typical for closely related avian congeners [J. C. Avise and R. M. Zink, *Auk* **105**, 516 (1988)]. The diagnostic enzymes Eco RV, Bgl II, and Ava I were selected for analysis of all individuals from each population across the region.

10. When DNA from all birds was analyzed with Bgl II, Ava I, and Eco RV, a few showed minor length variations in a single ~5-kb Eco RV band, but the haplotypes of all birds could be clearly related to one or the other parental type.
11. W. S. Moore, J. H. Graham, J. T. Price, *Mol. Biol. Evol.* **8**, 327 (1991); S. D. Ferris *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2290 (1983); N. Barton and J. S. Jones, *Nature* **306**, 317 (1983); J. R. Powell, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 492 (1983).
12. Tail length is not sexually dimorphic in these taxa, but only male birds have beards.
13. Using an approach similar to that of T. W. Quinn and B. N. White [in *Avian Genetics*, F. Cooke and P. A. Buckley, Eds. (Academic Press, London, 1987), pp. 163–198], we cloned restriction fragments of manakin DNA into bacteriophage λ or plasmid vectors and used randomly selected recombinant clones as DNA hybridization probes in Southern blotting to screen for RFLP markers. The λ5–Bgl II probe-enzyme combination detects a 5.6-kb band in *M. vitellinus* that is cleaved into a 3.7- and a 1.9-kb band in *M. candei*; the pSCN3–Ava II combination detects a 5.5-kb band in *M. candei* that is cleaved into a 4.8- and a 0.7-kb band in *M. vitellinus* (both combinations reveal additional bands that are not diagnostic). Heterozygotes were indicated by the presence of all three marker bands.
14. One-way analysis of variance (ANOVA) on sample data from localities 1 through 7 indicates no significant variation for beard length and tail length ($P > 0.5$) [Minitab Data Analysis Software, release 6.1.1, Minitab Inc., 1987]. Fisher's exact test on allele frequencies from localities 1 through 7 indicates no significant variation for pSCN3–Ava II ($P = 0.94$) or λ 5–Bgl II ($P = 0.07$) (SAS Procedures Guide, release 6.07, SAS Institute, Carey, NC, 1991).
15. Any spatially varying pattern within a population could in principle be due to a complex, spatially varying pattern of selection forces or environmental determinants. Given the genetic evidence of hybridization and the lack of any apparent habitat heterogeneity, our data are most simply explained by introgression of golden-collared traits as a result of hybridization.
16. A. Lill, *Z. Tierpsychol.* **36**, 1 (1974).
17. If female *M. candei* prefer golden collars, traits not originally present in males of their species, sexual selection may be based on preexisting sensory or cognitive biases, rather than on coevolution between female preference and the preferred trait [M. J. Ryan, in *Oxford Surveys in Evolutionary Biology*, D. Futuyma and J. Antonovics, Eds. (Oxford Univ. Press, New York, 1990), vol. 7, pp. 157–195; M. J. Ryan and A. S. Rand, *Evolution* **44**, 305 (1990); A. L. Basolo, *Science* **250**, 808 (1990)]. Intrasexual selection could also be based on such preexisting biases.
18. Beards are male secondary sexual characteristics in *Manacus*, yet the longer beard of *M. vitellinus* has not introgressed together with the plumage color traits. This suggests that selection for golden collars can predominate over any selection that may be occurring on beard length, and that the gene (or genes) for beard length are not closely linked to those for plumage color. Alternatively, the lack of introgression may be due to tight linkage of beard length to loci with alleles that are negatively selected in hybrids.
19. That introgression of golden collars stops at the Río Changuinola may be a coincidence or may indicate that the river creates a dispersal barrier over which golden-collared traits have not crossed. Genetic differentiation across wide rivers in South America has been documented in other manakins [A. P. Capparella, *Acta XIX Congressus Internationalis Ornithologici* **2**, 1658 (1988)].
20. Collar width was measured directly. Yellowness of underparts was ranked on a scale from 1 to 5, with reference specimens selected to define the integer points of the scale; a score of 5.0 was defined by an *M. candei* specimen with uniform yellow underparts, 1.0 by an *M. vitellinus* specimen with gray-green underparts (reference series United States National Museum (USNM) accession numbers: 5.0, 608192; 4.0, 608168; 3.0, 608186; 2.0, 606944; 1.0, 13636). Throat color was ranked similarly; the white throats of *M. candei* were assigned a score of 6.0, 5.0 was defined by a locality 3 bird with a lemon-yellow throat, and 1.0 by a locality 10 bird with a golden-orange throat (reference series USNM accession numbers: 6.0, 608192; 5.0, 608158; 4.0, 608167; 3.0, 606915; 2.0, 608146; 1.0, 608136).

21. We thank J. P. Angle, J. A. Blake, R. I. Crombie, J. P. Dean, F. M. Greenwell, C. O. Handley, Jr., E. S. Morton, M. Varn, and D. A. Wiedenfeld for assistance with field work; M. Arroyo, A. Arze, E. Bermingham, G. Maggiori, M. Morello, and staff members of the Smithsonian Tropical Research Institute for logistical support in Panama; the Chiriquí Land Company (C. Forsythe and M. Smith), Petroterminal de Panamá (C. Jurado), and L. Paget (Botel Tomás) for assistance in Bocas del Toro; R. M. Zink and S. J. Hackett for samples from Costa Rica; K. C. Parkes and R. Panza for samples from the Carnegie Museum of Natural History; M. K. Choo for technical assistance; A. Graybeal, G. R. Graves, R. G. Harrison, J. Mariaux, W. K. Moore, E. D. Sattler, J. F. Smith, D. L. Swofford, and E. A. Zimmer for comments; and J. D. Felley and R. W. Jernigan for assistance with statistics and data analysis. Research, collection, and export permits were granted by INRENARE, Republic of Panama. Supported by the Smithsonian Institution Research Opportunities Fund, the Alexander Wetmore Fund, and a Smithsonian Institution Molecular Evolution postdoctoral fellowship (to T.J.P.).

9 September 1992; accepted 5 April 1993

Evidence of DNA Bending in Transcription Complexes Imaged by Scanning Force Microscopy

William A. Rees, Rebecca W. Keller,* James P. Vesenka,† Guoliang Yang, Carlos Bustamante‡

Complexes of *Escherichia coli* RNA polymerase with DNA containing the λ P_L promoter have been deposited on mica and imaged in air with a scanning force microscope. The topographic images reveal the gross spatial relations of the polymerase relative to the DNA template. The DNA appears bent in open promoter complexes containing RNA polymerase bound to the promoter and appears more severely bent in elongation complexes in which RNA polymerase has synthesized a 15-nucleotide transcript. This difference could be related to the conformational changes that accompany the maturation of open promoter complexes into elongation complexes and suggests that formation of the elongation complex involves a considerable modification of the spatial relations between the polymerase and the DNA template.

Scanning force microscopy (SFM) has been used to generate high-resolution images of nucleic acids (*1–7*) and proteins (*8, 9*). Some of these images were obtained under conditions that preserve the state of hydration of the macromolecules, which suggests that SFM could be a powerful tool for investigation of the structure of complex macromolecular assemblies in their native conformation (*10*). This report describes the homogeneous deposition and reliable imaging of complexes of *E. coli* RNA polymerase and a DNA fragment containing the λ P_L promoter. In particular, we investigated the spatial relations between the RNA polymerase and the DNA template in open promoter complexes (OPCs) and in stable elongation complexes with a nascent 15-nucleotide transcript (C15 complexes).

Transcription complexes were prepared by a modification of the method of Levin *et al.* (*11*) and made use of a 681–base pair (bp) DNA fragment as the template (*12*). Complexes were imaged at low humidity in air with a Nanoscope II SFM (Digital Instruments) with tips that were modified with an electron beam (*1, 2, 13*). Samples were prepared in a low-salt buffer containing 5 mM $MgCl_2$ and deposited onto freshly cleaved, previously unmodified mica. The remainder of the deposition procedure was based on a previously published method (*1*). Images were obtained within 30 min of sample preparation.

Unbound DNA, unbound RNA polymerase, and RNA polymerase–DNA complexes can be seen against a relatively flat background in representative fields of complexes (Fig. 1). The DNA templates with RNA polymerase located about four-ninths of the distance from one end (at the λ P_L promoter) were identified as complexes. In images of individual OPCs (Fig. 2A) and C15 complexes (Fig. 2B), the RNA polymerase appears as a slightly asymmetric structure sitting astride the template with the flanking arms of the DNA meeting underneath the polymerase.

Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, OR 97403.

*Present address: Department of Chemistry, University of New Mexico, Albuquerque, NM 87131.
†Present address: Department of Zoology and Genetics, Iowa State University, Ames, IA 50010.
‡To whom correspondence should be addressed.

The ratio of the maximum lateral diameter to the perpendicular lateral diameter of the RNA polymerase was measured in 35 OPCs and 35 C15 complexes, with values (mean ± SD) of 1.24 ± 0.26 in OPCs and 1.25 ± 0.25 in C15 complexes (*14*). These values are consistent with those determined by other techniques for free RNA polymerase (*15, 16*) and RNA polymerase in complexes with DNA (*17*).

Significantly, the template appears systematically bent in the region bound to the polymerase (Fig. 2) (*18*). Studies of gel band shifts of OPCs and elongation complexes on DNA containing the *T7A1* promoter (*19*), as well as OPCs on the P_{gal} promoter (*20*), have been interpreted as evidence for RNA polymerase–induced bending of DNA or for highly flexible regions in the DNA. However, neutron diffraction studies showed no evidence of gross conformational changes in the *T7A1* promoter upon complex formation (*17*). Similarly, earlier studies of RNA polymerase–DNA complexes by electron microscopy (EM) showed that polymerase only slightly increased the propensity of the DNA to bend at the *lac* promoter (*21*).

Williams (*22*) has imaged negatively stained complexes of RNA polymerase and T7 DNA by EM. Although the micrographs show bending of the DNA in some fraction of the complexes, these data were neither discussed nor statistically analyzed. Recently, cryo-electron and regular electron micrographs of OPCs and elongation complexes on plasmid DNA (pUC9) showed the polymerase molecules located at the apices of loops on the supercoiled DNA (*23*). These observations indicate that the RNA polymerase may either bend the supercoiled DNA or increase its flexibility in these complexes.

To investigate whether bending of the template was induced by the binding of RNA polymerase to the λ P_L promoter, we performed a statistical analysis on images of 120 OPCs and 120 C15 complexes obtained from large scanned fields similar to and including those shown in Fig. 1. Lines were drawn through the axes of the DNA molecule on both sides of the polymerase, and the angle of their intersection (Φ) was measured (Fig. 3). The bend angle, θ, is the supplement of the measured angle (θ = 180° − Φ).

The mean bend angle in the preparation of OPCs is 54° (Fig. 4A). This is a significant amount of bending because it occurs over the 70 to 80 bp that RNA polymerase contacts on promoters (*24–28*). The DNA template is significantly more bent in C15 complexes, with a mean bend angle of 92° (Fig. 4B). For comparison, the bend angle distribution was determined at a position

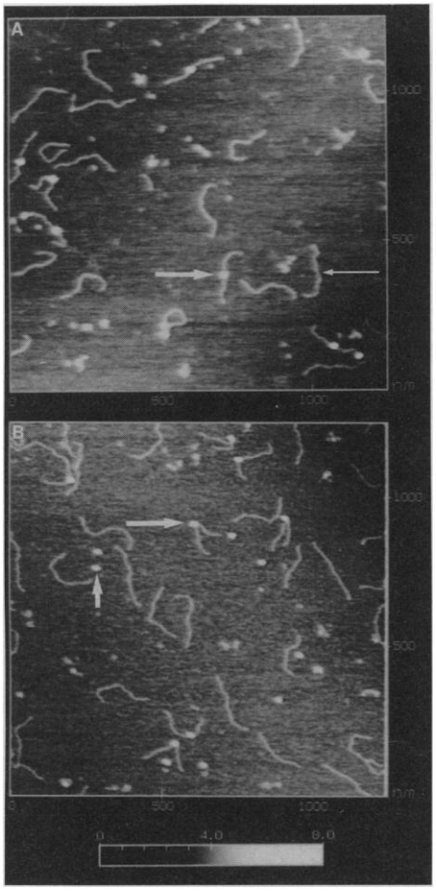

Fig. 1. Images from the SFM of *E. coli* RNA polymerase–DNA complexes. Long thin arrow, free DNA; short thick arrow, free RNA polymerase; and long thick arrows, complexes. (**A**) OPCs on a 681-bp DNA fragment. The start site of transcription from the λ P_L promoter is located 387 bp from the left end of the fragment, and transcription proceeds from left to right (*38*). (**B**) C15 elongation complexes. For details of preparation, see (*12*).

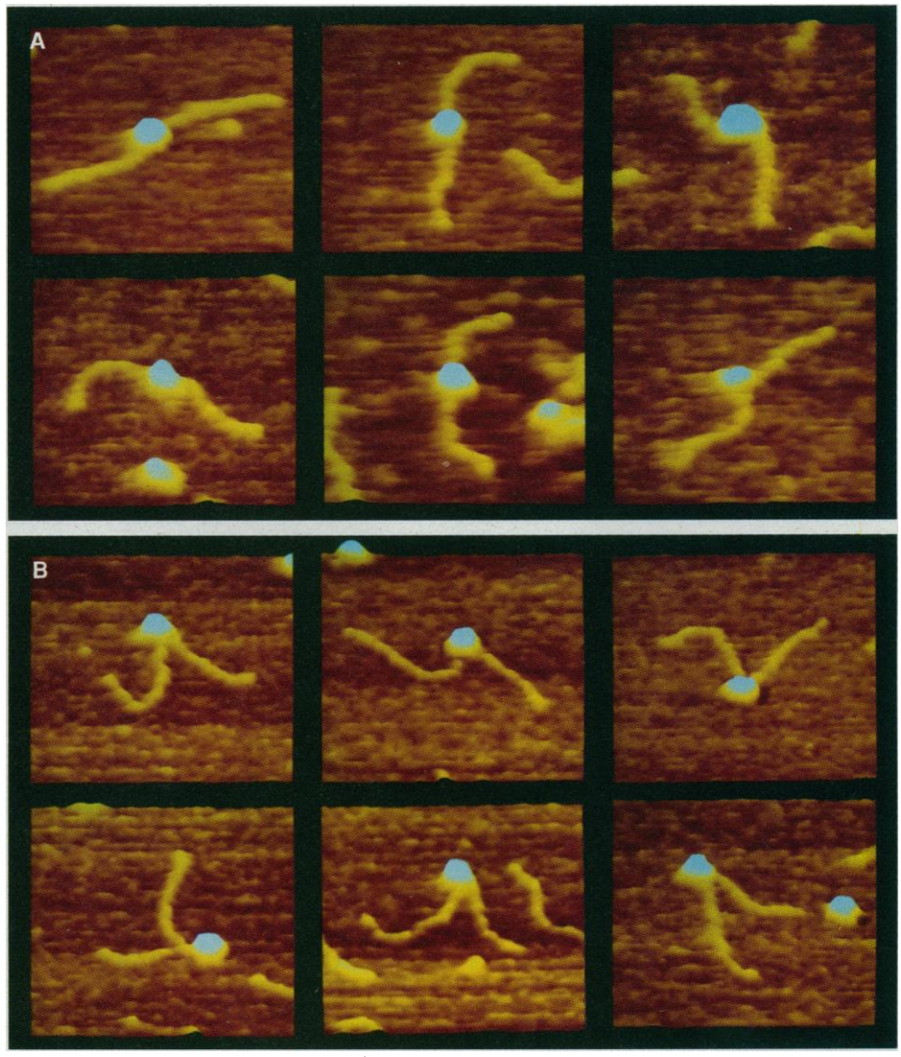

Fig. 2. Surface plots of enlarged images of (**A**) six OPCs and (**B**) six C15 complexes. The plane of the mica is inclined 60° to display the topography of the surface.

four-ninths of the distance from either end of DNA fragments that had been deposited in the absence of RNA polymerase. The distribution is Gaussian and centered at $\theta = 0°$ (Fig. 4C), indicating no intrinsic curvature of the fragments at these locations.

Several potential artifacts were eliminated as possible causes of these results:

1) The observed bend angle might have resulted from the added flexibility conferred to the molecule by the melted region of DNA transcription bubble. Although no unequivocal value is known for the bending rigidity of a melted DNA region, the expected angular distribution attributable to this "hinge" structure, unlike that observed, should be isotropic and thus centered at a mean value of $\theta = 0°$.

2) The observed bending could be accounted for by random thermal fluctuations of the template, which may have regained its double-stranded form upon deposition on the surface, instead of by its interaction with the polymerase. The observed bend angle distributions were compared with the distribution expected for double-stranded DNA by a standard statistical test. This analysis (*29*) indicated that the observed distributions were not a result of random (thermally induced) bending of the template. This is confirmed by the bend angle distribution of the DNA fragments alone (Fig. 4C).

3) The procedures for rinsing and removing excess liquid from the mica surface after deposition could have been responsible for the bent conformations. However, the binding of the complexes to the surface was fast, complete within 2 min after deposition. Moreover, no reorientation of the DNA molecules by the rinsing flow or the N_2 flow, used to remove excess water, was observed.

4) Finally, it is possible that adsorption forces altered the width of the bend angle distribution or shifted the mean of the distribution away from $\theta = 0°$ toward higher angles. To alter the width of the distributions, however, or induce a shift in their means, the surface must interact preferentially with highly bent molecules over those with straight ($\theta = 0°$) configurations. We consider this possibility unlikely in view of the nonspecific, electrostatic nature of the interaction between the complexes and the mica surface. It seems even more unlikely, although not impossible, that the significant difference in mean bend angles observed between OPCs and C15 complexes deposited under identical conditions was caused by a surface-binding selectivity between the two populations. Surface artifacts therefore seem unlikely to account for the data, which should reflect the solution conformations.

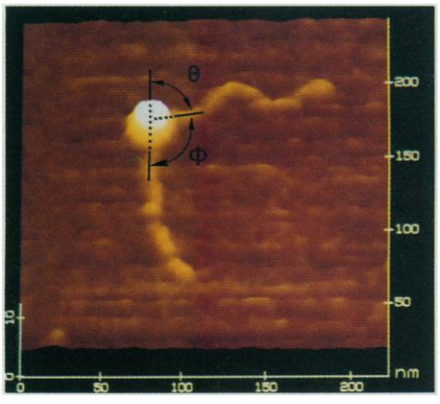

Fig. 3. Enlarged image of an OPC depicting the measured angle Φ and its relation to the bend angle θ.

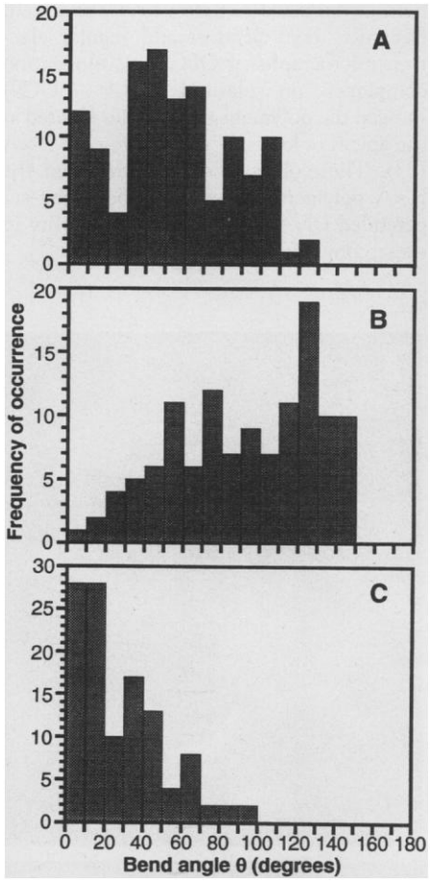

Fig. 4. Histograms of the frequency of occurrence of (**A**) OPCs and (**B**) C15 complexes with various bend angles (120 candidates for each were measured). Only complexes containing an RNA polymerase that was located about four-ninths of the distance from one end of the DNA fragment and was unobstructed by other DNA templates or complexes were chosen. Images revealing more than one polymerase bound in the interior of a template and overlapping complexes were discarded from the statistical analysis. Bend angle (mean ± SD): (A) 54° ± 31° (intermediates in the pathway to OPC formation may be responsible for the multimodal nature of the distribution) and (B) 92° ± 37°. (**C**) Histogram of the frequency of occurrence of bend angles four-ninths of the distance from either end of DNA fragments deposited in the absence of RNA polymerase. A circular mask with the average apparent dimensions of the polymerase was placed at these locations on 57 fragments, and the bend angles were measured.

The maturation of OPCs to stable elongation complexes is accompanied by many changes in the structure and physical properties of the complexes. These include (i) loss of the sigma factor (*28, 30*), (ii) increased resistance to dissociation by salts (*31–34*), and (iii) a drastic reduction in the size of the hydroxyl radical footprint of RNA polymerase on the DNA (*24, 25*). Structural changes in the complex could involve conformational changes in the protein or the DNA, or both. On the basis of band shift data, Heumann and co-workers (*19, 25*) proposed that changes in the bend angle of the template may contribute to structural differences between OPCs and elongation complexes. Alternatively, differential protease sensitivity of the two types of complexes has been interpreted as evidence for a conformational change in the polymerase (*35*).

We speculate that the reduction in the size of the footprint observed in elongation complexes could be as much a result of increased bending of the DNA template as of changes in the conformation of the polymerase. In particular, the significant difference in the mean bend angles between OPCs and C15 complexes suggests that the conformational changes that accompany the maturation of transcription complexes involve a modification in the spatial relation between the polymerase and the template.

Despite present resolution limitations, SFM is rapidly becoming a useful tool for the investigation of the conformation of complex macromolecules. As sharper tips become available, it will be possible, for examples, to use SFM to probe the quaternary structure of the polymerase in transcription complexes. These studies are a step toward imaging macromolecular assemblies under physiological buffers and, perhaps, following their interactions in real time.

REFERENCES AND NOTES

1. C. Bustamante *et al.*, *Biochemistry* **31**, 22 (1992).
2. H. G. Hansma *et al.*, *Science* **256**, 1180 (1992).
3. J. Vesenka *et al.*, *Ultramicroscopy* **42–44**, 1243 (1992).
4. S. M. Lindsay *et al.*, *SPIE* **1639**, 127 (1992).
5. E. Henderson, *Nucleic Acids Res.* **20**, 445 (1992).
6. T. Thundat *et al.*, *Ultramicroscopy* **42**, 1101 (1992).
7. J. Yang *et al.*, *FEBS Lett.* **301**, 173 (1992).
8. H. Arakawa, K. Umemura, A. Ikai, *Nature* **358**, 171 (1992).
9. A. L. Weisenhorn *et al.*, *Langmuir* **7**, 8 (1991).

10. Although it is possible to image macromolecules under aqueous environments (*36*), the yield of stably bound complexes on the mica surface with the use of the current deposition method is insufficient for the statistical analysis described here. This low yield is probably attributable to screening, by the ionic medium, of the electrostatic interactions that hold the complexes to the mica. Although the molecules are not fully bathed in water under the humidity-controlled conditions described here, the macromolecules are likely to retain strongly bound and structurally essential water (*37*). These conditions are then less harsh than the desiccating environments required for the imaging of macromolecules by more traditional electron microscopic methods.
11. J. R. Levin, B. Krummel, M. J. Chamberlin, *J. Mol. Biol.* **196**, 85 (1987).
12. The RNA polymerase holoenzymes (Epicentre Technologies) and DNA fragments (30 nM fRB2·BstU1 DNA fragment) (*38*) were incubated for 10 min at 30°C at an ~2:1 ratio in a buffer containing 20 mM tris-Cl (pH = 7.6), 0.1 mM EDTA, 100 mM NaCl, 0.5 mM dithiothreitol (DTT), and 5% glycerol. The reaction was then diluted ~10-fold in the above buffer without NaCl or glycerol and was incubated on Parafilm at ~37°C for an additional 5 min before being deposited on freshly cleaved mica. Deposited samples were processed and imaged as previously described (*1, 3*). The C15 elongation complexes were formed after incubation of OPCs with 100 mM dinucleotide ApU and 10 mM each of adenosine triphosphate (ATP), CTP, and GTP (final concentrations) (no UTP) for 10 min at 30°C and then incubation with heparin (75 μg/ml) for 10 min before dilution and incubation on Parafilm like the OPCs. Heparin was added to trap the RNA polymerase not involved in the elongation complexes; the highly bent molecules characteristic of C15 complexes were observed in both the presence and the absence of heparin (*39*). Because of the specific sequence of the initial region transcribed from the P_L promoter, the majority (>80%) of the elongation complexes formed by this missing nucleotide procedure contained a 15-nucleotide transcript (C15 complexes) (*11, 38*).
13. D. J. Keller and C. Chih-Chung, *Surf. Sci.* **268**, 333 (1992).
14. Because all lateral dimensions are overestimated by about twice the radius of curvature of the tip in SFM images (*1, 3*), the ratios determined here are underestimates of the actual values. From the apparent width of the DNA molecules (110 ± 15 Å), the average radius of curvature of the tip is estimated to be ~75 Å (*1, 3*). With use of the lateral dimensions of RNA polymerase, obtained by electron diffraction of two-dimensional crystals as an estimate of the actual values (160 Å by 100 Å) (*15*), and use of the above estimate of the radius of curvature of the tip, the ratio of lateral dimensions should be ~(160 Å + 2 × 75 Å)/(100 Å + 2 × 75 Å) = 1.24, similar to the observed values.
15. S. A. Darst, E. W. Kubalek, R. D. Kornberg, *Nature* **340**, 730 (1989).
16. W. Tichelaar *et al.*, *Eur. J. Biochem.* **135**, 263 (1983); O. Meisenberger, H. Heumann, I. Pilz, *FEBS Lett.* **112**, 117 (1980).
17. H. Heumann *et al.*, *J. Mol. Biol.* **201**, 115 (1988).
18. The term "bending" is used here to indicate the deviation of the flanking arms of the template at both sides of the polymerase from the straight configuration, even though all the DNA is not in a double-stranded form in the region that is in contact with the polymerase.
19. H. Heumann, M. Ricchetti, W. Werel, *EMBO J.* **7**, 4379 (1988).
20. G. Kuhnke, H.-J. Fritz, R. Ehring, *ibid.* **6**, 507 (1987).
21. J. Hirsh and R. Schleif, *J. Mol. Biol.* **108**, 471 (1976).
22. R. C. Williams, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2311 (1977).
23. B. ten Heggeler-Bordier, W. Wahli, M. Adrian, A. Stasiak, J. Dubochet, *EMBO J.* **11**, 667 (1992).
24. P. Schickor, W. Metzger, W. Werel, H. Lederer, H. Heumann, *ibid.* **9**, 2215 (1990).
25. W. Metzger, P. Schickor, H. Heumann, *ibid.* **8**, 2745 (1989).
26. A. J. Carpousis and J. D. Gralla, *J. Mol. Biol.* **183**, 165 (1985).
27. A. Spassky, K. Kirkegaard, H. Buc, *Biochemistry* **24**, 2723 (1985).
28. D. C. Straney and D. M. Crothers, *Cell* **43**, 449 (1985).
29. To determine if the observed bend-angle distributions can be accounted for by random thermal fluctuations, we compared the data with a Gaussian probability distribution

$$\mathcal{P}(\theta) = (\pi\ell/P)^{1/2}\exp(-P\theta^2/\ell)$$

where $\mathcal{P}(\theta)$ is the probability of finding a given bend angle θ over a length ℓ and P is the persistence length of double-stranded DNA (*40*). Here, ℓ is assumed to be the length covered by the polymerase as determined from protection studies (ℓ = 80 bp for OPCs and 40 bp for C15 complexes) (*25, 41*). The Gaussian is centered at 0°, as would be expected for a DNA fragment possessing no intrinsic bends or curvature. The DNA within 100 bp of the transcription start site of P_L does not appear to be intrinsically bent (*42*) [also confirmed by the SFM images (Fig. 4C)]. The width of the distribution is determined by the bending rigidity of the DNA template as characterized by its persistence length, P = 210 bp at 25°C (*43*). The Kolmogorov-Smirnov test was used to compare the observed and the theoretical distributions. This test indicated that the observed data are not likely to be drawn from the theoretical distribution (significance = 0.000 for OPC data and <0.000 for C15 data).
30. U. M. Hansen and W. R. McClure, *J. Biol. Chem.* **255**, 9564 (1980).
31. B. Krummel and M. J. Chamberlin, *Biochemistry* **28**, 7829 (1989).
32. D. Rhodes and M. J. Chamberlin, *J. Biol. Chem.* **249**, 6675 (1974).
33. R. Schafer, W. Zillig, K. Zechel, *Eur. J. Biochem.* **33**, 207 (1973).
34. K. Arndt and M. J. Chamberlin, *J. Mol. Biol.* **211**, 79 (1990).
35. R. L. Novak and P. Doty, *J. Biol. Chem.* **243**, 6068 (1968).
36. C. Bustamante, D. J. Keller, G. Yang, *Curr. Opin. Struct. Biol.*, in press.
37. J. T. Edsall, in *The Proteins*, H. Neurath and K. Bailey, Eds. (Academic Press, New York, 1953), vol. 1, part B, p. 549.
38. W. A. Rees, P. H. von Hippel, A. K. Das, unpublished data.
39. W. A. Rees, R. W. Keller, J. P. Vesenka, G. Yang, C. Bustamante, unpublished data.
40. M. D. Barkley and B. H. Zimm, *J. Chem. Phys.* **70**, 2991 (1979).
41. B. Krummel and M. J. Chamberlin, *J. Mol. Biol.* **225**, 239 (1992).
42. H. Giladi, M. Gottesman, A. B. Oppenheim, *ibid.* **213**, 109 (1990).
43. S. B. Smith, L. Finzi, C. Bustamante, *Science* **258**, 1122 (1992).

44. We thank S. E. Weitzel for technical assistance, G. P. Harhay for assistance with data analysis, and P. H. von Hippel, D. Erie, and A. Das for comments on the manuscript. Supported by U.S. Public Health Service (USPHS) research grants GM-32543 (C.B.) and GM-15792 and GM-29158 (P. H. von Hippel), National Science Foundation grant MCB-9118482 (C.B.), and a grant from the Lucille P. Markey Charitable Trust to the Institute of Molecular Biology. W.A.R. was a predoctoral trainee on USPHS institutional training grant GM-07759.

8 January 1993; accepted 13 April 1993

DNA Sequence Determination by Hybridization: A Strategy for Efficient Large-Scale Sequencing

R. Drmanac, S. Drmanac, Z. Strezoska,* T. Paunesku,* I. Labat,* M. Zeremski,* J. Snoddy,† W. K. Funkhouser, B. Koop,‡ L. Hood,§ R. Crkvenjakov‖

The concept of sequencing by hybridization (SBH) makes use of an array of all possible *n*-nucleotide oligomers (*n*-mers) to identify *n*-mers present in an unknown DNA sequence. Computational approaches can then be used to assemble the complete sequence. As a validation of this concept, the sequences of three DNA fragments, 343 base pairs in length, were determined with octamer oligonucleotides. Possible applications of SBH include physical mapping (ordering) of overlapping DNA clones, sequence checking, DNA fingerprinting comparisons of normal and disease-causing genes, and the identification of DNA fragments with particular sequence motifs in complementary DNA and genomic libraries. The SBH techniques may accelerate the mapping and sequencing phases of the human genome project.

The success of the human genome project will depend on whether DNA sequencing approaches can greatly increase throughput [at least 100-fold more than the current value of ~10^4 base pairs (bp) per day per machine] and decrease cost. Strategies that may help to accomplish this task include a greatly improved method of sequencing based on the conventional automated fluorescent DNA sequencers and the advent of new sequencing technologies (*1–3*).

Any linear sequence is an assembly of overlapping, shorter subsequences. Sequencing by hybridization (SBH) (*4–10*) is based on the use of oligonucleotide hybridization to determine the set of constituent subsequences (such as 8-mers) present in a DNA fragment. Unknown DNA samples can be attached to a support