Use of the Isocitrate Dehydrogenase Structural Gene for Attachment of e14 in *Escherichia coli* K-12

CHARLES W. HILL,^{1*} JANE ANTAL GRAY,¹ AND HOWARD BRODY²

Department of Biological Chemistry, The Milton S. Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania 17033,¹ and Department of Biology, University of California, Santa Barbara, California 93106²

Received 13 March 1989/Accepted 22 April 1989

The e14 element appears to be integrated into the *Escherichia coli* K-12 isocitrate dehydrogenase structural gene (*icd*). In being integrated, it replaced the last 52 codons of the gene with a closely related sequence. The two versions of the *icd* gene produce proteins of the same length but differ by 12 base substitutions that would cause two conservative amino acid replacements.

The genetic element e14 is a component of the *Escherichia coli* K-12 chromosome. Although it was present in the original isolate of strain K-12, e14 can be excised from the chromosome and is missing in a number of K-12 derivatives (4–6, 11). Excision is induced by UV irradiation, and both excision and reintegration occur by a site-specific mechanism (2, 4). A novel feature of the e14 attachment system is that it involves a large region of homology between the element and its host (3). This 216-base-pair (bp) homology is interrupted by 28 mismatches. The actual crossover for integration occurs within the first 11 bp at one end of the homology. A consequence of this skewing of the crossover position within the homology is that the *attL* homology at one end of integrated e14 is identical to the *attE* homology of free e14, while the *attR* homology is identical to the *attC* homology of the cured chromosome (Fig. 1).

The attachment site of e14 in *E. coli* K-12 is between the *purB* and *umuC* loci (2; Fig. 1). The isocitrate dehydrogenase locus, *icd*, maps in the same interval (1), but the relative positions of *icd* and e14 attachment have not been determined previously. A comparison of sequencing data for *icd* and e14 reveals a complex relation. First, we observed that the sequence of the last 320 bp of the *E. coli* K-12 *icd* coding region (10) is identical to that determined for a segment derived from the *attL* region of integrated e14 (3). More specifically, the last 165-bp portion of *icd* actually overlaps the first 165 bp of the *attL* homology (Fig. 2). Since the crossover between e14 and the host occurs within the first 11 bp of this homology (3), the last 52 codons of the *icd* protein coding sequence are provided by the integrated e14. Excision of e14 removes these codons and replaces them with a sequence from *attR*. Because of the homology of *attL* and *attR*, this event has minimal effect on the predicted protein. Both homologies provide a TAA stop codon at the same place, so the two versions of the protein are the same length. Furthermore, while the 165-bp segments are mismatched at 12 positions, only two of these substitutions lead to amino acid substitutions. In both of these cases, there is a conservative replacement of an aspartate by a glutamate upon e14 excision (Fig. 2). In fact, the isocitrate dehydrogenase specific activity in a strain carrying e14 did not differ significantly from an isogenic strain cured of its copy of e14. The derivation of the isogenic *E. coli* K-12 strains CH734 and CH1332 has been described previously (2). Isocitrate dehydrogenase assays were performed on sonicates of cells

grown in L broth containing 0.2% glucose and harvested during exponential growth. Specific activity was determined as described by Thorsness and Koshland (10). The specific activity of strain CH734 (with e14) was 0.14 U/mg of protein, whereas that of strain CH1332 (without e14) was 0.12 U/mg of protein.

The remaining 51-bp portion of the 216-bp homology contains 16 mismatches (Fig. 2). A GC-rich dyad symmetry, followed by a string of T bases, occurs in both versions near the end of the homology. These are potential transcription termination sites (9). It is interesting that while the perfect 11-bp dyad symmetry present in *attL* differs at five positions from the one in *attC*, four of these substitutions are involved in compensating changes that maintain the overall symmetry.

The e14 element is not present in either *E. coli* B/5 or *E. coli* C, two other natural isolates of *E. coli*. However, strains B/5 and C both show homology to a probe containing *attE* (2, 4). Furthermore, this homologous sequence can function as a site for e14 integration in *E. coli* C. It seems likely that the primordial state of *E. coli* is to not have e14 and, therefore, the *icd* type found in the cured derivative of *E. coli* K-12 is likely to be the native version. One can only speculate as to the origin of the *icd* carboxy-terminal sequence present in free e14. A possible advantage for selecting an internal

FIG. 1. Region of the *E. coli* K-12 chromosome containing *icd* and e14 (1, 2). The open and filled boxes indicate the 216-bp *att* homologies. The arrows indicate the positions involved in the crossover leading to excision. After e14 is excised, the *attC* homology is identical to that of *attR* but differs from those of *attL* and *attE* at 28 of 216 positions (3).

* Corresponding author.

FIG. 2. Comparison of the terminal sequences of *icd* with and without e14 integrated. The upper nucleotide sequence contains the *attL* homology of integrated e14 as well as flanking sequences (3) (GenBank accession M19683 and M19693). It is identical in sequence to the carboxy-terminal portion of the *icd* locus as determined by Thorsness and Koshland (10) (GenBank accession J02799). The last 63 amino acids of isocitrate dehydrogenase (10) are translated above the sequence. The lower nucleotide sequence contains the *attC* homology produced by e14 excision; -, the base at this position is identical to the one shown above it. The two glutamate-for-aspartate substitutions caused by e14 excision are shown below the *attC* sequence. The 216-bp homologies of *attL* and *attC* begin at coordinate 1 and are shown in capital letters, while the flanking regions are shown in lowercase letters. The 11-bp crossover region for e14 excision-integration is indicated along with a dyad symmetry that is a possible transcription termination signal (9).

The fact that the 55 codons contained in the homology segment have sustained 12 essentially neutral mutations suggests that the *icd* segment carried by e14 originated in a rather distant relative of *E. coli*. This degree of divergence falls between that observed in a comparison of homologous genes from *E. coli* and *Salmonella typhimurium* and that usually observed in comparison of different strains of *E. coli* (7). Conversely, the acquisition of e14 by *E. coli* K-12 must be recent since no mutations damaging the coding potential (i.e., non-neutral mutations) of the presumably superfluous *attR* have accumulated. Despite the implications of the above discussion concerning the mobility of e14, it should be emphasized that the manner in which e14 replicates independently of the host chromosome is unknown.

This work was supported by Public Health Service grant 16329 from the National Institutes of Health. Computer resources were provided by the BIONET National Computer Resource for Molecular Biology, which is funded by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health, grant no. P41 RR01685.

Stephanie Schneider and Allan Campbell (personal communication) have found that phage 21 integrates within *icd* at the same site as *e14*.

1. Apostolakis, D. P., A. Menter, B. J. Ramsph, H. C. Reeves, and E. A. Birge. 1982. Genetic map position of the cistron coding for isocitrate dehydrogenase in *Escherichia coli* K-12. *Curr. Microbiol.* **7**:45-47.
2. Brody, H., A. Greener, and C. W. Hill. 1985. Excision and reintegration of the *Escherichia coli* K-12 chromosomal element e14. *J. Bacteriol.* **161**:1112-1117.
3. Brody, H., and C. W. Hill. 1988. Attachment site of the genetic element e14. *J. Bacteriol.* **170**:2040-2044.
4. Greener, A., and C. W. Hill. 1980. Identification of a novel genetic element in *Escherichia coli* K-12. *J. Bacteriol.* **144**:312-321.
5. Kao, C., and L. Snyder. 1988. The *lit* gene product which blocks bacteriophage T4 late gene expression is a membrane protein encoded by a cryptic DNA element, e14. *J. Bacteriol.* **170**:2056-2062.
6. Maguin, E., H. Brody, C. W. Hill, and R. D'Ari. 1986. SOS-associated division inhibition gene *sfic* is part of excisable element e14 in *Escherichia coli*. *J. Bacteriol.* **168**:464-466.
7. Milkman, R., and I. P. Crawford. 1983. Clustered third-base substitutions among wild strains of *Escherichia coli*. *Science* **221**:378-380.
8. Pierson, L. S., III, and M. L. Kahn. 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*: DNA sequences of the phage and host regions involved in site-specific recombination. *J. Mol. Biol.* **196**:487-496.
9. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**:319-353.
10. Thorsness, P. E., and D. E. Koshland, Jr. 1987. Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. *J. Biol. Chem.* **262**:10422-10425.
11. van de Putte, P., R. Plasterk, and A. Kuijpers. 1984. A *Mu* *gin* complementing function and an invertible DNA region in *Escherichia coli* K-12 are situated on the genetic element e14. *J. Bacteriol.* **158**:517-522.