Biological Significance of Unwinding Capability of Nuclear Matrix–Associating DNAs

J. BODE, Y. KOHWI, L. DICKINSON, T. JOH, D. KLEHR, C. MIELKE, T. KOHWI-SHIGEMATSU*

Matrix attachment regions (MARs) are thought to separate chromatin into topologically constrained loop domains. A MAR located 5′ of the human β-interferon gene becomes stably base-unpaired under superhelical strain, as do the MARs flanking the immunoglobulin heavy chain gene enhancer; in both cases a nucleation site exists for DNA unwinding. Concatemerized oligonucleotides containing the unwinding nucleation site exhibited a strong affinity for the nuclear scaffold and augmented SV40 promoter activity in stable transformants. Mutated concatemerized oligonucleotides resisted unwinding, showed weak affinity for the nuclear scaffold, and did not enhance promoter activity. These results suggest that the DNA feature capable of relieving superhelical strain is important for MAR functions.

EUKARYOTIC CHROMATIN IS ORGANIZED into domains that may affect differential gene expression (*1, 2*). The DNA sequences that tightly associate with the nuclear matrix or scaffold in vitro after a combination of nuclease digestion and extraction have been called matrix or scaffold attachment regions (MARs or SARs) (*2*). MARs are common at the boundaries of transcription units where they may delimit the ends of an active chromatin domain (*3–5*). MARs often are close to or within known enhancerlike regulatory sequences (*2–7*). When stably integrated into genomic DNA, reporter genes flanked by certain MARs show position-independent, copy number–dependent expression (*4, 8*) and augmentation of the transcriptional activity (*4, 8–10*). MAR-like A+T-rich sequences also insulate the regulatory influences of adjacent domains (*11*).

MARs are typically 70% A+T and contain sequences similar to topoisomerase II consensus sequences (*2*). Structural features found in MARs include bending (*12*), a narrow minor groove due to oligo(dA) tracts (*13*), and single-strandedness (*14*). More recently, chemical probing and two-dimensional gel analysis of topoisomers containing the MARs flanking the immunoglobulin heavy chain gene (IgH) enhancer revealed that these MARs (*7*) readily relieve superhelical strain by being stably base-unpaired; the greater the superhelical strain, the more unwinding occurs (*15*). Within the MAR 3′ of the IgH enhancer, an AATATATTT motif remains base-unpaired, even in the presence of 2 mM Mg^{2+} ions, a situation in which the remainder of the plasmid DNA sequences are base-paired. Mutations from *AATATATTT* to ACTGTCTTT or to ACTGCTTT led to a complete inhibition of the extensive base-unpairing over the 200 bp of the MAR (*15*) and to a reduction of the affinity of the MAR for nuclear matrix in vitro (*9*).

An 800-bp DNA fragment located 5′ to the human β-interferon (huIFN-β) gene has properties similar to those of the IgH enhancer; it binds to the nuclear matrix (*15*), it contains sequences that readily become base-unpaired, and the extent of unpairing depends on the ionic strength (Fig. 1). This unpairing was detected by chloroacetaldehyde (CAA), which specifically reacts with unpaired DNA bases (*16, 17*). The same AATATATTT motif remained unpaired in the presence of 2 mM Mg^{2+}, and mutation of that motif resulted in loss of the base-unpaired region.

To test the effect of specific sequence changes on MAR function, we made two synthetic oligonucleotides: 5′-TCTTTAATT-TCT*AATATATTT*AGAA-3′ (wild-type, from the MAR 3′ of the IgH enhancer) and 5′-TCTTTAATTTCT*ACTGCTTT*A-GAA-3′ (three adenines mutated and one thymine deleted). Duplexes of both were multimerized to seven $(25bp)_7$ or eight repeats $(24bp)_8$ (Fig. 2). Supercoiled pUC18 plasmids that contained either wild-type $(25bp)_7$ or mutated $(24bp)_8$ sequences were analyzed with CAA for base-unpaired regions.

The wild-type $(25bp)_7$ was base-unpaired under varying ionic conditions (Fig. 2A). However, the mutated $(24bp)_8$, which is 74% A+T, remained fully base-paired even at low salt concentrations (Fig. 2B). The 100-bp portion of the IgH enhancer 3′ MAR sequence, from which the wild-type oligonucleotide sequence was derived, was 70% A+T, and it unwound. Therefore, in addition to high A+T content, particular sequence features are required for base-un-

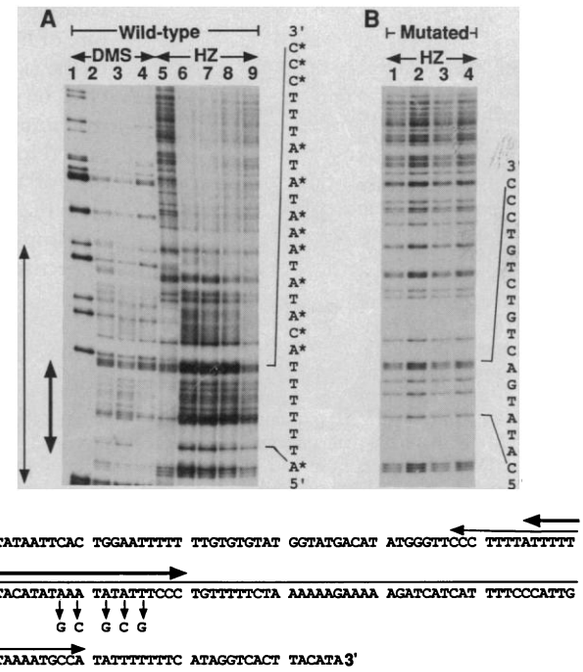

Fig. 1. Determination of the sequence responsible for base-unpairing within the MAR 5′ of the huIFN-β gene. (**A**) Supercoiled plasmid pCL, which contains the 800-bp DNA fragment IV (*9*) located 5′ of the huIFN-β gene that confers a strong binding to the nuclear scaffold, was modified with CAA (*16, 17*) in sodium acetate buffer (pH 5) containing 25 mM Na^+ (lanes 2 and 6), 37.5 mM Na^+ (lanes 3 and 7), 50 mM Na^+ (lanes 4 and 8), or 75 mM Na^+ (lane 9) (*28*). Untreated controls incubated in 25 mM Na^+ are shown in lanes 1 and 5. The Xmn I–Bam HI fragment, 5′ end-labeled at the Xmn I site with [γ-^{32}P]adenosine triphosphate ([γ-^{32}P]ATP) and kinase, was isolated from a native acrylamide gel. We mapped CAA-modified DNA sites by treating the DNA with either hydrazine (HZ) or dimethylsulfate (DMS), then cleaving the DNA with piperidine, and separating it by electrophoresis through an 8% urea-polyacrylamide gel. DNA cleavages not seen in control DNA indicate CAA-modified sites (*17*). (**B**) Supercoiled plasmid pCL, mutagenized at the bases indicated in (C), were analyzed as described above. Lane 1, untreated control DNA; lane 2, CAA-modified DNA at 25 mM Na^+; lane 3, 37.5 mM Na^+; lane 4, 75 mM Na^+. (**C**) Sequence of the huIFN-β MAR. In (A) and (C), regions reactive with CAA at 50 mM Na^+ are indicated by thin arrows, and regions reactive at 75 mM Na^+ are indicated by thick arrows. The five mutated positions are indicated in (C). CAA-modified DNA bases are indicated with (*).

J. Bode, D. Klehr, C. Mielke, Gesellschaft für Biotechnologische Forschung, mbH, Genetik von Eukaryoten, Mascheroder Weg 1, D-3300 Braunschweig-Stöckheim, Federal Republic of Germany.
Y. Kohwi, T. Kohwi-Shigematsu, L. Dickinson, T. Joh, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037.

*To whom correspondence should be addressed.

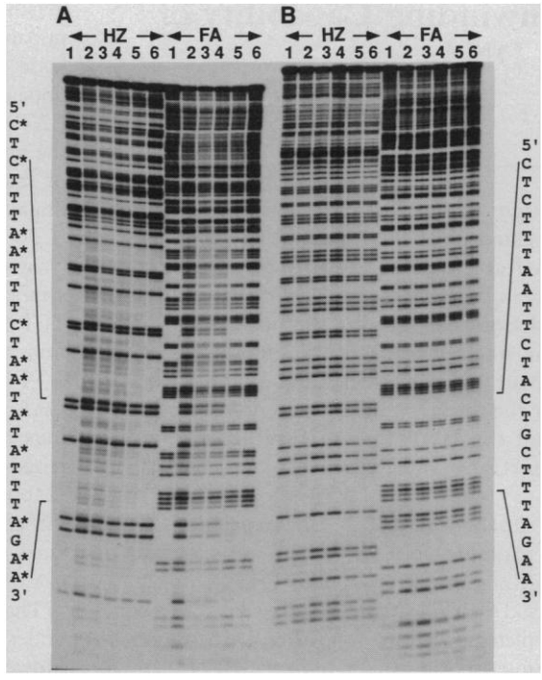

Fig. 2. Base-unpairing property of (**A**) wild-type and (**B**) mutated oligonucleotide multimers. Complementary oligonucleotides 5′-TCTTTAATTTC-TAATATATTTAGAAttc-3′ and 5′-TTCTAAATATATTAGAAATTAAA-GAgaa-3′ that contain the nucleation site of unwinding in the wild-type MAR 3′ of IgH enhancer were hybridized into double-stranded DNA, concatemerized by hybridization through overlapping single-stranded ends, and then digested with mung bean nuclease to remove single-stranded ends. The bases that were added to serve as single-stranded tails are indicated by small letters. Multimers contained one Eco RI site per monomer. The orientation of the sequence was the same throughout the multimer. Multimers were separated on an acrylamide gel, and heptamers and octamers were cloned into the Eco RV site of a bluescript vector. For the preparation of mutated multimers, we used 5′-TCTTTAATTTC-TACTGCTTTAGAAttc-3′ and 5′-TTCTAAAGCAGTAGAAATTAAAG-Agaa-3′ for hybridization. Supercoiled plasmid DNA (Bluescript DNA) with either (A) wild-type $(25bp)_7$ or (B) mutated $(24bp)_8$ sequence was reacted with CAA (*28*). CAA modification was performed in sodium acetate buffer (pH 5) that contained 25 mM Na^+ (lane 2), 50 mM Na^+ (lane 3), 75 mM Na^+ (lane 4), or 25 mM Na^+ plus 2 mM $MgCl_2$ (lane 5). Lanes 1 and 6 show unmodified control DNA. The CAA-modified sites in the Hind III–Bgl I fragment, 3′ end-labeled at the Hind III site, were determined as in Fig. 1. CAA-modified DNA bases are indicated with (*). FA, formic acid.

pairing. The sequence AATATATTT is apparently one such sequence with a strong base-unpairing property if it is located in an appropriate environment (*15*).

We next examined the correlation between nuclear matrix–binding and the unwinding potential of the sequence. A mixture of DNA fragments was allowed to bind to the extracted nuclear scaffold, then separated by centrifugation and agarose gel electrophoresis (Fig. 3). The wild-type $(25bp)_7$ bound to the scaffold to an extent similar to that of the 800-bp MAR fragment "IV" from the huIFN-β gene (on the basis of the densitomeric analysis, 75 and 80% bound, respectively) (Fig. 3). The mutated $(24bp)_8$ bound much less (45%) (Fig. 3), showing that the mutation that eliminates base-unpairing also severely reduces the association with the nuclear scaffold, even if the DNA fragment is rich in A+T and contains the same number of oligo(dT) tracts, a feature implied in certain MAR properties (*13*).

MAR sequences from the huIFN-β gene, from a light-inducible potato gene (ST-LS1), and from the chicken lysozyme gene augment the activity of heterologous promoters in stable transformants (*8–10*). We tested whether the wild-type $(25bp)_7$ sequence augments transcription when inserted 5′ of the SV40 early promoter fused with the luciferase indicator gene (Fig. 4). This DNA construct was transfected into mouse L cells under conditions that allowed, for the most part, integration of one copy per cell (*9*). Independent colonies (100 to 200) resistant to the neomycin analog G-418 were collected, and the luciferase gene activity was measured. The luciferase activity from the construct containing the wild-type $(25bp)_7$ was six times that from the vector construct (Lu) without the wild-type $(25bp)_7$ insert (Fig. 4). The 2.2-kb MAR 5′ of the huIFN-β gene (*9*) conferred an eightfold increase in activity (Fig. 4), the 800-bp core fragment (*9*) of the 2.2-kb MAR gave twofold enhancement (data not shown), and a 150-bp "sub-MAR" fragment (*9*) derived from the 800-bp fragment did not augment transcription (Fig. 4). After stable integration, the mutated $(24bp)_8$ did not augment transcription; rather, it reduced the luciferase gene activity to one-fourth of the original activity. As observed with the ST-LS1 and other MARs (*10*), the wild-type $(25bp)_7$ placed at the 3′ end of the reporter gene also augmented the promoter activity (data not shown). Thus, a synthetic 190-bp wild-type $(25bp)_7$, rich in the nucleation

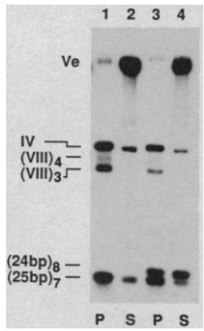

Fig. 3. The wild-type $(25bp)_7$ bound to nuclear scaffold. Wild-type $(25bp)_7$ and mutated $(24bp)_8$ were analyzed in vitro for their affinities for the nuclear scaffold in comparison with known MAR and non-MAR fragments: Ve, vector part (Cla I–Xba I fragment of pBluescript) as a non-MAR control; IV, 800-bp core fragment from the MAR upstream of huIFN-β; $(VIII)_4$ and $(VIII)_3$, multimers of a 150-bp "sub-MAR" fragment that resides in IV [designations for DNA fragments as in (*9*)]. All fragments were labeled with Klenow polymerase and [α-^{32}P]dATP (2′-deoxy 5′-triphosphate). To determine the affinity of these DNA sequences to scaffolds, we used an in vitro reconstitution protocol that selects for DNA fragments with an affinity for the extraction-resistant nuclear scaffold (*9*). A mixture of all fragments without (lanes 1 and 2) or with (lane 3 and 4) mutated $(24bp)_8$ was incubated at 37°C overnight with nuclear scaffold prepared from mouse L cells, then separated into a pellet (P) and supernate (S) fraction by centrifugation (4°C, 14000*g*, 45 min). DNA was then purified and dissolved in 200 μl of 10 mM tris-HCl (pH 7.5) and 1 mM EDTA buffer. DNA was applied (2000 cpm per slot) to a 3% agarose gel. After electrophoresis, the gel was dried and autoradiographed overnight.

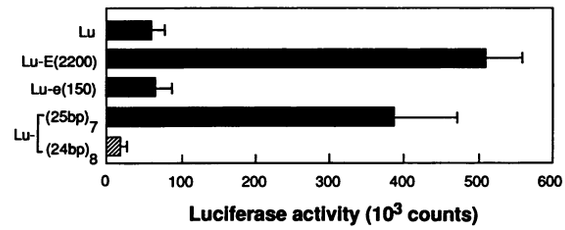

Fig. 4. Transcription-enhancing activity of wild-type $(25bp)_7$ and mutated $(24bp)_8$ in vivo. Plasmid Lu is described in (*10*). The wild-type $(25bp)_7$ or mutated $(24bp)_8$ sequence was excised as a Bam HI–Sal I fragment from a Bluescript vector and cloned into plasmid Lu at the Bam HI and Sal I restriction sites in a position upstream of the SV40 promoter to generate Lu-$(25bp)_7$ and Lu-$(24bp)_8$ plasmids. Plasmid Lu-E(2200) contains the 2.2-kb upstream MAR element of huIFN-β, and Lu-e(150) contains a 150-bp "sub-MAR" fragment; both were inserted upstream of the SV40 promoter (*10*). One microgram of each of the vector plasmids Lu, Lu-E(2200), Lu-$(25bp)_7$, and Lu-$(24bp)_8$ was linearized at the Pvu I site and transfected into mouse L cells as described (*9, 10*). This method gave a low number of integrations per cell and no indication of tandem integration. Stable transformant cells were selected with neomycin analog G-418 (700 μg/ml) for 10 days. The clones were counted on day 12. The cells from 100 to 200 independent colonies were grown close to confluence. Extracts were prepared from a defined number (0.8 × 10^6 to 1.4 × 10^6) of cells from the pools resistant to G-418 and were tested for luciferase activity (*29*). Luciferase fluorescence activity is reported per 10^6 cells and is normalized to a single copy per cell. The SDs of the luciferase activity, derived from several experiments, are shown by error bars.

site for base-unpairing, mimicked MAR properties and function.

The results from the MARs surrounding the IgH enhancer (*15*), the MAR 5′ of the huIFN-β gene, and synthetic oligonucleotides rich in A+T that show contrasting structural properties indicate that the potential for extensive unwinding is required for MAR function. This notion is also supported by the observation that the autonomously replicating sequence (ARS) of the yeast histone H4 gene unwinds under superhelical strain (*18*), and yeast ARS elements are MARs (*19*). The nucleation sites for unwinding vary depending on MARs; the sites did not necessarily contain the AATATATTT motif (*20*).

In prokaryotes, DNA supercoiling is important for various genetic functions (*21*). However, it is unclear whether unconstrained superhelical strain exists at eukaryotic loci. In vitro experiments show that a number of genes are optimally transcribed when supercoiled to various degrees (*22*); for example, formation of the fibroin gene transcriptional initiation complex was greatly facilitated by negative supercoiling (*23*). The transcriptional process generates local negative supercoiling 5′ of a transcribing gene and positive supercoiling 3′ of a transcribing gene (*24*). MARs, together with topoisomerases, may participate in controlling the topological state of chromatin domains. On the basis of the observed tendency of MARs to be base-unpaired, we speculate that negative superhelical strain generated in a chromatin loop domain could be relieved by unwinding of MARs. Thermodynamic energy could then be stored by an association of the unwound MAR with the proteinaceous scaffold and released at other times to generate negative supercoiling at a remote site within the domain. In support of this is the observation that the nuclear scaffold preferentially binds with supercoiled DNA (*25*). MARs could thus support the formation of a transcriptionally active complex, relax the positively supercoiled part of a transcriptional domain, help to accept any histones released from such a region (*26*), aid in the recruitment of topoisomerases (*27*), and prevent superhelical strain from being transmitted to neighboring domains.

REFERENCES AND NOTES

1. D. S. Gross and W. T. Garrard, *Trends Biochem. Sci.* **12**, 293 (1987).
2. S. M. Gasser and U. K. Laemmli, *Trends Genet.* **3**, 16 (1987); L. Phi-Van and W. H. Sträling, *Prog. Mol. Subcell. Biol.* **11**, 1 (1990).
3. J. Mirkovitch, M.-E. Mirault, U. K. Laemmli, *Cell* **39**, 223 (1984); S. M. Gassar and U. K. Laemmli, *ibid.* **46**, 521 (1986); A. Udvardy, E. Maine, P. Schedl, *J. Mol. Biol.* **185**, 341 (1985); B. Levy-Wilson and C. Fortier, *J. Biol. Chem.* **264**, 21196 (1990); J. Bode and K. Maass, *Biochemistry* **27**, 4706 (1988); P. A. Dijkwel and J. L. Hamlin, *Mol. Cell. Biol.* **8**, 5398 (1988); A. P. Jarman and D. R. Higgs, *EMBO J.* **7**, 3337 (1988).
4. L. Phi-Van, J. P. von Kries, W. Ostertag, W. H. Strätling, *Mol. Cell. Biol.* **10**, 2302 (1990).
5. W. H. Strätling, A. Dölle, A. E. Sippel, *Biochemistry* **25**, 495 (1986).
6. P. N. Cockerill and W. T. Garrard, *Cell* **44**, 273 (1986).
7. P. N. Cockerill, M.-H. Yuen, W. T. Garrard, *J. Biol. Chem.* **262**, 5394 (1987).
8. A. Stief, D. M. Winter, W. H. Strätling, A. E. Sippel, *Nature* **341**, 343 (1989).
9. C. Mielke, Y. Kohwi, T. Kohwi-Shigematsu, J. Bode, *Biochemistry* **29**, 7475 (1990).
10. D. Klehr, K. Maass, J. Bode, *ibid.* **30**, 1264 (1991).
11. R. Kellum and P. Schedl, *Cell* **64**, 941 (1991).
12. H. P. Homberger, *Chromosoma* **98**, 99 (1989); J. P. von Kries, L. Phi-Van, S. Diekmann, W. H. Strätling, *Nucleic Acids Res.* **18**, 3881 (1990).
13. Y. Adachi, E. Käs, U. K. Laemmli, *EMBO J.* **8**, 3997 (1989).
14. H. Probst and R. Herzog, *Eur. J. Biochem.* **146**, 167 (1985).
15. T. Kohwi-Shigematsu and Y. Kohwi, *Biochemistry* **29**, 9551 (1990).
16. N. K. Kochetkov, V. N. Shibaev, A. A. Kost, *Tetrahedron Lett.* **22**, 1993 (1971); J. R. Barrio, J. A. Secrist III, N. J. Leonard, *Biochem. Biophys. Res. Commun.* **46**, 597 (1972).
17. Y. Kohwi and T. Kohwi-Shigematsu, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3781 (1988).
18. R. M. Umek and D. Kowalski, *Cell* **52**, 559 (1988).
19. B. B. Amati and S. M. Gasser, *ibid.* **54**, 967 (1988).
20. T. Kohwi-Shigematsu, Y. Kohwi, J. Bode, unpublished results.
21. M. Gellert, *Annu. Rev. Biochem.* **50**, 879 (1981); J. C. Wang, *ibid.* **54**, 665 (1985); J. C. Sinden, *J. Chem. Educ.* **64**, 294 (1987).
22. S. Hirose and Y. Suzuki, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 718 (1988).
23. M. Mizutani, T. Ohta, H. Watanabe, H. Handa, S. Hirose, *ibid.* **88**, 718 (1991).
24. L. F. Liu and J. C. Wang, *ibid.* **84**, 7024 (1987); H.-Y. Wu, S. Shyy, J. C. Wang, L. F. Liu, *Cell* **53**, 433 (1988); G. N. Giaever and J. C. Wang, *ibid.* **55**, 849 (1988).
25. K. Tsutui, K. Tsutui, M. T. Muller, *J. Biol. Chem.* **263**, 7235 (1988); V. Kay and J. Bode, unpublished results.
26. P. Clark and G. Felsenfeld, *EMBO J.* **10**, 387 (1991).
27. D. S. Gilmore, G. Pflugfelder, J. C. Wang, J. T. Lis, *Cell* **44**, 401 (1986).
28. We modified DNA with CAA by incubating 2 μl of twice-distilled CAA in a total of 100 μl of reaction volume at 37°C for 1 hour and then precipitating DNA twice with ethanol.
29. T. M. Williams, J. E. Burlein, S. Ogden, L. J. Kricka, J. A. Kant, *Anal. Biochem.* **176**, 28 (1989).

30. Supported by the NIH (ROI CA39681 to T.K.-S. and ROI CA51377 to Y.K.), by funds provided by the Cigarette and Tobacco Surtax Fund of the State of California through the Tobacco-Related Disease Research Program of the University of California (1KT98 to Y.K.), by an American Cancer Society Faculty Award to T.K.-S., and by a grant from Bundesministerium für Forschung und Technologie to C.M. and J.B.

1 August 1991; accepted 24 October 1991

Site-Specific Incorporation of Novel Backbone Structures into Proteins

JONATHAN A. ELLMAN, DAVID MENDEL, PETER G. SCHULTZ*

A number of unnatural amino acids and amino acid analogs with modified backbone structures were substituted for alanine-82 in T4 lysozyme. Replacements included α,α-disubstituted amino acids, *N*-alkyl amino acids, and lactic acid, an isoelectronic analog of alanine. The effects of these electronic and structural perturbations on the stability of T4 lysozyme were determined. The relatively broad substrate specificity of the *Escherichia coli* protein biosynthetic machinery suggests that a wide range of backbone and side-chain substitutions can be introduced, allowing a more precise definition of the factors affecting protein stability.

SITE-DIRECTED MUTAGENESIS IS A powerful tool for probing the effects of amino acid structure on protein stability and folding. The ability to introduce amino acids with novel backbones and side chains not restricted by the genetic code would likely lead to a more detailed understanding of these phenomena. Substitutions might include amino acid analogs with altered cis-trans rotational barriers, with restricted conformations, or with modified hydrogen-bonding, van der Waals, electrostatic, or covalent interactions. We have recently developed a general biosynthetic method that allows site-specific incorporation of unnatural amino acids into proteins (*1, 2*). We report the application of this method to the preparation and characterization of T4 lysozyme (T4L) mutants containing modifications in the polypeptide backbone. These mutants not only provide an opportunity to assess the effects of backbone structure on protein stability, but they also provide insight into the selectivity of the *Escherichia coli* protein biosynthetic machinery.

T4 lysozyme is a structurally well-characterized protein that has served as a model for a number of mutational studies on protein stability (*3*). To survey the scope of novel backbone structures that can be incorporat-

Department of Chemistry, University of California, Berkeley, CA 94720, and the Center for Advanced Materials, Lawrence Berkeley Laboratory, Berkeley, CA 94720.

*To whom correspondence should be addressed.