11. S. Matsushita, F. Ibuki, A. Aoki, *Arch. Biochem. Biophys.* **102**, 446 (1963).
12. C. E. Wright, H. H. Tallan, Y. Y. Lin, *Annu. Rev. Biochem.* **55**, 427 (1986).
13. K. P. M. Heirwegh and S. B. Brown, Eds., *Bilirubin* (CRC Press, Boca Raton, FL, 1982), vols. 1 and 2; R. Schmid and A. F. McDonagh, in *The Metabolic Basis of Inherited Disease*, J. B. Stanbury, J. B. Wyngaarden, D. S. Fredrickson, Eds. (McGraw-Hill, New York, 1978), pp. 1221–1257.
14. M. Sarnat, in *Topics in Neonatal Neurology*, H. B. Sarnat, Ed. (Grune & Stratton, New York, 1984), pp. 109–136; G. B. Odell and H. S. Schutta, in *Cerebral Energy Metabolism and Metabolic Encephalopathy*, D. W. McCandless, Ed. (Plenum, New York, 1985), pp. 229–261.
15. K. Bernhard, G. Ritzel, K. U. Steiner, *Helv. Chim. Acta* **37**, 306 (1954); H. Beer and K. Bernhard, *Chimia* **13**, 291 (1959).
16. Y. Yamamoto, E. Niki, Y. Kamiya, *Bull. Chem. Soc. Jpn.* **55**, 1548 (1982).
17. Different solvents were used since biliverdin is not soluble in chloroform and bilirubin is insoluble in methanol. The solubility of oxygen in chloroform is only 20% greater than that in methanol (*25*).
18. F. R. Mayo, *Acc. Chem. Res.* **1**, 193 (1968).
19. A. F. McDonagh, L. A. Palma, F. R. Trull, D. A. Lightner, *J. Am. Chem. Soc.* **104**, 6865 (1982); A. F. McDonagh, L. A. Palma, D. A. Lightner, *ibid.*, p. 6867.
20. D. A. Lightner and A. F. McDonagh, *Acc. Chem. Res.* **17**, 417 (1984); A. F. McDonagh and D. A. Lightner, *Pediatrics* **75**, 443 (1985).
21. A. F. McDonagh, *N. Engl. J. Med.* **314**, 121 (1986).
22. G. W. Burton and K. U. Ingold, *J. Am. Chem. Soc.* **103**, 6472 (1981).
23. N. A. Porter *et al.*, *J. Am. Chem. Soc.* **102**, 5597 (1980); L. R. C. Barclay and K. U. Ingold, *ibid.* **103**, 6478 (1981); Y. Yamamoto *et al.*, *Biochim. Biophys. Acta* **795**, 332 (1984).
24. The phase transition temperature of soybean PC bilayer is below 0°C [J. De Gier, J. G. Mandersloot, L. L. M. Van Deenen, *Biochim. Biophys. Acta* **150**, 666 (1968)]. Therefore, we believe that there is no significant difference in the liposome structure between 37° and 50°C.
25. Chemical Society of Japan, Ed., *Kagaku Binran* (Maruzen, Tokyo, 1975), p. 769.
26. S. Srivastava, R. S. Phadke, G. Govil, C. N. R. Rao, *Biochim. Biophys. Acta* **734**, 353 (1983); A. N. Erin *et al.*, *ibid.* **774**, 96 (1984); A. N. Erin, V. K. Skrypin, V. E. Kagan, *ibid.* **815**, 209 (1985).
27. E. Colleran and P. O'Carra, in *Chemistry and Physiology of Bile Pigments*, P. D. Berk and N. I. Berlin, Eds. (Government Printing Office, Washington, DC, 1977), pp. 69–80.
28. J. A. T. P. Meuwissen and K. P. M. Heirwegh, in *Transport by Proteins*, G. Blauer and H. Sund, Eds. (de Gruyter, Berlin, 1978), pp. 387–401.
29. Y. Yamamoto, M. H. Brodsky, J. C. Baker, B. N. Ames, *Anal. Biochem.*, in press.
30. A. F. McDonagh and F. Assisi, *Biochem. J.* **129**, 797 (1972).
31. The observed lag in Fig. 1D is thought to reflect the gradual equilibration of the nitrogen atmosphere in the reaction tubes with air over the course of the experiment.
32. The impetus for this study was to provide a test of a general hypothesis put forward by one of us (A.N.G.) that end products of degradative metabolic pathways may play important roles as protective agents, and, that in this context, bilirubin is a likely candidate for such a role. This work was supported by National Cancer Institute Outstanding Investigator grant CA 39910 (B.N.A.), NIEHS (National Institute of Environmental Health Sciences) Center grant ES 01896 (B.N.A.), NIH grants AM 26307, MD 20551 (A.F.M.), and GM 28994 (A.N.G.), NSF grant DMB 85-18066 (A.N.G.), and University of California Toxic Substances Research and Teaching Program (Y.Y.). We also thank H. Hurd for help on earlier aspects of this study.

22 September 1986; accepted 6 January 1987

Construction of a General Human Chromosome Jumping Library, with Application to Cystic Fibrosis

FRANCIS S. COLLINS, MITCHELL L. DRUMM, JEFFERY L. COLE, WENDY K. LOCKWOOD, GEORGE F. VANDE WOUDE, MICHAEL C. IANNUZZI

In many genetic disorders, the responsible gene and its protein product are unknown. The technique known as "reverse genetics," in which chromosomal map positions and genetically linked DNA markers are used to identify and clone such genes, is complicated by the fact that the molecular distances from the closest DNA markers to the gene itself are often too large to traverse by standard cloning techniques. To address this situation, a general human chromosome jumping library was constructed that allows the cloning of DNA sequences approximately 100 kilobases away from any starting point in genomic DNA. As an illustration of its usefulness, this library was searched for a jumping clone, starting at the *met* oncogene, which is a marker tightly linked to the cystic fibrosis gene that is located on human chromosome 7. Mapping of the new genomic fragment by pulsed field gel electrophoresis confirmed that it resides on chromosome 7 within 240 kilobases downstream of the *met* gene. The use of chromosome jumping should now be applicable to any genetic locus for which a closely linked DNA marker is available.

THE USE OF LINKAGE ANALYSIS (*1*) and high-resolution cytogenetics has allowed chromosomal mapping in an increasing number of disorders of single human genes, even in situations where gene function is unknown. Recent examples include Huntington disease (*2*), adult polycystic kidney disease (*3*), cystic fibrosis (*4–7*), chronic granulomatous disease (*8*), Duchenne muscular dystrophy (*9*), and familial retinoblastoma (*10*). In the last three of these disorders mapping has led directly to the cloning of a candidate gene, raising hopes that this "reverse genetics" (*11*) approach may be widely successful for cloning disease-associated genes for which no protein product is known. In general, however, the closest DNA markers to a disease gene are often hundreds of kilobases away. In cystic fibrosis (CF), for example, the gene has been localized to the long arm of chromosome 7 by the finding that it is closely linked to two molecular probes, the *met* oncogene (*6*) and the DNA fragment pJ3.11 (*7*), both of which are placed at less than 0.5 centimorgan from the CF gene by current linkage estimates (*12*). Because, on the average, 1 centimorgan is represented by about 1000 kilobases (*13*), cloning the CF locus or any other disease locus by reverse genetics is likely to require crossing hundreds of kilobases. To address the general problem of cloning over such large distances, we have recently described a method of "chromosome jumping" (*13, 14*). In a model system we showed that this technique could be used to cross a distance of 45 kb (*14*). Lehrach and co-workers have independently described a similar scheme (*15*). To demonstrate the practical application of our technique, we have constructed a general jumping library and derived from it a cloned sequence of DNA that lies approximately 100 kb downstream from the *met* oncogene.

A human chromosome jumping library was constructed (*16*) from human lymphoblastoid cell DNA (see Fig. 1). In this instance, the distance of the jumps represented in the library (the "hopsize") was chosen to be approximately 100 kb by size selection of genomic DNA molecules in the size range of 80 to 130 kb. The principle of the technique, as previously outlined (*13–15*), depends on the formation of large genomic circles from size-selected DNA, bringing together the genomic fragments that were originally 100 kb apart. Note that a light partial digestion with the restriction enzyme Mbo I, which cuts frequently in genomic DNA, is performed prior to size selection, so that there is no significant bias for a particular sequence to occur at the junction of the circles. In each clone, the position of the joining of the two genomic fragments is marked by a suppressor transfer RNA or supF gene (*17*), which allows selection for these fragments after the circles are digested with Eco RI and ligated into a

F. S. Collins, Departments of Internal Medicine and Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109.
M. L. Drumm, Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109.
J. L. Cole, W. K. Lockwood, M. C. Iannuzzi, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109.
G. F. Vande Woude, National Cancer Institute–Frederick Cancer Research Facility, Frederick, MD 21701.

phage vector. The efficiency of each step must be high, since only the DNA immediately adjacent to the supF gene will be cloned. However, we found it possible to generate a library of approximately 2 million independent jumping clones from 4 μg of 100-kb size-selected DNA.

This type of library should contain sequences from the entire genome. A different variety of jumping library, generated by circularizing complete genomic digests with a restriction enzyme that cuts rarely (such as Not I), has also been proposed (*13, 15*). In such a library the DNA is not size-selected and many fewer clones are required to represent all possible junction fragments. It has the disadvantage, however, of not being usable unless one is starting at one of these rare restriction sites.

One million clones from our library were screened with the *met*G probe (*18–20*), which is a 2.2-kb Pst I genomic fragment near the 3′ end of the *met* oncogene (*19*). Four positive λ phage were plaque-purified, DNA was prepared, and the phage inserts were restriction-mapped. All four of the positive clones contained part or all of the *met*G sequence, one or two supF genes, and another fragment of DNA. One clone had a jumping fragment too small to be useful (<100 bp). Two other jumping fragments contained repetitive sequences. The fourth, denoted CF63 (Fig. 1), consisted of 0.7 kb of single-copy sequence and was chosen for further study.

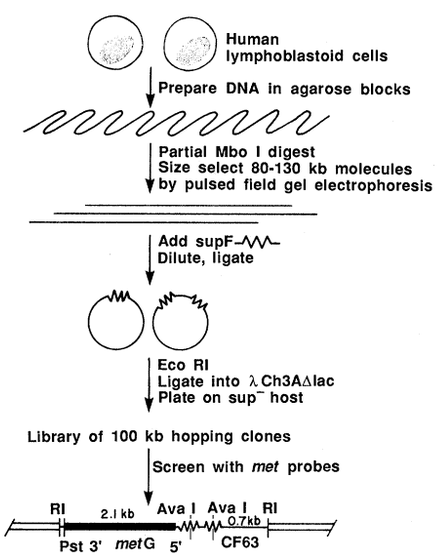

Fig. 1. Scheme for chromosome jumping. Circularization of genomic DNA is used to generate a jumping library (*16*). Each phage clone includes two genomic fragments that were originally located at opposite ends of an approximately 100-kb genomic fragment (*14*). One of the four clones obtained by screening 1 million phage from the primary library with *met*G is shown at the bottom. The heavy bar represents the *met*G sequence, the sawtooth mark represents the two supF genes with their internal Ava I sites that facilitate restriction mapping and subcloning, and the open bars are the phage arms. The 0.7-kb genomic fragment denoted CF63 should be located approximately 100 kb downstream from the 3′ end of *met*G.

By noting the orientation of the *met*G sequence in this clone (Fig. 1) and considering the topology of circle formation, we concluded that CF63 was located downstream to the 3′ end of the *met* transcription unit. Since the DNA size selection in preparing the library was rather broad, we could not specify the exact distance of CF63 from *met*G, but it should be between 80 and 130 kb. A major concern about the jumping approach is that if tandem ligation of unrelated DNA molecules occurs along with circle formation, this will give rise to junction fragments in which the two genomic sequences are completely unrelated. Although the ligation was performed at low DNA concentration to reduce this possibility, any damage to the DNA ends would have increased the likelihood of meaningless fragments. We therefore mapped the CF63 sequence using Southern blotting with a somatic cell hybrid that contains human chromosome 7 on a mouse background. CF63 was found to map to human chromosome 7, with no hybridization observed from mouse DNA (Fig. 2). Characterization of this jumping library with other probes indicated that the majority of clones have the desired properties; for example, of ten that have been mapped, nine have both the starting point and the end point of the jump on the same chromosome (*21*).

To further establish the validity of CF63, we used pulsed field gel electrophoresis (PFGE) techniques (*13, 22–27*). PFGE is capable of separating DNA molecules up to 2000 kb in size, and thus can be used in conjunction with Southern blotting to map relatively large regions of the genome. Figure 3A shows restriction digests of human DNA samples with the enzymes Not I and Sfi I. Each enzyme generates a wide distribution of DNA fragments from less than 50 kb to well above 1000 kb. The DNA fragments were transferred to a nylon membrane and hybridized, first with the CF63 probe (Fig. 3B), and subsequently (after stripping of the CF63 probe) with *met*G (Fig. 3C). The sizes of the hybridizing bands can be estimated by comparison to the λ phage size markers. For Not I digests probed with *met*G, there is a 550-kb band in lane 2, and a much larger fragment (>800 kb, in the nonresolving region of the gel) in lanes 1 to 4. This variation apparently represents a Not I restriction fragment length polymorphism (RFLP), and has been noted by others with this probe (*28*). The possibility that this is a partial digest or a methylation difference has not been excluded, but probing the Not I lanes with an unrelated probe did not result in partial digest patterns.

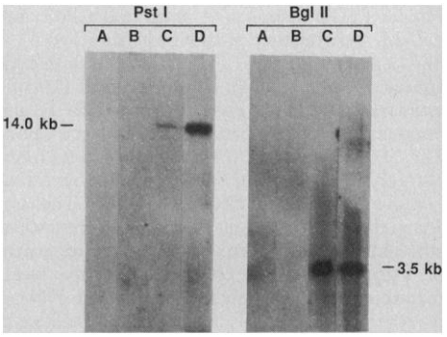

Fig. 2. Mapping of CF63 to chromosome 7. Southern blots digested with the enzymes Pst I and Bgl II are shown. In each panel, lane A is mouse DNA; lane B, a somatic cell hybrid containing parts of human chromosomes 4 and 5; lane C, a hybrid containing only human chromosome 7 on a tetraploid mouse background (HDm20s); and lane D, human leukocyte DNA. Equal amounts (6 μg) of DNA were loaded in each lane except for lane C in the Bgl II digest, where 25 μg of DNA was loaded in order to bring out the human chromosome 7 signal in this hybrid. Genomic DNA was digested with restriction enzymes according to the supplier's (New England Biolabs) recommendations and then subjected to electrophoresis in a 1% agarose gel, with Hind III–cut λ phage DNA being used as a size marker. DNA was transferred to GeneScreen Plus (New England Nuclear) and hybridized according to the manufacturer's recommendations. Probes were labeled by random hexanucleotide priming (*38*), and final washing of the blot was in 0.1× standard saline citrate (SSC) at 68°C for 15 minutes.

The Not I pattern detected by CF63 was the same as that detected by *met*G, both in the size of the fragments and the presence of two fragments in lane 2. This provides strong evidence that CF63 and *met*G reside within 550 kb of each other in the human genome. The Sfi I patterns, however, were different for the two probes: *met*G detected a 160-kb band and CF63 detected a 40-kb band. Restriction digests of *met* genomic clones (*18–20*) indicate that the next Sfi I site 3′ to *met*G is at least 20 kb away. Furthermore, from the two partial digests in lanes 6 and 7 it does not appear that the 160-kb and 40-kb Sfi I fragments are immediately adjacent, since one would expect to see a partial digest band of 200 kb with both probes if this were so. In fact, the major bands in lanes 6 and 7 show a doublet of approximately 20-kb spacing with both probes, suggesting that at least one additional Sfi I fragment of about 20 kb may lie in between *met*G and CF63. It is not possible to be certain of a complete Sfi I map with these data, however.

For the Southern blot with BssH II shown in Fig. 4, we used the field inversion

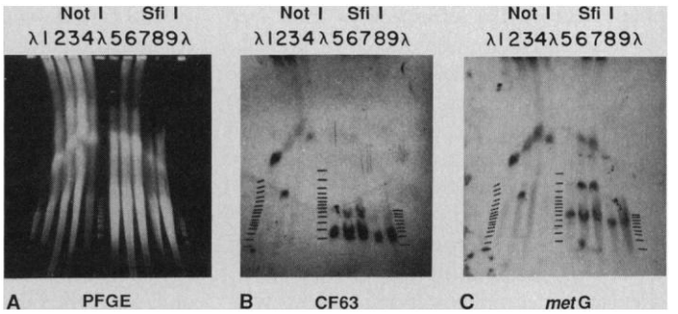

Fig. 3. PFGE mapping of *met*G and CF63. Samples of DNA from four humans were digested with Not I (lanes 1 to 4), and from five humans with Sfi I (lanes 5 to 9). (**A**) PFGE gel stained with ethidium bromide. A ladder of annealed λ phage DNA (λ) constructed from a 45-kb phage was used to estimate sizes. These bands are drawn onto (B) and (C) for clarity. (**B** and **C**) Southern blots of the bands in (A) with the CF63 (B) and *met*G (C) probes. The CF63 probe is the 0.7-kb Ava I–Eco RI fragment shown in Fig. 1. This also contains part of the supF gene, but hybridization with supF alone does not give detectable Southern blot signals. The larger Not I band in lanes 1 to 4 is in the part of the gel (>800 kb) where resolution breaks down, as in the ethidium-stained pattern. Lanes 6 and 7 are partial Sfi I digests. Sizes of the Sfi I fragments can be estimated from this blot, but more exact estimates require reducing the amount of genomic DNA per lane, because sizes are overestimated by PFGE if more than 2 to 3 μg per lane is used. The DNA was prepared in LMT agarose (*16, 24*), with 2×10^6 cells (13 μg of DNA) being used per 80-μl block. Digestions were performed for 5 hours on half-blocks (6.5 μg), in a volume of 300 μl; the enzyme (80 units, New England Biolabs) was used in the recommended buffer at a temperature of 37°C (Not I) or 50°C (Sfi I). The PFGE apparatus was constructed according to the OFAGE modification described by Carle and Olson (*23*), and electrophoresis was carried out at 14°C and 280 V for 20 hours, with a pulse time of 60 seconds. The DNA was then nicked by exposure to 254 nm of ultraviolet light for 2 minutes, transferred in 0.4*N* NaOH to GeneScreen Plus (New England Nuclear), and hybridized according to the manufacturer's recommendations. Probes were labeled by random hexanucleotide priming (*38*). The final washing of the blots was in 0.1× SSC at 68°C for 15 minutes. Autoradiography was at −70°C against Kodak XAR-5 film with two intensifying screens for 1 to 4 days.

gel electrophoresis (FIGE) technique (*26*). The hybridization of *met*G and CF63 to the same 240-kb BssH II fragment further confirms the validity of the jumping clone and indicates it cannot be further away from *met*G than 240 kb. The additional identity of the partial digest pattern in Fig. 4 is helpful in excluding the possibility of coincidental hybridization to BssH II fragments of similar size.

This mapping analysis indicates that the jumping fragment CF63 does in fact reside approximately 50 to 240 kb 3′ to the *met* gene, as expected on the basis of the plan of construction of this general human jumping library. Present evidence indicates that this is probably the correct direction to travel from *met* to reach the CF gene: linkage data favor the gene order *met*–CF–pJ3.11 (*12*), and chromosome-mediated gene transfer hybrids (*29*), prepared from a chromosome that has undergone a translocation within the *met* gene, suggest that pJ3.11 is on the 3′ side of *met*. This evidence is indirect, however, and the translocation may be a more complex rearrangement (*30*). By locating RFLPs with CF63 it will be possible to test individuals with recombinations between *met* and CF. If any of these are nonrecombinant between CF63 and CF, this would indicate definitively that CF is on the 3′ side of *met*.

Although we have not identified the exact region of the CF gene, this study demonstrates the feasibility of constructing a general human chromosome jumping library to locate such a gene. The preparation of these libraries is arduous, but once constructed they can be amplified and used repeatedly to jump from any starting point in the human genome. To cross a distance of 100 kb would require on the average five steps in a cosmid chromosome walk. Other creative schemes have been proposed to expedite this labor-intensive walking (*31, 32*), but the procedures can be slowed or halted at any point by recombinogenic sequences (*33*). Jumping allows one to hop over such troublesome sequences; however, the DNA fragments obtained with the procedure described here are small and may need to be expanded to obtain sufficient surrounding DNA to search for a transcript or to repeat the jumping process. In many situations, jumps even larger than the present 100-kb are desirable, and larger distances can be achieved without any fundamental changes in the procedure (*13, 14*). Thus chromosome jumping should be useful in the ongoing effort to clone genes for human diseases by reverse genetics.

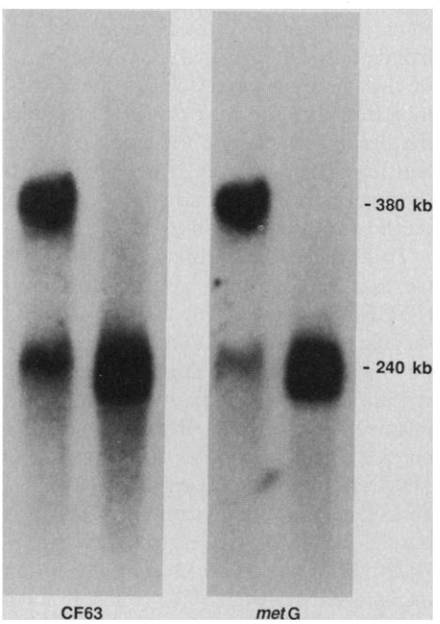

Fig. 4. Mapping of *met*G and CF63 to the same BssH II fragment. Partial (lane 1) and complete (lane 2) BssH II digests of human genomic DNA were subjected to electrophoresis in a field inversion gel apparatus (*26*), transferred to a nylon membrane, and probed with CF63. The blot was then stripped and reprobed with *met*G. A 240-kb BssH II fragment is seen with both probes in the complete digest, and the partial digest patterns are identical. DNA preparation, electrophoresis, transfer, and hybridization were as in the legend to Fig. 3 except that 3 μg of DNA per lane was used and electrophoresis was in a field inversion apparatus (*26*). The forward time interval was 9 to 60 seconds, and the backward interval was 3 to 20 seconds, ramped linearly over a 22-hour run.

Note added in proof: Poustka *et al.* (*34*) have just reported the construction of a human Not I jumping library.

REFERENCES AND NOTES

1. D. Botstein, R. L. White, M. Skolnick, R. W. Davis, *Am. J. Hum. Genet.* **32**, 314 (1980).
2. J. Gusella *et al.*, *Nature (London)* **306**, 234 (1983).
3. S. T. Reeders *et al.*, *ibid.* **317**, 542 (1985).
4. L.-C. Tsui *et al.*, *Science* **230**, 1054 (1985).
5. R. G. Knowlton *et al.*, *Nature (London)* **318**, 380 (1985).
6. R. White *et al.*, *ibid.*, p. 382.
7. B. J. Wainwright *et al.*, *ibid.*, p. 384.
8. B. Royer-Pokora *et al.*, *ibid.* **322**, 32 (1986).
9. A. P. Monaco *et al.*, *ibid.* **323**, 646 (1986).
10. S. H. Friend *et al.*, *ibid.*, p. 643.
11. S. H. Orkin, *Cell* **47**, 845 (1986); F. Ruddle, *Am. J. Hum. Genet.* **36**, 944 (1984); J. L. Marx, *Science* **228**, 1080 (1985); P. Little, *Nature (London)* **321**, 558 (1986).
12. A. Beaudet *et al.*, *Am. J. Hum. Genet.* **39**, 681 (1986).
13. C. L. Smith *et al.*, *Methods Enzymol.*, in press.
14. F. S. Collins and S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6812 (1984).
15. A. Poustka and H. Lehrach, *Trends Genet.* **2**, 174 (1986).
16. The jumping library was constructed as follows. Human lymphoblasts transformed with Epstein-Barr virus (cell line 3.1.0) were grown to 4×10^7 cells and embedded in low melting temperature (LMT) agarose by mixing equal volumes of 2% LMT agarose in 125 m*M* EDTA with cells suspended in phosphate-buffered saline. The mixture was then poured into a 2 by 8 mm by 13.5 cm mold (2 ml) for hardening. Several smaller (80 μl) blocks were also made at the same cell concentration. The cells were lysed, as described by Smith *et al.* (*24*), by placing the agarose blocks in 1% *N*-lauroylsarcosine, 0.5*M* EDTA (*p*H 9), and 2 mg per milliliter of proteinase K for 48 hours, and then the DNA remaining in the blocks was prepared for restriction enzyme digestion by washing several times with TE (10 m*M* tris, 1 m*M* EDTA); the first two washes also included 1 m*M* phenylmethylsulfonyl fluoride (PMSF). Analysis of these blocks by PFGE demonstrated that about 95% of the DNA remained at the origin. To eliminate the 5% that was sheared, the blocks were "pre-electrophoresed" in the PFGE box for 2 hours to remove DNA less than 1000 kb in size, and then the blocks were removed from the gel.

Several dilutions of Mbo I (New England Biolabs) were tested to obtain a range of DNA fragments in the 20- to 1000-kb size range; a concentration of 0.025 unit per microgram of DNA gave optimum results. A scaled-up digest was done on the large agarose block, which was then placed in a preparative 1% agarose PFGE gel for DNA sizing, using annealed λ phage DNA size markers (*24*). The gel band corresponding to the approximate size range of 80 to 130 kb was cut out, placed in a dialysis bag with electrophoresis buffer, and electroeluted in the PFGE box. The gel fragment was removed, and the DNA was extensively dialyzed against TE. Analysis of a portion by PFGE confirmed that the general range of DNA sizes was in the expected distribution, and that the DNA had not sheared. Four micrograms of the size-selected DNA was further diluted to a concentration of 0.22 μg/ml and 2 μg of a purified 219-bp Bam HI–ended supF gene (*14, 17*) was added. After equilibration in ligase buffer for 2 hours, T_4 DNA ligase (New England Biolabs) was added to a concentration of 1.6 U/μl and ligation was carried out for 12 hours at 4°C. A second aliquot of ligase was then added and another 12-hour ligation was carried out. After addition of 20 μg of carrier transfer RNA, the DNA was precipitated with ethanol, resuspended in 100 μl of TE, and digested to completion with Eco RI. After phenol extraction and ethanol precipitation, the genomic DNA was ligated into a 3:1 molar excess of Eco RI–cut λCh3AΔlac. This is an A*am* B*am* phage that will accept Eco RI inserts from 0 to 12 kb; this vector was generated by replacing the Eco RI–Bam HI fragment of λCh3A (*35*) with the Eco RI–Bgl II polylinker from the miniplasmid πVX (*36*). The ligated DNA was packaged in vitro (*37*) and plated on the sup$^-$ strain MC1061. A total of 2 million clones was obtained; a study of ten random clones indicated that all contained from one to three supF genes within the phage insert, in each instance flanked by genomic DNA segments.

17. R. J. Dunn *et al.*, *J. Biol. Chem.* **256**, 6109 (1981).
18. C. S. Cooper *et al.*, *Nature (London)* **311**, 29 (1984).
19. M. Dean *et al.*, *ibid.* **318**, 385 (1985).
20. M. Park *et al.*, *Cell* **45**, 895 (1986).
21. F. Collins, J. Cole, M. Iannuzzi, J. Richards, unpublished data.
22. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984).
23. G. F. Carle and M. V. Olson, *Nucleic Acids Res.* **12**, 5647 (1984).
24. C. L. Smith *et al.*, in *Genetic Engineering*, J. K. Setlow and A. Hollaender, Eds. (Plenum, New York, 1986), vol. 8, pp. 45–70.
25. C. L. Smith and C. R. Cantor, *Methods Enzymol.*, in press.
26. G. F. Carle, M. Frank, M. V. Olson, *Science* **232**, 65 (1986).
27. K. Gardiner, W. Laas, D. Patterson, *Somatic Cell Mol. Genet.* **12**, 185 (1986).
28. R. White *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
29. P. J. Scambler, H.-Y. Law, R. Williamson, C. S. Cooper, *Nucleic Acids Res.* **14**, 7159 (1986).
30. M. Park *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
31. A. Poustka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4129 (1984).
32. P. F. R. Little and S. H. Cross, *ibid.* **82**, 3159 (1985).
33. A. R. Wyman, L. B. Wolfe, D. Botstein, *ibid.*, p. 2880.
34. A. Poustka *et al.*, *Nature (London)* **325**, 353 (1987).
35. F. R. Blattner *et al.*, *Science* **196**, 161 (1977).
36. B. Seed, *Nucleic Acids Res.* **11**, 2427 (1983).
37. B. Hohn and K. Murray, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3259 (1977).
38. A. P. Feinberg and B. Vogelstein, *Analyt. Biochem.* **137**, 266 (1984).

39. We thank S. Weissman and D. Ginsburg for valuable discussions; B. Handelin, T. Lugo, K. Fournier, and D. Housman for supplying the hybrid cell line HDm20s; F. Blattner for the λCh3AΔlac phage vector; and M. Roth for assistance with PFGE. This research was supported by NIH grant GM34960 and a Hartford Foundation fellowship grant to F.S.C.

15 December 1986; accepted 29 January 1987

Solution of a Protein Crystal Structure with a Model Obtained from NMR Interproton Distance Restraints

AXEL T. BRÜNGER, ROBERT L. CAMPBELL, G. MARIUS CLORE, ANGELA M. GRONENBORN, MARTIN KARPLUS, GREGORY A. PETSKO, MARTHA M. TEETER

Model calculations were performed to test the possibility of solving crystal structures of proteins by Patterson search techniques with three-dimensional structures obtained from nuclear magnetic resonance (NMR) interproton distance restraints. Structures for crambin obtained from simulated NMR data were used as the test system; the root-mean-square deviations of the NMR structures from the x-ray structure were 1.5 to 2.2 Å for backbone atoms and 2.0 to 2.8 Å for side-chain atoms. Patterson searches were made to determine the orientation and position of the NMR structures in the unit cell. The correct solution was obtained by comparing the rotation function results of several of the NMR structures and the average structure derived from them. Conventional refinement techniques reduced the *R* factor from 0.43 at 4 Å resolution to 0.27 at 2 Å resolution without inclusion of water molecules. The partially refined structure has root-mean-square backbone and side-chain atom deviations from the x-ray structure of 0.5 and 1.3 Å, respectively.

IN PROTEIN CRYSTALLOGRAPHY, THE initial determination of phases by multiple isomorphous replacement may be difficult because heavy-atom derivatives of the crystal are not available. If the structure of a similar or homologous molecule is known, the "molecular replacement method" (*1–3*) can be tried for the initial phasing. It involves the rotation and translation of the known structure in the unit cell of the target crystal to obtain the best match between the Patterson function calculated from the model and that from the observed diffraction data. The optimally oriented and translated model structure is used to begin conventional x-ray crystallographic refinement; this approach may or may not succeed (*1–3*). Criteria for determining when the molecular replacement method should work are available only for specific cases (*3*).

With the use of nuclear magnetic resonance (NMR) techniques (*4, 5*) a large number of approximate interproton distances for certain proteins (up to 80 residues) can be obtained (*6, 7*) and used to build three-dimensional structural models. Because of the limitations on the number and the range (only <5 Å) of these distances, the NMR data must be complemented with information about covalent structure (derived from bond lengths, bond angles, and dihedral angles) and internal packing requirements (derived from van der Waals repulsions). Restrained molecular dynamics, in which the energy function is augmented by effective potentials obtained from interproton distances (*8–10*), can add sufficient information to determine the solution structure of a protein. The method has been shown to yield converged structures with root-mean-square (rms) deviations from the x-ray structure of 1.5 to 3.0 Å (*9, 10*). Similar results have been obtained with distance geometry algorithms (*11, 12*). Recently the structure of the α-amylase inhibitor tendamistat has been solved independently by x-ray crystallography and NMR spectroscopy (*13, 14*); a comparison of the two structures indicates reasonable agreement.

We show in this report that structures obtained for a protein in solution with NMR data can be used to solve the crystal structure of the same protein by molecular replacement (*1–3*). As a test we use crambin (*15*), a protein of 46 residues, for which NMR structures with simulated data have been obtained (*9, 10*). We show that the individual NMR structures are not sufficiently accurate to solve the rotation and translation functions uniquely. Use of the refined x-ray structure itself as a model demonstrates that crambin is a difficult case for the molecular replacement method. We obtained the correct solution when we averaged over an ensemble of NMR structures or used the average structure derived from them. We demonstrate that the averaged

A. T. Brünger and M. Karplus, Department of Chemistry, Harvard University, Cambridge, MA 02138.
R. L. Campbell and G. A. Petsko, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139.
G. M. Clore and A. M. Gronenborn, Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, Federal Republic of Germany.
M. M. Teeter, Department of Chemistry, Boston College, Chestnut Hill, MA 02167.