

AN AMAZING DISTORTION IN DNA INDUCED BY A METHYLTRANSFERASE

Nobel Lecture, December 8, 1993

by

RICHARD J. ROBERTS

New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

Much of my work in biology has been driven by my early training in chemistry. When studying a new chemical compound, the first and most important thing is to determine its detailed molecular structure. For a molecular biologist that usually means determining some DNA sequence, since an accurate knowledge of sequence will then allow the proper design of experiments to examine function. I was first exposed to the idea of macro-molecular sequences while I was a postdoctoral fellow with Jack Strominger at Harvard. During that time I briefly visited Fred Sanger's laboratory in Cambridge, England to learn the methodology of RNA fingerprinting and sequencing.

Shortly before moving to Cold Spring Harbor Laboratory, I learnt of restriction enzymes from a lecture by Dan Nathans and immediately decided that here was the key to DNA sequencing. The idea was that by mapping restriction sites and sequencing small fragments, longer gene-size sequences could be put together. My laboratory set out to isolate and characterize as many restriction enzymes as possible (Roberts, 1976). We began to use these enzymes to map Adenovirus-2 DNA (Mulder et al., 1974) and to identify small fragments that might be worth sequencing. One such fragment would contain the 5'-end of an Adenovirus-2 mRNA and the eukaryotic promoter that controlled its expression. It was the hunt for this promoter-containing fragment that led to our discovery of split genes and RNA splicing (Gelinis and Roberts, 1977; Chow et al., 1977; Gelinis et al., 1977; Broker et al., 1978). By this time, Fred Sanger and Walter Gilbert had developed DNA sequencing methods and we focused our attention on the sequence requirements for RNA splicing. Joe Sambrook, Walter Keller and others cloned the tripartite leader that was present on Adenovirus-2 late mRNAs and Sayeeda Zain determined its sequence (Zain et al., 1979). Soon we undertook the complete sequence of the Adenovirus-2 genome, which was eventually finished as a collaborative effort with Ulf Pettersson's laboratory (Roberts et al., 1986).

We realized early on that computational help would be essential for the sequencing project and we developed many programs for assembling and

analyzing DNA sequences (Gingeras et al., 1978; 1979; 1982; Gingeras and Roberts, 1980; Blumenthal et al., 1982; Keller et al., 1984). During this time we and others began to clone and sequence the genes for restriction enzymes and their companion methyltransferases (MTases) (Kiss et al., 1985; Bhagwat et al., 1986). Our initial naive assumption was that in any given restriction-modification system there would be a common region in the protein sequence of the restriction enzyme and the MTase that would pinpoint the region responsible for DNA sequence recognition. This was based on the fact that both enzymes had to recognize exactly the same DNA sequence. Once this common protein sequence had been recognized, we thought it likely that *in vitro* manipulation would allow us to change the DNA sequence recognized and so create new restriction enzymes by protein engineering. This proved hopelessly naive. Not only was there no similarity between the sequences for restriction enzyme and MTase genes, there was also no similarity between the genes for different restriction enzymes. There was, however, considerable sequence similarity among the genes for MTases, it being especially strong for the enzymes forming 5-methylcytosine. This observation has shaped the last few years of my research efforts and has led to our latest discovery, which is the topic of this lecture.

Methylation of adenine and cytosine residues is commonly found in prokaryotes and cytosine methylation is widespread in eukaryotes. In bacteria, DNA methylation serves as a component of restriction-modification systems (Roberts and Halford, 1993) and also as part of mismatch repair systems (Modrich, 1991). In higher eukaryotes, methylation of cytosine residues appears to be essential (Li et al., 1992) and is involved in the control of gene expression, developmental regulation, genomic imprinting and X-chromosome inactivation (Jost and Saluz, 1993).

There are three kinds of methylation that bacteria use to protect their DNA against the action of restriction enzymes: 5-methylcytosine (m5C), N4-methylcytosine (m4C), and N6-methyladenine (m6A). To date, 50 different m5C-MTase genes have been sequenced (Kumar et al., 1994). When we first began comparing the sequences of these genes in 1987, the available computer software was unable to provide good alignments of many of these sequences because the similarity between them was limited to short patches, which we now call motifs, that were separated by quite dissimilar regions. The similarity could be detected by eye however. Janos Posfai in my laboratory began developing algorithms that could find these small patches of similarity among proteins (Posfai et al., 1989). The program searched for the presence of small triplet patterns (Figure 1) and then combined them into motifs that represented well-conserved regions within each of the set of proteins being aligned. These motifs then anchored the initial alignment. By reducing the stringency of the triplet search and limiting it to the regions between the main motifs, weaker motifs could be found, enabling a more complete alignment. Finally, gaps could be introduced to complete the alignment. We have since refined these programs and included a user interface with graphic output that enables the overall architecture of a

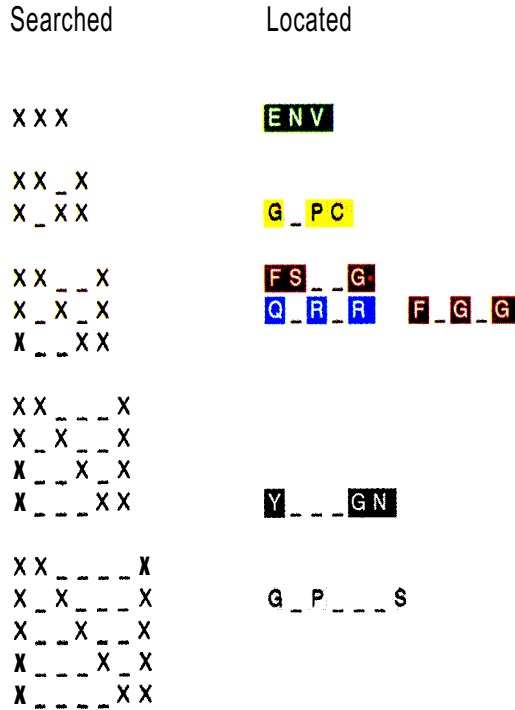


Fig. 1: Examples of the triplet patterns used to align multiple sequences. Each pattern contains three specific amino acids, which may be interrupted by positions where any amino acid is allowed. The simplest pattern, shown at the top, consists of all possible simple tripeptides (20 X 20 X 20 = 8,000 combinations). All sequences to be aligned were searched for the presence of each of these patterns (224,000 total within a frame of 10 amino acids) and their frequencies recorded. Examples of common patterns found in the m5C-MTases are shown in the right-hand column.

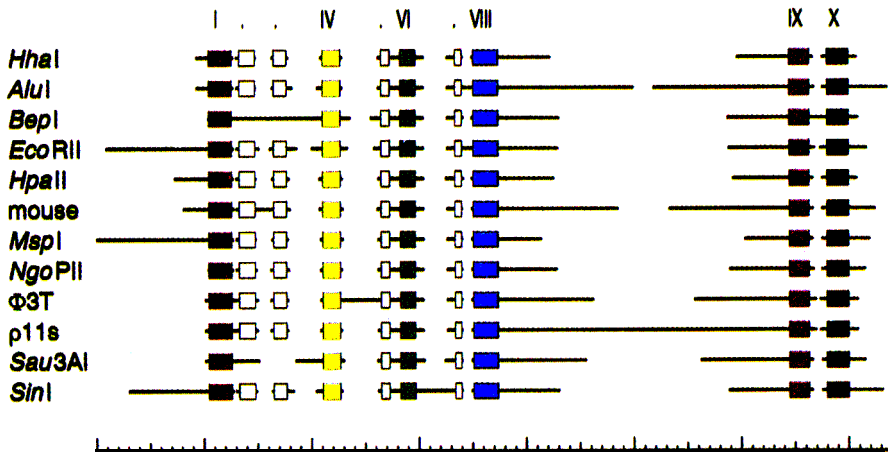


Fig. 2: Schematic showing the alignment of 12 representative m5C-MTases. The six well-conserved motifs that anchored the original alignment are shown in color (red: motif I, F-G-G; yellow: motif IV, PC; green: motif VI, ENV; cyan: motif VIII, Q-R-R; magenta: motif IX, RE; dark blue: motif X, GN). The open boxes represent the four weakly-conserved motifs.

group of proteins to be visualized. A schematic illustrating the relationships among the m5C-MTases is shown in Figure 2. There are six regions of strong similarity and four more regions of lesser similarity that can be used to anchor the alignments between these sequences. Connecting these motifs are regions that vary in length and sequence. Similar analyses of the m5C-MTases have been reported by others (Lauster et al., 1989).

Among the six well-conserved motifs, two could be assigned functional significance. Motif IV, which is shown in block form in Figure 3A, contains a cysteine residue that was postulated by Dan Santi to be a key catalytic residue (Santi and Danenberg, 1984). Santi had proposed that an initial step in the reaction involved the formation of a covalent complex between this cysteine residue in the MTase and the 6-position of cytosine in a Michael reaction. This activates the 5-position of cytosine and permits transfer of the methyl group from the cofactor S-adenosylmethionine (AdoMet). The proposed reaction pathway is shown in Figure 3B. Much biochemical evidence was available to support this mechanism, including the important observation that 5-fluorocytosine was a mechanism-based inhibitor (Wu and Santi, 1987). Later work has enabled the isolation of the covalent intermediate (Chen et al., 1991; Friedman and Ansari, 1992; Smith et al., 1992) and the cysteine within motif IV has been shown directly to be the site of covalent bond formation (Chen et al., 1991). Recent experiments have shown that mutation of this cysteine residue to glycine, serine or other

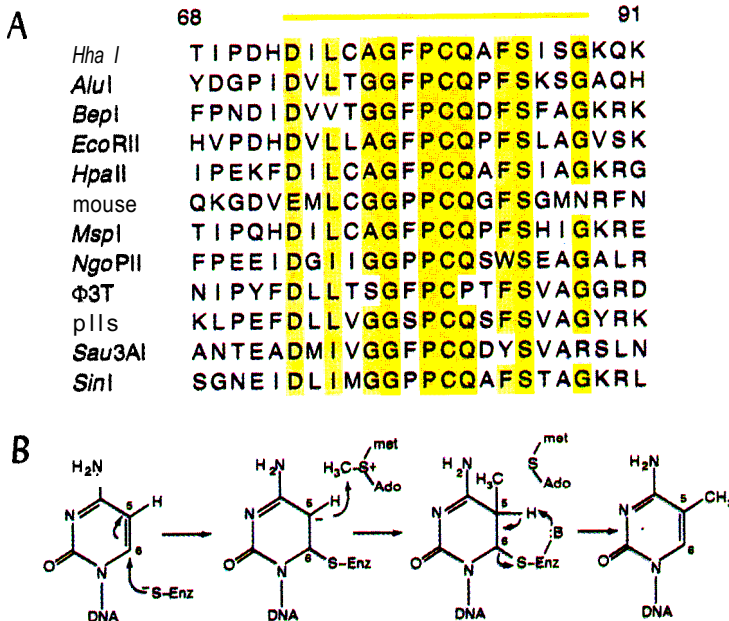


Fig. 3. A. Block diagram showing well-conserved motif IV (the PC motif). Residues conserved among all m5C-MTase sequences are strongly highlighted and those that show three or fewer variants are lightly shaded.

B. Schematic showing the reaction pathway.

amino acids blocks MTase action (Wyszynski et al., 1992; Mi and Roberts, 1993; Chen et al., 1993).

Motif I could also be assigned a functional role. It contained three highly conserved residues, FGG, as shown in Figure 4A. Importantly, it shows distinct similarities to motifs found in other MTases (Figure 4B). It can be seen that all three classes of DNA MTases, m5C, m4C and m6A, as well as protein and RNA MTases all contain a related motif. The common functional feature is that each MTase uses AdoMet as the methyl donor and so motif I was suggested to be involved in binding to the cofactor (Ingrosso et al., 1989; Rlimasauskas et al., 1989).

We were interested in trying to define the region within the m5C-MTases that was responsible for sequence-specific recognition of the DNA substrate. Because the shared function of this group of MTases was the chemistry of the reaction, we argued that each of the common regions, exemplified by motifs I and IV, must be directly involved, either by interacting with the cofactor or providing residues important for catalysis. Since the precise DNA sequence recognized by each enzyme was different we expected that this recognition would be mediated by a region that differed among the sequences. Only one region, the so-called variable segment located between motifs VIII and IX, appeared a reasonable candidate to mediate the sequence-specific recognition of DNA. Examination of these variable regions shows that not only amino acid sequence, but also the lengths of the segments vary considerably.

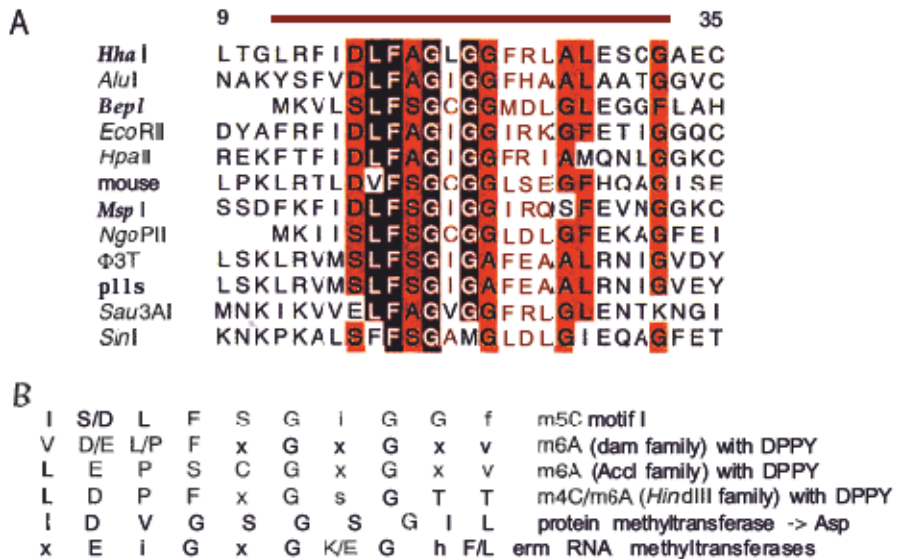


Fig. 4: A. Block diagram showing well-conserved motif I (the FGG motif). Residues conserved among all m5C-MTase sequences are strongly highlighted color and those that show three or fewer variants are lightly shaded.

B. Consensus sequences for motif I and its relatives in other MTases, based on Ingrosso et al. (1989).

Although most of the known m⁵C-MTases are mono-specific, that is they recognize a single sequence, some MTases are known that recognize more than one specific sequence. These are the so-called "multi-specific" MTases that are encoded by several *Bacillus* bacteriophages. These enzymes are single polypeptides like the monospecific enzymes, but have the unusual property of methylating several quite different sequences. One example is M.Φ3TI, which recognizes and methylates GGCC and GCNGC (Noyer-Weidner et al., 1983). It is included in Figure 2 from which it can be seen that its variable region is extraordinarily long. A series of elegant studies carried out in Tom Trautner's laboratory has showed convincingly that single mutations within the variable regions of M.Φ3TI and other multi-specific MTases knock out the ability of the enzyme to recognize one of the specific sequences, while still allowing the others to be recognized (Balganesh et al., 1987; Wilke et al., 1988). In the case of the mono-specific MTases, a comparable experiment was not possible, since mutations that prevented DNA recognition would also block MTase activity. Saulius Klimasauskas and Janise Nelson in my laboratory undertook a series of domain swap experiments between two of the mono-specific MTases (Klimasauskas et al., 1991). We chose M.HpaII (recognition sequence: CmCGG) and M.HhaI (recognition sequence: GmCGG) because they both recognize tetranucleotide sequences and the base methylated is located at an equivalent position within the recognition sequences. The variable regions, with or without flanking sequences, were swapped as illustrated in Figure 5. In several cases, active MTases resulted and comparison of the hybrids showed clearly that sequence specificity lay entirely within the variable region. Surprisingly, hybrids in which the variable region plus its adjacent motif IX were transferred showed much higher methylation activity than hybrids in which just the variable region had been swapped.

In another series of domain swap experiments, Mi Sha swapped the variable regions from M.HpaII and M.MspI. These enzymes both recognize the sequence, CCGG, but M.MspI transfers the methyl group to the first cytosine residue in the sequence (Walder et al., 1983) whereas M.HpaII transfers the group to the second cytosine residue (Mann and Smith, 1979). The results of these experiments showed that the choice of base to be methylated also depends upon the variable region (Mi and Roberts, 1992).

While pursuing our studies of the biochemistry and molecular biology of MTases, it became apparent that a crystal structure would be essential if we were to understand fully the reaction mechanism. Ashok Dubey, in my lab, began a collaboration with Xiaodong Cheng, working in Jim Pflugrath's laboratory at Cold Spring Harbor. They purified and attempted to crystallize M.MspI, which we had studied extensively (Lin et al., 1989; Dubey et al., 1992; Dubey and Roberts, 1992). Later they tried to crystallize M.HpaII, on which we had also worked (Card et al., 1990). In both cases, the efforts were unsuccessful. However, our third attempt, using M.HhaI was quite successful. This enzyme forms part of the HhaI restriction system from *Haemophilus haemolyticus*, which had been discovered in my laboratory (Roberts et al.,

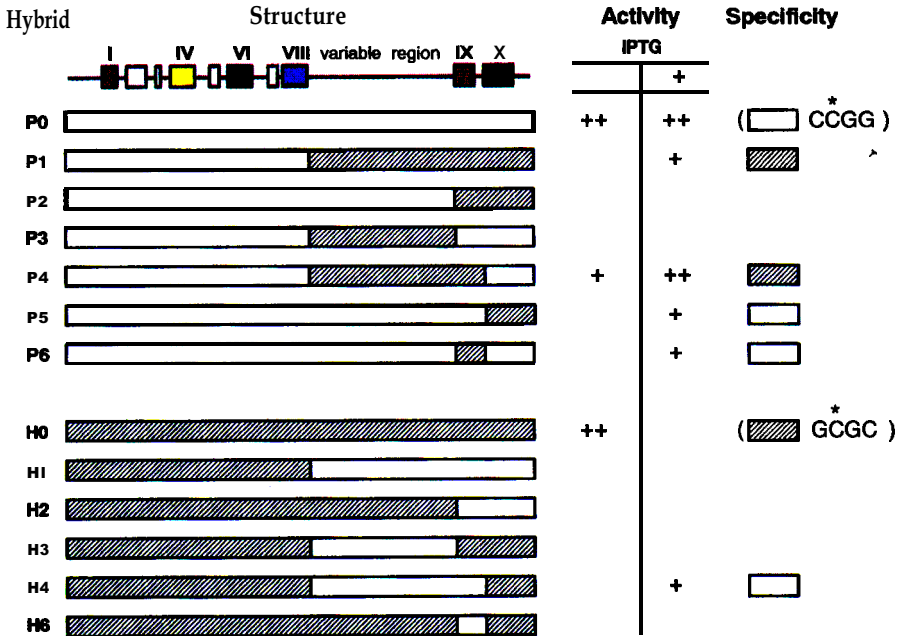


Fig. 5: Schematic showing the domain swaps between M.HpaII and M.HhaI. Various hybrids were constructed between M.HpaII (P0) and M.HhaI (H0) as indicated. The MTases were under the control of an IPTG-inducible promoter on a pBR322 derivative. Sequences derived from M.HpaII sequences are shown by open boxes; those from M.HhaI are shown by shaded boxes. The column labelled "Activity" shows the extent of protection of plasmid DNA *in vivo*, as measured by restriction enzyme cleavage *in vitro*, either before (-) or after IPTG-induction (- indicates no protection, + indicates weak protection, ++ indicates > 90 % protection). The specificity of the hybrid methylase is shown in the final column.

1976). It is one of the smallest of the m5C-MTases, containing 327 amino acids (M, 37 Kd). Its gene has been cloned and sequenced (Caserta et al., 1987) and the protein was overexpressed in *E. coli* (Wu and Santi, 1988; Klimasauskas et al., 1991) and subjected to detailed kinetic studies (Wu and Santi, 1987). Sanjay Kumar purified the enzyme and, in December 1991, it crystallized readily - a wonderful Christmas present! Within 11 months Xiadong Cheng had a structure for the binary complex between M.HhaI and AdoMet at 2.5 Å resolution (Cheng et al., 1993).

The structure was most revealing (Figure 6), being composed of a large domain and a small domain forming a cleft that appeared ideal to accommodate a DNA helix. Motif I (FGG) was confirmed to be involved in AdoMet binding, while another of the conserved motifs, X (GN), was also implicated in AdoMet binding. Motif IV (PC) was positioned close to the AdoMet-binding site on the same side of the cleft. Motif VIII (QRR) lay at the base of the cleft, where one could imagine that the positively charged residues of the motif could play a role in binding to the phosphodiester backbone of the DNA helix. Conserved motif IX (RE) threaded its way through the small domain which otherwise consisted of almost the entire variable region.

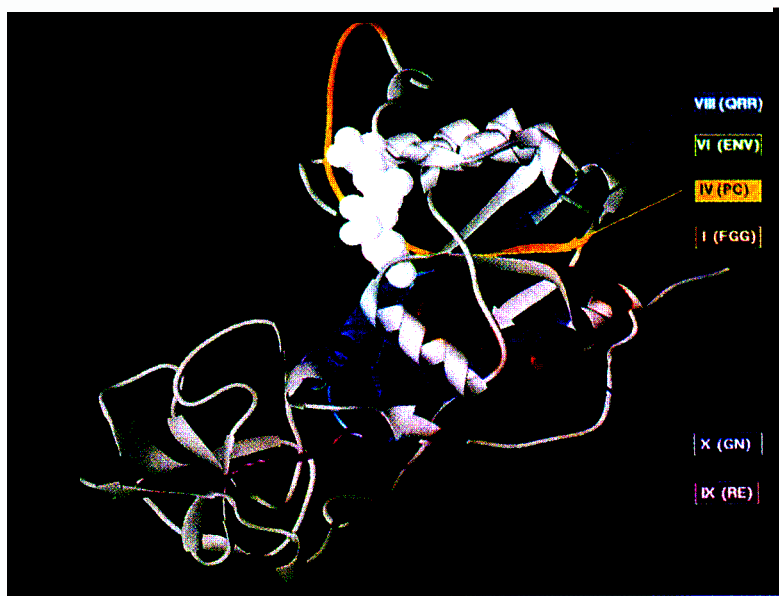


Fig. 6: The structure of *M.HhaI* in a binary complex with its cofactor AdoMet. The motifs are colored as in Figure 2. The co-factor, AdoMet, is shown in white as a space-filling representation.

Motif IX appeared to form a backbone responsible for bracing the structure of the small domain. Finally, conserved motif VI (ENV) was clearly involved in correctly positioning motif IV.

The structure provided a clear explanation for our earlier observation that domain swaps that included both motif IX and the variable region were more active as MTases than those involving just the variable region (Klimauskas et al., 1991; Mi and Roberts, 1992). Although motif IX is quite highly conserved among the MTases, there are differences from one MTase to another and these could be expected to influence the ability of a heterologous motif IX to brace the structure of the small domain.

Soon after we obtained crystals of the binary complex between *M.HhaI* and AdoMet, we set about trying to obtain co-crystals that would also include DNA. Two kinds of complexes were envisioned. One would contain *M.HhaI* together with native DNA and S-adenosylhomocysteine (AdoHcy), while another would contain *M.HhaI* with AdoMet and a DNA duplex substituted with 5-fluorocytosine at the target. The latter would be expected to form a covalent intermediate in which the methyl group had transferred to the 5-position of cytosine, but release of the DNA from the covalent complex with the protein would be inhibited by the presence of the fluorine atom, which is a very poor cleaving group. Such a complex could almost be viewed as a transition state intermediate.

Saulius Klimauskas joined the project for the co-crystallization and we were fortunate in obtaining good crystals fairly readily with both native

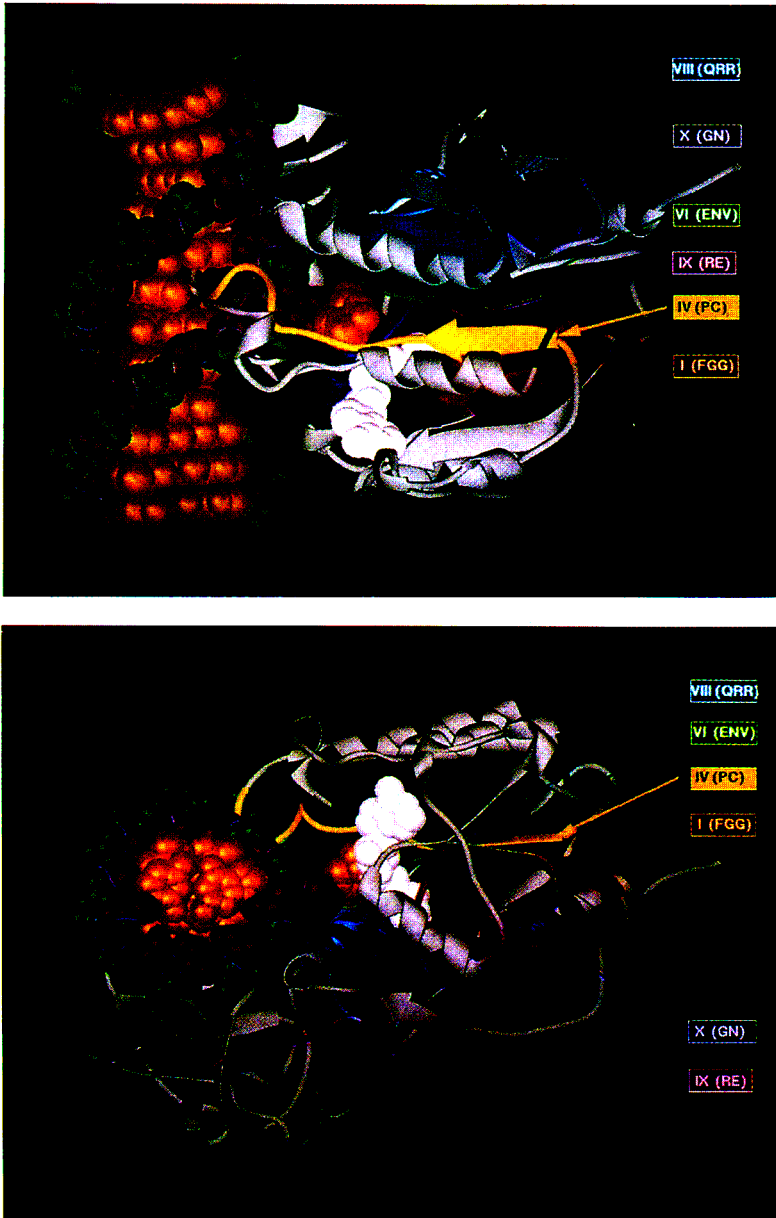


Fig. 7: The structure of *M.HhaI* in a ternary complex with a substrate duplex DNA oligonucleotide and the end product of the reaction AdoHcy. The motifs are colored as in Figure 2. AdoHcy is in white, the DNA bases are orange, the deoxyribose is purple and the phosphates are green. A. View from the side of the DNA axis. B. View looking down the DNA axis.

DNA and DNA containing 5-fluorocytosine. Xiaodong Cheng, who now has his own laboratory at Cold Spring Harbor, has solved the co-crystal structure at 2.8 Å resolution (Figure 7) (Klimasauskas et al., 1994). We had expected to find that the DNA helix would be distorted in some way to allow the chemical reaction to take place, because the chemistry requires that the

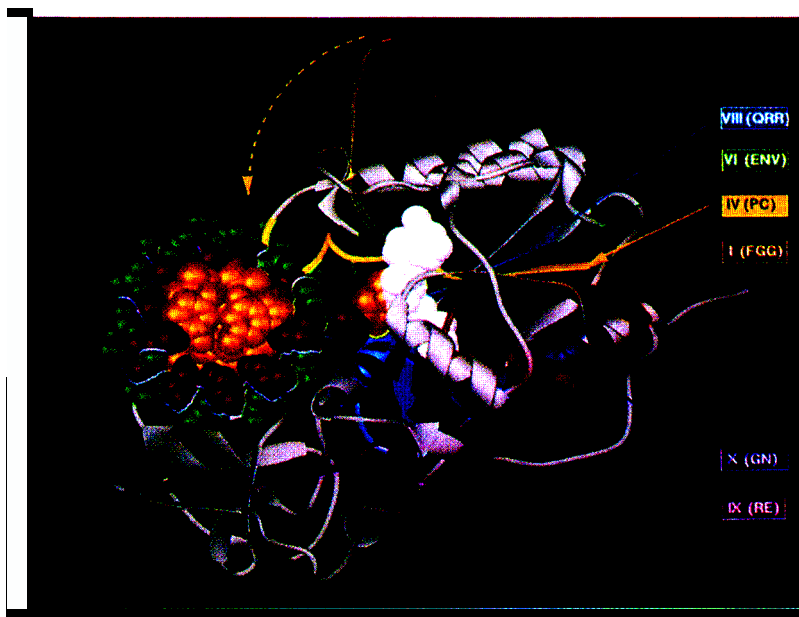


Fig. 8: Composite of the *M.HhaI* structures with and without DNA showing the conformational changes that take place upon binding to DNA. The colors are as in Figure 7. The active site loop (in yellow) is shown as a solid line to indicate its position in the ternary complex structure (with DNA) and as a triple-stranded ribbon in the binary complex structure (no DNA).

cysteine residue must approach the cytosine in a direction perpendicular to the plane of the ring. We imagined at most some extreme bending of the DNA. Unexpectedly, the distortion is much greater than a bend and much more elegant. The target cytosine flips right out of the axis of the DNA helix and into a pocket in the enzyme that contains the catalytic cysteine residue. The rest of the helix is relatively undistorted. Another large conformational shift has taken place in the active site loop which encompasses motif IV. The entire 20-residue loop rotates through almost 180 degrees from its original position to bring the catalytic cysteine into contact with the target cytosine (Figure 8).

The position in the DNA helix that was occupied by the target cytosine, is now filled by two residues. One is a glutamine residue from the small domain and the other is a serine residue from the active site loop. Surprisingly, neither of these residues is conserved among other m5C-MTases suggesting that many residues could assist in opening a helix in this manner. As can be seen from Figure 7, the interaction between *M.HhaI* and DNA can be thought of as the enzyme embracing the DNA with two arms infiltrating into the helix during the embrace and assisting the extrusion of the target cytosine.

The sequence specific interactions between *M.HhaI* and its substrate mainly involve two distinct loops (the sequence recognition loops) from the small domain. One loop is responsible for interactions with the orphan guanosine that is left after the cytosine has flipped and makes specific

contacts with the adjacent bases on the same strand. It also provides the glutamine residue that fills the hole left after the target cytosine is flipped. The second recognition loop interacts predominantly with the strand containing the target cytosine. However, the interactions in the covalent complex should be viewed as an end point of a more complicated interaction. They may not be an accurate reflection of the initial events which led to recognition. More information of these initial events may be provided by studies of mutant proteins or DNA analogs that can form complexes without flipping the target cytosine. It should be noted that none of the well-characterized DNA binding motifs (Harrison, 1991) appear to play a role in this system.

Previous studies of DNA protein interactions have shown quite dramatic distortions induced in DNA by proteins, but usually these have involved bends or kinks with one of the most dramatic being the flattening and bending of DNA induced by the TATA-box binding protein (Rim, Y. et al., 1993; Rim, J. et al., 1993). There have been no previous examples of proteins interacting with DNA and causing a base to flip out of the helix. Since the DNA bases lie buried in the inside of the helix in normal DNA, the mechanism presented here provides an elegant means by which complete access to the base becomes possible. We anticipate that other proteins performing chemistry on DNA bases will also use this mechanism. Obvious candidates would be the MTases that form N6-methyladenine or N4-methylcytosine. Some of the enzymes, such as DNA glycosylases, that repair DNA damage might also flip the damaged base out of the helix prior to its excision. Many proteins that interact with DNA need to open up the helix. Some examples would be topoisomerases, helicases, DNA polymerases and/or their auxiliary proteins that operate at replication origins, RNA polymerases and recombination enzymes. In the structure shown in Figure 9 the phosphodiester bonds, adjacent to the target cytosine, are distorted from their positions in normal B DNA and one might imagine the continued unzipping of the helix would be easy.

Surprisingly, the interaction between *M.HhaI* and DNA requires no external energy source. Conformational rearrangements within the protein combined with specific interactions between protein and DNA likely provide the energy to open the helix. The energy is then stored for use during the return of the target cytosine. It would be surprising if this mechanism is not used elsewhere. Split genes were discovered unexpectedly, but proved to be easily found once we knew they were there. This mechanism of flipping a base out of the DNA helix might also prove to be of widespread importance as a first step in opening up a DNA helix. We should lose no time in exploring the possibility.

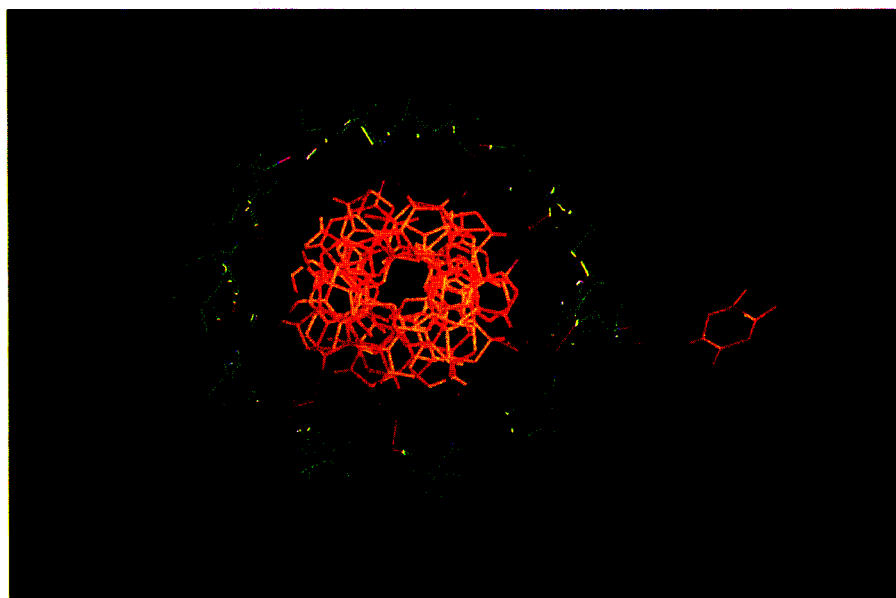
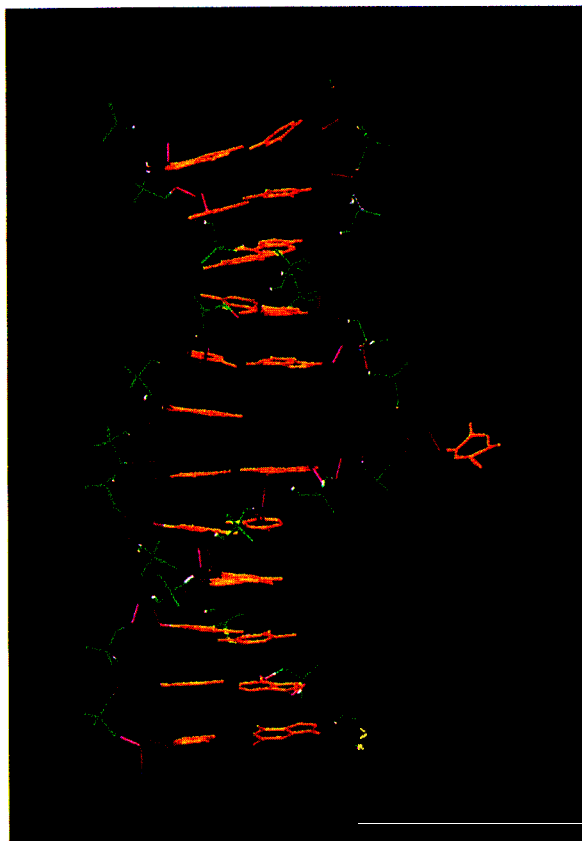


Fig. 9: Views of the DNA in the ternary complex with the protein omitted. The bases are yellow, the deoxyribose ring is purple and the phosphates are green. A. View from the side of the helix axis. B. View looking down the helix axis.

ACKNOWLEDGEMENTS

I am most grateful to all of my colleagues who have worked so hard in my laboratory to ensure our success. None of this would have been possible without their input. In addition to those mentioned explicitly in the text I thank Phyllis Myers, Kathy O'Neill, Margaret Wallace, Louise D'Alessandro, Jodie Freyer, Carol Marcincuk, Dana Macelis and Ching Lin whose dedication has been especially important in making my life easier. I thank Jim Watson for persuading me to go to Cold Spring Harbor Laboratory and for many stimulating conversations. Among many colleagues, Ashok Bhagwat, Mike Botchan, Don Comb, Greg Freyer, Rich Gelin, Tom Gingeras, David Helfman, Winship Herr, Adrian Krainer, Stu Linn and Ira Schildkraut have been good friends and valuable sources of criticism and advice. I thank NIH and NSF for much support over the years. Finally, I am indebted to my wife, Jean, and my children, Alison, Andrew, Christopher and Amanda who have been supportive and understanding of my addiction to science.

REFERENCES

- Balganesh, T.S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K. and Trautner, T.A. (1987). *EMBO J.* 6, 3543-3549.
- Bhagwat, A.S., Sohail, A. and Roberts, R.J. (1986) *J. Bacteriol.* 166: 751-755.
- Blumenthal, R.M., Rice, P.R. and Roberts, R.J. (1982) *Nucl. Acids Res.* 10, 91-101.
- Broker, T.R., Chow, L.T., Dunn, A.R., Gelin, R.E., Hassell, J.A., Klessig, D.F., Lewis, J.B., Roberts, R.J. and Zain, B.S. (1978) *Cold Spring Harbor Symp. Quant. Biol.* 42, 531-553.
- Card, C.O., Wilson, G.G., Weule, K., Hasapes, J., Kiss, A. and Roberts, R.J. (1990) *Nucl. Acids Res.* 18, 1377-1383.
- Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G.G. and Wells, R.D. (1987). *J. Biol. Chem.* 262, 4770 - 4777.
- Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) *Biochemistry* 30, 11018-11025.
- Chen, L., MacMillan, A.M. and Verdine, G.L. (1993) *J. Amer. Chem. Soc.* 115, 5318-5319.
- Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W. and Roberts, R.J. (1993) *Cell* 74, 299-307.
- Chow, L.T., Gelin, R.E., Broker, T.R. and Roberts, R.J. (1977) *Cell* 12, 1-8.
- Dubey, A.K., Mollet, B. and Roberts, R.J. (1992) *Nucl. Acids Res.* 20, 1579-1585.
- Dubey, A.K. and Roberts, R.J. (1992) *Nucl. Acids Res.* 20, 3167-3173.
- Friedman, S. and Ansari, N. (1992) *Nucl. Acid-s Res.* 20, 3241-3248.
- Gelin, R.E. and Roberts, R.J. (1977) *Cell* 11, 533-544.
- Gelin, R.E., Chow, L.T., Roberts, R.J., Broker, T.R. and Klessig, D.F. (1977) In: *Brookhaven Symposium in Genetic Interaction and Gene Transfer, Brookhaven Symposium in Biology* 29, 345 - 347.
- Gingeras, T.R., Milazzo, J.P. and Roberts, R.J. (1978) *Nucl. Acids Res.* 5, 4105-4127.
- Gingeras, T.R., Milazzo, J.P., Sciaky, D. and Roberts, R.J. (1979) *Nucl. Acids Res.* 7, 529-545.
- Gingeras, T.R., Rice, P.I. and Roberts, R.J. (1982) *Nucl. Acids Res.* 10, 103-114.
- Gingeras, T.R. and Roberts, R.J. (1980) *Science* 209, 1322-1328.

- Harrison, S.C. (1991) *Nature* **353**, 715-719.
- Ingrosso, D., Fowlcr, A.V., Bleibaum, J. and Clarke, S. (1989) *J. Biol. Chem.* **264**, 20130-20139.
- Jost, J.P. and Saluz, H.P. (1993) "DNA methylation: molecular biology and biological significance" Birkhauser, Basel.
- Keller, C., Corcoran, M. and Roberts, R.J. (1984) *Nucl. Acids Res.* **12**, 379-386.
- Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993) *Nature* **365**, 520-527.
- Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993) *Nature* **365**, 512-520.
- Kiss, A., Posfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) *Nucl. Acids Res.* **13**, 6403-6421.
- Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) *Cell* **76**, 357-369 (1994).
- Klimasauskas, S., Nelson, J.E. and Roberts, R.J. (1991) *Nucl. Acids Res.* **19**, 6183-6190.
- Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. and Janulaitis, A.A. (1989) *Nucl. Acids Res.* **17**, 9823-9832.
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. and Wilson, G.G. (1994) *Nucl. Acids Res.* **22**, 1 - 10 (1994).
- Kumar, S., Chng, X., Pflugrath, J.W. and Roberts, R.J. (1992) *Biochemistry* **31**, 8648-8653.
- Lauster, R., Trdutner, T.A. and Noycr-Wcidner, M. (1989) *J. Mol. Biol.* **206**, 305-312.
- Li, E., Bcstor, T.H. and Jacnisch, R. (1992). *Cell* **69**, 915-926.
- Lin, P.M., Lee, C.H. and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 3001- 3011.
- Mann, M.B. and Smith, H.O. (1979) *Nucl. Acids Res.* **4**, 421 1-4221.
- Mi, S. and Roberts, R.J. (1992) *Nucl. Acids Res.* **20**, 4811-4816.
- Mi, S. and Roberts, R.J. (1993) *Nucl. Acids Res.* **21**, 2459-2464.
- Modrich, P. (1991) *Annu. Rev. Genet.* **25**, 229 -253.
- Muldr, C., Arrand, J.R., Dclius, H., Keller, W., Petterson, U., Roberts, R.J. and Sharp, P.A. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **39**, 397-400.
- Noyer-Weidncr, M., Jentsch, S., Pawlck, B., Gunthert, U. and Trautner, T.A. (1983) *J. Virol.* **46**, 446-453.
- Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 2421- 2435.
- Roberts, R.J.. (1976) *CRC Critical Reviews in Biochemistry*, **4**, 123 - 164.
- Roberts, R.J., Akusjarvi, G., Alestrom, P., Gclinas, R.E., Gingeras, T.R., Sciaky D. and Petterson, U. (1986) In: "Adenovirus DNA. The Viral Genome and Its Expression", ed. W. Doerfler, Martinus Nijhoff, Boston, Mass. pp 1-51.
- Roberts, R.J. and Halford, S.S. (1993). In *Nucleuses* S.M. Linn, R.S. Lloyd and R.J. Roberts, eds. Cold Spring Harbor Press. pp. 35 -88.
- Roberts, R.J., Myers, P.A., Morrison, A. and Murray, K. (1976) *J. Mol. Biol.* **103**, 199-208.
- Santi, D.V. and Dancnberg, P.V. (1984). In: "Folates and Pterins", R.L. Blakely, and S.J. Bcnkovic, eds. (New York, Wiley-Interscience), Vol 1, pp 343-396.
- Smith, S.S., Kaplan, B.E., Sowers, L.C. and Newman, E.M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4744-4748.
- Waldr, R.Y., Langtim, C.J., Catterjee, R. and Walder, J.A. (1983) *J. Biol. Chem.* **258**, 1235-1241.
- Wilke, K., Rauhut, E., Noyer-Weidener, M., Lauster, R., Pawlck, B., Bchrens, B. and Trautner, T.A. (1988) *EMBO J.* **7**, 2601-2609.
- Wu, J.C. and Santi, D.V. (1987) *J. Biol. Chem.* **262**, 4778-4786.
- Wu, J.C. and Santi, D.V. (1988) *Nucl. Acids Res.* **16**, 703-717.
- Wyszynski, M.W., Gabbara, S. and Bhagwat, A.S. (1992) *Nucl. Acids Res.* **20**, 319-326.
- Zain, B.S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979) *Cell* **16**, 851-861.