Intervening Sequences in Human Fetal Globin Genes Adopt Left-handed Z Helices*

Michael W. Kilpatrick†, Jan Klysik‡§, Charles K. Singleton†, David A. Zaring‡, Thomas M. Jovin∥, Lawrence H. Hanau**‡‡, Bernard F. Erlanger**‡‡, and Robert D. Wells‡

From the †Department of Biochemistry, University of Alabama in Birmingham, Schools of Medicine and Dentistry, Birmingham, Alabama 35294, the ‡Max-Planck-Institute for Biophysical Chemistry, Department of Molecular Biology, Goettingen, D3400, West Germany, and the **Department of Microbiology, Columbia University, College of Physicians and Surgeons, New York, New York, 10032

(Received for publication, January 24, 1984)

The large intervening sequences (IVS2) of three human fetal globin genes contain tracts of alternating purine-pyrimidine sequences approximately 40–60 base pairs in length which adopt left-handed Z DNA helices under the influence of negative supercoiling. The amount of negative supercoiling (~0.045) required for the right- to left-handed transitions is within the physiological range. The structural aberrations between the right- and left-handed helices were mapped by sequencing the S1 nuclease cleavage sites. Two-dimensional gel electrophoretic analyses of the supercoil-induced relaxation served to characterize the type and length of left-handed structure. Furthermore, binding studies with several types of antibodies confirmed the presence of left-handed helices. Since these simple sequences appear to be hotspots for recombination and gene conversion, unusual DNA conformations may participate in genetic expression.

Left-handed helices have been demonstrated in synthetic DNA polymers (reviewed in Refs. 1–3) and in segments of DNA restriction fragments (2, 4) and recombinant plasmids (2–17). In general, these fragments and plasmids contained the simple sequence (dG-dC) which seems to occur infrequently in nature. Recent studies (5, 10, 13, 16, 17) demonstrated that the perfectly alternating purine-pyrimidine (dT-dA), (dC-dG), also adopts a left-handed structure. This result is significant since (dT-dC), (dC-dA), is an abundant sequence that is ubiquitous in eukaryotic DNA (18–21); however, the function of this simple sequence is uncertain.

The role of the static and dynamic structures and properties of DNA in gene expression has been reviewed (1, 22–24). Since the realization (2, 9) that left-handed Z DNA can coexist in close proximity with right-handed B DNA, it has been important to determine if a natural DNA sequence with a well defined genetic function can adopt different conformations, such as B and Z. Prior studies (16, 25–33) showed the binding of specific antibodies (raised versus left-handed Z DNA polymers which were chemically modified to stabilize the Z structure) to synthetic, recombinant, or natural DNAs and to chromosomes. However, definitive proof and precise localization of left-handed structures requires the combined use of several different direct structural determinations on DNAs of established sequence including the mapping of single-strand-specific nuclease cleavage sites at the junctions between helices with different chiralities (11–15), the supercoil- or salt-induced topoisomerase relaxation (2, 6, 9, 11, 13, 15, 17, 34), and the binding of different types of antibodies raised versus left-handed DNAs (6, 25–33), as well as other physical procedures (1–3, 5, 23, 24).

We have investigated the conformational properties of the large intervening sequences (IVS2) of variable length between codons 104 and 105 in human fetal globin genes (36, 37) since (a) they offer the opportunity to further evaluate the prediction (38, 39) that (purine-pyrimidine), sequences can adopt a Z structure and (b) they are involved in genetic recombination and intergenic exchange (gene conversion) (36, 37). The notable sequences are (TG)\(_5\)(CG)\(_5\)(TG)\(_5\), in the \(\gamma\) gene of chromosome B, (TG)\(_{11}\)(CG)\(_5\)(TG)\(_3\), (TG)\(_3\)(TG)\(_4\), in the \(\gamma\) gene of chromosome A, and (TG)\(_{18}\)(CG)\(_4\)(AC)(TG)\(_2\), (TG)\(_2\)(TG)\(_5\), in the \(\gamma\) gene of chromosome B.

We have determined that the global free energy of negative supercoiling can cause localized right- to left-handed transitions in these sequences. The structural aberrations between the right- and left-handed helices were mapped by sequencing the S1 nuclease cleavage sites. Also, two-dimensional gel electrophoretic analyses of the supercoil-induced relaxation and antibody binding studies confirmed the presence of left-handed helices.

**MATERIALS AND METHODS**

The construction and characterization of pRW751 was described (6). pRW780, -781, and -782 are pBR322 derivatives where the pBR322 EcoRI-BamHI 375-bp fragment was replaced by the EcoRI-BamHI intervening sequence of the human fetal globin genes (36, 37). pRW780 contains a 943-bp insert from the \(\beta\) gene of chromosome B containing the (TG)\(_5\)(CG)\(_5\)(TG)\(_5\) sequence. pRW781 contains a 953-bp insert from the \(\delta\) gene of chromosome A containing the (TG)\(_{11}\)(CG)\(_5\)(TG)\(_3\)(TG)\(_5\) sequence. pRW782 has a 971-bp insert from the \(\delta\) gene of chromosome B containing the (TG)\(_{18}\)(CG)\(_4\)(AC)(TG)\(_2\), (TG)\(_2\)(TG)\(_5\) sequence. All three clones have almost identical sequence background outside the purine-pyrimidine tracts specified above. pRW780–782 were characterized by restriction mapping with BamHI, EcoRI, Ball, Mbol, Hhal, HphI, HaeIII, XhoI, and Aul. Also, the EcoRI-BamHI insert in pRW780 was partially sequenced. The abbreviation used is: bp, base pair.

*This work was supported by National Institutes of Health Grants CA20578 and AI 06860, National Science Foundation Grant PCM-8002622, the Alexander von Humboldt Society, and the Max-Planck Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

‡ Present address, Department of Biochemistry, Medical Academy, 30-131 Lodz, ul. Lindeliny 6, Poland.

§ Supported by Postdoctoral Fellowship PF-1904 from the American Cancer Society. Present address, Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706.

∥ Supported by National Institutes of Health Predoctoral Training Grant 5T-32-GM07367.

† Supported by National Institutes of Health Postdoctoral Fellowship PF-1904 from the American Cancer Society. Present address, Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706.
plasmid DNAs were isolated as described (6). Topoisomer of the plasmids were prepared essentially as described previously (11, 43). Restriction enzymes were from Bethesda Research Laboratories except for BstNI (New England Biolabs), and S1 nuclease was isolated and characterized as described (44).

S1 Nuclease Reactions—S1 nuclease reactions were performed as follows. 6 μg of plasmid DNA in 150 μl of 40 mM Na acetate, 50 mM NaCl, 1 mM ZnSO4, (pH 4.6) were incubated at 37 °C for 20 min in the presence of 0.24 Vogt units of S1 nuclease. The reaction was terminated by addition of 7 μl of stop mixture (2 volumes of 0.5 M EDTA (pH 8.0) + 4.5 volumes of 1 M Tris base). The DNA was precipitated with 3 volumes of ethanol, redissolved with an appropriate buffer, and digested with a restriction enzyme as indicated in the figure legends. Acrylamide gel electrophoresis was carried out on 6% gels (4) in 40 mM Tris-HCl, 2 mM EDTA (pH 8.2). Gels were stained with ethidium bromide, and bands were visualized by UV illumination.

Mapping of S1 Nuclease-sensitive Sites at the Nucleotide Level—The general strategy for the mapping of S1 nuclease-sensitive sites at the nucleotide level for pRW780, pRW781, and pRW782 (Fig. 3) and for pRW751 (Fig. 4) is as follows. 50 μg of the supercoiled plasmid were digested with S1 nuclease and dephosphorylated with alkaline phosphatase (Sigma). The 5' ends were labeled with T4 polynucleotide kinase (P-L Biochemicals) and [γ-32P]ATP (New England Nuclear). The labeled plasmid was digested with the designated restriction enzymes to release restriction enzyme fragments. After acrylamide gel electrophoresis, the labeled fragments were extracted and recut with another restriction enzyme to obtain a suitable range of sizes for separation on the sequencing gel (45). The digest was denatured (45) and run next to the four sequencing lanes of the restriction fragment with the known sequence. In this way, we were able to determine precisely the length of fragments that correspond to the distance from the 3' restriction end to the 5' labeled site, and to correlate it with the known sequence. For pRW780, pRW781, and pRW782, the DNAs were labeled after S1 nuclease digestion. The isolated SI/EcoRI fragments from acrylamide gels (1.5% agarose) were cut with EcoRI and BamHI. The labeled SI/EcoRI fragments were extracted together and the two shorter bands were extracted separately. The longer set of bands was redigested with HpaII and the shorter set redigested with TaqI and electrophoresed on a sequencing gel next to the sequencing lanes of their restriction fragment of known sequence. This protocol allowed the mapping of the SI sites on the bottom strand of the sequence shown. In order to map SI sites on the upper strand, pRW751 cut with S1 nuclease was labeled and redigested with SalI. The longer doublet bands were extracted separately from the shorter set from the acrylamide gel and were recut with HpaII and HaeIII, respectively. The rest of the study was done identically as described for the mapping experiments of the bottom strand. Thus, four sequencing gels were required for each complete set of data.

Two-dimensional Gel Electrophoresis of Topoisomers—For each sample, 2 μg of a family of topoisomers, generated as described (43), were loaded onto a 1.5% agarose gel, and electrophoresis in the first dimension was carried out in 80 mM Tris acetate, 40 mM Na acetate, 2.5 mM EDTA (pH 7.6) at 125 V for 65 h. The ladder of topoisomers was then cut from the gel, soaked for 1 h in the same buffer plus 1.25 μM chloroquine (Sigma), and then embedded in a 1.5% agarose gel also containing 1.25 μM chloroquine. Electrophoresis in the second dimension was carried out in the chloroquine-containing buffer at 100 V for 45 h. The gel was stained with ethidium bromide, and the DNA was visualized under UV light.

RESULTS

Restriction Mapping of S1 Nuclease Cleavages—The three supercoiled plasmids (pRW780, -781, and -782) containing the large intervening sequences (IVS2) of the human fetal globin genes were each treated with S1 nuclease. Fig. 1 (lanes 1) shows that pRW780 was linearized; identical results were obtained with the other two supercoiled plasmids. Since each of these recombinants contains a single EcoRI site, digestion of the S1 nuclease-treated DNAs with EcoRI released sets of fragments that can be mapped on the completely sequenced plasmids. Fig. 1 (lanes 2–4) show such mapping studies for pRW780 at negative supercoil densities of 0.055, 0.075, and 0.14, respectively. Comparison of the resulting fragments with the size standards localized the S1 nuclease-sensitive sites near the (TG)n(CG)n(TG)n blocks. Lanes 3 and 4 reveal that the sizes of the released fragments changed significantly as the −α increased, which indicated a change in the position of the S1-sensitive sites. A similar effect of supercoil density on the nuclease-sensitive sites was found for pRW781 containing the (TG)n(CG)n(TG)n (TG)n block. This change in position of nuclease cleavage as a function of supercoil density was not observed previously for (dG-dC)-containing plasmids (11, 12) under similar conditions.

Mapping of the S1 nuclease cleavage sites on pRW782 by EcoRI digestion of the treated plasmid is shown in Fig. 1 (lane 5). Again, the cleavage sites are near the (TG)n(CG)nAC(TG)n(TG)n block. However, the S1 cleavage sites on pRW782 were always found at the same distances from the EcoRI site irrespective of the supercoil density (up to −0.104), in contrast to the results found for pRW780 and -781.

The S1 nuclease sensitivity of these three plasmids was strongly dependent on negative supercoiling since preparations with few superhelical turns showed no specific cleavage. Similar results were reported (11, 12) for five recombinants containing (dG-dC) tracts. These data strongly indicate that the purine-pyrimidine tracts in pRW780, -781, and -782 are adopting left-handed conformations.

The effect of negative supercoil density on the amount of the fragments produced by S1 and EcoRI treatment of pRW780 was evaluated. Preparations of the plasmid containing different populations of topoisomers were digested, and the polyacrylamide gel analyses were quantitated by microdensitometer scanning. Fig. 2 shows that the 350- and 332-bp fragments were initially found at low supercoil densities. As −α increased, the 363- and 318-bp fragments became visible. Further increases in the number of supercoils corresponded with formation of the 292- and 278-bp bands and complete disappearance of the 332- and 318-bp bands. Fig. 2 does not include the formation of the 268- and 261-bp bands that were
All observations were made relative to the EcoRI site, which is defined as location 0.

A quantitative measure was obtained by determining the relative amount of each labeled fragment observed in each gel. The relative amount of each labeled fragment was calculated as the intensity of the band divided by the sum of the intensities of all labeled bands in the same gel.

The results of the electrophoretic mobility experiments are shown in Fig. 1. The mobility of each fragment is given in base pairs (bp).

The results show that the mobility of each fragment is dependent on the presence of left-handed DNA. For example, the fragment with a mobility of 125 bp in the control gel has a mobility of 100 bp in the gel containing left-handed DNA.

The results also show that the mobility of each fragment is dependent on the presence of right-handed DNA. For example, the fragment with a mobility of 120 bp in the control gel has a mobility of 105 bp in the gel containing right-handed DNA.

The results further show that the mobility of each fragment is dependent on the presence of single-stranded DNA. For example, the fragment with a mobility of 110 bp in the control gel has a mobility of 95 bp in the gel containing single-stranded DNA.

The results indicate that the mobility of each fragment is dependent on the presence of both left-handed and right-handed DNA. The mobility of each fragment is greatest in the gel containing both types of DNA.

The results also show that the mobility of each fragment is dependent on the presence of single-stranded DNA. For example, the fragment with a mobility of 105 bp in the gel containing single-stranded DNA has a mobility of 90 bp in the gel containing both types of DNA.

The results indicate that the mobility of each fragment is dependent on the presence of both left-handed and right-handed DNA. The mobility of each fragment is greatest in the gel containing both types of DNA.

The results also show that the mobility of each fragment is dependent on the presence of single-stranded DNA. For example, the fragment with a mobility of 90 bp in the gel containing single-stranded DNA has a mobility of 75 bp in the gel containing both types of DNA.

The results indicate that the mobility of each fragment is dependent on the presence of both left-handed and right-handed DNA. The mobility of each fragment is greatest in the gel containing both types of DNA.
Left-handed IVS2 Sequences

FIG. 3. Map of S1 nuclease cleavage sites of fetal globin gene intervening sequences. For pRW780 and pRW781, mapping was done for two different superhelical densities (−a = 0.055 and 0.14 for pRW780 and −a = 0.076 and 0.12 for pRW781). In both cases, the outermost designation of cleavage frequencies represents the data for the greater negative supercoil densities. For pRW782 the −a = 0.075. The lengths of the vertical bars between the nucleotides represent the intensities of the bands on the sequencing gels. The relative intensities (vertical bars) for one set of data (horizontal lines) are very reproducible; two separate experiments resulted in virtually identical distribution data around the purine-pyrimidine blocks. The limit of resolution on the sequencing gels is indicated by the horizontal arrows. Thus, it was not possible to determine the extent of cleavage across both strands for all junction regions. The BamHI site is located on the left of the BstNI site and the EcoRI site is on the right of the BstNI site.

For pRW780 and pRW781, the S1 cleavage data (Fig. 3) revealed a longer left-handed sequence at higher (−a = 0.13) superhelical density. Although the limit of resolution of the two-dimensional gel analysis prevented a thorough study of this phenomenon, there was some indication of further abnormal mobility of topoisomers with such high degrees of supercoiling.

The combined results on the lengths of left-handedness for the three plasmids agree well from the S1 cleavage and the extent of supercoil relaxation determinations. This suggests that regions of 10.4 bp/turn B helix are converted to 12.0 bp/turn Z helix. However, other interpretations are possible, especially considering the breadth of the junction aberrations (Figs. 3 and 4). Indeed, the left-handed regions may contain several types of helical structures, and the junctions may represent a continuum of conformations between right-handed B and left-handed Z.

Anti-Z DNA IgG Binding Studies—A further test that these regions of intervening sequence can adopt left-handed Z helices was provided by anti-Z DNA binding measurements. The capacity of pRW780-782 to bind IgGs, both poly- and monoclonal, raised versus brominated (dG-dC)n,(dG-dC)n, was
tested by the gel retardation assay of Pohl et al. (28) as modified (31). When each DNA was supercoiled at a density of $-0.075$, anti-Z IgG binding was observed using purified polyclonal rabbit T-4 IgG (31). In these experiments, all the Form I DNA was retarded in electrophoretic mobility. The small amount of Form II (nicked) and Form III (linear) DNAs in the preparations served as an internal control, and their electrophoretic mobilities were not affected. Furthermore, when each plasmid was completely relaxed with topoisomerase I, no binding was found. Identical results were found when polyclonal antibodies raised against N-acetylamino-fluorene-modified (dG-dC), (dG-dC)$_n$ were employed using a double antibody competition radioimmunoassay (46). When pRW780-782 (at $-\dot{\alpha}$ of either 0 or 0.075) were tested (in a single experiment) for their capacity to bind a (dG-dC)$_n$ sequence-dependent monoclonal IgG D-11 (28, 47) (a kind gift from R. Thomae and F. Pohl, University of Constance, West Germany), electrophoretic gel retardation was found for supercoiled pRW780, but comparatively less retardation was observed for supercoiled pRW781 and -782. The stronger interaction with pRW780 may be due to its longer segment of (dC-dG) sequence.

In other control experiments with pRW751 (6-9) at the same negative superhelical density ($-0.075$), the (dG-dC)$_n$ sequence-specific D-11 monoclonal IgG (as well as the polyclonal anti-Z (T-4) IgG) was specifically bound to the supercoiled (Form I) but not nicked (Form II) or to experimentally linearized (Form III) DNAs. In contrast, pRW777 (10-13), containing 64 bp of alternating (dT-dG)-(dC-dA), at a superhelical density of $-0.075$ does not bind D-11 IgG but does bind the Z sequence-independent anti-Z IgG T-4 (35). The lack of binding of D-11 to (dT-dG)-(dC-dA) sequences also holds for all known pyrimidine-substituted (methyl or halogen) (dT-dG)$_n$-(dC-dA) family members (35). This is in marked contrast to T-4 IgG which does bind all known (dG-dC) (dG-dC) or (dT-dG)-(dC-dA) family members (31). Thus, the sequence specificity of anti-Z DNA antibodies established with synthetic polynucleotides is preserved in complexes with corresponding inserts in plasmid DNAs. It will be interesting to determine if contiguous IgG binding exists between adjacent (dG-dC)-(dA-dC) regions.

**DISCUSSION**

The purine-pyrimidine type regions which are $\sim\!40-60$ bp in length in the large intervening sequences (IVS2) of human
whereas the nonallelic genes differ considerably. They proposed that the 5' two-thirds of the $\alpha\gamma$ gene on chromosome A was "converted" by an intergenic exchange to become more like the $\alpha\beta$ gene on its own chromosome A than it is like the allelic $\alpha\gamma$ gene on the other chromosome B. Interestingly, no convincing evidence was found (37) for a gene conversion close to the (dA-dC)$_{20}$ sequence located at position 6261 in the $\alpha\gamma$ part of the duplication. The short length of this sequence may not permit it to adopt the required left-handed conformation; a (dG-dC)$_{20}$ tract will adopt a left-handed structure (1a), but this sequence undergoes the R to L transition much easier than the (dT-dG)$_{20}$, (dC-dA)$_{20}$, sequence (2, 5, 13). Furthermore, we have previously documented (8) the unusual recombination behavior of (dG-dC)-containing plasmids in E. coli. The unusual types of conformations adopted by these simple sequences may serve as recognition sites for the enzymes involved in recombination as well as other genetic processes. In addition, the apparent nonhelical regions at the junctions may serve as entry points for strand interactions (42).

Acknowledgments—We thank Dr. Oliver Smithies, Department of Genetics, University of Wisconsin, Madison for the original clones containing the intervening sequences of the human fetal globin genes and Dr. James Wang for providing a manuscript in press.

REFERENCES


Fig. 5. Analysis of supercoiling-induced changes in the structure of pRW780 and a control plasmid pRW451 by two-dimensional gel electrophoresis. The figure shows the mobilities in the first dimension of topoisomers of the two plasmids as determined by this technique. Mobilities were determined relative to the most relaxed topoisomer. Topoisomer mobilities from four separate analyses, each focusing on different regions of the gel, are combined in this figure. Topoisomer mobilities from different experiments were normalized using the distance between topoisomer -10 and the most relaxed topoisomer as a normalization factor. ×, pRW451, a control plasmid (6) containing the 174-bp BamHI fragment from pBR322 cloned into the filled in BamHI site of pBR322; ●, pRW780.

Fig. 6. Extent of relaxation of plasmids containing intervening sequences from human fetal globin genes as a function of number of supercoils. pRW780-782 were analyzed by the two-dimensional gel assay (40) shown in Fig. 5. No relaxation was seen for the control plasmid pRW451 which showed a smooth progression (no break in the curve as in Fig. 5).

The demonstration that the simple purine-pyrimidine sequence in IVS2 between codons 104 and 105 can adopt left-handed helices is particularly significant since Slightom et al. (36) and Shen et al. (37) have indicated that they are "hot-spots" for recombination. On the 5' side of this simple sequence, the allelic $\alpha\gamma$ genes differ considerably in IVS2 whereas the nonallelic $\alpha\beta$ and $\alpha\gamma$ genes from chromosome A differ only slightly. However, on the 3' side of the simple sequence, the allelic genes differ only slightly, whereas the nonallelic genes differ considerably.
Left-handed IVS2 Sequences