Sequence Variations in the Non-Coding Sequence of CTX Phages in Vibrio cholerae

Eun Jin Kim¹², Hyun Jin Yu¹², and Dong Wook Kim¹²

¹Department of Pharmacy, College of Pharmacy, Hanyang University, Ansan 15588, Republic of Korea
²Institute of Pharmacological Research, Hanyang University, Ansan 15588, Republic of Korea

Introduction

The CTX phage (cholera toxin phage) that provides the cholera toxin gene to Vibrio cholerae strains, making them toxigenic bacteria, has been classified into two major types, classical type CTX phage (CTX-cla) and El Tor type CTX phage (CTX-1) [10]. The CTX phage consists of 10 genes, rstR, rstA, rstB, psh, cep, orfU, ace, zot, ctxA, and ctxB. Whereas the genes from rstA to ctxB are homologous between two phages, rstR is phage type-specific and has been used to distinguish two phages [11]. ctxB is homologous in two phages; however, two SNPs (single nucleotide polymorphisms, 115th and 203rd nucleotides) have been noticed as phage type-specific [20]. Although the DNA sequences of the CTX phage genes have been used to discriminate different CTX phages, the non-coding sequence between ctxB and rstR has not been extensively studied.

Atypical El Tor V. cholerae strains that harbor the mosaic CTX phages between the classical and El Tor CTX phages have emerged since 1991 and, currently, most of the clinical isolates of V. cholerae are atypical El Tor variants that contain the mosaic CTX phage [19]. Various mosaic CTX phages have been classified among the strains belonging to Wave 2 and 3 within the 7th cholera pandemic [17]. CTX-2 in Wave 2 strains contained classical rstR and ctxB, while the rest of the phage genome is similar to CTX-1. CTX-3 ~ CTX6 in Wave 3 strains contained a CTX-1-like sequence except that they contain classical ctxB [12].

CTX phage is a single strand DNA phage that undergoes a replicative form double strand circular DNA, pCTX, during replication [15]. The integration of CTX phage genome on the V. cholerae chromosome is mediated by the host Xer recombination machinery [3]. XerC and XerD recombinases originally mediate the chromosomal dimer resolution after the chromosomal DNA replication via a 28 bp site, dif (deletion-induced filamentation), and CTX phage exploits XerCD for integration in the chromosomal DNA [3]. A molecular mechanism that explains the integration of pCTX on the dif site has been demonstrated [2]. The attP sequence (or XerC binding sequence) is

This study focused on the variations in the non-coding sequences between ctxB and rstR of various CTX phages. The non-coding sequences of CTX-1 and CTX-cla are phage type-specific. The length of the non-coding region of CTX-1 and CTX-cla is 601 and 730 nucleotides, respectively. The non-coding sequence of CTX phage could be divided into three regions. There is a phage type-specific Variable region between two homologous Common regions (Common regions 1 and 2). The non-coding sequence of RS1 element is similar to CTX-1 except that Common region 1 is replaced by a short RS1-specific sequence. The non-coding sequences of CTX-2 and CTX-cla are homologous, indicating the non-coding sequence of CTX-2 is derived from CTX-cla. The non-coding region of CTX-O139 is similar to CTX-cla and CTX-2; however, it contains an extra phage type-specific sequence between Common region 2 and rstR. The variations in the non-coding sequences of CTX phages might be associated with the difference in the replication efficiency and the directionality in the integration into the V. cholerae chromosome.

Keywords: V. cholerae, CTX phage, non-coding sequence
located in the non-coding sequence of pCTX that spans from ctxB and rstR of the circular pCTX phage genome [16].

The attL and attR sites, generated by the integration of the CTX phage, have been described as the “end repeat” empirically [7]. As shown in Fig. 1, the integration of pCTX in the chromosome via attP and the dif sequence splits the non-coding sequence of pCTX into two parts; (i) ig-1 (intergenic sequence-1, previously used to describe the sequence between attL and rstR), and (ii) 3’ UTR (untranslated region) sequence between ctxB and attR (Fig. 1). This process is similar in the integration of pRS1.

V. cholerae strains have two chromosomes (chromosome 1 of about 3 M bp and chromosome 2 of about 1 M bp) and the CTX phages can be integrated in either of the two chromosomes or both [8]. It was suggested that the CTX-2 ~ CTX-6 could have been generated in a strain that contained different CTX phages on each chromosome; for example, a strain contained the CTX-1 in one chromosome and the CTX-cla in the other. It was demonstrated that CTX-3 ~ CTX-6 could be generated in a strain, V212-1, which contained CTX-1 on chromosome 1 and CTX-2 on chromosome 2 [12].

From the analysis of the sequences of mosaic CTX-3 ~ CTX-6 phages produced in V212-1 by the inter-chromosomal recombination, the generation of pCTX phages with variant non-coding sequences was noticed [12]. These variant sequences contained a non-coding sequence of CTX-2 on chromosome 2, which could be distinguished from the non-coding region of CTX-1 on chromosome 1. Sequence analyses of the non-coding region of CTX prophages and pCTXs produced in various V. cholerae strains showed that CTX-cla and CTX-1 had a major different region in the middle of the non-coding sequence that was flanked by common regions. The phage type-specific regions in the non-coding sequences of pRS1 and pCTX-O139 have also been identified in this study.

Materials and Methods

Non-Coding Sequences of pCTXs and pRS1

The non-coding sequences of CTX-1 prophage in two Wave 1 El Tor strains, N16961 and V212-1, were obtained from the genome sequence data base under the accession number of AE003852/ AE003853 and ERS013132, respectively [17]. V. cholerae serogroup O1 biotype El Tor strain N16961 contains a CTX-1:RS1 array on chromosome 1. By comparison of the CTX phage sequence of N16961 with a strain that does not contain any element on chromosome 1, such as PM4 (GenBank Accession No. KF471410), and a pCTX-1 (GenBank Accession No. KF664569) produced from strain V212-1, the non-coding sequence of pCTX-1 could be deduced (Table 1, GenBank Accession No. KU877492, Fig. 2) [12]. Similarly, the non-coding sequence of the replicative form of the satellite phage RS1, pRS1, could also be deduced (GenBank

<table>
<thead>
<tr>
<th>Non-coding sequences</th>
<th>GenBank Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRS1 non-coding sequence</td>
<td>KU877491</td>
</tr>
<tr>
<td>pCTX-1 non-coding sequence</td>
<td>KU877492</td>
</tr>
<tr>
<td>pCTX-1-hybrid non-coding sequence</td>
<td>KU877493</td>
</tr>
<tr>
<td>pCTX-2 non-coding sequence</td>
<td>KU877494</td>
</tr>
<tr>
<td>pCTX-cla non-coding sequence of chromosome 2 of strain O395</td>
<td>KU877495</td>
</tr>
<tr>
<td>pCTX-O139 non-coding sequence</td>
<td>KU877496</td>
</tr>
<tr>
<td>pCTX-3 non-coding sequence</td>
<td>KU877497</td>
</tr>
</tbody>
</table>

Table 1. Non-coding sequences of the replicative form of CTX phages and RS1.
Non-Coding Sequences of V. cholerae CTX Phages 2016

Vol. 26 | No. 8

Accession No. KU877491).

Wave 2 strain B33 contains a tandem repeat of CTX-2 on chromosome 2 (GenBank Accession No. GQ485644) [13]. By comparing the diff sequence of chromosome 2 of a strain that does not contain any element on chromosome 2 (GenBank Accession No. FJ449754, strain IB4122) with the CTX-2 sequence of B33, the non-coding sequence of the replicative form of CTX-2, pCTX-2, could be deduced (GenBank Accession No. KU877494) [13]. The DNA sequences of the non-coding region of pCTX-1 and pCTX-2 are shown in Fig. 3.

A classical biotype strain, O395, contains a truncated classical CTX followed by a full-length classical CTX phage on chromosome 1 and a classical CTX phage on chromosome 2 [13]. The non-coding sequences of the CTX-cla prophage in two classical biotype strains, O395 and GP8, were obtained from the genome sequence information under the accession number CP000626/CP000627 and ERS013128, respectively [17]. Since the non-coding sequence of pCTX-cla from the CTX-cla’s on chromosome 1 could be a mix of two variant classical phages, the non-coding sequences of CTX-cla on chromosome 2 could be considered the authentic non-coding sequence of pCTX-cla (GenBank Accession No. KU877495).

pCTX-1-Hybrid Intergenic Sequence

The mosaic pCTX phage generated by the inter-strand recombination between CTX-1 on chromosome 1 and CTX-2 on chromosome 2 in strain V212-1 usually contained the non-coding sequence of CTX-1 on the chromosome. However, a variant of the mosaic CTX phage that contained the mixed non-coding sequences of CTX-1 and CTX-2 was identified and designated as pCTX-1-hybrid.

Results

Comparison of the Non-Coding Sequences of CTX-1, CTX-2, and CTX-cla

The non-coding sequences of prophages CTX-1 in strains N16961 and V212-1, CTX-2 in B33, and CTX-cla in O395 and GP8 were deduced from the genome sequence data, as shown in Fig. 2. The non-coding sequences of CTX-1 in strains N16961 and V212-1 were identical and the CTX-cla’s in strains O395 and GP8 contained the same non-coding sequences. These non-coding sequences were compared with the non-coding sequence of the replicative form pCTX-1 (GenBank Accession No. KF664579) [12].

The non-coding sequence of CTX-1 prophage in strain N16961 was identical to that of pCTX-1, which implies that the DNA sequence of the non-coding sequence is not altered by the integration or excision process [1]. The length of the non-coding sequence of CTX-1 was 601 bp (between the termination codons of rstrR and ctxB).

Two CTX-2 prophages in the tandem repeat array on chromosome 2 of strain B33 had identical sequence not only in the ORFs of the CTX phage genome but also in the non-coding sequences. The non-coding sequences of CTX-2 and CTX-cla were homologous and the length was 730 bp (Fig. 3). There are four SNPs between non-coding sequences of CTX-2 and CTX-cla at positions 131, 213, 330, and 432, and the 131st position belongs to the XerD binding site (Fig. 2). The non-coding sequences of the replicative form of pCTX-2 and pCTX-cla are not available since they have not been experimentally demonstrated yet; however, they could be assumed to be identical to the non-coding sequence of CTX-2 and CTX-cla prophages, respectively, as
Alignment of the non-coding sequences of CTX-1 and CTX-cla (CTX-2) showed that the non-coding sequence of CTX could be divided into three parts. The first 270bp (from the termination codon of ctxB) are homologous between CTX-1 and CTX-cla, while five SNPs were found (Fig. 3). This part can be designated as Common region 1 of the non-coding sequence. Next to Common region 1, a Variable region with a different length is located. The Variable regions are 136 bp and 267 bp long in CTX-1 and CTX-2, respectively. There is no sequence similarity between the two Variable regions, and therefore, they could not be aligned (Fig. 3). The Variable region is followed by another Common region (Common region 2) of 185 and 193bp long in CTX-1 and CTX-2, respectively. Next to Common region 2 is the termination codon of rstR (rstR El Tor in CTX-1 and rstR cla in CTX-cla). The difference in length of Common region 2 is due to a T addition between the 5th and 6th nucleotides in CTX-2, and the last nucleotide G of CTX-1 is replaced with TTGATTAC in CTX-2 (Fig. 3).

The non-coding sequence of RS1 element was 511bp long and contained a similar sequence as CTX-1. Although the Variable region and Common region 2 of the non-coding sequence of RS1 are identical to those of CTX-1, the first 90bp of Common region 1 are missing in the non-coding sequence of RS1, and nine SNPs in the next 15bp (Table 1).
Non-Coding Sequence of \( pCTX-O139 \)

The non-coding sequence of \( pCTX-O139 \) was deduced from the PCR product of the primer set ctxB/rstRcalR from two O139 serogroup strains [13]. This sequence includes part of the non-coding sequence (between \( \text{ctxB} \) and \( \text{attR} \)) of the preceding CTX prophage, not the CTX-O139 origin sequence. However, this sequence is identical among CTX-1, CTX-cla, and CTX-2, and it was assumed that CTX-O139 also contained the same sequence. The non-coding sequence of \( pCTX-O139 \) was similar to the non-coding sequence of \( pCTX-2 \) and \( pCTX-cla \), since it contained Common regions 1 and 2 and the Variable region of \( pCTX-2 \) (Fig. 3). However, the non-coding sequence of \( pCTX-O139 \) contained additional 194 nucleotides instead of the last 11 nucleotides of Common region 2 of \( pCTX-2 \). These additional nucleotides in the non-coding sequence of \( pCTX-O139 \) was also reported previously [4].

Non-Coding Sequence of \( pCTX-1\text{-Hybrid} \)

\( pCTX-3 \sim pCTX-6 \) that are found in Wave 3 strains have been shown to be generated by an inter-strand homologous recombination between two different CTX prophages as shown in an example of strain V212-1 [12]. The recombination could occur between CTX-1 (followed by an RS1) on chromosome 1 and the first CTX-2 of a tandem repeat array on chromosome 2 in V212-1. The recombination occurs anywhere upstream of \( \text{ctxB}s \) of two phages (position (1) in Fig. 4) and downstream of \( \text{ctxB}s \) to generate a new mosaic CTX phage genome that contains classical \( \text{ctxB} \) on chromosome 1. When the second recombination occurs downstream of \( \text{ctxB} \), the recombination can occur via Common region 1 or 2 (recombination position (2) or (3) in Fig. 4, respectively). When the recombination is mediated via Common region 1, the newly generated CTX phage should have the same non-coding sequence as CTX-1. (There are five SNPs between the Common regions 1 of CTX-1 and CTX-2; therefore, five minor variants could be generated via these five SNPs, Fig. 3.) For example, the \( pCTX-3 \) found in Wave 3 strain 01.07.VP (IB4122) contains the same non-coding sequence as CTX-1 in the Wave 1 strain N16961 (Table 1), which implies that this \( pCTX-3 \) was generated by recombination between the Common regions 1 of CTX-1 and CTX-2.

However, a hybrid non-coding sequence could be generated when the recombination occurred via Common region 2, and the CTX phage generated by this recombination should contain Common region 1, the Variable region of \( pCTX-2 \), and a hybrid of the Common regions 2 of CTX-1 and CTX-2. In this hybrid non-coding sequence, the length is 723 bp (7 bp shorter than CTX-2 since the 6th nucleotide “T” is from CTX-2 and the TTGATTAC sequence at the end of CTX-2 is replaced with G of CTX-1) (Fig. 3).

The recombination via Common region 1 is more likely to happen, since the recombination occurs in the homologous region of CTX phage (\( \text{ctxB} \) and Common region 1). The recombination position (3) in Fig. 4 is less likely because Common region 2 is only 193 bp long and flanked by two diverse sequences (Variable region in the non-coding sequence and \( \text{rstR} \)). However, this recombination does occur via Common region 1, and \( pCTX-1\text{-hybrid} \) was indeed

---

**Fig. 4.** Schematic diagram of generation of \( pCTX-1\text{-hybrid} \).

The first homologous recombination between CTX-1 on chromosome 1 and CTX-2 on chromosome 2 occurs in any part of CTX phages (recombination position (1)) and the second recombination can occur between Common region 1 or Common region 2. When the recombination occurs on Common region 1 (recombination position (2)), the CTX phage produced on chromosome 1 contains the same intergenic sequence as CTX-1. However, the CTX phage produced on chromosome 1 contains an intergenic sequence similar to CTX-2, when the recombination occurs on Common region 2 (\( pCTX-1\text{-hybrid} \), recombination position (3)).
generated. The pCTX-1-hybrid could be integrated on the genome of the transduced recipient strain [12].

Discussion

The molecular mechanism of the integration of pCTX phage genome via the dif site in V. cholerae chromosomes has been demonstrated [2, 16]. The excision of CTX phage has not been documented extensively, since the replication mechanism of CTX phage is unique among the lysogenic phages [6]; however, the excision of intact CTX phage genome has been recently demonstrated, which can leave the authentic dif site on the V. cholerae chromosome [9, 12]. We deduced non-coding sequences of pCTX-cla, pCTX-1, pCTX-O139, and pRS1.

The single-stranded CTX phage forms a stem-loop structure for integration in the V. cholerae chromosomes, and the stem-loop structure sequence is located entirely on Common region 1 of the non-coding sequence [2]. The XerC binding site is located at the positions 111–121 and 252–262 of Common region 1, while the XerD binding site is at 134–144 and 234–244. During the integration of pCTX, the non-coding sequence splits into two parts. The first half of Common region 1 (from the first nucleotide to the XerCD binding site) becomes the 3’ UTR and the rest of the non-coding sequence constitutes ig-1. There are three SNPs that are located in the bulge of the stem-loop structure of the non-coding sequence of CTX-1, CTX-2, and CTX-cla and a proposed stem-loop structure is shown in Fig. 5 [16].

The non-coding sequence of pCTX-cla has not been described because the replicative form of pCTX-cla has not been experimentally revealed [5]. However, the non-coding sequence of pCTX-cla could be deduced from the integrated CTX-cla by combining the ig-1 sequence and 3’ UTR sequence in a strain that harbors CTX-cla.

While we have shown the generation of pCTX-3 ~ pCTX-6 from V212-1 that contained CTX-1 on chromosome 1 and a tandem repeat of CTX-2 on chromosome 2, we have noticed that a derivative of each pCTX was also generated [12]. The derivative of each mosaic CTX contained the same CTX genome of the newly generated mosaic CTX phage, but contained a different non-coding sequence. We found that the non-coding sequences of pCTX-1 on chromosome 1 and pCTX-2 on chromosome 2 are different.

Previously, the non-coding sequence between the attR site and the termination of rstR has been named as ig-1 [4]. The molecular mechanism of the integration of pCTX on the chromosomes of V. cholerae via the XerCD recombination

![Fig. 5](image-url)
system on the dif sequence has been unveiled and we now understand the non-coding sequence between ctxB and rstR on pCTX splits into the ig-1 and the 3′ UTR sequences next to ctxB of the prophage.

The non-coding sequences of CTX-2 and CTX-cla are homologous, containing the same Variable region. Although most of the CTX genome of CTX-2 is identical to CTX-1, the ctxB, non-coding sequence, and rstR of CTX-2 are of CTX-2, which indicates that CTX-2 was perhaps generated by recombination between CTX-1 and CTX-cla.

There have been no experimental evidence of what the functions of the various non-coding sequences in different types of CTX phages are. The CTX phage replication initiation/termination point is shown to be located in Common region 2 [4]; however, the roles of Common region 1 and the Variable region remain to be investigated. The different non-coding sequences, especially the Variable region, might be important for the difference in the biological activities of the CTX phages; for example, the replication efficiency of the CTX phage and directionality of the integration into chromosome 1 or chromosome 2 of V. cholerae.

The replication of CTX-cla and CTX-2 has not been demonstrated, while the replication of CTX-1, CTX-O139, and RS1 element has been well-described [6, 9, 20]. The Variable regions in CTX-cla and CTX-O139 are homologous, but the non-coding sequence of CTX-O139 contains an additional 194 nucleotides. The replication efficiencies of CTX-cla and CTX-O139 might be altered by the Variable region and the additional sequence.

In addition to the efficiency of the replication of CTX phages, the chromosomal integration of CTX phage might be governed by the non-coding sequence. CTX-cla or CTX-2 have been found to be integrated on chromosomes 1 and 2, whereas CTX-1, CTX-O139, and RS1 have been identified only in chromosome 1 [13]. The exact role of the non-coding sequence in the replication and integration of CTX phage could be elucidated if the replication of CTX-cla and CTX-2 is established.

Acknowledgments

This work was supported by the research fund of Hanyang University (HY-2012-P).

References


