

H-DNA AND RELATED STRUCTURES

Sergei M. Mirkin

Department of Genetics, University of Illinois at Chicago, Chicago, Illinois
60612

*Maxim D. Frank-Kamenetskii*¹

Institute of Molecular Genetics, Russian Academy of Sciences, Moscow
123182, Russia

KEY WORDS: DNA structure, triple-helical DNA, nuclease
hypersensitivity, DNA polymerase, promoters,
telomeres

CONTENTS

PERSPECTIVES AND OVERVIEW	541
CANONICAL H-DNA	544
*H-DNA	553
H-DNA-LIKE-STRUCTURES	557
<i>Nodule DNA</i>	558
<i>Eclectic DNA</i>	559
<i>H-Like Structure Formed by Parallel-Stranded DNA</i>	561
DETECTION OF H-DNA IN VIVO	562
H-DNA AND DNA POLYMERIZATION	564
H MOTIFS IN EUKARYOTIC PROMOTERS	568
CONCLUSIONS	570

PERSPECTIVES AND OVERVIEW

The discovery of the crystal form of Z-DNA by Alex Rich and his coworkers in 1979 (123) and the subsequent demonstration of Z-DNA

¹Present address: Center for Advanced Biotechnology and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215.

and cruciform formation in supercoiled plasmids (71, 97, 116) stimulated interest in unusual DNA conformations and their possible biological role(s). Observations by Larsen & Weintraub (68) and Hentchel (41) that promoters of eukaryotic genes in both active chromatin and supercoiled plasmids show hypersensitivity toward S1 nuclease indicated to many researchers that unusual structures might be formed at those sites (127). Numerous studies revealed that S1 sensitivity was associated with homopurine-homopyrimidine stretches (reviewed in 128). This finding was surprising because homopurine-homopyrimidine sequences could adopt neither cruciform nor Z-DNA structures. Slip-page loops (41, 80, 87, 112), left-handed helices (19, 84, 105), triple helices (20, 69), and quarter helices (47) had been discussed in the literature; but the controversial data on fine mapping of S1 cleavage sites did not permit definite conclusions about the nature of any unusual structures in the homopurine-homopyrimidine tracts.

Such conclusions were finally enabled by two-dimensional gel electrophoresis of supercoiled DNAs carrying a homopurine-homopyrimidine insert (75, 76, 91). These studies revealed beyond any doubt that an unusual structure was formed and, remarkably, that this structure was stabilized by hydrogen ions—hence the name "H form" for the unusual structure (Figure 1). The main element of the H form is an intramolecular triple helix formed by the entire pyrimidine strand and half of the purine strand; the other half of the purine strand remains single stranded. The triple helix is stabilized by CG^*C^+ and TA^*T base

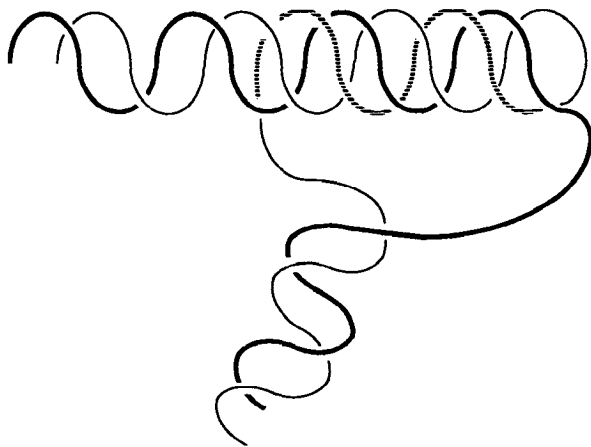


Figure 1 H-DNA model. Bold line, homopurine strand; thin line, homopyrimidine strand; dashed line, the half of the homopyrimidine strand donated to the triplex.

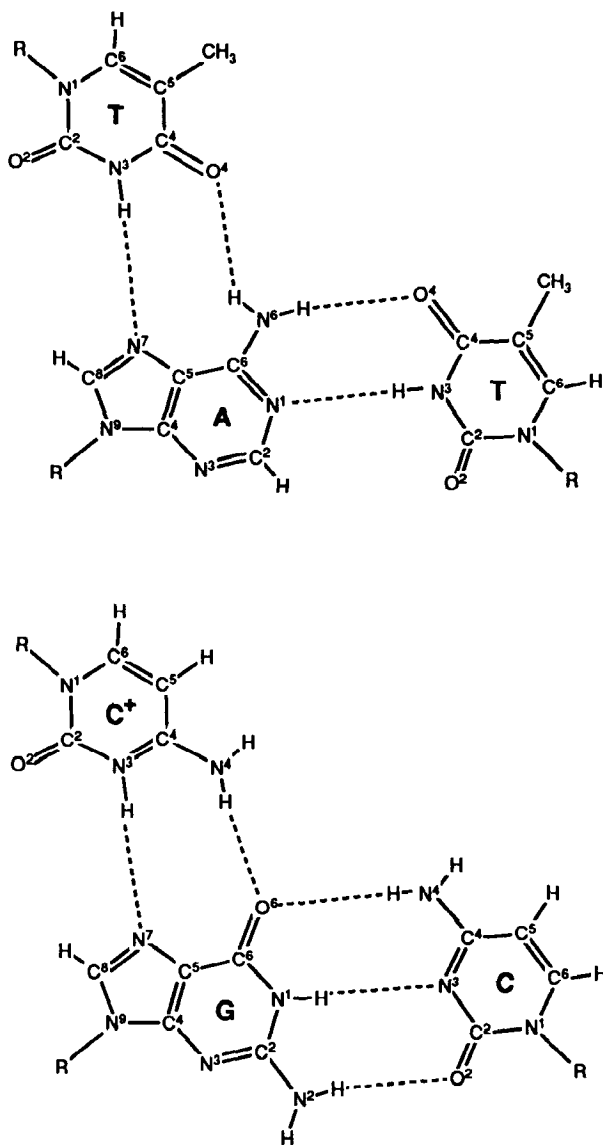


Figure 2 TA*T and CG*C⁺ base triads.

triads (Figure 2). Numerous studies using mutational analysis and chemical probing have substantiated this model (see reviews 32, 33, 96, 128, 129, and references therein).

H-DNA was the first example of an intramolecular DNA triplex. It has become clear in recent years that an entire family of H-like DNA structures exists whose members differ in the chemical nature of their triplexes; such structures can exist in different isomeric forms depending on ambient conditions and sequences. Similarly, the variety of sequences known to be able to adopt the H-DNA structure has significantly increased during the last few years. We review these recent findings and discuss factors affecting the structure and stability of H-DNA.

The discovery of H-DNA stimulated speculations about its possible biological role. However, until very recently, efforts by many groups to find H-DNA *in vivo* remained unsuccessful. Now several reports have appeared that provide the first reliable indications that H-DNA may exist *in vivo*. Intriguing biochemical data are available that show the influence of H-DNA on replication and transcription. We discuss these data along with other biological applications.

CANONICAL H-DNA

In order to study a structure of homopurine-homopyrimidine stretches in supercoiled DNA Lyamichev et al (75) chose a method called two-dimensional (2-D) gel electrophoresis of DNA topoisomers. This method, first described by Wang et al (125), is based on the dependence of the mobility of circular DNA in the gel on its torsional tension. The mobility of DNA topoisomers, i. e. circular DNA molecules that are chemically identical but differ in their supercoiling, will increase until saturation with an increase of the number of supercoils in the first direction of gel electrophoresis. If a structural transition occurs, accompanied by the release of some superhelical stress, the mobility of the corresponding topoisomers decreases. Thus, the topoisomer under transition co-migrates with a less supercoiled one. Electrophoresis in a second direction, perpendicular to the first, is used to resolve co-migrating topoisomers. At this stage an intercalating dye (usually chloroquine) is added to the buffer. The dye intercalates into the DNA duplex and relaxes negative superhelical tension, thus converting non-B-DNA segments into B conformation. As a result, the mobility of previously co-migrated topoisomers becomes different according to their actual linking difference. In the final picture, one can see a gradual increase of topoisomer mobility until a sharp drop appears, reflecting

the transition (Figure 3). From its 2-D pattern, one can determine two important characteristics of a structural transition. First, it is easy to calculate the number of supercoils released during transition. If the length of the sequence adopting a new conformation is known, one can deduce the topological status of the new conformation. Second, it is easy to calculate the number of supercoils necessary for the transition. This makes it possible to estimate the free energy of a transition under given ambient conditions.

Lyamichev et al (75) studied a cloned sequence from a spacer between the histone genes of the sea urchin *P. miliaris*. It contained a $d(G-A)_{16}$ stretch that had been found to be hypersensitive to S1 nuclease (41). First, the structural transition was demonstrated without any enzymatic or chemical modification. The pH dependency of the transition was remarkable (Figure 4). One can see that at acidic pH, the transition occurs under low torsional tension, while at neutral pH it is almost undetectable. Because pH-dependence had never been observed before for non-B-DNA conformations (cruciforms, Z DNA, bent DNA, etc), the investigators concluded that a novel DNA conformation had been formed, which they called H-DNA.

Treatment of the DNA topoisomer mixture with S1 nuclease before gel electrophoresis removed from the 2-D pattern topoisomers that

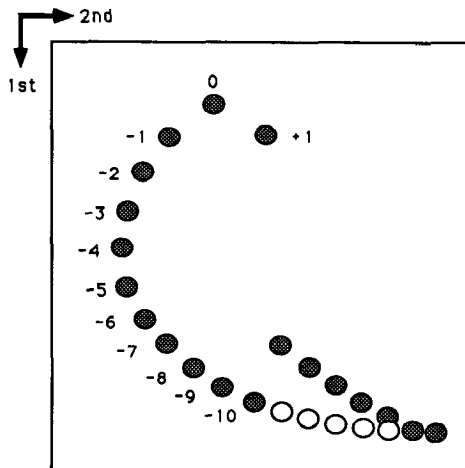


Figure 3 Schematic representation of two-dimensional gel electrophoresis. The structural transition occurs in topoisomers starting from number -11 and accompanies the release of four supercoils. Filled circles show the DNA topoisomers where the transition took place; empty circles show the mobility of corresponding topoisomers where no transition occurred.

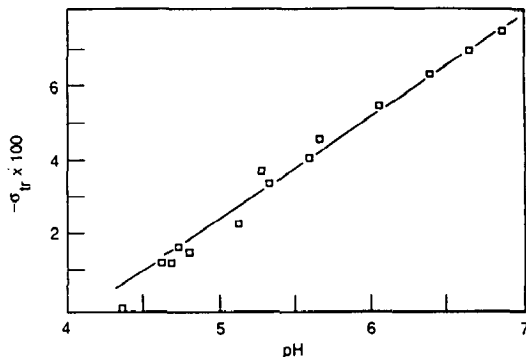


Figure 4 Dependence of the superhelical density of the B-to-H transition on pH for the $d(TC)_n \cdot d(GA)_n$ sequence (75).

adopted the H conformation. S1 cleavage sites are located in a small $d(G-A)_{16}$ stretch inside the whole 509-bp insert, so it seemed reasonable to conclude that the observed structural transition was due to this DNA segment. Because the mobility drop accompanying the transition corresponded to three superhelical turns, and an unusual structure was formed in the 32-bp-long stretch, the investigators concluded that the H form must be topologically equivalent to unwound DNA.

The N3 position of cytosine seemed the most probable protonation site. Among free nucleotides it has the highest pK_t value for protonation, and this pK_t value had been known to increase substantially when protonated cytosines were involved in different structures.

A theoretical consideration of B-to-H transitions noted by Lyamichev et al (75) allowed predictions about the number of protonation sites in the structure. Let m be the number of base pairs within DNA undergoing transition from B- to H-DNA. The free energy of transition will be:

$$\Delta F = \Delta F_o - (RT/r) \ln(1 + 10^{pK_t - pH}) \quad 1.$$

where r is the number of base pairs per protonated site in H-DNA, pK_t is the pK value of cytosine within the structure, and ΔF_o is the free energy of H-DNA formation with all protonation sites unoccupied. At pH values below pK_t , the stability of the protonated structure is strongly pH-dependent:

$$\Delta F = \Delta F_o^* + pH (RT/r) \quad 2.$$

Equation 2 predicts a linear dependence of the free energy of H-DNA stabilization on pH. As one can see from Figure 4, a linear pH-

dependence of supercoiling density required for H-DNA formation was indeed observed experimentally. Equation 2 leads to a simple equation for the superhelical density of transition:

$$\sigma_{tr} = (0.1/r) (pH_0 - pH), \quad 3.$$

where pH_0 is the value of the pH at which the H form is extruded in linear, topologically unconstrained DNA. The slope of the linear part of the experimental dependence curve (Figure 4) directly showed that the variable, r , was equal to 4, i.e. there was one protonation site per 4 bp of the insert. Because this curve represents data for the $d(G-A)_n$ stretch, half of the cytosines must be protonated for H-DNA formation to occur.

A model of H form DNA was proposed by Lyamichev et al (76) (Figures 1 and 5). It consists of an intramolecular triple helix formed by the pyrimidine strand and half of the purine strand; the other half of the purine strand is single stranded. As Figure 1 shows, this structure is topologically equivalent to unwound DNA. Two isoforms of H form are possible: one single-stranded in the 5' part of the purine strand and the other single-stranded in the 3' part (Figure 5). The existence of single-stranded purine stretches in H-DNA may explain its hyperre-activity to S1 nuclease.

TA*T and CG*C⁺ base triads stabilize the triple helix (Figure 2). Thymines or protonated cytosines from the third strand interact with adenines or guanines, respectively, from AT or GC base pairs via Hoogsteen rules (43). The protonation of cytosines is crucial for the formation of CGC⁺ base triads. This observation explains the pH dependency of the structural transition. It is also clear that only one half of the cytosines must be protonated for Hoogsteen hydrogen bonding, while the remaining cytosines form Watson-Crick hydrogen bonds. It

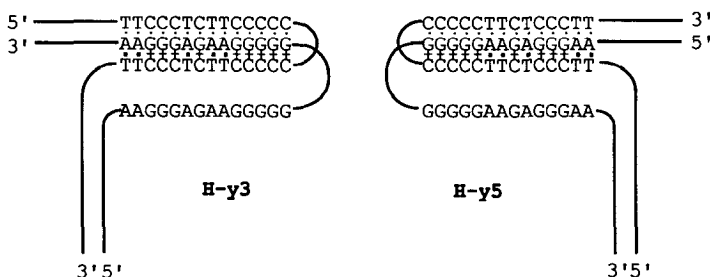


Figure 5 Two isoforms of H-DNA (91). Watson-Crick hydrogen bonds are labeled by points, nonprotonated Hoogsteen hydrogen bonds are shown by squares, and protonated Hoogsteen hydrogen bonds are shown by plus symbols.

is important that the triads be isomorphous such that good stacking in a triple helix is possible. The formation of triplexes, first suggested by Felsenfeld et al (28) for mixtures of homopurine and homopyrimidine polyribonucleotides, has been documented further in many studies (69, 92).

The H-DNA model suggested several obvious predictions. First, it should be true for several simple repeats including $d(G)n \cdot d(C)n$ and $d(A)n \cdot d(T)n$. However, the features of the transition for these two sequences should be different. For $d(G)n \cdot d(C)n$ one must expect a substantial pH dependence, and since r should be 2 rather than 4, the slope of the experimental dependency of σ_{tr} on pH should be twice that observed for $d(G-A)n \cdot d(T-C)n$. Using 2-D gel electrophoresis for cloned $d(G)n \cdot d(C)n$ sequences, the pH-dependent structural transition was indeed observed (77), and the maximal slope of the experimental pH-dependency curve actually corresponded to the predicted value $r = 2$ at 200 mM Na^+ .

For $d(A)n \cdot d(T)n$, one would expect a pH-independent structural transition, since TAT triads do not require base protonation. However, for a long time all attempts to detect this transition remained unsuccessful. Using single-strand-specific nucleases and chemicals as probes, Fox (30) found that $d(A)n \cdot d(T)n$ stretches adopt the H conformation under the influence of DNA supercoiling at pH 8. It turned out that even those very sensitive approaches detected H-DNA formation for only very long (69 bp) stretches. It is not yet clear why shorter sequences refuse to form H-DNA. This refusal may be due to an unusual helical conformation (B'-DNA) of $d(A)n \cdot d(T)n$ which is characterized by high propeller twist, contains additional bifurcated hydrogen bonds (22, 95), and does not wrap around nucleosomes (66). The rigidity and enhanced stability of this structure may prevent initiation of H-DNA formation for short tracts.

A less trivial prediction regarding sequence requirements of H-DNA formation is based on the importance of TA^*T and CG^*C^+ triad isomorphism. When the pyrimidine-rich strand folds back to form a triplex, cytosines from one half of the homopurine-homopyrimidine sequence should interact with GC but not AT base pairs in its other half. Conversely, thymines in one half should interact with AT but not GC base pairs from the other half. Thus, a homopurine-homopyrimidine sequence must be a mirror repeat to form H-DNA. Regular sequences $d(G-A)n \cdot d(T-C)n$, $d(G)n \cdot d(C)n$, and $d(A)n \cdot d(T)n$ are mirror repeats. One would expect that irregular homopurine-homopyrimidine sequences with mirror symmetry must adopt H conformation as well.

This hypothesis was proved by Mirkin et al (91) in studies of cloned

mirror repeated homopurine-homopyrimidine sequences with different point substitutions:

- AAGGGAGAAAGGGGTATAGGGGAAAGAGGGAA a.
 AAGGGAGAAAGGGGTATAGGGGGAAGAGGGAA b.
 AAGGGAGAAGGGGTATAGGGGAAAGAGGGAA c.
 AAGGGAGAGGGGGTATAGGGGGAAGAGGGAA d.

The first sequence is a perfect mirror repeat with the exception of the central TATA box. Both (b) and (c) contain an A-to-G substitution in either the left or right half of the mirror repeat (underlined). Though the homopurine-homopyrimidine nature of the stretch is undisturbed, both (b) and (c) are no longer mirror repeats. The last sequence combines both point substitutions, such that the mirror symmetry is restored. The first and last sequences must adopt the H conformation, while the second and the third should not—or should form it only at higher negative superhelicity. The results of 2-D gel-electrophoretic study of the corresponding supercoiled plasmid DNAs (Figure 6) were in full agreement with the above predictions.

It was concluded, therefore, that any mirror repeated homopurine-homopyrimidine sequence could form H-DNA. Such a sequence was called an H palindrome. Most natural S1-hypersensitive sequences show no sequence homology but contain prominent H palindromes (91). Thus, it is likely that the H form is the structural basis for DNA S1 hypersensitivity.

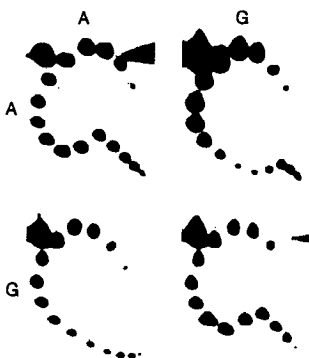


Figure 6 Two-dimensional gel electrophoresis of plasmids carrying H palindromes with point substitutions (91). In DNA samples with mirror repeated homopurine-homopyrimidine inserts (*top left, bottom right* corners), transition to the H form occurred at much lower superhelicity than in the other two DNAs, where mirror symmetry was destroyed by point mutations.

Final proof of the correctness of the H-form model was obtained by chemical probing of H-DNA (39, 40, 44, 48, 57, 120, 121). Several reagents were used that are reactive toward different bases and sensitive to particular DNA conformations (for review see 72, 128). Briefly, diethyl pyrocarbonate (DEPC) carboxyethylates purines at the N7-position in the single-stranded or Z conformation. Thus, it should react with single-stranded purines in the H form. Dimethyl sulfate (DMS) methylates the N7-position of guanines in the double- and single-stranded states. The N7-position of some guanines in H-DNA are involved in Hoogsteen hydrogen bonding. Methylation protection of guanines when in the H form is expected. Osmium tetroxide (OsO_4) forms osmate esters with the C5-C6 double bond of single-stranded thymines. It should interact with the thymines in the central part of the pyrimidine strand that are looped out in the H form. Chloroacetaldehyde (CAA) forms ethenoderivatives with the base-pairing positions of adenines, cytosines and, less prominently, guanines. It must interact with single-stranded purines and cytosines in H-DNA. All of these modified residues are detectable at a nucleotide resolution after piperidine cleavage, followed by sequencing gel electrophoresis.

Several groups found a unique pattern of chemical modification of different H palindromes in supercoiled DNA under acidic pH. The 5' portion of the purine-rich strand was hypersensitive towards DEPC and CAA, while its 3' half was relatively resistant (39, 40, 44, 48, 57, 120, 121). Conversely, for DMS, clear methylation protection of guanines in the 3' part of the purine strand, compared to its 5' half was observed (48, 121). Finally, OsO_4 and CAA were found to modify specifically the central part of the pyrimidine strand (39, 40, 44, 57, 120). These results support the H form model. They also reveal that different sequences preferentially adopt one of the two possible isomeric forms of the H-DNA in which the 5' part of the purine strand is unstructured.

The structural features responsible for the difference between the two isoforms have been identified by Htun & Dahlberg (45). They have shown that the isoform with the looped out 5' half of the purine strand (designated H-y3) is preferentially formed at high superhelical densities. (Note that DNA samples with a high level of DNA supercoiling were used in most chemical probing experiments.) The other isoform with the looped out 3' half of the purine strand (designated H-y5) is observed at lower superhelical density. A simple three-dimensional modeling of H-DNA formation showed that formation of the H-y3 releases one superhelical turn more than the H-y5 isoform, such that the former becomes more favorable at high superhelical density. Recent studies show that the mechanisms underlying preferential isomeriza-

tion into the H-y3 conformation are more complex. Apparently, the presence of bivalent cations makes the H-y5 isoform preferable (50). What is more surprising, the loop sequence plays an important role for the direction of isomerization (49, 114). Systematic studies of factors contributing to isomerization are yet to be provided.

The energetics of the B-to-H transition is still poorly understood. A simple thermodynamic treatment shows that when the cooperative transition of the n bp insert into the H form occurs, the superhelical density of the transition is determined from the equation (78):

$$\Delta F(n - 3) + F_n + \Delta G = 0, \quad 4.$$

where $n - 3$ base pairs actually form a triplex (due to the existence of a loop), ΔF is the free energy per base pair of the triplex part of H-DNA, F_n is the length-independent energy of nucleation of the H form, and ΔG is the change in superhelical energy accompanying the transition. ΔG may be also determined as (31):

$$\Delta G = 10RTN[(\sigma + n/N)^2 - \sigma^2], \quad 5.$$

where N is the total number of base pairs in DNA and σ is a superhelical density. Equations 4 and 5 yield:

$$-\sigma_{tr} = n/2N + \Delta F/20RT + (F_n - 3\Delta F)/20RTn. \quad 6.$$

This consideration allows one to estimate both ΔF and F_n by comparing experimentally determined superhelical densities of B-to-H transitions for regular homopurine-homopyrimidine repeats of varying length. Studies of this question for $d(G-A)_n \cdot d(T-C)_n$ and $d(G)_n \cdot d(C)_n$ inserts showed that the nucleation energy of the B-to-H transition is 18 kcal/mol (78), which is close to the corresponding value for cruciform extrusion.

To what extent may DNA tolerate deviations from the mirror symmetry and/or the homopurine-homopyrimidine character of a sequence? This issue was addressed by studying sequences like:



where X and Y are any DNA bases. For X = Y = A or X = Y = G, the sequences corresponded to perfect mirror repeats and easily adopted H conformation (see above). Using two-dimensional gel electrophoresis and S1 mapping, Belotserkovskii et al tested H-DNA formation for all other mismatches (10). They showed that the H conformation is actually possible for all X and Y, though in cases other than X = Y = A/G the transition requires greater superhelical stress. Quantitative analysis of the data made it possible to estimate the energy cost

of triplex formation due to all possible mismatched base triads (Table 1). These data were later confirmed in studies of intermolecular DNA triplexes with different noncanonical triads (79, 89). The only notable contradiction between the two series of data is the AT*G triad, which appeared advantageous in intermolecular triplexes (36, 124) but is not among the favorable triads in H-DNA (10). The reason for this difference is not clear. Energies are within the range of 3–6 kcal/mol—i.e. similar to the energy cost of mismatches in a B-DNA. Sequence requirements for triplex formation are thus similar to the sequence requirements for complementary recognition in a duplex. Though costly, the incorporation of noncanonical triads into H-DNA would significantly increase the number of sequences that could adopt this conformation.

As noted above, H-DNA formation requires DNA supercoiling and/or acidic pH. The positive effect of these factors is evident: DNA supercoiling compensates for the high nucleation energy of H-DNA formation, while protonation of cytosines makes the CG*C⁺ base triads favorable. Other stabilization factors have recently been revealed.

Belotserkovskii et al (9) studied the influence of oligonucleotides complementary to the single-stranded homopurine stretch in H-DNA on the stability of an intramolecular triplex. They found that such oligonucleotides stabilized H-DNA under high pH values (pH 7.0) where H-DNA alone rapidly flops into the B conformation. To explain this finding it is useful to consider the elementary steps in the H-to-B transition. The transition involves two energetically unfavorable processes: a disruption of a Hoogsteen pair, and an increase of negative superhelical stress. These processes are at least partly compensated by the energetically favorable formation of the Watson-Crick pair. In the case of oligonucleotide-associated H-DNA, formation of a Watson-Crick pair occurs at the cost of disruption of duplex base pairs formed by the oligonucleotide. In this case every step in H-to-B transition is less

Table 1 Energy cost for H-DNA formation for mismatched triads (10)

Triad	ΔE kcal/mol	Triad	ΔE kcal/mol
CG*T	2.9	AT*T	5.1
CG*A	2.9	TA*A	5.1
GC*T	3.6	AT*A	5.1
AT*C	3.6	GC*G	5.1
GC*C	4.4	TA*C	5.1
CG*G	4.4	TA*G	5.1
AT*G	5.1	GC*A	5.8

favorable by the energy of one Watson-Crick base pair, δF_0 . For a transition consisting of many sequential steps, the lifetime of the oligonucleotide-associated H-DNA is (2):

$$\vartheta' = \vartheta s^m, \quad 7.$$

where ϑ is the lifetime of H-DNA, $s = \exp(\delta F_0/kT)$, and m is the length of the oligonucleotide. Because the δF_0 for a Watson-Crick pair is about 10, association of a 14-nucleotide-long oligonucleotide with H-DNA would increase its lifetime by a factor of 10^{14} . Thus, addition of oligonucleotides kinetically traps DNA in the H conformation (9).

An interesting hypothesis of Reaban & Griffin (108; see also discussion in 109, 117) was that the H (or *H) form may be transiently extruded as a result of transcriptionally driven supercoiling and stabilized by the uptake of newly synthesized RNA molecules (see the section below on Detection of H-DNA *in vivo*). The data in Ref. 9 show that such stabilization is actually possible.

Another stabilizing factor was described by Lee and coworkers (37, 38). They found that pyrimidine-rich oligonucleotides complementary to the single-stranded purine stretch in H-DNA causes dimerization of plasmids carrying the corresponding H-forming sequences under conditions favorable for triplex extrusion. As discussed above, these oligonucleotides may associate with H-DNA or they may form an intermolecular triplex with the homopurine-homopyrimidine stretch in a B conformation (9). The authors suggested that dimerization is caused by the interaction of a single oligonucleotide with two independent plasmid molecules. It appeared that the presence of the polyamines spermine and spermidine favors plasmid dimerization under neutral pH. Independent studies showed that polyamines stabilize putative intermolecular triplexes as well (93). Because the density of a negative charge is higher in triplexes than in duplexes, the stabilizing effect of polyamines on triplex formation is probably caused by the reduction of repulsion between phosphate backbones upon binding of polyamines to DNA. Polyamines are normal components of eukaryotic cells, present in the nucleus (reviewed in 118). Thus, their stabilizing effect may play an important role in H-DNA formation *in vivo*.

*H-DNA

A related conformation was first described by Kohwi & Kohwi-Shigematsu (57). They studied the structure of $d(G)_n \cdot d(C)_n$ stretches in supercoiled DNA in the presence of Mg^{2+} ions using the single-stranded DNA-specific reagent chloroacetaldehyde. They observed an

unusual pattern of chemical hyperreactivity under neutral pH in the presence of magnesium ions. In this case, the central part of the purine strand and the 5' half of the pyrimidine strand were hyperreactive. The authors concluded that an H-DNA-like structure was formed in which the entire purine strand and half of the pyrimidine strand formed Py·Pu·Pu triplex. (They called it H'-DNA, but in current literature it is more often called *H-DNA, the designation we use here.) This conclusion was later supported by 2-D gel electrophoresis (98). The structural transition for the d(G)₄₆·d(C)₄₆ insert was observed at neutral pH in the presence of Mg²⁺ ions. The transition occurred at high superhelical stress ($\sigma_{tr} = -0.06$) and was accompanied by the release of 5 supercoils. Thus, like H-DNA, *H-DNA is topologically equivalent to unwound DNA.

For several years d(G)_n·d(C)_n tracts remained the only sequences shown to adopt *H conformation (57, 62). It was not clear whether this type of triplex could be formed by other sequences. Surprisingly, however, recent studies have shown that *H-DNA is much more versatile than canonical H-DNA with respect to sequence requirements.

Using nuclease and chemical probing, Azorin and his coworkers (12, 13) found that d(G-A)_n·d(T-C)_n stretches adopt *H configuration when in a supercoiled state. The structural units of such triplex are CG*G and TA*A base triads (Figure 7). In this case, Zn²⁺ rather than Mg²⁺ ions stabilize the structure (see below). In all of the above cases, the H-r3 isoform of *H-DNA dominated.

The analysis of *H-DNA benefited significantly from numerous studies of the formation of intermolecular Py·Pu·Pu triplexes during recent years. In particular, it was shown that nonorthodox TA*T and protonated CG*A⁺ base triads may successfully incorporate into such triplexes (5, 82). Less stable, but still possible, are TA*C and GC*T triads (6). These observations permitted the design of experiments which showed that *H-like structures can be formed by sequences that are neither homopurine-homopyrimidine nor mirror repeats.

One example is *H-DNA consisting of intervening CG*G and TA*T triads. To form such a structure, guanines in the purine-rich strand should be arranged in a mirror repeated way, while thymines concentrated in one half of the purine strand should be reflected by adenines in the other half (Figure 8). As a result, such sequences would be neither homopurine-homopyrimidine (because they contain thymines in otherwise purine-rich strands) nor mirror repeated (because A and T bases are positioned in an inverted repeated way). The asymmetric character of such sequences leads to the existence of two subclasses with adenines in either the 5' part or the 3' part of the purine-rich strand. This

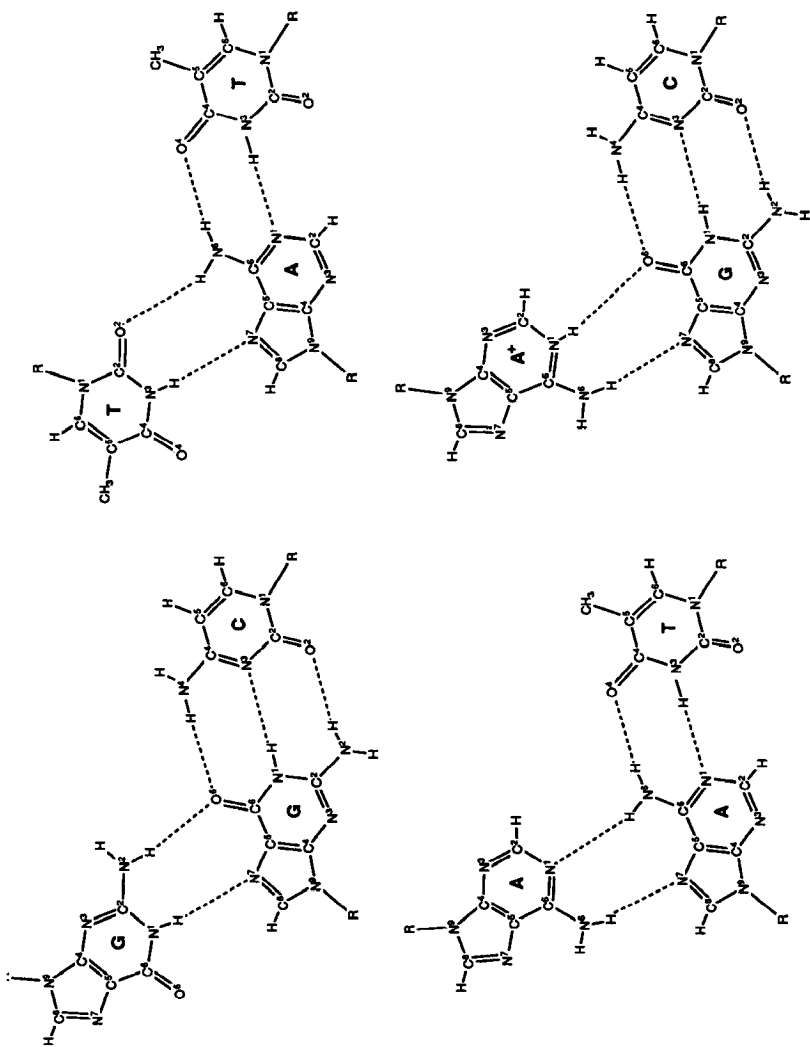


Figure 7 *H-DNA-forming triads: CG*G, TA*A, TA*T, and CG*A⁺.

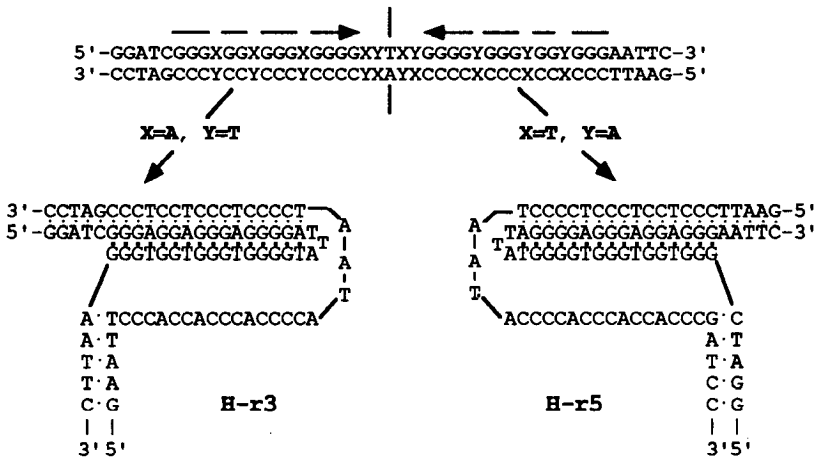


Figure 8 *H-DNA consisting of CG*G and TA*T base triads. GC base pairs are arranged as mirror repeats (shown by arrows flanking a vertical line representing the pseudosymmetry site), while AT base pairs are arranged as inverted repeats. Depending on the relative location of the adenines and thymines, either the H-r3 or the H-r5 isomer of *H-DNA is formed (27). Points, Watson-Crick hydrogen bonds; squares, Hoogsteen hydrogen bonds.

causes differences in triad composition between the two isoforms of *H-DNA for these subclasses (Figure 8). Using fine chemical probing of supercoiled plasmids with corresponding sequences in the presence of Mg^{2+} Dayn et al (27) showed that they indeed form the *H conformation. It is interesting that here the sequence content of the triplex appeared to be more important than topological differences between isoforms. Both sequences form triplexes composed of CG*G and TA*T triads; thus, H-r3 and H-r5 isoforms were stable for different sequences.

Malkov et al (82) observed *H-DNA formation for the sequence $G_{10}TTAA(AG)_5$ which is not mirror repeated. This asymmetry again leads to differences in triads between different isoforms: CG*G and CG*A triads in the H-r3 configuration and CG*G and TA*G triads in the H-r5 configuration. Chemical probing showed that only the H-r3 isoform was formed. Thus, *H-DNA is formed by CG*G and CG*A triads. It is surprising that this transition occurred under acidic pH and did not require bivalent cations—a combination of phenomena never observed before for *H-DNA. It was suggested, therefore, that protonation of adenines is crucial for stability of CG*A⁺ triads. Figure 7 represents all triads shown to be involved in *H-DNA formation. Be-

Table 2 Formation of Py·Pu·Pu triplexes in the presence of different bivalent cations (81)

Bivalent metal cation	$dC_n \cdot dG_n \cdot dG_n$		$d(TC)_n \cdot d(GA)_n \cdot d(GA)_n$	
	*H-DNA	Intermolecular triplex	*H-DNA	Intermolecular triplex
—	—	—	—	—
Ba ²⁺	No data	—	No data	—
Ca ²⁺	+	+	—	—
Mg ²⁺	+	+	—	—
Cd ²⁺	No data	+	+	+
Co ²⁺	—	+	+	+
Mn ²⁺	+	+	+	+
Zn ²⁺	—	+	+	+
Ni ²⁺	No data	+	—	+

cause those triads are not strictly isomorphous, triad isomorphism is not so crucial for the formation of *H-DNA as for that of H-DNA.

One of the most complicated questions about *H-DNA and Py·Pu·Pu triplexes in general is their dependence on bivalent cations. First of all, it is not clear why some bivalent cations are so efficient in stabilizing these triplexes, while others are not. This situation with H-DNA and Py·Pu·Py triplexes is quite different; in that case the requirement for hydrogen ions is evident. Second, cation requirements are different for different sequences (Table 2) (11–13, 23, 57, 59, 62, 85). While *H-DNA formed by $d(G)_n \cdot d(C)_n$ sequences is stabilized by Ca²⁺, Mg²⁺, and Mn²⁺, the same structure formed by $d(G-A)_n \cdot d(T-C)_n$ is formed in the presence of Zn²⁺, Mn²⁺, Cd²⁺, and Co²⁺. Similar effects are observed for intermolecular triplexes (81). It is not yet clear whether differences in cation requirements are caused by variations in neighboring triads or by changes in the GC content between different triplexes. Recent data indicate that moderate changes in GC content (from 75% to 63%) switch cation requirement from Mg²⁺ to Zn²⁺ for a particular sequence to form *H-DNA (110). However, because changing GC-content necessarily changes the sequence itself, only studies of many different *H-forming sequences can answer this question.

H-DNA-LIKE STRUCTURES

Triplex-containing structures, originally represented only by canonical H-DNA, actually comprise an entire family of structures (113) formed by various sequences, not necessarily homopurine-homopyrimidine.

Under superhelical stress they are stabilized by a variety of means, including acidic pH, bivalent cations, etc. We have already discussed the best-studied representatives of this family, namely H- and *H-DNA. Even those structures exhibit significant versatility in sequence requirements, in isomeric forms, and in conditions favoring their formation. In this section, we discuss other members of this family which are more unusual yet.

Nodule DNA

Two independent groups (63, 99) described a composite DNA triplex formed by sequences $d(G)_n \cdot d(C)_n$ and $d(A-G)_n \cdot d(T-C)_n$ (Figure 9). To adopt such a structure two thirds of the pyrimidine strand form an H-like triplex with one third of the purine strand, while the rest of the pyrimidine strand is involved in an *H-like triplex with two thirds of the purine strand. The resulting structure, called nodule DNA, contains only few unpaired DNA bases at the tips of both triplexes. Formation of nodule DNA was detected by chemical and nuclease probing, which showed sites reactive towards single-strand-specific agents and nucleases situated at one third the distance from both ends of the inserts.

As discussed above, H-DNA is stable at acidic pH while *H-DNA is stabilized by bivalent cations. Therefore, one would expect nodule DNA to exist within a narrow range of ambient conditions; and indeed nodule DNA formed by $d(G)_n \cdot d(C)_n$ is detected under mild acidic pH in the presence of bivalent cations (63). For $d(A-G)_n \cdot d(T-C)_n$ this structure was only observed at neutral pH in the presence of cobalt hexamine (99).

Nodule DNA formation was detected for relatively long regular homopurine-homopyrimidine inserts, while shorter inserts adopted orthodox intramolecular triplexes. This could be easily explained. Compared with H- and *H-DNA, the nodule conformation has a higher nucleation energy owing to a larger number of junctions between different conformations, whereas it has a lower elongation energy because the single-stranded region (inherent in H and *H forms) is lacking.

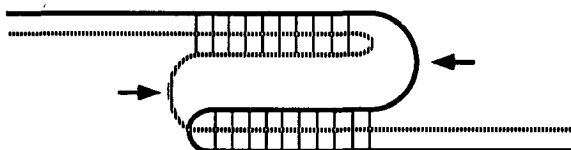


Figure 9 Nodule DNA (63, 99). Bold line, homopyrimidine strand; dashed line, homopurine strand. Arrows show single-stranded stretches that demonstrate nuclease and chemical hyperreactivity.

Eclectic DNA

Several studies described the formation of unusual DNA structures by telomeric repeats. Telomeric ends of eukaryotic chromosomes contain the redundant consensus $d[T_{1-3}(T/A)G_{1-4}]$, which is tandemly repeated up to 15 kb in length in the double-stranded state, followed by a 3' overhang containing two repeats of the G-rich strand alone (reviewed in 130). The structure of the ends of chromosomes is beyond the scope of this paper and is reviewed by Williamson in this volume (128a). We concentrate here on a special problem—an unusual structure adopted by telomeric sequences cloned within supercoiled plasmids.

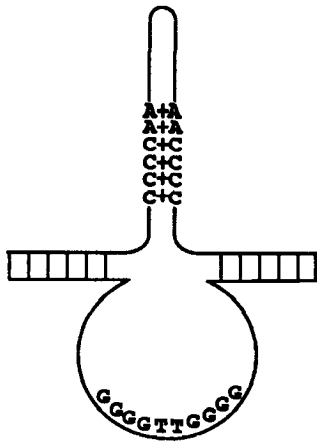
Interest in this problem was initiated by Budarf & Blackburn (17), who found that the cloned *Tetrahymena* telomeric repeat, $d(G_4T_2)_n$, exhibits hypersensitivity to S1 nuclease. Because these data looked similar to those that led to the discovery of H-DNA, this problem was attacked using 2-D gel electrophoresis, chemical and enzymatic probing, and oligonucleotide binding—i.e. all the methods previously successful for H-like DNA structures (9, 74, 122).

These studies of cloned repeats of *Tetrahymena* (T_2G_4) $_n$, human (T_2AG_3) $_n$, and yeast (TG_3TGTG) $_n$ revealed that all these sequences behave similarly. 1. Two-dimensional gel electrophoresis proved that these sequences adopt an unusual structure(s) topologically equivalent to unwound DNA. 2. As in the case of H-DNA, the transition was strongly pH dependent. 3. In all cases, complementary Py-rich, but not Pu-rich, oligonucleotides bound with plasmids carrying unusual structures. Thus, an unusual protonated DNA structure was formed, and at least some portion of the Pu-rich strand was single stranded.

These results were puzzling. They reflected what one would expect if telomeric sequences adopted an H-like structure, but the sequences themselves were not homopurine-homopyrimidine mirror repeats. In the first attempt to explain these data, for the case of the *Tetrahymena* telomeric sequence, Lyamichev et al proposed a so-called C,A-hairpin model (74). The main component of this structure is a hairpin formed by the pyrimidine-rich strand and stabilized by antiparallel C^*C^+ and A^*A^+ pairs. The purine-rich strand is mainly unstructured (Figure 10A). The requirement for base protonation explains the pH-dependence of the transition. The possibility of the formation of an antiparallel C^*C^+ hairpin was later confirmed in a study of single-stranded telomeric sequences using absorbance thermal denaturation, chemical probing, and gel electrophoresis (1).

However, the results of more detailed investigation by chemical and enzymatic probing of the human telomeric repeat proved to be incon-

A



B

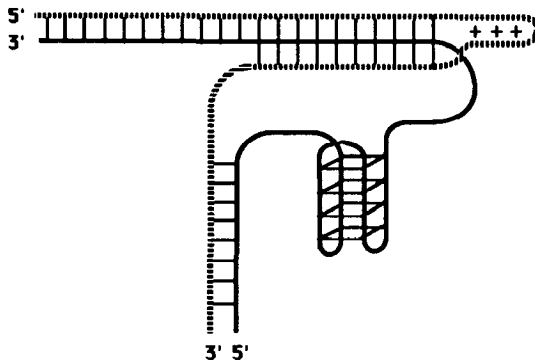


Figure 10 Models of unusual structures formed by telomeric repeats at acidic pH under superhelical stress. (A) C,A-hairpin (74). The C-rich strand forms a hairpin stabilized by C*C⁺ and A*A⁺ base pairs, while the G-rich strand is unstructured. (B) Eclectic DNA (122). The structure includes a nonorthodox triplex, quadruplex, C,A-hairpin, and unstructured regions. Bold line, G-rich strand; dashed line, C-rich strand.

sistent with the C,A-hairpin model. Clear-cut asymmetry in chemical reactivity within each strand was observed, whereas the C,A-structure implies a symmetrical pattern of modification. As a result, an alternative model, called "eclectic DNA," was proposed (Figure 10B; 122). It was called eclectic because it combines two unusual elements: 1. a

nonorthodox triplex containing both regular CG^*C^+ triads and mismatched AT^*A and AT^*T triads, and 2. a quadruplex formed by part of the Pu-rich strand (a detailed discussion of quadruplex DNA is presented elsewhere in this volume). As discussed above, both AT^*A and AT^*T triads discourage triplex formation. Nevertheless, the proposed model fits Voloshin et al's chemical and nuclease modification data best (122).

Two-dimensional gel-electrophoretic studies of cloned *Tetrahymena* repeats showed that the nucleation energy for the formation of the unusual structure is unexpectedly low: 7 kcal/mol (8). This low energy requirement may result from the formation of the quadruplex, which makes initiation of the eclectic structure more favorable than that of the H-DNA containing unstructured single-stranded region. The lower nucleation energy can make the formation of H-like triplexes with mismatched triads possible, especially for long sequences that dramatically release superhelical stress upon transition.

Elucidation of the protonated DNA structures formed by telomeric sequences awaits further studies. The exact configuration for a given telomeric repeat may depend on its length, base composition, ambient conditions, etc. Mutational analyses similar to those provided for H-DNA would also be useful in addressing the problem.

H-Like Structure Formed by Parallel-Stranded DNA

A highly unusual *H-like structure was described by Klysik et al (56). They studied a $d(A)_{15}\cdot d(T)_{15}$ segment in which the orientation of the two strands was parallel rather than antiparallel as in normal double-helical DNA (reviewed in 109). The authors synthesized a DNA duplex in which a central parallel-stranded segment was flanked by normal antiparallel-stranded regions, and inserted it into plasmid DNA. Circular DNAs with varying levels of DNA supercoiling were obtained in vitro, and the influence of torsional stress on the structure of the $ps\ d(A)_{15}\cdot d(T)_{15}$ stretch was studied by 2-D electrophoresis and chemical probing. Both methods revealed a structural transition within the parallel-stranded segment under the influence of DNA supercoiling. Chemical modification data were consistent with the formation of an intramolecular triplex stabilized by TA^*A base triads. (As in *H DNA, two homopurine strands in such a triplex are antiparallel; as in *H-DNA, only one of two possible isoforms dominated.) The nucleation energy of this unusual structure was estimated as 10 kcal/mol, which is lower than the H-DNA value of 18 kcal/mol. This reduction may result from the lower stability of the parallel-stranded insert relative to antiparallel sequences.

DETECTION OF H-DNA IN VIVO

Sequences that can form H-DNA are widespread throughout the genome of eukaryotes (7, 83) but are not common among eubacteria. However, the direct detection of H-DNA in eukaryotic cells is difficult owing to the great complexity of genomic DNA. Therefore, most of the studies of H-DNA in vivo exploit *Escherichia coli* as a convenient model system and artificially design constructs containing triplex-forming sequences.

The first direct detection of H-DNA in vivo was carried out by Karlovsky et al (51). Investigators there studied the structure of a d(T-C)₁₆-d(A-G)₁₆ stretch cloned into a plasmid vector. They employed OsO₄ as a tool for H-DNA detection in vivo. It modifies the central part of the homopyrimidine strand when in H-DNA but not when in B-DNA. It can also penetrate bacterial cells, making possible chemical probing in vivo. *E. coli* cells containing the corresponding plasmid were incubated with OsO₄, then plasmid DNA was isolated, and the sites of in vivo modification were detected by sequencing. H-DNA could be formed in *E. coli* cells under a specific set of conditions. First, it was detected only in cells treated by the antibiotic chloramphenicol prior to modification. Chloramphenicol, an inhibitor of bacterial protein synthesis, causes (among other effects), an increase in superhelical density of plasmids within cells up to -0.055 (26). This density increase was shown to provoke the formation of unusual DNA structures, including Z-DNA and cruciforms (26, 115). Thus, it is reasonable to suggest that this same increase in superhelicity caused H-DNA formation. Second, H-DNA was observed only if cells were additionally pre-incubated in a buffer with a pH below 5.2. Such incubation should drop the intracellular pH somewhat, making H-DNA formation more likely.

A more detailed study of H-DNA formation in *E. coli* cells was provided by Usery & Sinden (119). These authors used chemical probing of a plasmid bearing an H-forming d[(G-A)₇TA(G-A)₇] insert. The central TA dinucleotide in this sequence is a target for trimethylpsoralene photobinding in double-helical DNA but not in the H form. Since this chemical penetrates *E. coli* cells, it is a convenient tool for the detection of H-DNA in vivo. Using this approach Usery & Sinden found that an H-y3 isomer of H-DNA may form in normal cells without chloramphenicol treatment, but the higher level of DNA supercoiling in topoisomerase I mutants enhances its formation. Growth conditions were also significant: H-DNA formation was most prominent in bacterial cells grown in a synthetic medium at pH 5.0. These authors proposed that this pH supposedly decreases intracellular pH to 7.1 instead of

the standard 7.8. Finally, H-DNA occurrence depended on the stage of *E. coli* culture growth, being more pronounced in stationary than in exponential phase.

Similar approaches were used for the detection of ³H-DNA. Kohwi et al (60) used chloroacetaldehyde for the detection of this structure in *E. coli* cells. Chloroacetaldehyde specifically modifies one half of the homopyrimidine strand in ³H conformation and may enter bacterial cells. Kohwi et al found that d(G)_n·d(C)_n stretches within bacterial plasmids may form ³H-DNA. This structure was detected for inserts longer than 35 bp capable of adopting ³H conformation under physiological ionic strength at reasonable supercoiling density (58). ³H-DNA was observed only after chloramphenicol treatment of cells.

In summary, these studies show that H- and ³H-DNA may in principle exist within bacterial cells. Clearly the level of DNA supercoiling in vivo is the major limiting factor in the formation of these structures. Environmental conditions during *E. coli* growth also significantly contribute to the appearance of H-DNA.

Recently it became clear that, while the steady-state level of DNA supercoiling is determined by the balance of DNA gyrase and Topo I (reviewed in 126), the local level of supercoiling strongly depends on transcription. Elongating RNA polymerase creates domains of high negative and positive supercoiling upstream and downstream of itself, respectively (73) which may influence the formation of unusual DNA structures. Indeed, transcription was found to drive cruciform and Z-DNA formation when corresponding DNA sequences were located upstream of promoters (25, 106). Quite recently similar observations were made for ³H-DNA (61). d(G)_n·d(C)_n stretches were cloned upstream from a regulated promoter in an *E. coli* plasmid, and the structure of this stretch was studied by chloroacetaldehyde probing. Investigators found that induction of transcription provokes ³H-DNA formation for stretches longer than 32 bp. ³H-DNA formation in turn stimulated homologous recombination between direct repeats flanking the d(G)_n·d(C)_n insert. The authors suggested that the change in DNA geometry accompanying H-DNA formation brings flanking DNA sequences into close proximity, stimulating strand exchange.

There are also indirect indications of H-DNA formation in vivo. Kly-sik and coworkers (100) found that a GATC site situated in an H-forming sequence is undermethylated in vivo by *dam* methyltransferases. In vitro, formation of H-DNA prevented *dam* methylation. The authors suggested that the formation of H-DNA in vivo may be responsible for the observed undermethylation.

Two other studies concerned the influence of H motifs on transcrip-

tion in vivo. A homopurine-homopyrimidine mirror repeat artificially designed in the transcribed portion of a bacterial gene significantly decreased gene expression due to premature transcriptional termination (111). Sarkar & Brahmachari speculated that H-DNA formation is responsible for transcriptional termination. Reaban and Griffin suggested that a possible mechanism of such termination is the interaction of an RNA chain with the H-DNA transiently extruded by local negative supercoiling upstream of RNA polymerase (108). Conversely, an H motif placed upstream of the β -lactamase promoter increased its activity (52). Because this promoter's activity strongly depends on DNA topology, the authors suggested that changes in topology associated with H-DNA formation are responsible for the elevation of transcription.

Other interpretations of these results are also possible. For example, chloramphenicol treatment completely abolished the undermethylation effect described above. Because in most cases chloramphenicol stimulated H-DNA formation (see above), this result does not support the proposed model. An alternative explanation is that some proteins in *E. coli* cells can bind H motifs that in turn effect different genetic processes in vivo. Additional studies are required to clarify this issue.

The only data on triplex DNA detection in eukaryotic cells were obtained using antibodies against triplex-helical DNA (70). It was found that these antibodies interact with eukaryotic chromosomes. Note, however, that interpretation of these interesting results is hindered by the fixation procedure used for antibody staining.

H-DNA AND DNA POLYMERIZATION

The biological role of H-DNA, if any, remains to be established. Recently, it has become clear that intramolecular triplexes significantly influence DNA polymerase activity in vitro. Dayn et al (27) studied DNA polymerization on DNA templates containing sequences that form *H-DNA consisting of CG*G and TA*T base triads (Figure 8). As discussed above, in supercoiled DNA, either the H-r3 or the H-r5 isomer of *H-DNA is formed, depending on the design of a particular sequence. It turned out that DNA polymerase terminates at specific sites on both DNA chains within supercoiled templates containing such sequences. The location of the termination sites was different for distinct *H-forming stretches but in all cases coincided precisely with triplex boundaries as defined by chemical probing. Dayn et al suggested, therefore, that formation of *H-DNA prior to DNA synthesis causes DNA polymerase to terminate. The difference in the location

of termination sites is not surprising because the H-r5 and the H-r3 isomers contrast in chain polarity. Since DNA polymerase moves in a 3' to 5' direction along the template strand, it would face the triplex either at the end or the middle of an *H motif, depending on the nature of the isomer (Figure 11).

An unexpected twist in DNA polymerase–triplex relations came from studies showing that H-DNA may be formed in the course of DNA polymerization. The first data obtained by Manor and his co-authors showed that $d(G-A)_n$ or $d(C-T)_n$ inserts within single-stranded

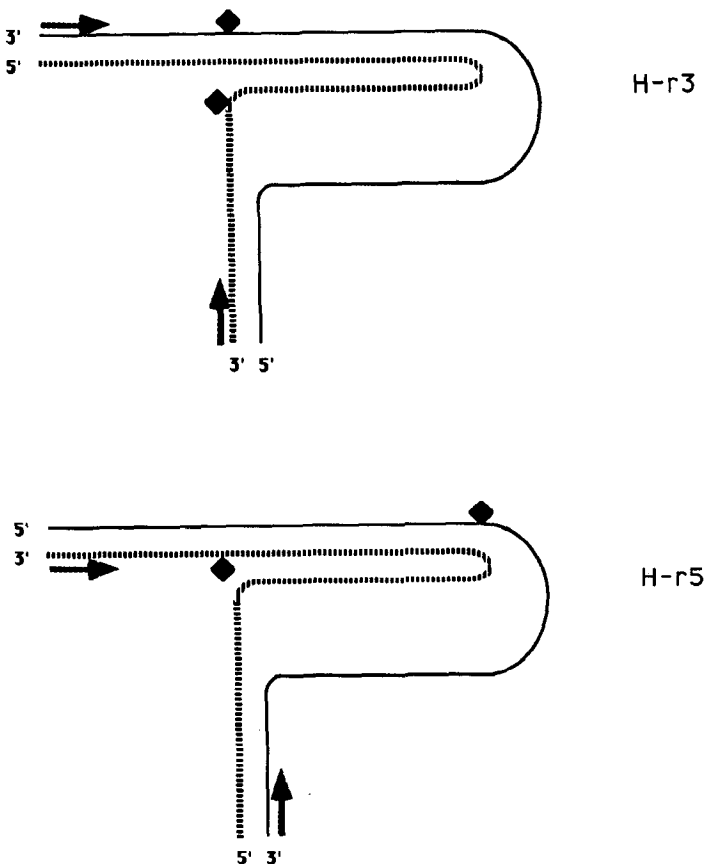


Figure 11 *H-DNA within a template causes DNA polymerase to terminate (27). Location of termination sites (shown by diamonds) differs between the H-r3 and the H-r5 isomers. Solid line, Py-rich strand; dashed line, Pu-rich strand. Growing DNA chains are shown by arrows.

DNA templates cause partial termination of DNA polymerases (4, 67). Termination sites were always located in the center of the insert. In explanation of these results, Baran et al (4) suggested that when the newly synthesized DNA chain reaches the center of the homopolymer sequence, the remaining homopolymer stretch folds back, forming a stable triplex (Figure 12A). As the result, DNA polymerase finds itself in a kind of trap and is unable to continue elongation.

For double-helical DNA templates another mechanism of polymerase-driven triplex formation is also possible. Many DNA polymerases can displace the nontemplate DNA strand in the course of DNA synthesis (reviewed in 65). The displaced strand may fold back, promoting the formation of an intramolecular triplex downstream of the replication fork. Conditions for DNA polymerization *in vitro*—i. e. neutral pH and high magnesium concentration—are optimal for the formation of *H-DNA. Thus, if a DNA polymerase meets a potential *H-forming sequence, displacement of the purine-rich strand could provoke triplex formation (Figure 12B) which, in turn, could lead to termination of DNA synthesis. This hypothesis was recently proven by Samadashwily et al (110) in their studies of DNA polymerization on open circular DNAs. T7 DNA polymerase terminated exactly at the pseudosym-

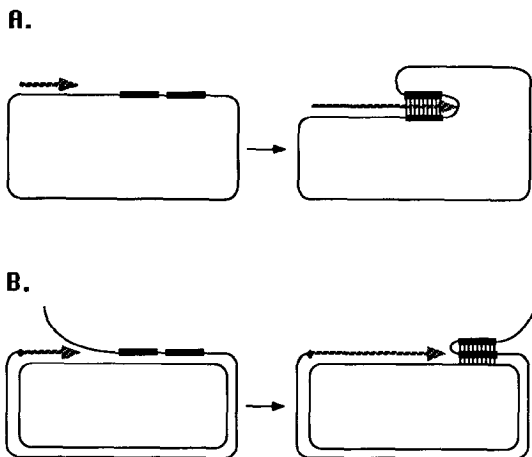


Figure 12 DNA polymerase-driven triplex formation blocks polymerization. Black boxes represent the two halves of a homopurine-homopyrimidine mirror repeat involved in the formation of an intramolecular triplex; the striated arrow depicts the newly synthesized DNA chain. (A) Single-stranded DNA template (4). (B) Double-stranded DNA template (110). The diamond shows the original nick in the double-helical template providing a 3'-OH end for DNA polymerase.

metry site of *H-motifs when the purine-rich strand was displaced. When the pyrimidine-rich strand was displaced, no termination occurred. The model gained support from the observation that mutations within such motifs which destroyed *H potential released polymerization termination, while compensatory mutations restoring the *H-forming ability restored termination as well.

Prevention of further elongation by DNA polymerase-driven triplex formation may serve as an efficient mechanism of DNA polymerase self-termination. Consequently, H motifs may be considered suicidal sequences for DNA polymerization. Clearly, the actual replication fork contains not only DNA polymerase but a complex of replication proteins, including helicases, SSB proteins, primases, topoisomerases, etc. Recently it was found that the DNA-helicase activity of the SV40 T-antigen is inhibited by triplex formation (101). However, detailed studies on the role of H-DNA in reconstituted replication systems have not yet been provided.

Only fragmentary data are available on the role of H motifs in the regulation of replication *in vivo*. One example concerns polyomavirus-transformed rat cells. Polyomavirus DNA integrates in a particular chromosomal site (88), and treatment of the cells with mitomycin C leads to the amplification of virus DNA and adjacent cellular sequences (3). The boundary of the amplified DNA segment lies within a homopurine-homopyrimidine stretch, d(G-A)₂₇-d(C-T)₂₇, located 2 kb from the viral DNA. Because this sequence causes premature termination of replication *in vitro*, it was suggested that this DNA motif could be a natural replication terminator. This hypothesis gained support from the observation that when the corresponding DNA segment was cloned into SV40 DNA it caused a pausing in replication fork progression (107).

Another case is the *dhfr*-amplicon. It contains an unusual cluster of simple-repeats, including d(A-C)₁₈, d(A-G)₂₁, d(G)₉, and d(A-G)₂₇, which is located 2 kb 3' of the *dhfr* origin of replication (18, 90). This cluster, when cloned at either side of the SV40 origin in the pSV011 episome, reduced its replication up to twofold, and, when placed on both sides of the origin, blocked replication almost completely (15). Based on *in vitro* data, the authors speculated that the triplex-forming repeats within this cluster may play a key role in termination of DNA replication *in vivo* (15).

Although these data make the idea of H-DNA involvement in the regulation of replication promising, this hypothesis is far from proved. Alternative ideas (e.g. protein binding, or changes in chromatin structure) may explain the results as well. Future studies, including direct

detection of H-DNA in vivo, are crucial for the evaluation of this hypothesis.

H-MOTIFS IN EUKARYOTIC PROMOTERS

As noted above, the studies that led to the discovery of H-DNA were initiated by the detection of S1 hypersensitivity within eukaryotic promoters—hypersensitivity associated with homopurine-homopyrimidine stretches (reviewed in 128). In many cases, it was then found that the formation of H-DNA did indeed account for nuclease hypersensitivity. In other cases, however, the formation of canonical H-DNA seems unlikely because of the insufficient length of homopurine-homopyrimidine motifs or the lack of proper symmetry within them. Thus other structures, which may be related to H-DNA, must be responsible for nuclease hypersensitivity. There is a growing body of data on the structure-functional dissection of H motifs in eukaryotic promoters. We describe below several of the best-studied cases showing the significance of H motifs in promoter function.

One well-characterized example is the promoter of the *Drosophila* heat-shock gene *hsp26*. At approximately the -100 position of the promoter it contains a $d(C-T)_{10} \cdot d(A-G)_{10}$ stretch (34) crucial for the heat shock response of the promoter in vivo (35). Careful nuclease and chemical probing led to the conclusion that this sequence adopts an H conformation in vitro (34). However all attempts to detect H-DNA in vivo using genomic footprinting have been unsuccessful (35). A *Drosophila* nuclear protein that binds preferentially to the pyrimidine-rich strand of this sequence was found (34). It is not yet clear whether or not the H-forming potential of the target sequence is important for protein binding.

Another case is the promoter of the chicken β^A -globin gene, which contains a $d(G)_{18} \cdot d(C)_{18}$ string approximately 200 bp upstream of the transcriptional start site. Detailed structural studies of this stretch (57, 62) revealed that it can exist in either H or *H conformation, depending on environmental conditions in vitro. No attempt has yet been made to determine the structure of this string in vivo. Chicken erythroid cells contain a zinc-dependent protein called BGP1 that specifically targets this sequence (21). This protein binds to the G-rich strand of the stretch.

The human *c-myc* gene contains a nuclease-hypersensitive element, called NHE or NSE, located 125 bp upstream of the P1 promoter start site (14, 54, 104). Several groups have suggested that formation of H-like DNA is responsible for nuclease sensitivity in this promoter (29, 104). Two different structures have been discussed, both containing

numerous mismatches and looped out single- and double-stranded DNA stretches. The NSE is important for *c-myc* transcription (24, 104). This sequence serves as a binding site for a protein(s), presumably a transcriptional activator (24, 104). Cloning and sequencing of a gene coding for a protein that binds to the *c-myc* NSE, called NSEP-1, has been reported (64). This protein binds preferentially to the pyrimidine-rich strand—not only to the NSE in the *myc* promoter, but also to several other H-forming sequences from different promoters, including *c-Ki-ras*, EGF-R, etc. The authors suggested that NSEP-1 recognizes structural rather than sequence similarities of the homopurine-homopyrimidine stretches, presumably their ability to form H-DNA. Postel et al have described a different protein binding to the same element, called PuF (104). This protein has recently been cloned and sequenced. It appears to be identical to the previously described nm23-H2 nucleoside diphosphate kinase, which is a candidate suppressor of tumor metastasis (103).

Homopurine-homopyrimidine stretches showing S1 hypersensitivity are common in so-called TATA-less promoters. These promoters are unusually GC-rich, have no TATA or CAAT boxes, and contain numerous transcriptional start sites. It is not immediately clear whether the elevated frequency of the homopurine-homopyrimidine stretches has a functional significance or is a simple result of high sequence redundancy. One well-studied case is the mouse *c-Ki-ras* oncogene promoter. The S1-hypersensitive element is located approximately 100 bp upstream of the multiple transcriptional start-sites area and is important for promoter function (42). A nuclear protein(s) that specifically binds to this element was identified by band-shift assay (42). Pestov et al (102) found that H-DNA formed by a 27-bp-long H palindrome is responsible for nuclease cleavage. It is thus reasonable to believe that the H-forming motif in the *c-Ki-ras* promoter is a positive *cis*-acting transcriptional element. Similar conclusions have been reached for several other TATA-less promoters, including human *c-ets-2* (86) and EGF-R (46).

All the cases described here share significant similarities. H-forming sequences appear to be important for transcriptional regulation. These sequences serve as targets for nuclear proteins, presumably transcriptional activators. And these proteins, when characterized, often bind preferentially to either the purine-rich or the pyrimidine-rich strand of the H motifs. This unusual binding pattern may dramatically influence the equilibrium between different DNA conformations of the promoter.

Numerous hypotheses have been offered about the role of H-motifs in promoter functioning. In our view it is reasonable to suggest that

local changes in DNA structure may regulate the interaction between promoter DNA and specific regulator proteins. In support of this hypothesis are numerous reports describing eukaryotic proteins preferentially binding to either the homopurine- or the homopyrimidine strand of H motifs (16, 53, 94). Recently, the partial characterization of a protein that preferentially binds to triple-helical DNA has also been reported (55). A study of Kohwi & Kohwi-Shigematsu (58) additionally support this idea. The authors studied the influence of $d(G)_n$ stretches of varying length on the activity of a downstream minimal promoter. It turned out that an initial increase in the length of the $d(G)_n$ stretch caused a progressive activation of transcription, but a further increase in the length restored the original level of promoter activity. There was a clear reverse correlation between the ability of a stretch to form *H configuration in vitro and its ability to activate transcription in vivo. The authors concluded that short $d(G)_n$ stretches served as binding sites for a transcriptional activator; while longer stretches adopted a triplex configuration that prevented activator binding.

Despite the wealth of data and hypotheses, no direct evidence indicates that the structural features of H motifs are involved in transcriptional regulation in vivo. Further studies addressing this issue are needed.

CONCLUSIONS

As in the cases of other unusual DNA structures (Z-DNA, parallel-stranded DNA, cruciforms, quadruplexes), the triplex H form and its related structures have been extensively studied in solution in recent years using a variety of physical and chemical methods. A large body of data has been accumulated about the versatility of H-like DNA structures and the factors affecting their stability. A variety of intramolecular triplexes can be formed by quite different sequences under a wide range of ambient conditions. In coming years we will likely encounter many more such structures.

As in the cases of other unusual structures, the biological roles of H-DNA remain obscure in spite of numerous speculations and repeated attempts to attack the problem. In recent years, accurate in vitro studies have demonstrated clear-cut effects of H-DNA-forming sequences on the function of different purified enzymes, the first of which was DNA polymerase. Both the crucial role of intramolecular triplexes in these effects and the possibility of H-DNA formation in vivo have been demonstrated; but we still lack a single example of the direct involvement of H-DNA in any biological process. Nevertheless, the versatility

of the intramolecular triplexes makes it unlikely, in our opinion, that these structures play no role in normal or pathological biological processes.

ACKNOWLEDGMENTS

We thank our colleagues who have participated in the discovery and studies of H-DNA, including VI Lyamichev, ON Voloshin, BP Belotserkovskii, VA Malkov, AG Veselkov, ON Danilevskaya, AV Vologodskii, VN Dobrynin, VN Soyfer, A Dayn, DG Pestov, GM Samadashwily, M Vojtiskova, and E Palecek; Randal Cox for proofreading the manuscript; and the NIH, ACS, and Soros Foundation for financial support.

Any Annual Review chapter, as well as any article cited in an Annual Review chapter, may be purchased from the Annual Reviews Preprints and Reprints service.
1-800-347-8007; 415-259-5017; email: arpr@class.org

Literature Cited

1. Ahmed S, Henderson E. 1992. Formation of novel hairpin structures by telomeric C-strand oligonucleotides. *Nucleic Acids Res.* 20:507-11
2. Anshelevich VV, Vologodskii AV, Lukashin AV, Frank-Kamenetskii MD. 1984. Slow relaxational processes in the melting of linear biopolymers: a theory and its application to nucleic acids. *Biopolymers* 23:39-58
3. Baran N, Lapidot A, Manor H. 1987. Unusual sequence element found at the end of an amplicon. *Mol. Cell. Biol.* 7:2636-40
4. Baran N, Lapidot A, Manor H. 1991. Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)_n and d(GA)_n tracts. *Proc. Natl. Acad. Sci. USA* 88:507-11
5. Beal PA, Dervan PB. 1991. Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science* 251:1360-63
6. Beal PA, Dervan PB. 1992. The influence of single base triplet changes on the stability of a pur-pur-pyr triple helix determined by affinity cleaving. *Nucleic Acids Res.* 20:2773-76
7. Beasty AM, Behe MJ. 1988. An oligopurine sequence bias occurs in eukaryotic viruses. *Nucleic Acids Res.* 16:1517-28
8. Belotserkovskii BP, Frank-Kamenetskii MD. 1993. Study of pH-dependent structural transition in telomeric sequences by two-dimensional gel electrophoresis. *Electrophoresis* 14:266-70
9. Belotserkovskii BP, Krasilnikova MM, Veselkov AG, Frank-Kamenetskii MD. 1992. Kinetic trapping of H-DNA by oligonucleotide binding. *Nucleic Acids Res.* 20:1903-8
10. Belotserkovskii BP, Veselkov AG, Filippov SA, Dobrynin VN, Mirkin SM, Frank-Kamenetskii MD. 1990. Formation of intramolecular triplex in homopurine-homopyrimidine mirror repeats with point substitutions. *Nucleic Acids Res.* 18:6621-24
11. Beltran R, Martinez-Balbas A, Bernues J, Bowater R, Azorin F. 1993. Characterization of the zinc-induced structural transition to *H-DNA at a d(GA·CT)₂₂ sequence. *J. Mol. Biol.* 230:966-78
12. Bernues J, Beltran R, Casanovas JM, Azorin F. 1989. Structural polymorphism of homopurine-homopyrimidine sequences: the secondary DNA structure adopted by a d(GA·CT)₂₂ sequence in the presence of zinc ions. *EMBO J.* 8:2087-94
13. Bernues J, Beltran R, Casanovas JM, Azorin F. 1990. DNA sequence and metal-ion specificity of the formation of *H-DNA. *Nucleic Acids Res.* 18:4067-73
14. Boles TC, Hogan ME. 1987. DNA

572 MIRKIN & FRANK-KAMENETSKII

- structure equilibria in the human *c-myc* gene. *Biochemistry* 26:367-76
15. Brinton BT, Caddle MS, Heintz NH. 1991. Position and orientation-dependent effects of a eukaryotic Z-triplex motif on episomal DNA replication in COS-7 cells. *J. Biol. Chem.* 266:5153-61
 16. Brunel F, Alzari PM, Ferrara P, Zakin MM. 1991. Cloning and sequencing of PYBP, a pyrimidine-rich specific single strand DNA-binding protein. *Nucleic Acids Res.* 19:5237-45
 17. Budarf M, Blackburn EH. 1987. S1 nuclease sensitivity of a double-stranded telomeric DNA sequence. *Nucleic Acids Res.* 15:6273-92
 18. Caddle MS, Lussier RH, Heintz NH. 1990. Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon. *J. Mol. Biol.* 213:19-33
 19. Cantor CR., Efstratiadis A. 1984. Possible structures of homopurine-homopyrimidine S1-hypersensitive sites. *Nucleic Acids Res.* 12:8059-72
 20. Christophe D, Cabrer B, Bacolla A, Targovnik H, Pohl V, Vassart G. 1985. An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream from the human thyroglobulin gene. *Nucleic Acids Res.* 13:5127-44
 21. Clark SP, Lewis CD, Felsenfeld G. 1990. Properties of BGPI, a poly(dG)-binding protein from chicken erythrocytes. *Nucleic Acids Res.* 18:5119-26
 22. Coll M, Frederick CA, Wang AH, Rich A. 1987. A bifurcated hydrogen-bonded conformation in the d(A·T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. USA* 84:8385-89
 23. Collier DA, Wells RD. 1990. Effect of length, supercoiling, and pH on intramolecular triplex formation. Multiple conformers at pur-pyr mirror repeats. *J. Biol. Chem.* 265:10652-58
 24. Davis TL, Firulli AB, Kinniburgh A. 1989. Ribonucleoprotein and protein factors bind to an H-DNA-forming *c-myc* DNA element: Possible regulators of the *c-myc* gene. *Proc. Natl. Acad. Sci. USA* 86:9682-86
 25. Dayn A, Malkhosyan S, Mirkin SM. 1992. Transcriptionally driven cruciform formation *in vivo*. *Nucleic Acids Res.* 20:5991-97
 26. Dayn A, Malkhosyan SR, Duzhii DI, Lyamichev VI, Panchenko YA, Mirkin SM. 1991. Formation of (dA-dT)_n cruciforms in *Escherichia coli* cells under different environmental conditions. *J. Bacteriol.* 173:2658-64
 27. Dayn A, Samadashwily GM, Mirkin SM. 1992. Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. *Proc. Natl. Acad. Sci. USA* 89:11406-10
 28. Felsenfeld G, Davies DR, Rich A. 1957. Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* 79:2023-24
 29. Firulli AB, Maibenko DC, Kinniburgh AJ. 1992. The identification of a tandem H-DNA structure in the *c-myc* nuclease sensitive promoter element. *Biochem. Biophys. Res. Commun.* 185:264-70
 30. Fox KR. 1990. Long (dA)_n(dT)_n tracts can form intramolecular triplexes under superhelical stress. *Nucleic Acids Res.* 18:5387-91
 31. Frank-Kamenetskii MD, Vologodskii AV. 1984. Thermodynamics of the B-Z transition in superhelical DNA. *Nature* 307:481-82
 32. Frank-Kamenetskii MD. 1990. DNA supercoiling and unusual structures. In *DNA Topology and Its Biological Effects*, ed. NR Cozzarelli and J Wang, pp. 185-215. Cold Spring Harbor Lab. Press
 33. Frank-Kamenetskii MD. 1992. Protonated DNA structures. *Methods Enzymol.* 211:180-91
 34. Gilmour DS, Thomas GH, Elgin SCR. 1989. *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science* 245:1487-90
 35. Glaser RL, Thomas GH, Siegfried E, Elgin SCR, Lis JT. 1990. Optimal heat-induced expression of the *Drosophila hsp26* gene requires a promoter sequence containing (CT)_n(GA)_n repeats. *J. Mol. Biol.* 211:751-61
 36. Griffin LC, Dervan PB. 1989. Recognition of thymine adenine base pairs by guanine in a pyrimidine triple helix motif. *Science* 245:967-71
 37. Hampel KJ, Burkholder GD, Lee JS. 1993. Plasmid dimerization mediated by triplex formation between polypyrimidine-polypurine repeats. *Biochemistry* 32:1072-77
 38. Hampel KJ, Crosson P, Lee JS. 1991. Polyamines favor DNA triplex formation at neutral pH. *Biochemistry* 30:4455-59
 39. Hanvey JC, Klysik J, Wells RD. 1988.

- Influence of DNA sequence on the formation of non-B right-handed helices in oligopurine-oligopyrimidine inserts in plasmids. *J. Biol. Chem.* 263:7386-96
40. Hanvey JC, Shimizu M, Wells RD. 1988. Intramolecular DNA triplexes in supercoiled plasmids. *Proc. Natl. Acad. Sci. USA* 85:6292-96
 41. Hentchel CC. 1982. Homopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S1 nuclease. *Nature* 295:714-16
 42. Hoffman EK, Trusko SP, Murphy M, George DL. 1990. An S1 nuclease-sensitive homopurine homopyrimidine domain in the c-Ki-ras promoter interacts with a nuclear factor. *Proc. Natl. Acad. Sci. USA* 87:2705-9
 43. Hoogsteen K. 1963. The crystal and molecular structure of a hydrogen-bonded complex between 1 methylthymine and 9 methyladenine. *Acta Crystallogr.* 16:907-16
 44. Htun H, Dahlberg JE. 1988. Single strands, triple strands and kinks in H-DNA. *Science* 241:1791-96
 45. Htun H, Dahlberg JE. 1989. Topology and formation of triple-stranded H-DNA. *Science* 243:1571-76
 46. Johnson AC, Jinno Y, Merlino GT. 1988. Modulation of epidermal growth factor receptor proto-oncogene transcription by a promoter site sensitive to S1 nuclease. *Mol. Cell. Biol.* 8:4174-84
 47. Johnson D, Morgan AR. 1978. Unique structures formed by pyrimidine-purine DNAs which may be four-stranded. *Proc. Natl. Acad. Sci. USA* 75:1637-41
 48. Johnston BH. 1988. The S1-sensitive form of d(C-T)_n-d(A-G)_n: chemical evidence form a three-stranded structure in plasmids. *Science* 241:1800-4
 49. Kang S, Wells RD. 1992. Central non-pur-pyr sequences in oligo (dG-dC) tracts and metal ions influence the formation of intramolecular DNA triplex isomers. *J. Biol. Chem.* 267:20887-91
 50. Kang S, Wohlrab F, Wells RD. 1992. Metal ions cause the isomerization of certain intramolecular triplexes. *J. Biol. Chem.* 267:1259-64
 51. Karlovsky P, Pecinka P, Vojtiskova M, Makaturova E, Palecek E. 1990. Protonated triplex DNA in *E. coli* cells as detected by chemical probing. *FEBS Lett.* 274:39-42
 52. Kato M, Shimizu N. 1992. Effect of potential triplex DNA region on the in vivo expression of bacterial β -lactamase gene in superhelical recombinant plasmids. *J. Biochem.* 112:492-94
 53. Kennedy GC, Rattner JB. 1992. Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates an insulin promoter in heterologous cells. *Proc. Natl. Acad. Sci. USA* 89:11498-11502
 54. Kinniburgh AJ. 1989. A cis-acting transcription element of the c-myc gene can assume an H-DNA conformation. *Nucleic Acids Res.* 17:7771-78
 55. Kiyama R, Camerini-Otero D. 1991. A triplex DNA-binding protein from human cells: purification and characterization. *Proc. Natl. Acad. Sci. USA* 88:10450-54
 56. Klysik J, Rippe K, Jovin TM. 1991. Parallel-stranded DNA under topological stress: rearrangement of (dA)₁₅-(dT)₁₅ to a d(A-A-T)_n triplex. *Nucleic Acids Res.* 19:7145-54
 57. Kohwi Y, Kohwi-Shigematsu T. 1988. Magnesium ion-dependent, novel triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled DNA. *Proc. Natl. Acad. Sci. USA* 85:3781-88
 58. Kohwi Y, Kohwi-Shigematsu T. 1991. Altered gene expression correlates with DNA structure. *Genes Dev.* 5:2547-54
 59. Kohwi Y, Kohwi-Shigematsu T. 1993. Structural polymorphism of homopurine-homopyrimidine sequences at neutral pH. *J. Mol. Biol.* 231:1090-1101
 60. Kohwi Y, Malkhosyan SR, Kohwi-Shigematsu T. 1992. Intramolecular dG-dG-dC triplex detected in *Escherichia coli* cells. *J. Mol. Biol.* 223:817-22
 61. Kohwi Y, Panchenko Y. 1993. Transcription-dependent recombination induced by triple helix formation. *Genes Dev.* 7:1766-78
 62. Kohwi Y. 1989. Cationic metal-specific structures adopted by the poly(dG) region and the direct repeats in the chicken adult b^A globin gene promoter. *Nucleic Acids Res.* 17:4493-4502
 63. Kohwi-Shigematsu T, Kohwi Y. 1991. Detection of triple-helix related structures adopted by poly(dG)-poly(dC) sequences in supercoiled plasmid DNA. *Nucleic Acids Res.* 19:4267-71
 64. Kolluri R, Torrey TA, Kinniburgh AJ. 1992. A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. *Nucleic Acids Res.* 20:111-16
 65. Kornberg A, Baker TA. 1992. DNA

- Replication*. New York: W. H. Freeman. 2nd ed.
66. Kunkel GR, Martinson HG. 1981. Nucleosomes will not form on double-stranded RNA or over poly(dA)-poly(dT) tracts in recombinant DNA. *Nucleic Acids Res.* 9:6859-88
 67. Lapidot A, Baran N, Manor H. 1989. (dT-dC)_n and (dG-dA)_n tracts arrest single stranded DNA replication *in vitro*. *Nucleic Acids Res.* 17:883-900
 68. Larsen A, Weintraub H. 1982. An altered conformation detected by S1 nuclease occurs at specific region in active chick globin chromatin. *Cell* 29: 609-22
 69. Lee JS, Johnson DA, Morgan AR. 1979. Complexes formed by (pyrimidine)_n-(purine)_n DNAs on lowering the pH are three-stranded. *Nucleic Acids Res.* 6:3073-91
 70. Lee JS, Latimer LJ, Haug BL, Pulleyblank DE, Skinner DM, Burkholder GD. 1989. Triplex DNA in plasmids and chromosomes. *Gene* 82:191-99
 71. Lilley DMJ. 1980. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc. Natl. Acad. Sci. USA* 77:6468-72
 72. Lilley DMJ. 1992. Probes of DNA structure. *Methods Enzymol.* 212:133-39
 73. Liu LF, Wang JC. 1987. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* 84: 7024-27
 74. Lyamichev VI, Mirkin SM, Danylevskaya ON, Voloshin ON, Balatskaya SV, et al. 1989. An unusual DNA structure detected in a telomeric sequence under superhelical stress and at low pH. *Nature* 339:634-37
 75. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. 1985. Structure of homopurine-homopyrimidine tract in superhelical DNA. *J. Biomol. Struct. Dynam.* 3:327-38
 76. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. 1986. Structures of homopurine-homopyrimidine tract in superhelical DNA. *J. Biomol. Struct. Dynam.* 3:667-69
 77. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. 1987. Structure of (dG)_n-(dC)_n under superhelical stress and acid pH. *J. Biomol. Struct. Dynam.* 5:275-82
 78. Lyamichev VI, Mirkin SM, Kumarev VP, Baranova LV, Volodogskii AV, Frank-Kamenetskii MD. 1989. Energetics of the B-H transition in supercoiled DNA carrying d(CT)_x-d(AG)_x and d(C)_n-d(G)_n inserts. *Nucleic Acids Res.* 17:9417-23
 79. Macaya RF, Gilber DE, Malek S, Sinsheimer JS, Feigon J. 1991. Structure and stability of X-G-C mismatches in the third strand of intramolecular triplexes. *Science* 254:270-74
 80. Mace HAF, Pelham HRB, Travers A. 1983. Association of an S1 nuclease-sensitive structure with short direct repeats 5' of *Drosophila* heat shock gene. *Nature* 304:555-57
 81. Malkov VA, Soyfer VN, Frank-Kamenetskii MD. 1992. Effect of intermolecular triplex formation on the yield of cyclobutane photodimers in DNA. *Nucleic Acids Res.* 20:4889-95
 82. Malkov VA, Voloshin ON, Veselkov AG, Rostapshov VM, Jansen I, et al. 1993. Protonated pyrimidine-purine-purine triplex. *Nucleic Acids Res.* 21: 105-11
 83. Manor H, Sridhara Rao B, Martin RG. 1988. Abundance and degree of dispersion of genomic d(GA)_n-d(TC)_n sequences. *J. Mol. Evol.* 27:96-101
 84. Margot JB, Hardison RC. 1985. DNase I and nuclease S1 sensitivity of the rabbit β1 globin gene in nuclei and in supercoiled plasmids. *J. Mol. Biol.* 184: 195-210
 85. Martinez-Balbas A, Azorin F. 1993. The effect of zinc on the secondary structure of d(GA·TC) DNA sequences of different length: a model for the formation *H-DNA. *Nucleic Acids Res.* 21:2557-62
 86. Mavrothalassitis GL, Watson DK, Papas TS. 1990. Molecular and functional characterization of the promoter of ETS2, the human *c-ets-2* gene. *Proc. Natl. Acad. Sci. USA* 87:1047-51
 87. McKeon C, Schmidt A, de Crombrughe BA. 1984. A sequence conserved in both the chicken and mouse α2(I) collagen promoter contains sites sensitive to S1 nuclease. *J. Biol. Chem.* 259:6636-40
 88. Mendelsohn E, Baran N, Neer A, Manor H. 1982. Integration site of polyoma virus DNA in the inducible LPT line of polyoma-transformed rat cells. *J. Virol.* 41:192-209
 89. Mergny JL, Sun JS, Rougee M, Montenay-Garestier T, Barcelo F, et al. 1991. Sequence specificity in triple-helix formation: experimental and theoretical studies of the effect of mismatches on triplex stability. *Biochemistry* 30:9791-98
 90. Milbrandt JD, Heintz NH, White WC,

- Rothman SM, Hamlin JL. 1981. Methotrexate-resistant Chinese hamster ovary cells have amplified a 135-kilobase-pair region that includes the dihydrofolate reductase gene. *Proc. Natl. Acad. Sci. USA* 78:6043-47
91. Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetskii MD. 1987. DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* 330:495-97
92. Morgan AR, Wells RD. 1968. Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *J. Mol. Biol.* 37:63-80
93. Moser HE, Dervan PB. 1987. Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* 238:645-50
94. Muraiso T, Nomoto S, Yamazaki H, Mishima Y, Kominari R. 1992. A single-stranded DNA binding protein from mouse tumor cells specifically recognizes the C-rich strand of the (AGG:CCT)_n repeats that can alter DNA conformation. *Nucleic Acids Res.* 20:6631-35
95. Nelson HCM, Finch JT, Bonaventura FL, Klug A. 1987. The structure of an oligo(dA)-oligo(dT) tract and its biological implications. *Nature* 330:221-26
96. Palecek E. 1991. Local supercoil-stabilized DNA structures. *Crit. Rev. Biochem. Mol. Biol.* 26:151-226
97. Panayotatos N, Wells RD. 1981. Cruciform structures in supercoiled DNA. *Nature* 289:466-70
98. Panyutin IG, Kovalsky OI, Budowsky EL. 1989. Magnesium-dependent supercoiling-induced transition in (dG)_n-(dC)_n stretches and formation of a new G-structure by (dG)_n strand. *Nucleic Acids Res.* 17:8257-71
99. Panyutin IG, Wells RD. 1992. Nodule DNA in the (GA)₃₇-(CT)₃₇ insert in superhelical plasmids. *J. Biol. Chem.* 267:5495-5501
100. Parniewski P, Kwinkowski M, Wilk A, Klysik J. 1990. Dam methyltransferase sites located within the loop region of the oligopurine-oligopyrimidine sequences capable of forming H-DNA are undermethylated *in vivo*. *Nucleic Acids Res.* 18:605-10
101. Peleg M, Manor H. 1993. DNA triplex formation inhibits DNA unwinding by the SV40 large T-antigen helicase. *J. Biomol. Struct. Dynam.* 10:a149
102. Pestov DG, Dayn A, Siyanova EYu, George DL, Mirkin SM. 1991. H-DNA and Z-DNA in the mouse c-Ki-ras promoter. *Nucleic Acids Res.* 19:6527-32
103. Postel EH, Berberich SJ, Flint SJ, Ferrone CA. 1993. Human c-myc transcription factor PuF identified as nm23-H2 nucleoside diphosphate kinase, a candidate suppressor of tumor metastasis. *Science* 261:478-83
104. Postel EH, Mango SE, Flint SJ. 1989. A nuclease-hypersensitive element of the human c-myc promoter interacts with a transcription initiation factor. *Mol. Cell. Biol.* 9:5123-33
105. Pulleyblank DE, Haniford DB, Morgan AR. 1985. A structural basis for S1 nuclease sensitivity of double-stranded DNA. *Cell* 42:271-80
106. Rahmouni AR, Wells RD. 1992. Direct evidence for the effect of transcription on local DNA supercoiling *in vivo*. *J. Mol. Biol.* 223:131-44
107. Rao S, Manor H, Martin RG. 1988. Pausing in simian virus 40 DNA replication by a sequence containing (dG-dA)₂₇-(dT-dC)₂₇. *Nucleic Acids Res.* 16:8077-94
108. Reaban ME, Griffin JA. 1990. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* 348:342-44
109. Rippe K, Kuryavy VV, Westhof E, Jovin TM. 1992. Polymorphism and possible biological functions of parallel-stranded DNA. In *Structural Tools for the Analysis of Protein-Nucleic Acids Complexes. Advances in Life Sciences*, ed. DMJ Lilley, H Heumann, D Suck, pp. 81-107. Basel: Birkhauser
110. Samadashwily GM, Dayn A, Mirkin SM. 1993. Suicidal nucleotide sequences for DNA polymerase. *EMBO J.* 12:4975-83
111. Sarkar PS, Brahmachari SK. 1992. Intramolecular triplex potential sequence within a gene down regulates its expression *in vivo*. *Nucleic Acids Res.* 20:5713-18
112. Shen C-K. 1983. Superhelicity induces hypersensitivity of a human polypyrimidine-polypurine DNA sequence in the human $\alpha 2$ -1 globin intergenic region to S1 nuclease digestion. *Nucleic Acids Res.* 11:7899-10
113. Shimizu M, Hanvey JC, Wells RD. 1990. Multiple non-B-DNA conformations of polypurine-polypyrimidine sequences in plasmids. *Biochemistry* 29:4704-13
114. Shimizu M, Kubo K, Matsumoto U,

- Shindo H. 1993. The loop sequence plays crucial roles for isomerization of intramolecular DNA triplexes in supercoiled plasmids. *J. Mol. Biol.* 235:185-97
115. Sinden RR, Kochel TJ. 1987. Reduced 4,5,8-trimethylpsoralene cross-linking of left-handed Z-DNA stabilized by DNA supercoiling. *Biochemistry* 26:1343-50
116. Singleton CK, Klysik J, Sturdivant SM, Wells RD. 1982. Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* 299:312-16
117. Stavnezer J, Reaban ME, Griffin JA. 1991. Triple-helix stabilization? *Nature* 351:447-48
118. Tabor CW, Tabor H. 1984. Polyamines. *Annu. Rev. Biochem.* 53:749-91
119. Usery DW, Sinden RR. 1993. Environmental influences on the *in vivo* level of intramolecular triplex DNA in *Escherichia coli*. *Biochemistry* 32:6206-13
120. Vojtiskova M, Mirkin S, Lyamichev V, Voloshin O, Frank-Kamenetskii M, Palecek E. 1988. Chemical probing of the homopurine-homopyrimidine tract in supercoiled DNA at single-nucleotide resolution. *FEBS Lett.* 234:295-99
121. Voloshin ON, Mirkin SM, Lyamichev VI, Belotserkovskii BP, Frank-Kamenetskii MD. 1988. Chemical probing of homopurine-homopyrimidine mirror repeats in supercoiled DNA. *Nature* 333:475-76
122. Voloshin ON, Veselkov AG, Belotserkovskii BP, Danilevskaya ON, Pavlova MN, et al. 1992. An eclectic DNA structure adopted by human telomeric sequence under superhelical stress and low pH. *J. Biomol. Struct. Dynam.* 9:643-52
123. Wang AH, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, et al. 1979. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* 282:680-86
124. Wang E, Malek S, Feigon J. 1992. Structure of a G-T-A triplet in an intramolecular DNA triplex. *Biochemistry* 31:4838-46
125. Wang JC, Peck LJ, Becherer K. 1983. DNA supercoiling and its effects on DNA structure and function. *Cold Spring Harbor Symp. Quant. Biol.* 47:85-91
126. Wang JC. 1985. DNA topoisomerases. *Annu. Rev. Biochem.* 54:665-97
127. Weintraub H. 1983. A dominant role for DNA secondary structure in forming hypersensitive structures in chromatin. *Cell* 32:1191-03
128. Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F. 1988. The chemistry and biology of unusual DNA structures adopted by oligopurine/oligopyrimidine sequences. *FASEB J.* 2:2939-49
- 128a. Williamson JR. 1994. G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* 23:703-30
129. Yagil G. 1991. Paranemic structures of DNA and their role in DNA unwinding. *Crit. Rev. Biochem. Mol. Biol.* 26:475-559
130. Zakian VA. 1989. Structure and function of telomeres. *Annu. Rev. Genet.* 23:579-604



CONTENTS

STRUCTURAL PRINCIPLES

- Hydration and Steric Pressures Between Phospholipid Bilayers, *Thomas J. McIntosh and Sidney A. Simon* 27
- Membrane Proteins: From Sequence to Structure, *Gunnar von Heijne* 167
- Global Statistics of Protein Sequences: Implications for the Origin, Evolution, and Prediction of Structure, *Stephen H. White* 407
- Conformational and Thermodynamic Properties of Supercoiled DNA, *Alexander V. Vologodskii and Nicholas R. Cozzarelli* 609
- The β -Ribbons DNA Recognition Motif, *Simon E. V. Phillips* 671
- G-Quartet Structures in Telomeric DNA, *James R. Williamson* 703
- Mechanical Properties of the Red Cell Membrane in Relation to Molecular Structure and Genetic Defects, *Narla Mohandas and Evan Evans* 787

STRUCTURE AND FUNCTION

- DNA Branched Junctions, *Nadrian C. Seeman and Neville R. Kallenbach* 53
- Annexin Structure and Membrane Interactions: A Molecular Perspective, *Manal A. Swairjo and Barbara A. Seaton* 193
- Molecular Diversity and Functions of Glutamate Receptors, *Shigetada Nakanishi and Masayuki Masu* 319
- Protein Structure-Based Drug Design, *Peter J. Whittle and Tom L. Blundell* 349
- Perspectives on the Physiology and Structure of Inward-Rectifying K^+ Channels in Higher Plants: Biophysical Implications for K^+ Uptake, *Julian I. Schroeder, John M. Ward, and Walter Gassmann* 441
- Evolution of the EF-Hand Family of Proteins, *Susumu Nakayama and Robert H. Kretsinger* 473

The Bacterial Flagellar Motor, <i>Stephan C. Schuster and Shahid Khan</i>	509
H-DNA and Related Structures, <i>Sergei M. Mirkin and Maxim D. Frank-Kamenetskii</i>	541
Voltage-Dependent Gating of Ionic Channels, <i>Francisco Bezanilla and Enrico Stefani</i>	819

DYNAMICS

Field Gradient ESR and Molecular Diffusion in Model Membranes, <i>J. H. Freed</i>	1
Alamethicin: A Peptide Model for Voltage-Gating and Protein-Membrane Interactions, <i>D. S. Cafiso</i>	141
Cyclic Nucleotide-Gated Ion Channels and Sensory Transduction in Olfactory Receptor Neurons, <i>F. Zufall, S. Firestein, and G. M. Shepherd</i>	577
Polypeptide Interactions with Molecular Chaperones and Their Relationship to In Vivo Protein Folding, <i>Samuel J. Landry and Lila M. Gierasch</i>	645
Molecular Dynamics Simulations of the Gramicidin Channel, <i>Benoît Roux and Martin Karplus</i>	731
Molecular Mechanics in Biology: From Structure to Function, Taking Account of Solvation, <i>W. F. van Gunsteren, F. J. Luque, D. Timms, and A. E. Torda</i>	847

EMERGING TECHNIQUES

Biomolecular Imaging with the Atomic Force Microscope, <i>Helen G. Hansma and Jan H. Hoh</i>	115
Nonresonance Raman Difference Spectroscopy: A General Probe of Protein Structure, Ligand Binding, Enzymatic Catalysis, and the Structures of Other Biomacromolecules, <i>Robert Callender and Hua Deng</i>	215
Biological Applications of Optical Forces, <i>Karel Svoboda and Steven M. Block</i>	247
High Pressure NMR Spectroscopy of Proteins and Membranes, <i>J. Jonas and A. Jonas</i>	287
Mass Spectrometry of Macromolecules: Has Its Time Now Come?, <i>M. W. Senko and F. W. McLafferty</i>	763

BIOTECHNOLOGY

- The Light-Addressable Potentiometric Sensor: Principles and Biological Applications, *John C. Owicki, Luc J. Bousse, Dean G. Hafeman, Gregory L. Kirk, John D. Olson, H. Garrett Wada, and J. Wallace Parce* 87
- Molecular Nanomachines: Physical Principles and Implementation Strategies, *K. Eric Drexler* 377

INDEXES

- Subject Index 865
- Cumulative Index of Contributing Authors, Volumes 19–23 881
- Cumulative Index of Chapter Titles, Volumes 19–23 883