Reprint Series
10 December 1992, Volume 360, No. 6404, pp. 606-610.

Nature

A Whole Genome Approach to *In Vivo* DNA-Protein Interactions in *E. coli*

Ming X. Wang*, M.D., Ph.D. and George M. Church**, Ph.D.

*Laboratory of Oncology Research, Research Division,
Wills Eye Hospital and Jefferson Medical College of
Thomas Jefferson University, Philadelphia, Pennsylvania, 19107, USA
**Department of Genetics, Harvard Medical School and
Howard Hughes Medical Institute, Boston, Massachusetts, 02115, USA

A whole genome approach to *in vivo* DNA–protein interactions in *E. coli*

Ming X. Wang* & George M. Church†

* Laboratory of Oncology Research, Research Division, Wills Eye Hospital and Jefferson Medical College of Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA
† Department of Genetics, Harvard Medical School and Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

THE increasingly rapid pace at which genomic DNA sequences are being determined has created a need for more efficient techniques to determine which parts of these sequences are bound *in vivo* by the proteins controlling processes such as gene expression, DNA replication and chromosomal mechanics. Here we describe a whole-genome approach to identify and characterize such DNA sequences. The method uses endogenous or artificially introduced methylases to methylate all genomic targets except those protected *in vivo* by protein or non-protein factors interfering with methylase action. These protected targets remain unmethylated in purified genomic DNA and are identified using methylation-sensitive restriction endonucleases. When the method was applied to the *Escherichia coli* genome, 0.1% of the endogenous adenine methyltransferase (Dam methylase) targets were found to be unmethylated. Five foreign methylases were examined by transfection. Database-matched DNA sequences flanking the *in vivo*-protected Dam sites all fell in the non-coding regions of seven *E. coli* operons (*mtl*, *cdd*, *flh*, *gut*, *car*, *psp* and *fep*). In the first four operons these DNA sequences closely matched the consensus sequence that binds to the cyclic AMP-receptor protein. The *in vivo* protection at the Dam site upstream of the *car* operon was correlated with a downregulation of *car* expression, as expected of a feedback repressor-binding model.

Our method is based on three assays: (1) characterizing the genomic DNA patterns of the *in vivo*-protected methylation targets using end-labelling and gel fractionation; (2) cloning and sequencing of these sites to identify both exact DNA matches and more degenerate protein-binding motifs in databases; (3) direct physiological tests to determine partial *in vivo* protection of the target as a function of genetic and environmental changes by means of filter hybridization assays. The *E. coli* genome provides a particularly good testing system for our method as over 40% of the genome has been sequenced. We examined the genomic DNA sequence GATC by using the endogenous Dam methylase which methylates the N6 position of adenines in GATC sequences. Figure 1 shows an end-labelling experiment which demonstrates that about 20 GATC sites are completely unmethylated (0.1% of the total targets in the genome) and that many sites are partially methylated[1-6]. Methylation targets remain unmethylated as a result of binding of

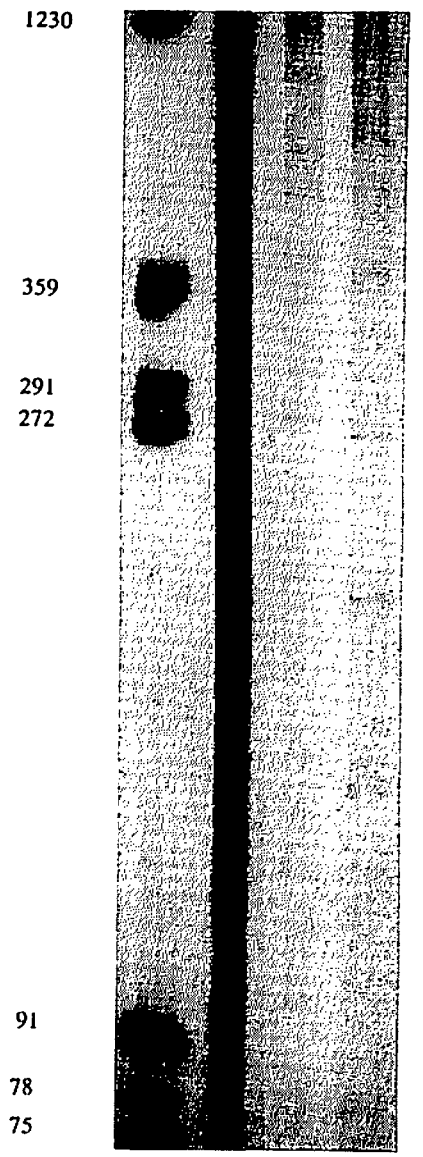

FIG. 1 Autoradiograph of an end-labelling experiment. Lane 1 is a Plex00 molecular weight standard (sizes given in nucleotides); other lanes represent *E. coli* genomic DNA cut by different combinations of endonucleases as follows: 5 μg genomic DNA was digested with restriction endonuclease I, end-labelled with 1 μl reverse transcriptase (10 U μ^{-1}; Stratagene) and 0.5 μl [α-^{32}P]dNTP (1 mCi in 50 μl), chosen to correspond to the base following the restriction site 3' terminus, which helped to reduce background labelling at random breaks. After incubation at 37 °C for 40 min, the reaction was chased with 1 μl of each dNTP at 100 mM, digested with 2 μl endonuclease II (chosen to cut the genome frequently), then electrophoresed on a 6% non-denaturing polyacrylamide gel. The gel was dried and autoradiographed. Restriction endonucleases *Mbo*I, *Sau*3A1 and *Not*I were used. *Mbo*I specifically cleaves only GATC sequences with unmethylated adenines; *Sau*3A1 acts as a control, cleaving at this sequence regardless of its adenine methylation status (about 21,000 times in the *E. coli* genome based on current sequence data); *Not*I is a single-copy intensity control which we used to help estimate the fraction of bands in the *Mbo*I digest that are especially intense owing to comigration of multiple bands or are noticeably weak as a result of partial site protection. Restriction endonucleases I and II are used in different lanes as follows: lane 2: *Sau*3A1/*Sau*3A1, representing all genomic DNA fragments containing a GATC sequence at each end; lane 3: *Not*I/*Sau*3A1, indicating single-copy intensity control; lane 4: *Mbo*I/*Sau*3A1, representing genomic fragments bearing an unmethylated GATC (*Mbo*I) site at one end and a *Sau*3A1 site at the other. About 20 prominent bands corresponding to completely unmethylated GATC sites are seen; many partially methylated sites (faint bands) are also evident. Unmethylated GATC sites were cloned as follows: 100 μg *E. coli* genomic DNA was cut with *Mbo*I, extracted twice with phenol, mixed with 1 μg *Bam*HI-cut Bluescript SK vector (Stratagene), precipitated with ethanol and resuspended in 50 μl H_2O; 2 μl each of ATP (100 mM) and T4 DNA ligase (2,000 U per μl) were then added. The ligation mixture was size-fractionated on a 1% agarose gel and any products larger than the major vector bands were excised and purified using Geneclean (Bio101), then resuspended in 50 μl H_2O. A *Cla*I reaction liberated small (6 kb on average) vecotor-genomic ligation products, and after phenol extraction and ethanol precipitation, a ligation reaction was done to circularize the products. *E. coli* cells (BRL Max Efficiency DH5α) were then transformed to produce the libraries, from which random plasmid clones were purified and 5 μl of each plasmid was treated with 0.5 μl RNaseA (10 μg μl^{-1}), denatured with 1 μl NaOH (2 M), precipitated with ethanol and sequenced using Sequenase (USB).

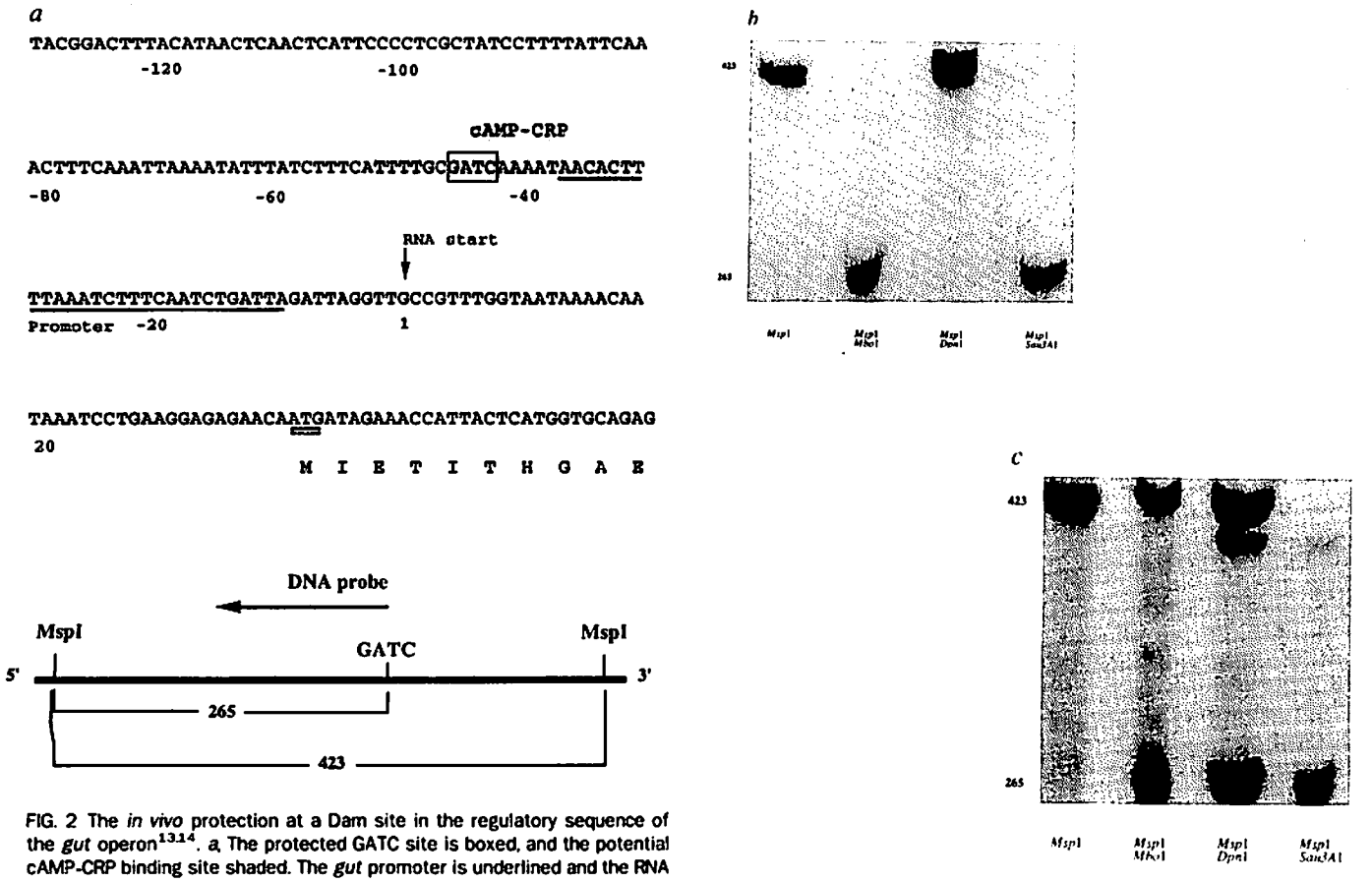

FIG. 2 The *in vivo* protection at a Dam site in the regulatory sequence of the *gut* operon[13,14]. *a*, The protected GATC site is boxed, and the potential cAMP-CRP binding site shaded. The *gut* promoter is underlined and the RNA start site is indicated with an arrow; the translational start site ATG is underlined twice. Nucleotide numbering starts with the RNA start site. The diagram illustrates the relationship between the *in vivo*-protected GATC site, the flanking restriction endonuclease sites, and the DNA probe which are used for filter hybridization (*b* and *c*). *b*, Autoradiograph of a filter hybridization experiment using a DNA probe complementary to the regulatory region of the *gut* operon. DNA was extracted from *E. coli* K-12 EMG2 grown in rich medium (1% yeast extract, 2% tryptone) to early stationary phase. After endonuclease digestion, the samples were electrophoresed on 6% denaturing polyacrylamide gels and electrotransferred to nylon, crosslinked by ultraviolet irradiation and hybridized[37,38]. The restriction enzymes used for each sample are indicated under the lanes, see *a* for explanation of the hybridization bands. Numbers indicate sizes in nucleotides. The *gut* DNA probe was constructed by 3' extension from a 20-nucleotide primer annealed to the vector near the cloning site: 10 pmol of the Bluescript plasmid (Stratgene) containing the *gut* inserts were treated with 2 μl RNase A (10 U μl^{-1}), denatured with 4 μl 2M NaOH, precipitated with ethanol and air-dried. It was then incubated with annealing mix (2 μmol 20-nucleotide primer per 4 μl USB 10× Sequenase buffer and 12 μl water) at 37 °C for 30 min. The probe was labelled with 4 μl (20 pmol) of [α^{32}P]dATP, 1 nmol each of dCTP, dGTP and dTTP using 2.5 μl (32 U) of USB Sequenase. The reaction products were denatured in formamide at 90 °C and fractionated on a 6% acrylamide-urea gel. A narrow region of the gel (80–120 nucleotides) was excised and eluted by grinding in hybridization buffer (7% SDS, 10% polyethylene glycol, 0.25 M NaCl)[37,38]. *c*, Autoradiograph of filter hybridization experiments using an *E. coli* strain with a deletion in the *crp* gene which inactivates the CRP[15]. *E. coli* K12 CGSC 7043 (λ⁻ *relA1 spoT1 thi1 rpsL136 Δcrp45*) cells were grown in minimal A medium with 1% glycerol to early stationary phase. DNA was prepared and hybridized as *b*. The additional weaker bands in lanes 3 and 4 were due to low-stringency hybridization and were not normally seen. Further comparison using two *E. coli* strains (isogeneic except the *crp* locus) show 82% *in vivo* protection in the *crp*⁺ strain and 6% protection in the *crp*⁻ strain.

protein or non-protein factors, DNA conformational steric hindrance[7] or demethylation by DNA repair. These we refer to collectively as protecting factors. A typical *E. coli* DNA-binding protein sterically blocks 10 to 70 base pairs of target DNA from access to DNase I (refs 8, 9). Hence each GATC site could lie within footprints anywhere in a region of 22 to 142 base pairs and the sum of these footprints covers about 10–60% of the *E. coli* genome. To expand the genomic portion that could be assayed, we studied foreign methylases introduced by transfection. Prokaryotic methylases, with 111 types characterized and cloned[10], make a rich source of target sequences experimentally accessible. Perturbing effects of methylases on cellular processes might confine experiments to lower intracellular methylase concentrations or to brief methylase induction periods. We explored the possible use of exogenous methylases by examining five *E. coli* strains transfected with one of the following foreign methylase genes each: *HhaI*, *HhaII*, *HpaII*, *MspI* and *TaqI*, having corresponding target DNA sequences: GCGC, GANTC, CCGG, CCGG and TCGA respectively. Methylase overproduction resulted in no obvious phenotypic effect, and the end-labelling experiments produced bands similar in number but distinct in pattern from that of Dam.

The cloning strategy (Fig. 1 legend) eliminated the initially high levels of background clones, such that essentially all genomic DNA fragments cloned contained an unmethylated GATC site each. The flanking DNA sequences of seven of the first nine such GATC sites cloned matched to the *E. coli* database in seven different operons: *mtl*, *cdd*, *flh*, *gut*, *car*, *psp* and *fep*. Remarkably, all seven *in vivo*-protected GATC sites fall in the 5' non-coding regions of these genes. As the probability that this will occur by chance is extemely small (1 in 10^6 on the basis of current *E. coli* sequence data), our finding indicates that the protecting factors have a strong and non-random preference for the upstream regulatory regions of *E. coli* genes. Table 1 lists these DNA sequences and shows that the *in vivo*-protected GATC sites are located between −230 and +5 base pairs relative to the transcription start sites. DNA sequences flanking these GATCs in four operons (*mtl*, *cdd*, *flh* and *gut*) show close matches to the consensus sequence that binds to cAMP-receptor protein (CRP). In the *cdd* operon, the putative CRP-binding

sequence, previously shown to bind to CRP strongly *in vitro*[11], is found here to contain an *in vivo*-protected GATC site. In the *mtl* operon, the protected GATC locus overlaps the most significant CRP consensus match in that region. In contrast to the above four operons, the regulatory sequences of the *car*, *psp* and *fep* operons, which were not thought to be under CRP control, lack significant matches to the CRP box. The protection at the GATC loci in the regulatory regions of these operons may be due to interactions by other factors that, for example, regulate carbamoyl phosphate synthesis (*car*), stress response (*psp*), and iron transport (*fep*). The fraction of each Dam site protected in a cell population was quantitated by filter hybridization, a method similar to those used to study vertebrate CG methylation[12]. The genomic DNA from cell cultures was analysed by appropriate methylation-sensitive restriction endonucleases and specific DNA probes were used to visualize selectively the *in vivo* protection of each methylation target in the genomes directly and without cloning. These results are also shown in Table 1.

The *gut* operon is responsible for glucitol uptake in *E. coli*[13,14]. The undermethylated GATC locus in the *gut* regulatory region (Fig. 2*a*) shows the strongest *in vivo* protection (95%; Fig. 2*b*),

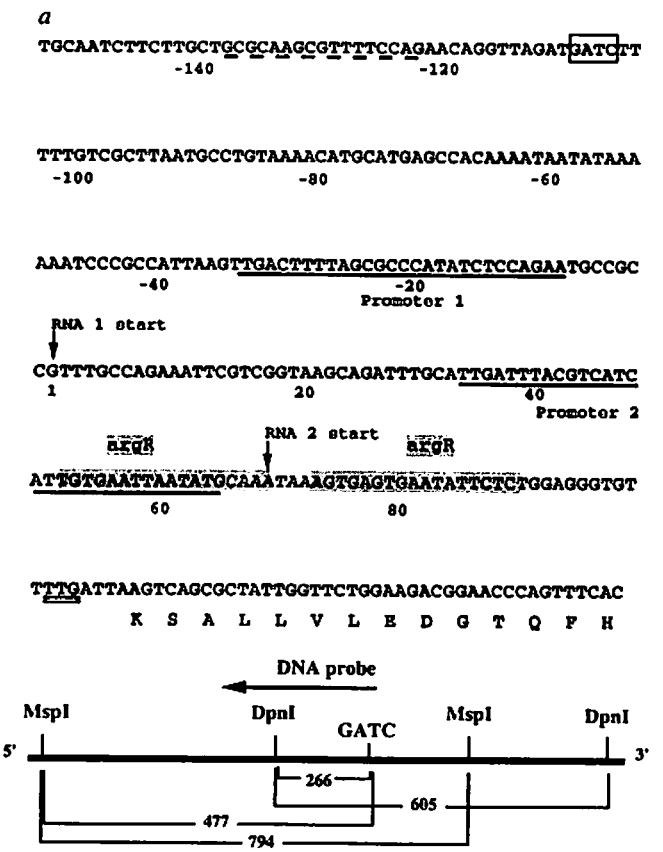

FIG. 3 The *in vivo* protection at a Dam site in the regulatory region of the *car* operon. *a*, The protected GATC site is boxed. A *pur*R box is shown with a dashed underline. Promoters P1 and P2 are underlined, and the corresponding RNA start sites are indicated by arrows. Two *arg*R boxes are represented by tandem shaded rectangles. The translational start site for *car*A initiation is underlined twice. Nucleotide numbering starts with the RNA1 start site. The diagram shows the relationship between the undermethylated GATC site, the flanking restriction endonuclease sites and the DNA probe which were used in the hybridization reactions (*b* and *c*). *b*, Autoradiograph of filter hybridization reaction using a DNA probe complementary to the *car* regulatory sequence. Strains and media were as described in Fig. 2*b*; hybridization bands are explained in *a*. Numbers indicate sizes in nucleotides. The DNA probe was constructed from a plasmid clone containing the appropriate *car* insert using methods described for Fig. 2*b*. Restriction enzymes used are indicated under the appropriate lanes. *c*, Hybridization band patterns arising from *E. coli* EMG2 cells grown in nutrient conditions with different availability of pyrimidine and arginine. Lane 1, minimal A medium; lane 2, minimal A plus uracil (500 $\mu g\ ml^{-1}$) and cytidine (1 $mg\ ml^{-1}$); lane 3, minimal A plus L-arginine (1 $mg\ ml^{-1}$); lane 4, minimal A plus uracil, cytidine and arginine. The degrees of *in vivo* protection at this *car* GATC locus (see *a*) are 0.05%, 3%, 0.07% and 15%, respectively. Hence, without pyrimidines the GATC site is essentially unprotected and independent of arginine. Pyrimidines have a significant effect and seem to be synergistic with arginine. The finding that arginine availability by itself does not affect the methylation status of the GATC site is consistent with the previous assignment[16] of arginine repressor interaction at the P2 promoter, which is located at quite a distance (138 bp) downstream from the GATC site that is pyrimidine-responsive.

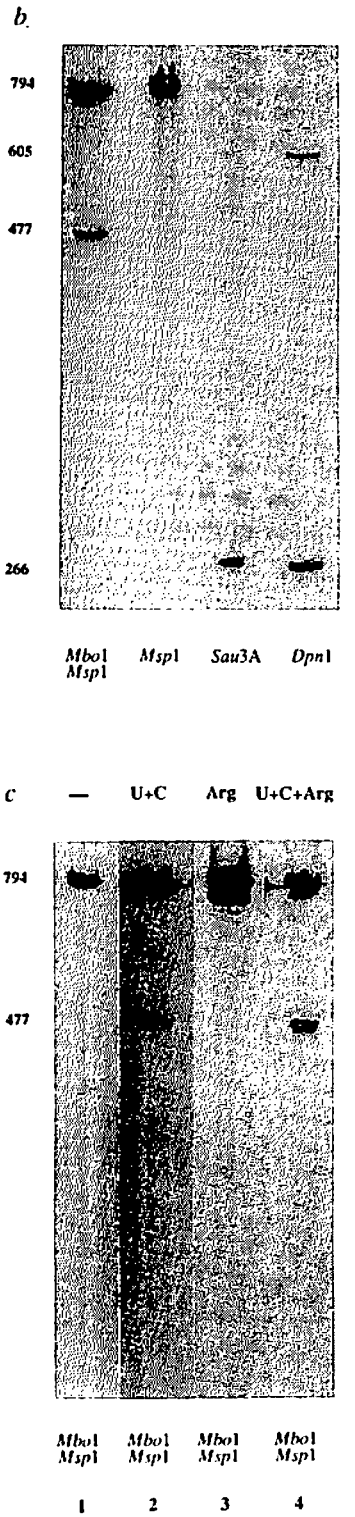

TABLE 1 DNA sequences flanking *in vivo*-protected GATC sites in the regulatory regions of seven *E. coli* operons and comparison with the consensus CRP-binding sequence

Gene	Map	Unmethylated (%) *Mbo*I	Unmethylated (%) *Dpn*I	Position	CRP consensus matches		Match score (s.d. units)
mtl	81	—		−261	5′ TAACATGCTGT	AGATCACATCA	2.7
		14	20	−218	5′ TCTTGTGATTC	AGATCACAAAT	6.2
		—		−180	5′ AAATGTGACAC	TACTCACATTT	6.6
		—		−107	5′ TTTTGTGATGA	ACGTCACGTCA	5.0
		—		−63	5′ TTATGTGATTG	ATATCACACAA	5.5
cdd	46	—		−97	5′ ATTTGCGATGC	GTCGCGATTTT	0.6
		18	22	−41	5′ TAATGAGATTC	AGATCACATAT	4.4
flh	42	38	46	−230	5′ ------GATCT	GTCATCACGAA	−0.3 to 4.2
		—		−13	5′ TGCGGTGAAAC	CGCTAAAAATA	0.2
gut	58	89	99	−47	5′ TTTTGCGATCA	AAATAACACTT	1.2
car	1	14	21	−107	5′ TTAGATGATCT	TTTTGTCGCTT	−1.9
psp	29	ND		−16	5′ ATTCTTCAATC	AGATCTTTATA	−2.2
fep	13	ND		+5	5′ ATATCCAAATA	AGATCGATAAC	−2.8
				CRP consensus:	5′ aaaTGTGAtct	agaTCACAttt	8.5

The first column gives the name of each genetic locus, the next column the location on the genetic map in minutes. The third column shows the per cent unmethylation (protection) assayed by *Mbo*I and *Dpn*I respectively through densitometry of filter hybridization bands. We selected *Dpn*I whose specificity is in a sense complementary to *Mbo*I in that it recognizes only GATC sites methylated on both strands[26]. In every case the *Mbo*I estimate is consistently lower by 4 to 10% than that of *Dpn*I, which could be due to partial enzyme cleavage or to hemimethylation (as neither enzyme cleaves hemimethylated DNA[26]). We determined the upper bounds to partial enzyme cleavage to be 3% based on the band intensities of adjacent GATC sites. The 'position' column lists the position of the G in GATC relative to the RNA start site (for segments lacking a GATC, it refers to the seventh base from the left of the CRP box). Bases matching the GATC sites in the CRP consensus are in bold. Undermethylated GATC sites are underlined. The rightmost column lists the score for a match to the consensus CRP-binding sequence. The scores are in standard deviation units (s.d.) above the mean for all *E. coli* sequences normalized to a length of 22 bp using the GCG version of the program ProfileSearch[27]. A reasonable indicator of a significant match is +2 s.d. The range covers from −2.8 to +8.5 s.d. All CRP sites noted previously[11,13,28,29-35] are included. Promoters for *flh* and *psp* lack S1 or reverse transcriptase mapping of the transcription start, instead putative −10 sequence motifs were used for alignment. A symmetric CRP consensus matrix was built by including both orientations of known CRP sites[11,13,28,29-35], which excluded all sites analysed here. In the CRP consensus binding sequence, the most commonly found base is given and the upper-case letters represent the most highly conserved base pairs which have been modelled to be in contact with the symmetric CRP protein dimer[9,36].
ND, not determined.

even though the CRP box spanning this Dam site has the lowest match score among the four CRP-regulated operons. To investigate the possible involvement of CRP, we examined this GATC locus in an *E. coli* strain with a deletion in the *crp* gene which inactivated CRP[15] and found a significantly reduced level of protection (50%; Fig. 2*c*). The change could be due to strain or environmental variations, or to an indirect effect of CRP on other proteins or nearby DNA conformations, but probably represents direct CRP binding.

The *car* operon encodes carbamoyl phosphate synthetase, a common element of the arginine and pyrimidine biosynthetic pathways. Pyrimidines and arginine repress the transcription of *car*, although the locations of protein-binding sites were uncertain[16]. We found an *in vivo*-protected GATC locus in the *car* regulatory region (Fig. 3*a* and *b*; 18% protection). Evidence that DNA-protein binding around this locus regulates *car* expression was obtained by examining *in vivo* protection as a function of *E. coli* growth conditions which differed in pyrimidine and arginine availability. As shown in Fig. 3*c*, there is no protection at this *car* GATC locus without pyrimidines, 3% protection with pyrimidines alone, and 15% when both nutrients are present. Our finding indicates that *in vivo* protection at this *car* upstream regulatory sequences is probably due to the binding of pyrimidine repressor(s). The arginine repressor acts (argR; Fig. 3*a*) synergistically with pyrimidines in down-regulating *car* expression.

The degree of methylation protection can be correlated with that of protein-binding through kinetic modelling because protein factors need to bind to DNA targets persistently to prevent methylation, whereas methylases need only a brief contact with targets to succeed in methylation. If one assumes that the methylase acts on target sequences in a genome at random, then the fraction of targets methylated per unit time is constant and the kinetic interaction between methylases and protecting factors is characterized by Poisson first-order decay: $U = 2^{-[(1-P)G/T]}$, where P is the fraction of a cell generation with protein factors that bind to the target sequence, U the unmethylated fraction of the target site at steady state, T the half-life of methylase action[17,18], and G the cell generation time. For the *in vivo*-protected *gut* GATC locus, $U = 0.95$, so $P = 0.99$, that is, protein factors bind to this DNA sequence for 99% of the cell cycle. For the GATC site in the *car* operon, $U = 0.18$, and hence $P = 78\%$.

Our method of studying *in vivo* DNA-protein interactions has several advantages over previous *in vitro* and *in vivo* footprinting techniques. *In vitro* experiments allow fractionation and modification of the interacting components but reflect artefactual deviations from intracellular states. *In vivo* footprinting techniques[19-22] typically require chemistry hazardous to cells and hence may alter chromatin structure. Our enzymatic method is less perturbing to cell integrity. Furthermore, it avoids assumptions about protein-binding sites and so is applicable for genome-scale analysis. Such methods can be applied[18,23] to the study of regulation of individual genes, as well as to overall changes in pattern in the genome and in chromosomal structures. Extension to other organisms and target sequences is dependent on the efficiency of expression of transfected methylase genes and on the compatibility of the methylation with cellular physiology. In some cases tolerance of exogenous methylases[24] can be enchanced by mutations in repair genes such as *Saccharomyces cerevisiae rad2*, or restriction genes such as *E. coli mcr* and *mrr*. We have tested five *E. coli* strains transfected with different foreign methylase genes. Application to eukaryotic genomes would seem to be feasible because methylase genes have been transfected into yeast and mammalian cells, where high levels of *in vivo* methylation of host DNA have been achieved without phenotypic consequences[24,25]. With the approaching completion of several genome sequences, whole genome approaches like ours will become increasingly important and more efficient in defining biologically functional domains in the genome. □

Received 25 June; accepted 15 September 1992.

1. Razin, A. *et al. Nucleic Acids Res.* **8**, 1783-1792 (1980)
2. Geier, G. & Modrich, P. *J. biol. Chem.* **254**, 1408-1413 (1979)
3. Blyn, L. B., Braaten, B. A. & Low, D. A. *EMBO J.* **9**, 4045-4054 (1990)

4. Ringquist, S. & Smith, C. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4539-4543 (1992).
5. Braaten, B. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 4250-4254 (1992).
6. Campbell, J. L. & Kleckner, N. *Gene* **74**, 189-190 (1988).
7. Jaworski, A. *et al. Science* **238**, 773-777 (1987).
8. Yang, C. C. & Nash, H. *Cell* **57**, 869-880 (1989).
9. Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253**, 1001-1007 (1991).
10. Wilson, G. G. *Gene* **74**, 281-289 (1988).
11. Valentin-Hansen, P. *et al. Molec. Microbiol.* **3**, 1385-1390 (1989).
12. Bird, A. P. & Southern, E. M. *J. molec. Biol.* **118**, 27-47 (1978).
13. Yamada, M. & Saier, M. *J. biol. Chem.* **262**, 5455-5462 (1987).
14. Lengeler, J. & Steinberger, H. *Molec. gen. Genet.* **164**, 163-170 (1978).
15. Sabourin, D. & Beckwith, J. *J. Bact.* **122**, 338-340 (1975).
16. Piette, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4134-4138 (1984).
17. Lyons, S. M. & Schendel, P. F. *J. Bact.* **159**, 421-423 (1984).
18. Campbell, J. L. & Kleckner, N. *Cell* **62**, 967-979 (1990).
19. Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. *Science* **227**, 134-140 (1985).
20. Becker, M. M. & Wang, J. C. *Nature* **309**, 682-687 (1984).
21. Cartwright, I. & Kelly, S. E. *BioTechniques* **11**, 188-203 (1991).
22. Saluz, H. P., Wiebauer, K. & Wallace, A. *Trends Genet.* **7**, 207-211 (1991).
23. Singh, J. & Klar, A. J. S. *Genes Dev.* **6**, 186-196 (1992).
24. Feher, Z., Schlagman, S. L., Miner, Z. & Hattman, S. *Curr. Genet.* **16**, 461-464 (1989).
25. Kwoh, T. J. *et al. Nucleic Acids Res.* **16**, 11489-11505 (1988).
26. Vovis, G. F. & Locks, S. J. *J. molec. Biol* **115**, 525-538 (1977).
27. Gribskov, M., Luthy, R. & Eisenberg, D. *Meth. Enzym.* **183**, 146-159 (1990).
28. Stormo, G. D. & Hartzell, G. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1183-1187 (1989).
29. Jiang, W. *et al. Molec. Microbiol.* **4**, 2003-2006 (1990).
30. Davis, T., Yamada, M., Elgort, M. & Saier, M. H. *Molec. Microbiol.* **2**, 405-412 (1988).
31. Silverman, M. & Simon, M. *J. Bact.* **120**, 1196-1203 (1974).
32. Bartlett, B. H., Frantz, B. B. & Matsumura, P. *J. Bact.* **170**, 1575-1581 (1988).
33. Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. *Genbank* **69.0** (1991).
34. Shea, C. M. & McIntosh, M. A. *Genbank* **69.0** (1991).
35. Ebright, R. in *Molecular Structure and Biological Activity* (eds Griffen, J. & Duax, W.) (Elsevier Scientific, New York, 1982).
36. Gunasekera, A., Ebright, Y. W. & Ebright, R. H. *Nucleic Acids Res.* **18**, 6853-6856 (1990).
37. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
38. Church, G. M. & Kieffer-Higgins, S. *Science* **240**, 185-188 (1988).

ACKNOWLEDGEMENTS. We thank R. Chin, G. Wilson, B. Bachmann, J. Roth, C. Smith, M. Rubenfield, A. Mian, R. Baldarelli and S. Kieffer-Higgins for strains, help and discussion. This work was supported by a grant from the Department of Energy. M.W. is the H. & C. Bower Fellow and is grateful to L. A. Donoso, W. S. Tasman, the Pennsylvania Lions Foundation, Research to Prevent Blindness, E. C. King Trust, and Crippled Children's Vitreo Retinal Research Foundation for support.