JOURNAL OF BACTERIOLOGY, Feb. 1989, p. 1214–1218
0021-9193/89/021214-05$02.00/0

Vol. 171, No. 2

Construction of an Ordered Cosmid Collection of the *Escherichia coli* K-12 W3110 Chromosome

SATOSHI TABATA,[1]* ATSUSHI HIGASHITANI,[1] MITURU TAKANAMI,[2] KIYOTAKA AKIYAMA,[3] YUJI KOHARA,[3] YUKINOBU NISHIMURA,[4] AKIKO NISHIMURA,[4] SEIICHI YASUDA,[4] AND YUKINORI HIROTA[4]

Department of Biology[1] and Department of Molecular Biology,[3] Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Institute for Chemical Research, Kyoto University, Uji, Kyoto-Fu 611,[2] and National Institute of Genetics, Mishima 411,[4] Japan

Received 6 June 1988/Accepted 24 October 1988

A cosmid library of the *Escherichia coli* K-12 W3110 chromosome was constructed in which clones were assigned to locations on the chromosome map by hybridization and genetic marker complementation tests. Approximately 70% of the genome was represented by this library. The identified clones can be maintained in the homologous system and would facilitate genetic studies of *E. coli*.

Escherichia coli is one of the best-investigated microorganism in terms of gene structure and function, and more than 1,000 genes have been identified and localized on the circular chromosome map (2). Nevertheless, the number of genes characterized corresponds to only about one-third of the entire number estimated to be present on the *E. coli* chromosome. To understand the whole genetic system of *E. coli*, further isolation of genes with known genetic markers as well as a systematic search for new genes would be necessary. The first gene library of the *E. coli* chromosome was constructed by Clarke and Carbon in 1975 (10). Since then, the library has been screened to isolate clones which complement a variety of mutations at known locations on the genetic map. So far, about 300 clones have been mapped and analyzed by using this library (2, 36). In the work described in this paper, we attempted to construct an ordered cosmid library of the *E. coli* chromosome in which clones are assigned to their specific locations. The cloning vectors cosmid pHC79 and its derivative pHC79EBE were adopted for the following reasons: (i) DNA fragments as large as 40 kilobases can be inserted; (ii) the copy number is about five when maintained as a hybrid plasmid (24), so that adverse effects of increasing gene dosage would be relatively small; and (iii) DNA can be introduced and maintained in a variety of mutant cells for marker complementation analysis. pHC79EBE has been constructed by replacement of the *Eco*RI-*Bam*HI region of pHC79 with a synthetic linker

5'AATTCGGATCCGAATTC
GCCTAGGCTTAAGCTAG5'

that creates a *Bam*HI site between two *Eco*RI sites.

Two sets of genomic libraries were constructed by the following procedures. High-molecular-weight DNA was extracted from *E. coli* W3110 (3) and partially digested with either *Eco*RI or *Sau*3AI to various extents. A pool of the *Eco*RI partial digests and a separate pool of the *Sau*3AI digests were fractionated by sucrose density gradient centrifugation, and DNA fractions of 35 to 45 kilobases were pooled. *Eco*RI fragments were ligated to the *Eco*RI-digested pHC79 DNA, and the *Sau*3AI fragments were ligated to the *Bam*HI-digested pHC79EBE. Ligated DNAs were packaged into bacteriophage lambda capsids in vitro (23) and used to transfect an *E. coli recA* strain, DH1 (29). Transformants which appeared on ampicillin-containing agar plates were subjected to screening.

For screening of clones, the dot hybridization procedure involving isolated plasmid DNA was adopted to reduce the background level in hybridization. Although the colony hybridization method is generally used for this purpose, we noted that nonspecific binding of probes to colonies often made it difficult to distinguish the positive signals from the negatives. Assuming that the genome size of *E. coli* W3110 is 4,500 kilobases and that fragments of an average of 40 kilobases are cloned, a total of 458 independent clones are necessary to cover 99% of the *E. coli* genome with minimal overlaps (11). On the basis of this estimate, 800 random clones obtained from *Eco*RI digests (*Eco*RI clones) and 500 clones obtained from *Sau*3AI digests (*Sau*3AI clones) were suspended and grown in 20 ml of medium (L broth), from which plasmid DNA was extracted by the alkaline extraction method (5). Portions (about 300 ng) were blotted on nylon membranes and hybridized to various probes by standard procedures (31).

To locate the clones on the circular chromosome map, chromosome walking was initiated from several known points. Inserts were extracted from the pLC plasmids indicated in Fig. 1, which had been screened from the Clarke-Carbon collection (10). The DNA fragments were labeled with ^{32}P by nick translation and used as probes. After hybridization with random gene libraries on nylon membranes, each membrane was washed under stringent conditions and exposed to X-ray films. Positive clones thus identified were used as probes for the next walks. Linkage between overlapping clones was confirmed by the method of Coulson et al. (13). Plasmid DNAs to be compared were digested with *Eco*RI, end labeled with ^{32}P, and redigested with the second enzyme, which recognizes a shorter sequence, e.g., *Hae*III or *Hpa*II. The resulting fragments were resolved by polyacrylamide gel electrophoresis and detected by autoradiography. The common terminal fragment(s) should be observed on autoradiograms if given clones share the common *Eco*RI recognition site(s). As a result of intensive screening, we constructed 31 independent "islands," in which constituent clones partially overlapped each other.

Correlation of these islands with the *E. coli* chromosome map was carried out on the basis of the following three criteria: (i) hybridization with *Not*I-generated fragments of the whole chromosome, (ii) hybridization to the ordered

* Corresponding author.

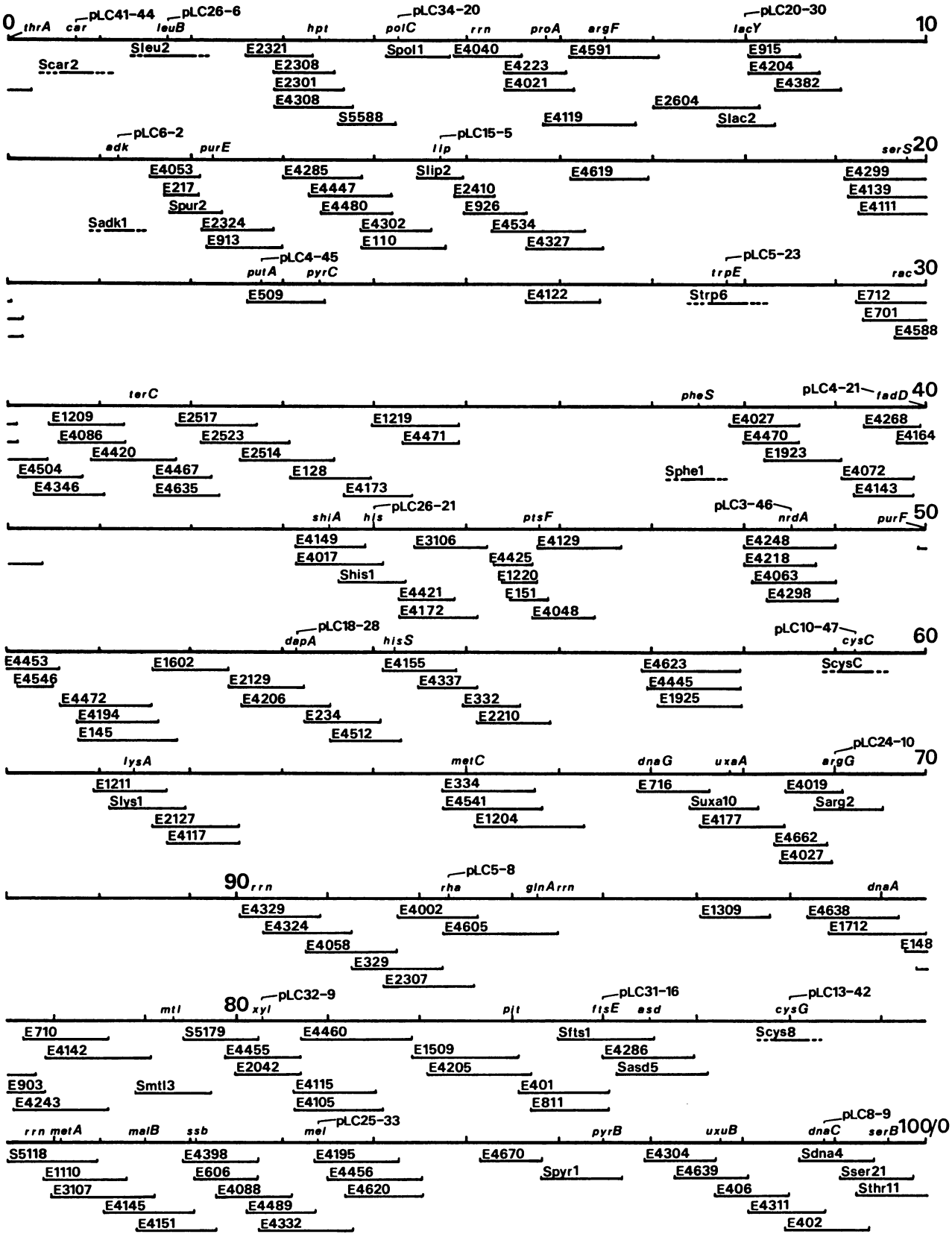

FIG. 1. Location of the cosmid clones on the genetic map of *E. coli*. Mutations which were complemented by the cloned DNA and the pLC plasmids which were used as the starting materials for chromosome walking are shown on the map from reference 2. The DNA regions which hybridized to rRNA, lambda- phage DNA (15) and cloned *terC* segment (1) are also shown and referred to as *rrn*, *rac*, and *terC*, respectively. E and S before the clone number stand for the *Eco*RI- and *Sau*3AI clone, respectively.

gene library which had been constructed by using the lambda phage vector (ordered lambda gene library) (28), and (iii) complementation tests with known genetic markers. First, the *Not*I digests of the W3110 DNA were resolved by pulse-field gel electrophoresis (43). DNA was extracted from each band, and hybridization with the cosmid libraries described above was performed to confirm the linkage between islands. Second, fragments from representative clones in each island were labeled with ^{32}P, and correlation to the ordered lambda gene library was examined by hybridization. Most of the clones were localized on the map, except for clones E332 and E2210, which occurred at approximately 55 to 56 min. The linkage of these clones has been confirmed by both the end-labeling method and hybridization with *Not*I-generated fragments. However, the fragments from these clones hybridized to the clones at 64.5 to 65.5 min of the ordered lambda gene library. Since both libraries were constructed from the same strain, it is likely that translocation of this region occurred during the passages in different laboratories. By combining the two independent sets of hybridization data with the *Not*I digests and ordered lambda gene library, we located about 700 clones along the circular chromosome map. To correlate this alignment with actual gene locations, we tested the complementation ability of genetic markers by using known chromosomal mutations. Plasmid DNA from candidate clones was introduced into competent cells of each mutant, and complementation was assayed under conditions that favored detection of marked changes of phenotype. Of 90 mutants tested, 70 were found to convert the phenotypes to the wild type. Plasmid DNA was prepared from each type of transformant, and marker rescue was confirmed for 52 mutants (Table 1) by retransformation.

By using the procedures described above, we assigned 325 independent clones to locations on the chromosome map. The regions covered by these clones extend to about 70% of the whole *E. coli* genome. Representative clones are shown in Fig. 1. The DNA regions covered by each clone are aligned under the genetic map scale. The distances between the markers on the genetic map do not correspond to those of the physical map (28, 43), so that the assignment of the bar length, as well as terminal positions of each clone, is only relative. The large inversion between *rrnD* (72.1 min) and *rrnE* (91.5 min) has been reported to exist in W3110 (21). Since we have also confirmed this inversion, the region is placed in an inverted orientation on the map.

Despite intensive efforts, we were unable to cover the entire chromosome. There are several possible reasons for this difficulty. At the beginning of this study, we noted that although most clones were stably maintained in the *recA* strain, DH1, some clones were lost or partly deleted during propagation even after the establishment of clones. Although the copy number of pHC79 clones has been estimated to be as low as five, increase of some target sequences or genes which play key roles for regulation of cell growth might result in harmful effects on the host. There is another possibility, i.e., that *recA*-independent recombination is involved in deletion, as has been suggested by other workers (45). Another possibility is that the restriction sites used in cloning are extremely unequally distributed. According to the data from the ordered lambda gene library (28), *Eco*RI generates many fragments too large to be packaged in a lambda phage capsid. On the other hand, if the restriction sites for the same enzyme occur frequently, it would be difficult to generate fragments covering such regions even under controlled digestion conditions.

TABLE 1. Bacterial strains

Strain	Relevant genotype	Map position (min)	Source or reference
W3110	Wild type		Laboratory collection
DH1	*recA*		Laboratory collection
CBK044	*thrA*	0	41
CBK031	*car*	1	41
LC102	*leuB*	2	
	lacY	8	
	purE	12	
	his	44	
	argG	45	
	xyl	80	4
JE5440	*hpt*	3	22
E486	*polC*	4	47
AB1157	*proA*	6	25
GIA37	*argF*	7	49
607plsA	*adk*	11	19
AT1325	*lip*	15	20
KL235	*serS*	20	
	pit	77	30
NFB203	*putA*	23	
	pyrC	23	7
PLK1273	*trpE*	28	6
JP1112	*pheS*	38	40
K27	*fadD*	40	35
AB2880	*shiA*	43	37
JM477	*ptsF*	46	26
E101	*nrdA*	49	16
NK6073	*purF*	50	33
JE7094	*dapA*	53	12
IQ399	*hisS*	54	42
JC2915	*cysC*	59	44
AT713	*lysA*	61	B. J. Bachmann
JE5661	*metC*	65	N.I.G.[a]
PC3	*dnaG*	67	9
TG11	*uxaA*	68	18
CBK103	*cysG*	74	48
JE6889	*asd*	76	N.I.G.
MFT1181	*ftsE*	76	38
RK1041	*mtl*	81	27
LC173	*dnaA*	83	32
TH16	*glnA*	87	46
JE6089	*rha*	88	W. Epstein
PA505	*metA*	91	
	malB	92	M. Schwarz
JGC155	*ssb*	92	17
PA3092	*mel*	93	Y. Hirota
CBK090	*pyrB*	97	41
MP1	*uxuB*	98	39
PC2	*dnaC*	99	8
AT2459	*serB*	100	B. J. Bachmann

[a] N.I.G., National Institute of Genetics, Japan.

During hybridization analysis, we often encountered the problem that even under highly stringent conditions, a DNA region hybridized to a number of clones which spread all over the genome. This false hybridization caused serious problems, especially when neighboring clones were picked up in walking experiments. We assume that this is due to the short repeated sequences which were dispersed on the chromosome.

During the past few years, projects aimed at construction of ordered sets of DNA clones covering the entire genomes of *Caenorhabditis elegans* (13), *Saccharomyces cerevisiae* (34), and *E. coli* (14, 28) have been undertaken. In the last case, complete coverage of the genome has been made by using a lambda phage cloning system (28). Compared with

virulent phage clones, those with plasmid vectors have the advantage that DNA to be tested can be easily introduced into host cells and subjected to genetic analysis, since a variety of mutant cells are available. In this regard, the cosmid libraries reported here could be useful in a study of the genetic system of *E. coli* as well as providing a DNA source for the search for new genes.

We are grateful to Y. Hotta for help in the preparation of the manuscript.

This work was partly supported by a grant from the Ministry of Education.

LITERATURE CITED

1. **Asada, K., S. Nakatani, and M. Takanami.** 1985. Cloning of the contiguous 165-kilobase-pair region around the terminus of *Escherichia coli* K-12 DNA replication. J. Bacteriol. **163**:398–400.
2. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
3. **Bachmann, B. J.** 1987. Deriviations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190–1219. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
4. **Berg, C. M., and R. Curtiss III.** 1967. Transposition derivatives of an Hfr strain of Escherichia coli K-12. Genetics **56**:503–525.
5. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA Nucleic Acids Res. **7**:1513–1523.
6. **Bitner, R. M., and P. L. Kuempel.** 1981. P1 transduction map spanning the replication terminus of *Escherichia coli* K-12. Mol. Gen. Genet. **184**:208–212.
7. **Bohin, J.-P., and E. P. Kennedy.** 1984. Mapping of a locus (*mdoA*) that affects the biosynthesis of membrane-derived oligosaccharides in *Escherichia coli*. J. Bacteriol. **157**:956–957.
8. **Carl, P. L.** 1970. *Escherichia coli* mutants with temperature-sensitive synthesis of DNA. Mol. Gen. Genet. **109**:107–122.
9. **Chen, P. L., and P. L. Carl.** 1975. Genetic map location of the *Escherichia coli dnaG* gene. J. Bacteriol. **124**:1613–1614.
10. **Clarke, L., and J. Carbon.** 1975. Biochemical construction and selection of hybrid plasmids containing specific segments of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **72**: 4361–4365.
11. **Clarke, L. and J. Carbon.** 1976. A colony band containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
12. **Cooper, S.** 1966. Utilization of D-methionine by *Escherichia coli*. J. Bacteriol. **92**:328–332.
13. **Coulson, A., J. Sulston, S. Brenner, and J. Karn.** 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **83**:7821–7825.
14. **Daniels, D. L., and F. R. Blattner.** 1987. Mapping using gene encyclopaedias. Nature (London) **325**:831–832.
15. **Evans, R., N. R. Seeley, and P. L. Kuempel.** 1979. Loss of *rac* locus DNA in merozygotes of *Escherichia coli* K12. Mol. Gen. Genet. **175**:245–250.
16. **Fuchs, J. A., and H. O. Karlström.** 1976. Mapping of *nrdA* and *nrdB* in *Escherichia coli* K-12. J. Bacteriol. **128**:810–814.
17. **Glassberg, J., R. R. Meyer, and A. Kornberg.** 1979. Mutant single-strand binding protein of *Escherichia coli*: genetic and physiological characterization. J. Bacteriol. **140**:14–19.
18. **Goss, T. J., and P. Datta.** 1984. *Escherichia coli* K-12 mutation that inactivates biodegradative threonine dehydratase by transposon Tn*5* insertion. J. Bacteriol. **158**:826–831.
19. **Henson, J. M., A. Blinkowa, and J. R. Walker.** 1982. The *Escherichia coli dnaW*-mutation is an allele of the *adk* gene. Mol. Gen. Genet. **186**:488–492.
20. **Herbert, A. A., and J. R. Guest.** 1968. Biochemical and genetic studies with lysine + methionine mutants of *Escherichia coli*: lipoic acid and α-ketoglutarate dehydrogenase-less mutants. J. Gen. Microbiol. **53**:363–381.
21. **Hill, C. W., and B. W. Harnish.** 1981. Inversions between ribosomal RNA genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78**:7069–7072.
22. **Hoffmeyer, J., and J. Neuhard.** 1971. Metabolism of exogenous purine bases and nucleosides by *Salmonella typhimurium*. J. Bacteriol. **106**:14–24.
23. **Hohn, B.** 1979. *In vitro* packaging of λ and cosmid DNA. Methods Enzymol. **68**:299–309.
24. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11**:291–298.
25. **Howard-Flanders, P., E. Simson, and L. Theriot.** 1964. A locus that controls filament formation and sensitivity to radiation in *Escherichia coli* K-12. Genetics **49**:237–246.
26. **Jones-Mortimer, M. C., and H. L. Kornberg.** 1974. Genetical analysis of fructose utilization by *Escherichia coli*. Proc. R. Soc. London Sec. B **187**:121–131.
27. **Kadner, R. J., and H. H. Winkler.** 1973. Isolation and characterization of mutations affecting the transport of hexose phosphates in *Escherichia coli*. J. Bacteriol. **113**:895–900.
28. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
29. **Low, B.** 1968. Formation of merodiploids in mating with a cross of rec⁻ recipient strains of *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **60**:160–163.
30. **Low, B., F. Gates, T. Goldstein, and D. Soll.** 1971. Isolation and partial characterization of temperature-sensitive *Escherichia coli* mutants with altered leucyl- and seryl-transfer ribonucleic acid synthetases. J. Bacteriol. **108**:742–750.
31. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 387–389. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Nishimura, Y., and L. Caro.** 1971. Chromosome replication in *Escherichia coli*. IV. Control of chromosome replication and cell division by an integrated episome. J. Mol. Biol. **55**:441–456.
33. **Oliver, D. B.** 1985. Identification of five new essential genes involved in the synthesis of a secreted protein in *Escherichia coli*. J. Bacteriol. **161**:285–291.
34. **Olson, M. V., J. E. Dutchik, M. Y. Graham, G. M. Brodeur, C. Helms, M. Frank, M. MacCollin, R. Scheinman, and T. Frank.** 1986. Random-clone strategy for genomic restriction mapping in yeast. Proc. Natl. Acad. Sci. USA **83**:7826–7830.
35. **Overath, P., G. Pauli, and H. U. Schairer.** 1969. Fatty acid degradation in *Escherichia coli*. Eur. J. Biochem. **2**:559–574.
36. **Phillips, T. A., V. Vaughin, P. L. Bloch, and F. C. Neidhardt.** 1987. Gene-protein index of *Escherichia coli* K-12, edition 2, p. 919–966. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
37. **Pittard, J., and B. J. Wallace.** 1966. Gene controlling the uptake of shikimic acid by *Escherichia coli*. J. Bacteriol. **92**:1070–1075.
38. **Ricard, M., and Y. Hirota.** 1973. Process of cellular division in *Escherichia coli*: physiological study on thermosensitive mutants defective in cell division. J. Bacteriol. **116**:314–322.
39. **Robert-Baudouy, J., and R. C. Portalier.** 1974. Mutations affetant le catabolisme du glucuronate chez *Escherichia coli* K12. Mol. Gen. Genet. **131**:31–46.
40. **Russell, R. R. B., and A. J. Pittard.** 1971. Mutants of *Escherichia coli* unable to make protein at 42°C. J. Bacteriol. **108**: 790–798.
41. **Shaw, K. J., and C. M. Berg.** 1979. *Escherichia coli* K-12 auxotrophs induced by insertion of the transposable element Tn*5*. Genetics **92**:741–747.
42. **Shibata, K., K. Ito, and T. Yura.** 1984. Mutation that suppresses the protein export defect of the *secY* mutation and causes cold-sensitive growth of *Escherichia coli*. J. Bacteriol. **160**: 696–701.
43. **Smith, C. L., J. G. Econome, A. Schutt, S. Klco, and C. R. Cantor.** 1987. A physical map of the *Escherichia coli* K12 genome. Science **236**:1448–1453.
44. **Stacy, K. A., and R. G. Lloyd.** 1976. Isolation of rec⁻ mutants

from an F-prime merodiploid strain of *Escherichia coli* K-12. Mol. Gen. Genet. **143**:223–232.

45. **Symington, L. S., P. Morrison, and R. Kolodner.** 1985. Intramolecular recombination of linear DNA catalyzed by the *Escherichia coli recE* recombination system. J. Mol. Biol. **186**:515–525.
46. **Ueno-Nishio, S., K. C. Backman, and B. Magasanik.** 1983. Regulation at the *glnL*-operator-promoter of the complex *glnALG* operon of *Escherichia coli*. J. Bacteriol. **153**:1247–1251.
47. **Wechsler, J. A., and J. D. Gross.** 1971. *Escherichia coli* mutants temperature-sensitive for DNA synthesis. Mol. Gen. Genet. **113**:273–284.
48. **Whalen, W. A., and C. M. Berg.** 1984. Gratuitous repression of *avtA* in *Escherichia coli* and *Salmonella typhimurium*. J. Bacteriol. **158**:571–574.
49. **Wijsman, H. J. W., and H. C. Pafort.** 1974. Pleiotropic mutations in *Escherichia coli* conferring tolerance to glycine and sensitivity to penicillin. Mol. Gen. Genet. **128**:349–357.