28. U. Hansen and P. A. Sharp, *EMBO J.* **2**, 2293 (1983).
29. B. Wasylyk and P. Chambon, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 921 (1983).
30. R. Tjian, *Cell* **26**, 1 (1981).
31. J. P. McGrath *et al.*, *Nature (London)* **304**, 501 (1983).
32. D. W. Melton, D. S. Konecki, J. Brennand, C. T. Casky, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2147 (1984).
33. G. A. Reynolds *et al.*, *Cell* **38**, 275 (1984).
34. D. Valerio *et al.*, *EMBO J.* **4**, 437 (1985).
35. D. Gidoni, W. S. Dynan, R. Tjian, *Nature (London)* **312**, 409 (1984).
36. G. T. Merlino, J. S. Tyagi, B. deCrombrugghe, I. Pastan, *J. Biol. Chem.* **257**, 7254 (1982).
37. G. Vogeli, H. Ohkubo, M. Sobel, Y. Yamada, I. Pastan, B. deCrombrugghe, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5334 (1981).
38. P. S. Thomas, *ibid.* **77**, 5201 (1980).
39. We thank D. Lowy for providing NIH 3T3 cells transformed by pEJ6.6, J. Evans for manuscript preparation, and R. Steinberg for photography. We are grateful to R. Pozzatti and A. Johnson for reading this manuscript and for offering criticisms.

5 August 1985; accepted 30 October 1985

Sequence of the Immunodominant Epitope for the Surface Protein on Sporozoites of *Plasmodium vivax*

Abstract. Plasmodium vivax *is one of the four malaria parasites that cause disease in humans. The structure of the immunodominant repeating peptide of the circumsporozoite (CS) protein of* P. vivax *was determined. A fragment of* P. vivax *DNA that encodes this tandemly repeating epitope was isolated by use of an oligonucleotide probe whose sequence is thought to be conserved in CS protein genes. DNA sequence analysis of the* P. vivax *clone indicates that the CS repeat is nine amino acids in length (Gly-Asp-Arg-Ala-Asp-Gly-Gln-Pro-Ala). The structure of the repeating region was confirmed with synthetic peptides and monoclonal antibodies directed against* P. vivax *sporozoites. This information should allow synthesis of a vaccine for* P. vivax *that is similar to the one being tested for* P. falciparum.

THOMAS F. MCCUTCHAN
ALTAF A. LAL
VIDAL F. DE LA CRUZ
LOUIS H. MILLER
W. LEE MALOY
Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205
YUPIN CHAROENVIT
RICHARD L. BEAUDOIN
PATRICIA GUERRY
RICHARD WISTAR, JR.
Infectious Diseases Program Center, Naval Medical Research Institute, Bethesda, Maryland 20814-5055
STEPHEN L. HOFFMAN
WAYNE T. HOCKMEYER
Department of Immunology, Walter Reed Army Institute of Research, Washington, D.C. 20307
WILLIAM E. COLLINS
Malaria Branch, Division of Parasitic Diseases, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333
DYANN WIRTH
Department of Tropical Public Health, Harvard School of Public Health, Boston, Massachusetts 02114

Of the four human malarias, those due to the parasites *Plasmodium vivax* and *Plasmodium falciparum* are the most common and are a major cause of disease in the tropics. Controlling these two infections would improve health in this region. One research approach has been the development of vaccines against the parasite at different stages in its life cycle (*1*). Sporozoites, the form inoculated by mosquitoes to initiate the human infection, are covered with a protein known as the circumsporozoite (CS) protein (*2*). Antibodies to the CS protein block infection in vivo (*3*). The gene for the CS protein of *P. falciparum* has been cloned by using an antibody to the CS protein to screen a genomic library constructed with mung bean nuclease–digested fragments (*4*) or a complementary DNA (cDNA) library from sporozoite messenger RNA (*5*). Recombinant (*6*) and synthetic peptide antigens (*7*) of the repeat region in the middle of the CS protein of *P. falciparum* induce antibodies that block sporozoite invasion of liver cells in vitro. These antigens will be used for vaccine testing in man. We now report the sequence of the immunodominant repeat region of the CS protein of *P. vivax*.

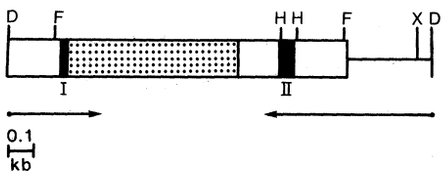

Fig. 1. General structure and partial restriction map of the CS protein gene of *P. vivax*. DNA sequence analysis (arrows) established the orientation of the gene (box) (the amino-terminus is to the left, the carboxyl-terminus to the right) and yielded the nucleotide sequence of the repeating epitope as shown in Fig. 2. Regions I and II are indicated by black boxes. The repeat region is indicated by stippling. Restriction enzyme sites indicated: D, Dra I; F, Fok I; H, Hinf I; X, Xba I.

Our strategy for cloning the gene for the CS protein of *P. vivax* with an oligonucleotide probe from a region of nucleotide homology between *P. falciparum* and *P. knowlesi* (region II) (*4*) was based on the observation (*8*) that *P. vivax* is more closely related to *P. knowlesi* than to *P. falciparum*. Therefore, if a sequence is conserved in *P. knowlesi* and *P. falciparum*, it should also be conserved in *P. vivax*. A 27-base oligonucleotide probe to region II (*9*) hybridized to a single fragment 1.3 kilobases in length of Dra I–digested genomic DNA (*10*) of an El Salvador strain (Sal-I) of *P. vivax* (*11*). The same probe was then used to screen a Dra I genomic library cloned in the bacterial plasmid pUC9.

The general structure of the CS protein gene from *P. vivax* as deduced from the pPv1 insert is shown in Fig. 1. The nucleotide sequence of a segment of the gene containing the repeats of the CS protein of *P. vivax* is shown in Fig. 2. It contains tandem repeats of nine amino acids (Gly-Asp-Arg-Ala-Asp-Gly-Gln-Pro-Ala). The peptide on the amino-terminal side of the repeat region has an overall structure similar to those of the CS proteins of *P. knowlesi* (*12*) and *P. falciparum* (*4*), including a charged region and the homologous region I (Fig. 3). Because of the high degree of homology between *P. vivax* and *P. knowlesi* (see below), we predicted that the 5′ coding region of the gene missing from our Dra I clone would encode the start codon for translation, a signal sequence, and a few amino acids between the signal sequence and the Dra I recognition sequence.

Monoclonal antibodies against sporozoites of *P. knowlesi*, *P. falciparum*, and *P. cynomolgi* reacted with tandemly repeated peptides in the CS proteins (*4, 5, 12, 13*). Therefore, to prove that the sequence was indeed the CS protein of *P. vivax*, we investigated how monoclonal antibodies to *P. vivax* sporozoites would react to a synthetic peptide of two tandem repeats (*14*) in the deduced *P. vivax* CS peptide. The peptide (at 100 μg/ml) blocked the binding of two different horseradish peroxidase–labeled monoclonal antibodies (219c and 427) to *P. vivax* sporozoites in an enzyme-linked immunosorbent assay. As a control for specificity, we determined the effect of the repeat peptide (four repeats) of the CS protein of *P. falciparum* on the binding of these two antibodies to *P. vivax* sporozoites. The *P. falciparum* peptide did not block the binding of antibody to *P. vivax* sporozoites. Furthermore, antibody 219c reacted with the *P. vivax* peptide in a two-sided immunoassay

(*15*), thus indicating the presence of two epitopes within two tandem repeats of the peptide.

The regions of tandemly repeated amino acids of CS proteins are immunodominant (*16*) and one target for vaccine development (*6, 7*). The amino acid sequence of the repeats for *P. vivax*, Gly-Asp-Arg-Ala-Asp-Gly-Gln-Pro-Ala, differs from those for (i) *P. falciparum*, Asn-Ala-Asn-Pro and Asn-Val-Asp-Pro (*4, 5*); (ii) the H strain of *P. knowlesi*, Gly-Gln-Pro-Gln-Ala-Gln-Gly-Asp-Gly-Ala-Asn-Ala (*12*); (iii) the Nuri strain of *P. knowlesi*, Glu-Gln-Pro-Ala-Ala-Gly-Ala-Gly/Arg-Gly (*17*); and (iv) the Gombak strain of *P. cynomolgi*, Asp/Gly-Gly-Ala-Ala-Ala-Ala-Gly-Gly-Gly-Gly-Asn (*13*).

The amino acid sequence at the amino-terminal end of the region of repeats was highly conserved in *P. knowlesi* (*12*) and *P. vivax* (Fig. 2). Sequence conservation was maintained except for the charged region on the amino-terminal side of the repeats, and even in this region both proteins retained charged amino acids. The region of high sequence homology between *P. vivax* and *P. falciparum* was limited to region I (Fig. 3). Thus, the CS protein of *P. vivax*, a human malaria, is more like that of a monkey malaria, *P. knowlesi*, than that of another human malaria, *P. falciparum*. This finding is consistent with our earlier report that *P. vivax* is more closely related evolutionarily to *P. knowlesi* than to *P. falciparum* (*8*). We found that three gene probes (chicken actin, *Chlamydomonas* tubulin, and mouse dihydrofolate reductase) that hybridized to the DNA's of *P. knowlesi*, two other monkey malarias, and the human malaria *P. vivax* did not hybridize to the DNA of *P. falciparum*. Another probe, yeast thymidylate synthetase, which hybridized to the DNA of *P. falciparum*, did not hybridize to the DNA's of *P. knowlesi* or *P. cynomolgi*, a *P. vivax*-like parasite of monkeys.

The sequence of the gene for the CS protein of *P. vivax* will facilitate development of a vaccine against the parasite at this stage. The repeat region is the obvious first target (*6, 7*). The finding (*17*) that there is no homology between the repeat regions of two strains of *P. knowlesi* from Malaysia isolated 7 years apart raises the possibility that a vaccine prepared against the repeat region may select mutant parasites with immunologically unrelated sequences. However, monoclonal antibodies to the repeating epitopes in *P. falciparum* and *P. vivax* all react with the few isolates studied to date (*18*). Furthermore, the structure of the gene for the *P. falciparum* CS protein from 17 other Asian, African, Central American, and South American isolates has been analyzed by nucleic acid hybridization with a probe from the repeat region and found to be conserved in the strains examined (*19*). Nonetheless, if mutant repeats exist in the human parasite populations, an important consideration is whether a vaccine should also be directed against the conserved regions, such as region I (*4, 20*). Another consideration is whether the efficacy of a sporozoite vaccine may be extended by combining it with a vaccine directed toward other stages of the parasite's life cycle.

```
                    *           *           *           *           *           *
          AAT GGA GTA AAC TTC AAT AAT GTA GAC GCC AGT TCA CTT GGC GCG GCA CAC GTA GGA CAA
P. vivax  Asn Gly Val Asn Phe Asn Asn Val Asp Ala Ser Ser Leu Gly Ala Ala His Val Gly Gln

          AAT GGA GTA AGC TTC AAT AAT GTA GAC ACC AGT TCA CTT GGC GCA CAG CAG GTG AGA CAA
P. knowlesi Asn Gly Val Ser Phe Asn Asn Val Asp Thr Ser Ser Leu Gly Ala Gln Gln Val Arg Gln

                    *           *           *           *           *           *
          AGT GCT AGC CGA GGC AGA GGA CTT GGT GAG AAC CCA GAT GAC GAG GAA GGA GAT GCT AAA
P. vivax  Ser Ala Ser Arg Gly Arg Gly Leu Gly Glu Asn Pro Asp Asp Glu Glu Gly Asp Ala Lys

          AGT GCT AGC CGA GGC AGA GGA CTT GGT GAG AAG CCA AAA GAA GGA GCT GAT AAA GAA AAG
P. knowlesi Ser Ala Ser Arg Gly Arg Gly Leu Gly Glu Lys Pro Lys Glu Gly Ala Asp Lys Glu Lys

                    *           *           *           ¦<----------Region I--
          AAA AAA AAG GAT GGA AAG --- AAA --- GCA GAA CCA AAA AAT CCA CGT GAA AAT AAG CTG
P. vivax  Lys Lys Lys Asp Gly Lys --- Lys --- Ala Glu Pro Lys Asn Pro Arg Glu Asn Lys Leu

          AAA AAA GAA AAA GGA AAA GAA AAA GAA GAA GAA CCA AAG AAG CCA AAT GAA AAT AAG CTG
P. knowlesi Lys Lys Glu Lys Gly Lys Glu Lys Glu Glu Glu Pro Lys Lys Pro Asn Glu Asn Lys Leu

          ---------->¦        *           *           *           *           *
          AAA CAA CCA GGA GAC AGA GCA GAT GGA CAG CCA GCA GGA GAC AGA GCA GAT GGA CAG CCA
P. vivax  Lys Gln Pro Gly Asp Arg Ala Asp Gly Gln Pro Ala Gly Asp Arg Ala Asp Gly Gln Pro
                    ¦-----------------------------------¦ ¦-----------------------------

          AAA CAA CCG AAT GAA GGA CAA CCA CAA GCA CAG GGT GAT GGA GCA AAT GCA GGA CAA CCA
P. knowlesi Lys Gln Pro Asn Glu Gly Gln Pro Gln Ala Gln Gly Asp Gly Ala Asn Ala Gly Gln Pro
                    ¦-------------------------------------------¦ ¦----------

                *           *           *           *           *           *
          GCA GGA GAC AGA GCA GAT GGA CAG CCA GCA GGT GAT AGA GCA GAT GGA CAA CCA GCA GGA
P. vivax  Ala Gly Asp Arg Ala Asp Gly Gln Pro Ala Gly Asp Arg Ala Asp Gly Gln Pro Ala Gly
          --¦ ¦-----------------------------------¦ ¦-----------------------------------¦ ¦--

          CAA GCA CAA GGA GAT GGA GCA AAT GCA GGA CAA CCA CAA GCA CAG GGT GAT GGA GCA AAT
P. knowlesi Gln Ala Gln Gly Asp Gly Ala Asn Ala Gly Gln Pro Gln Ala Gln Gly Asp Gly Ala Asn
          ----------------------------------¦ ¦-----------------------------------------
```

Fig. 2. The nucleotide sequence of the immunodominant region of the CS protein of *P. vivax*. A partial DNA sequence of the pPv1 insert was determined by the dideoxy procedure (*21*) as shown in Fig. 1. The deduced amino acid sequence of the CS protein is given beneath the nucleotide sequence. The amino acid sequence of the comparable region of the *P. knowlesi* (H) CS protein gene from amino acid 35 to amino acid 134 (*12*) is also given for comparison. Dashes interrupting the *P. vivax* sequences represent gaps in comparison to the *P. knowlesi* sequence. The bounds of region I are indicated. The repeated sequences are underlined.

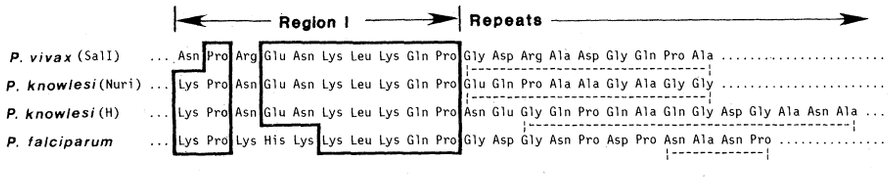

Fig. 3. Region of sequence homology among the CS proteins of *P. vivax*, *P. falciparum*, and *P. knowlesi* (H and Nuri strains). A region of amino acid homology between the CS proteins of *P. knowlesi* and *P. falciparum* has been reported and labeled region I (*4*). The extent of amino acid sequence homology delineating region I was later reduced (*17*) as shown. Amino acid sequence homologies are boxed. Repeated sequences are underlined.

References and Notes

1. L. H. Miller, P. H. David, T. J. Hadley, *Phil. Trans. R. Soc. London Ser.* B **307**, 99 (1984).
2. R. S. Nussenzweig and V. Nussenzweig, *ibid.*, p. 117.
3. P. Potocnjak, N. Yoshida, R. S. Nussenzweig, V. Nussenzweig, *J. Exp. Med.* **151**, 1504 (1980).
4. J. B. Dame *et al.*, *Science* **225**, 593 (1984).
5. V. Enea *et al.*, *ibid.*, p. 628.
6. J. F. Young *et al.*, *ibid.* **228**, 958 (1985).
7. W. R. Ballou *et al.*, *ibid.*, p. 996 (1985); F. Zavala *et al.*, *ibid.*, p. 1437.
8. T. F. McCutchan, J. B. Dame, L. H. Miller, J. Barnwell, *ibid.* **225**, 808 (1984).
9. The following four oligonucleotide probes from region II common to *P. knowlesi* and *P. falciparum* (*4*) were used: CCATG$^{T}_{C}$AGTGTAAC$^{T}_{C}$T-GTGGAAATGGT.
10. *Plasmodium vivax* DNA (*11*) was digested in the presence of Dra I restriction endonuclease, and

fragments, 800 to 1300 base pairs in length, were isolated after agarose gel electrophoresis. Fragments were ligated into the plasmid pUC9. Recombinant plasmids were transfected into the *Escherichia coli* strain JM83. Colonies were transferred to nitrocellulose and probed with radiolabeled oligonucleotide (*9*). The filters were hybridized with the probe at 42°C for 18 hours in 4× Denhardt's (0.08 percent polyvinyl pyrolidone, 0.08 percent Ficoll and 0.08 percent bovine serum albumin), 2× standard saline citrate (SSC), and 0.1 percent sodium dodecyl sulfate (SDS). The filters were washed with 0.5× SSC and 0.2 percent SDS at 37°C. A positive colony (pPv1) was detected by autoradiography.

11. W. E. Collins, P. G. Contacos, W. A. Krotoski, W. A. Howard, *J. Parasitol.* **58**, 332 (1972). The Sal-I strain of *P. vivax* was isolated from an infected human in the Cangrejera area of La Paz, El Salvador. The parasite was adapted to grow in *Aotus lemurinus griseimembra*. Parasite DNA was isolated from infected blood as described by J. B. Dame and T. F. McCutchan, *Mol. Biochem. Parasitol.* **8**, 263 (1983).
12. L. S. Ozaki, P. Svec, R. S. Nussenzweig, V. Nussenzweig, *Cell* **34**, 815 (1983).
13. V. Enea *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7520 (1984).
14. The peptide GQPAGDRADGQPAGDRAD (designated with the one-letter notation for amino acids) was prepared by the solid-phase method of peptide synthesis [R. B. Merrifield and A. Marglin, *Annu. Rev. Biochem.* **39**, 841 (1970)] as described in (*4*).
15. Monoclonal antibodies 219c and 427 were derived from Balb/c mice immunized with sporozoites of the ONG strain of *P. vivax* from Vietnam and react with *P. vivax* sporozoites from North Korea, Thailand, and Colombia. Antibody 219c gave a circumsporozoite precipitin test with *P. vivax* sporozoites and identified proteins of 46 to 57 kD in an immunoblot assay with SDS-solubilized *P. vivax* sporozoites. Immulon II plates (Dynatech) were coated with purified 219c. Serial tenfold dilutions of synthetic peptide (*14*) from 100 μg/ml to 10 ng/ml were incubated with the antibodies. After overnight incubation, the plates were washed and incubated with horseradish peroxidase-labeled antibody 219c for 2 hours. Substrate was then added to determine the amount of enzyme-linked antibody bound. There was a strongly positive reaction down to 100 ng/ml of peptide with the monoclonal antibody to *P. vivax*, 219c. As a control for specificity, a monoclonal antibody to the *P. falciparum* sporozoite was coated on the plate. No *P. vivax* synthetic peptide bound to the *P. falciparum* antibody.
16. F. Zavala, A. H. Cochrane, E. H. Nardin, R. S. Nussenzweig, V. Nussenzweig, *J. Exp. Med.* **157**, 1947 (1983).
17. S. Sharma, P. Svec, G. H. Mitchell, G. N. Godson, *Science* **229**, 779 (1985).
18. F. Zavala, A. Masuda, R. S. Nussenzweig, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **43**, 1808 (1984).
19. J. L. Weber and W. T. Hockmeyer, *Mol. Biochem. Parasitol.* **15**, 305 (1985).
20. U. Vergara, A. Ruiz, A. Ferreira, R. S. Nussenzweig, V. Nussenzweig, *J. Immunol.* **134**, 3445 (1985).
21. F. Sanger, S. Nicklen, A. R. Coulsen, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
22. Supported by a Rockefeller Foundation fellowship (to A.A.L.) and in part by the Naval Medical Research and Development Command, work unit 3M463750D808AD061. We thank P. David, K. Mendis, and J. Dame for providing *P. vivax* DNA for preliminary experiments.

22 August 1985; accepted 21 October 1985

Systematic Variation in DNA Length Yields Highly Ordered Repressor-Operator Cocrystals

Abstract. *Crystals have been grown that contain the operator-binding domain of the λ repressor and the λ operator site O_L1. Crystallization conditions were tested with a set of DNA fragments, ranging in length from 17 to 23 base pairs. The best crystals were grown with a 20–base pair DNA fragment. These crystals have space-group symmetry $P2_1$, with unit cell dimensions* a = *37.1 Å*, b = *68.8 Å*, c = *56.8 Å*, *and a* β *angle of 91.5°. They diffracted to at least 2.5 Å resolution. High resolution data from these crystals should allow the direct determination of how a repressor recognizes its operator site.*

STEVEN R. JORDAN
TIMOTHY V. WHITCOMBE
JEREMY M. BERG
CARL O. PABO
Department of Biophysics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

The crystal structures of four gene-regulatory proteins—the lambda cro protein, the NH_2-terminal domain of the λ repressor, and the CAP protein and the *trp* repressor of *Escherichia coli*—have been reported (*1–4*). Each of these proteins binds to its operator site(s) as a dimer and forms a complex that is approximately twofold symmetric. Each protein has a conserved helix-turn-helix motif (*5, 6*). Model-building studies suggest that the second helix of this bihelical unit fits into the major groove of right-handed B-form DNA and makes site-specific contacts with the base pairs (*7–9*). Genetic and biochemical studies are generally consistent with the proposed models (*10*). In the case of λ repressor, mutations that disrupt binding to DNA have been identified at each of the residues that is presumed to make contacts in the major groove (*11, 12*). These models are also supported by analysis, at 7-Å resolution, of cocrystals that contain the NH_2-terminal domain of a related phage repressor, the 434 repressor, and its 14–base pair operator site (*13*). Further work on these crystals is in progress (*13*), but the crystals do not appear to be well ordered at high resolution (*14*).

It is necessary to have cocrystals that diffract to high resolution in order to understand the detailed molecular basis of protein-DNA recognition and to detect any conformational changes that may occur as the complexes form. Such cocrystals have been reported for the Eco RI restriction enzyme and its recognition site, and preliminary analysis has revealed conformational changes in the DNA (*15*). We attempted to crystallize a λ operator site with the NH_2-terminal fragment (residues 1–92) of λ repressor. The protein fragment was produced by cleaving the repressor with papain and was purified by methods similar to those reported previously (*16, 17*). The first DNA used for cocrystallization was a 17–base pair duplex containing the sequence found in the operator site O_L1 (Fig. 1). In spite of a thorough search of crystallization conditions, we were unable to grow useful cocrystals with this DNA fragment, and we began to systematically test the effect of DNA length on crystallization.

DNA fragments often stack when DNA (*18*) or protein-DNA complexes (*14, 15*) are crystallized, and, consequently, we expected that the length of the DNA would be a critical variable in the search for cocrystals. When DNA fragments stack to form a continuous helix, changing the length of the DNA alters the distance between neighboring complexes along the helix and also changes the azimuthal relationship of neighboring protein molecules around the double-helical axis. Such changes should have major effects on the crystallization. To continue our search for suitable cocrystals, we synthesized the DNA fragments shown in Fig. 1. Each contains the 17–base pair λ operator site O_L1, which is the highest affinity site for repressor binding (*19*). However, the length of the fragments and, concomitantly, the nature of the ends were varied. Most of the fragments had one or two unpaired bases on each end, and these were chosen to allow base-pairing with a neighboring fragment in the cocrystals. We expected that base-pairing between the ends might accentuate the natural tendency of the DNA to stack and might provide very stable contacts between the ends of the helices.

Crystallization conditions were screened by the hanging drop vapor diffusion method (*20*), and we used DNA that had been purified by gel electrophoresis (*21*). Each duplex was tested with polyethylene glycol (PEG) 400 and PEG 4000, with organic solvents, with several polycations, and with concentrated salt solutions. Crystals were obtained under widely different conditions and with a number of the DNA fragments. Three fragments gave crystals large enough for us to take diffraction photographs: cocrystals with the 18-nucleotide-long fragment (18-mer) were grown by adding 10 m*M* $CaCl_2$, cocrystals with the 20-mer were grown from 20 percent PEG 400, and cocrystals with the 22-mer were grown by adding 15 m*M* $Co(NH_3)_6Cl_3$.