THE ROLE OF DNA STRUCTURE IN GENETIC REGULATION

Authors: **Robert D. Wells**
Robert W. Blakesley
Stephen C. Hardies
Glenn T. Horn
Jacquelynn E. Larson
Erik Selsing
Department of Biochemistry
University of Wisconsin
Madison, Wisconsin

John F. Burd
Ames Company
Division of Miles Laboratories, Inc.
Elkhart, Indiana

Hardy W. Chan
Laboratory of Biology and Viruses
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland

Jerry B. Dodgson
Department of Chemistry
California Institute of Technology
Pasadena, California

Kaj Frank Jensen
Universitets Institut for Biologisk Kemi B
Copenhagen, Denmark

Ingolf F. Nes
Department of Biochemistry
University of Bergen
Bergen, Norway

Roger M. Wartell
Department of Physics and Biology
Georgia Institute of Technology
Atlanta, Georgia

Referee: Peter H. von Hippel
Institute of Molecular Biology
University of Oregon
Eugene, Oregon

I. INTRODUCTION

Fundamental Features of Gene Expression

Substantial advances have occurred in the past 15 years towards understanding various facets of genetic regulation. During that time our knowledge has progressed from vague genetic concepts to a clearer biochemical and physical comprehension of certain control elements. For example, detailed information is presently available on the molecular interactions involved between certain operators and repressors, as well as between some promoters and RNA polymerases.

At the heart of a complete and fundamental understanding of gene regulation is the precise interaction of a specific region of DNA with a regulatory protein. For years it has been known that this interaction is exquisitely specific and leads to very tight binding in some cases. In order for a region of DNA to be recognized as a "specific site" it must have structural features which make it distinctive from other regions of DNA. A complete understanding of this interaction requires detailed knowledge of the kinetics and thermodynamics of this process and detailed information on the structural features involved in the recognition interaction.

Numerous excellent review articles have appeared on the genetic aspects of gene regulation.[1-3] In addition, other reviews have dealt extensively with the proteins involved in gene expression.[4-6] Due to space limitations this review will not deal primarily with these topics.

This review will place special emphasis on consideration of the role of DNA conformation in gene expression. The hypothesis at the foundation of this concept is that DNA does not have the same conformation throughout its entire length; instead, it has interesting structural variations at certain positions. A corollary to this hypothesis is that DNA structure is not unalterable but may be perturbed when certain proteins are bound. At the outset it should be emphasized that proof for this hypothesis is only beginning to emerge. Only recently data have been presented which indicate that chromosomal DNA does not have the same static conformation throughout its entire length. Likewise, recent data suggest that neighboring nucleotide sequences can mutually influence the dynamic properties and conformations of each other. Rigorous proof of these concepts must await further studies, but they are of fundamental importance to our eventual understanding of the details of DNA conformation in gene regulation. Emphasis will be placed on the work from this laboratory since only a few other groups have addressed the problem at this place on the spectrum between the pure physical chemistry of nucleic acid structures through DNA biochemistry to pure genetic analyses. Also, this review is not intended to be comprehensive. On some topics it is necessarily speculative; however, it is hoped that the hypotheses presented will provoke revealing experiments.

In past years, this problem has attracted only modest interest from an experimental standpoint. This is due to the lack of appropriate probes for determining unique configurations, and to the extent of sensitivity required for detection. Since it appears certain that structural anomalies in chromosomal DNA cannot be detected with composite measurements (such as circular dichroism, nuclear magnetic resonance [NMR], spectroscopy, or X-ray diffraction on high molecular weight molecules [see below]), it follows that such regions must be present in very small amounts per molecule. This is consistent with the notion that they may exist at regulatory sites. In order to determine the existence of such regions, sensitive monitoring agents, such as repressor-DNA interactions, are necessary. Also, very sensitive probes are required such as the single-strand-sensitive-specific nucleases which have high specificity for single-stranded, compared to double-stranded, nucleic acids.

DNA Structural Features Which May Be Recognized by Regulatory Proteins

At the outset it will be useful to consider the possible types of nucleic acid structural features which may be recognized by special proteins. Some of these structural features have been discussed previously by von Hippel[7] in a short article which showed considerable insight into the possible role of nucleic acid configuration in protein recognition.

Single-stranded DNA

True single-stranded regions – If a polynucleotide chain is single stranded, e.g., completely devoid of ordered structure, it would contain an abundance of specific sites to which a regulatory protein may attach. These would include the

heterocyclic bases as well as the phosphate and sugar moieties. However, most DNA genomes are not single stranded but are double stranded; thus, we shall not consider this possibility further. In addition, it is apparent that even single-stranded polynucleotides are not completely devoid of ordered structure;[8] however, in any case, an abundance of structural information is readily available for protein recognition.

Double-helical regions within "single-stranded" genomes – As described below, at least three DNA viral genomes which are considered to be primarily "single stranded" (i.e., do not exist as a double-stranded or replicative form within the virus particle) contain substantial amounts of double-helical hairpin loop regions according to their sensitivity to certain DNA restriction endonucleases. These DNAs are from ϕX174, M13, and fd viruses. In addition, recent sequence studies[9] on single-stranded RNA genomes suggest that the RNAs may be folded into highly ordered configurations. The double-helical regions in these single-stranded viral genomes could serve as recognition benchmarks for biological processes such as recombination[10] or replication.[11] Additional specificity in the system could be due to "frayed" polynucleotide regions near the helical segments; protein recognition of these frayed regions near the helical segments would provide for a greater degree of specificity than might be found in a completely double-stranded DNA.

Double-stranded DNA

Recognition of double helix – If there are no structural anomalies in duplex DNA genomes, e.g., base pairs in chromosomal DNA only exist in a typical *B* configuration (see below), regulatory proteins must recognize structural features of the double-helical molecule. Examination of a molecular model of DNA reveals that the unique features (base sequences) of regulatory regions of DNA are well protected within the helical structure. In the major groove, only the six and eight positions of purines and the four and five positions of pyrimidines are available. In the minor groove, even fewer structural features are available for specific interactions with proteins. It is conceivable that there could be sufficient specificity in the interaction between a duplex *B*-DNA site (providing a sufficiently large number of base pairs were involved) and a regulatory protein (of approximate molecular weight 100,000). However, more specificity would be provided by a partially nonhelical structure.

Partially nonhelical DNA segment – If a regulatory region of DNA was partially nonhelical, perhaps due to increased thermolability, this segment could be substantially more easily recognized by regulatory proteins than helical *B*-DNA structure. This thermolabile region (Figure 1A) was previously proposed as a possibility for explaining the high degree of sensitivity of repressor binding to λp*lac* DNA after treatment with S_1 or mung bean nucleases. This site could be generated by low GC content or by the presence of modified nucleotides which destabilize the helix. Direct studies (see below) have been performed to distinguish between models containing single-stranded regions and those which are completely helical.

Cruciform region – An alternate model which contains nonpaired nucleotides is the cruciform structure which can be generated from a number of the regulatory regions which have been sequenced to date (see below). However, to the best of our knowledge, there is no direct evidence supporting the existence of this type of configuration in a natural chromosome for any substantial period of time. Indeed, some evidence[12,13] suggests that it does not exist.

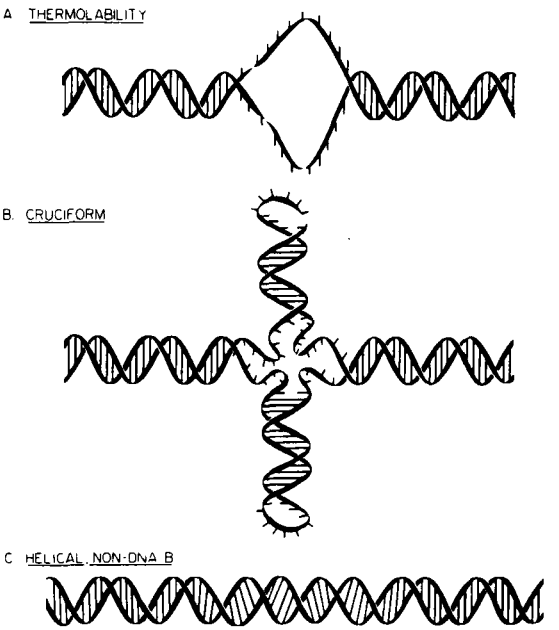

FIGURE 1. Possible models for structural irregularities in DNA.

However, if a cruciform exists transiently, it could be stabilized by a suitable protein.

Helical, non-*B*-DNA configuration – In this model a unique region may be totally helical but in a different geometric configuration (not *B*-DNA) than the majority of the DNA chromosome. For example, tilt of the base pairs or the helical dimensions could be somewhat altered; such helical variances have been recognized with DNA polymers (described below). In addition, a previously unrecognized configuration was observed at the junction of the AT blocks and GC blocks of $d(C_{15}A_{15})$ and $d(T_{15}G_{15})$ by high resolution NMR studies (see below).

Protein-induced Distortion of DNA

When a regulatory protein binds to a specific site on DNA, it may distort that region and/or nearby sequences. The capacity of a region of DNA to be distorted may be an important component in the recognition mechanism. Also, the binding of a second protein may be influenced by the prior binding of another protein; the influence may be transmitted through the DNA molecule which acts like a conduit of dynamic and/or static structural changes. However, this hypothesis does not resolve the uncertainty of the mechanism of recognition by the first protein. A portion of this model has been termed the "director protein" concept,[7] and a specific hypothesis was formulated to explain the capacity of the catabolite gene activator protein to influence transcription of the *lac* and *gal* operons.[14,15]

II. STUDIES WITH DNA POLYMERS OF DEFINED REPEATING NUCLEOTIDE SEQUENCE

In the mid 1960s, our laboratory became interested in the concept of DNA nucleotide sequence influencing DNA properties as an outgrowth of our studies on the genetic code.[16] High molecular weight DNA polymers of defined repeating nucleotide sequences were synthesized and used as templates for forming RNAs of defined repeating nucleotide sequence. Subsequently, these RNAs were used as messenger molecules to form polyamino acids with specified composition and sequence.

Early studies with these DNAs by several spectral and physical techniques demonstrated that DNA polymers with the same nucleotide composition but different sequences had fundamentally different properties. The first investigations which revealed these notions were absorbance-temperature and absorbance-pH profiles on $(dT\text{-}dC)_n \cdot (dG\text{-}dA)_n$ and $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$. In 1965 and 1967,[17,18] analytical buoyant density analyses in CsCl gradients demonstrated that sequence isomeric* DNAs had different and characteristic properties.

The notion that sequence isomeric DNAs should have different properties was unexpected at that time, since many investigations between 1955 and 1965 on natural DNAs showed that DNAs with the same nucleotide composition had virtually identical properties. Some of the properties studied were X-ray diffraction (see below), absorbance-temperature transitions[19] and buoyant density analyses.[20] Thus, it was not recognized that the studies on natural DNAs were composite measurements on various types of sequences. Alternatively, the effect of nucleotide sequence on DNA properties is apparent in the case of DNA polymers, presumably because of their alternating and repetitious sequence. Thus, any effect of sequence is greatly magnified and not masked, or cancelled out, by other types of sequences.

Table 1 lists the duplex DNAs which have been synthesized. The techniques of synthesis and characterization of these molecules have been reviewed previously.[21] DNAs of the repeating homo-, di-, tri-, and tetranucleotide sequences have been prepared and studied. The majority of these DNAs were formed by a combination of chemical and enzymatic techniques; however, five of the most simple DNAs are synthesized by *de novo* reactions. Thus, no preformed template is necessary[22] making their synthesis rather simple. However, after any of these DNAs has been synthesized and fully characterized, it is then possible to rereplicate it so that a continuing supply is available in principle.

Over the past 10 years, various studies have been performed on the physical, spectral, and biological properties of these DNAs in an effort to fully evaluate the influence of DNA structure on properties (Table 2). Since the results of virtually all of these measurements have been previously reported and these studies have been previously

*Sequence isomeric DNAs are defined as DNAs with the same nucleotide composition but different sequences. Thus, the arrangement of nucleotides on the complementary strands is different, but the overall composition is identical.

TABLE 1

Double-stranded DNA Polymers with Repeating Sequences[a]

Repeating homopolymers
- $(dA)_n \cdot (dT)_n$
- $(dG)_n \cdot (dC)_n$
- $(dI)_n \cdot (dC)_n$

Repeating sequences with self-complementary structures
- $(dA\text{-}dT)_n \cdot (dA\text{-}dT)_n$
- $(dG\text{-}dC)_n \cdot (dG\text{-}dC)_n$
- $(dI\text{-}dC)_n \cdot (dI\text{-}dC)_n$

Repeating dinucleotide polymers
- $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$
- $(dT\text{-}dC)_n \cdot (dG\text{-}dA)_n$

Repeating trinucleotide polymers
- $(dT\text{-}dT\text{-}dG)_n \cdot (dC\text{-}dA\text{-}dA)_n$
- $(dT\text{-}dA\text{-}dC)_n \cdot (dG\text{-}dT\text{-}dA)_n$
- $(dA\text{-}dA\text{-}dT)_n \cdot (dA\text{-}dT\text{-}dT)_n$
- $(dT\text{-}dT\text{-}dC)_n \cdot (dG\text{-}dA\text{-}dA)_n$
- $(dA\text{-}dT\text{-}dC)_n \cdot (dG\text{-}dA\text{-}dT)_n$
- $(dT\text{-}dC\text{-}dC)_n \cdot (dG\text{-}dG\text{-}dA)_n$

Repeating tetranucleotide polymers
- $(dT\text{-}dT\text{-}dA\text{-}dC)_n \cdot (dG\text{-}dT\text{-}dA\text{-}dA)_n$
- $(dT\text{-}dA\text{-}dT\text{-}dC)_n \cdot (dG\text{-}dA\text{-}dT\text{-}dA)_n$

[a] DNA polymers containing nucleotide analogues of some of these DNAs have been synthesized and characterized and were reviewed previously.[21] The I-containing DNAs are included in this table because of their relative importance.

TABLE 2

Types of Studies Performed with Duplex DNAs of Defined Repeating Nucleotide Sequences

	Reference
Physical determinations	
Absorbance-temperature transition	17, 23–25
Absorbance-pH	17, 23, 26, 27
Buoyant density in CsCl and Cs_2SO_4 gradients	18, 23, 25, 27, 28
Binding of actinomycin D	29, 30
Binding of netropsin	31, 32
The specific formation of three-stranded complexes between double-stranded DNA and single-stranded RNA and their selective inhibition of transcription	33
Viscosity	24, 26, 34
X-ray diffraction	see section on structures of synthetic homopolymer duplexes
Spectral determinations	
Circular dichroism	23, 35, 39, 47
Ultraviolet spectra	23, 24
High-resolution proton nuclear magnetic resonance studies	36–38
Biological studies	
In vitro replication	17, 24, 40–44
In vitro transcription	16, 33, 45, 46
lac repressor binding	48, 49
Interferon induction	34, 50–52

reviewed in part,[21] they will only be summarized at this point. (The X-ray results are discussed separately below.)

Polymer nomenclature is in accordance with the IUPAC-IUB recommendations. Thus, poly(dT-dG)·poly(dC-dA) is a double-stranded DNA with precisely alternating dG and dT moieties in one strand and alternating dC and dA moieties in the complementary strand. An alternative form of nomenclature for this DNA is $(dT\text{-}dG)_n\cdot(dC\text{-}dA)_n$. However, the value of n is not necessarily the same in all cases, neither for this example nor for the DNAs shown in Table 1.

The polymer nomenclature was designed to show the repeating, complementary, and antiparallel nature. Thus, poly(dT-dA-dC)·poly(dG-dT-dA) stands for the polymer which has the deoxy sequence ...pTpApCpTpA... in one strand and ...pGpTpApGpT... in the complementary strand, both of which are written with the 5′-hydroxyl on the left and the 3′-hydroxyl on the right. The nucleotides at the beginning and ends of the polymers are unknown. Thus, the above example could equally well be designated as $(dA\text{-}dC\text{-}dT)_n\cdot(dA\text{-}dG\text{-}dT)_n$.

The general conclusion from all of these determinations is that sequence isomeric DNAs do not have identical properties; instead, they have different and characteristic properties. In each group of sequence isomeric DNAs a polymer with both purines and pyrimidines in each complementary strand usually has properties closer to those of naturally occurring DNAs than the polymer containing all purines in one strand and all pyrimidines in the complementary strand. Moreover, there is greater variance in virtually every measurement between properties of the more simple sequence isomers [i.e., between $(dG\text{-}dC)_n\cdot(dG\text{-}dC)_n$ and $(dG)_n\cdot(dC)_n$] than between the more complex sequence isomers [i.e., between $(dT\text{-}dT\text{-}dC)_n\cdot(dG\text{-}dA\text{-}dA)_n$ and $(dT\text{-}dT\text{-}dG)_n\cdot(dC\text{-}dA\text{-}dA)_n$]. The repeating tri- and tetranucleotide polymers resemble each other and natural DNA more closely than do polymers with only two complementary nucleotides.

Two of the early key determinations performed on solutions of DNA polymers were absorbance-temperature transitions and analytical buoyant density gradient studies. A great deal of information was obtained from absorbance-temperature transition studies with the DNA polymers, and this information has been reviewed.[21] One of the more striking conclusions is that sequence isomeric DNAs do not have identical helix-coil transitions. Instead, sequence isomeric polymers have T_m values which differ by as much as 9°C.

In addition, analytical density-gradient centrifugation studies in both CsCl and Cs_2SO_4 solutions have been performed at both pH 7.3 and 12.5.[18,23,27] Each polymer has its own characteristic and unique density value. This behavior was unexpected since studies with natural DNAs had demonstrated direct relationships between buoyant density and nucleotide composition in both salt solutions.[20,53] Also, natural DNAs with the same nucleotide composition had the same buoyant density.

Figure 2 shows a composite of helix-coil transitions and CsCl buoyant density determinations for 12 DNA polymers. This empirical relationship demonstrates that the isomer with the greatest T_m value always has the smallest density (data points for sequence isomeric DNAs are connected). In addition, the points for the sequence isomers do not fall in the line generated from values with natural DNAs, but tend to straddle this line with the exception of $(dT\text{-}dA\text{-}dC)_n\cdot(dG\text{-}dT\text{-}dA)_n$ which has an unusually high density.[21,23] The detailed explanation of buoyant density and helix-coil transition data requires a quantitative knowledge of the polymer properties governing these measurements. A complete understanding is not yet available for either measurement; thus, it is not possible to fully define the reason for this behavior. However, these composite data clearly indicate that the data for natural DNAs are an approximate average of the DNA polymer values.

Most of the determinations listed in Table 2 are inadequate for providing fundamental information on the absolute conformation of a DNA polymer. Instead, most determinations measure differences in the property in question for the DNAs under study. Only X-ray diffraction measurements on oriented fibers provide static structural information. The results of these studies are described below. Also, high-resolution NMR studies show promise for revealing detailed conformational properties of DNA.[36] However, these studies are in their infancy; comparative determinations with a variety of DNAs and the interpretation of these results are not yet available.[36,165] However, as described below, high-resolution NMR studies on the DNA block polymer $(dC_{15}A_{15})\cdot(dT_{15}dG_{15})$ revealed that three to four A-T pairs at the

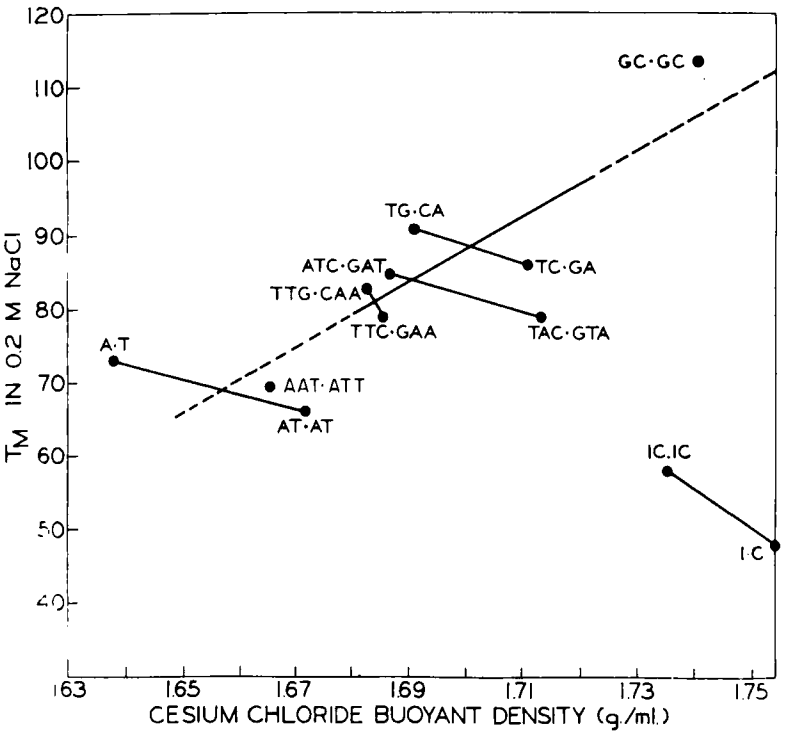

FIGURE 2. T_M as a function of CsCl buoyant density for synthetic DNAs. The solid diagonal line (dotted for extrapolation) represents the expected values for natural DNAs of various base compositions. The density standard is *Escherichia coli* DNA (1.703 g/ml^{-1}). The point for AAT·ATT is extrapolated from the published information[25] and the T_M data for the polymers containing G and C are extrapolated from the observed data.[23] Whereas it is realized that large errors may be made in the latter extrapolations, these points are included solely for the sake of comparison. The extrapolated T_M value for $(dG)_n \cdot (dC)_n$ at this salt concentration is 106°C. Cesium chloride density studies on this DNA have not been successful due to aggregation problems; similarly Cs_2SO_4 measurements are beset with difficulties but it is clear that the values[23] are greater than those found for $(dG\text{-}dC)_n \cdot (dG\text{-}dC)_n$. Thus, since the same ordering of densities is found in both CsCl and Cs_2SO_4 solutions, $(dG)_n \cdot (dC)_n$ is probably more dense than the alternating polymer in CsCl solution. Hence, this pair of sequence isomers also conforms to the observation described in the text. (From Wells, R. D., Larson, J. E., Grant, R. C., Shortle, B. E., and Cantor, C. R., *J. Mol. Biol.*, 54, 465, 1970. With permission.)

junction of the A-T and G-C blocks have a configuration different from the remainder of the A-T pairs in the block polymer, as well as those in any other A-T polymer examined to date.[36] In the future, this technique may be useful for obtaining a detailed structure of nucleic acids.

III. THE STRUCTURE OF DNA DUPLEXES: X-RAY STUDIES

Classical DNA Structures (A, B, and C)

To date, most of the information concerning the three-dimensional conformations adopted by DNA duplexes has resulted from X-ray diffraction studies of oriented fibers of high molecular weight DNA. For instance, such studies have shown that the Watson-Crick base-paired double helix can be observed in two different conformational forms, designated as *A*-DNA and *B*-DNA (Figure 3), depending on the environment of the DNA molecules in the fiber.[54,55] At high relative humidities (most likely corresponding to high levels of hydration of DNA in the fiber), the *B*-DNA conformation is usually observed, whereas the *A*-DNA helix is seen at lower humidities (Table 3). A reversible transition between these two forms can be induced by appropriately raising or lowering the hydration of the fiber.[55,56] Salt

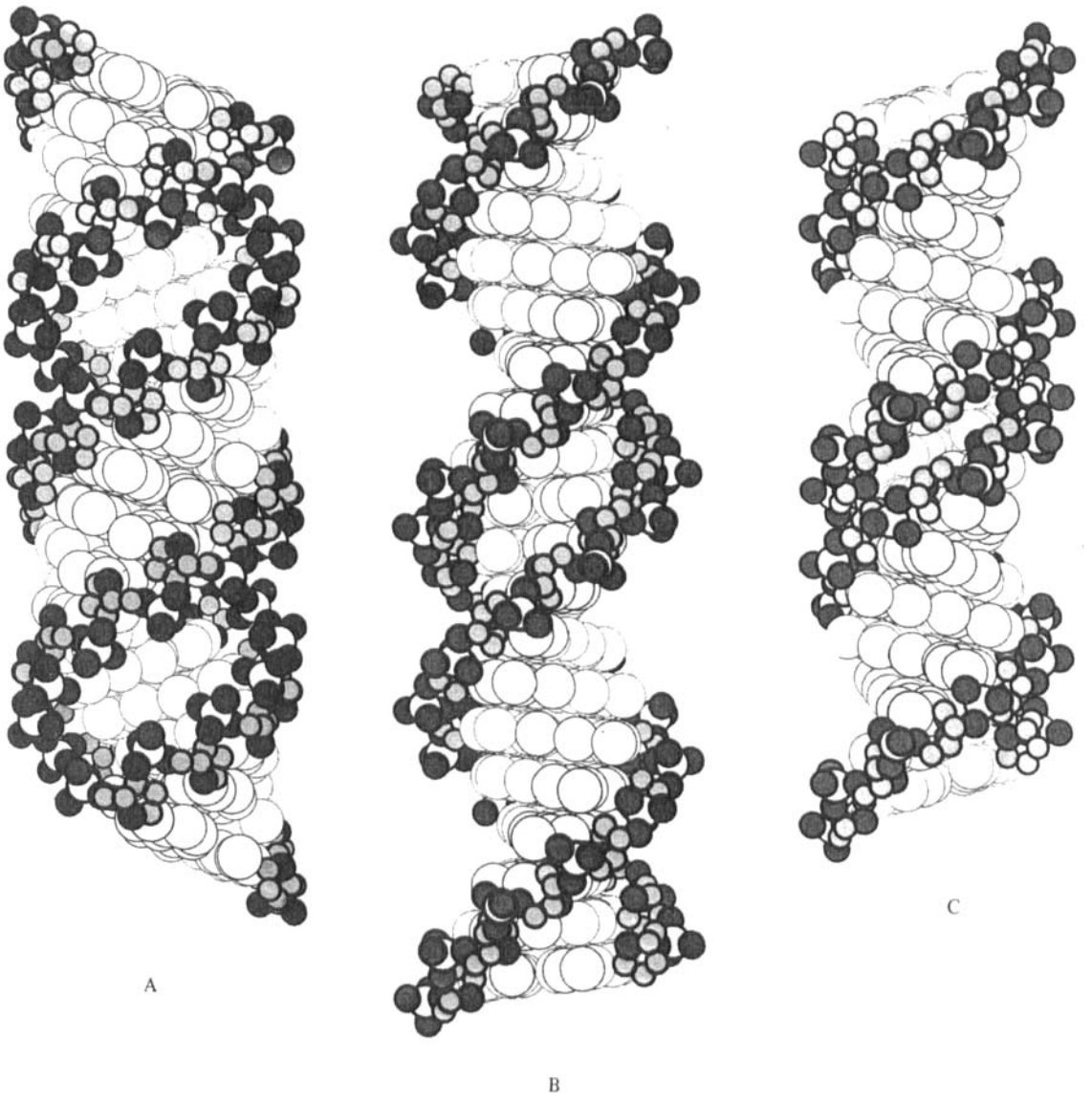

FIGURE 3. Helical projections for (A) *A*-DNA, (B) *B*-DNA and (C) *D*-DNA. Two helical turns are represented for each duplex conformation. All atoms of the bases are indicated as large circles, whereas atoms of the sugar-phosphate backbones are shown as smaller circles with carbon atoms smallest, oxygen atoms intermediate in size and shaded, and phosphorus atoms largest and unshaded.

content and hydration affect the allowable conformations of DNA molecules in fibers. At sufficiently low ion concentrations, only *A*-DNA is seen independent of the relative humidity, while at high salt contents only *B*-DNA is seen. As with relative humidity, increasing the ion content of a DNA fiber containing duplexes in the *A*-DNA form induces the *A*-DNA → *B*-DNA transition.

X-ray diffraction data from the *A*-DNA and *B*-DNA conformations have been analyzed extensively,[57-60] and well-defined molecular models for both of these helices are available.[61] The two forms are quite dissimilar in three-dimensional structure. The *B*-DNA helix has a rise per residue (h) (the distance between successive base pairs in the helix measured along the helix axis) of 3.38 Å with ten base pairs per turn of helix (turn angle per residue (t) equal to 36°), whereas *A*-DNA has h = 2.56 Å and eleven residues per turn (t = 32.7°) (Table 4). The shorter rise per residue in *A*-DNA reflects the positioning of the bases in the molecule. The base pairs are out from the helix axis and

TABLE 3

Effect of Humidity and Salt Concentration on DNA Conformations

Environment	Calf thymus DNA	$(dA)_n \cdot (dT)_n$	$(dG)_n \cdot (dC)_n$	Nonclassical DNAs
Low salt, all humidities	A	B′	A	D
Slight salt, low humidity	A	B′	A	D
Slight salt, high humidity	B	B′	(A + B)	(D)
Intermediate salt, low humidity	C	(B′)	(A + B)	(B)
Intermediate salt, high humidity	B	(B′)	(A + B)	B
High salt, all humidities	B	Three stranded	B	B

Note: DNA conformations are listed as their single letter prefixes. Parentheses imply that the listed conformation has not been extensively studied under these conditions and should be regarded as a tentative assignment. A + B means that the diffraction patterns show a mixture of A and B forms. Nonclassical DNAs are defined in the text. Calf thymus DNA should be considered representative of most DNAs having relatively random base sequences. Salt concentrations are difficult to characterize in fiber diffraction experiments. In general, low salt implies 0 to 2% excess Na^+ over the amount needed to counter the DNA negative charges in the fiber, slight salt is taken as 2 to 4% excess salt, intermediate salt is 4 to 6% excess and high salt is 6% excess and above. Humidities are more well defined; low humidity is taken as 0 to 75% relative humidity and high humidity is taken as 75 to 100% relative humidity.

TABLE 4

Structural Parameters of Polynucleotide Duplex Conformations

Helix conformation	Helix symmetry	Rise per residue (A)	Base pair parameters			
			Tilt(°)	Twist(°)	γ(°)	R(A)
A-DNA	11_1	2.56	19.3	−3.2	20.2	4.5
B-DNA	10_1	3.38	−5.9	−2.1	6.3	−0.2
C-DNA[a]	28_3	3.31	−8.0	1.0	8.0	−0.9
D-DNA	8_1	3.03	−16.0	5.6	16.4	−1.8
B′-DNA	10_1	3.29	−7.9	−1.0	8.0	−0.1
A-RNA	11_1	2.82	16.0	−6.9	17.4	4.4
A′-RNA	12_1	3.00	10.0	−7.6	12.5	4.5

[a]Values are for a representative *C*-DNA structure (see text).

Note: Selected conformational features of DNA and RNA double-stranded helices are listed. Helix symmetry is given as N_m meaning N residues per m turns of helix. Marvin et al.[64] describe tilt and twist angles in greater detail. Tilt is essentially the angle base pair planes make to the helix axis, while twist is the angle between base planes in a base pair (bases are not coplanar in a base pair, but are related like the blades of a propeller). Tilt and twist angles can be easily visualized in Figure 4. γ is a composite of tilt and twist and is described in the text. R is the x coordinate of purine C8 in the helix (as described in Reference 65) and represents the distance of base pairs from the helix axis. Positive and negative values of R denote opposite directions from the helix axis, as can easily be seen in Figure 4.

are skewed (Figure 4) so that a line perpendicular to the plane of any base in the duplex makes an angle (designated as γ)[62] of 20.2° with respect to the helix axis. On the other hand, in *B*-DNA the base pairs are positioned astride the helix axis and are more closely perpendicular to it (γ = 6.3°). In addition to differences in base positioning, the two duplex geometries exhibit differences in the conformations adopted by the furanose ring groups in the molecule. In *A*-DNA the sugar rings are puckered in the C3-*endo* manner (sugar ring pucker nomenclature is described in 63), whereas in *B*-DNA the rings adopt the C3-*exo* pucker.

All polynucleotide structures analyzed to date

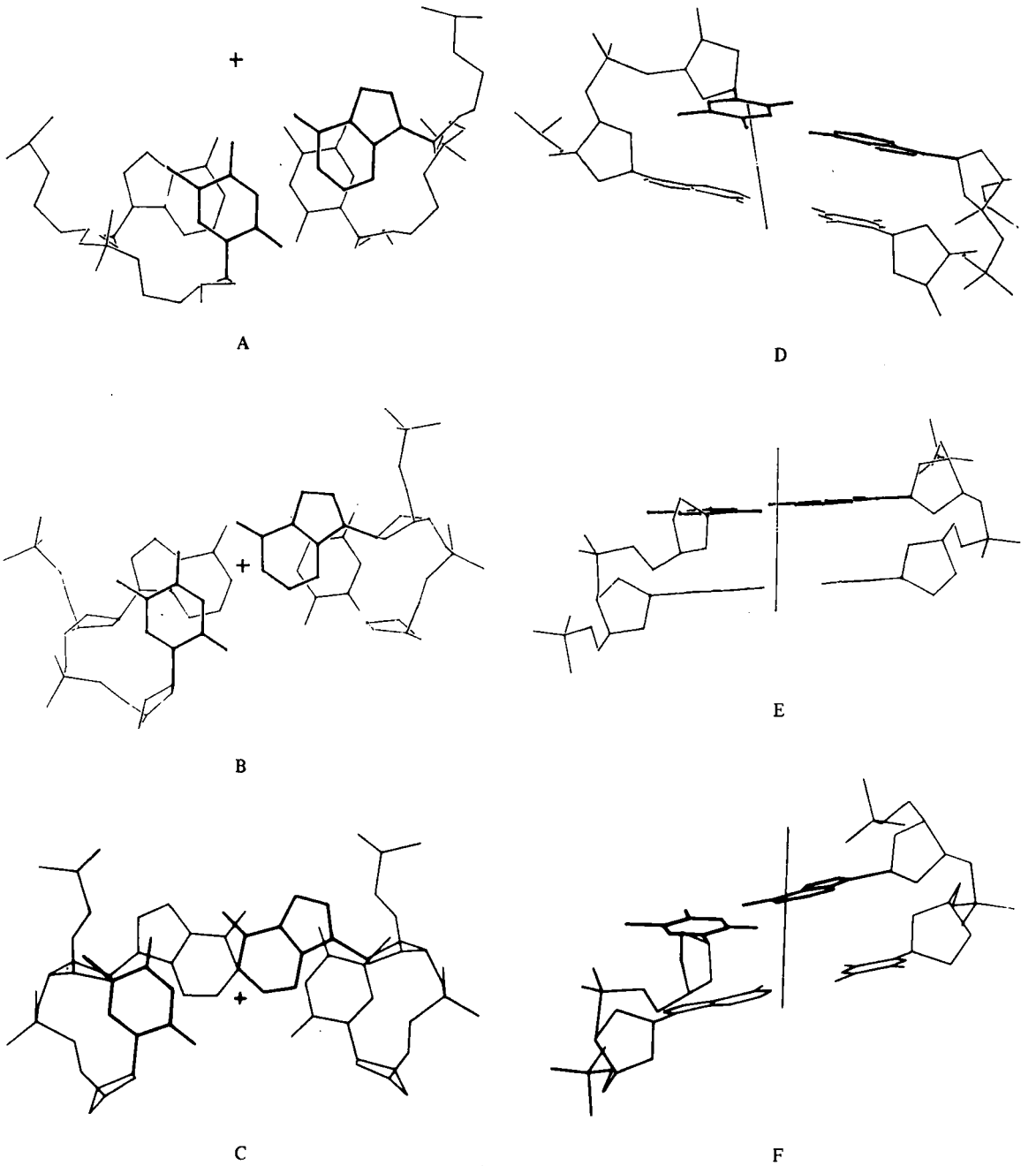

FIGURE 4. Dinucleotide residue projections for polynucleotide helix conformations. Two base pair residues are shown in projection both parallel and perpendicular to the helix axes for *A*-DNA (A and D), *B*-DNA (B and E) and *D*-DNA (C and F). The upper base pair is accentuated in each illustration for clarity. Helix axes in the projections are represented as crosses in A, B, and C and vertical lines in D, E, and F. (Courtesy of Drs. Andrew Leslie and Struther Arnott.)

(including one-, two-, three-, and four-stranded complexes of both RNA and DNA) can be conveniently classified according to sugar ring pucker into two separate families of conformationally similar helices.[62] One, designated as the *B* family and including *B*-DNA, has C3-*exo* (or very similar C2-*endo*) sugar ring puckers; the other, denoted as the *A* family, has C3-*endo* rings as in *A*-DNA. Interestingly, only DNA-DNA double-stranded helices have thus far been found to display *B* geometries. All complexes containing one or more RNA strands display *A* conformations.[66-73]

The relative base pair positioning in *A*- and *B*-DNA drastically affects the overall morphologies of the two helices (Figure 3). In *A*-DNA the major groove of the molecule is quite deep and the minor groove is shallow, whereas in *B*-DNA the major and minor grooves are relatively equal in size and depth.

At low relative humidities, DNA fibers having salt contents intermediate to those favoring *A*-DNA and *B*-DNA forms exhibit X-ray diffraction patterns similar to but distinct from *B*-DNA patterns.

Under these conditions, analysis of these patterns has indicated that DNA duplexes can adopt a series of closely related structures (collectively designated as *C*-DNA forms), all of which are essentially slightly distorted *B*-DNA helices.[64,65] *C*-DNA structures exhibit ranges of rises per residue (h = 3.31 to 3.4 Å) and turn angles per residue (t = 36 to 40°), but are all examples of *B* geometry, having C3-*exo* sugar rings and bases near the helix axis. Transitions from *C*-DNA forms to *B*-DNA can be affected by increasing the hydration of DNA fibers.

Structures of Synthetic Homopolymer Duplexes

While the *A*-, *B*-, and *C*-DNA conformations described above were originally observed with calf thymus DNA, studies have indicated that these conformations are common to all DNAs having relatively random nucleotide sequence and composition.[74] More recently, however, X-ray studies of DNA duplexes containing simple, repeating sequences indicate that certain of these helices can exhibit geometries different from the classical forms.

The synthetic homopolymer duplex $(dA)_n \cdot (dT)_n$ has been found to adopt only one conformation – a tenfold helix having a rise per residue of 3.29 Å.[75] This form, designated as B′-DNA, is extremely similar to *B*-DNA, with the bases being only slightly more skewed to the helix axis (γ = 8°). No *A*-DNA form has been found for $(dA)_n \cdot (dT)_n$ at low relative humidities or salt contents, whereas at high salt contents $(dA)_n \cdot (dT)_n$ fibers disproportionate to a triple-stranded helix, $2(dT)_n \cdot (dA)_n$, plus residual $(dA)_n$.[75] This triplex is conformationally similar to previously analyzed triplexes, $(A)_n \cdot 2(U)_n$, $(dA)_n \cdot 2(U)_n$[76] and $(A)_n \cdot 2(I)_n$,[77] since all display *A* geometry with C3-*endo* sugar rings and bases out from the helix axis. The bases of the "extra" strands in the complexes hydrogen bond to the adenine polynucleotides in a Hoogsteen manner.[75-77] Interestingly, the possibility of isogeometrical base triplets between adenine and thymine (or uracil) and between guanine and cytosine[76] implies that triple-stranded structures should be allowed between any polypurine and two polypyrimidines of appropriate sequence. The finding that $(dA\text{-}dG)_n \cdot (dC\text{-}dT)_n$ can form a triplex with $(C\text{-}U)_n$ and that DNAs with both purines and pyrimidines in both complementary strands do not form triplexes with their complementary RNAs strongly supports this contention.

In contrast to the single nonorthodox conformation of the $(dA)_n \cdot (dT)_n$ duplex, $(dG)_n \cdot (dC)_n$ complexes display only the classical forms, *A*-DNA and *B*-DNA.[78] However, it is noteworthy that, as opposed to the lack of any *A*-like geometry for $(dA)_n \cdot (dT)_n$, the $(dG)_n \cdot (dC)_n$ helix seems to prefer the *A*-DNA form. While a transition to *B*-DNA can apparently be effected at higher salt concentrations, only a partial transition occurs under conditions that would normally result in a complete conversion of calf thymus DNA from *A* to *B* forms.[78]

The homopolymer duplex $(dI)_n \cdot (dC)_n$ is chemically identical to $(dG)_n \cdot (dC)_n$ except for the C2 amino substituent on guanine, which results in an additional hydrogen bond between G:C base pairs. Interestingly, this duplex displays a geometry similar to that of $(dA)_n \cdot (dT)_n$ (B′-DNA) and does not behave structurally like $(dG)_n \cdot (dC)_n$.[166]

Nonclassical DNA Structures

Synthetic DNA duplexes with base sequences somewhat more complex than homopolymers have also been investigated by X-ray diffraction. Like the homopolymer helices, several of these have been found to demonstrate conformational

behavior distinct from natural DNAs. At high relative humidities and high salt contents, DNA fibers of the alternating copolymer duplexes $(dA\text{-}dT)_n \cdot (dA\text{-}dT)_n$, $(dG\text{-}dC)_n \cdot (dG\text{-}dC)_n$ and $(dI\text{-}dC)_n \cdot (dI\text{-}dC)_n$ all exhibit classical *B*-DNA geometries.[79,80] However, under low salt and low humidity conditions the *A*-DNA form is not observed. Instead, these duplexes adopt a novel helical structure, designated as *D*-DNA, which has only eight base pairs per turn (t = 45°) and a rise per residue of 3.03 Å.[28,29] *D*-DNA is clearly a member of the *B* family of nucleic acid structures (having C3-*exo* sugar rings), yet it displays a molecular architecture that is distinct from other *B* duplexes.[80] As in *A*-DNA, the base pairs in *D*-DNA are positioned out from and are skewed with respect to the helix axis. However, the *D*-DNA base positioning, relative to *B*-DNA, is opposite in sense to that seen in *A*-DNA (Figure 4).[80] This results in a helix morphologically different from either *A*- or *B*-DNA (Figure 3).

It should be noted that an *A*-DNA structure for $(dA\text{-}dT)_n \cdot (dA\text{-}dT)_n$ has been observed at low salt concentrations.[79] However, this form appears to be metastable and always undergoes a transition to a *D*-DNA form, which is apparently the preferred structure under these conditions.

For ease of discussion, the structural behavior of the alternating copolymer duplexes (*B*-DNA at high and *D*-DNA at low relative humidities) will be termed "nonclassical" behavior, in contrast to the "classical" behavior (*A*-, *B*- and *C*-DNA) of calf thymus DNA.

Recent studies indicate that the *D*-DNA helix may be more than an isolated structural anomaly of synthetic repeating trinucleotide DNA duplexes. $(dA\text{-}dA\text{-}dT)_n \cdot (dA\text{-}dT\text{-}dT)_n$, $(dA\text{-}dI\text{-}dC)_n \cdot (dI\text{-}dC\text{-}dT)_n$ and $(dA\text{-}dC\text{-}dT)_n \cdot (dA\text{-}dI\text{-}dT)_n$ all exhibit nonclassical structural behavior similar to the alternating copolymers described above.[81,167] On the other hand, the synthetic DNA duplexes $(dA\text{-}dG)_n \cdot (dC\text{-}dT)_n$, $(dA\text{-}dG\text{-}dT)_n \cdot (dA\text{-}dC\text{-}dT)_n$, $(dA\text{-}dG\text{-}dC)_n \cdot (dG\text{-}dC\text{-}dT)_n$, $(dA\text{-}dT\text{-}dC)_n \cdot (dG\text{-}dA\text{-}dT)_n$, and $(dA\text{-}dC\text{-}dC)_n \cdot (dG\text{-}dG\text{-}dT)_n$ display only classical behavior.[166]

It is noteworthy that, despite the nonclassical conformations exhibited by some DNA duplexes in fibers under certain conditions, all of the DNAs studied to date, with the exception of $(dA)_n \cdot (dT)_n$, can exhibit *B*-DNA geometries at high salt concentrations and relative humidities. Thus, extrapolating from fiber studies, this is the conformation that may be expected to be the time-average structure occurring in solution. Therefore, with the possible exception of $(dA)_n \cdot (dT)_n$ which may adopt the B' structure in solution, many of the anomalous solution properties of certain synthetic DNAs may not be attributed to differences in conformation. However, sequence-dependent conformations, like *D*-DNA, may be relevant in vivo when DNAs enter specialized environments, such as during interaction with proteins. Also, at least one helical conformation different from those mentioned above was demonstrated by NMR at the AT-GC junction of a synthetic block polymer (see below). Thus, different types of helices in one polynucleotide duplex may be mutually perturbing, generating a third conformational type.

In view of the results described above, it should be mentioned that DNAs having unusual base compositions, such as greater than 66% A,T, but also having more random base sequence (i.e., no homopolymer or alternating copolymer tracts) than the synthetic DNAs mentioned, apparently demonstrate only classical *A*-DNA and *B*-DNA structures.[74,82-84] Nonetheless, unusual base compositions do seem to affect DNA structural behavior, since A,T-rich DNAs are reluctant to undergo the $B \rightarrow A$ transition. This effect has been seen with both synthetic DNAs[22,26,30] and with natural DNAs isolated from organisms such as *Clostridium perfringens* and *Cytophaga johnsonii*, with A,T-rich genomes.[85]

Tendencies for small tracts of a DNA duplex to favor either *A*-DNA (G,C-rich tracts) or *B*-DNA (A,T-rich tracts) forms may be biologically relevant during initiation and termination of DNA transcription and replication. Enzymes with different affinities for the *A* and *B* geometries of DNA, the preference of RNA·DNA hybrid complexes for *A*-like forms, and the forementioned conformational tendencies of DNA duplexes with differing base compositions could participate in these processes without the necessity of precise sequence recognition sites. A scheme based on such conformational arguments has already been put forth for steps in the replication of ϕX174.[86] In accord with the notion of the role of DNA conformation in gene regulation, without precise sequence homology, is the presence of alternating blocks of A·T-rich and G·C-rich regions near certain control elements (see Section V).

IV. DOES NATURAL DNA CONTAIN STRUCTURAL ANOMALIES?

As an outgrowth of our studies on DNA polymers of defined nucleotide sequence, we wanted to determine if natural DNA chromosomes contain small segments of structural irregularities. The hypothetical contrasting view would be that every base pair in natural DNA is in the same conformation (presumably *B*-DNA).

The *Lac* Operator-repressor System

The problem of ascertaining DNA structural irregularities in high molecular weight genomes is formidable. DNA structural irregularities must be present in very small quantities in high molecular weight chromosomes since composite measurements (such as buoyant density measurements or X-ray diffraction) do not reveal these features[19,20] (see Section III). Thus, highly specific and sensitive probes must be used; very few suitable probes currently exist. We reasoned that if DNA structural anomalies exist in small quantities in natural chromosomes, this was consistent with their presence at regulatory sites. The interaction between the *lac* repressor and *lac* operator containing DNA was chosen because of the high degree of specificity and tightness of binding of this protein to the DNA, as well as the apparent simplicity of the interaction (no translocation or polymerizing functions as with RNA polymerase). The single-strand-specific nucleases, mung bean nuclease and S_1 nuclease, were chosen as probes because of their pronounced specificity for random coil compared to helical DNA. In addition, chemical probes were used (see below).

The structural uniqueness of the lactose operator was demonstrated[13,87] by the ability of single-strand-specific nucleases to influence *lac* repressor binding. λp*lac* DNA was treated with various concentrations of nucleases, and the modified DNA was analyzed for its *lac* repressor binding activity as well as the number of nicks, as judged by alkaline sucrose gradient sedimentation or by gel electrophoresis.

The repressor binding capacity of λp*lac* DNA was maximally reduced by as few as 2 to 5 cuts per molecule by either S_1 or mung bean nucleases. These nucleases are highly specific for random coil DNA. However, approximately 300 cuts per molecule by any three less specific nicking agents (micrococcal nuclease, pancreatic DNase or sonication) were necessary to produce the same effect. In addition to the repressor-operator equilibrium measurements, the same effect has recently been demonstrated with measurements of the rates of association and dissociation.[13] Also prior binding of the *lac* repressor, before treatment with the nuclease, protected the operator from selective cleavage. The dissociation kinetics showed that mung bean nuclease damaged 30 to 40% of the operators.

Mapping experiments were recently performed[13] on λp*lac* DNA treated with various amounts of mung bean nuclease by cleaving the DNA with *Hin* II and III and then fractionating the fragments by polyacrylamide gel electrophoresis. Statistical analysis of the loss of duplex restriction fragments as a function of nuclease concentration demonstrated that the *lac* operator fragment (789 base pairs) was not uniquely sensitive. However, when the fragments were analyzed on denaturing gels, a specific nick was revealed at approximately 100 nucleotides from the end of the 789 base-pair fragment. The amount of nicked fragment was commensurate with the extent of loss of repressor binding. Other studies showed that this fragment was uniquely nicked compared to other fragments of approximately the same size.

These studies clearly demonstrate that a DNA region near the final position of binding of the repressor is uniquely sensitive to single-strand-specific nucleases. The reason for this sensitivity is unknown at present, but is currently being studied by other techniques. In addition, this work demonstrates the necessity of "outside" regions of DNA for maximum repressor binding activity since DNA containing several random nicks is not a suitable substrate.[13] This notion is also buttressed by the finding that repressor binding was reduced after treatment of λp*lac* DNA with three different DNA restriction endonucleases.[13] The extent of reduction was a function of the molecular weight of the operator-containing fragment, e.g., the smaller the operator-containing fragment, the lower the binding constant.[13]

The structural peculiarity which is recognized by the single-strand-specific nucleases is unknown at present, but is currently being studied by several other techniques. However, it has recently been argued[13] that neither model A nor B (thermolability or cruciform models, Figure 1) adequately explains these data. In addition, even though the *lac* operator region can be written into a two-

dimensional "loop-back" structure, it is unlikely that the nucleases specifically cleave this configuration since the DNA is cut at least 100 base pairs away from this sequence.

Three lines of evidence[13,88] indicate that the unique feature of the *lac* operator is lost when a few cuts are introduced at other loci on the DNA chain. It is not known whether a particular nucleotide sequence of composition is required for the maintenance of the unique feature. Similarly, the necessary minimum length of the adjoining sequence is uncertain at the present time.

Number of Nonpaired Nucleotides Necessary for Cleavage by S_1 or Mung Bean Nucleases

Two of the models (Figure 1) invoked to explain the unusual DNA structure near the lactose operator contain nonpaired nucleotides. Recent work has provided information on the substrate structural requirements for these nucleases. In addition to the use of single-strand-specific nucleases to probe the *lac* operator, these nucleases have also been used to detect regions of instability in supercoiled DNA.[89,90] Furthermore, from studies on genetic heteroduplex (wild-type hybridized to temperature-sensitive mutant) SV40 DNA molecules, Shenk et al.[91] have suggested that S_1 nuclease may be useful in the isolation of specific blocks of DNA bounded by single-base mismatches.

To examine S_1 and mung bean nuclease action on more defined substrates, a variety of defined heteroduplex regions were synthesized flanked by blocks of G-C base pairs.[92] Using calf thymus terminal transferase and *Escherichia coli* polynucleotide phosphorylase, the oligonucleotides of the form $d(C_{12}A_xC_{\bar{y}})$ (Figure 5) were synthesized where x = 1, 2, 3 . . . 6 and y = (y − 5, . . . y + 5). y was between 10 and 30 in different preparations. These oligomers were annealed to $(dG)_n$ to give dA·dG heteroduplex regions of one to six base pairs. The products of S_1 or mung bean nuclease action on these defined heteroduplexes were analyzed by electrophoresis on 20% polyacrylamide gels.[93] Heteroduplexes three or more base pairs in length were cleaved by high concentrations of both nucleases as judged by the appearance of $dC_{12}A_{0,1...x}$ peaks on the analytical gels (Figure 6). The heteroduplex regions one and two bases in length were weakly attacked by high concentrations of the nucleases, as judged by a decrease in the amount of substrate oligomer, and very low

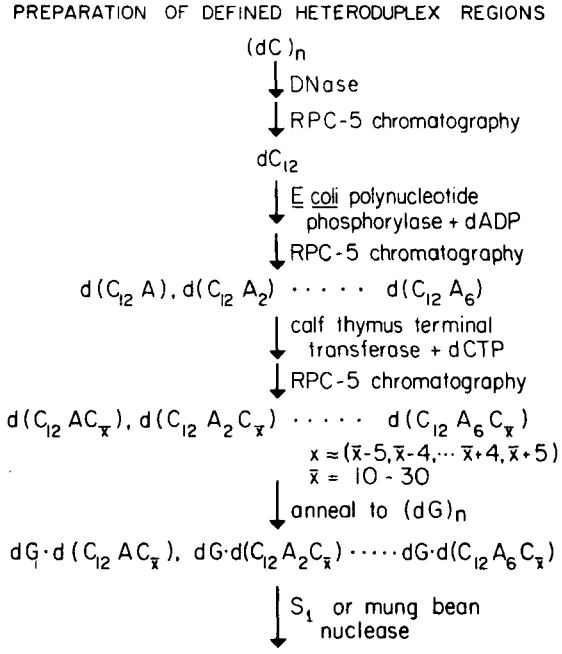

FIGURE 5. Scheme of synthesis of defined heteroduplex DNAs of the form $(dG)_n \cdot d(C_{12}A_mC_{\bar{x}})$. The oligomer dC_{12} was purified from a pancreatic DNase digest of $(dC)_n$ by RPC-5 chromatography.[149] The dC_{12} was extended from the 3′-hydroxyl end with *Escherichia coli* polynucleotide phosphorylase and dADP.[163] After purification of the individual $d(C_{12}A_m)$ product oligomers, a block of dC residues was added to the 3′-hydroxyl end of each oligomer with calf thymus terminal transferase.[149] In each case, the product oligomers were purified and annealed to $(dG)_n$ to give a series of DNAs with regions of dA·dG mismatched base pairs of defined length. These were tested for sensitivity to S_1 nuclease by analytical gel electrophoresis.[93]

levels of product oligomer were seen on the gels. Thus, the cleavage of the single or double heteroduplex sites occurred only at very low levels, similar to the low levels of "nibbling" (i.e., cleavage from the 5′ and 3′ ends of the oligomers). Similar results have been obtained with a series of dG·dG heteroduplexes one, three, four, and five nucleotides in length.[92]

These results are subject to two major qualifications. The first is that heteroduplex regions surrounded by dG·dG blocks had to be used due to the nuclease susceptibility of the homoduplex $(dA)_n \cdot (dT)_n$. While the results appear to be unaffected by the changes in the overall stability of the oligomer-polymer complex resulting from changing the length of the 3′ C stretch ($\bar{y}$ above), it is not certain that heteroduplexes would behave

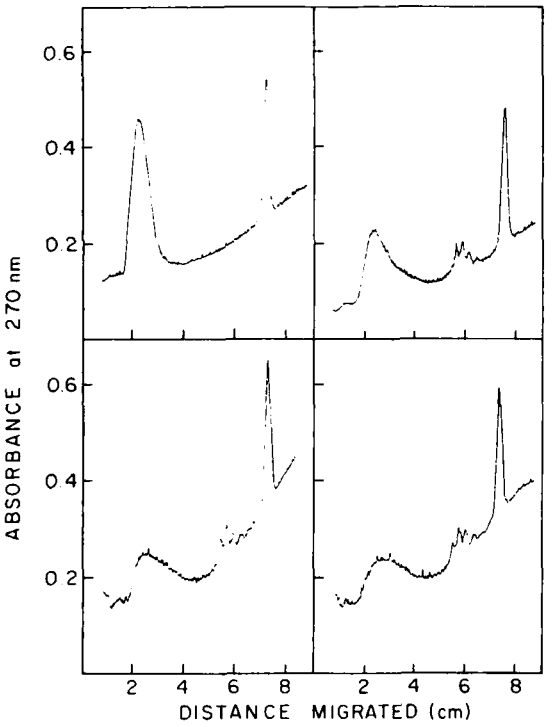

FIGURE 6. Reaction of S_1 nuclease with $(dG)_n \cdot d(C_{12}A_5C_{\overline{29}})$. 0.34 A_{250} units of $(dG)_n \cdot d(C_{12}A_5C_{\overline{29}})$ were incubated in a 0.34-ml reaction mixture containing 0.03 *M* sodium acetate buffer, pH 4.6, 5% glycerol, 0.05 *M* NaCl; 1.5 m*M* $ZnSO_4$ and 0 (upper left), 34 (upper right), 85 (lower left) and 850 (lower right) units of S_1 nuclease.[92] Incubations were for (30 min at 37° and were terminated by making the mixture 8 m*M* in EDTA and 15 m*M* in tris base. 0.95 A_{270} units of $(dC)_n$ were added and the mixture was boiled for 3 minutes and then cooled on ice. After the addition of sucrose and 0.04 A_{270} units of dC_9, the solution was electrophoresed on 20% polyacrylamide gels, and the gels were scanned for absorbance at 270 nm.[93] Electrophoresis is from left to right; the sharp peak on the right side of each panel is the $(dC)_9^9$ marker oligonucleotide. Oligonucleotides produced by cleavage of the heteroduplex DNA appear between the uncleaved oligonucleotide peak (upper left panel) and the $(dC)_9$ marker. Note especially the four product peaks in the lower right-hand panel, whose mobilities correspond to $d(C_{12}A_3)$, $d(C_{12}A_2)$, $d(C_{12}A)$, and $(dC)_{12}$ (from left to right).

similarly when situated in blocks of complex DNA sequences. Furthermore, while the results of dA·dG and dG·dG heteroduplexes appear quite similar, other heteroduplex combinations may behave differently. (Only certain heteroduplex combinations can be examined in our system due to the constraint of using surrounding blocks of dG·dC homoduplex.)

With the above reservations in mind, these results influence the interpretations of previous experiments using single-strand-specific nucleases as structural probes. If the region of single-strand nuclease susceptibility near the lactose operator[13,87] results from unpaired bases, its unusual sensitivity suggests that several unpaired bases would be involved. Since the results of iodination studies (discussed below) indicate that there can be few unpaired cytidine residues, the combined results of our experiments suggest that the sensitive site may involve an unusual duplex structure (Figure 1C).

These results also suggest that the nuclease-sensitive sites in supercoiled DNAs[89,90] may involve several unpaired bases. We cannot be sure whether these would be transiently unpaired or stably unpaired, as suggested by Dean and Lebowitz.[94]

Finally, these results suggest that the use of single-strand-specific nucleases for isolation of a block of DNA between two single base-pair heteroduplex sites would be rather difficult to accomplish with reasonable yield. Recently, Dodgson et al.[168] tested such an experiment by isolating the 789 base pair *Hin* II and III fragments containing the lactose promoter-operator[15] isolated from λ*plac* 5 or λ (*lac-trp*) fusion strains that were either of the *wild type* or contained a single base mismatch (or single base deletion) in the promoter region. The *lac* promoter mutations studied were L305, L241, and p^r1a. The promoter regions of both the wild type and the mutant strains have been sequenced[15] so the nature of the mutant site was clearly defined. Three different heteroduplexes arising from the cross hybridization of two base substitution mutants and one base deletion mutant with the *wild-type* promoter fragment were tested for sensitivity to S_1 and mung bean nucleases. No observable cleavage at the heteroduplex site was observed when analyzed by polyacrylamide gel electrophoresis. While the nature of the *lac* fragment heteroduplex system would preclude the detection of cleavage if less than about 10% of the initial heteroduplex fragments were cut, these experiments certainly suggest that this reaction would not normally be useful in isolating defined gene fragments. Shenk et al.[91] observed that these results do not necessarily imply that the rather low percentage of cleavage is due to other than single base-pair heteroduplexes.

Chemical Probes for Determining if the *lac* Operator Contains Nonpaired Nucleotides

The three models (Figure 1) invoked to explain

the structural uniqueness of the *lac* operator differ since models A and B contain nonpaired nucleotides, whereas model C is helical but contains other structural irregularities. Attempts have been made, using iodination as a selective probe for nonpaired nucleotides, to further distinguish between these models.

Conditions were established where the iodination (catalyzed by thallium chloride) of heat-denatured DNA proceeds linearly (for several hours at 40°) at a rate approximately 150 times the rate with native DNA.[169] Kinetic studies were performed with both helical and random-coil DNA at 40°, 50° and 60°C; the presence of 5 m*M* $MgCl_2$ substantially enhanced the selectivity for random-coil DNA.

This technique was used to determine if the *lac* operator contained nonpaired nucleotides. Intact λp*lac* DNA was iodinated with ^{125}I under conditions which should cause selective reaction with nonpaired nucleotides. The labeled DNA was degraded with *Hin* II and III restriction endonucleases, and the specific radioactivity of the operator-containing fragment (789 base pairs)[95] was compared with the specific activity of 15 other fragments. No difference in specific activity was observed under any experimental condition tested. If as few as three dCMP residues in the operator-containing fragment were in a nonpaired state, this fragment would have been iodinated with a specific activity twice that of the average fragment. Of course, this negative result depends on the rate of iodination of those dCMP residues which are equal to that in nonhelical DNA. Hence, within this level of sensitivity, it is concluded that the thermolability and cruciform models (Figure 1) are less likely to exist than the helical non-*B*-DNA model.

Studies have been performed to determine the minimum number of nonpaired nucleotides which are detectable by these iodination procedures. A heteroduplex was constructed between φ80 p*lac* 1 DNA and the same genome containing the L_1 deletion. The sequence deleted in the mutant DNA is known to be 73 base pairs (38 are GC pairs), and the deletion extends from the carboxyl terminal end of the i gene into the promoter of the *lac* operon.[170] φ80 p*lac* 1 DNA and φ80 p*lac* 1 DNA containing the L_1 deletion were restricted with *Hin* II and III. The operator-containing fragments were isolated by binding to the *lac* repressor.[96] Subsequently, the hetero- and homoduplexes were prepared according to established procedures.[97] The three possible duplexes were well separated by polyacrylamide gel electrophoresis, and the heteroduplex was identified by its extreme sensitivity to single-strand-specific endonucleases.

When subjected to the iodination conditions described above, the heteroduplex incorporated 12 times as much iodine as the corresponding homoduplexes. Assuming that all of the excess iodine is located in the single-stranded "loop" of the longer strand of the heteroduplex, this indicates a 200- to 210-fold preferential iodination of the single-stranded fragment over double-stranded DNA on a nucleotide residue basis. Thus, approximately 2 to 3 nonpaired dCMP residues in the operator fragment of λp*lac* DNA should have resulted in a doubling of the specific radioactivity of this fragment compared to other fragments and, thus, should have been easily detectable. However, if the base pairs in models A or B (Figure 1) are nonhydrogen bonded a small fraction of the time, they cannot be eliminated. Such structures may have been recognized by the single-strand-specific nucleases, since they are substantially more specific for nonpaired nucleotides than the iodination procedure.[98]

Do "Single-stranded" Viral DNAs Contain Double-helical Regions?

The genomes of the small icosahedral (φX174, G4, and S13) and filamentous (M13, fd, and f1) bacteriophages are single, circular polynucleotide chains of DNA. These "single-stranded" viral DNAs may at first be considered as random coils, seemingly inappropriate in a review on DNA structure and genetic regulation. However, considering the extensive secondary structure of single-stranded RNAs (for example phage MS2 RNA),[9,99] it should not be unexpected to find structure in these viral DNAs. Recent results indicate that "single-stranded" viral DNAs indeed possess double-helical regions; hence, they may have interesting properties, as discussed above.

Restriction endonucleases can be used as probes for duplex structure in DNA.[100] Most restriction enzymes are endonucleases that recognize and cleave twofold symmetrical sequences of four to six nucleotides in duplex DNA.[101] Using the restriction reaction as an assay for helical regions in a DNA molecule is justified since only duplex DNA is cleaved. *Haemophilus aegyptius* III restriction endonuclease readily cleaved λp*lac* 5

DNA into more than 50 fragments. However, no cleavage was observed when the DNA was heat denatured prior to incubation with the enzyme.[100] Oligonucleotides containing appropriate restriction enzyme recognition sequences for either *Haemophilus parainfluenzae* II[102] or *Escherichia coli*[103] R_I were cleaved only under conditions where the complementary oligonucleotides were in a duplex form. Also, the restriction reaction is sensitive to perturbants of duplex structure (see below).

Under standard conditions, *Haemophilus aegyptius* III cleaved the viral DNA of ϕX174, M13, and f1 into specific fragments. Figure 7 is a densiometric scan of stained *Haemophilus aegyptius* III-ϕX174 viral DNA fragments electrophoresed on polyacrylamide gels under denaturing conditions. By comparison, these fragments are identical in size to those obtained from a *Haemophilus aegyptius* III digestion of the double-stranded replicative form (RF) DNA. Thus, there was site-specific cleavage of the viral DNA. Similar results were obtained for the viral and RF DNAs of M13[100] and f1.[104]

That the duplex-specific restriction enzyme was cleaving only duplex sequences in these viral DNAs was confirmed by studying the effect of perturbants of helical structure on the reaction.[105] Temperature affected the initial rate of reaction of *H. aegyptius* III with ϕX174 viral and RF DNAs (Figure 8, note the difference in the scales of the ordinate axes). When the incubation temperature was raised above a certain critical temperature (57°), cleavage of the viral DNA no longer occurred. However, as evidenced by the reaction with RF DNA, *H. aegyptius* III remained active to at least 72°. Thus, increasing the reaction temperature inactivated the viral DNA substrate by destroying duplex regions. Analysis of the products of these reactions by polyacrylamide gel electrophoresis revealed a differential thermal stability between cleavage sites. Thus, there are several different duplex regions within the ϕX174 viral DNA.

Actinomycin, a drug which perturbs duplex DNA structure by binding at GC pairs,[30] inhibited the *H. aegyptius* III cleavage of ϕX174 viral and RF DNAs.[105] If the viral DNA did not contain

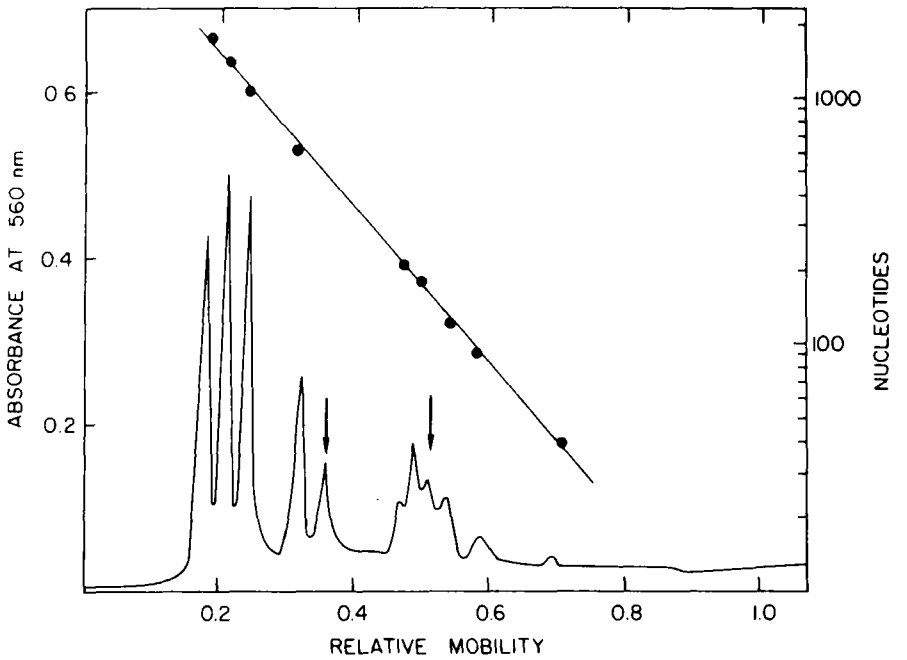

FIGURE 7. Densiometric scan of *Haemophilus aegyptius* III fragments of ϕX174 (+)-strand DNA separated by formamide-polyacrylamide gel electrophoresis. ϕX174 (+)-strand (–) and RF (o) DNAs were incubated under standard conditions, and the products were electrophoresed in parallel 5% polyacrylamide gels containing 98% formamide and 20 m*M* sodium phosphate (pH 7.5). The bands were stained with Stain's All and scanned. The RF fragment peaks are plotted against the logarithm of their molecular weights. For details, see Reference 100. (From Blakesley, R. W. and Wells, R. D., *Nature*, 257, 421, 1975. With permission.)

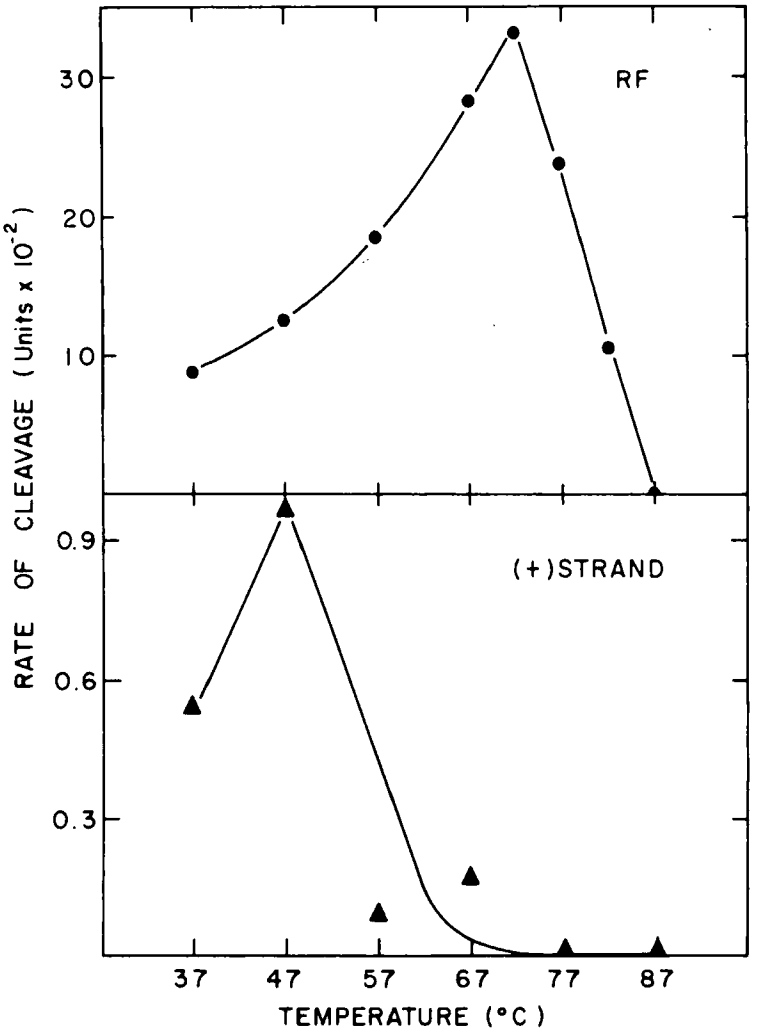

FIGURE 8. Effect of temperature on cleavage of ϕX174 (+)–strand and RF DNAs. ^{32}P-labeled ϕX174 (+)–strand or RF DNA was incubated with *Haemophilus aegyptius* III in the standard reaction mixture[100] at various temperatures. The initial rate of cleavage at each temperature was measured by following the disappearance of RF I, II, and III DNAs into smaller fragments. Aliquots were removed from the reaction at 30-sec intervals during the 3-min period following enzyme addition. Quantitation of the unrestricted DNA was essentially as described.[103] The units are defined as the microgram of unit lenth DNA degraded per hour per milliliter of *Haemophilus aegyptius* III.

duplex structure at the restriction sequences, actinomycin should not bind and no interference in the restriction reaction would be expected. Therefore, the observed inhibition of cleavage of ϕX174 viral DNA was a result of the binding of actinomycin to cleavage sites within duplex regions of DNA.

There are other indications of structure in "single-stranded" viral DNAs. Early studies[106] suggested the presence of a hairpin structure based on the inability of an exonuclease to totally degrade the DNA. More recently, Schaller et al.[107] digested fd DNA with the single-strand-specific nucleases *E. coli* exonuclease I and *Neurospora crassa* endonuclease to identify possible helical regions. A nuclease-resistant core was isolated, comprising about 2% of the total viral DNA. The core was 40 nucleotides in length, G + C rich, double-strand-like in its spectra, and rapidly renaturable. Bartok et al.[108] also obtained a resistant DNA core by treating ϕX174 viral DNA with *N. crassa* endo- and exonuclease. Again, the

core comprised about 2% of the viral genome and was characterized as double-strand like. In contrast to the fd DNA core, the two sequences of the ϕX174 DNA core, 16 and 24 base pairs in length, were A,T rich. In both cases the cores are presumed to be hairpin structures, apparently with enough stability to prevent further cleavage by these nucleases. The cores are believed to originate from unique points in the viral DNA sequence and may have some important biological function. One hairpin has been tentatively mapped in the *H. aegyptius* III restriction fragment 3, which contains an RNA polymerase promoter.[108]

The design of these nuclease experiments has only allowed for the detection of the longer, more stable duplex or hairpin structures in the viral DNA. From the restriction endonuclease results, it seems certain that there are several duplex regions in the DNA, each with a different degree of stability. Digestion with the single-strand-specific T4 endonuclease IV[109] has permitted the detection of less stable structures by using low temperatures and high ionic strength. At least ten site-specific fragments were obtained from ϕX174 viral DNA. However, when the temperature was increased and the ionic strength lowered, these fragments did not appear, but total digestion was achieved.

Another approach to detect structural features in the viral DNAs was to protect specific regions of the DNA from nuclease digestion. In the absence of unwinding protein, *E. coli* RNA polymerase binds to several places on fd viral DNA. However, the polymerase can be directed to bind only to the origin of replication, if the single-stranded portions of the DNA are first covered with *E. coli* unwinding protein. Nuclease digestion of this complex yields a protected fragment with double-strand-like properties. As opposed to the non-specific polymerase binding sites, this fragment is resistant to the single-strand nucleases S_1 and exonuclease I.[110]

Another site-specific protection experiment was the binding of *E. coli* ribosomes to f1 or ϕX174 DNA.[111] A DNA fragment was protected from pancreatic DNase digestion, then isolated and sequenced.[112] Although the sequence can be arranged to form a hairpin loop,[112,113] no evidence for duplex structure was obtained.

That helices, hairpins, and/or unique structures exist in the "single-strand" viral DNAs should be no surprise when one considers the single-stranded genomes of the RNA bacteriophages.[99,114] Approximately 75 to 85% of the RNA of MS2 and R17 is in a base-paired or stacked arrangement.[115,116] Fiers et al.[9] combined low-temperature nuclease digestion patterns with the most thermodynamically stable arrangement of nucleotides (based on the complete primary sequence) to draw an extensively folded and base-paired model. Confidence in accepting this structure in the RNA phage molecules is based on a wider variety of analytical procedures (e.g., References 114–119).

The most important but most speculative aspect of the structural features in these viral genomes is their correlation with a biological function. A hairpin structure was postulated[11] in a model of DNA replication for the circularization of nascent linear DNA molecules. Localization of the RNA polymerase protection fragment at the origin of replication[110] gives support to the presence of such a hairpin. In addition, Eisenberg et al.[120] found a single gap in the viral strand of the replicative intermediate, occurring near the origin of replication. Such a gap would be uniquely located when RNA polymerase creates an RNA primer at its binding site at the origin of replication.

The use of DNA restriction endonucleases as sensitive probes of helical segments promises to provide new means for assessing biological function of regions in "single-stranded" DNA genomes.

Structural Peculiarities in Supercoiled DNAs

As briefly discussed above, there is a large body of evidence suggesting that unique structural sites occur in superhelical closed circular DNA molecules, at least in those of fairly high superhelix density. Whereas space considerations preclude an extensive review of this work, the nature of the evidence for unique structural sites is briefly summarized below. Much of this evidence involves the use of single-strand-specific nucleases to detect sensitive sites in supercoiled DNA that are absent in untwisted DNA. Such sites have been found in ϕX174 RF I,[90,121] SV40 DNA,[89,122] polyoma DNA,[123] PM2 DNA,[124] and M13 RF.[125]

These nuclease-sensitive regions have also been investigated by other techniques designed to detect either transiently or stably unpaired regions of DNA. These techniques include the use of single-strand DNA binding protein coded for by T4 gene

32,[126,127] the binding of complementary oligodeoxynucleotides,[128] and a variety of chemical probes such as carbodiimide and formaldehyde.[94,129-132]

The nature of the site of action of these probes in supercoiled DNA is still uncertain. Chen et al.[132] argue that the sites involve hairpin loops in the two DNA strands composing a palindromic region of the DNA duplex or some other stably unpaired block of DNA. Hsieh and Wang argue against such a proposal from a thermodynamic aspect,[133] the inability of supercoiled PM2 to initiate DNA synthesis in the presence of oligonucleotide primers, and an examination of the rate of action of mung bean nuclease on supercoiled PM2 DNAs.[124] Whereas recent upward revisions in the free energy of supercoiling have negated the thermodynamic arguments, it is not yet possible to determine the actual unique structure(s) that results in the observations summarized above.

V. SEQUENCE IRREGULARITIES IN DNA

The nucleotide sequences of a number of regions of procaryotic and eucaryotic genomes have been elucidated including several operators, more than a dozen promoters and several structural genes (reviewed in part in References 136–138). Several types of symmetry have been revealed in these sequences and palindromic elements in particular were discussed by the original authors of the sequencing papers as possible sites of interaction with regulatory proteins.

Whereas nucleotide sequence data provide new insights from both biochemical and genetic standpoints, they do not provide information on the three-dimensional structure of the DNA. Even though several DNA sequences can be drawn into a reasonable two-dimensional "loop-back" (or cruciform) structure, this does not prove that a conformation of this type actually exists in high molecular weight DNA. Hence, elucidation of a DNA sequence provides little or no information on conformational problems.

A few sequences of selected regulatory sites are presented in Table 5. The list is not intended to be complete; however, the sequences were chosen to point out the occurrence of contiguous blocks of AT followed by GC pairs. That DNAs containing contiguous blocks of AT and GC pairs may have unusual properties is discussed in Section VI.

Table 5 shows that numerous examples of "block-polymer"-like structures exist at or near genetic control regions. Some of the runs of nearly pure AT (or GC) regions are quite long, more than two turns of DNA helix. For example, the SV40 sequence has a 100% AT region 17 base pairs long joined to an 83% GC region 24 base pairs long; the *trp* attenuator contains a "five-block" region with the longest an 83% AT region 30 base pairs long; and a sequence near the leftward λ operator-promoter has a 96% AT region 23 base pairs long joined to a 90% GC block 10 base pairs long joined to an 86% AT block 7 base pairs long.

An additional feature of note is the bias for $(dpur)_m \cdot (dpyr)_m$* type sequences within some of these blocks. In the 17 base pair AT block of the SV40 sequence, 76% of the As are on one strand. In the 96% AT block of the sequence near the leftward λ operator, 77% of the As are on one strand and in the contiguous 90% GC block, 78% of the Gs are on one strand. The unusual properties of $(dpur)_n \cdot (dpyr)_n$ polymers were discussed in Sections II and III.

Dykes et al.[138] previously determined the probability of occurrence of a number of types of symmetries and other sequence irregularities. They conclude that the expected frequency of occurrence of adjacent alternating regions of high GC and AT blocks (three or more) is as rare or rarer than palindromes. Table 5 shows several sequences, published after the Dykes et al. calculations, which have as many as seven contiguous blocks; also, the longer stretches of AT (or GC) blocks would render these sequences even more improbable. In addition, if the $(dpur)_m \cdot (dpyr)_m$ bias is considered, this type of sequence becomes an even rarer event. On the other hand, Dykes et al.[138] calculate that the probability of occurrence of many types of palindromic sequences is quite high.

Thus, it may be useful to devote more attention to the presence of alternating AT and GC blocks in natural DNA sequences as they become available. The possible biological role of these types of sequences is discussed in Section VI.

VI. DYNAMIC STRUCTURE OF DNA

In addition to the studies described above on

*pur, purine; pyr, pyrimidine.

TABLE 5

Occurrence of Contiguous AT- and GC-rich Regions in Nucleic Acid Sequences

Source of DNA sequence	Sequence (5′ → 3′)	Blocks (% of pure AT or GC in brackets)							Reference
		AT	GC	AT	GC	AT	GC	AT	
Portion of promoter for *Escherchia coli* tyrosine tRNA	GCG_4CGC: $ATCATATCA_3TGA$: GGCGCCGC		7/7 (100)	11/14 (79)	8/8 (100)				139
Portion of promoter for *lac* operon	$GGCAC_4AGGC$: T_3ACACT_3AT: GCTTCCGGCTCG		10/12 (83)	10/12 (83)	9/12 (75)				15, 140
Sequence near origin of DNA replication for SV40	$TA_2T_8AT_3AT$: $GCAGAG_2C_2GAG_2$ $C_2GC_2UCG_2C_2$			17/17 (100)	20/24 (83)				141
Portion of attenuator between operator and first structural gene of *E. coli* tryptophan operon	ATTGTTATTCTCT AAT_4GTTCA_8:GC_3 GCG_3CGC:TCATTA: $GGCTG_3$:TATCTGAT $TGCT_3A$	25/30 (83)	12/12 (100)	5/6 (83)	6/7 (86)	11/15 (73)			142
Portion of rightward operator-promoter of λ	AACCATTAT:CACCGC CAGAGG:TA_4TAGTC AA:CACGCACGGTG: $TTAGATAT_3AT$:C_3T $TGCGGTG$:$ATAGAT_3$ A_2	7/9 (78)	9/12 (75)	10/12 (83)	8/11 (73)	11/12 (92)	8/11 (73)	9/10 (90)	143, 144
Portion of leftward operator of λ	ATAGT:GGCGGTCA CC:ATA_3TA			4/5 (80)	8/10 (80)	7/7 (100)			145
Portion of SV40 early promoter	TTTATT:GCAGC: TTATAAT:GG:TT ACA_3UA_3:GC:AA UA			6/6 (100)	4/5 (80)	21/26[a] (81)			146
Portion of fd promoter	$TATAATA$:$GACAG_3$: TA_3:GACCTG:AT_5	7/7 (100)	5/7 (72)	4/4 (100)	4/6 (66)	6/6 (100)			147
Sequence near leftward operator of λ	A_3CA_2T:GC_6TGC: $A_6TA_3T_2CATATA_6$			6/7 (86)	9/10 (90)	22/23 (96)			148

[a]This region can be further divided into a 7/7 (100%) AT block followed by a 2/2 (100%) GC block followed by a 10/11 (91%) AT block followed by a 2/2 (100%) GC block followed by a 4/4 (100%) AT block.

the static structure of DNA, interest has developed on the dynamic properties of helical polynucleotides and the possible role of these dynamic transitions in gene regulation. A portion of these notions has developed from physical and enzymatic studies on the three duplex block polymers shown in Figure 9. These DNAs were synthesized[93,149] using a combination of DNA polymerases and exonucleases coupled with a high-resolution preparative scale fractionation technique which gave high-resolution separations according to chain length. These studies led to the telestability model for gene regulation at a distance. More recently, studies have been performed on the high-resolution 300-MHz nuclear magnetic resonance of one of these polymers, $d(C_{15}A_{15})\cdot d(T_{15}G_{15})$.[36] These studies suggest that the alternating AT and GC blocks present in some regulatory regions (see above) may act at several levels in defining these regions as specific sites for protein recognition.

Telestability in DNA

The duplex block polymers were synthesized with 10 or 15 AT base pairs attached to either 15 or 20 GC base pairs (Figure 9). The thermal denaturation of these polymers was studied at various sodium ion concentrations.[14,149] These thermal denaturation analyses showed that a regional cooperativity existed in the melting of these polymers. The thermal denaturation of $d(C_{20}A_{15})\cdot d(T_{15}G_{20})$ clearly illustrates this point. Figure 10 presents this analysis in 0.01 *M* sodium ion along with a tabulation of the calculated T_m of the isolated portions of this block polymer. As shown, this molecule melted in a single monophasic transition with a T_m of 51°. The GC portion of the block polymer stabilized the AT portion of 34° above the calculated T_m of isolated $d(A)_{15}\cdot d(T)_{15}$. The observed T_m was even 3° above that found for high molecular weight $d(A)_n\cdot d(T)_n$. In addition, the 20 GC base

Block DNA Polymers

$d(C_{15}A_{15})\cdot d(T_{15}G_{15})$

CCCCCCCCCCCCCCCAAAAAAAAAAAAAAA
GGGGGGGGGGGGGGGTTTTTTTTTTTTTTT

$d(C_{20}A_{15})\cdot d(T_{15}G_{20})$

CCCCCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAA
GGGGGGGGGGGGGGGGGGGTTTTTTTTTTTTTTT

$d(C_{20}A_{10})\cdot d(T_{10}G_{20})$

CCCCCCCCCCCCCCCCCCCCAAAAAAAAAA
GGGGGGGGGGGGGGGGGGGTTTTTTTTTT

FIGURE 9. Block DNA polymers used in the study of telestability. The synthesis and characterization of these DNAs were described elsewhere.[93,149]

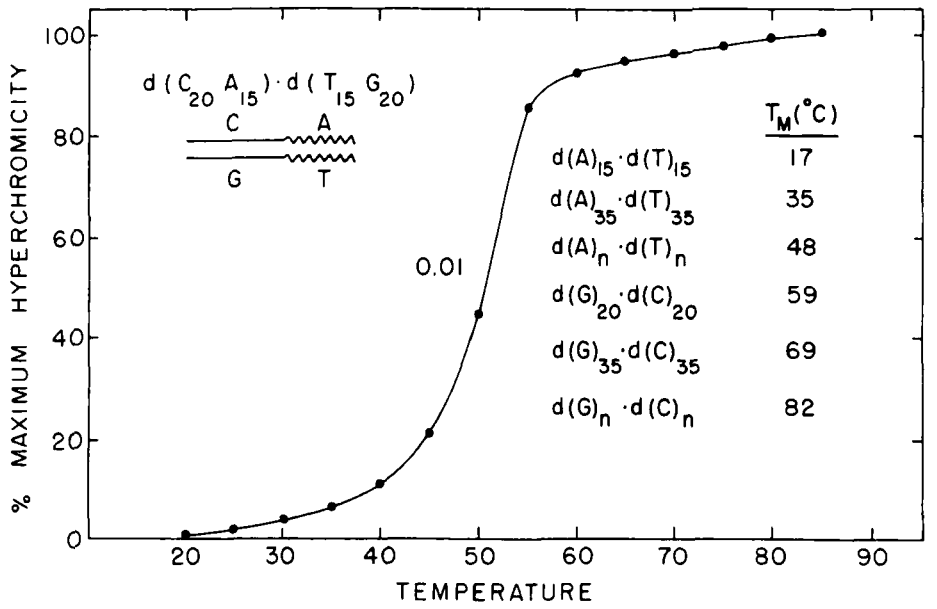

FIGURE 10. Helix-coil transition for $d(C_{20}A_{15})\cdot d(T_{15}G_{20})$ in 0.01 *M* sodium ion. The T_M values of related DNAs in the same salt concentration are listed. Further details are described elsewhere.[149]

pairs were destabilized by the AT portion of the helix. This block polymer melted 8° below the calculated T_m of isolated $d(G)_{20}\cdot d(C)_{20}$.

Telestability (the stabilization or destabilization of one region of DNA helix by an adjacent region) may be important in DNA-protein interactions. The studies described above show that the ability of a portion of a natural DNA helix to breathe can be influenced by the nucleotide sequence of adjacent regions. Hence, alterations in the nucleotide sequence (via deletions or base pair changes) or regions adjacent to an actual DNA-protein interaction site could influence the recognition of that site, particularly if the frequency of breathing is important for protein binding (see below).

Additional information on telestability was obtained using base-pair-specific drugs which bound to only one end of the block polymers. Actinomycin bound only to the GC portion, and netropsin bound only to the AT portion of the synthetic polymers.[14,149] The increased thermostability caused by drug binding extended to the nonbinding portions of these DNAs. For example, thermal denaturation of a block polymer in the presence of actinomycin increased the Tm by 12° but still yielded a single monophasic transition, indicating that both ends of the molecule melted simultaneously. Degradation of the AT portion of $d(C_{15}A_{15})\cdot d(T_{15}G_{15})$ by *E. coli* exonuclease I (which recognizes single-stranded character in DNAs) was inhibited when AT was bound to the GC end. This finding was consistent with the thermal denaturation studies, e.g., ligand binding to one region of a DNA can influence the dynamic properties (e.g., breathing frequency) of adjacent regions of the DNA.

Can Nearest-neighbor Interactions Explain Telestability?

The extent to which one DNA region can influence the thermostability of adjacent regions will depend on the range and magnitude of their cooperative interactions. The melting curves of the DNA block polymers were analyzed[150] to quantify the interactions between the AT and GC regions. The nearest neighbor base pair interaction model of the DNA helix-coil transition was used.[151] The merit of this model is that only nearest-neighbor base pair interactions are needed to account for observed transitions of nucleic acids. Although the model is a simplified description of DNA melting, it provides a framework to measure the influence of one DNA region on adjacent regions. The aim of this work was to determine if nearest-neighbor interactions alone could describe the block polymer melting curves and, if not, what the results would imply about the interactions between the two regions.

The melting of a block DNA polymer can be described by essentially three equilibrium constants. κ is the equilibrium constant for the association of two separate strands by an isolated base pair. S_{AT} is the equilibrium constant assigned to the reaction of forming an intact AT base pair next to an intact AT pair. S_{GC} is the equilibrium constant for forming an intact GC pair next to an intact GC pair. κ is generally assumed to be independent of temperature. S_{AT} and S_{GC} can be expressed in terms of the free energy change of forming a base pair by the equation

$$-RT \log S = \Delta G = \Delta H - T\Delta S \quad (1)$$

where

R = the gas constant;
T = temperature;
ΔH and ΔS = enthalpy and entropy changes of forming a base pair.

Estimates for ΔH and ΔS can be obtained from the melting temperatures of $d(A)_n \cdot d(T)_n$ and $d(G)_n \cdot d(C)_n$ and appropriate calorimetric measurements. The AT and GC pairs at the ends and at the AT/GC junction would be expected to have altered equilibrium constants due to their nearest neighbors. Changes in these parameters were included in the analysis but did not have a significant effect.

The analysis of the block DNA melting curves showed that nearest-neighbor interactions could not predict the experimental transitions of all three DNAs.[150] This was observed in 0.01 and 0.1 *M* sodium ion solutions. The nearest neighbor model provided good agreement with the melting curves of $d(C_{20}A_{15}) \cdot d(T_{15}G_{20})$ and $d(C_{15}A_{15}) \cdot d(T_{15}G_{15})$, but not $d(C_{20}A_{10}) \cdot d(T_{10}G_{20})$. Adjustments of the theoretical model to account for the phosphate repulsion terms did not alter this conclusion. However, if all 10 AT pairs in $d(C_{20}A_{10}) \cdot d(T_{10}G_{20})$ were arbitrarily stabilized 4°C above the melting temperature for the 15 AT pairs of the other block DNAs, a good agreement was observed on both high and low salt. This result is consistent with a substantial stabilization of the AT region nearest the GC block.

Further evidence for telestability in the block polymers was provided by the analysis of the effect of actinomycin on the melting of $d(C_{20}A_{15}) \cdot d(T_{15}G_{20})$ and $d(C_{15}A_{15}) \cdot d(T_{15}G_{15})$. Actinomycin binds tightly to GC regions, but does not bind DNAs containing only AT pairs.[30] Saturating actinomycin concentrations shifted the entire melting curves of $d(C_{20}A_{15}) \cdot d(T_{15}G_{20})$ and $d(C_{15}A_{15}) \cdot d(T_{15}G_{15})$ by 12°C in 0.01 *M* Na^+ solutions. No significant change in the breadth of the transition was noted. A test was conducted to determine if the nearest-neighbor theory could predict these results (Figure 11). In the theory, the stability of the G·C base pairs was increased by 18.5°C. This is the increase observed for $d(G)_n \cdot d(C)_n$. The predicted curve disagreed substantially with the experimental curve. The only way to produce agreement between the calculated and experimental curves was to increase the stability of all AT pairs by 9°C. The possibility of actinomycin binding and stabilizing several AT pairs near the GC/AT junction was ruled out as an explanation of the observations.[150] Changes in the single strand to duplex association constant, κ, could not explain the results. Hence, the entire AT end of the block DNAs appeared to be stabilized by 9°C due to actinomycin bound to the GC end.

In summary, interactions extending over a range greater than nearest nucleotide neighbors must be invoked to explain the helix-coil behavior of the block polymer. Thus, the novel feature of telestability is the range over which interactions must extend. At present, we know the range must extend over 35 base pairs; future work should provide further information on the extent of the range.

Sequence-dependent Conformational Differences

In addition to substantiating the results of the thermal denaturation studies on the block polymers described above, the high-resolution (300 MHz) proton magnetic resonance analysis of $d(C_{15}A_{15}) \cdot d(T_{15}G_{15})$ yielded several results which have implications regarding the function of the AT- and GC-rich blocks found in DNA regulatory regions.[36]

The NMR results were as follows. The AT portion of this duplex block polymer existed in two conformations; the major portion (11 to 12 base pairs) had the same conformation as $d(A)_n \cdot d(T)_{25}$ while 3 to 4 base pairs at the junction of the block were induced into an altered conformation by the adjacent GC block. The GC base pairs existed in only one conformation.

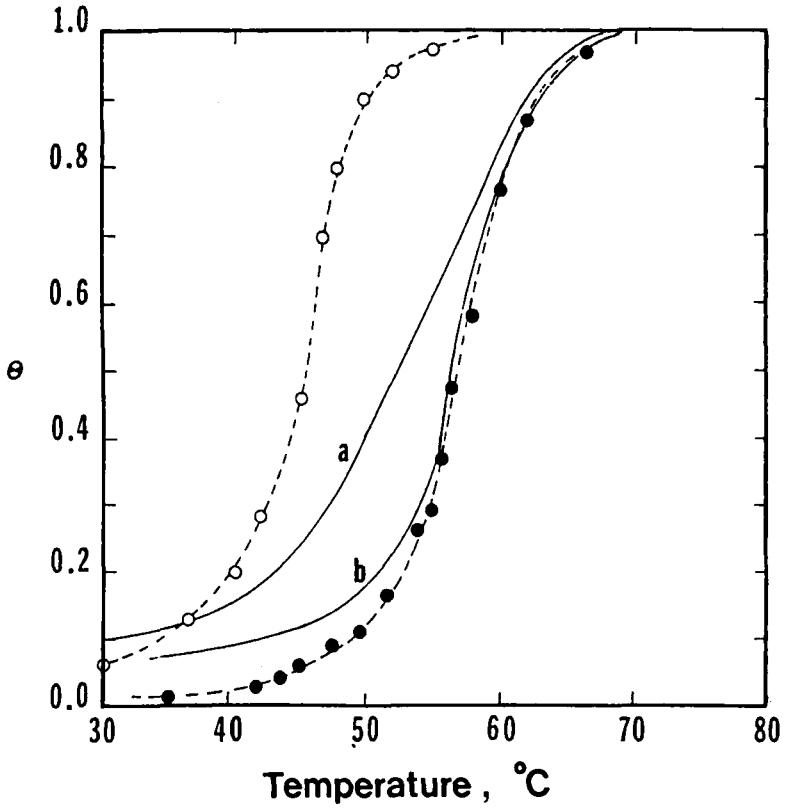

FIGURE 11. Effect of actinomycin on helix-coil transition of $d(C_{15}A_{15})\cdot d(T_{15}G_{15})$. Experimental curves are shown for the block DNA polymer in 0.01 *M* sodium ion in the absence of actinomycin (-0-0) and in the presence of 0.4 mol AM/DNA nucleotide (-●-●-). Curve a employs the theoretical parameters which fit the block DNA transition in the absence of actinomycin, except that the stability of G·C pairs was increased 18.5°C. Curve b shows the result of also increasing the stability of the A·T pairs by 9°C.

However, the conformation of the GC portion varied at different salt concentrations. The AT portion was not influenced by variation in the salt concentration. Actinomycin binding to the GC portion of the block polymer did not alter the conformation of the majority of AT base pairs.

Possible Recognition and Regulatory Functions of AT- and GC-rich Blocks in DNA

Telestability

The studies with the synthetic duplex block DNA polymers suggest several ways in which the AT- and GC-rich blocks may function in the recognition and control of DNA regulatory regions. The thermal denaturation studies are relevant to DNA regulation if the dynamic properties ("breathing") of DNA are important in its recognition and/or if it is necessary for a protein to "melt out" the helix after it binds. It is clear from a variety of studies[4,13] that in natural DNA there are, at least transiently, stretches of open single-stranded nucleotides. The ability to form such melted regions must be defined by the nucleotide sequence in and adjacent to these regions. Recent direct data show that an AT-rich region would be more likely to exist in an open form than a GC-rich region.[13] As discussed above, such open regions offer a wealth of potential interaction sites, particularly when compared to the sites afforded by the helical double-stranded form.

A possible recognition process might involve the interaction of melted regions of DNA with a specific protein. Transcription of certain messenger RNAs could involve the recognition by RNA polymerase of a defined number and sequence of nucleotides in the open form. One mechanism for regulating mRNA production would be to vary the amount of time that this recognition sequence for different mRNAs existed in the open form. This regulation could be accomplished by varying the

stability (e.g., nucleotide sequence) around this region. For example, an operon which required high levels of mRNA production at all times would have a high AT content adjacent to the recognition site. In contrast, an operon which required low levels of mRNA production might have a high GC content adjacent to the recognition site. This would work against the open conformation, thus allowing recognition by the RNA polymerase less often than in the case with the AT-rich adjacent region. The same arguments hold true for the case where the protein must "melt out" the DNA helix as it proceeds down the DNA. Regions of high GC content will impair progress, while high AT regions will offer lowest resistance. The processes involved in the termination of mRNA synthesis may require blocks of high GC content.

An extension of the principle of telestability is the possibility that a protein binding at one region of DNA alters the binding of a second protein at adjacent regions. The telestability model for the mechanism of action of catabolite gene activator protein (CAP) has been previously described.[14,15] This model states that CAP binds to a region of DNA adjacent to the RNA polymerase binding site, destabilizing the DNA and, hence, promoting the initiation of transcription on the nonhelical DNA template. One aspect of repressor binding to operator DNA may be to stabilize the adjacent DNA region, thereby disallowing RNA polymerase binding or progression. This seems reasonable in operons such as the arabinose operon where the operator is on the distal side of the promoter. However, in all of these systems there is the possibility of competition for overlapping sites leading to steric hindrance as well as direct protein-protein interaction on adjacent binding sites.

Several recent studies suggest that if the dynamic properties of DNA are important in its recognition and interaction with proteins, then these interactions at one region of DNA could be influenced by both the adjacent nucleotide sequence and by protein binding at adjacent regions.[13,152,153] Hirsh and Schleif[154] indicate that the "repression complex" for the arabinose regulatory region is located 108 base pairs upstream from the site of the "induction complex." The mechanism by which repression is accomplished from this distance and on the "wrong" side of the genome is unknown at present but may be an effect transmitted through the DNA helix, such as telestability.

Also, sequence studies on the *lac* messenger RNA formed in vitro[155] showed a GC-rich block of eight to ten bases directly followed by six to eight Us at the 3′ end. That the GC-rich region on the DNA template may be important in the termination of transcription was demonstrated by sequence studies on a number of short transcripts from abortively terminated RNAs directed by mutant DNAs.[140,155] When GC pairs (in the GC-rich region) were changed to AT pairs, there was little or no abortive termination. However, when the single AT pair in the string of GC pairs was converted to a GC pair (to give a string of 6 GCs), a stronger pause was found. Thus, a mutation ten base pairs upstream influenced the termination of transcription.

Unique DNA Conformations

The high-resolution NMR studies with the duplex block polymer $d(C_{15}A_{15})\cdot d(T_{15}G_{15})$ provide evidence for the existence of unique conformations within the regulatory regions of DNA. Furthermore, structural perturbants may differentially alter DNA conformation in such regulatory regions. Studies on high molecular weight synthetic double-helical DNAs (described above) show that they had a variety of properties and sometimes had different conformations, depending on the base sequence. The high-resolution NMR results showed that two different DNA conformations can exist with a transition zone of three to four base pairs possessing some intermediate conformation.

The promoter region of the gene for *Escherichia coli* tyrosine transfer RNA provides an example of how the NMR results may relate to a biological system and suggests several unique structural possibilities for this promoter. This sequence (Table 5) contains three blocks: 9/9 GC followed by 11/14 AT followed by 8/8 GC. Since the AT portion of the block polymer existed in two distinct conformations (one conformation for three to four base pairs being induced by the adjacent GC block), the tyrosine tRNA promoter AT-rich block could have an altered structure adjacent to the GC blocks on either side. Due to the proximity of the GC blocks, the entire AT-rich region may exist in a structure totally different from the structure of such a sequence when not bounded by such GC segments. The flexibility of DNA structure has been suggested from a variety of studies,[4,32] thus allowing different conformations to coexist. These unique con-

formations may provide for the specific recognition by proteins.

The block polymer studies also suggest that the GC blocks may exist in different conformations, depending on the solution environment. An intracellular environmental change could alter the conformation of one portion of a regulatory region, allowing (or disallowing) the recognition of that region by a specific protein.

VII. LARGE-SCALE ISOLATION OF DEFINED SEGMENTS OF CHROMOSOMES

In order to study the influence of nucleotide sequence on DNA properties and conformations, it will be necessary to obtain sizeable quantities of small genetically defined regions of DNA. One approach involves direct chemical synthesis of the sequence of interest.[156,157] Although laborious this method allows the construction of variant sequences not obtainable by genetic techniques. The alternative approach is to isolate certain DNA restriction fragments. This requires carrying the fragment on a vector molecule which may be obtained in quantity. Additionally, the fragment must be easily purified on a large scale from a restriction digest of the vector.

Past preparative fractionation procedures usually involved gel electrophoresis or ligand binding. Recently, reversed phase column chromatography (RPC) on RPC-5 resin has been investigated as a high-capacity separatory method for the fractionation of DNA fragments.[158,159] Milligram quantities of fragments generated by restriction endonuclease (either *Haemophilus aegyptius* III, *H. influenzae* II + III, or *Escherchia coli* R1) cleavage of λ*plac* 5 DNA have been purified on high-pressure columns. A fractionation of a *H. influenzae* II + III digest of λ*plac* 5 DNA is shown in Figure 12.

Figure 13A shows that the fragments generated by *H. aegyptius* III digestion of ϕX174 replicative form DNA are each purified to homogeneity by a single chromatography step on RPC-5. Figure 13B demonstrates the extent of resolution of fragments Z8, Z7, and Z6.

Considering the results from all restriction digests fractionated by RPC to date, the following generalizations can be stated. Fractionation is approximately according to increasing molecular weight. A 30 to 40% difference in molecular weight between any two fragments is sufficient for essentially complete resolution of the DNAs. Fragments in the range of 6×10^4 to 14×10^6 daltons have been fractionated, and it appears probable that both smaller and larger DNAs will be amenable to study by RPC. DNAs of the same molecular weight are resolved according to the presence or absence of four base single-stranded ends. In many cases, fragments of the same electrophoretic mobility are resolved by RPC. The property of DNA responsible for this behavior is unknown, but is not correlated with base composition.

Cloning into a more convenient vector can greatly simplify the subsequent large-scale separation of a certain fragment from the vector DNA. In general, a suitable DNA carrying the target sequence is cleaved by a restriction enzyme, and the resulting fragments are randomly joined to an autonomously replicating vehicle, such as the colicinigenic factor, a drug-resistant plasmid, or bacteriophage λDNA. Both sticky ends[160] and blunt ends[161] can be successfully joined. Next, a suitable *E. coli* host is either transformed or transfected with the recombinant DNA molecule and cloned to single colonies or plaques. A suitable selection or screening procedure based on the properties of the inserted sequence must then be applied.

Figure 14 shows the technique recently used for the large-scale preparation of the controlling elements of the *E. coli* lactose operon.[171] Fragments produced by *H. influenzae* d II + III were joined to a *H. influenzae* d II-treated minicolicin E_1 DNA.[162] The minicolicin E_1 DNA possesses only one *H. influenzae* II site.[172] After the T_4 DNA ligase reaction, *E. coli* was transformed with the composite plasmids. The resulting colicin-immune colonies were screened for *lac* constitutivity. The rationale for this assay is that the composite DNA sequesters the *lac* repressor as suggested by Abelson. After isolation, the plasmid DNA was cleaved with *H. influenzae* d II to give a 2×10^6 colicin segment and a 0.5×10^6 *lac* segment. These segments have been preparatively separated by chromatography on RPC-5.[173]

VIII. CONCLUDING REMARKS

A full understanding of the role of DNA structure in gene regulation will require the combined use of genetic, biochemical, and

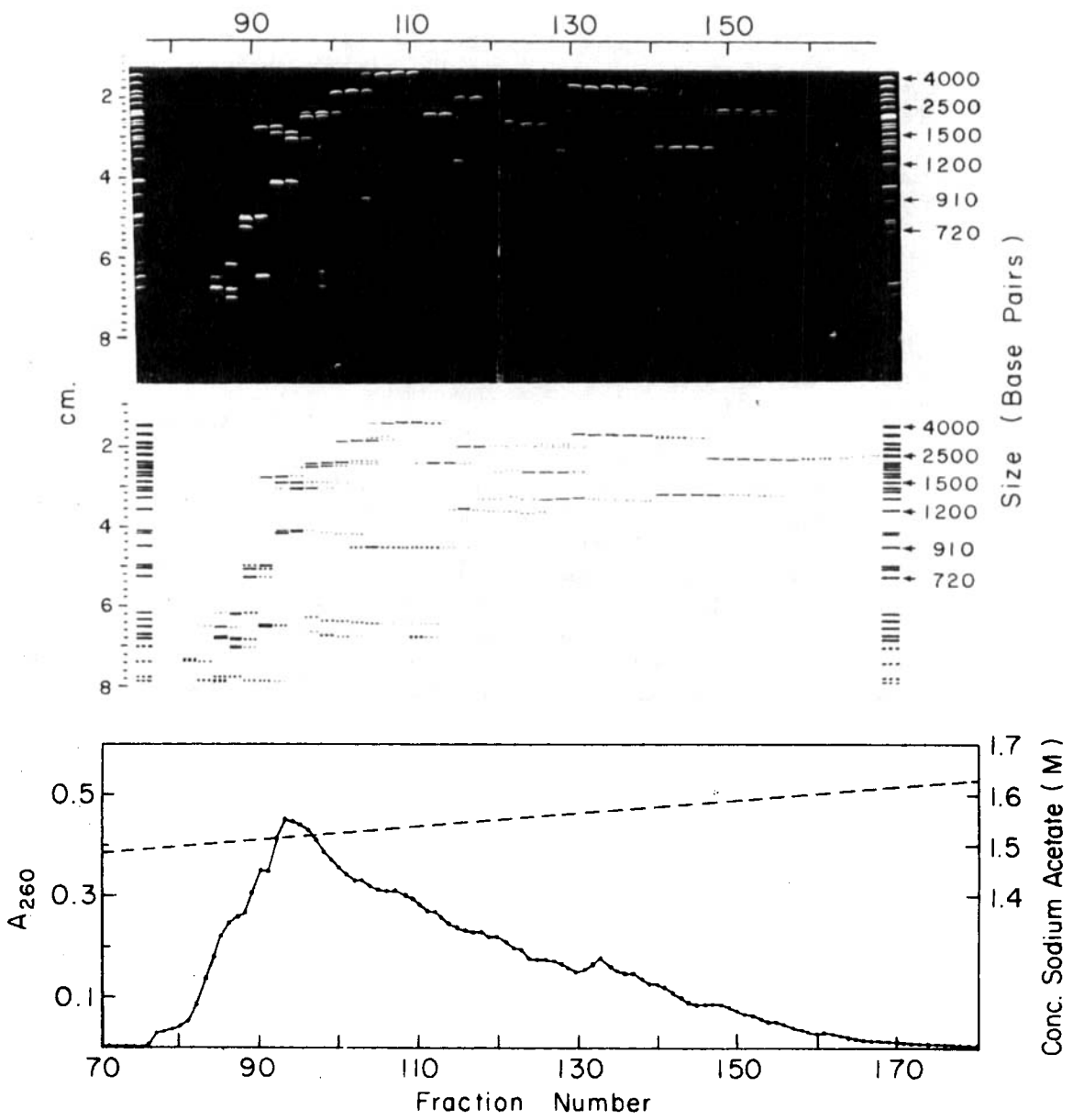

FIGURE 12. Fractionation of *Haemophilus influenzae* d II + III digest of λ*plac* DNA by RPC-5 column chromatography. Ninety-six A_{260} units of the digest were loaded onto a 135 × 0.9 cm column; the column was run at 200 to 400 psi. Four-milliliter fractions were taken; other details were described.[159] Upper panel: 4% polyacrylamide gel electrophoretic analysis of odd-numbered fractions from the column. The gels at the extreme right and left are on unfractionated digests. The scale at the left is an arbitrary measure for ascertaining the position of fragments. Middle panel: artist's conception of upper panel. Dotted lines indicate faint bands. Lower panel: A_{260} elution profile. (From Hardies, S. C. and Wells, R. D., *Proc. Natl. Acad. Sci. U.S.A.*, 73, 3117, 1976. With permission.)

physical techniques. Developments over the past 15 years in the area of fine structure mapping of various types of mutations and rapid advances in the past 2 years in the sequencing of DNAs justify high expectations for the future. Also, the biochemical and genetic characterization of several genetic controlling elements has provided new understanding of the mechanism of gene regulation.

In addition, the recent characterization of a variety of DNA restriction endonucleases, in combination with sequence analyses, permits the

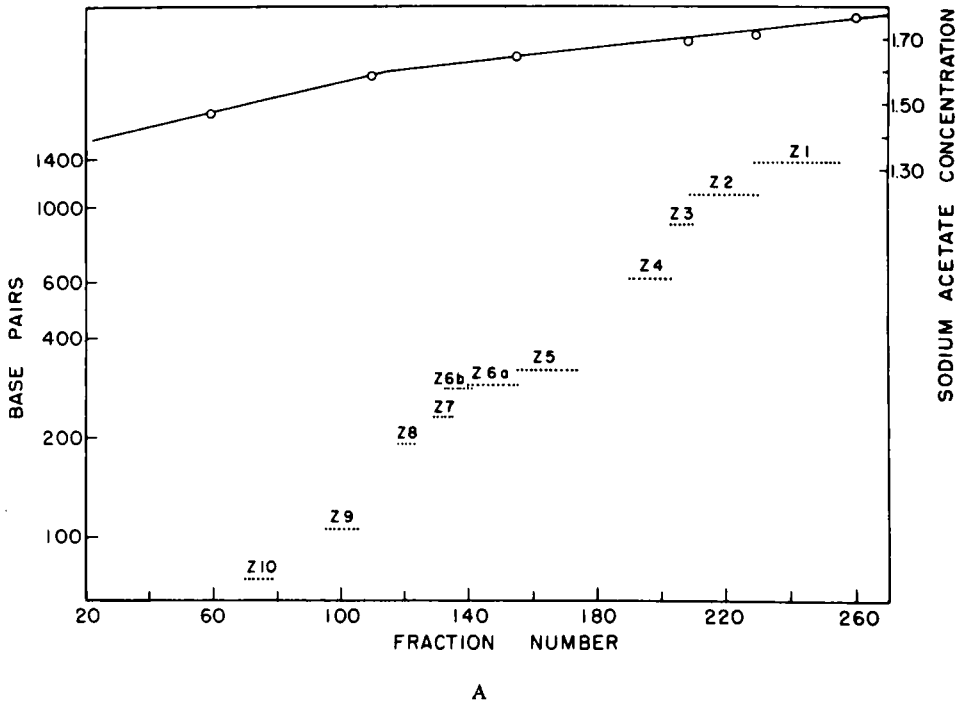

FIGURE 13. Fractionation of *Haemophilus aegyptius* III fragments of ϕX174 RF DNA by high-pressure RPC-5 column chromatography. ϕX174 RF DNA (approximately 20 OD_{260}) was digested to completion with *H. aegyptius* III under standard conditions.[100] The 11 fragments[164] range from 1350 to under 100 base pairs and are designated Z1 to Z10, respectively. (Z6a and b are a doublet and are poorly resolved.) The fragments were fractionated on a 0.6 × 60 cm column of RPC-5 at 200 psi essentially as described previously.[159] Fraction sizes were 4 ml from Fractions 20 to 110 and 3 ml thereafter; the flow rate was 60 ml/hr. (A) The column profile. The dots represent visible bands on the analytical gel electrophoresis. The size of the fragment[164] and the salt concentration at which it elutes (determined by conductance, open circles) are plotted. (B) A photograph of the analytical polyacrylamide gels for fractions 116 through 136; unfractionated total digests are shown on the extreme right and left. Note the complete resolution of fragment Z8 on the left and the partial overlap of fragments Z7 and Z6b on the right.

dissection of defined gene segments. The use of DNA cloning and/or column purification techniques will facilitate the large-scale purification of a variety of gene segments.

However, a simple and unambiguous physical or biochemical assay for different types of nucleic acid conformational features has been, and still is, a serious problem. New techniques must be explored.

ACKNOWLEDGMENTS

A portion of this work was supported by funds from the National Institutes of Health (CM-12275) and the National Science Foundation (BMS74-21420 to R. D. W. and pcm76-04565 to R. M. W.). Two of the authors (J. B. D. and S. C. H.) were supported by National Science Foundation Predoctoral Fellowships. One author (R. W. B.) was supported by a National Institutes of Health Viral Oncology Postdoctoral Training Grant and J. F. B. and J. B. D. were supported, in part, by Peterson and Wharton Fellowships from this department. G. T. H. was supported by a National Institutes of Health Predoctoral Training Grant, K. F. J. by the University of Copenhagen, and E. S. and R. M. W. by National Institutes of Health Postdoctoral Fellowships.

We are grateful to our friends and colleagues who provided information prior to publication.

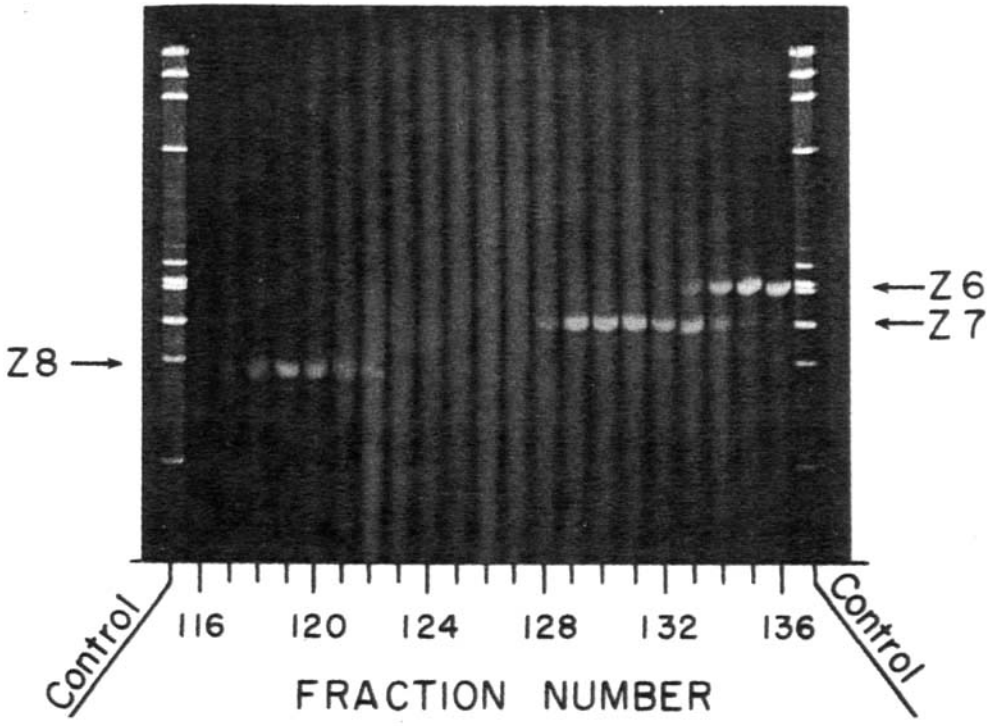

FIGURE 13B

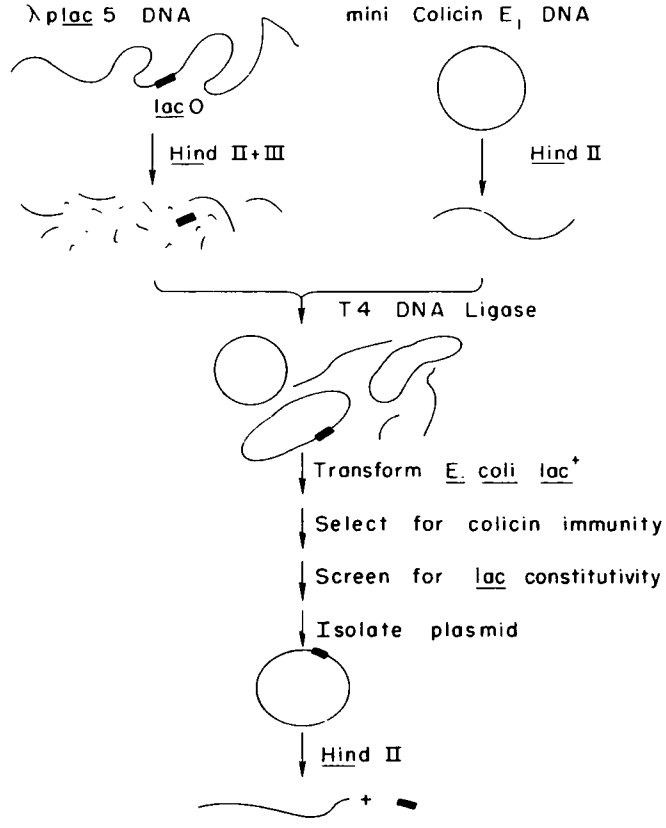

FIGURE 14. Procedure used for the formation of a recombinant DNA molecule for generating large quantities of the *Haemophilus influenzae* d II fragment (789 base pairs in length) containing the *lac* controlling elements.[171]

REFERENCES

1. Reznikoff, W. S., The operon revisited, *Annu. Rev. Genet.*, 6, 133, 1972.
2. Beckwith, J. and Rossow, P., Analysis of genetic regulatory mechanisms, *Annu. Rev. Genet.*, 8, 1, 1974.
3. Gots, J. S. and Benson, C. E., Biochemical genetics of bacteria, *Annu. Rev. Genet.*, 8, 79, 1974.
4. von Hippel, P. H. and McGhee, J. D., DNA-protein interactions, *Annu. Rev. Biochem.*, 41, 231, 1972.
5. Chamberlin, M. J., The selectivity of transcription, *Annu. Rev. Biochem.*, 43, 721, 1974.
6. Sund, H. and Blauer, G., Eds., *Protein-Ligand Interactions*, Walter de Gruyter, Berlin, 1975.
7. von Hippel, P. H., Conformational aspects of the nucleic acid – protein recognition problems, *J. Cell. Physiol.*, 74 (Suppl. 1), 235, 1969.
8. Inners, L. D. and Felsenfeld, G., Conformation of polyribouridylic acid in solution, *J. Mol. Biol.*, 50, 373, 1970.
9. Fiers, W., Contreras, R., Duernick, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., and Ysebaert, M., Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene, *Nature*, 260, 500, 1976.
10. Benbow, R. M., Hutchinson, C. A., III, Fabricant, J. D., and Sinsheimer, R. L., Genetic map of bacteriophage φX174, *J. Virol.*, 7, 549, 1971.
11. Geider, K. and Kornberg, A., Conversion of the M13 viral single strand to the double-stranded replicative forms by purified proteins, *J. Biol. Chem.*, 249, 3999, 1974.
12. Wang, J. C., Barkley, M. D., and Bourgeois, S., Measurements of unwinding of *lac* operator by repressor, *Nature*, 251, 247, 1974.
13. Chan, H. W., Dodgson, J. B., and Wells, R. D., The influence of DNA structure on the operator-repressor interaction, unpublished.
14. Burd, J. F., Wartell, R. M., Dodgson, J. B., and Wells, R. D., Transmission of stability (telestability) in DNA, *J. Biol. Chem.*, 250, 5109, 1975.
15. Dickson, R. C., Abelson, J., Barnes, W. M., and Reznikoff, W. S., Genetic regulation: the *lac* control region, *Science*, 187, 27, 1975.
16. Khorana, H. G., Büchi, H., Ghosh, H., Gupta, N., Jacob, T. M., Kössel, H., Morgan, R., Narang, S. A., Ohtsuka, E., and Wells, R. D., Polynucleotide synthesis and the genetic code, *Cold Spring Harbor Symp. Quant. Biol.*, 31, 39, 1966.
17. Wells, R. D., Ohtsuka, E., and Khorana, H. G., Studies on polynucleotides. L. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: a new double-stranded DNA-like polymer containing repeating dinucleotide sequences, *J. Mol. Biol.*, 14, 221, 1965.
18. Wells, R. D. and Blair, J. E., Studies on polynucleotides. LXXI. Sedimentation and buoyant density studies of some DNA-like polymers with repeating nucleotide sequences, *J. Mol. Biol.*, 27, 273, 1967.
19. Marmur, J. and Doty, P., Determination of the base composition of DNA from its thermal denaturation temperature, *J. Mol. Biol.*, 5, 109, 1962.
20. Schildkraut, C. L., Marmur, J., and Doty, P., Determination of the base composition of DNA from its buoyant density in CsCl, *J. Mol. Biol.*, 4, 430, 1962.
21. Wells, R. D. and Wartell, R. M., The influence of nucleotide sequence on DNA properties, in *Biochemistry of Nucleic Acids*, Vol. 6, Kornberg, H. L., Phillips, D. C., and Burton, K., Eds., University Park Press, Baltimore, 1974, 41.
22. Burd, J. F. and Wells, R. D., The effect of incubation conditions on the nucleotide sequence of DNA products of unprimed DNA polymerase reactions, *J. Mol. Biol.*, 53, 435, 1970.
23. Wells, R. D., Larson, J. E., Grant, R. C., Shortle, B. E., and Cantor, C. R., Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences, *J. Mol. Biol.*, 54, 465, 1970.
24. Grant, R. C., Kodama, M., and Wells, R. D., Enzymic and physical studies on $(dI\text{-}dC)_n \cdot (dI\text{-}dC)_n$ and $(dG\text{-}dC)_n \cdot (dG\text{-}dC)_n$, *Biochemistry*, 11, 805, 1972.
25. Ratliff, R. L., Williams, D. L., Hayes, F. N., Martinez, E. L., Jr., and Smith, D. A., Preparation and properties of the repeating sequence polymer $d(A\text{-}A\text{-}T)_n \cdot d(A\text{-}T\text{-}T)_n$, *Biochemistry*, 12, 5005, 1973.
26. Inman, R. B. and Baldwin, R. L., Helix-random coil transitions in DNA homopolymer pairs, *J. Mol. Biol.*, 8, 452, 1964.
27. Wells, R. D. and Larson, J. E., Buoyant density studies on natural and synthetic DNAs in neutral and alkaline solutions, *J. Biol. Chem.*, 247, 3405, 1972.
28. Inman, R. B. and Baldwin, R. L., Helix-random coil transitions in synthetic DNAs of alternating sequences, *J. Mol. Biol.*, 5, 172, 1962.
29. Wells, R. D., Actinomycin binding to DNA: inability of a DNA containing guanine to bind actinomycin D, *Science*, 165, 75, 1969.
30. Wells, R. D. and Larson, J. E., Studies on the binding of actinomycin D to DNA and DNA model polymers, *J. Mol. Biol.*, 49, 319, 1970.
31. Wartell, R. M., Larson, J. E., and Wells, R. D., Netropsin: a specific problem for A-T regions of duplex DNA, *J. Biol. Chem.*, 249, 6719, 1974.
32. Wartell, R. M., Larson, J. E., and Wells, R. D., The compatibility of netropsin and actinomycin binding to natural DNA, *J. Biol. Chem.*, 250, 2698, 1975.

33. Morgan, A. R. and Wells, R. D., Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences, *J. Mol. Biol.*, 37, 63, 1968.
34. De Clercq, E., Wells, R. D., Grant, R. C., and Merigan, T. C., Thermal activation of the antiviral activity of synthetic double-stranded polyribonucleotides, *J. Mol. Biol.*, 56, 83, 1971.
35. Gray, D. M. and Ratliff, R. L., Circular dichroism spectra of DNA, RNA and hybrid polymers in the presence of ethanol, *Biopolymers*, 14, 487, 1975.
36. Early, T. A., Kearns, D. R., Burd, J. F., Larson, J. E., and Wells, R. D., High resolution proton NMR investigation of the structural and dynamic properties of $d(C_{1\,5}A_{1\,5})\cdot d(T_{1\,5}G_{1\,5})$, *Biochemistry*, in press.
37. Patel, D. J., Proton and phosphorus NMR studies of d-CpG$(pCpG)_n$ duplexes in solution. Helix-coil transition and complex formation with actinomycin-D, *Biopolymers*, 15, 533, 1976.
38. Patel, D. J. and Canuel, L., Nuclear magnetic resonance studies of the helix-coil transition of poly (dA-dT) in aqueous solution, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 674, 1976.
39. Pohl, F. M. and Jovin, T. M., Salt-induced cooperative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC), *J. Mol. Biol.*, 67, 375, 1972.
40. Byrd, C., Ohtsuka, E., Moon, M. W., and Khorana, H. G., Synthetic deoxyribo-oligonucleotides as templates for the DNA polymerase of *Escherichia coli*: new DNA-like polymers containing repeating nucleotide sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 53, 79, 1965.
41. Wells, R. D., Jacob, T. M., Narang, S. A., and Khorana, H. G., Studies on polynucleotides. LXIX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: DNA-like polymers containing repeating trinucleotide sequences, *J. Mol. Biol.*, 27, 237, 1967.
42. Wells, R. D., Büchi, J., Kössel, H., Ohtsuka, E., and Khorana, H. G., Studies on polynucleotides. LXX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of *Escherichia coli*: DNA-like polymers containing repeating tetranucleotide sequences, *J. Mol. Biol.*, 27, 265, 1967.
43. Morgan, A. R., Wells, R. D., and Khorana, H. G., Studies on polynucleotides. LXXIV. Direct transition in vitro of single-stranded DNA-like polymers with repeating nucleotide sequences in the presence of neomycin B, *J. Mol. Biol.*, 26, 477, 1967.
44. Wells, R. D., Flugel, R. M., Larson, J. E., Schendel, P. F., and Sweet, R. W., Comparison of some reactions catalyzed by deoxyribonucleic acid polymerase from avian myeloblastosis virus, *Escherichia coli* and *Micrococcus luteus*, *Biochemistry*, 11, 621, 1972.
45. Morgan, A. R., Wells, R. D., and Khorana, H. G., Studies on polynucleotides. LIX. Further codon assignments from amino acid incorporations directed by ribopolynucleotides containing repeating trinucleotide sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 56, 1899, 1966.
46. Morgan, A. R., Studies on polynucleotides. XCIV. Transcription of DNA's with repeating nucleotide sequences, *J. Mol. Biol.*, 52, 441, 1970.
47. Mitsui, Y., Langridge, R., Shortle, B. E., Cantor, C. R., Grant, R. C., Kodama, M., and Wells, R. D., Physical and enzymatic studies on poly d(I-C)·poly d(I-C), an unusual double-helical DNA, *Nature*, 228, 1166, 1970.
48. Riggs, A. D., Lin, S., and Wells, R. D., *Lac* repressor binding to synthetic DNAs of defined nucleotide sequence, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 761, 1972.
49. Lin, S. Y. and Riggs, A. D., *Lac* repressor binding to operator analogues, *Biochem. Biophys. Res. Commun.*, 45, 1542, 1971.
50. De Clercq, E., Wells, R. D., and Merigan, T. C., Increase in antiviral activity of polynucleotides by thermal activation, *Nature*, 226, 364, 1970.
51. Merigan, T. C., De Clercq, E., Eckstein, F., and Wells, R. D., Molecular requirements for synthetic RNA to act in interferon stimulation, in *Miles Symposium on Biological Effects of Polynucleotides*, Beers, R. F. and Braun, W., Eds., Springer-Verlag, New York, 1970, 67.
52. De Clercq, E., Wells, R. D., and Merigan, T. C., Studies on the antiviral activity and cell interaction of synthetic double-stranded polyribo- and polydeoxyribonucleotides, *Virology*, 47, 405, 1972.
53. Erikson, R. L. and Szybalski, W., The Cs_2SO_4 equilibrium density gradient and its application for the study of T-even phage DNA, *Virology*, 22, 111, 1964.
54. Wilkins, M. H. F., Gosling, R. G., and Seeds, W. E., Physical studies of nucleic acid. Nucleic acid: an extensible molecule?, *Nature*, 167, 759, 1951.
55. Franklin, R. E. and Gosling, R. G., The structure of sodium thymonucleate. I. The influence of water content, *Acta Crystallogr.*, 6, 673, 1953.
56. Cooper, P. J. and Hamilton, L. D., The *A-B* conformational change in the sodium salt of DNA, *J. Mol. Biol.*, 16, 562, 1966.
57. Crick, F. H. C. and Watson, J. D., The complementary structure of deoxyribonucleic acid (DNA), *Proc. R. Soc. London Ser. A*, 223, 80, 1954.
58. Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F., and Hamilton, L. D., The molecular configuration of deoxyribonucleic acid. II. Molecular models and their fourier transforms, *J. Mol. Biol.*, 2, 38, 1960.
59. Fuller, W., Wilkins, M. H. F., Wilson, H. R., Hamilton, L. D., and Arnott, S., The molecular configuration of deoxyribonucleic acid. IV. X-ray diffraction study of the *A*-form, *J. Mol. Biol.*, 12, 60, 1965.

60. Arnott, S., Dover, S. D., and Wonacott, A. J., Least-squares refinement of the crystal and molecular structures of DNA and RNA from X-ray data and standard bond lengths and angles, *Acta Crystallogr. Sect. B*, 25, 2192, 1969.
61. Arnott, S. and Hukins, D. W. L., Optimised parameters for A-DNA and B-DNA, *Biochem. Biophys. Res. Commun.*, 47, 1504, 1972.
62. Arnott, S., Chandrasekaran, R., and Selsing, E., The variety of polynucleotide helices, in *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions*, Sundaralingam, M. and Rao, S. T., Eds., University Park Press, Baltimore, 1975, 577.
63. Sundaralingam, M. and Jensen, L. H., Stereochemistry of nucleic acid constituents. II. A comparative study, *J. Mol. Biol.*, 13, 930, 1965.
64. Marvin, D. A., Spencer, M., Wilkins, M. H. F., and Hamilton, L. D., The molecular configuration of deoxyribonucleic acid. III. X-ray diffraction study of the *C*-form of the lithium salt, *J. Mol. Biol.*, 3, 547, 1961.
65. Arnott, S. and Selsing, E., The conformation of *C*-DNA, *J. Mol. Biol.*, 98, 265, 1975.
66. Spencer, M., Fuller, W., Wilkins, M. H. F., and Brown, G. L., Determination of the helical configuration of ribonucleic acid molecules by X-ray diffraction study of crystalline amino-acid-transfer ribonucleic acid, *Nature*, 194, 1014, 1962.
67. Langridge, R. and Gomatos, P. J., The structure of RNA, *Science*, 141, 694, 1963.
68. Sato, T., Kyogoku, Y., Higuchi, S., Mitsui, Y., Iitaka, Y., Tsuboi, M., and Miura, K., A preliminary investigation on the molecular structure of rice dwarf virus ribonucleic acid, *J. Mol. Biol.*, 16, 180, 1966.
69. Arnott, S., Hutchinson, F., Spencer, M., Wilkins, M. H. F., Fuller, W., and Langridge, R., X-ray diffraction studies of double helical RNA, *Nature*, 211, 227, 1966.
70. Fuller, W., Hutchinson, F., Spencer, M., and Wilkins, M. H. F., Molecular and crystal structures of double-helical RNA. I. An X-ray diffraction study of fragmented yeast RNA and a preliminary double helical RNA model, *J. Mol. Biol.*, 27, 507, 1967.
71. Arnott, S., Wilkins, M. H. F., Fuller, W., and Langridge, R., Molecular and crystal structures of double helical RNA. III. An eleven-fold molecular model and comparison of the agreement between the observed and calculated three dimensional diffraction data from ten- and eleven-fold models, *J. Mol. Biol.*, 27, 535, 1967.
72. Milman, G., Langridge, R., and Chamberlain, M., The structure of a RNA-DNA hybrid, *Proc. Natl. Acad. Sci. U.S.A.*, 57, 1804, 1967.
73. O'Brien, E. J. and MacEwan, A. W., Molecular and crystal structure of the polynucleotide complex: polyinosinic acid plus polydeoxycytidylic acid, *J. Mol. Biol.*, 48, 243, 1970.
74. Hamilton, L. D., Barclay, R. K., Wilkins, M. H. F., Brown, G. L., Wilson, H. R., Marvin, D., Ephrussi-Taylor, H., and Simmons, N. S., Similarity of the structure of DNA from a variety of sources, *J. Biophys. Biochem. Cytol.*, 5, 397, 1959.
75. Arnott, S. and Selsing, E., Structures for the polynucleotide complexes poly(dA)·poly(dT) and poly(dT)·poly(dA)·poly(dT), *J. Mol. Biol.*, 88, 509, 1974.
76. Arnott, S. and Bond, P. J., Structures for poly(U)-poly(A)-poly(U) triple-stranded polynucleotides, *Nat. New Biol.*, 244, 99, 1973.
77. Arnott, S. and Bond, P. J., Triple-stranded polynucleotide helix containing only purine bases, *Science*, 181, 68, 1973.
78. Arnott, S. and Selsing, E., The structure of polydeoxyguanylic acid·polydeoxycytidylic acid, *J. Mol. Biol.*, 88, 551, 1974.
79. Davies, D. R. and Baldwin, R. L., X-ray studies of two synthetic DNA copolymers, *J. Mol. Biol.*, 6, 251, 1963.
80. Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C., and Watts, L., Structural details of a double-helix observed for DNAs containing alternating purine and pyrimidine sequences, *J. Mol. Biol.*, 88, 523, 1974.
81. Selsing, E., Arnott, S., and Ratliff, R. L., Conformations of poly[d(A-T-T)]·poly[d(A-A-T)], *J. Mol. Biol.*, 98, 243, 1975.
82. Premilat, S. and Albiser, G., X-ray diffraction study of three DNA fibers with different base compositions, *J. Mol. Biol.*, 99, 27, 1975.
83. Selsing, E. and Arnott, S., Conformations of A,T-rich DNAs, *Nucleic Acids Res.*, 3, 2443, 1976.
84. Selsing, E., Leslie, A. G. W., Arnott, S., Gall, J. G., Skinner, D. M., Southern, E. M., Spencer, J. H., and Harbers, K., Conformations of satellite DNAs, *Nucleic Acids Res.*, 3, 2451, 1976.
85. Pilet, J. and Brahms, J., Dependence of the B-A conformational change in DNA on base composition, *Nat. New Biol.*, 236, 99, 1972.
86. Denhardt, D. T., Eisenberg, S., Harbers, B., Lane, H. E. D., and McFadden G., Replication of ϕX174 in *Escherichia coli*: structure of the replicating intermediate and the effect of mutations in the host *lig* and *rep* genes, in *DNA Synthesis and Its Regulation*, ICN-UCLA Symposium on Molecular and Cellular Biology, Vol. 3, Goulian, M., Hanawalt, P., and Fox, C. F., Eds., W. A. Benjamin, Menlo Park, 1976, 398.
87. Chan, H. W. and Wells, R. D., Structural uniqueness of the lactose operator, *Nature*, 252, 205, 1974.
88. Marians, K. J. and Wu, R., Structure of the lactose operator, *Nature*, 260, 360, 1976.
89. Beard, P., Morrow, J. F., and Berg, P., Cleavage of circular, superhelical simian virus 40 DNA to a linear duplex by S_1 nuclease, *J. Virol.*, 12, 1303, 1973.
90. Kato, A. C., Bartok, K., Fraser, M. J., and Denhardt, D. T., Sensitivity of superhelical DNA to a single-strand specific endonuclease, *Biochim. Biophys. Acta*, 308, 68, 1973.

91. Shenk, T. E., Rhodes, C., Rigby, P. W. J., and Berg, P., Biochemical method for mapping mutational alterations in DNA with S_1 nuclease: the location of deletions and temperature-sensitive mutations in simian virus 40, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 989, 1975.
92. Dodgson, J. B. and Wells, R. D., The action of single strand specific nucleases on model DNA heteroduplexes of defined size and sequence, submitted, 1976.
93. Burd, J. F. and Wells, R. D., Synthesis and characterization of the duplex block polymer $d(C_{1\,5}A_{1\,5})\cdot d(T_{1\,5}G_{1\,5})$, *J. Biol. Chem.*, 249, 7094, 1974.
94. Dean, W. W. and Lebowitz, J., Partial alteration of secondary structure in native superhelical DNA, *Nat. New Biol.*, 231, 5, 1971.
95. Allet, B. and Bukhari, A. F., Analysis of bacteriophage Mu and λ-Mu hybrid DNAs by specific endonucleases, *J. Mol. Biol.*, 92, 529, 1975.
96. Riggs, A. D., Suzuki, H., and Bourgeois, S., *Lac*-repressor-operator interaction. I. Equilibrium studies, *J. Mol. Biol.*, 48, 67, 1970.
97. Davis, R. W., Simon, M., and Davidson, N., Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids, in *Methods in Enzymology*, Vol. 21, Grossman, L. and Moldave, K., Eds., Academic Press, New York, 1971, 413.
98. Johnson, P. H., and Laskowski, M., Sr., Mung bean nuclease. I. Resistance of double stranded deoxyribonucleic acid and susceptibility of regions rich in adenosine and thymidine to enzymatic hydrolysis, *J. Biol. Chem.*, 245, 891, 1970.
99. Weissmann, C., Billeter, M. A., Goodman, H. M., Hindley, J., and Weber, H., Structure and function of phage RNA, *Annu. Rev. Biochem.*, 42, 303, 1973.
100. Blakesley, R. W. and Wells, R. D., "Single-stranded" DNA from φX174 and M13 is cleaved by certain restriction endonucleases, *Nature*, 257, 421, 1975.
101. Boyer, H. W., Restriction and modification of DNA: enzymes and substrates, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 33, 1125, 1974.
102. Baumstark, B., Roberts, R., and RajBhandary, U., unpublished.
103. Greene, P. J., Poonian, M. S., Nussbaum, A. L., Tobias, L., Garfin, D. E., Boyer, H. W., and Goodman, H. M., Restriction and modification of a self-complementary octanucleotide containing the *Eco*RI substrate, *J. Mol. Biol.*, 99, 237, 1975.
104. Horuichi, K. and Zinder, N. D., Site-specific cleavage of single-stranded DNA by a *Hemophilus* restriction endonuclease, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 2555, 1975.
105. Blakesley, R. W., Duplex regions in "single-stranded" φX174 DNA, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 35, 1518, 1976.
106. Fiers, W., and Sinsheimer, R. L., The structure of the DNA of bacteriophage φX174. III. Ultracentrifugal evidence for a ring structure, *J. Mol. Biol.*, 5, 424, 1962.
107. Schaller, H., Voss, H., and Gucker, S., Structure of the DNA of bacteriophage fd. II. Isolation and characterization of a DNA fraction with double strand-like properties, *J. Mol. Biol.*, 44, 445, 1969.
108. Bartok, K., Harbers, B., and Denhardt, D. T., Isolation and characterization of self-complementary sequences from φX174 viral DNA, *J. Mol. Biol.*, 99, 93, 1975.
109. Galibert, F., Sedat, J., and Ziff, E., Direct determination of DNA nucleotide sequences: structure of a fragment of bacteriophage φX174 DNA, *J. Mol. Biol.*, 87, 377, 1974.
110. Schaller, H., Uhlmann, A., and Geider, K., A DNA fragment from the origin of single-strand to double-strand DNA replication of bacteriophage fd, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 49, 1976.
111. Robertson, H. D., Isolation of specific ribosome binding sites from single-stranded DNA, *J. Mol. Biol.*, 92, 363, 1975.
112. Barrell, B. G., Weith, H. L., Donelson, J. E., and Robertson, H. D., Sequence analysis of the ribosome-protected bacteriophage φX174 DNA fragment containing the gene G initiation site, *J. Mol. Biol.*, 92, 377, 1975.
113. Air, G. M., Blackburn, E. H., Sanger, F., and Coulson, A. R., The nucleotide and amino acid sequences of the N(5') terminal region of gene G of bacteriophage φX174, *J. Mol. Biol.*, 96, 703, 1975.
114. Kozak, M., and Nathans, D., Translation of the genome of a ribonucleic acid bacteriophage, *Bacteriol. Rev.*, 36, 109, 1972.
115. Thomas, G. J., Jr., Prescott, B., McDonald-Ordzie, P. E., and Hartman, K. A., Studies of virus structure by Laser-Raman spectroscopy. II. MS2 phage, MS2 capsids and MS2 RNA in aqueous solutions, *J. Mol. Biol.*, 102, 103, 1976.
116. Boedtker, H., The reaction of ribonucleic acid with formaldehyde. I. Optical absorbance studies, *Biochemistry*, 6, 2718, 1967.
117. Gralla, J., Steitz, J. A., and Crothers, D. M., Direct physical evidence for secondary structure in an isolated fragment of R17 bacteriophage mRNA, *Nature*, 248, 204, 1974.
118. Jacobson, A. B., Studies on secondary structure of single-stranded RNA from bacteriophage MS2 by electron microscopy, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 307, 1976.
119. Weber, H., The binding site for coat protein on bacteriophage Qβ RNA, *Biochim. Biophys. Acta*, 418, 175, 1976.
120. Eisenberg, S., Harbers, B., Hours, C., and Denhardt, D. T., The mechanism of replication of φX174 DNA. XII. Non-random location of gaps in nascent φX174 RF II DNA, *J. Mol. Biol.*, 99, 107, 1975.

121. **Bartok, K. and Denhardt, D. T.,** Site of cleavage of superhelical φX174 replicative form DNA by the single strand-specific *Neurospora crassa* endonuclease, *J. Biol. Chem.*, 251, 530, 1976.
122. **Mechali, M., De Recondo, A.-M., and Girard, M.,** Action of the S_1 endonuclease from *Aspergillus oryzae* on simian virus 40 supercoiled component 1 DNA, *Biochem. Biophys. Res. Commun.*, 54, 1306, 1973.
123. **Germond, J.-E., Vogt, V. M., and Hirt, B.,** Characterization of the single-strand-specific nuclease S_1 activity on double-stranded supercoiled polyoma DNA, *Eur. J. Biochem.*, 43, 591, 1974.
124. **Wang, J. C.,** Interactions between twisted DNAs and enzymes: the effects of superhelical turns, *J. Mol. Biol.*, 87, 797, 1974.
125. **Mitra, S., Allison, D. P., and Dasgupta, S.,** Base-unpaired regions in covalently closed supercoiled replicative form DNA of coliphage M13, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 35, 1517, 1976.
126. **Morrow, J. F. and Berg, P.,** Location of the T4 gene 32 protein binding site on simian virus 40 DNA, *J. Virol.*, 12, 1631, 1973.
127. **Brack, C., Bickel, T. A., and Yuan, R.,** The relation of single-stranded regions in bacteriophage PM2 supercoiled DNA to the early melting sequences, *J. Mol. Biol.*, 96, 693, 1975.
128. **Holloman, W. K., Wiegand, R., Hoessli, C., and Radding, C. M.,** Uptake of homologous single-stranded fragments by superhelical DNA: a possible mechanism for initiation of genetic recombination, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 2394, 1975.
129. **Beerman, T. A. and Lebowitz, J.,** Further analysis of the altered secondary structure of superhelical DNA. Sensitivity of methylmercuric hydroxide, a chemical probe for unpaired bases, *J. Mol. Biol.*, 79, 451, 1973.
130. **Woodworth-Gutai, M. and Lebowitz, J.,** Introduction of interrupted secondary structure in supercoiled DNA as a function of superhelix density: consideration of hairpin structures in superhelical DNA, *J. Virol.*, 18, 195, 1976.
131. **Lebowitz, J., Garon, C. G., Chen, M. C. Y., and Salzman, N. P.,** Chemical modification of simian virus 40 DNA by reaction with a water-soluble carbodiimide, *J. Virol.*, 18, 205, 1976.
132. **Chen, M., Lebowitz, J., and Salzman, N. P.,** *Hin* D restriction mapping of unpaired regions in simian virus 40 superhelical DNA I: Considerations regarding structure-function relationships, *J. Virol.*, 18, 211, 1976.
133. **Hsieh, T.-S. and Wang, J. C.,** Thermodynamic properties of superhelical DNAs, *Biochemistry*, 14, 527, 1975.
134. **Depew, R. E. and Wang, J. C.,** Conformational fluctuations of DNA helix, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 4275, 1975.
135. **Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J., and Vosberg, H.-P.,** Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed circular DNA: formation of a Boltzmann distribution of topological isomers, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 4280, 1975.
136. **Salser, W. A.,** DNA sequencing techniques, *Annu. Rev. Biochem.*, 43, 923, 1974.
137. **Wu, R., Bambara, R., and Jay, E.,** Recent advances in DNA sequence analysis, *CRC Crit. Rev. Biochem.*, 2, 455, 1975.
138. **Dykes, G., Bambara, R., Marians, K., and Wu, R.,** On the statistical significance of primary structural features found in DNA-protein interaction sites, *Nucleic Acids Res.*, 2, 327, 1974.
139. **Sekiya, T., van Ormondt, H., and Khorana, H. G.,** The nucleotide sequence in the promoter region of the gene for an *Escherichia coli* tyrosine transfer ribonucleic acid, *J. Biol. Chem.*, 250, 1087, 1975.
140. **Gilbert, W., Gralla, J., Majors, J., and Maxam, A.,** Lactose operator sequences and the action of *lac* repressor, in *Protein-Ligand Interactions*, Sund, H. and Blauer, G., Eds., Walter de Gruyter, Berlin, 1975, 193.
141. **Weissman, S.,** personal communication.
142. **Bertrand, K., Korn, L., Lee, F., Platt, T., Squires, C. L., Squires, C., and Yanofsky, C.,** New features of the regulation of the tryptophan operon, *Science*, 189, 22, 1975.
143. **Maniatis, T., Jeffrey, A., and Kleid, D. G.,** Nucleotide sequence of the rightward operator of phage λ, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 1184, 1975.
144. **Walz, A. and Pirrotta, V.,** Sequence of the P_r promoter of phage λ, *Nature*, 254, 118, 1975.
145. **Maniatis, T., Ptashne, M., Barrell, B. G., and Donelson, J.,** Sequence of a repressor-binding site in the DNA bacteriophage λ, *Nature*, 250, 394, 1974.
146. **Dhar, R., Weissman, S. M., Zain, B. S., Pan, J., and Lewis, A. M.,** The nucleotide sequence preceding RNA polymerase initiation site on SV40 DNA. II. The sequence of the early strand transcript, *Nucleic Acids Res.*, 1, 595, 1974.
147. **Schaller, H., Gray, C., and Herrmann, K.,** Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 737, 1975.
148. **Pirrotta, V.,** personal communication.
149. **Burd, J. F., Larson, J. E., and Wells, R. D.,** Further studies on telestability in DNA: the synthesis and characterization of the duplex block polymers $d(C_{20}A_{10})\cdot d(T_{10}G_{20})$ and $d(C_{20}A_{15})\cdot d(T_{15}G_{20})$, *J. Biol. Chem.*, 250, 6002, 1975.
150. **Wartell, R. M. and Burd, J. F.,** Evidence for long range interactions in DNA: analysis of melting curves of block polymers $d(C_{15}A_{15})\cdot d(T_{15}G_{15})$, $d(C_{20}A_{15})\cdot d(T_{15}G_{20})$ and $d(C_{20}A_{10})\cdot d(T_{10}G_{20})$, *Biopolymers*, 15, 1461, 1976.
151. **Crothers, D. M.,** On the mechanism of DNA unwinding, *Acc. Chem. Res.*, 2, 225, 1969.
152. **Nakanishi, S., Adhya, S., Gottesman, M., and Pastan, I.,** Selective effects of $MgCl_2$ and temperature on the initation of transcription at *lac*, *gal*, and λ promoters, *J. Biol. Chem.*, 250, 8202, 1975.

153. Travers, A., Baillie, D. L., and Pedersen, S., Effect of DNA conformation on ribosomal RNA synthesis in vitro, *Nat. New Biol.*, 243, 161, 1973.
154. Hirsh, J. and Schleif, R., Electron microscopy of gene regulation: the L-arabinose operon, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 1518, 1976.
155. Maizels, N. M., The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of *Escherichia coli, Proc. Natl. Acad.. Sci. U.S.A.*, 70, 3585, 1973.
156. Khorana, H. G., Agarwal, K. L., Buchi, H., Caruthers, M. H., Gupta, N. K., Kleppe, K., Kumar, A., Ohtsuka, E., RajBhandary, U. L., Van de Sande, J. H., Sgaramella, V., Terao, T., Weber, H., and Yamada, T., Studies on polynucleotides. CIII. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast, *J. Mol. Biol.*, 72, 452, 1972.
157. Bahl, C. P., Wu, R., Itakura, K., Katagiri, N., and Narang, S. A., Chemical and enzymatic synthesis of lactose operator of *Escherichia coli* and its binding to lactose repressor, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 91, 1976.
158. Hardies, S. C., Large scale fractionation of double stranded DNA restriction fragments by reversed phase column chromatography, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 35, 1369, 1976.
159. Hardies, S. C. and Wells, R. D., Preparative fractionation of DNA restriction fragments by reversed phase column chromatography, *Proc. Natl. Acad. Sci. U.S.A.*, 73, 3117, 1976.
160. Cohen, S. N., Chang, A. C. Y., Boyer, H. W., and Helling, R. B., Construction of biologically functional bacterial plasmids *in vitro, Proc. Natl. Acad. Sci. U.S.A.*, 70, 3240, 1973.
161. Sgaramella, V. and Khorana, H. G., Studies on polynucleotides. CXVI. A further study of the T4 ligase-catalyzed joining of DNA at base-paired ends, *J. Mol. Biol.*, 72, 493, 1972.
162. Hershfield, V., Boyer, H. W., Chow, L., and Helinski, D. R., Characterization of mini-Col E_1 plasmid, *J. Bacteriol.*, 126, 447, 1976.
163. Gillam, S., Waterman, K., Doel, M., and Smith, M., Enzymatic synthesis of deoxyribo-oligonucleotide synthesis, *Nucleic Acids Res.*, 1, 1649, 1974.
164. Lee, A. S. and Sinsheimer, R. L., A cleavage map of bacteriophage ϕX174 genome, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 2882, 1974.
165. Early, T. A., Kearns, D. R., Burd, J. F., and Wells, R. D., unpublished data.
166. Leslie, A. G. W., Arnott, S., and Tsivoglou, S., personal communication.
167. Arnott, S., personal communication.
168. Dodgson, J. B., Reznikoff, W. S., and Wells, R. D., unpublished data.
169. Jensen, K. F., Nes, I. F., and Wells, R. D., Iodination as a probe for small regions of disrupted secondary structure in double-stranded DNAs, *Nucleic Acid Res.*, 3, 3143, 1976.
170. Gilbert, W. and Maxam, A., personal communication.
171. Hardies, S. C., Reznikoff, W. S., and Wells, R. D., unpublished data.
172. Barnes, W. M. and Sanger, F., personal communication.
173. Hardies, S. C. and Wells, R. D., unpublished data.