitatively inhibited by excessive Zn$^{2+}$, that is common for Zn-containing peptidases. The enzymes were also low-active in the potassium phosphate buffer, that could arise from low solubility of phosphates of catalytic zinc. Surprisingly, EndoRB43 and EndoRB49 were not affected substantially by Triton X-100, which often increases the activity of endolysins by interacting with the substrate [11].

**Thermal resistance**

One of the objectives of this study was to analyse resistance of the enzyme preparations to high temperatures.
The conservative residues H62, D69, D124 and H127 in EndoRB43 and homologous residues H62, D69, D114 and H117 in EndoRB49 should coordinate zinc. However, in contrast to EndoT5, whose activity is regulated by Ca\(^{2+}\) and is effectively inhibited by Ca\(^{2+}\) chelators (EDTA, EGTA and BAPTA), EndoRB43 and EndoRB49 are not Ca\(^{2+}\)-dependent. Nevertheless, they still can be attributed to the same family of small lytic L-alanoyl-D-glutamate peptidases of phage origin.

**Evolutionary position of RB43 and RB49 endolysins and their orthologues**

A BLAST query for homologous amino acid sequences in the Caudovirales phage genomes (nr database) reveals about 200 evidently similar sequences (E-value < 10\(^{-11}\)). We have analysed the extent of the similarity of 96 homologous sequences of endolysins of tailed phages selected as described in Methods and the evolution of these sequences. The results are given in the form of a tree in Fig. 7.

EndoRB43 is located on the branch formed by enzymes of closely related phages, infecting enterobacteria (RB16, Lw1) and *Klebsiella* (KP15, KP27). It is interesting that some of these phages infect bacteria from quite remote groups, e.g. *Cronobacter* phage vB_CsaM_leB and *Citrobacter* phages Miller and Margaery (see Fig. 7).

EndoRB49 is located on the same branch as enzymes of the closest pseudo T-even phages, Phi1 and JSE. Interestingly, close to the branch of proteins of RB49-type phages is a small branch of proteins of *Salmonella* siphoviruses. Among them are lytic siphovirus 9NA and two closely related bacteriophages of *S. enterica* serovar Newport, which are still poorly studied. These phages recognize O-antigens at the surface of *Salmonella* and are very attractive targets for control of various salmonelleses, including the most severe lethal infections.

It is interesting that close to these enzymes is the putative endolysin of *Moraxella catarrhalis* prophage of the moderate siphovirus Mcat9. The length of this protein is 218 amino acid residues, apart from the catalytic domain it contains a putative peptidoglycan-binding domain (pfam01471). It is obvious that this protein is a high-molecular peptidase, which has a modular structure and is homologous to EndoRB43 and EndoRB49.

The modular structure is a feature of the absolute majority of homologous endolysins of phages represented on the tree that infect the Gram-positive bacteria of *Listeria*, *Bacillus* and *Geobacillus* genera. Homology is manifested in the catalytic domains of these modular proteins.

If summarized, it could be supposed that the L-alanoyl-D-glutamate peptidase gene is a gene-traveller, which was actively transferred between the genomes of phages of different taxonomic groups but remained within the same ecological phage community.

**DISCUSSION**

In this work we have characterized two novel enzymes, endolysins of bacteriophages RB43 and RB49 (EndoRB43 and EndoRB49), which are components of lytic systems required for the cell lysis from within and the release of phage progeny. EndoRB43 and EndoRB49 proved to belong to the group of L-alanoyl-D-glutamate peptidases, zinc-containing enzymes of the M15 family. Earlier we characterized
Fig. 4. Thermal resistance of EndoRB43 (a) and EndoRB49 (b). Relative activity units are given as percentage values relative to initial sample activity defined as 100%. Measurements were performed in triplicate. Results are stated as mean values±SD.
biochemical properties [8] and evaluated structure [13], as well as thermal stability and bacteriolytic effect [14], of an orthologous protein of T5 syphovirus (EndoT5). It is reasonable to compare these three enzymes, which belong to the same family and are evidently of the same origin. They have a lot in common, but they also have significant differences.

Their common properties include small size (from 129 to 137 amino acid residues), substrate specificity, evidently the same Zn-binding site (one per molecule), as well as resistance of their structure and enzymatic activity to high temperatures.

Thermal resistance is often found in different endolysins: there are peptidases [15], amidases [16] and muramidases [17], which possess such a property. Most often the thermal resistance is evaluated by preserving the enzymatic activity after heating [15, 18]. It was shown that EndoT5 and EndoRB43 retain 2/3, EndoRB49 – 1/4 of the initial enzymatic activity after 30 min of heating at 90 °C, renaturing 80, 82 and 65 % of the secondary structure respectively after recooling. The half transition temperature $T_{1/2}$ for both EndoRB43 and EndoRB49 was found to be 43±2 °C, which is somewhat less than the value measured for EndoT5 – 56 ±1 °C [14], but almost identical to Tm=44 °C of another thermostable endolysin Lys68 [19]. These relatively low $T_{1/2}$ values suggest that the resistance of these proteins to temperature arises from the ability of the secondary structure and functional activity to restore after heating and recooling.

Among the differences between the proteins, one should mention, first of all, different values of their maximal specific activity (EndoRB43>EndoT5>EndoRB49). A very distinct difference concerns their ionic strength optima: EndoRB43 and EndoT5 favour low levels of ionic strength (25 mM), whereas EndoRB49 operates best at the levels close to physiological values (100 mM). Surprisingly, there is no connection between the optima of ionic strength and the calculated values of the enzyme pK (8.31 for T5, 9.5 for both RB). A radical difference is that among the three Zn-containing peptidases, only EndoT5 is a Ca$^{2+}$-activated protein, according to its sensitivity to Ca$^{2+}$ chelators EDTA, BAPTA and EGTA. Earlier we supposed that binding of paramagnetic ions (which Ca$^{2+}$ belongs to) by EndoT5 is mediated by a long loop formed by 111–130th amino acid residues [13]. As evident from the multiple alignment data (Fig. 1a), this region of EndoT5 has significant differences.

Fig. 5. Far-UV CD spectra of EndoRB43 (a) and EndoRB49 (b) at 20 °C before heating (white circles), at 90 °C (triangles), and at 20 °C after heating (black circles). Curves of thermal denaturation of EndoRB43 (c) and EndoRB49 (d) registered by CD at the wavelength of 220 nm (black circles), 215 nm (white circles) and 210 nm (black triangles).
The first pieces of evidence of peptidoglycan hydrolases date back to the 1970s [21, 22]. However, neither any data on the amino acid sequence of those enzymes, nor any information about their genes were available yet. Nevertheless, it was clear that endolysins with this type of substrate specificity should not be rare, since the bond between L-alanine and D-glutamic acid is common for peptidoglycan in general and for peptidoglycan of Gram-negative bacteria in particular. The first L-alanoyl-D-glutamate peptidases, for which amino acid sequence was determined, were two-domain endolysins of moderate siphoviruses infecting the Gram-positive bacteria Listeria [23]. Since then, researchers have characterized a few modular L-alanoyl-D-glutamate peptidases from Gram-positive sources [24, 25], as well as the small, single-domain peptidase EndoT5 from syphovirus T5, infecting a Gram-negative host [8]. Recently, another L-alanoyl-D-glutamate peptidase with Gram-negative specificity has been characterized: the peptidase from Klebsiella bacteriophage K27 [26]. At the same time, the number of orthologous sequences in phage genomes is much larger than the number of proteins characterized. Orthologues can be predominantly found in phages of Gram-negative bacteria such as Yersinia, Klebsiella, Escherichia, Salmonella, Pseudomonas. Thus, identification and characterization of new L-alanoyl-D-glutamate peptidases with Gram-negative specificity is just a matter of time.

Taking into account the systematic place of bacteriophages RB43 and RB49, the discovery of EndoT5-orthologous endolysins seems quite intriguing. These phages belong to the Myoviridae family, which represents another taxonomic group than the lytic syphovirus T5 and the above-mentioned moderate listeria phages. Phages RB43 and RB49 are often called pseudo T-even due to their significant morphological similarity with T-even phages and, at the same time, a global difference in the DNA composition (DNA of pseudo T-even phages contains cytosine, in contrast to DNA of T-even phages which contains a hydroxymethylated and glycosylated derivative of cytosine) [27]. The closest relatives of phages RB43 and RB49 are T4-type and similar phages. Recently, a new classification of Myoviridae has been suggested, according to which RB43 and RB49 belong to different groups (RB43-type and RB49-type) but the same genus of T4-like phages of the subfamily Teequatrovirinae [28]. The common origin and close relativity of T4-like phages is indicated by the similarity of large structural and replicative modules in their genomes. Modular organization plays a significant role in the evolution of these phages [29].

An overwhelming majority of Teevenvirinae phages, including phage T4, carries sequences orthologous to glycosyl hydrolase (lysozyme T4, the product of gene e). At the same time, only 12 representatives of the taxon, including RB43 and RB49, have sequences orthologous to EndoT5.

How and when did the gene of L,D-peptidase evolve to get into the genomes of the above-mentioned myoviruses? The source of new genes in phage genomes can be either genomes of lysogenic cells of bacterial hosts or coinfected...
Fig. 7. Phylogenetic analysis of 97 endopeptidases. Endolysins of bacteriophages of Gram-positive bacteria are marked by red circles, endolysins of bacteriophages of Gram-negative bacteria are marked by blue circles, ADAM7 is marked by a black circle. The tree with the highest log likelihood (−3444.3, 3288) is shown. Numbers reflect the percent support values from 1000 bootstrap replicates. Abbreviations denote proteins of the following bacteriophages (GenBank accession numbers are given in brackets): RB49, Enterobacteria phage RB49 (NP_891673.1); Phi1, Enterobacteria phage Phi1 (YP_001469446.1); ECD7, Escherichia phage ECD7 (ASJ80195.1); JSE, Enterobacteria phage JSE (YP_002922178.1); SP069, Salmonella phage FSL SP069 (AGF89556.1); Sergei, Salmonella phage vB_SenS_Sergei (APU92900.1); 9NA, Salmonella phage 9NA (YP_009101227.1); P88, Enterobacteria phage P88 (YP_009113074.1); PY100, Yersinia phage PY100 (CAJ284446.1); phiTE, Pectobacterium phage phiTE (YP_007392609.1); phiP27, Enterobacteria phage phiP27 (NP_543082.1); MP1, Morganella phage vB_MmoM_MP1 (YP_009279959.1); PM2, Proteus phage PM2 (ASZ76368.1); Pm461, Proteus phage vB_Pm461 (YP_009195522.1); eAU, Edwardsiella phage eAU-183 (YP_0009004687.1); S-CBS4, Synechococcus phage S-CBS4 (AGN30423.1); SPC35, Salmonella virus SPC35 (YP_004306522.1); SP1, Salmonella phage SP1 (ATI18557.1); 3/49, Shewanella sp. phage 3/49 (YP_009103932.1); NR01, Salmonella phage NR01 (YP_009283467.1); Shivani, Salmonella phage Shivani (YP_009196850.1); ST64T, Salmonella phage ST64T (NP_720320.1); PG3, Salmonella enteric phage PG3 (C5A09701.1); Miller, Citrobacter phage Miller (NC_025414.1); RB43, Enterobacteria phage RB43 (YP_239135.1); Lw1, Escherichia phage Lw1 (YP_000306075.1); RB16, Enterobacteria phage RB16 (YP_0003858447.1); KP15, Klebsiella phage KP15 (YP_003580002.1); KP27, Klebsiella phage KP27 (YP_007348726.1); leB, Cronobacter phage vB_CsaM_leB (AOG16285.1); PaMx74, Pseudomonas phage PaMx74 (YP_009195477.1); Margaery, Citrobacter phage Margaery (YP_009194771.1); phiJL001, alpha proteobacteria phage phiJL001 (YP_224014.1); Mcat9, Moraxella phage Mcat9 (AKI27396.1); G11, Enterobacteria phage phiEcoM-G11 (YP_001595916.1); AAT-1, Pseudomonas phage AAT-1 (AME18059.1); Pm34, Proteus phage vB_Pm34_Pm34 (APU92697.1); PaMx28, Pseudomonas phage PaMx28 (YP_009210649.1); Ss1, Cronobacter phage vB_CsaP_Ss1 (AKI67543.1); Xoo-sp2, Xanthomonas phage Xoo-sp2 (ANT45272.1); H61, Pseudalteromonas phage vB_PspS-H61 (ANJ65301.1); H40/1, Pseudalteromonas phage vB_PspS-H40/1 (ANI22027.1); sal2, Salmonella phage 100268_sal2 (YP_009344175.1); Stitch,
Salmonella virus Stich (YP_009145980.1); Sttp1, Salmonella phage Sttp1 (ARQ96246.1); phiR, Yersinia phage phiR201 (YP_007237012.1); 7–11, Salmonella phage 7–11 (YP_004782425.1); VvAW1, Vibrio phage VvAW1 (YP_007513836.1); GAP52, Cronobacter phage vB_CsaP-GAP52 (YP_006987695.1); IME260, Klebsiella phage vB_Kpn_IME260 (APTA1082.1); phiST, Cellulophaga phage phiST (YP_007673455.1); eiDWF, Edwardsiella phage eiDWF (ADV36485.1); Xp15, Xanthomonas phage Xp15 (YP_239293.1); My1, Pectobacterium phage My1 (YP_006906290.1); GBK2, Geobacillus phage GBK2 (YP_009010493.1); PR1, Providencia phage vB_Pre5, PR1 (AQT25280.1); SP-10, Bacillus phage SP-10 (YP_007003388.1); LP-101, Listeria phage LP-101 (YP_009044822.1); A118, Listeria phage A118 (CA59362.1); E3, Geobacillus virus E3 (YP_009223760.1); LP-026, Listeria phage LP-026 (YP_009044833.1); LP-114, Listeria phage LP-114 (YP_009045157.1); Deep Blue, Bacillus phage Deep Blue (YP_009285532.1); P70, Listeria phage P70 (YP_006905871.1); BM5, Bacillus phage BM5 (YP_009223175.1); B4, Bacillus phage B4 (YP_006908235.1); Stahl, Bacillus phage Stahl (YP_009203662.1); B025, Listeria phage B025 (YP_001468664.1); Spock, Bacillus phage Spock (YP_008770279.1); Riley, Bacillus phage Riley (YP_009055819.1); BigBertha, Bacillus phage BigBertha (YP_008771084.1); Troll, Bacillus phage Troll (YP_008430843.1); Pascal, Bacillus phage Pascal (YP_009151490.1); Palmer, Bacillus phage Palmer (YP_009210060.1); Pony, Bacillus phage Pony (YP_008771342.1); Page, Bacillus phage Page (YP_008770531.1); Stills, Bacillus phage Stills (YP_009196491.1); poppyseed, Bacillus phage poppyseed (AGY58041.1); Pavlov, Bacillus phage Pavlov (YP_009197493.1); A006, Listeria phage A006 (YP_001468860.1); Slash, Bacillus phage Slash (YP_008771960.1); A500, Listeria phage A500 (YP_001468411.1); phiLM4, Listeria phage phiLM4 (ABG75905.1); SPO1, Bacillus virus SPO1 (YP_002300379.1); EPS57, Escherichia virus EPS57 (YP_001836966.1); P35, PlyP35 Listeria phage P35 (AY53213.1); SP-15, Bacillus phage SP-15 (YP_009302456.1); Moonbeam, Bacillus phage Moonbeam (YP_009151593.1); Eldridge, Bacillus phage Eldridge (YP_009274740.1); P12002S, Poliarbacter phage P12002S (YP_009195706.1); 1/32, Flavobacterium sp. phage 1/32 (AHK11392.1); ADAM7, Homo sapiens ADAM7 protein (AAH58037.1).

For T4-related phages, the first source is considered unlikely, since these phages destroy DNA of the host at an early stage of the infection [30]. A more probable explanation implies the existence of an unknown phage, whose coinfection with the common ancestor of the discussed myoviruses resulted in the horizontal transfer of the genome into its genome.

The role of horizontal gene transfer in the evolution of phage genomes has been discussed many times and is accepted to be essential [27, 31, 32]. In the case of the phages discussed, we have an example of replacement ‘by position’, when the gene of a lytic enzyme is replaced with a non-orthologous functional analogue. It should be noted that all the lytic genes of bacteriophages T4, RB43 and RB49 are located in so-called hyperplastic regions (HPRs) of their genomes. HPRs are rich in nonessential genes and these genes are not clustered in structural and replicative modules, characteristic for T4 relatives [33]. The location of lytic genes in HPRs could provide one of the ways of phage adaptation, allowing host cell lysis efficiency, yield, dissemination and host range to be varied without affecting life-sustaining functions. As for the mechanisms, by which gene translocation could happen, there is a few of them: homologous, nonhomologous or site-specific recombination. It seems very probable, though, that the translocation is mediated by promoter early stem-loops (PESLs), mobile regulatory elements which have been recently discovered in T4-like phages. They can serve as vectors for horizontal transfer and are predominantly localized in the hyperplastic regions of the genome [34, 35].

Another important aspect is the comparative evolutionary analysis of endolysin sequences in the genomes of phages T4, RB43 and RB49. We know that the genome of RB49 is very similar to that of T4 in respect to gene composition and location. In contrast, the RB43 genome is more different from the T4 one, with the differences being primarily related to the inversion of a large part of the genome and some rearrangements [30]. On the basis of the evolutionary tree (Fig. 7), one can suppose that the gene of lytic peptidase was transferred into the genome of a common ancestor of RB49 and RB43 myoviruses, whereas the partial inversion of the RB43 genome occurred at a later evolutionary stage. Comparison of evolutionary trees built for single proteins sometimes gives information on the succession of evolutionary processes in the genomes which differs from the information taken from the trees built on the basis of general sequence similarity. This may be a consequence of modular evolution, which is common in T4 phages and implies single-event translocations of large parts of the genome. Anyway, in the case of T4, RB43, RB49 and their relatives, our data is consistent well with the tree based on the similarity of fibrinectins [36], but it poorly agrees with the trees built on the basis of general sequence similarity [28, 32]. The latter trees indicate an earlier divergence of RB43-type phages from their common branch with T4 and RB49. However, such an early divergence would imply two independent cases of transfer of the peptidase gene into the genomes of phages of the RB43 and RB49 groups, which seems unlikely.

Finally, it should not be left aside that phage endolysins are gradually becoming biotechnologically important proteins due to their capacity to serve as an alternative to antibiotics. Destroying peptidoglycan, they can cause osmotic lysis of bacterial cells, and this is very important in view of the problem of resistance to antibiotics, which has become a matter of serious concern in the world community. There are many examples of successful use of peptidoglycan hydrolases for lysis of Gram-positive bacteria [37–39]. A technology has been developed using complex chimeric modular endolysins against Gram-negative bacteria. Such proteins carry a membrane-active antimicrobial peptide [40]. It has also been demonstrated by a few laboratories [19, 41, 42], including our group [14], that cultures of Gram-negative bacteria can be lysed by endolysins in the presence of agents permeabilizing the outer membrane. The
list of such agents includes cationic detergents, cationic peptides, weak acids, complex-forming agents [43]. The enzymes examined in this work have an important advantage: being highly active, they possess an evident thermal resistance. Thermal resistance implies a high conformational stability, and this is important for medical preparations. We believe that small size, high activity and thermal resistance open good perspectives in the use of these novel proteins for pharmaceutical purposes, for treatment and prophylaxis of external infections.

METHODS

Materials

*Escherichia coli* strains B, XL10 Gold, BL21(DE3), C41 (DE3) and bacteriophages RB43<sup>+</sup>, RB49<sup>+</sup> were taken from the collection of the Laboratory of Molecular Microbiology of the Institute of Biochemistry and Physiology of Microorganisms (IBPM RAS). Plasmid pET30a was provided by Novagen (USA). Bacteria and phages were grown either in liquid LB broth or on agarized LB media. Selection of clones chased, unless otherwise stated, from either ICN (Irvine, CA) or Sigma (St. Louis, MO).

Cloning of genes RB43ORF159c and RB49p102

Genes RB43ORF159c and RB49p102 were amplified using the PCR method. The phage genome DNA was used as a matrix; oligonucleotides containing sites hydrolysed by restriction endonucleases *Nde* I and *Xho* I were used as primers. The reverse primers included the own translational stop codons of the ORFs, the recombinant proteins did not contain the His-tag, and their amino acid sequences were completely identical to the natural ones. The fragments were cloned in pET30b vector at *Nde* I and *Xho* I sites in *E. coli* XL10 Gold cells. The clones carrying the insert were selected by PCR and electrophoresis in 1 % agarose. The constructions were validated by sequencing. The constructions obtained were named pEndoRB43 and pEndoRB49 and were further used to transform cells of *E. coli* strains BL21(DE3) or C41(DE3). The synthesis of endolysins was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at a culture density corresponding to the absorbance A<sub>550</sub> = 1.0; the cells were harvested by centrifugation 3 h later.

Isolation and purification of endolysins

Proteins were isolated and purified from cells of *E. coli* strain C41(DE3) cultured in 250 ml of liquid LB medium (−1 g of cell mass per isolation). Purification was conducted using chromatography techniques as described earlier [8], on Toyopearl DE50M (TosoHaas, Stuttgart, Germany) and phosphocellulose (Whatman, UK) columns.

Enzyme activity assay

The lytic activity of the enzymes was measured on cells of overnight *E. coli* B culture, which were preliminary permeabilized with chloroform as described earlier [8]. The substrate-providing cells were suspended in the reaction buffer (25 mM Tris-HCl, pH 8.0, containing 37.5 mM NaCl and 0.1 % Triton X-100). The reaction was initiated by the addition of the enzyme. The activity was determined spectrophotometrically, by the decrease of initial optical density (1.0) at 450 nm, in 1 cm acrylic cuvettes at room temperature. An activity unit was defined as the quantity of enzymes that provides the rate of optical density decrease of 1.0 optical unit min<sup>−1</sup>. All activity data were calculated from three independent measurements.

Determination of optimal conditions for operation of the enzymes’ pH optima of the enzymatic reactions were determined at 25 °C using a series of Tris-HCl buffers (50 mM) with pH ranging from 6.0 to 9.5. To evaluate the effect of salts, cations and chelators, the corresponding agent was added, at a necessary concentration, to the standard reaction medium (50 mM Tris-HCl, pH 8.0, containing 0.1 % Triton X-100) and, after mixing, the reaction was started by the injection of the enzyme.

Evaluation of thermal resistance

Samples of the enzyme preparation – 100 µl of the protein solution in the storage buffer (25 mM Tris/HCl, 200 mM NaCl and 1 mM EDTA; pH 8.0) – were incubated in a solid state thermostat (DNA-Technology, Russia) at different temperatures for 5, 10 and 30 min. After the incubation, the samples were cooled on ice for 10 min and assayed for the enzyme activity by the turbidimetric method. The control samples were incubated for the same time on ice.

Measurement of CD spectra

CD spectra in the far UV region (250–190 nm) were recorded using a spectropolarimeter JASCO J-815 with a thermostated cell (Jasco, Japan). The measurements were performed in 1 mm cuvettes; the protein concentration in samples was 0.18 mg ml<sup>−1</sup>. After subtraction of the buffer baseline, the spectra obtained were converted into molar ellipticity ([θ]) according to the following formula:

\[
[\theta] = \frac{\text{data} \times \text{Mr}}{l \times c}
\]

where [θ] is the molar ellipticity, deg cm<sup>2</sup> dmol<sup>−1</sup>; data are the data from the spectropolarimeter, mdeg; Mr is the average molar mass of a single amino acid residue (g mol<sup>−1</sup>);

l is the optical length, mm;

c is the concentration of protein, mg ml<sup>−1</sup>.

The secondary structure of proteins was calculated using the CONTINLL module of the CDPro program [44].
The heating-induced changes in the secondary structure of the protein were monitored by measuring the ellipticity at 210, 215 and 220 nm; temperature was raised with the rate of 1 °C min⁻¹. After heating to 90 °C, samples were cooled to 20 °C for 30 min, and then spectra were registered again. The relative ellipticity changes upon temperature increasing at the chosen wavelengths were calculated using the following formula:

\[
\frac{[\theta_i] - [\theta_{\text{min}}]}{[\theta_{\text{max}}] - [\theta_{\text{min}}]} \times 100\% 
\]

where \([\theta_i]\) is the ellipticity value at a given temperature and wavelength, \(\text{deg cm}^2 \text{dmol}^{-1}\);

\([\theta_{\text{min}}]\) is the minimal ellipticity value at a given wavelength, \(\text{deg cm}^2 \text{dmol}^{-1}\);

\([\theta_{\text{max}}]\) is the maximal ellipticity value at a given wavelength, \(\text{deg cm}^2 \text{dmol}^{-1}\).

**Extraction of peptidoglycan**

Peptidoglycan was extracted by Streshinskaia’s method [45], with small modifications. *E. coli* cells (11.8 g, wet weight) were resuspended in 100 ml of 10 mM Tris-HCl buffer (pH 8.0), and then 75 ml of 10% SDS solution was added. The mixture was boiled for 45 min and kept at 4 °C for 2 h. Then the suspension was sonicated (75W, 1 min) and centrifuged at 30000 \(\times\) g for 10 min. Supernatant was then centrifuged at 28000 \(\times\) g for 1 h. The pellet was resuspended in 200 ml of 10% trichloracetic acid and incubated at room temperature under stirring for 48 h, this was followed by centrifugation at 20000 \(\times\) g for 30 min. The pellet was washed three times with warm water followed by centrifugation at 20000 \(\times\) g for 30 min, resuspended in 10 ml of distilled water and lyophilized.

Cell walls were acetylated with acetic anhydride as described earlier [46]. The oxidized redox groups were reduced with NaBH₄ by Ward’s method [47].

**Determination of substrate specificity of the enzymes**

The substrate specificity of the enzymes was determined with acetylated or reduced *E. coli* peptidoglycan as a substrate (2 mg ml⁻¹). The enzymes (none in the control samples; egg lysozyme, endolysins of bacteriophages T5, RB43 and RB49 in other samples) were added at concentration 3 µg ml⁻¹. The reaction was carried out in 0.025 M Tris-HCl buffer (pH 8.0) at room temperature (the buffer for EndoRB49 additionally contained 75 mM NaCl). After hydrolysis, the samples were centrifuged, and the supernatant was assayed for reducing and amino groups released in the course of hydrolysis. The quantity of reducing groups was determined by Park and Johnson’s method [48]. The quantity of free NH₂ groups was measured according to Ghuysen’s protocol [49].

To determine the exact site of hydrolysis, peptidoglycan was suspended in 1 ml of 25 mM Tris-HCl buffer (pH 8.0) (the buffer for EndoRB49 additionally contained 75 mM NaCl).

Five 200 µl samples were taken; the first, second and third samples were supplemented with RB43, RB49 and T5 peptidases respectively (3 µg each), the fourth sample was supplemented with the same amount of egg muramidase, and the fifth sample was used as a control. After 12 h, the samples were treated with dinitrofluorobenzene and then subjected to acidic decomposition and ether extraction as described above [49]. 2,4-dinitrophenyl (DNF) derivatives of amino acids were analysed by thin-layer chromatography on Kieselgel 60F254 plates (Merck, Germany) in a system of chloroform:methanol:benzyl alcohol:concentrated ammonium:water (30:30:30:6:2) [49]. The qualitative and quantitative analysis of amino acids remaining in the solution was performed with a Microtech T-339 amino acid analyzer (Microtech, Czech Republic).

**Standard analytical techniques**

The protein concentration was determined by the Warburg and Christian method [50], using white egg lysozyme as the standard. Protein samples were analysed by denaturing electrophoresis in a 12% polyacrylamide gel by the Laemmli method [51], using a standard kit of marker proteins containing β-galactosidase (116 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), lactate dehydrogenase (35 kDa), REase Bsp981 (25 kDa), β-lactoglobulin (18.4 kDa) and hen egg white lysozyme (14.4 kDa). Electrophoresis was carried out for 1.0 h at room temperature (field intensity, 15 V cm⁻¹). Gels were stained with Coomassie Brilliant Blue G-250 (Sigma, St. Louis, MO) and washed with distilled water.

**Software**

Nucleotide and amino acid sequences were analysed using Gene Runner (v. 3.0) (Hastings software). For analysis of homologous amino acid sequences, we used the BLAST search program [52] and COBALT multiple alignment tool [53] from the server of the National Center of Biotechnology Information (National Library of Medicine, USA; www.ncbi.nlm.nih.gov/blast/). The peptidases were classified using the database MEROPS [12], which can be accessed by the following link: http://merops.sanger.ac.uk/.

**Phylogenetic analysis**

Using the PSI-BLAST program, the amino acid sequences of the bacteriophages RB43 and RB49 endolysins were compared with the amino acid sequences of caudate bacteriophages of the order Caudovirales (tailed phages (taxid: 28883)) in the nr databases (GenBank+PDB+SwissProt+PIR+ PRF). Two successive iterations were made. From the sequences found, 84 sequences were selected with E-values below 10⁻¹¹. Repeated sequences (completely identical or differing by two–three amino acids) were deleted from a file intended for further analysis (only one sequence from each group was left). Multiple alignment of the 38 remaining amino acid sequences was performed using the Clustal W program [54]. The construction of the tree was carried out using the MEGA6 software package [55]. Cladistic analysis was carried out using the maximum likelihood method based on the JTT model [56]. The bootstrap
consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed [57]. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The initial tree for the heuristic search was obtained automatically by applying the maximum parsimony method. The analysis involved 97 amino acid sequences. All positions containing gaps and missing data were eliminated. At the tree inference following options were used: subtree–pruning–regrafting – fast (SPR level 3), branch swap filter – strong. The far-related amino acid sequence was chosen to determine the root of the phylogenetic tree. For this purpose, a sequence of the ADAM7 protein from the family of metalloendopeptidases and disintegrins (ADAMs) [58] was used. ADAM7 is the protein closest to the root of the phylogenetic tree of the ADAMs family [58]. The root of the endolysins tree was placed by the mid-point rooting method at the mid-point of the distance between the ADAM7 sequence branch and other branches.

References

Funding information
This work was supported in part by the Russian Foundation for Basic Research (RFBR) grant No 18-04-00492.

Acknowledgements
We are very grateful to Dr Odinokova I.V. (Institute of Theoretical and Experimental Biophysics RAS) for kind help with the electronic artwork and valuable comments.

Conflicts of interest
The authors declare that there are no conflicts of interest.


32. Filé J, Baptiste E, Susko E, Krisch HM. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol Biol Evol 2004;21:1688–1696.


