1993

Recombination Triple Helix, R-form DNA. A Stereochemical Model for Recognition and Strand Exchange

V. B. Zhurkin, *National Institutes of Health*
G. Raghunathan, *National Institutes of Health*
N. B. Ulyanov, *University of California, San Francisco*
R. D. Camerini-Otero, *National Institutes of Health*
R. L. Jernigan, *National Institutes of Health*

Structural Biology: The State of the Art, ISBN 0-940030-42-X
Proceedings of the Eighth Conversation, State University of New York, Albany, NY 1993
Eds. Ramaswamy H. Sarma and Mukti H. Sarma, ©Adenine Press 1994

Recombination Triple Helix, R-form DNA. A Stereochemical Model for Recognition and Strand Exchange

V.B. Zhurkin[1], G. Raghunathan[1], N.B. Ulyanov[2], R.D. Camerini-Otero[3] and R.L. Jernigan[1]

[1] Laboratory of Mathematical Biology
NCI, NIH
Bethesda, MD 20892

[2] Department of Pharmaceutical Chemistry
University of California
San Francisco, CA 94143

[3] Genetics & Biochemistry Branch
NIDDK, NIH
Bethesda, MD 20892

Abstract

Based on the energy calculations, a model for the recombination triplex is proposed, which is stereochemically possible for any arbitrary sequence and consistent with the chemical modification data for the RecA mediated joint molecules. The detailed mechanisms of the recognition and strand exchange are suggested, the key elements of which are the isomorphism of the triplets and the coordinated rotations of bases in the three strands.

Each base from the third strand recognizes both purine and pyrimidine from the duplex and is located in the center of the major groove. Partial charges on the third strand bases are strictly complementary to the charges of the Watson-Crick pairs. Hence, the initial selection of the homologous sequence may occur through electrostatic interactions of the third strand with the closed WC pairs rather than with the open bases. This recognition code may be applicable to both recombination and transcription.

In the collapsed triplex with a rise of about 3.4 Å, the bases from the third strand can interact with two neighboring WC base pairs simultaneously, that could lead to recognition errors. In the triplex extended up to 5.1 Å in the presence of RecA protein, these mismatches are impossible. Therefore, we suggest that an extended DNA structure increases the fidelity of recognition, and RecA plays the role of a "chaperone".

Introduction

Homologous recombination involves the exchange of genetic information between two sister chromosomes. As such, it is universal and to a first approximation sequence independent (or nearly independent). Our current understanding of the stereochemistry of the strand exchange reaction is based on the *in vitro* studies carried out with the *E. coli* recombination protein, RecA (1). The elementary act of exchange involves the three DNA strands: one single strand (ssDNA) and two strands paired in a duplex (dsDNA); the ssDNA is identical to one of the duplex strands. This process consists of at least three consequtive steps: (i) binding of RecA to ssDNA; (ii) binding of the duplex and formation of a homologous synaptic complex (RecA+ssDNA+dsDNA); (iii) exchange of the strands.

1. Definitions

The single third strand is denoted as *R-strand* (where R stands for Recombination), and the identical duplex strand is called arbitrarily W-strand (the opposite strand is called as usual, C-strand). Then the exchange reaction can be written as

$$(W:C)+R \rightarrow W+(R:C).$$

Here (W:C) is for the initial Watson-Crick duplex, and (R:C) is for the resulting heteroduplex. When recombination protein pairs dsDNA with ssDNA, then a transient high energy DNA structure is formed, which is extended and underwound by 50% with respect to B-DNA (2-4); we call such a complex the *extended R-form*. After deproteinization of the joint molecule *in vitro*, stacking is restored, and the collapsed *R-form* is obtained (5,6).

The collapsed R-form remains relatively stable up to 60-65 °C (5). Although the collapsed triplex does not have any known biological function, one of its virtues is that it can be studied by chemical probing in the absence of the recombination proteins that otherwise would protect the extended R-DNA (7-9). Thus, the studies of the collapsed form are crucial for elucidating the geometry of hydrogen bonding in the parallel triplex, which, in turn, can help in understanding the organization of the extended form since there is a strong reason to believe that the extended and collapsed R-forms are structurally related.

Using the term *"collapsed R-form"*, we mean the experimentally studied triple stranded structure obtained after deproteinization of the synaptic complex (5,6). Following the lead of Hsieh et al.(5) we assume that the three strands are linked together by hydrogen bonds. Speaking about the *"extended R-form"*, we refer to the putative intermediate formed in the center of the RecA filament (10,11).

Currently, there are two opposite points of view presented in the literature: (i) recognition occurs by virtue of formation of the transient triple helix stabilized (at least temporarily) by hydrogen bonds (5,6,12); (ii) recognition occurs through opening of WC pairs (13,14). The question is extremely tricky, since the DNA duplex is breathing (opening) even under "standard" physiological conditions; in complex

with recombination proteins, where the DNA is severely extended and underwound, it is expected to have an increased probability of opening. It may well be that the extended R-form is a loose association with the base triplets dynamically opening and closing all the time. Therefore, the term *"extended R-form"* is applied to the hypothetical triple stranded structure formed in the presence of RecA; at least part of the time it comprises the triplex, stabilized by the hydrogen bonds.

2. Parallel and Antiparallel Triplexes

Historically, the notion of DNA triple helix with the third strand buried in the major groove of the duplex, originated shortly after the double helix was discovered, see review (15). This triplex was assumed to participate both in transcription and recombination.; it was postulated to be *parallel* in a sense that its two identical strands had to be oriented in the same way. In the case of recombination, the parallel orientation of the two identical strands is a direct consequence of the biological role played by the triplex, namely pairing and exchanging genetic information at *any arbitrary* sequence on the chromosomal homologues. But the first triplex [poly(U):poly(A)]:poly(U) observed experimentally by Felsenfeld, Davies and Rich (16), proved to be *antiparallel*: the two poly(U) strands had opposite orientations (17). The other homopurine-homopyrimidine triplexes, including H-form (18,19), are also antiparallel. (These triplexes, obtained in the absence of proteins, are usually called "non-enzymatic".) Thus, the anti-parallel orientation of the two identical strands in these triplexes makes them crucially different from the parallel recombination R-form.

The other difference between the recombination R-form and the non-enzymatic triplexes is the base pairing scheme. In the non-enzymatic triplexes the third strand bases usually interact with the strongly electronegative atoms N7 of purines from the WC duplex. This scheme, however, is less satisfactory for the mixed sequences with purines and pyrimidines in each strand, since the third strand would have to alternate its pairing with the Watson and Crick strands, and such a zig-zag path of the sugar-phosphate backbone is energetically unfavorable (20). Besides, irregular zig-zagging, depending on the purine-pyrimidine sequence, is improbable if the triplex were to interact with the recombination protein assembly in a regular way. Hence, the pairing scheme in the recombination triplex has to be different from pairing in the non-enzymatic triplexes (16-19).

So far, the detailed structure of the RecA-mediated triple helix remains unknown. Several tripling schemes for the bases have been proposed (5,7,11,15,21-22), but no conformational calculations were used to discriminate between the possible alternatives. In an attempt to resolve this issue, we have performed the energy calculations (20,23-25). We tried to answer the following questions: (i) What is the structure of the parallel triple stranded DNA that can accommodate any arbitrary sequence? (ii) What is the stereochemical mechanism, ensuring recognition of the ssDNA and the duplex? (iii) Why is DNA so severely distorted in the complex with RecA? In other words, what are the functional advantages of the extended and unwound DNA in recognition and/or strand exchange? The logical scheme of our study is given below.

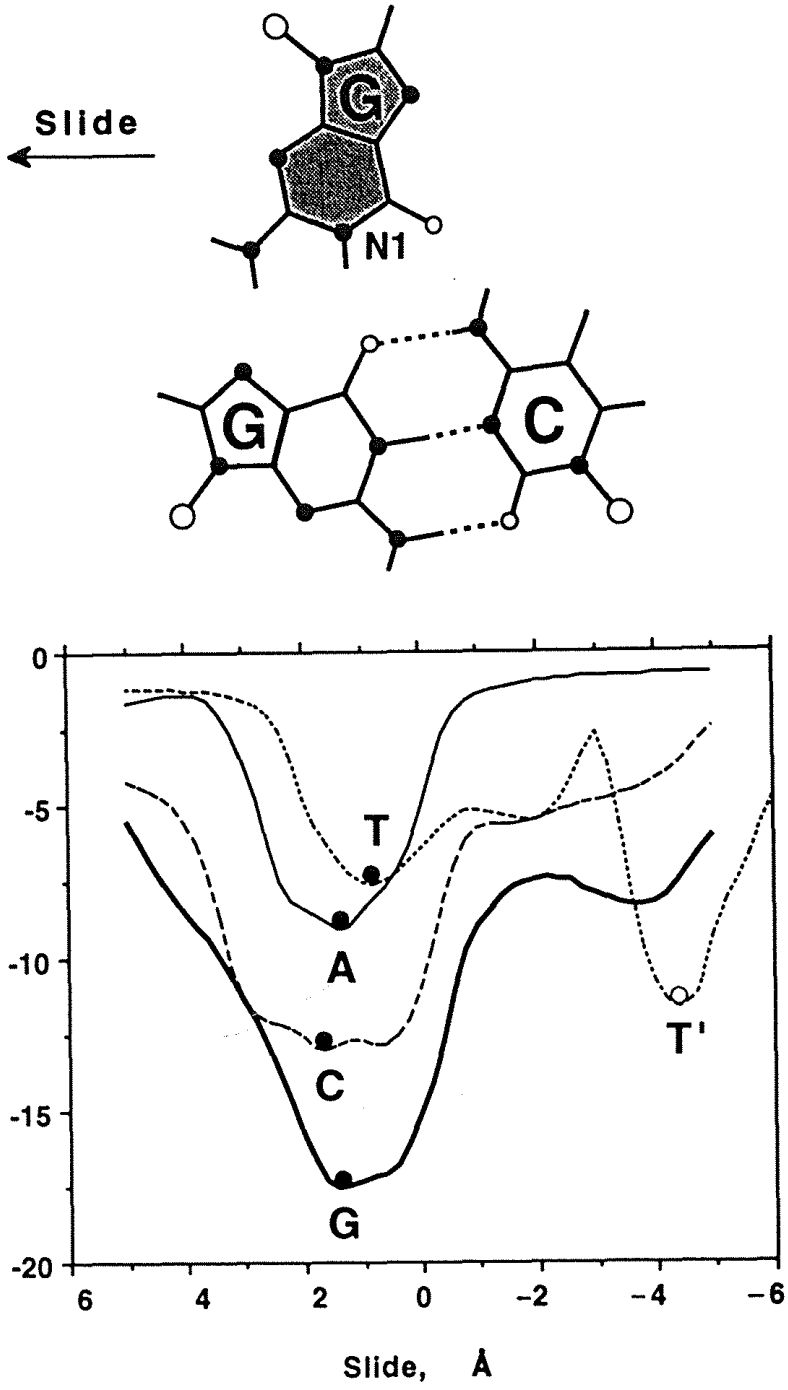

Figure 1: *Above.* Triad of bases in the parallel triple helix. Here and in the other figures W-strand is located on the left, C-strand on the right, and the shaded R-strand is placed in the major groove of the WC duplex. Position of the third base (R) in the co-planar triad is defined by three parameters: projection of the N1(Pur)/N3(Pyr) atom on the long axis of WC base pair (horizontal slide), the distance of N1/N3 atom from this axis (vertical shift), and orientation angle. In the unfrozen (non-planar) triad, the position of the third base is defined by one more translation and two more rotations (see Figure 5).

Below. Energy of interaction of the third base with the WC pair as a function of *slide.*
T stands for triplet (TA):T, A is for (AT):A, C for (CG):C and G for (GC):G.

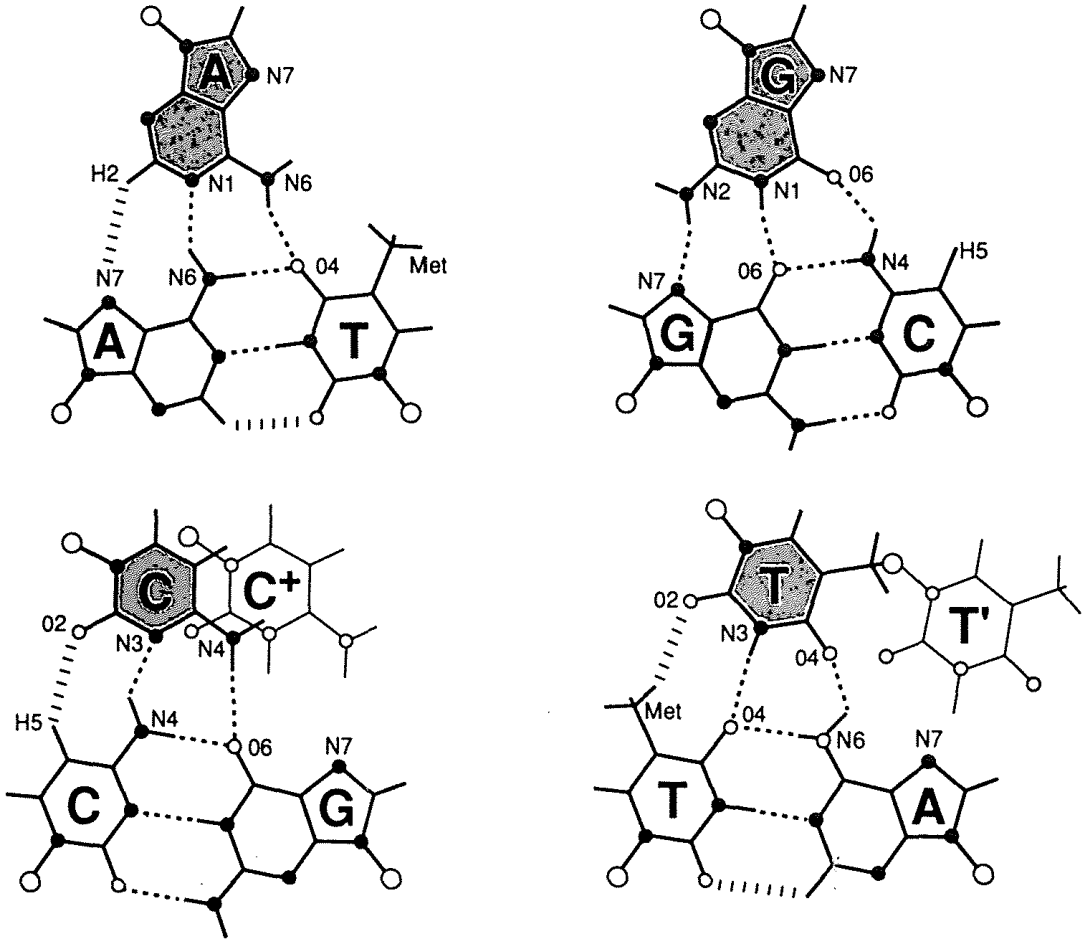

Figure 2: Hydrogen bonding schemes for the energetically optimal triplets. The isomorphic positions of the third strand bases are shaded; they correspond to the local minima of the energy profiles in Figure 1 shown by black dots. Locations of C and T proposed by Hsieh et al. (5) are presented in white (denoted C^+ and T' respectively). Hydrogen bonds are shown by broken lines. The non-hydrogen-bond electrostatic interactions are shown as perpendicular bars. These electrostatic interactions involving the leftmost N7(Pur)/C5(Pyr) groups, stabilize the triplets shown in the same way, as the interaction between H2(Ade) and O2(Thy) stabilizes the AT pair.

3. Logical Scheme

(1) With the energy calculations, iso-geometric base triplets are found, consistent with experiment. These triplets suggest a basis for the electrostatic specificity of ssDNA-dsDNA recognition.
(2) Then the parallel triplexes are calculated with the sugar-phosphate backbone stereochemically feasible for any sequence of bases.
(3) Finally, we consider a hypothetical mechanism for strand exchange. Here, as well as in recognition, the isomorphism of the triplets is one of the key premises for the construction of models.

I. Base Triplets

First, we analyzed how the energy of a triplet depends on the position of the third strand base, assuming that it is located in the major groove of the duplex, and is oriented parallel to the identical Watson strand (Figure 1). The horizontal slide of the third base served as an independent parameter, whereas the vertical shift and orientation of the third base were energy-minimized. The WC pairs retained their standard co-planar geometry. The potential energy of base triplets was calculated in a standard way as a sum of van der Waals, electrostatic and hydrogen bond interactions (26,27).

All the energy profiles obtained in this way, have a minimum at essentially the same slide values (slide=0.8-1.6 Å; see the black dots on the curves in Figure 1). Positions of the third bases corresponding to these minima are shaded in Figure 2.
Note that the third base in these triplets is located in the center of the major groove and interacts with both duplex strands. In the (AT):A, (CG):C and (TA):T triplets these interactions occur by virtue of two hydrogen bonds and in addition, by favorable electrostatic interactions between the W-and R-strand bases (Figure 2). The triplet (GC):G, which has the lowest energy, is stabilized by three hydrogen bonds. The strength of the interaction increases with the following order: T, A, C and G.
To demonstrate isomorphism of the four base triplets, they are superimposed in Figure 3.

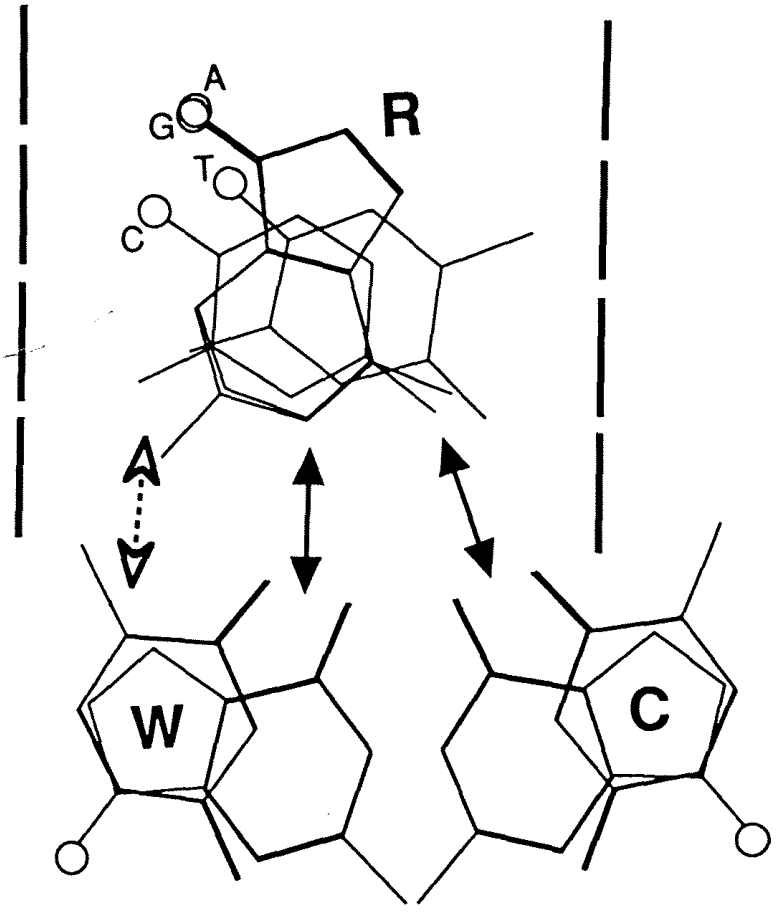

Figure 3: Superposition of the four isomorphic triplets from Figure 2. The C1′ atoms are shown by circles. Recognition between the WC pairs and the third strand bases can occur by virtue of hydrogen bonds (solid arrows) or by favorable electrostatic interactions (open arrow). The broken lines depict the vertical "corridor", in which the third strand moves with respect to the WC duplex in the field of the recombination proteins (see Figures 9 and 10).

The positions of A and G in the third strand are nearly identical. The largest distance between the C1′ atoms in the R-strand, 1.6 Å, is between the C1′ of adenine (A) and C1′ of cytosine (C). Such a close position of the glycosidic bonds in all four triplets implies that for any arbitrary sequence the sugar-phosphate backbone of R-form DNA would be nearly as regular as in the WC double helix. Note that the difference in size between our triplets is comparable to the difference between the W:C and the purine:purine pairs when both purines are in the *anti*-orientation. It is known that the I:A and G:A pairs incorporated into the double helix do not cause large distortions in the DNA backbone (28,29). So, we may expect that our triplets can likewise be as easily accommodated into the triple helix, just as the purine:purine pairs into the duplex.

On the other hand, the mentioned isomorphic positions of the third bases have suboptimal energies for most of the bases (A, C and T). For example, the right-hand position T′ shown in Figure 2, the "reverse Hoogsteen" position proposed by Hsieh et al.(5), is more preferable by 3-4 kcal/mol than the triplet proposed here shown in black (see minima T and T′ in Figure 1). The energetic disadvantage of the new triplets originates from the imperfect matching of the interacting donor and acceptor groups in the major groove of the WC pairs with the groups from the third strand bases (25). In particular, the N-H...O and N-H...N angles deviate significantly from 180° (Figure 2). Another reason for the relatively high energies of the new triplets is that the strongly electronegative atom N7 of purines is not involved in the hydrogen-bonding scheme for any of the triplets but (GC):G. However, the energetic disadvantage of these triplets *per se* is compensated by a favorable regular backbone for mixed sequences, as discussed below.

Comparison with experiment

Two sets of data on the chemical modification of the deproteinized joint molecules (collapsed R-form) were obtained recently by Radding and coworkers (7) and by Camerini-Otero and colleagues (8,9). The results of our calculations can be compared with these experiments. First, the triplets proposed here are certainly consistent with the destabilizing effect of methylation of adenines and cytosines at positions N6 and N4 respectively (7). Second, they are in accord with the data on dimethylsulfate modification and deaza-substitution at the N7 position in the duplex purines (8,9).

In the (GC):G triad, 7-deaza substitution in guanines in the Watson strand decreases the stability of the joint triplex; and the N7 groups of these guanines are protected from modification by dimethylsulphate (8,9). These results agree with our scheme for (GC):G (Figure 2), where N7 is hydrogen bonded to the amino group of guanine from the third strand. By contrast, in the (CG):C triad the central position of cytosine, C, shown in Figure 2, leaves N7 of guanine free, which is consistent with the accessibility of the N7 of the same guanine to chemical modification by dimethylsulphate (8,9).

Stability of the AT-containing triplets is decreased upon 7-deaza substitution of adenines (A→A*) in both duplex strands (8,9). In the (AT):A triplet, the N7 position of adenine in the duplex is not directly involved in hydrogen bonding (Figure 2), a

result in seeming contradiction to the decreased stability of the substituted (A*T):A triplet (8,9). However, this N7 nitrogen and the H2 proton from the R-strand are in close proximity, so that their interaction is electrostatically favorable. When N7 is substituted by C7-H, the two protons H7 and H2 "clash", and the (A*T):A triplet becomes less stable than (AT):A (9,25).

The data for the (TA):T triplet (8,9) can be explained by a specific hydration pattern in the (TA):T triad: a uniquely stable "structural" water molecule can bridge the N7 of adenine and the O4 of thymine (9), see Figure 2. If N7 of adenine (A) is substituted by C7-H (A*), the "bridging" water could not stabilize this triplex, and as a consequence, the modified triplex (TA*):T would be less stable than the standard (TA):T (9, 25).

Thus, we conclude that the isomorphic base triads proposed here are consistent with the available experimental data for the mixed sequences (6-9). As for oligo(purine):oligo (pyrimidine) sequences, the possibility of the rightward shift of pyrimidines to the positions T′ and C^+ (Figure 2) is discussed elsewhere (25).

II. Collapsed Triple Helices

At the next stage we proved the stereochemical feasibility of the parallel triple helices with mixed sequences. For this purpose the triplexes with the alternating purine-pyrimidine sequence ATGCATGC were calculated (Figure 4). The parameters of bases and sugar ring pseudorotation were used as independent variables (30,31), and the sugar-phosphate chain was closed using the standard values of bond lengths and angles. During energy minimization the periodic boundary conditions were imposed (32), with the tetramer ATGC as the repeating unit. The program for calculation of the DNA duplexes, *"DNAminiCarlo"* (31,32), was modified correspondingly to include the third strand (25).

Two sets of structures were obtained: with all the sugar rings in the C3′-endo conformation, and with sugars in the C2′-endo--C1′-exo conformation. The latter case is illustrated in Figure 4 (left). This structure is obtained assuming that the base triplets are planar and frozen as in Figure 2 (see the shaded positions of the third strand bases). The axial rise is about 3.4 Å. All the dihedral angles lie in the local minima typical for B-DNA. The strongest sequence dependent variation is observed for the ζ[P-O3′] and ε[O3′-C3′] angles; but even for these angles the difference between Pur-Pyr and Pyr-Pur steps does not exceed 30° (25). As a result, orientation of the phosphates is nearly sequence independent (Figure 4). Note that the structures of the anti-parallel triplexes, described recently (33,34), also have a majority of sugars in all three strands belonging to the C2′-endo domain, and their backbone on the whole is similar to the B-form duplex DNA.

When the C3′-endo triplexes are minimized with the planar base triads, the axial rise increases to 4.1 Å (25). This stretching of DNA relieves the steric clash between the (C2′)-H_2 group of a sugar and the (C8/C6)-H group of the 3′-neighboring base; this clash is known to be stronger for the C3′-endo sugar pucker than for C2′-endo

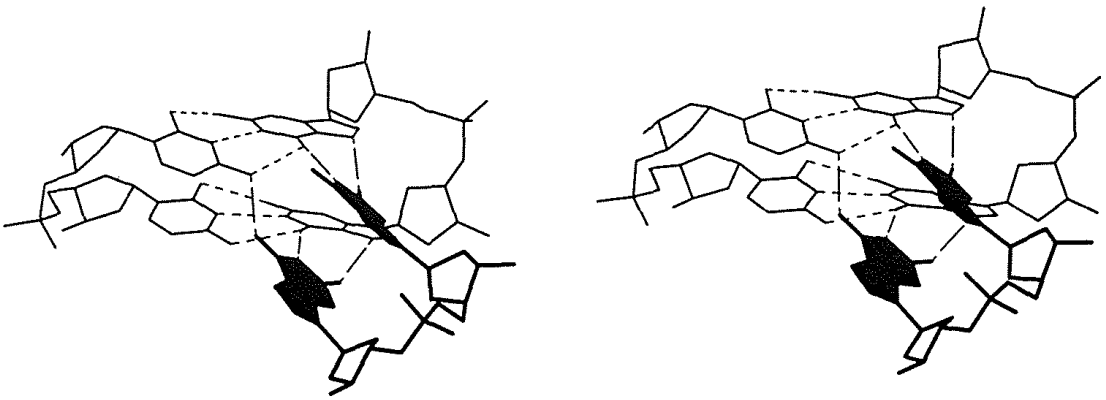

Figure 5: Energetically optimal structure of [poly(dG):poly(dC)]:poly(dG) with non-planar base triads and sugars in C3′-endo conformation. Bases from the third R-strand, shown in black, form hydrogen bonds with two layers of the duplex GC-pairs. The duplex GC-pairs are buckled by 5°, and have propeller twists of 10° (half-angle values). The third strand guanines are rotated by 32° around their long axes (going approximately toward the viewer), and by 12° around their short axes. Such hydrogen bonding would cause errors of recognition (see text).

(30). If the base triads were unfrozen, using such degrees of freedom as *buckle, propeller twist, stagger*, the separation between bases retains its standard value 3.3 Å. But the resulting conformations are very unusual for certain sequences. In particular, for [poly(dG):poly(dC)]:poly(dG) the energetically optimal structure has a non-planar hydrogen bonding geometry as shown in Figure 5. In this structure each guanine from the third strand interacts with two layers of WC pairs; two hydrogen bonds are formed with O6 and N7 of the duplex guanine, and one with N4 of cytosine. Due to the distortion of the hydrogen bonds, their energy is increased. However, since in the "non-planar" structure the intra-strand stacking interactions are more preferable, the resulting energy is favorable overall.

Although the details of the energetically optimal structure depend on the potential functions used (26,27), the possibility of formation of such non-planar structures with "distorted" inter-layer hydrogen bonds has been proven experimentally. The hydrogen bonding scheme shown in Figure 5 is remarkably similar to the low temperature X-ray structure of the CACA:TGTG tetramer (35), which is characterized by the inter-layer A G and C T hydrogen bonds in the major groove. These bonds are distorted approximately to the same extent as in the non-planar triad (GC):G shown in Figure 5. A similar arrangement of hydrogen bonds is possible for other sequences as well: for example in 5′ApT, where adenine in the third strand can interact with two adenines in W- and C-strands (not shown).

C2′-endo and C3′-endo structures

Our calculations show that the deproteinized collapsed triplex is a polymorphic structure, having at least two families of forms, with feasible C3′-endo and C2′-endo sugar conformations, similar to the canonical A and B-families for the duplex DNA (36). According to the energy calculations *in vacuo*, the C3′-endo non-planar structures are the most preferable ones. Among the planar structures, the C2′-endo ones

are better for (Pur:Pyr):Pur homopolymers, but the C3′-endo structures are better for (Pyr:Pur):Pyr homopolymers partly because the sugar-base interaction within a pyrimidine nucleoside is better for C3′-endo sugar than for C2′-endo (37,38). The energetic preference of one of these structures over the others is not significant, differing by only several kcal/mol.

Most likely, the equilibrium between the two families of forms depends both on sequence and environmental conditions, as is the case for the duplex DNA (39). But, the sequence dependence of this equilibrium is likely to be more complicated for triplexes than for duplexes: in triplexes in addition to stacking, the hydrogen bonds also determine the sequence specific base-base interactions (Figure 5). This matter can only be clarified after experimental studies.

However, rather than emphasizing the differences between the C2′-endo and the C3′-endo forms, we pursue another aim: to show that the collapsed triplexes are possible in principle for any arbitrary sequence within the framework of the present scheme (Figures 2 and 4). This scheme does not contradict the stereochemical restrictions imposed by the sugar-phosphate backbone and thus can be used as a structural basis for interpreting the experimental data on deproteinized triplexes (5-9).

III. Extended Triplex

According to electron microscopic measurements (2-4), DNA is extended and underwound by 50% in the complex with RecA: its axial rise increases up to 5.1 Å, and helical twisting decreases to 20°. These estimates agree with recent X-ray data on the RecA filament without DNA (40). In addition, the linear flow dichroism shows that in the RecA-DNA complex the bases are nearly perpendicular to the DNA helical axis: the average angle of inclination of bases does not exceed 20-30° (41). It has been suggested that the DNA extension occurs by intercalation (42).

Several putative models of DNA extension can be proposed now (Figure 6). The loops L2 and L1, disordered in the X-ray structure of RecA, are assumed to interact with ssDNA and dsDNA respectively (40). If RecA residues were indeed intercalating between the DNA bases, then the conserved phenylalanine/tyrosine 204 from the loop L2 would be the first candidate for intercalation into the ssDNA, or R-strand. But in the loop L1 there are no aromatic residues. So, binding of the loop L1 to the dsDNA (WC duplex) cannot be explained by a "conventional" intercalation mechanism, such as that of ethidium bromide (43).

The degree of extension of DNA in the complex with RecA also testifies against the well known intercalation scheme, where the DNA helical rise is increased up to 6.8 Å at the sites of intercalation, and remains 3.4 Å at other steps. In this case the rise per RecA monomer (each of which covers three DNA steps), would be $3.4 \bullet 2 + 6.8 = 13.6$ Å instead of the 15.4 Å from electron microscopy data for the so called "active" RecA-DNA filament (4). Therefore, the base pairs (or base triplets) in the RecA-DNA complex are either inclined more strongly than now anticipated (41), or stretched in some dynamic fashion so that the rise of the filament per RecA monomer is increased to 15.4 Å compared to 13.6 Å for the static intercalation.

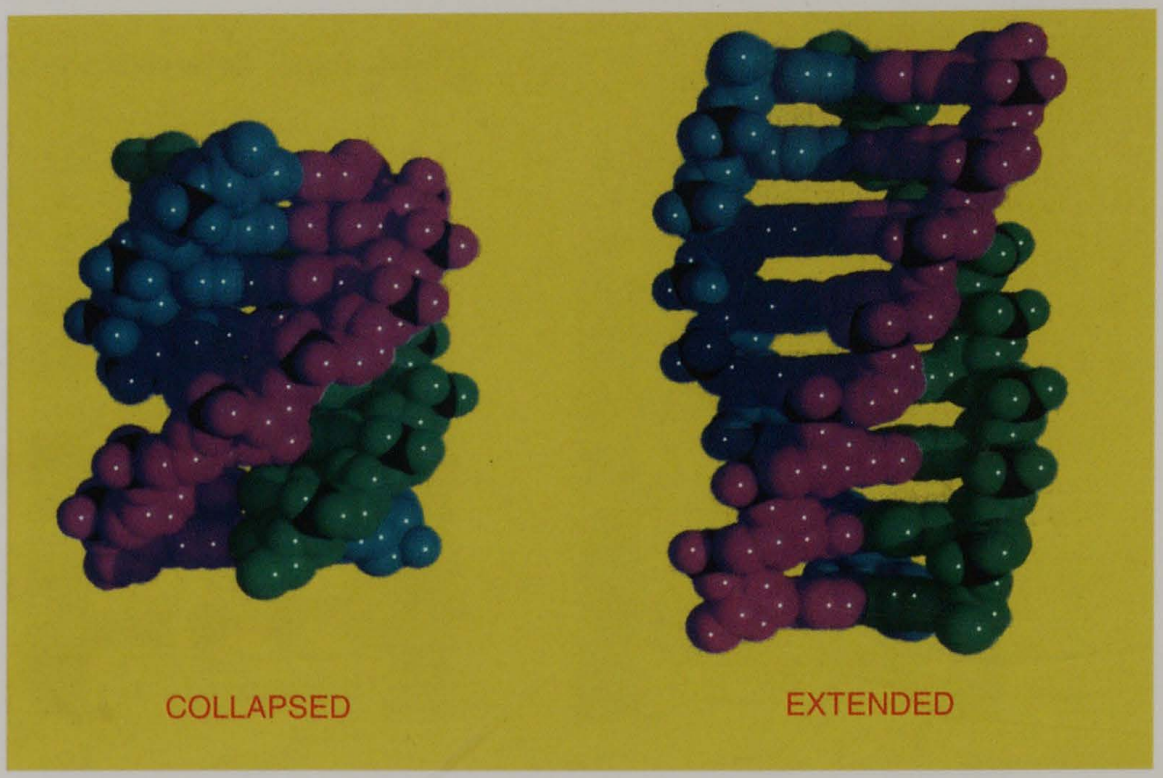

Figure 4: The space filling representation of the collapsed R-form (left) and extended R-form (right). Note that the R-strand (magenta) is parallel to the identical W-strand (green); their sequence is ATGCATGC; the C-strand (blue) has the complementary sequence, GCATGCAT. The structures are minimized with the assumption that base triads are planar and fixed as shown in Figure 3. The collapsed R-form has sugar rings in the C2'-endo domain; in the extended R-form all sugars are in the C3'-endo conformation.

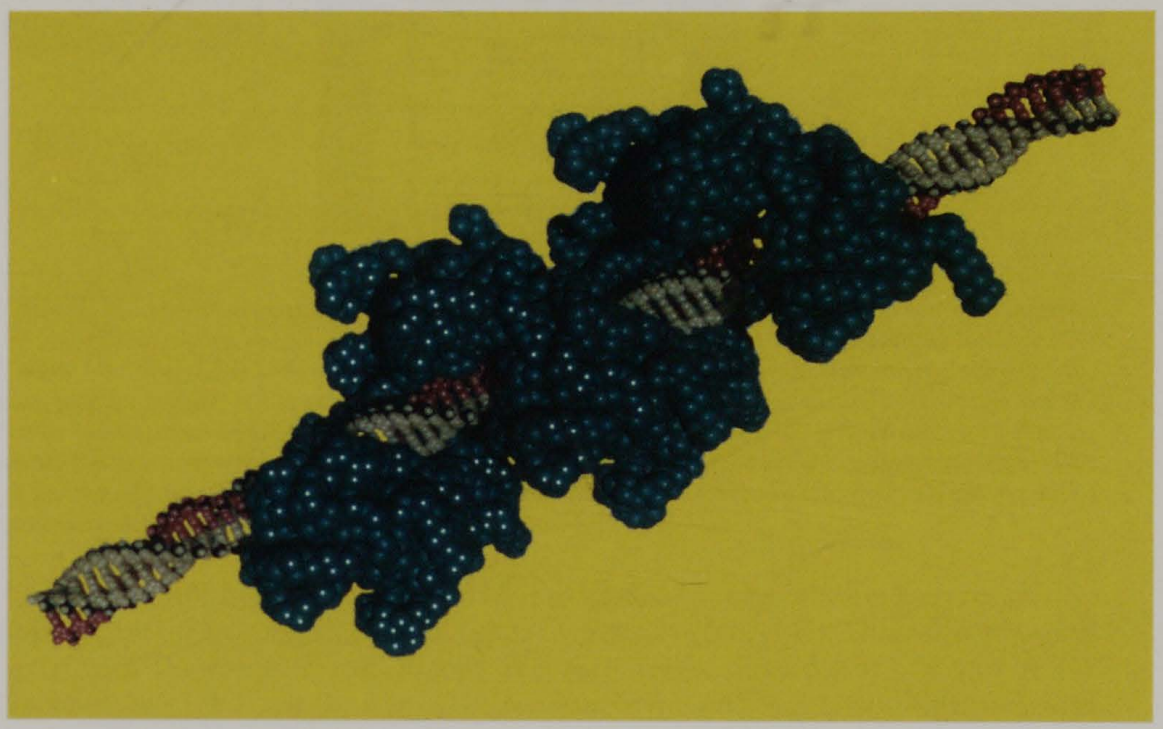

Figure 10: Model of the RecA-DNA filament. The RecA structure (green) is taken from the X-ray study (40); extended R-form DNA is obtained by energy calculations. The WC duplex is shown in yellow, and the third R-strand is in pink. DNA-protein interactions do not include any steric clashes (55).

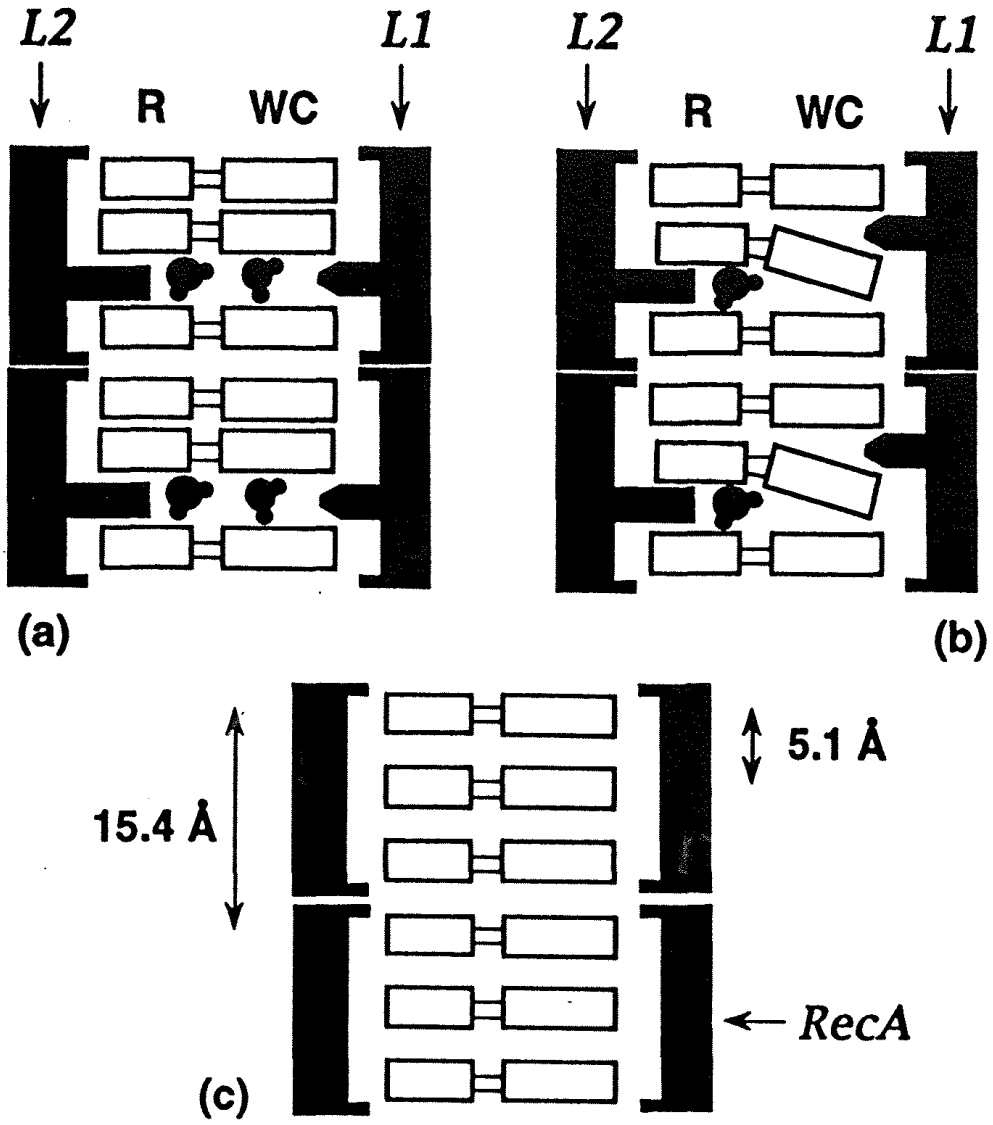

Figure 6: Models of the extended DNA triplex in the complex with RecA. One RecA monomer interacts with three base triplets — (a,b) Two possible schemes of extending DNA by intercalation. Residue Phe/Tyr-204 from loop L2 is assumed to intercalate into the ssDNA, or R-strand. The intercalating residue from loop L1 is either juxtaposed exactly across the residue from L2 (a), or shifted by one base step (b). Notice the water molecules which are expected to fit between the stretched DNA bases since Phe/Tyr aromatic ring is relatively small (compared to purines), and in the loop L1 there are no aromatic residues (40). (c) The simplified structure, in which DNA is uniformly stretched and has a rise of 5.1 Å. This way, DNA extension is consistent with the electron microscopy data (2-4).

Another possibility might be that DNA is extended by intercalation at every second step, so that the average rise would be (3.4 + 6.8) / 2 = 5.1 Å (43). This scheme would agree with the electron microscopy measurements, since 5.1 • 3 ≈ 15.4 Å (4). However, in such a case the RecA-DNA filament must be highly non-uniform, with the RecA monomers having "alternating" rise configurations, 1-2-1-2... (because periodicity of DNA is two steps due to intercalation, whereas the periodicity of RecA-DNA filament is three DNA steps). To the best of our knowledge, this assumption is not confirmed by the electron microscopy or X-ray data.

Above, the mutual positions of DNA bases and RecA residues were discussed. Nothing is known about the specific interactions of DNA backbone with RecA. Therefore, due to the paucity of experimental data, it would be premature to build a detailed model of extended DNA. Moreover, it might well be that any single static structure would not be representative of the real "fluid" R-form DNA, fluctuating among the numerous conformations like those two shown in Figure 6 (a,b). Instead, here we may be considering an "average" structure, in which DNA is uniformly stretched and has a rise of 5.1 Å per step (Figure 6c). This simplified model allows us, first, to analyze the most general conformational features of the extended DNA, and second, to address the possible functional role of the extended DNA.

The optimal extended conformation of DNA shown in Figure 4 (right) has the C3′-endo sugar pucker. Preference for the C3′-endo sugar pucker in triplexes extended by 50%, is consistent with the observation made in the previous section, that sugar repuckering from C2′-endo to C3′-endo leads to an increase in the axial rise. The energetic advantage of the C3′-endo pucker in the extended triplex agrees also with the recent NMR structure of 5′-d(TC_5) with intercalated base pairs (44), where the majority of the sugar rings have C3′-endo conformations. This tendency has a simple mechanistic explanation. When DNA is extended, its sugar-phosphate backbone changes its orientation: it is rotated counter-clockwise (compare the two structures in Figure 4). This is achieved mainly by rotation of the sugar rings around the C1′-N bonds by about 30° in the positive direction (25). For the sugar rings with C2′-endo pucker it means that the glycosidic angle χ would be shifted from the *anti*-region to *high anti*, which is energertically unfavorable (especially for pyrimidines). If the sugars have C3′-endo pucker, the increase in χ leads to a transition from *low anti* to *anti*-region, which is more preferable. That is why, under a stretching tension the sugar pucker is likely to change from C2′-endo to C3′-endo.

The sugar-phosphate backbone in the extended DNA is typical for A-DNA. In particular, the sequence dependent variation of the dihedral angles (25) is comparable with that for the X-ray A-DNA structures. As a consequence, the phosphates in extended triplex are oriented in a quite uniform way, which would facilitate interactions between the DNA and the recombination protein (Figures 4, *right*, and 10).

IV. Recognition of Duplex by Single Strand

The major groove of the WC duplex (both in A and B-forms) comprises a two-dimensional pattern of donor and acceptor groups, where distances between the groups in a horizontal direction are 2.9-3.1 Å, and in a vertical direction 3.4 Å or less (Figure 7). The vertical separation is usually decreased compared to 3.4 Å due to propeller twist in the WC pairs (Figure 5). On the other side, the distances between the groups in one base of the third strand are 2.3 Å. It follows from these values that if interaction between a duplex and the homologous third strand occurs while the bases are stacked, there is no strong preference for a certain motif arranged horizontally (within a single WC pair) over all other motifs. That is, the third strand bases can easily form triplets with the bases of two neighboring WC pairs simultaneously (the "diagonal" interactions shown in Figures 5 and 7a). These "mis-triplings" pose a

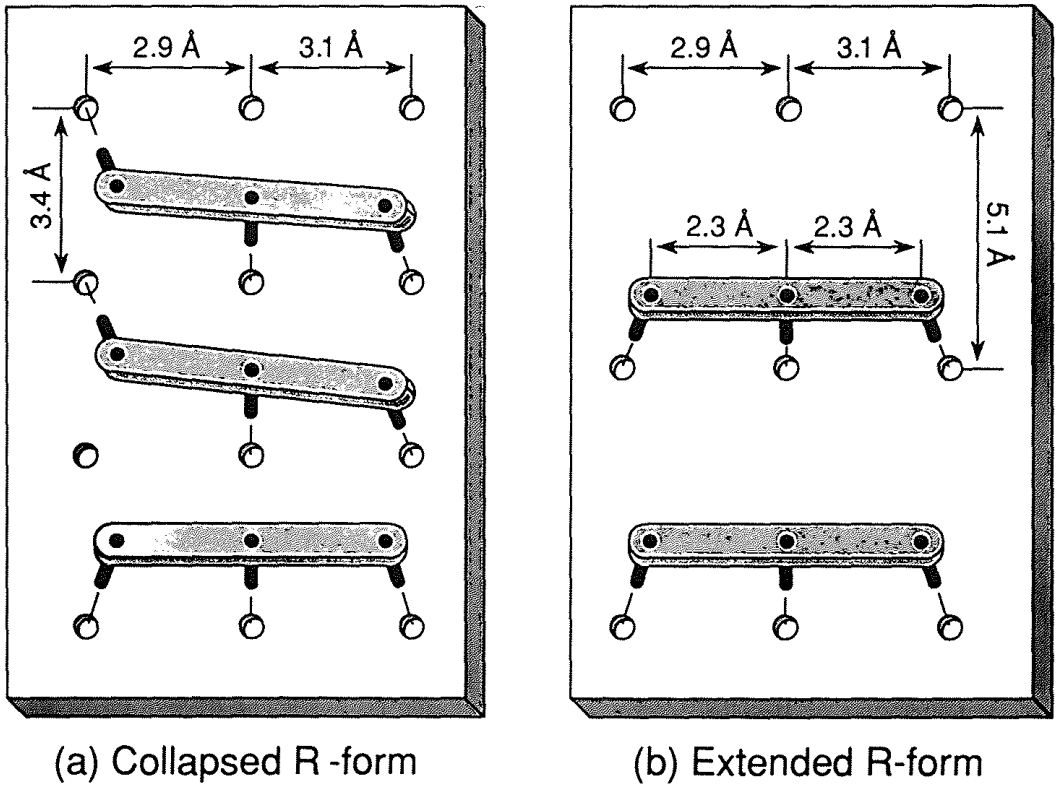

Figure 7: Schematic representation of the hydrogen bonds between the duplex and the third strand. The third strand bases are shown as rigid elements with protruding pegs, and the major groove surface of the duplex is depicted as a board with holes (either pegs or holes represent the donor or acceptor groups of the bases). The scheme explains why the extended R-form is preferable for the stringent mutual recognition of ssDNA and dsDNA. (a) Collapsed R-form with axial rise 3.4 Å. Distances between the donor or acceptor groups in the duplex ("holes in the board") are approximately equal in both the vertical and horizontal directions, so the third strand bases ("rigid elements with pegs") can interact either with bases from the same base pair (horizontal orientation of the lower base), or with bases from the two neighboring pairs (inclined positions of the upper two bases). (b) Extended R-form with axial rise 5.1 Å. The distance between "the holes" in vertical direction is increased by 50% compared to (a), and the only acceptable scheme for interaction is the horizontal position of the bases.

serious problem, since they would delay formation of the homologous base triplets and thus slow the process of recognition.

1. Advantage of the Extended Triplex for Recognition

When DNA is extended by a recombination protein, the inter-layer "diagonal" interactions are unlikely, and binding of the ssDNA to dsDNA can occur exclusively through the intra-layer "horizontal" mechanism shown in Figure 7b. In our opinion, this is the key element for understanding the functional advantage of the extended triplex. We suggest that by stretching DNA, the recombination proteins act to increase the fidelity of the ssDNA--dsDNA recognition and to accelerate it. At least so far, these proteins are required as mediators for the formation of the parallel triple helices with mixed sequences (5-9).

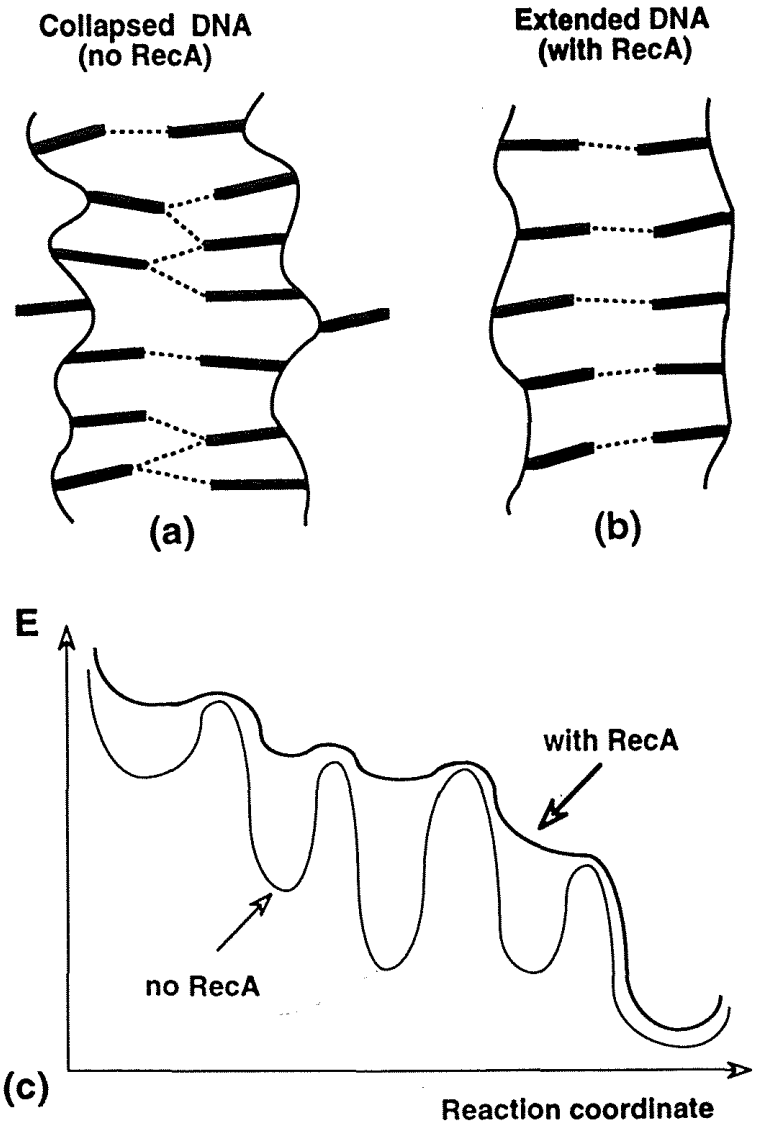

Figure 8: Illustration explaining the slow kinetics of DNA renaturation in the absence of RecA (a; thin line in c), and the fast kinetics in the presence of RecA (b; thick line in c), see text.

The situation with the triple helices is somewhat similar to the case of dsDNA, where kinetics of DNA reassociation is accelerated about 1000-fold in the presence of RecA (45). This effect can be explained by a simple scheme shown in Figure 8 (a-c). Without an extending agent a complicated random net of non-planar hydrogen bonds is formed (a), which creates multiple local minima in the energy profile (c), and thereby slows finding the global minimum, corresponding to the helical state of the dsDNA. Extension of DNA abolishes these mismatches (b), and thus accelerates the recognition process (c). The regularizing scheme can be related to both duplex DNA and triplex R-form. In the case of duplex, possible mismatches include the non-WC pairing observed in CACA:TGTG sequences (35); for the R-form triplexes the feasible mismatches are shown in Figure 5.

This scheme is not directly applicable to the non-enzymatic anti-parallel triplexes, such as H-form (18,19). Indeed, in H-form the third strand bases are shifted from the center of the major groove and interact with only one strand containing purines. The two adjacent purines from this strand are nearly parallel to each other, even when the propeller twist deviates from zero (see the two guanines in WC duplex, Figure 5). The vertical separation between the two purines also remains unchanged, ≈ 3.4 Å. Accordingly, the inter-layer hydrogen bonding is much less probable in H-form compared to the parallel triplex considered here. Therefore, in principle, the kinetics of formation and/or dissociation of a parallel R-form triplex is expected to be slower than that of H-form. That the conformational restrictions imposed on DNA by an extending agent (e.g., RecA) might influence the kinetics of DNA association, opens an interesting perspective for the molecular dynamics studies of the processes involved in the more general chemical recognition of nucleic acids.

2. "Chaperone" Role of RecA.

The recombination protein restricts movements of the bases, both in the WC duplex and in the third R-strand. These restrictions are of two kinds: lateral and vertical.

Lateral restrictions are shown schematically in Figure 3 as a "corridor" between the broken lines. When the sugar-phosphate moieties of the first (W) and the third (R) strands are bound to the protein, the "horizontal" in-plane fluctuations of the bases are decreased (Figure 9a). In particular, we assume that the bases of R-strand cannot move to the right relative to their isomorphic position by more than about 2 Å. This enforces the formation of the homologous triplets and prevents mismatching (see "Electrostatic Recognition" below).

Vertical restrictions. Increase in the average base-to-base vertical separation up to 5.1 Å would eliminate formation of the "diagonal" hydrogen bonds in the triplex, as shown in Figure 7b. Note that the two restrictions (lateral and vertical) are not completely independent. When the DNA backbone is extended, the in-plane movements of its bases are automatically limited as well. Thus, the "corridor" for the bases becomes even narrower.

In our opinion, by binding to ssDNA and stretching it, RecA protein serves as a *"chaperone"*, strongly guiding the mutual approach of ssDNA and dsDNA, and drastically decreasing the number of possible interactions between them.

The 50% unwinding of duplex DNA by RecA may also be advantageous for recognition. When the DNA helical pitch is increased from 10-10.5 bp in solution to 18 bp in the RecA-DNA complex (4), the length of dsDNA accessible from one side is also increased two-fold: from 5 to 9 bp. Hence, 9 bp segment of the duplex can interact simultaneously with 9 nucleotides long fragment of ssDNA buried in the RecA filament. This estimate agrees, perhaps incidentally, with the data by Hsieh et al.(46) indicating that 8-10 bases are involved in initiation of pairing between ssDNA and dsDNA.

Table I
Electrostatic Code of Recognition

W-strand				C-strand		R-strand			
A	-	+	-	(+)	T	A	(+)	-	+
T	(+)	-	+	-	A	T	-	+	-
G	-	-	+	(+)	C	G	+	+	-
C	(+)	+	-	-	G	C	-	-	+

For the WC pairs, the partial charges in major groove are given, going from left to right as in Figure 2. The R-strand base charges are facing the major groove as shown in Figure 2. Minuses denote acceptors of protons (N, O); pluses are for the donors (NH_2 and NH-groups); (+) are for the C5-H and C5-Met groups of pyrimidines. *Electrostatic code:* The three leftmost groups in a WC pair are complementary to the three groups from the corresponding R-strand base (see Figure 3).

3. Electrostatic Recognition Code

Charge complementarity is an important factor in achieving specific binding of various ligands to DNA. The electrostatic interactions are operative both in the course of the initial approach of a ligand (47,48), and in ensuring the sequence specific binding at the final stages of recognition, as the hydrogen bonds are formed (49). Consideration of the donor-acceptor pattern of base pairs by Seeman et al. has been useful in understanding the specificity of protein-DNA recognition (50). Based on a similar approach for the DNA-DNA interactions, we are now presenting a simple mechanism for how the third R-strand can recognize the major groove surface of the WC pairs.

Consider the partially charged groups in the major groove of the WC pairs, going from left to right as in Figure 2. For example, in AT pair these are N7, $(N6)H_2$, O4 and CH_3. Let us mark them with + and -; then we obtain the patterns shown in the left side of Table I. Next, for each base in the third R-strand we mark the three groups directed toward the WC pair, also going from left to right (Figure 2). In adenine these are H2, N1, $(N6)H_2$; they are presented on the right side of Table I.

Now we can formulate the recognition code. For every base the triple pattern is strictly complementary to the three leftmost positions in the corresponding pair (e.g. (+ - +) of A is complementary to the leftmost (- + -) in the AT pair), see Table I. If the RecA protein restricts the approach of the third R-strand to the double helix so that the R-strand can interact only with the "left" part of the pair, then the above complementarity ensures perfect *electrostatic recognition*.

Electrostatic interactions might allow a fast and efficient search of homologous sequences in the recombination process. During the initial approach of R-strand to the WC pair, at a distance between donors and acceptors too long for formation of strong hydrogen bonds, the electrostatic repulsion is sufficient to prevent mismatches. The WC duplex could then "slide" further along the third strand searching for the homologous sequence, without forming hydrogen bonds with "wrong" oligonucleotides and getting into kinetic "traps".

If, for example, G or T instead of A, were approaching the AT pair, they would be easily rejected because of the repulsion between the central groups shown in Table I. If, however, cytosine were approaching the AT pair, its N3 and H(N4) groups would match perfectly the H(N6) of adenine and O4 of thymine. In the two central positions capable of donating and accepting protons, the pairs AT and CG, TA and GC are identical (50), therefore some additional feature is necessary to distinguish among them. We suggest that this role is being played by the leftmost position: e.g., in the case of (AT):C mismatch there would be an unfavorable repulsion between O2(Cyt) and N7(Ade), see Table I and Figure 2.

The most unorthodox idea in this scheme is the assumption that the protons H2, H5 and H(Met) are important for recognition (their interactions in the leftmost position are shown by perpendicular bars in Figure 2). Obviously, the positive charge of a hydrogen attached to a carbon is significantly less than the charge of an amino proton, and these protons in the 5-th position of pyrimidines and 2-nd position of adenine cannot form strong "classical" hydrogen bonds. Nevertheless, one must admit that attraction of H5 or H2 to N7 or O2 atoms is more favorable than repulsion between the charged groups of the same sign.

In this way, considering interactions between the third strand bases and the duplex pairs in all the three positions (see the arrows in Figure 3), the discrimination between A and C, G and T can be explained unambiguously (25). Thus, we conclude that the electrostatic interactions in general, and at the N7(Pur)/C5(Pyr) position in particular, may serve as a necessary device for eliminating mismatches.

Recently, Rao and Radding (51) described the so-called "self-recognition" of two identical strands in the presence of RecA ("R-W recognition" in our notations). Clearly, our scheme is entirely consistent with this observation. The two charges in a W-base are complementary to the two leftmost charges in a R-base (Table I). Besides, these two charges comprise four different combinations of plus and minus for all the four bases. So, interaction between the two strands positioned as R and W in Figure 3, is indeed sufficiently discriminatory to secure the "self-recognition" of A by A, C by C, G by G, and T by T.

V. Putative Mechanism of Strand Exchange

Isomorphism of the base triplets supposedly involved in recognition (Figure 3), implies that the strand exchange can also be a universal, sequence independent process. Consider the base-base interactions before and after the strand exchange (Figure 9). The groups from the R-strand and C-strand bases, recognizing each other, retain their interaction after the R:C pair is formed (these are the O6 and N4 groups in Figure 9). Hence, movements of the R- and C-strands during the strand exchange can be envisioned as coordinated rotations in the opposite directions during which the two donor-acceptor groups will remain hydrogen bonded for all sequences.

Location of the homologous R-strand in the immediate vicinity of the C-strand facilitates opening of the initial WC pairs. Indeed, in this case after opening of a pair only one base, W, is exposed to solvent (Figure 9b), whereas in the absence of the R-strand, both the W- and C-strand bases would be exposed, which is energetically

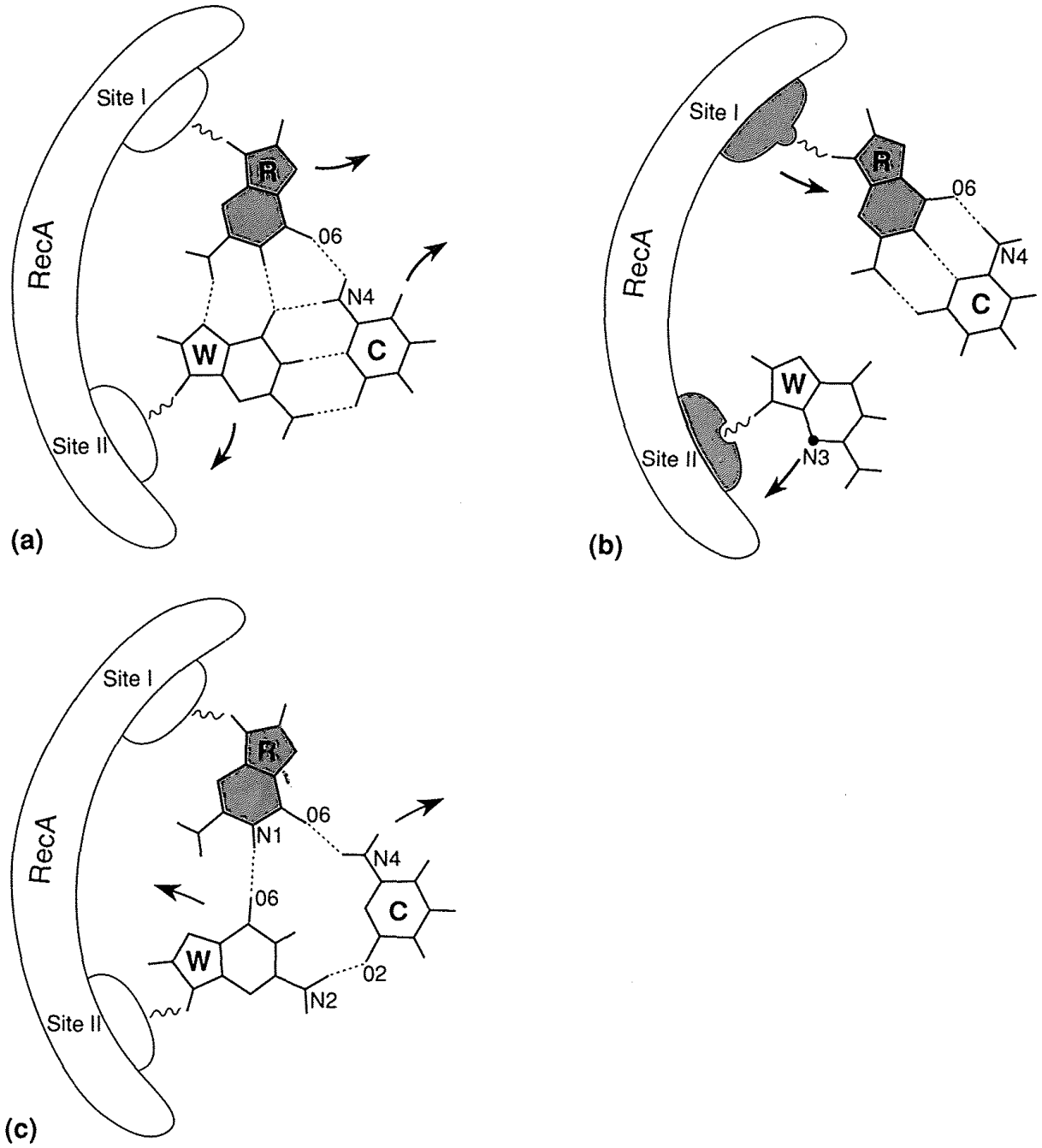

Figure 9: Putative model of strand exchange. Movements of bases shown here, might be applicable to the process of RNA transcription as well (see text). (a) Initial interaction of the third R-strand with site I in RecA, and of the duplex W-strand with site II, stabilizes the isomorphic geometries of triplets shown in Figure 3. The arrows show hypothetical motions of the bases during the strand exchange. Sites I and II (54) correspond to the loops L2 and L1 respectively (40). The C-strand does not interact directly with the RecA protein (62). (b) Base pair R:C is formed after the strand exchange. Note that the O6--N4 interaction remains the same as in (a). The convex and concave shapes of sites I and II depict conformational changes in the RecA protein, which are assumed to induce the motions of the R- and W-bases shown by arrows. (c) Intermediate stage showing how the partial opening of the WC base pair toward the major groove might not prevent the third strand interacting with both the W- and C-strands. Note that here, contrary to scheme (b), the W-strand base would move counter-clockwise.

unfavorable. Thus, opening of WC pairs is expected to occur easier *after* recognition by R-strand (55).

The exchange mechanism might be thought of as analogous to a train of three gear wheels (bases). The protein might act as a drive wheel in this set of gears (25). For example, conformational changes in RecA in the vicinity of site I could lead to the rightward movement of the R-strand (Figure 9b); as a result the C-strand would follow along, and the two bases would rotate in the opposite directions: R counter-clockwise and C clockwise. In the end of this hypothetical process the R- and C-strands would be hydrogen bonded, while the W-strand would be separated. The motions described could occur cooperatively, at least in all the three triplets bound to the same RecA subunit (40).

In addition to the "expansion" of the site I and rightward motion of the R-strand, the "contraction" of the site II could "pull" the W-strand down- and leftward, thus relieving the C-strand and facilitating its pairing with the R-strand (Figure 9b). The universal N3(Pur)/O2(Pyr) acceptor can serve as an "attractor" to the protein and thus assist the strand exchange. Going further in this speculation, we hypothesize that the ATP hydrolysis, necessary to dissociate the W-strand from RecA protein (52,53), is used, in particular, to break this N3/O2--RecA bond. This is the difference between the deproteinized collapsed R-form and the RecA-covered extended R-form: the equilibrium (a) $\Leftrightarrow$ (b) might be shifted toward (b) in the extended structure (13) due to interaction between N3/O2 and the protein.

The scheme shown in Figure 9 is in principle similar to the mechanism proposed earlier by Howard-Flanders, West & Stasiak (54). Yet another scheme might have the R-strand as a whole rotating counter-clockwise around the helical axis of the DNA triplex, substituting for the W-strand (23). All these mechanisms are tentative, pending further experiments. The bottom line is that as a result of the central positioning of the third strand bases, the strand exchange could be comprised of simple rotations of bases without any substantial translational motions.

The triple helix proposed here can easily fit into the channel in the RecA filament deduced from the X-ray structure (40), and there is enough space for the postulated movements of the strands (55), see Figure 10. Although the structures of the loops L1 and L2 that have been proposed to interact with the two DNAs are not determined yet (40), simple model building shows that the scheme presented in Figure 7 is stereochemically feasible (55).

VI. Dynamic Nature of the RecA-DNA complex

In a duplex under the "standard" conditions, the lifetimes of base pairs are of the order of milliseconds (56). When DNA is stretched, the sliding and opening of bases would become much easier due to decrease of "friction" from their neighbors (Figure 6). This is similar to the case of DNA bending, where disruption of base stacking eases base pair opening (57). If we assume that base pairs in the duplex are transiently opened into the major groove (at least partially, as in Figure 9c), then the

pairing of bases from the R- and C-strands might be involved in the process of recognition. In effect, the third R-strand can recognize both the *open* and the *closed* structures of the duplex, since the *open C-base* and the *closed WC pair* have the same topography of the donor-acceptor groups (Table I).

Two consequences follow from this consideration. First, the actual structure of the complex (RecA+ssDNA+dsDNA) could be an equilibrium between all the three stages shown in Figure 9. This notion is consistent with the observation by Adzuma (13) that in the region of ssDNA-dsDNA pairing, the "outgoing" W-strand was susceptible to dimethylsulfate and potassium permanganate. To explain this experiment it is not necessary to assume that the strand exchange occurs irreversibly, as soon as the synaptic complex is formed (13). The alternative interpretation is also possible. Specifically, if there is a dynamic equilibrium in the RecA-DNA complex (Figure 9), then the W-strand bases would be easily accessible to such small chemical reagents, when the R-and C-strands are paired as in Figures 9b and 9c, and the W-strand is slightly removed from them.

Second, in the strand exchange model shown in Figure 9, the three bases are assumed to move in the plane of the triplet. It is clear, however, that the out-of-plane motions of the bases could also facilitate the strand exchange, since rotations of the bases around the glycosidic axes would disrupt the WC pair and might help in creating the new R:C pair. It is evident that these rotations of bases are less restricted in the extended R-triplex because of the "looser" stacking (Figure 6).

Note that this reasoning involving the dynamic nature of the RecA-DNA complex, provides a structural explanation for seemingly conflicting results reported by several groups, compare refs.(5-9) and (13,14). The data of Adzuma (13) have been discussed above. As to the paper by Jain et al.(14), the authors observed that strand exchange occurs when guanines in the W-strand are substituted with 7-deaza guanines (G*). They concluded that the N7 of guanine from the W-strand is not involved in interaction with the third strand (14). However, based on the idea of an extremely fluid structure of the extended DNA, we can give another interpretation to their results.

Most likely, 7-deaza substitution in the Watson strand leads to formation of a non-planar (G*C):G triplet, with the third strand guanine rotated around its long axis to avoid the (C7)H--H(N2) clash (Figure 2). The modified triad (G*C):G would be less stable than the standard (GC):G, since (G*C):G is stabilized by two hydrogen bonds instead of three. So, the 7-deaza substitution G → G* does not exclude completely formation of the triad (G*C):G, but rather makes the triad less stable. In fact, the decrease in stability of (G*C):G compared to (GC):G is consistent with an increase in the rate of strand exchange, detected by Jain et al.(14).

At present, it is impossible to conclude whether or not "homology recognition is mechanistically coupled with exchange of DNA strands" (13). In principle, *the recognition can occur through the formation of triplets* (Figure 9a). At the later "proof-reading" stage, verification of recognition might indeed be coupled with disruption

of the initial WC pairs (Figure 9b). However, the transition between the initial state, (W:C):R, and the final one, W+(R:C), is expected to be so subtle, that further experiments are indispensable to elucidate the thermodynamics and kinetics of this process.

Conclusion

Based on theoretical conformational analysis, we propose a novel parallel triple helix, the R-form DNA. It is actually a family of structures, rather than a single conformation. Our study deals with both the extended R-form stabilized in the complex with RecA (average rise ≈ 5.1 Å), and the protein-free collapsed R-form with a rise of ≈ 3.4 Å. Unlike the previously known triple helices, the R-form is stereochemically possible for any arbitrary sequence of nucleotides. The proposed triads of bases are nearly iso-geometric. As a result, the novel structure is almost independent of sequence, which should be functionally important: virtually no adjustments of DNA conformations are required, as the universal protein machinery promotes the recognition and strand exchange of DNAs with any sequence.

Note that the present modelling effort is a non-conventional approach to computer analysis of the macromolecular structure in which compelling functional reasoning is utilized to choose among several classes of alternative models. Based on the limited amount of experimental information, we are attempting to elucidate an appropriate model of the parallel DNA triplex, which would be advantageous for interaction with recombination proteins. Therefore, rather than minimizing the free energy of DNA with the given sequence in solution (e.g., taking into account the details of solvation, increasing the precision of the electrostatic energy calculations, etc.), we are paying attention primarily to the regularity restrictions imposed on DNA by the protein (4,40).

The isomorphic triplets of bases, set forth here, are generally consistent with available experimental data for the collapsed deproteinized triplexes with mixed sequences (7-9). We assume, however, that essentially the same system of hydrogen bonds is formed in the biologically important extended triplex. The model of electrostatic recognition and strand exchange, based on the isomorphism of base triplets, provides strong, although indirect, support for this assumption.

This model proved to be consistent with the "self-recognition" of two identical strands, observed recently (51). Our postulate on the critical role of the leftmost N7/C5 position for recognition is experimentally testable: the most straightforward way would be make the two-strand experiments (51) with the 7-deaza substitutions in purines in the W-strand (Figure 2), in the same way as it was done in the three-strand experiments (8,9).

Now we can try to visualize in general, how recombination proteins facilitate the recognition of ssDNA and dsDNA in the system with an astronomically large number of possible interactions. First, RecA binds to ssDNA and fixes it in the extended underwound conformation (2-4). Then dsDNA is "threaded through" the binary complex (58). The "Brownian" sliding of dsDNA along ssDNA is facilitated by the

extended conformation of DNA (Figure 7). At this stage the hydrogen bonds are not necessarily formed: the selection can occur by virtue of longer range electrostatic interactions (Figure 3); the R-form can be a looser, more expanded association rather than a tight stable triplex. After the sufficiently long homology between ssDNA and dsDNA is found, the hydrogen bonds are formed; the ternary complex (dsDNA+ssDNA+RecA) looks as shown in Figure 10. Then the two identical strands (W and R in Figure 9) are exchanged. One should be mindful that the equilibrium between several structures is not excluded (Figure 9).

The novel *parallel* triads of bases are relatively weak compared to the conventional *antiparallel* triads found in numerous non-enzymatic DNA triple helices (16-19). This might be one of the reasons why the parallel R-like triplexes have not yet been reproducibly formed in the absence of recombination proteins (at least so far). But from the biological point of view, *the parallel triple helices do not have to be very stable.* The function of these triplexes is not a storage of information or stabilization of some unusual secondary structure. Instead, they serve as transient intermediates to aid recognition and the subsequent strand exchange. If these triplexes were too stable, movement of bases would be hampered. By analogy with the "molten globule" of proteins (59) we term the highly dynamic extended DNA a "molten helix" (9).

Although the extended R-triplex is energetically unfavorable itself, it is stabilized by the recombination proteins (with perhaps intercalation between the bases and interaction with sugar-phosphate backbone). The functional advantages of the extended DNA structure were discussed above: (i) accelerating the kinetics of recognition between ssDNA and dsDNA; (ii) increasing its fidelity; (iii) facilitating movements of bases during strand exchange. Thus, we conclude that RecA plays the role of a *chaperone* and *enzyme*, in inducing a conformation in the ligand (DNA), favorable for the relevant reaction (recognition and strand exchange).

Recombination and Transcription

Finally, we wish to point out a possible similarity between these two fundamental processes. According to the "conventional" model, widely accepted now, the DNA duplex in the complex with RNA polymerase has to be locally opened, so the growing RNA chain can recognize the coding strand of DNA (60). However, following the other models, RNA transcription could take place without opening the WC pairs, with the RNA chain located in the wide groove of the DNA, so that the transient parallel triplex is formed, see (15,61) and references therein. We suggest that recognition of the WC duplex by ribo-triphosphates may occur, at least at the early stages, through the isomorphic triplets proposed in this study (Figure 3). Subsequently, these triplets facilitate opening of the initial WC pairs and formation of the new pairs (55), as shown in Figures 9b and 9c. Unwinding and stretching of DNA in the complex with RNA-polymerase might be one of the factors promoting rigorous recognition of DNA by ribo-triphosphates (Figures 5, 7). Hence, stereochemical analogy between recombination and transcription proves to be deeper than is commonly recognized, and the conformational features of the parallel triple helices reported here may be applicable to both these processes.

Acknowledgements

The authors are grateful to Drs. N.R. Cozzarelli, D.M. Crothers, V.I. Ivanov, Y.L. Lyubchenko, S.M. Mirkin, B. Norden, C.M. Radding and A. Stasiak for stimulating discussions and constructive critical remarks. We appreciate help of Dr. S. Durell with computer graphics.

References and Footnotes

1. Smith, G., *Cell 64*, 19-27 (1991).
2. Stasiak, A., DiCapua, E. and Koller, T., *J. Mol. Biol. 151*, 557-564 (1981).
3. Stasiak, A. and DiCapua, E., *Nature 299*, 185-186 (1982).
4. Egelman, E., *Curr. Opin. Struct. Biol. 3*, 189-197 (1993).
5. Hsieh, P., Camerini-Otero, C.S. and Camerini-Otero, R.D., *Genes and Devel. 4*, 1951-1963 (1990).
6. Rao, B.J., Detreix, M. and Radding, C.M., *Proc. Natl. Acad. Sci. USA 88*, 2984-2988 (1991).
7. Rao, B.J., Chiu, S.K. and Radding, C.M., *J. Mol. Biol., 229*, 328-343 (1993).
8. Kim, M.G., Hsieh, P., Camerini-Otero, C.S., Raghunathan, G., Jernigan, R.L., Zhurkin, V.B. and Camerini-Otero, R.D., *Biophys. J. 61*, a481 (1992).
9. Kim, M.G., Zhurkin, V.B., Jernigan, R.L. and Camerini-Otero, R.D., submitted for publication (1993).
10. West, S., *BioEssays 13*, 37-38 (1991).
11. Stasiak, A., *Molec. Microbiol. 6*, 3267-3276 (1992).
12. Umlauf, S.W., Cox, M.M. and Inman, R.B., *J. Biol. Chem. 265*, 16898-912 (1990).
13. Adzuma, K., *Genes and Devel. 6*, 1679-1694 (1992).
14. Jain, S.K., Inman, R.B. and Cox, M.M., *J. Biol. Chem. 267*, 4215-4222 (1992).
15. Stent, G.S., *Adv. Virus Res. 5*, 95-149 (1958).
16. Felsenfeld, G., Davies, D.R. and Rich, A., *J. Amer. Chem. Soc. 79*, 2023-2024 (1957).
17. Arnott, S. and Selsing, E., *J. Mol. Biol. 88*, 509-521 (1974).
18. Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D., *J. Biomol. Struct. Dynam. 3*, 667-669 (1986).
19. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank-Kamenetskii, M.D., *Nature 330*, 495-497 (1987).
20. Zhurkin, V.B., Raghunathan, G., Camerini-Otero, R.D. and Jernigan, R.L., *J. Biomol. Struc. Dyn. 8*, a268 (1991).
21. Lacks, S., *Genetics 53*, 207-235 (1966).
22. Umlauf, S., Ph. D. Thesis, Univ. Wisconsin- Madison, Program in Cell. and Molec. Biol. (1990).
23. Zhurkin, V.B., Raghunathan, G., Camerini-Otero, R.D. and Jernigan, R.L., *Biophys. J. 61*, a482 (1992).
24. Zhurkin, V.B., Raghunathan, G., Ulyanov, N.B., Camerini-Otero, R.D. and Jernigan, R.L., *J. Biomol. Struc. Dyn. 10*, a221 (1993).
25. Zhurkin, V.B., Raghunathan, G., Ulyanov, N.B., Camerini-Otero, R.D. and Jernigan, R.L., submitted for publication (1993).
26. Zhurkin, V.B., Poltev, V.I. and Florent'ev, V.L., *Mol. Biol. (Engl. Transl.) 14*, 882-895 (1980).
27. Poltev, V.I. and Shulyupina, N.V., *J. Biomol. Struct. Dynam. 3*, 739-765 (1986).
28. Chuprina, V.P. and Poltev, V.I., *Nucl. Acids Res. 11*, 5205-5222 (1983).
29. Prive, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. and Dickerson. R.L., *Science 238*, 498-504 (1987).
30. Zhurkin, V.B., Lysov, Y.P. and Ivanov, V.I., *Biopolymers 17*, 377-412 (1978).
31. Ulyanov N.B. and Zhurkin V.B., *J. Biomol. Struct. Dynam. 2*, 361-385 (1984).
32. Zhurkin, V.B., Ulyanov, N.B., Gorin, A.A. and Jernigan, R.L., *Proc. Natl. Acad. Sci. USA 88*, 7046-7050 (1991).
33. Macaya, R., Wang, E., Schultze, P., Sklenar, V. and Feigon, J., *J. Mol. Biol. 225*, 755-773 (1992).
34. Raghunathan, G., Miles, H.T. and Sasisekharan, V., *Biochemistry 32*, 455-462 (1993).
35. Timsit, Y., Vilbos, E. and Moras, D., *Nature 354*, 167-170 (1991).
36. Ivanov, V.I., Minchenkova, L.E., Schyolkina, A.K. and Poletayev, A.I., *Biopolymers 12*, 99-110 (1973).

37. Zhurkin, V.B., Gorin, A.A., Charakhchyan, A.A. and Ulyanov, N.B. in: *Theor. Biochem. & Molec. Biophysics*, (Beveridge, D.L. and Lavery, R. Eds), Adenine Press, New York, pp. 411-431 (1990).
38. Poncin, M., Hartmann, B. and Lavery, R., *J. Mol. Biol. 226*, 775-794 (1992).
39. Ivanov, V.I., Minchenkova, L.E., Minyat, E.E., Frank-Kamenetskii, M.D. and Schyolkina, A.K., *J. Mol. Biol. 87*, 817-833 (1974).
40. Story, R.M., Weber, I.T. and Steitz, T.A., *Science 355*, 318-325 (1992).
41. Norden, B., Elvingston, C., Kubista, M., Sjoberg, B., Ryberg, H., Ryberg, M., Mortensen, K. and Takahashi, M., *J. Mol. Biol. 226*, 1175-1191 (1992).
42. Kim, S.K., Norden, B. and Takahashi, M., *J. Biomol. Struct. Dynam. 10*, a091 (1993).
43. Sobell, H.M., Tsai, C.C., Jain, S.C and Gilbert, S.G., *J. Mol. Biol. 68*, 21 (1977).
44. Gehring, K., Leroy, J.-L. and Gueron, M., *Nature 363*, 561-565 (1993).
45. McEntee, K and Weinstock, G.M., in: *The Enzymes, Vol. XIV*, Nucleic Acids, Part A (Boyer, P.D., ed.), Academic Press, New York, pp. 445-470 (1981).
46. Hsieh, P., Camerini-Otero, C.S. and Camerini-Otero, R.D., *Proc. Natl. Acad. Sci. USA 89*, 6492-6496 (1992).
47. Zakrzewska, K., Lavery, R. and Pullman, B., in: *Proc. of the Second SUNYA Conversation in the Discipline Biomol. Stereodyn.*, (Sarma, R.H., ed.), Adenine Press, New York, Vol I, pp.327-343 (1981).
48. Steitz, T.A., Weber, I.T and Matthew, J.B., *Cold Spring Harbor Symp. Quant. Biol. 47*, 435-440 (1983).
49. Sharp, K.A., Honig, B. and Harvey, S.C., *Biochemistry 29*, 340-346 (1990).
50. Seeman, N. C., Rosenberg, J. M. and Rich, A., *Proc. Natl. Acad. Sci. USA 73*, 804-808 (1976).
51. Rao, B.J. and Radding, C.M., *Proc. Natl. Acad. Sci. USA 90*, 6646-6650 (1993).
52. Menetski, J.P., Bear, D.G. and Kowalczykowski, S.C., *Proc. Natl. Acad. Sci. USA 87*, 21-25 (1990).
53. Rosselli, W. and Stasiak, A., *J. Mol. Biol. 216*, 333-350 (1990).
54. Howard-Flanders, P., West, S.C. and Stasiak, A., *Nature 309*, 215-220 (1984).
55. Zhurkin, V.B. and Jernigan, R.L., unpublished observation (1993).
56. Leroy, J.-L., Charretier, E., Kochoyan, M. and Gueron, M., *Biochemistry 27*, 8894-8898 (1988).
57. Ramstein, J. and Lavery, R., *Proc. Natl. Acad. Sci. USA 85*, 7231-7235 (1988).
58. Howard-Flanders, P., West, S.C., Cassuto, E., Hahn, T.-R. and Egelman, E., in: *DNA Replication and Recombination*, Alan R. Liss, Inc. pp.609-617 (1987).
59. Ptitsyn, O.B., *J. Prot. Chem. 6*, 273-293 (1987).
60. Florentiev, V.L. and Ivanov. V.I., *Nature 228*, 519-522 (1970).
61. Riley, P.A., *Nature 228*, 522-525 (1970).
62. Chow, S.A., Honigberg, S.M., Bainton, R.J. and Radding, C.M., *J. Biol. Chem. 261*, 6961-6971 (1986).