Chromosomal replication origin from the marine bacterium Vibrio harveyi functions in Escherichia coli: oriC consensus sequence

[initiation/high copy lethal (HCL) genes/gene–polypeptide relationships/in vitro initiation]

JUDITH W. ZYSKIND*,†, JOSEPH M. CLEARY*,‡, WILLIAM S. A. BRUSILOW§, NANCY E. HARDING*,†, AND DOUGLAS W. SMITH*

*Department of Biology, C-016, University of California at San Diego, La Jolla, California 92037, and †Department of Biological Sciences, Stanford University, Stanford, California 94305

Communicated by Bruno H. Zimm, October 25, 1982

ABSTRACT The chromosomal replication origin (oriC) of Vibrio harveyi has been isolated on a plasmid and shown to function as an origin in Escherichia coli. The nucleotide sequence of the V. harveyi oriC was determined. From a comparison of this sequence with oriC sequences of five enteric bacteria, we derived a consensus sequence of bacterial origins that function in E. coli. This consensus sequence identifies 192 positions within oriC where nucleotide substitutions can occur without loss of origin function. These positions are clustered rather than scattered. Four inter-related nine-base-pair repeats and eight of the dam methylation G-A-T-C sites are conserved in the consensus sequence. Very few relative insertion-deletion changes occur, and these are localized to one region of oriC. The genes for three polypeptides linked to the V. harveyi oriC were identified by using in vitro protein synthesis directed by deletion derivative plasmid templates. One of these genes, coding for a 58,000 M r protein and located 3.0 kilobase pairs from the V. harveyi oriC region, is lethal to E. coli when many copies (approximately 40 per cell) are present (high copy lethal or HCL gene). In addition, nucleotide sequence analysis showed that a different gene, the gid gene to the left of oriC, is highly conserved between E. coli and V. harveyi, whereas the coding region to the right of oriC is much less conserved.

By comparing the nucleotide sequences of different bacterial DNA replication origins (oriC) that function in Escherichia coli, we hope to determine which positions within oriC will tolerate changes and which require an invariant nucleotide to maintain origin function. This approach is similar to the one used for identifying the conserved −10 and −35 regions of promoters (1). Such a comparison will also narrow the number of possible intraspecies secondary structures the origin might assume during the initiation process. The pattern of homology observed when the origin sequences from E. coli (2, 3), Salmonella typhimurium (4), Enterobacter aerogenes (5), Klebsiella pneumoniae (5), and Erwinia carotovora (6), all members of the family Enterobacteriaceae, were compared (7, 8) is a clustering of nucleotide substitution-type differences interspersed with conserved regions up to 19 base pairs (bp) long. We show here that the replication origin from the marine bacterium Vibrio harveyi (9), thought to be distantly related to the Enterobacteriaceae (10, 11), functions in E. coli. Study of its sequence should contribute significantly to determining which nucleotides need not be strictly conserved for oriC function in E. coli.

To isolate origins from V. harveyi, we used a cloning vector that requires DNA polymerase I, the product of the polA gene, for replication. V. harveyi restriction fragments that allow this vehicle to replicate in an E. coli polA mutant were selected.

None of the three different classes of V. harveyi DNA sequences that were obtained in this manner hybridize to the E. coli oriC region, nor do they complement E. coli mutations in genes known to be linked to oriC. One class of plasmids most likely contains the true V. harveyi chromosomal replication origin for the following reasons: (i) They alone hybridized to DNA from marine organisms closely related to V. harveyi. The chromosomal replication origin region is expected to be highly conserved in related marine bacteria just as it is among members of the Enterobacteriaceae. (ii) Only plasmids in this class replicated in the in vitro oriC-specific initiation system, in a manner dependent on the presence of dnaA protein (ref. 12; R. Fuller, J. Kaguni, and A. Kornberg, personal communication). (iii) Subsequent sequence determination showed that this V. harveyi origin is homologous to the E. coli chromosomal DNA replication origin. This sequence analysis locates the V. harveyi oriC between protein coding regions similar to those flanking the E. coli origin. This class of plasmids contains a high copy lethal (HCL) gene, as do certain other oriC-containing plasmids (13). The V. harveyi HCL gene, however, appears to be unrelated to the HCL gene(s) linked to oriC in species of the Enterobacteriaceae family (13).

MATERIALS AND METHODS

Bacterial Strains and General Procedures. The E. coli strains used for cloning (13); growth conditions for V. harveyi B392 (14), other marine organisms (15), and E. coli (16); selection of antibiotic resistance (16); genetic procedures (16); and DNA isolation techniques (6) have been described.

Restriction Endonuclease Digestion, Plasmid Construction, and Gel Electrophoresis Analysis. Reaction conditions for restriction endonucleases and phage T4 DNA ligase (Bethesda Research Laboratories) were previously described (16). Molecular weights of plasmids and restriction fragments were determined by using either agarose or polyacylamide gel electrophoresis (16).

Hybridization Studies. DNA was transferred from agarose gels to nitrocellulose filters (Schleicher & Schuell) as described (17). Hybridizations were performed at 50°C in 2× Denhardt's solution (18) either alone for 24 hr (Fig. 1) or in the presence of 10% dextran sulfate (19) for 4 hr (Fig. 2). All probes were radiolabeled in vitro by nick-translation (20) using [32P]dCTP.

In Vitro Protein Synthesis. An in vitro coupled transcription–translation system previously described (21) was used to assay DNA-dependent protein synthesis.

Abbreviations: bp, base pair(s); HCL, high copy lethal.
* Present address: Biology Dept., San Diego State Univ., San Diego, CA 92182.
† Present address: Kelco, Div. of Merck and Co., Inc., 8355 Aero Dr., San Diego, CA 92123.
DNA Sequence Determination. Nucleotide sequences were determined by using the Maxam–Gilbert chemical approach (22). Sequences were analyzed in part by the SEQ computer program procedures made available by the Stanford Molgen Project and the National Institutes of Health SUMEX-AIM Facility.

RESULTS
Isolation and Properties of the V. harveyi Origin of Replication. The previously described procedure (6, 13) using the ColE1-like cloning vehicle pMK2004 (23) was used for cloning V. harveyi DNA sequences that act as origins. Restriction endonuclease analysis of the isolated plasmids divided these DNA sequences into three classes, represented by pNH23, pNH24, and pNH26 (Fig. 1A). The V. harveyi Sal I fragments were re-cloned, using the nonreplicating Sal I restriction fragment of 5.6 kilobase pairs (kbp) containing the kan gene from plasmid pDF11 (23), thereby confirming that these fragments carry functional origins. Plasmids in all three classes were unstable, in approximately five generations, 50% of the cells had lost the plasmid in the absence of selection.

When used as probes, none of the plasmids hybridized to the E. coli oriC region (Fig. 1, lanes c). Also, none of the plasmids complemented E. coli asA or uncB mutations, genes known to bracket other bacterial origins (4, 6, 13, 25). However, the inability of the class of plasmids carrying the 6.0-kbp insert, represented by pNH23 (Fig. 1), to transform E. coli polA+ strains is shared by certain oriC-containing plasmids from E. coli, Enterobacter aerogenes, and K. pneumoniae (13). In particular, the unc promoter plus the first few genes of the unc operon, located near oriC, confers lethality on many of which copies of the region are present. The plasmid pNH23 is maintained at an apparent low copy number (1–2 per cell, data not shown) in polA− strains. The replication origin of pMK2004 would be used in polA+ strains, resulting in a copy number of approximately 40 (23); thus the 6.0-kbp Sal I insert from V. harveyi presumably contains a gene (HCL) that is lethal when many copies per cell are present.

One of the original plasmids isolated, pNH25, was shown to have an insertion in the HCL gene of pNH23. This insertion introduces a new Sal I restriction site and inactivates the conditional lethality of pNH23. This interesting spontaneous insertion derivative illustrates the need to use a low copy number plasmid as a cloning vehicle for cloning HCL genes, or any nearby genetic element. When a high copy number cloning vehicle is used, the only plasmids recovered would be those in which the lethal region has in some way been inactivated, such as in pNH25.

DNA from marine bacteria closely related to V. harveyi (10, 26) is expected to cross-hybridize to the V. harveyi oriC region. Plasmid pNH23 hybridized to DNA from V. harveyi, two Photobacterium species, and V. fischeri; however, pNH24 and pNH26 hybridized only to DNA from V. harveyi (Fig. 2). Thus, only the pNH23 class of plasmids is homologous to DNA in species related to V. harveyi.

The plasmids pNH24 and pNH26 appear to be derived from cryptic prophage in the V. harveyi chromosome rather than from resident plasmids. No plasmids have been detected in V. harveyi B392 except for a 1.6-kbp plasmid-like band observed in an Eckhardt gel (27); no cross-hybridization of this band with pNH23, pNH24, or pNH26 was detected.

Although not hybridizing to E. coli chromosomal DNA (Fig. 2), these cloned V. harveyi origins function in E. coli cells and, therefore, use its DNA replication components. Of the three plasmid classes, only plasmids containing the V. harveyi origin found in pNH23 replicated in a manner dependent on the presence of the dnaA protein in the soluble enzyme system that is specific for oriC-containing plasmids (refs. 12 and 28; R. Fuller, J. Kaguni, and A. Kornberg, personal communication). From this evidence and the sequence analysis presented below, we conclude that this replication origin is the V. harveyi chromosome origin.

![Fig. 1. Cross-hybridization of V. harveyi origin plasmids. Plasmid DNA was digested with either Sal I for pNH23 (lanes a), pNH24 (lanes b), pNH25 (lanes c), and pNH26 (lanes d), or EcoRI for pAL3 (ref. 24; lanes e). Restriction fragments were separated by electrophoresis on a 1.0% agarose gel (A, ethidium bromide staining) in quadruplicate and transferred to nitrocellulose (B, C, D, and E). Probes used were pNH23 (B), pNH24 (C), pNH25 (D), and pNH26 (E). The sizes given on the left are in kbp. The cloning vehicle pMK2004 is 5.2 kbp. Sizes in kbp of Sal I restriction fragments in A are 6.0 and 5.2 (lane a); 5.2 and 0.7 (lane b); 5.2, 3.85, and 2.75 (lane c); and 18.0 and 5.2 (lane d). The plasmid pAL3 (24) consists of two EcoRI fragments (lane e); the larger one of 4.0 kbp consists of pACYC184, and the smaller one of 3.6 kbp is derived from the E. coli chromosome and includes oriC.](image1)

![Fig. 2. Hybridization of pNH23 to Sal I digests of bacterial DNA. Chromosomal DNA (2.5 µg per lane) from E. coli C600 (lanes a), V. fischeri MJ1 (lanes b), P. leiognathi PL-721 (lanes c), P. phosphoreum NZ11 (lanes d), and V. harveyi B392 (lanes e) was digested with Sal I followed by electrophoresis in a 0.8% agarose gel (A, ethidium bromide staining) in quadruplicate. Probe used was pNH23 (B). Data are not shown for probes pNH24 and pNH26 but are summarized in the text. When pNH25 was used as a probe, results were the same as for pNH23 (data not shown). The sizes given on the left are in kbp.](image2)
Identification of the HCL Gene of pNH23. Plasmids missing the HCL gene were constructed by deleting HindIII fragments from pNH23 (Fig. 3A). In vitro transcription-translation experiments (21) with these plasmids tested whether or not a polypeptide is encoded in the region of the deleted DNA (Fig. 4). Three polypeptides (molecular weights 58,000, 50,000, and 10,000) were observed with plasmid pNH23 as template (Fig. 4, lane c) that are not encoded by the cloning vehicle pMK2004 (Fig. 4, lane b).

The 58,000 M, polypeptide is the only polypeptide absent when pNH23-derived plasmids that transform E. coli polA+ strains are used as templates (Fig. 4, lanes d, e, f, and g). Overproduction of this polypeptide apparently causes the lethality at high copy number observed with pNH23. A gene found in this region in E. coli is trkD (25); however, an E. coli trkD mutant is not complemented by this region of V. harveyi DNA. Because the molecular weight of this polypeptide is different from the molecular weights of any of the eight polypeptides encoded by the unc operon (Fig. 4, lane a) or any of the polypeptides encoded by the E. coli oriC-containing plasmid pAL1 (ref. 24; Fig. 4, lane b), the HCL gene on pNH23 is apparently different from the HCL gene(s) carried by pAL1 (13).

Analysis of these deletion plasmids shows that the gene encoding the 50,000 M, polypeptide must be at the left end of the SalI fragment in pNH23 (Fig. 3A). The coding region replaces the comparable colinear position in E. coli of the asNA gene, which encodes asparagine synthetase of 39,000 M, (25). E. coli has two asparagine synthetase genes, asNA and asNB (25). The asNA gene may have evolved into a new gene in V. harveyi, with V. harveyi still preserving a functional asNB gene.

Sequence Analysis of Regions Adjacent to oriC of V. harveyi. Between oriC and asNA in E. coli lies a region coding for a 15,500 M, polypeptide (2, 29, 30). The sequence of over 70% of the coding region overlap in V. harveyi has been determined and compared to the E. coli sequence (2, 30): 47 of 104, or 45%, of the amino acids are conserved, while 156 of 312 (50%) of the nucleotides are conserved. The V. harveyi region contains an

![Fig. 3. Genetic and physical map of the V. harveyi oriC region, with strategy of sequence determination. (A) Physical map of pNH23 and derivative plasmids containing the origin of V. harveyi. The plasmids pJC235, pJC236, and pJC238 are HindIII deletion derivative plasmids of pNH23 and pJZ58 is a Pat I deletion plasmid of pNH23. Open regions designate the V. harveyi chromosomal DNA and regions with a horizontal line indicate the cloning vehicle pMK2004. The ColEl-like origin of pMK2004 is indicated by rep. Restriction sites are designated by H, HindIII; P, Pat I; S, Sal I; C, Cia I; Bg, Bgl II. Roman numerals are used to designate HindIII fragments. Coding regions are given above the map; four encoded proteins are identified by M, × 10^3. The protein coding region indicated by an asterisk appears to be responsible for the conditional lethality of pNH23. Quotation marks enclose E. coli coding regions showing sequence homology with V. harveyi. (B) Strategy for Maxam–Gilbert sequence determination of the oriC region of V. harveyi. Tails of arrows, 32P-labeled end of each fragment; heads of arrows, extent to which sequence was determined from that fragment.](image-url)
inserted extension relative to *E. coli*, lengthening the polypeptide by 38 amino acids with a stop codon at positions 260–262 within the minimal origin (Fig. 5).

The gene to the unc side of oriC in *E. coli*, gid, codes for a polypeptide of 70,000 M₉, (29). Sequence comparison shows that this gene is more highly conserved than the gene coding for the 15,500 M₉ polypeptide. Of the 17% of the gid gene contained on the 6.0-kbp Sal I fragment was determined from deletion plasmids of pNH23 that retained oriC function, the smallest being pJCC228 (Fig. 3A). When the 178-bp Bgl II fragment was deleted from pJZ58 (Fig. 3B), the resulting plasmid, pJZ61, no longer transformed *E. coli* polA strains, indicating loss of oriC function.

The strategy for determining the nucleotide sequence of the *V. harveyi* origin is shown in Fig. 3B. The nucleotide sequence of the *V. harveyi* origin region is compared to the nucleotide sequences of the origin regions from five enteric bacteria in Fig. 5, which shows a consensus sequence derived from these six *oriC* sequences. The six sequences were aligned to minimize the number of nucleotide changes, insertions, or deletions. All six origins, including the *Escherichia carotovora* oriC (R. Fuller, J. Kaguni, and A. Kornberg, personal communication), initiate replication in *E. coli* cells and are dnaA-dependent templates (ref. 13 and this paper) in the oriC-specific in vitro initiation system (12, 28).

**DISCUSSION**

Sequence analysis and in vitro replication studies demonstrate that we have isolated the *V. harveyi* chromosomal origin of replication equivalent to oriC of *E. coli*. The inclusion of this origin in the oriC consensus sequence introduces 25 changes within the minimal origin (34) of 245 bp (Fig. 5). Although the extent of homology of the *V. harveyi* origin is less than that of the other bacterial origins to the *E. coli* origin, the nucleotide differences are grouped in clusters similar to those found previously (4–7).

One group of conserved clusters identifies four interrelated 9-bp repeats (7), underlined with arrows in Fig. 5. The inverted repeats R1 and R4 (Fig. 5) may be essential in the formation of intranrarr secondary structure within the origin. The helix formed between these two repeats has the lowest free energy of any dyad repeat within five of the minimal *oriC* sequences and is the second lowest in that of *V. harveyi*.

Oka et al. (34), by relating a series of plasmid deletions to *oriC* function, delimited the size of the functional *oriC* of *E. coli* to between 237 and 245 bp (Fig. 5, box). Because the two As to the left of the box and the A to the right of the box are conserved in all six *oriC* sequences, they may be important nucleotides with respect to replication initiation, especially when *oriC* is functioning in the chromosome rather than on a plasmid.

To best align the *oriC* sequences from the six different species, a few relative deletions are introduced; all of these are within the first 35 nucleotides of the minimal sequence. With the *V. harveyi* oriC sequence, we include an insertion of two nucleo-

---

**Fig. 5.** Consensus sequence of the minimal origin of the bacterial chromosome. The consensus sequence is derived from six bacterial origin sequences, those of *E. coli* (2, 3), *S. typhimurium* (4), *Enterobacter aerogenes* (5), *K. pneumoniae* (5), *E. carotovora* (6), and *V. harveyi*. The alignment of the six sequences is such that the least number of changes are introduced into the consensus sequence. The two bases inserted (at positions 239 and 240) and deleted (at positions 246 and 247) are counted as two rather than four changes for calculations in the text. In the consensus sequence, a large capital letter means that the same nucleotide is found in all six origins; a small capital letter means the nucleotide is present in five of the six sequences; a lowercase letter is used when that nucleotide is present in three or four of six bacterial origins but only two different nucleotides are found at that site; and where three or four of the four possible nucleotides, or two different nucleotides plus a deletion, are found at a site, the letter n is used. Bold capital letters locate positions 149, 242, and 267, where single-base substitutions produce an *oriC* phenotype in *E. coli* (31). The arrow at position 166 represents the first ribonucleotide of a RNA transcribed by RNA polymerase in vitro (32). The arrows ending between positions 71 and 72 and positions 107 and 108 mark the major RNA–DNA transition sites (33). The directions of these three arrows show the direction of transcription. In the individual origin sequences, - means a deletion relative to the consensus sequence is present and , indicates that the nucleotide in the bacterial genome sequence is the same as the nucleotide in the consensus sequence. G-A-T-C sites are underlined in the consensus sequence and certain *E. coli* restriction sites are noted. The minimal origin of *E. coli* (34) is enclosed within the box. The numbering of the nucleotide positions is that used for *E. coli* (2, 3), and the upper left end is the 5' end. The four related 9-bp repeats, R1, R2, R3, and R4, are indicated by the arrows, with the 5' to 3' sequence given below the arrow.
tides followed by a deletion of two nucleotides between positions 139 and 147 so that the fewest changes are introduced into the consensus sequences. Nevertheless, the number of nucleotides between the positions 54 and 268 is exactly the same in all six sequences. This invariance in size suggests that a functional bacterial origin can tolerate no insertions or deletions within the entire region between nucleotides 54 and 268, independent of the local degree of nucleotide conservation.

Of the 245 nucleotides within the consensus sequence of the minimal origin, 122 are different in at least one species. In the clustering of these differences, the longest conserved length is 12 bp and the longest nonconserved length is 16 bp. Although insertions or deletions in the BamHI (at nucleotide 95), Aca II, or HindIII sites are oriC- (34), nucleotides differ at every position in these sites. This result suggests that the role of the nonconserved clusters in oriC is to provide a specific distance, perhaps between binding sites of initiation proteins; their role may be similar to that of the nonconserved spacer region (35) in promoters, which is found between the conserved -10 and -35 regions.

The nucleotides at three positions in oriC have been found to be essential (31) and are located at positions 149, 242, and 267 (Fig. 5). Single-base substitutions at these sites produce an oriC- phenotype. Base substitutions in another oriC- mutant suggest that position 134 is also important (31).

RNA polymerase is required for initiation of DNA replication at oriC (12, 36), but the role of the synthesized RNA is unknown. Two studies have located transcription products within oriC. A RNA that is transcribed in vitro from within the minimal origin by RNA polymerase has been isolated (32). The first nucleotide of this RNA is at position 167, with the direction of transcription from right to left (Fig. 5). In the second study, two transition sites from RNA to DNA within the minimal origin were detected in vivo experiments (33). These transition sites at positions 72 and 109 (Fig. 5) have the same polarity as the RNA mentioned above (32) and, therefore, may be termination sites for this RNA. This RNA may function as a primer for DNA synthesis in one direction.

Of the 14 G-A-T-C sites found within the V. harveyi minimal origin, 8 appear in the consensus sequence. The potential importance of the high frequency of the G-A-T-C sequence within oriC, a sequence specifically recognized and methylated by the E. coli dam methylase, is discussed elsewhere (4). The adenine in G-A-T-C sites is methylated in V. harveyi DNA, although no hybridization of V. harveyi DNA with a dam-containing probe from E. coli has been detected (J. E. Brooks, personal communication).

Of the two regions directly adjacent to the origin, gid is the more highly conserved; gid is also more highly conserved than oriC. The as-yet-unknown function of this gene must, therefore, be important for the survival of the bacterial species studied here, even though a mutant with a deletion in this gene is viable (29).

We thank Ken Nealon and Charles Bieger for the strains of marine bacteria, Adela Baer for her critical reading of the final manuscript, and Hans-Jorg Buhk for giving us a copy of his dissertation, which has the complete sequence of pCM959. This work was supported by Postdoctoral Traineeship GM07199 from the National Institutes of Health to J.M.C. and Public Health Service Grant GM21978 to J.W.Z.