VOLUME 1, NUMBER 2, FEBRUARY 2001

NANO LETTERS

A Self-Assembling Nanoscale Camshaft: Implications for Nanoscale Materials and Devices Constructed from Proteins and Nucleic Acids

Steven S. Smith†

Department of Cell and Tumor Biology, City of Hope National Medical Center, 1500 East Duarte Road, Duarte, California 91010

Received October 6, 2000

ABSTRACT

The methyltransferase-directed addressing of fusion proteins to DNA scaffolds offers an approach to the construction of protein/nucleic acid biostructures with potential in a variety of applications. In this report I describe the use of this technology in the design and construction of a molecular camshaft with implications in the design of independently addressable trisubstituted components for the construction of additional nanoscale materials and devices.

The central chemical principle of the technology studied here is the capacity of the DNA (cytosine-5) methyltransferases to specifically attack their DNA recognition sequences and form a covalent bond between the enzyme protein and the target base in DNA. This is a consequence of their biological role in the methylation of these sites in DNA. It results from the normal cyclical (i.e., catalytic) chemical process in which the chemical activation of the substrate by the enzyme is followed by methyltransfer and then release of the product methylated DNA. When 5F-cytosine is substituted for cytosine at the target site[1] or when uracil is substituted for cytosine,[2] the cycle cannot be completed and the enzyme molecule forms a Michael adduct[3] with cytosine, stalls at the targeted site and becomes stabily bound to the DNA at the targeted recognition site (Figure 1).

The catalytic reaction is confined to cytosines in the targeted recognition site. For the bacterial DNA (cytosine-5) methyltransferase M•*Hha*I this site has the sequence 5′GCGC3′. Although the tetrameric sequence is the only portion of the DNA molecule that is required for recognition, the observed reaction rate is strongly influenced by the size of the DNA flanking the recognition sequence. As can be seen from Figure 2 M•*Hha*I does not catalyze significant methylation of duplex oligodeoxynucleotides in which the centrally placed recognition site is contained in a sequence that is less than 20 bp in length. Larger sequences are methylated more efficiently.

Although the enzyme makes an efficient covalent attack on a 5FdC when it is at the target base, the chemical deactivation of the enzymatically formed dihydrocytosine intermediate induced by the presence of the fluorine sub-

† Telephone: 626-301-8316. Fax: 626-301-8972. E-mail: ssmith@coh.org.

10.1021/nl005511y CCC: $20.00
Published on Web 01/24/2001

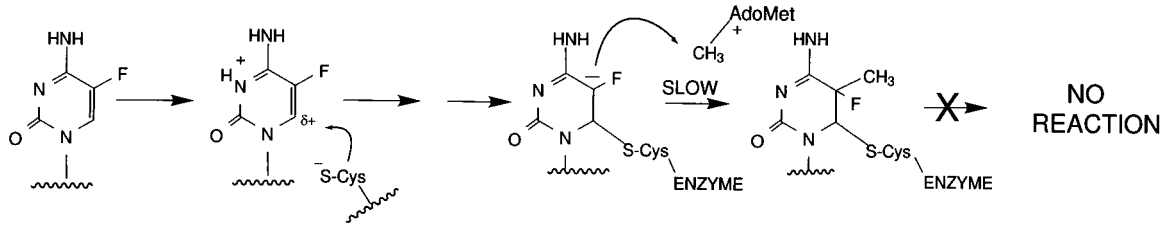

Figure 1. Attack of 5-fluorodeoxycytosine by DNA (cytosine-5) methyltransferases. Protonation of cytosine-N3 by a group on the enzyme activates the ring for nucleophilic attack by a cysteine residue at the active site. Once nucleophilic attack occurs, a 5−6, saturated intermediate attacks the methyl group on *S*-adenosylmethionine to generate a dihydrocytosine intermediate. In the biological reaction with cytosine, this intermediate undergoes β-elimination to generate a 5-methylcytosine product and active enzyme. With the 5-fluorodeoxycytosine shown here, conversion of the complex to the 5-methyl-, 5-fluoro-dihydrocytosine is very slow, and subsequent β-elimination is blocked. Both effects are attributed to the substitution of fluorine for hydrogen at C5. Fluorine substitution deactivates the dihydrocytosine intermediate, making it less probable that methyl transfer will occur. In those cases where it does occur, the fluorine at C5 cannot be abstracted and β-elimination cannot occur.

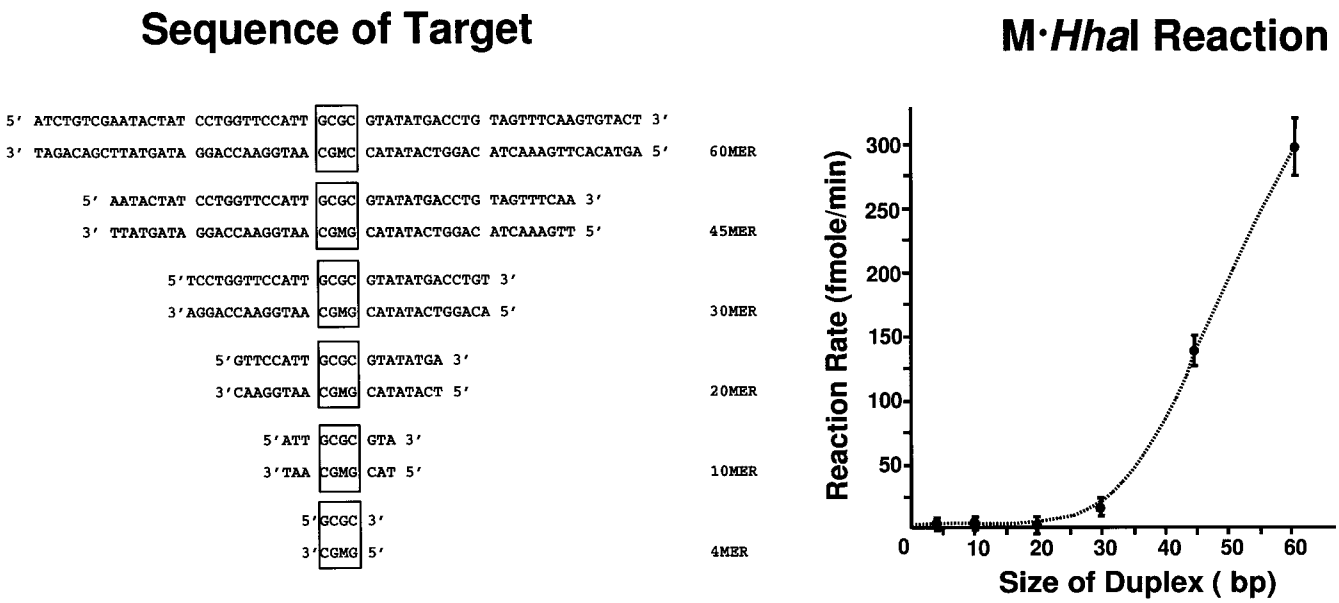

Figure 2. Effect of flanking DNA on the M•*Hha*I reaction rate. An homologous series of oligo-deoxynucleotides was constructed so as to expand the flanking sequence around the central recognition site. Each of the sequences used is listed on the left. The homologous series of oligo-deoxynucleotides, each with a single centrally located *Hha*I recognition site (GCGC/GMGC), was labeled with M•*Hha*I (2U) using 4 mM oligodeoxynucleotide for each length. The reaction rate at each length is plotted in the panel on the right.

stituent slows the methyltransfer step dramatically.[4] The course of the reaction for the 30 mer with and without 5FdC is depicted in Figure 3.

These considerations suggest that the efficiency of formation of covalently bound product and the product yield in the construction of a desired DNA assembly will not only be affected by the usual mass action considerations on the reactants but will also be affected by the spacing between the target sites and the duration of the reaction.[4] In previous work we constructed a doubly substituted 60 mer in which the placement of the recognition sites is expected to orient the N-termini of the M•*Hha*I molecules at about 180° about the *Z*-axis of the DNA helix with center to center spacing of 35 bp between the recognition sites (Figure 4, bottom). This placement was designed to present the two target sites as roughly centered in regions occupying at least 20 bp of DNA. This was intended to take advantage of the more rapid reaction rate observed when the methyltransferase attacks sequences that are greater than 20 bp in length. These same considerations were carried over into the design of a 60 mer having three recognition sites. In this case the positioning of the M•*Hha*I sites was such that the N-termini of the M•*Hha*I molecules are expected to be oriented about the *Z*-axis at intervals of 108°, 108°, and 144° (Figure 4, top).

Experimental detection of the structures formed in these reactions was easily accomplished by gel electrophoretic separation of the reaction products.[4] ^{32}P end-labeled DNA was detected by autoradiography. The gel system separates on the basis of molecular weight. Since the molecular weight is proportional to the reciprocal of the distance traveled in these gels, larger molecules migrate more slowly in the electric field.[6] The observation of discrete bands indicates the presence of species having a discrete molecular weight. Unreacted DNA is also separated on the basis of molecular weight with the 30 mer running ahead of the 60 mer in identically prepared gels. In each case unsubstituted DNA runs most rapidly, with mono-, di-, and trisubstituted molecules ordered by molecular weight. In the data shown in Figure 5 (left) the monosubstituted 30 mer is retarded

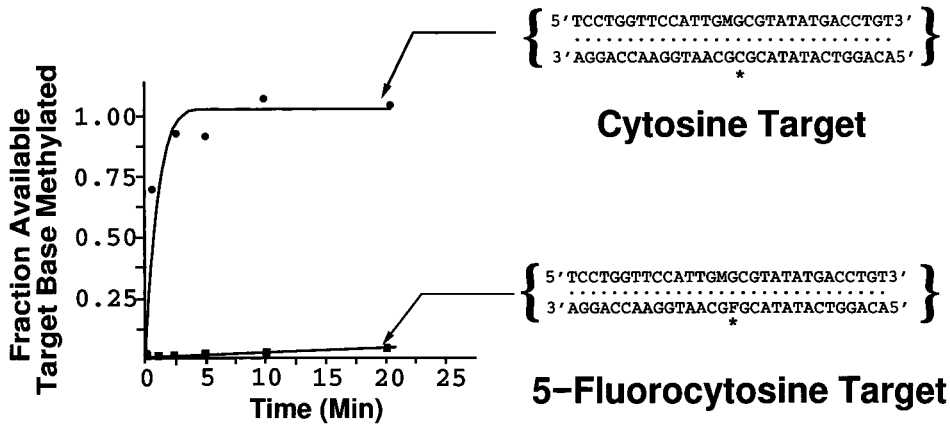

Figure 3. Effect of FdC substitution on the reaction rate. Time course of the reaction rate for M•*Hha*I with FdC (lower curve) or without FdC (upper curve).

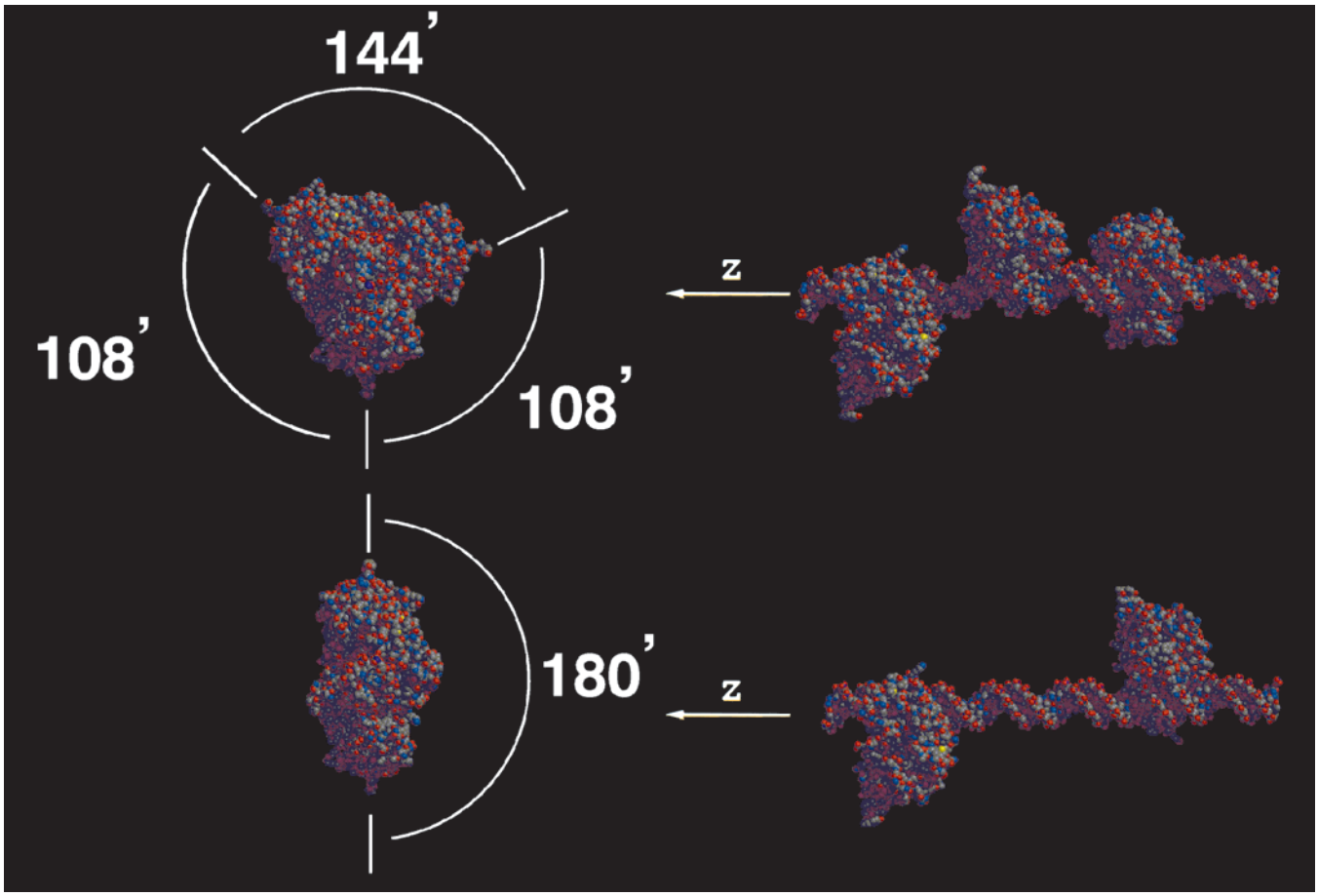

Figure 4. Molecular model of the three-lobed camshaft. Top, left: end view of the three-lobed camshaft (trisubstituted 60 mer). Top, right: side view of the three-lobed camshaft (trisubstituted 60 mer). Bottom, left: end view of the disubstituted 60 mer. Bottom, right: side view of the disubstituted 60 mer. Molecular models were constructed in Biograf 3.21 (MSI, San Diego, CA). The initial conformation of M•*Hha*I was that determined by Klimasauskas et al. for the crystalline protein complexed with 5-fluorocytosine at its target site.[5] The structures were minimized in molecular mechanics to 0.1 (kcal/mol)/Å and rendered in MidasPlus (Computer Graphics Laboratory, UCSF). The model of the doubly substituted system is shown for comparison. Each model assumes a linear DNA molecule with the pitch of 10.0 bp/turn (i.e., a helical twist of 36.0°/bp) derived from fiber diffraction. With this assumption, the N-termini of the three enzymes assume dihedral angles of 108°, 108°, and 144°, respectively, measured down the helical axis, as shown in the model. Although the DNA is not bent by the enzymes, unwinding could result in a twist angle of 31.6°/bp for those base pairs in the binding site. This consideration, coupled with the possibility that the twist angles for the DNA outside the binding sites could be as low as 34.3°/bp, based on solution measurements, suggests that the true dihedral angles could vary considerably.

behind the free DNA. The mono- and disubstituted 60 mer designed for 180° orientation of the N-termini are retarded behind the free 60 mer in the expected positions[4] (Figure 5, center). These products serve as control markers for the construction of the camshaft observed as the slowest band in the rightmost panel of Figure 5. Clearly the spacing chosen

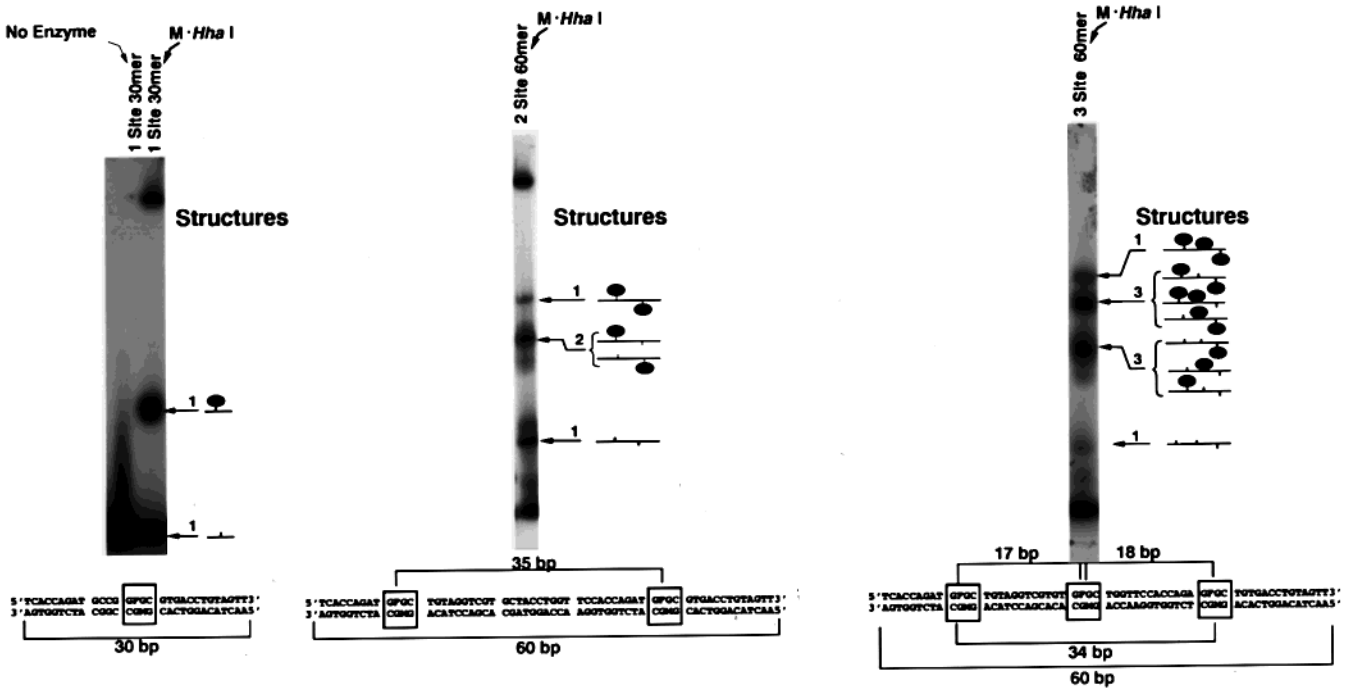

Figure 5. Production of mono-, di-, and trisubstituted DNAs. Discrete gel retardation products observed with autoradiography of ^{32}P end-labeled DNA identify mono-, di-, and trisubstituted DNAs. Substitution products are depicted to the right of the autoradiogram (numbers to the left of each set indicate the number of ways each substitution product can be formed). DNA sequences for the 1, 2, and 3 site oligodeoxynucleotides employed in the constructs are given at the base of each panel.

for the placement of the sites in the camshaft permits the production of an adequate level of the trisubstituted product under the conditions shown in Figure 5; however it is important to point out that increasing the amount of added methyltransferase in an attempt to increase the yield of the product by mass action increased the yield of mono- and disubstituted product but did not increase the yield of trisubstituted product significantly (Figure 6).

As can be seen from Figure 6, the DNA is chased into the monosubstituted form at low concentrations of M•*Hha*I and the monosubstituted form is chased into the disubstituted form as the concentration is increased, but the trisubstituted form does not increase significantly over the concentration range studied here. Two interpretations of the result can be made. One interpretation is suggested by the model of the camshaft in Figure 4 and the data on the effect of flanking DNA depicted in Figure 2. These data suggest that steric hindrance in the design of the trisubstituted 60 mer may favor the accumulation of disubstituted molecules that are substituted at both ends. Thus, the most favored disubstituted form may not be capable of efficiently accepting a third methyltransferase molecule due to steric hindrance in this design. The other interpretation is that much higher protein concentrations are required to chase the substitution level beyond the 3:3:1 ratio for mono-:di-:trisubtituted forms that would be predicted by diffusion-controlled binding without steric hindrance. The data do not distinguish between these possibilities.

These considerations also point out the interesting competition in these designs between the enzymatic requirement for a minimum kinetic footprint that is larger than the structural footprint of the bound enzyme. This effect is

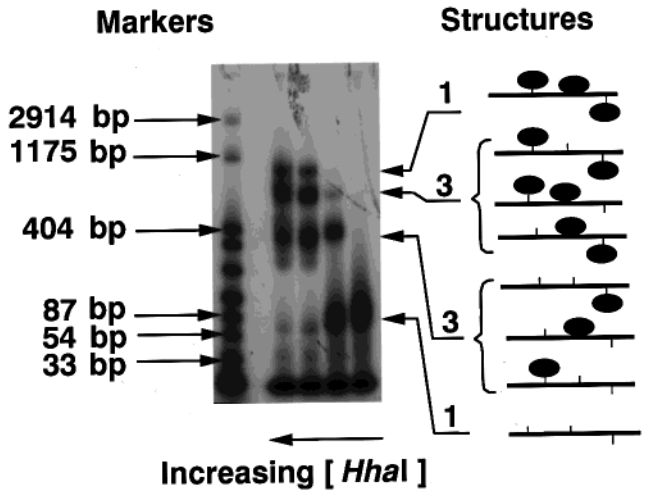

Figure 6. Effect of increasing concentrations of M•*Hha*I on the yield of mono-, di-, and trisubstituted products from the three-site 60 mer. A 50 pmol sample of 60 mer was reacted with 0, 34, 135, or 279 pmol of M•*Hha*I in a 50μL final volume of reaction mixture. Input M•*Hha*I concentration increases from right to left in the figure. Duplex DNA length markers comprise the leftmost lane of the autoradiogram. Mono-, di-, and trisubstituted products are depicted to the right of the autoradiogram (numbers to the left of each set indicate the number of ways each substitution product can be formed).

usually interpreted in terms of the general weak affinity of DNA binding proteins for DNA that does not contain a binding recognition site. This weak affinity hinders diffusion away from a DNA molecule once one has been encountered and thus facilitates a search for the recognition site and enhances the reaction rate. The minimum size of a molecule that facilitates such a search, based on the data given in

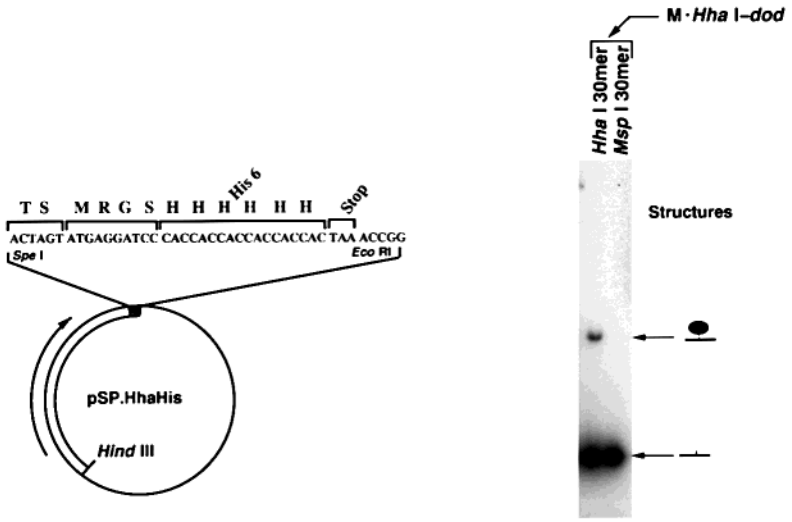

Figure 7. Structure and function of the M•*Hha-dod* fusion. Sequence of the fusion protein plasmid construct is given in the left panel. Gel retardation demonstrates that the fusion protein binds to its target sequence uniquely.[4]

Figure 2, is about 20 bp. Thus, efficient construction of multiply substituted DNAs requires that M•*Hha*I molecules be placed with center to center distances no closer than 6.8 nm. On the other hand the structure of the molecule suggests that low efficiency placements could be accomplished with center to center distances of about 14 bp or about 4.8 nm without physical contact between adjacent molecules (see Figure 4).

In previous work, it was pointed out that the total number of structures and devices that can be approached with this nanoscale addressing system is very large because of the large number of DNA (cytosine-5) methyltransferases that can be utilized in the construction of any one type of assembly superimposed on the large number of possible arrangements in producing scaffolds with precisely located points of attachment. Hogg[7] has proposed that an assessment of the designability of a structure can be made by defining the term G as the desired global structure and L as the collection of components available for assembly of that structure. Interactions between structures yield an internal energy E(G,L) that is a function of the interactions between components.

Under this proposal, the designability of the structure is defined by the count of all component sets that self-assemble to the same global structure:

$$d(G) = |\{L|L \rightarrow G\}|$$

This definition of designability is dependent on the definition of functionality in the concept of global structure. For example, the essential functional properties of a camshaft are 2-fold. The first property is, the linear and axial placement of lobes (cams) on the shaft so that when it is rotated about the axis in the center of the shaft (here the Z-axis of the helix), the peak of each cam appears at the time and place required for the application. The second property, the size of the cam, must be selected for the application so as to provide the cam size (lobe peak height) necessary for the application. Given that the camshaft constructed here is composed of two identical components that were designed to array the peaks of the cams at intervals corresponding to a 108°, 144°, and 108° arrangement around the shaft with cam dimensions depicted in Figure 4, it is highly designable with a designability of $4^{(60-12)} + 1$ or $4^{48} + 1$. This corresponds to the number of DNA duplexes with different sequences outside the 3 tetrameric recognition sites that might form the idealized DNA shaft depicted in Figure 4 and the single methyltransferase that would form the lobes of the camshaft.

For many applications, however, exact lobe height may not be important, or it may be a property that can be best optimized by selection in a directed evolution of the functionality of the component. For these applications a pool of camshafts with identical positioning of different methyltransferases could be constructed from a single reaction mixture containing DNA duplexes each with three target sequences that were identical along one duplex but different between duplexes. When mixed with a collection of appropriate methyltransferases, only methyltransferases of the appropriate type would assemble on a given DNA duplex.

Given the experimental results above, it is possible to produce a variety of nanoscale camshafts approximating a 180° placement around the Z-axis of the helix by placing tetrameric recognition sites at integral multiples of 5 bp along the helix. To accommodate methyltransferases with molecular weights that are larger than the M•*Hha*I example studied here, the available space between molecules should be expanded by placing them at least 25 bp apart. A 75 mer with these spacings could accomplish this. With these parameters, a pool of camshafts with differing lobe dimensions would have a designability of at least $4^{63} + 20$, since the available set of cytosine methyltransferases with unique recognition sequences is more than 20, and there is no limitation on providing matched DNA components for the system.

The construction of targetable fusion proteins was demonstrated in previous work.[4] In that study, the sequence **thr-ser-met-arg-gly-ser-his$_6$** was cloned immediately down-

stream of the *M•Hha*I sequence where the amino acids (Figure 7, left) represent an epitope tag recognized by the ^{MRGS}his antibody.[8] The complex formed between a *Hha*I target-containing oligodeoxynucleotide and the fusion protein could be detected either by retardation of the ^{32}P-end labeled oligodeoxynucleotide (Figure 7, right) or immunologically with western blots using antisera to the fusion peptide.[4] Moreover, when the *Hha*I target sequence was replaced with the *Msp*I target sequence, the M•*Hha-dod* fusion protein did not produce a retardation product (Figure 7, right).

These experiments coupled with the construction of the nanoscale camshaft described above clearly demonstrate that two- and three-address macromolecular assemblies carrying fusion proteins can be produced using the biospecificity of the DNA methyltransferases. With this capacity, it is now possible to consider the construction of materials that self-assemble into two-dimensional and three-dimensional macromolecular arrays. Moreover, simple two-state devices (switches) whose conformation is dependent on the concentration of a peptide ligand or a protein kinase can be easily designed. Accomplishing these goals requires simple extension of the technologies already described. The use of peptide donor−acceptor systems like the Src homology 3 (SH3) domain[9−11] as fusion acceptors and the associated proline-rich peptide ligands as fusion donors might provide a workable system.

The principle here is simply to prepare a donor structure and its cognate acceptor structure independently. Self-assembly of higher order structures would be achieved by mixing the donor with the acceptor in equimolar amounts. Linear two-dimensional assemblies, or in rare cases closed cyclic assemblies, are expected for divalent systems (i.e., those composed of two methyltransferase fusion proteins like that shown in Figure 4, bottom). Three-dimensional assemblies are expected for trivalent and multivalent systems like the camshaft described here (Figure 4 top). Three-dimensional assemblies for such systems could also be based on Y-junction DNA scaffolds,[12] which we have already been able to produce with appropriately placed fluorines.[13] The extension of the technology in these directions must await further experimentation.

Acknowledgments. I thank David Baker and John Wendel for technical assistance. This work was supported in part by grants from the Office of Naval Research (N0014-94-1-1116) and the National Library of Medicine (1G08 LM 006722).

References

(1) Santi, D. V.; Garrett, C. E.; Barr, P. J. *Cell* **1983**, *33* (1), 9−10.
(2) Wendel, J. A.; Smith, S. S. *Nanotechnology* **1998**, *9*, 297−304.
(3) Ivanetich, K. M.; Santi, D. V. *Prog. Nucl. Acids Res. Mol. Biol.* **1992**, *42*, 127−156.
(4) Smith, S. S.; Niu, L.; Baker, D. J.; Wendel, J. A. Kane, S. E.; Joy, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2162−2167.
(5) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357−369.
(6) Smith, S. S.; Gilroy, T. E.; Ferrari, F. A. *Anal. Biochem.* **1983**, *128* (1), 138−151.
(7) Hogg, T. *Nanotechnology* **1999**, *10*, 300−307.
(8) Pogge von Strandmann, E.; Zoidl, C.; Nakhei, H.; Holewa, B.; Pogge von Strandmann, R.; Lorenz, P.; Klein-Hitpass, L.; Ryffel, G. U. *Protein Eng.* **1995**, *8*, 733−735.
(9) Rickles, R. J.; Botfield, M. C.; Zhou, X. M.; Henry, P. A.; Brugge, J. S.; Zoller, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10909−10913.
(10) A Combinatorial Library Approach to Understanding the Sequence Determinants of the SH3 Domain Fold. In *Biological and Biomedical Science and Technology Division 1996 Programs*; Eisenstadt, E., Ed.; U.S. Navy Publication ONR 34196-19; Office of Naval Research: Arlington, VA, 1996; Vol. 1, pp 3−4.
(11) Chen, Y. J.; Lin, S. C.; Tzeng, S. R.; Patel, H. V.; Lyu, P. C.; Cheng, J. W. *Proteins* **1996**, *26*, 465−471.
(12) Seeman, N. C.; Kallenbach, N. R. *Annu. Rev. Biophys. Biomol Struct.* **1994**, *23*, 53−86.
(13) Laayoun, A.; Smith, S. S. *Nucl. Acids. Res.* **1995**, *23*, 1584−1589.

NL005511Y