Geometry and Energetics of DNA Basepairs and Triplets from First Principles Quantum Molecular Relaxations

James P. Lewis and Otto F. Sankey
Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287-1504

ABSTRACT A first principles model for calculating hydrogen bonding interactions, previously applied to water, is here applied to the more difficult problem of interactions between DNA bases. We first consider the energetics and geometry for the A-T and the G-C basepairs, comparing our results to other calculated results as well as to experiment. Next, we study the interactions of isomorphic DNA base triplet structures, which are important because of their suggested role in the recombination process. We find that energetically the third base in the triplet tends to favor a position along the dyadic axis, where it is hydrogen bonded to both bases in the duplex.

INTRODUCTION

Within 10 yr after the discovery of Watson-Crick basepairing, a theory of the central role of such basepairing in general recombination was proposed by Holliday (1964), among others (Camerini-Otero and Hsieh, 1993). Subsequent genetic experiments led to the development of two models, the Meselson-Radding model (Meselson and Radding, 1975) and the double-strand break repair model (Szostak et al., 1983). In both models, only three strands are required for homologous recognition because only one strand is used in scanning the homologous duplex. It was established (Stasiak 1992; West et al., 1982; 1983) that the most productive homologous recognition followed by strand exchange occurs between a single-stranded region in one molecule and a homologous duplex region of the second DNA molecule.

Felsenfeld et al. (1957) reported the first experimentally observed triple helix in nucleic acids. Much of the initial investigations of triple-stranded polynucleotide complexes occurred in the following two decades, but more recently there has been a renewed interest (for example, Hsieh et al., 1990; Pilch et al., 1991; Durland et al., 1991; Zhurkin et al., 1994a,b). Theoretical studies of nitrogen base interactions performed before 1979 can be found within two review articles (Rein, 1979; Danilov et al., 1979). In addition, there have been many theoretical studies, using both classical and quantum mechanical methods, of the interactions between DNA basepairs (for example, Langlet et al., 1981; Zhurkin et al., 1981; Poltev and Shulyupina, 1986). More recently, theoretical studies of the interactions of DNA base triplets have been completed both by using pair potentials (Zhurkin et al., 1994a,b) and on the ab initio level (Jiang et al., 1994). Our motivation for this work is twofold: to test our electronic structure method on the more complex case of hydrogen-bonding interactions between DNA bases and to provide additional information on the energetics of these DNA base triplets.

In biological systems, four different bonding mechanisms play a role: covalent, ionic, van der Waals, and hydrogen bonding. Many techniques (too numerous to reference) have been developed and used for the difficult task of describing the strong covalent and ionic interactions. In biological systems, hydrogen bonding is one of the most important, if not the most important, type of interaction. However, few techniques address the hydrogen bonding interactions at the ab initio level, which are not enormously computationally demanding.

In an earlier work (Ortega et al., 1994), we introduced a truly unique first principles quantum mechanical approach to handle the hydrogen bond or other weak interactions between closed shell molecules. This method was applied to an initial "test case" of H_2O dimer and trimer systems. The results were comparable to those of ab initio methods applied by others. However, the method we developed is orders of magnitude faster than most other ab initio methods and can therefore, for the first time, be used to simulate hydrogen bonding within large systems such as biological molecules. The intermolecular interactions are handled in a quantum mechanical manner, and there is no fitting of parameters. The interactions between entities are true many-body interactions, as they involve the electronic states of the complete system that in general are delocalized.

We use this method to tackle the difficult problem of hydrogen bonding interactions between DNA bases. In particular, we study the interactions between DNA triplets, which are indicated to have biological significance in homologous recombination. DNA triplets involve the standard Watson-Crick basepair, whose geometry we keep fixed with respect to each other, plus an additional base. Each molecule can be treated as a closed shell molecule with weak hydrogen bonds between molecules. Our major goal is to obtain a good description of the weak intermolecular interactions. We keep each DNA base molecule rigid, so that the strong covalent intramolecular interactions are kept constant.

Received for publication 13 April 1995 and in final form 20 June 1995.

Address reprint requests to Dr. James P. Lewis, Box 871504, Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287-1504. Tel.: 602-965-0667; Fax: 602-965-7954; E-mail: lewis@bardeen.la.asu.edu.

0006-3495/95/09/1068/09 $2.00

Computational methods

In a system of weakly bound molecules, the intramolecular interactions are relatively unchanged as the structure of the system is changed. However, the weak intermolecular interactions are drastically modified by the structure of the system. The intramolecular distances range between 1.0 Å and 1.5 Å, whereas the intermolecular distances are approximately 2.0 Å. Our method, thoroughly developed in Ortega et al. (1994) and summarized in Lewis et al. (Lewis, J. P., O. F. Sankey, and J. Ortega. 1994 Proceedings of the First World Congress on Computational Medicine, Public Health and Biotechnology, submitted for publication), takes advantage of this situation by establishing a fixed Hamiltonian for the isolated molecule. Evaluation of the energetics is greatly simplified by keeping these Hamiltonians constant and by including perturbations to describe the interactions between molecules. The purpose of this simplification is to concentrate on accurately describing the weak interactions between molecules so that differences between large internal energies are avoided (as much as possible) and to make the method computationally fast.

We use a linear combination of atomic orbitals as the basis for the electronic structure. The orbitals on each molecule overlap with its neighboring molecules, but because this intermolecular overlap is small, the effects of overlap are Taylor-series expanded and are treated only to the second order. The interactions between molecules that we include are electrostatic, exchange, and one-electron "hopping" interactions. The one-electron hopping interactions include a repulsive overlap interaction and an attractive rehybridization contribution. In addition, attractive interactions between molecules occur because of electrostatic and exchange contributions.

We write the expectation value of the total energy Hamiltonian for the N molecules system as

$$\langle H_{\mathrm{total}} \rangle = \sum_{\mathrm{i}} \langle H(i) \rangle + \partial U_{\mathrm{N}}^{\mathrm{binding}} \tag{1}$$

The first term, involving $\langle H(i) \rangle$, is the sum of the total energies, within the local density approximation, over all N noninteracting molecules, and the second term, $\partial U_{\mathrm{N}}^{\mathrm{binding}}$, represents the intermolecular interaction between the N molecules. The intramolecular interactions are relatively large in energy (on the order of 1000 eV for a DNA base), but remain constant. The intermolecular interaction is quite small compared with these intramolecular terms (on the order of 1.0 eV or 20 kcal/mol for a DNA basepair) and changes according to the geometry of the system. Because of the large difference of scale between $\langle H(i) \rangle$ and $\partial U_{\mathrm{N}}^{\mathrm{binding}}$, this theory fixes $\langle H(i) \rangle$ and concentrates only on the calculation of the latter term.

The energy of interaction for a system of N molecules is expressed as

$$\partial U_{\mathrm{N}}^{\mathrm{binding}} = \partial E_{\mathrm{N}}^{\mathrm{eigenvalue}} + \sum_{(\mathrm{i,j})} \{[\partial U_{\mathrm{ij}}^{\mathrm{ion\text{-}ion}} - \partial U_{\mathrm{ij}}^{\mathrm{ee}}] - \partial U_{\mathrm{ij}}^{\mathrm{exch}} - \partial U_{\mathrm{ij}}^{\mathrm{vdW}}\} \tag{2}$$

where the sum i, j is over all unique pairs of molecules. The terms $\partial U_{\mathrm{ij}}^{\mathrm{ion\text{-}ion}}$, $\partial U_{\mathrm{ij}}^{\mathrm{ee}}$, $\partial U_{\mathrm{ij}}^{\mathrm{exch}}$, and $\partial U_{\mathrm{ij}}^{\mathrm{vdW}}$ are pairwise, and they describe the nuclear-nuclear, electron-electron, exchange, and van der Waals interactions, respectively. The term $\partial E_{\mathrm{N}}^{\mathrm{eigenvalue}}$ contains the electronic structure information, which is calculated from the occupied eigenvalues of the single particle Hamiltonian, and is obtained from

$$\partial E_{\mathrm{N}}^{\mathrm{eigenvalue}} = 2 \sum_{\mathrm{i\ occ}} E_{\mathrm{i}}(N) - 2 \sum_{\mathrm{j}} \sum_{\mathrm{i\ occ}} E_{\mathrm{i}}(j) \tag{3}$$

This term is the difference of the sum of occupied eigenvalues of the single-particle Hamiltonian corresponding to a system of N interacting molecules and the sum of the occupied eigenvalues of each isolated molecule summed over all the molecules in the system.

The pairwise nuclear-nuclear, electron-electron, and exchange interactions have been developed previously by Ortega et al. (1994). Partial charges, for each atom on the four different DNA bases, were shown to agree well with calculations by others as well as with experimental dipole moments (Lewis, J. P., O. F. Sankey, and J. Ortega. 1994 Proceedings of the First World Congress on Computational Medicine, Public Health and Biotechnology, submitted for publication). These partial charges are used in the electrostatic and exchange interactions. Upon review of calculations by others, we note that van der Waals energies, r^{-6}, sometimes referred to as dispersion energies, may account for almost 50% of the total binding energy (Langlet et al., 1981). In previous calculations of water, we did not include van der Waals energies because for this system it accounts for only ~10% of the binding energy. The fact that van der Waals energies are important in molecules such as DNA bases, and less important in water, is most likely attributable to the more complex charge distributions that make up the former.

Dispersion energies are due to correlation effects and cannot be obtained from a mean field electronic structure theory such as local density approximation, as we use here. A recent review article discusses empirical van der Waals energies based on the Slater-Kirkwood approximation (Halgren, 1992). This approximation is based on a weighted average of the dispersion C_6 coefficients, which is attributable to the polarizability of each individual atom and its effective number of electrons. With the use of this model, we write the van der Waals interactions as

$$\partial U_{\mathrm{mn}}^{\mathrm{vdW}} = \frac{C_{6\mathrm{mn}}}{\mathrm{r}_{\mathrm{mn}}^{6}} \tag{4}$$

where

$$C_{6\mathrm{mn}} = \frac{2\alpha_{\mathrm{m}}\alpha_{\mathrm{n}}C_{6\mathrm{mm}}C_{6\mathrm{nn}}}{\alpha_{\mathrm{m}}^{2}C_{6\mathrm{nn}} + \alpha_{\mathrm{n}}^{2}C_{6\mathrm{mm}}} \tag{5}$$

The subscripts m and n signify atoms on molecule i and j, respectively. The α values are the atomic polarizabilities that were obtained by least-squares fitting to the experimental molecular polarizabilities, assuming additivity. The atomic C_6 coefficients are obtained from the effective electron number and the atomic polarizability. Table 1 shows the values of α and C_6 that were used for H, C, N, and O (Halgren, 1992).

For all intermolecular interactions, a basis set of atomic orbitals of the Slater type were used. These orbitals are of the form (Kern, 1972)

$$\Psi_{\alpha}^{0} = \Psi_{\mathrm{nlm}} = N_{\mathrm{nlm}} \mathrm{r}^{\mathrm{n}-1} Y_{\mathrm{l}}^{\mathrm{m}}(\hat{\mathrm{r}}) e^{-\varsigma \mathrm{r}} \quad (6)$$

where n, l, and m are the appropriate quantum numbers, the functions $Y_{\mathrm{l}}^{\mathrm{m}}(\hat{\mathrm{r}})$ are the spherical harmonics, and N_{nlm} is the normalization constant. The set of Slater type orbitals that we use are the 1s orbitals of hydrogen and the 2s and 2p orbitals of carbon, nitrogen, and oxygen. The decay constants for the 2p orbitals were based on results from Hehre et al. (1969), which lists optimum exponents for several molecules. We note the approximation that the decay constant falls off as the square of the orbital energy, $\varsigma \sim \sqrt{\mathrm{E}}$. Thus $\varsigma_{\mathrm{p}}/\varsigma_{\mathrm{s}} = \sqrt{\mathrm{E}_{\mathrm{p}}/\mathrm{E}_{\mathrm{s}}}$ was used to determine the decay constants for the 2s orbitals, with the energy level values obtained from Harrison (1989). The values chosen for the decay constants are $\varsigma_{\mathrm{s}} = 1.27$ for hydrogen, $\varsigma_{\mathrm{s}} = 2.45$ and $\varsigma_{\mathrm{p}} = 1.75$ for carbon, $\varsigma_{\mathrm{s}} = 2.76$ and $\varsigma_{\mathrm{p}} = 1.95$ for nitrogen, and $\varsigma_{\mathrm{s}} = 3.23$ and $\varsigma_{\mathrm{p}} = 2.25$ for oxygen. For spherical symmetry on each atom, we average the charges in each of the three p-state orbitals to be equal for each atom (e.g., for carbon $\mathrm{n}_{\mathrm{p_x}}^{\mathrm{C}} = \mathrm{n}_{\mathrm{p_y}}^{\mathrm{C}} = \mathrm{n}_{\mathrm{p_z}}^{\mathrm{C}}$).

Forces for each atom at position $\vec{\mathrm{r}}_{\mathrm{i}}$ are determined by taking the analytic derivative of the energy with respect to $\vec{\mathrm{r}}_{\mathrm{i}}$

$$\begin{aligned}\vec{F}_{\mathrm{i}} &= -\frac{\partial E_{\mathrm{tot}}}{\partial \vec{\mathrm{r}}_{\mathrm{i}}} \\ &= -\frac{\partial E_{\mathrm{eigenvalue}}}{\partial \vec{\mathrm{r}}_{\mathrm{i}}} - \left[\frac{\partial U_{\mathrm{nn}}}{\partial \vec{\mathrm{r}}_{\mathrm{i}}} - \frac{\partial U_{\mathrm{ee}}}{\partial \vec{\mathrm{r}}_{\mathrm{i}}}\right] + \frac{\partial U_{\mathrm{exch}}}{\partial \vec{\mathrm{r}}_{\mathrm{i}}} + \frac{\partial U_{\mathrm{vdW}}}{\partial \vec{\mathrm{r}}_{\mathrm{i}}} \quad (7)\end{aligned}$$

The derivative of the electronic term is determined on the basis of the Hellmann-Feynman theorem (Hellmann, 1937; Feynman, 1939). A sixth-order interpolation algorithm obtains the derivatives of the electron single-particle Hamiltonian matrix elements, which exactly includes Pulay corrections, from the precalculated integral data tables. All calculations throughout this paper are performed on a DEC Alpha 3000/600. From the resulting forces, molecular dynamics simulations can be carried out by solving the equations of motion using the leap-frog algorithm for quaternions (Allen and Tildesley, 1987), which is used in a specialized molecular dynamics for rigid molecules. In searching for a minimum energy configuration, a dynamical quenching process is used. In this process the equations of motion are solved, and the resulting kinetic energy is determined. When the kinetic energy reaches a maximum, the velocities are quenched (set to zero), and the process is repeated until a zero force configuration is obtained.

TABLE 1 Atomic polarizabilities and C_6 parameters used in the Slater-Kirkwood approximation for the van der Waals interactions (Halgren, 1992)

Atom	α_1	C_{6ii}
H	2.60	2.8
C	6.38	19.1
N	6.90	22.8
O	5.42	16.8

RESULTS AND DISCUSSION

DNA basepairs

The DNA basepair A-T is shown in Fig. 1 *a*, and the basepair G-C is shown in Fig. 2 *a*. The geometries of the DNA base molecules were kept rigid for simplicity, and the energies were calculated as the separation distance between the two molecules was changed. For the A-T basepair, the energy is plotted as a function of the nitrogen-nitrogen [N1(A)-N3(T)] distance, r_{NN}, whereas for the G-C basepair, the energy is plotted as a function of the oxygen-nitrogen [O6(G)-N4(C)] distance, r_{ON}. The results of these calcula-

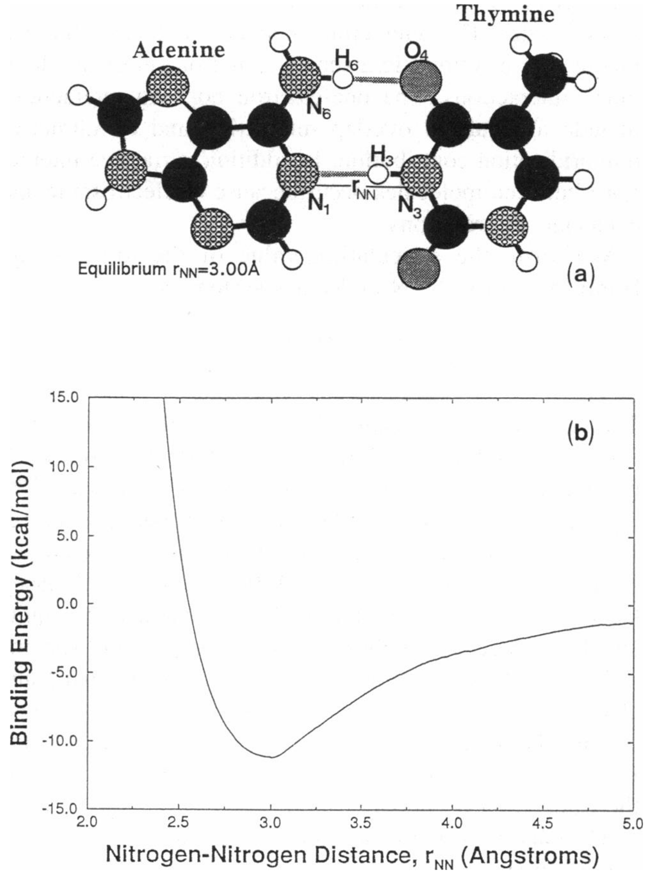

FIGURE 1 (*a*) The adenine-thymine basepair. (*b*) Energy plotted as a function of the nitrogen-nitrogen [N1(A)-N3(T)] distance, r_{NN}, for the A-T basepair. The result of our calculations yields a binding energy of 11.16 kcal/mol with an equilibrium separation distance of 3.00 Å.

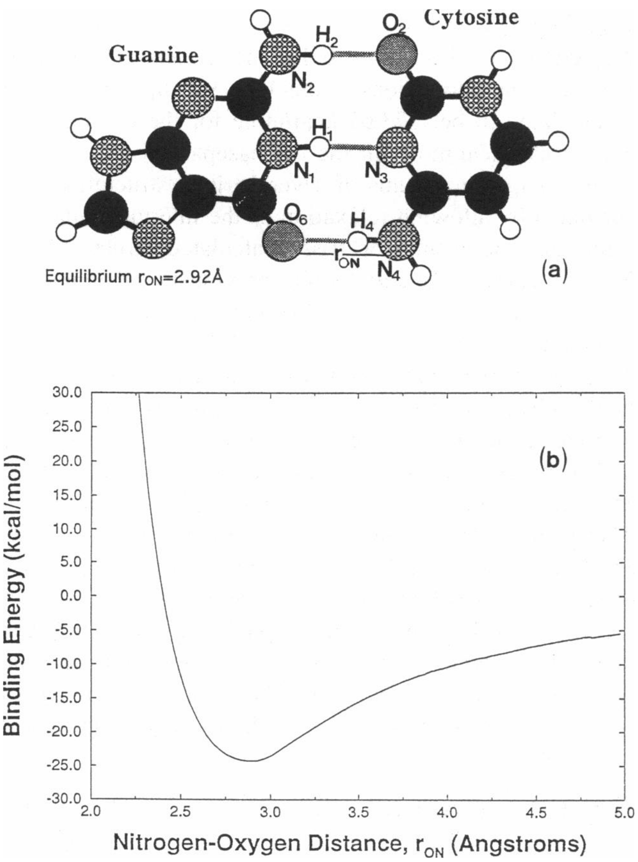

FIGURE 2 (*a*) The guanine-cytosine basepair. (*b*) Energy plotted as a function of the oxygen-nitrogen [O6(G)-N4(C)] distance, r_{ON}, for the G-C basepair. The result of our calculations yields a binding energy of 24.25 kcal/mol with an equilibrium separation distance of 2.90 Å.

tions are shown in Figs. 1 *b* and 2 *b* for the A-T and G-C basepairs, respectively.

We compare our results for the DNA basepairs (Table 2) with results calculated by others on the basis of quantum chemistry (Langlet et al., 1981; Jiang et al., 1994) and classical potentials (Poltev and Shulyupina, 1986; Zhurkin et al., 1994a,b). Also listed are results obtained experimentally (Yanson et al., 1979). For the A-T basepair, we calculate a binding energy of 11.16 kcal/mol with an equilibrium separation distance of 3.00 Å, and for the G-C basepair, we calculate a binding energy of 24.25 kcal/mol with an equilibrium separation distance of 2.90 Å. The binding energies for both DNA basepairs are similar for all the methods listed in Table 2. Generally, theoretical energies are underbound compared with experiment for the A-T basepair and overbound compared with experiment for the G-C basepair. This trend is likely attributable to the fact that the extra hydrogen bond in the G-C leads to too much additional increase in the electrostatic interaction compared with experiment.

TABLE 2 Binding energy results for DNA basepairs

Base pair	This work	Langlet *et al.*	Jiang *et al.* STO-3G	Poltev and Shulyupina	Zhurkin *et al.*	Experiment Yanson *et al.*
A-T	−24.3	−23.0	−26.6	−25.5	−22.8	−21.0
G-C	−11.2	−12.9	−14.1	−11.3	−11.9	−13.0

All energies are listed in units of kcal/mole. For comparison, several results coming from quantum chemistry calculations (Langlet et al., 1981; Jiang et al., 1994), classical potential calculations (Poltev and Shulyupina, 1979; Zhurkin et al., 1994a,b), and experiment (Yanson et al., 1979) are listed.

For small displacements away from the equilibrium position, a harmonic potential was fit to the data from Figs. 1 *b* and 2 *b*. This results in a harmonic spring constant k of 0.601 mdyn/Å for A-T and 0.832 mdyn/Å for G-C. These values correspond to vibrational frequencies of 125.14 cm^{-1} and 148.64 cm^{-1}, respectively. (These frequencies were obtained by using $\omega = \sqrt{k/\mu}$, where μ is the reduced mass.) By way of comparison, a self-consistent field calculation made for the A-T basepair in the minimal basis set (SCF-MINI-1) yields diagonal intermolecular force constants of 0.609 mdyn/Å for the N1(A)-H3(T) stretch mode and 0.396 mdyn/Å for the O4(T)-H6(A) stretch mode (Hrouda et al., 1993). Our results, which are based on rigid molecules, are more comparable to the former mode because in the latter the NH_2 group on adenine contributes an extra degree of flexibility in the O4(T)-H6(A) stretch mode compared with the N1(A)-H3(T) stretch mode.

As a basis of comparison for the A-T base pair, polarized infrared and Raman spectroscopies on the Hoogsteen-type methyladenine-methylthymine1:1 single crystal yields results of 106 cm^{-1} for the base pair stretch mode (Harada and Lord, 1970). In addition, low frequency vibration almodes of polynucleotide shave been experimentally and theoretically examined by Powell et al. (1987). The theoretical method used is that of alattice model of DNA developed by Prohofsky and co-workers (Kimetal., 1986; Lindsay et al., 1984; Meietal., 1981; Eyster and Prohofsky, 1977). For Na·poly (dA)·poly(dT), it is reported that the stretch modes with the large stoscillator strengths are 100 cm^{-1} and 140 cm^{-1}. In addition, the 142 cm^{-1} mode in poly(dG)·poly(dC) corresponds to a stretch modes imilar to the 140 cm^{-1} modeinpoly(dA)·poly(dT). Although we cannot directly compare our results with those of others, overall it seems that our G-C and A-T stretch modes a requite reasonable.

DNA base triplets

Our hydrogen-bonding model seems to work well for water and for DNA basepairs. We now turn to our main results, which concern DNA base triplets. It is commonly accepted that the RecA nucleoprotein filament forms a triple-stranded DNA structure that serves as an intermediate in homologous recombination (West, 1991; Stasiak, 1992; Camerini-Otero and Hsieh, 1993; Kowalczykowski and Eggleston, 1994). As shown by electron microscopy experiments, the RecA substantially extends and underwinds the single-stranded DNA and the duplex DNA strand (Stasiak et al., 1981; Egelman, 1993). In almost all models of putative recombination triplexes, starting with that of Lacks (1966), the third strand lies in the major groove. The extended R-form of DNA accommodates these models, inasmuch as it has an axial rise per base of 5.1 Å compared with 3.4 Å for the B-form of DNA. Also, experiments with modified analogs of bases have recently shown the homologous single-stranded DNA positioned in the major groove of a regular duplex DNA strand (Kim et al., 1992; Rao et al., 1992, 1993). In actuality, these distortions of DNA are extremely energetically unfavorable, but the recombinase protein promotes the three-stranded hybridization by lowering the activation energy that would be required for this structural transition (Camerini-Otero and Hsieh, 1993). As an additional observation, the R-form from the third strand (called R-strand) must be parallel to a strand in the duplex that is identical (denoted as the W-strand). This parallel orientation between the R-strand and the W-strand is a direct consequence of the biological role of the RecA (Zhurkin et al., 1994).

Given the biological evidence of the importance of DNA triplexes, we consider the energetics of several DNA base triplets. Note that studying energetics of DNA base triplet structures does not presuppose complete knowledge of the larger system in question. Factors such as including the backbone into the calculations may modify the final conclusions. These triplets consist of the standard Watson-Crick basepair (WC) interacting with a third base (X), which can be either a purine or a pyrimidine. The notation we use to represent this triplet is (WC):X. Experimental evidence suggests that certain constraints may be imposed on our calculations. Some constraints were proposed and justified earlier (Hsieh et al., 1990) on the basis of the biological significance of DNA base triplets in the recombination process, and additional constraints have also been imposed in previous calculations. On the basis of these, we choose to impose constraints as follows. 1) The third base (X) is bound in the major groove of the duplex. 2) The third base (X) can be either a purine or a pyrimidine. 3) The orientation of the third base (X) is parallel with respect to the duplex (WC). 4) The third base (X) is usually hydrogen bonded to the purine base, but as in previous calculations (Jiang et al., 1994; Zhurkin et al., 1994a,b), we also consider cases in which the third base is hydrogen bonded to both bases in the duplex (WC). The third base (X) is linked to the duplex (WC) by at least two hydrogen bonds. (5) We choose the third base (X) identical to one of the bases in the duplex (WC).

As in the case for the DNA basepair, we complete all calculations keeping all intramolecular coordinates and intramolecular interactions constant. This is valid because for these systems intermolecular interactions determine the major contribution in the binding as the intermolecular coordinates are varied. Both the intramolecular and intermolecular geometry of the duplex (WC) was kept fixed, based on the structure of Arnott (1970), whereas the third base was allowed to move freely. We choose the Arnott structure rather than our energy minimum geometries in order to more directly compare our results with those of others. The interaction energies recorded for the triplets are energies relative to the interactions of this lone WC basepair, which is calculated to be −23.65 kcal/mole for the GC basepair and −10.96 kcal/mole for the AT basepair.

The energy minimums of several triplet structures were calculated by allowing relaxation of the intermolecular coordinates. The results of those calculated structures are listed in Table 3. Energy results previously calculated by others (Jiang et al., 1994) are also shown for comparison. In the notation of Jiang et al. (1994), (r)H denotes the (reverse) Hoogsteen triplet, and m denotes the triplet structure where the third base (X) is located near the dyadic axis and thus is hydrogen bonded with both the purine and the pyrimidine of the WC basepair. Structures prefixed by R signify that these structures are the proposed isomorphic triplets by Zhurkin et al. (1994a,b). The equilibrium geometries for these triplet structures is shown in Figs. 3 *a–j*, where we choose to use a nonstandard shading scheme for the atoms.

Our results show that for the (TA):T and (GC):G triplets, the third base tends to energetically favor locations near the dyadic axis of the WC basepair. The trends from highest to lowest in the binding energies of the calculated structures are as follows:

$$\text{(TA):T(H)} > \text{(TA):T(rH)} > \text{R} - \text{(TA):T(m)} > \text{(TA):T(rm)}$$

and

$$\text{(GC):G(rH)} > R - \text{(GC):G(H)}$$

For the lower energy structures [(TA):T(rm), R-(TA):T(m), and R-(GC):G(H)], the third base is hydrogen bonded to both bases of the WC duplex. Contrarily, the higher energy structures [(TA):T(rH), (TA):T(H), and (GC):G(rH)] have the third base hydrogen bonded to only one base of the WC duplex. Regarding the energetics of the (AT):A triplet group, there is no real difference between the (AT):A(m) and the (AT):T(rm) structures, taking into account thermal fluctuations at kT_o (T_o = 300K). In addition, regarding the energetics of the (CG):C triplet group, calculations per-

TABLE 3 Minimum energy values for various DNA triplet structures

Triplet	This work	Jiang et al.
(AT):A(rm)	−9.75	−10.29
R-(AT):A(m)	−9.47	
(TA):T(rm)	−10.45	
R-(TA):T(m)	−8.77	−5.60
(TA):T(rH)	−7.14	−11.10
(TA):T(H)	−5.94	−11.00
R-(GC):G(H)	−25.84	−17.94
(GC):G(rH)	−24.99	−15.70
R-(CG):C(m)	−17.24	

The energies are listed in units of kcal/mol and are reference energies with respect to the WC basepair energy at the equilibrium structure, as defined by Arnott (1970) (−10.96 kcal/mole for A-T and −23.65 kcal/mole for G-C in this work, and −14.10 kcal/mole for A-T and −26.60 kcal/mole for G-C in Jiang et al.).

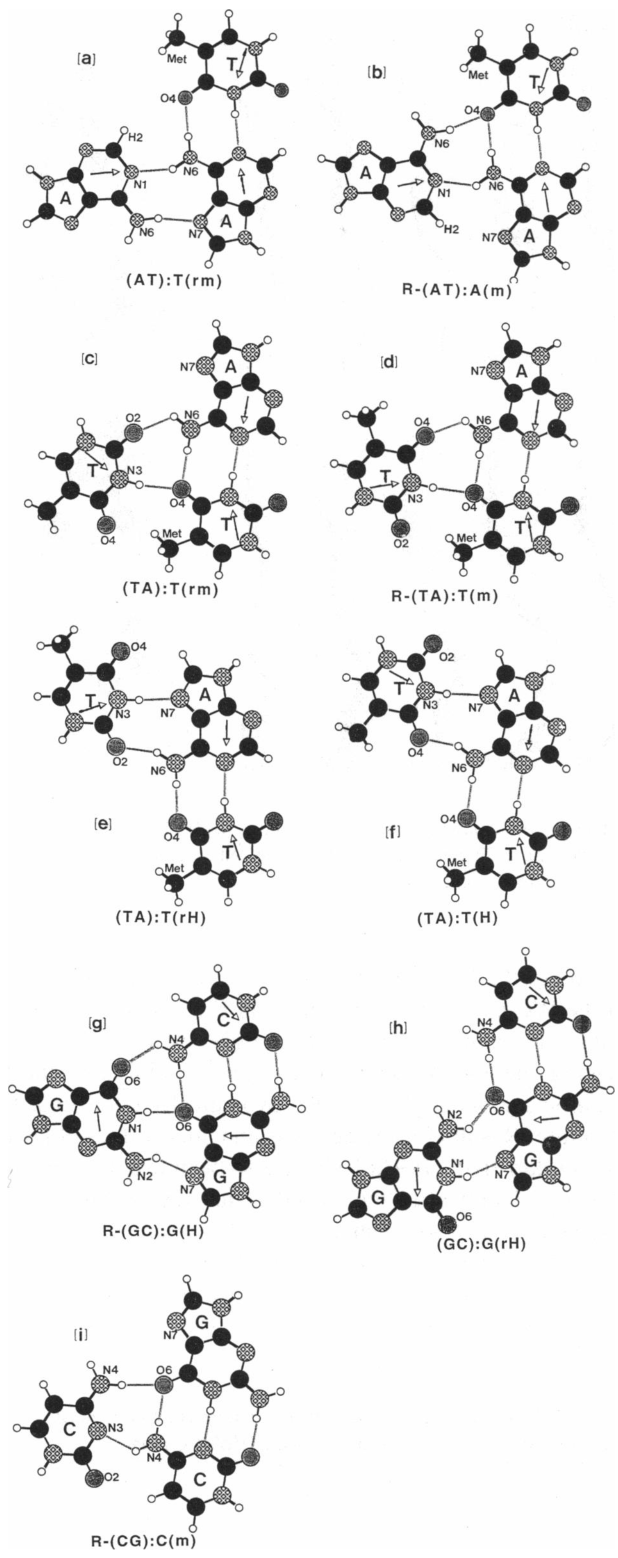

FIGURE 3 Equilibrium configurations for various triplet structures are shown in the following order: (*a*) (AT):T(rm), (*b*) R-(AT):A(m), (*c*) (TA):T(rm), (*d*) R-(TA):T(m), (*e*) (TA):T(rH), (*f*) (TA):T(H), (*g*) R-(GC):G(H), (*h*) (GC):G(rH), and (*i*) R-(CG):C(m). The shading scheme used is not standard but is easy to follow because some of the atoms are labeled. Only hydrogen bonds within the strong range (1.8 Å-2.2 Å) are shown.

formed for the (CG):C$^+$(H) and (CG):C$^+$(rH) structures (not shown) indicate rather large binding energies compared with R-(GC):C(m), which is similar to results obtained by others (Pullman et al., 1967; Jiang et al., 1994).

As an additional investigation of the triplet structures, we study the energetics of the R-triplet structures [R-(AT): A(m), R-(TA):T(m), R-(GC):G(H), R-(CG):C(m)]. We perform calculations for different geometries of the third base in the region near the dyadic axis of the WC basepair. For these calculations, the WC basepair was kept fixed while the R-base was horizontally and vertically translated to determine the potential surface of the third base with respect to the duplex. A series of contour plots, shown in Figs. 4 *a–d*, was created for each triplet structure system. For each figure, five contour plots were generated, each representing different clockwise rotations of the R-base with respect to its identical W-base, given by the angle θ (i.e., $\theta = 0°$ implies that the R-base is not rotated). For these figures, the WC basepair and the contour plot for $\theta = 70°$ are at the correct distance from each other and have the correct scale; the remaining contour plots ($\theta = 110°, 100°, 90°, 80°$) should be overlaid on the $\theta = 70°$ plot for correct scaling with the WC basepair. The position and geometry of the R-base is defined according to the swing, slide, and shift that were discussed in previous publications (Zhurkin et al., 1994a,b). In these contour plots, the solid black region represents a large repulsive barrier, and the circular regions represent areas of minimum energy. The solid white regions represent where the net binding energy, $\partial U^{\text{binding}}$, is positive. For the five contour plots of each triplets structure case, the extremum minimum energy values, labeled with a • in the figures, are shown in Table 4. This illustrates the relation between the energetics and the swing of the R-base. The computational efficiency of our method is demonstrated by the fact that each of these contour plots is made up of 18,000 points requiring approximately 18 h CPU time, or 3.6 s per point.

The main results of these contour plots show that for each geometry of the triplet structures the extremum minimum energy regions are located near the dyadic axis of the WC basepair. This implies that for the geometries based on these angles, and any interpolated angle (70°-100°), the third base finds it more energetically favorable to be hydrogen bonded to both the W-base and the C-base. The trend of the extremum minimum energy values, given in Table 4, is that the global minimum energy for all R-triplet structures is located near where the R-base is rotated clockwise with respect to the W-base by approximately $\theta = 90°$.

For the R-(GC):G and R-(CG):C triplet structures, there is only one extremum minimum energy found in each contour plot [with the exception of $\theta = 110°$ for R-(GC):G]. The regions associated with these minima are very wide compared with the R-(AT):A and R-(TA):T triplet structures. Results from Zhurkin et al. (1994a,b) also show, for values within kT_o ($T_o = 300K$), very wide potential wells for the R-(GC):G and R-(CG):C triplet structures. This indicates that the r-base, guanine or cytosine, has a larger

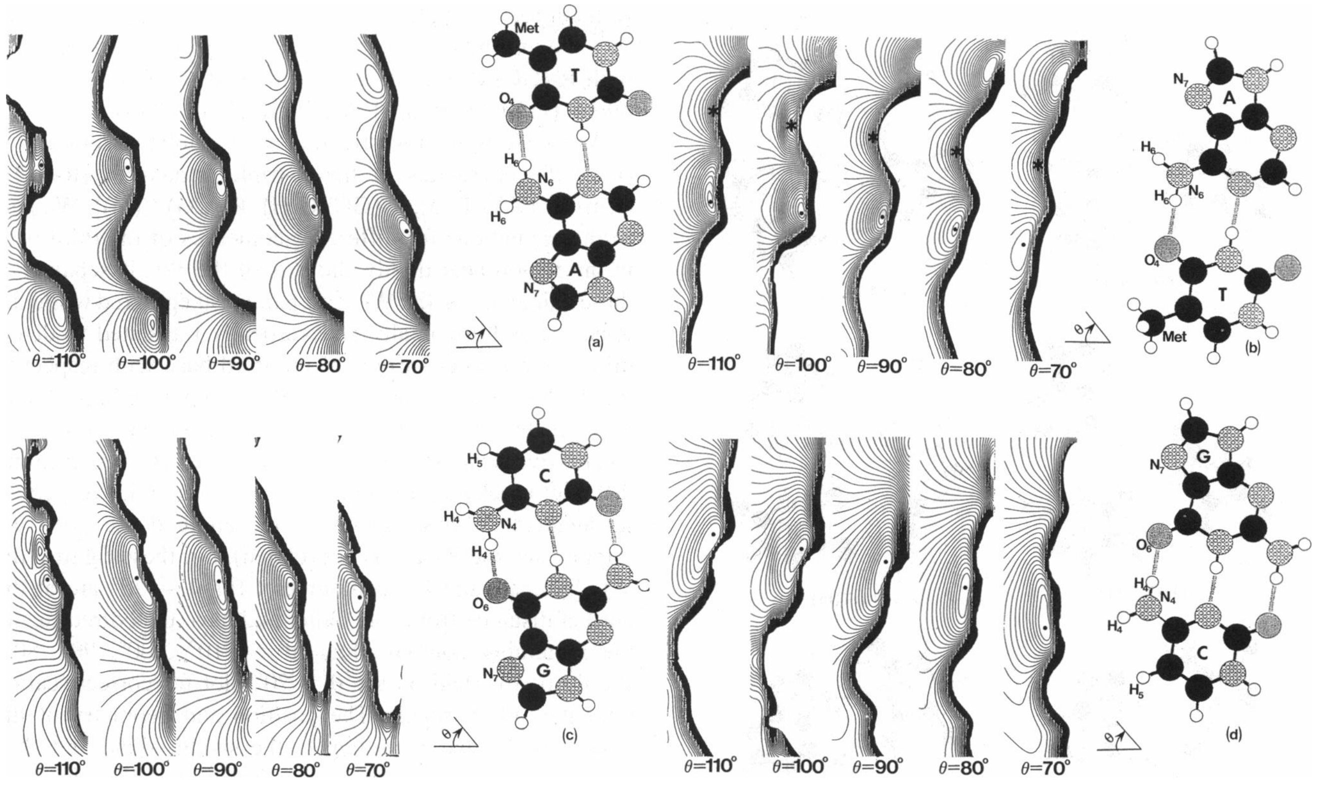

FIGURE 4 Contour plots of the potential surface for the R-triplet structures are shown in the following order: (*a*) R-(AT):A, (*b*) R-(TA):T, (*c*) R-(GC):G, and (*d*) R-(CG):C. The R-base is rotated by a clockwise angle, θ, with respect to the W-base as shown. The R-base is translated horizontally and vertically to create the contour. Points marked as · represent the extremum minimum energy value for the corresponding angle.

region to explore under thermal vibrations. Other minima exist for the R-(AT):A and the R-(TA):T triplet structures, but these are far from the dyadic axis. In addition, these positions would not be structurally favored for the R-strand confined by the recombination protein. Similar to results shown by Zhurkin et al. (1994a,b) for the R-(TA):T triplet structure, there seems to be another minimum energy region, labeled with * in Fig. 4 *b*, which is located near the dyadic axis. This region is less energetically favorable compared with the extremum minimum energy region.

Results of the global minimum energy values, with corresponding angles and slide values of the R-base, are listed in Table 5. The slide parameter is defined as the distance of the R-base from the dyadic axis, measured to be the projection of the N-1(pur)/N-3(pyr) atom onto the vertical axis of the WC basepair, as viewed from Fig. 4 *a–d*. The global minimum configurations and energies were obtained by placing the R-base near the extremum minimum energy location for the angle θ = 90° dynamically quenching the structure. Results for the energies and slide values obtained by Zhurkin et al. (1994a,b) have also been listed in Table 5. We find, identical to their results, that the strength of the interaction in these isomorphic triplets increases with the following order of bases in the R-strand: T, A, C, G.

The results in Table 5 also show that the global minimum energy exists at angles near θ = 90°. The slide results (also

TABLE 4 Listed are extremum energy values as a function of the angle

Triplet	θ = 70°	θ = 80°	θ = 90°	θ = 100°	θ = 110°
(AT):A	−6.52	−7.14	−8.88	−7.09	−4.43
(TA):T	−6.37	−7.16	−8.06	−6.46	−4.98
(GC):G	−20.01	−22.31	−22.70	−18.64	−17.55
(CG):C	−13.13	−13.87	−15.79	−14.14	−12.01

The values are taken from results of the contour plots shown in Fig. 4 (*a*)-(*d*). The energy values, measured in kcal/mole, are the net binding energy below that of the WC pair (i.e., −10.96 kcal/mole for AT and −23.65 kcal/mole for GC).

TABLE 5 Results obtained after performing dynamical quenching of the R base in the four triplet cases

	This work			Zhurkin et al.a,b	
Triplet	E_{min}	Slide Å	Angle θ	E_{min}	Slide Å
(AT):A	−9.27	1.22	87.01°	−9.1	≈1.75
(TA):T	−8.57	1.47	81.95°	−7.6	≈1.00
(GC):G	−25.24	1.01	87.02°	−17.6	≈1.60
(CG):C	−16.64	1.31	85.64°	−13.1	≈1.75

Energies are measured in kcal/mole and represent the net binding energy below that of the WC pair (i.e., −10.96 kcal/mole for AT and −23.65 kcal/mole for GC). The slide distances, in Å, are measured from the dyad axis, defined to be the projection of the N-1(pur)/N-3(pyr) atom onto the horizontal axis of the WC pair (Zhurkin et al., 1994).

shown in Table 5) for each of the triplet structures are all within 1.5 Å from the dyadic axis. Our results also show, as in Zhurkin et al. (1994a,b), that there are two hydrogen bonds for the R-(AT): A, R-(TA): T, and R-(CG): C triplet structures within the strong range (1.8 Å-2.2 Å), as seen in Figs. 3*b*, *d*, and *i*. A third hydrogen bond for these triplet structures exists, but it is greater than the strong range (>2.2 Å). For the R-(GC): G triplet structure, three hydrogen bonds exist, as shown in Fig. 3*g*, within the strong range. These results for the global minimum configurations support, as do the contour plots, that in all these triplet structures the R-base is located very close to the WC basepair and interacts with both the W-base and the C-base.

SUMMARY

We have extended our electronic structure method to handle hydrogen bonding in biological molecules and have shown that the method gives reasonable results for several hydrogen bonded systems. The method is simple enough to handle large systems of hydrogen bonded molecules and is orders of magnitude faster than other ab initio methods. The results of the method for the A-T and the G-C basepairs have been compared with experiment and other calculations, and the results are in excellent agreement.

We find specifically that the R-triplet structures are energetically favorable structures, compared with other similar triplet structures, when a comparison can be made. The potential surface of the R-triplet structures has been determined by using contour plots of the potential surface for various geometries of the R-base. On the basis of these results, we have found that the global minimum energy positions are located near the dyadic axis and that at these positions the R-base is rotated nearly 90° from the W-base. We have studied the energy trends in the triplet structures and the potential surface of the R-triplet structures, and on the basis of these, our results indicate that more energetically favorable structures tend to be ones where the third base is hydrogen bonded to both the W-base and the C-base.

We are currently developing methods to handle large scale problems by using this model for the hydrogen bonding picture along with another model that will handle the strong covalent interactions from first principles. Future work will relax the intramolecular coordinates and will include the backbones into the calculations in order to see how these might further enhance or hinder the energetics of triplet structures. Because the R-triplet structures are energetically favorable structures, addition of this backbone may yield results where these structures are indeed the most energetically favorable structures.

We acknowledge support from the Office of Naval Research, ONR-N00014–90–5–1304. We would also like to thank the following people for stimulating discussions: Alex Demkov, Kristine Dolwick, Matt Grumbach, José Ortega, and Victor Zhurkin.

REFERENCES

Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation in Liquids. Oxford, New York.

Arnott, S. 1970. The geometry of nucleic acids. *Prog. Biophys. Mol. Biol.* 21:265–319.

Camerini-Otero, R. D., and P. Hsieh. 1993. Parallel DNA triplexes, homologous recombination, and other homology-dependent DNA interactions. *Cell.* 73:217–223.

Danilov, V. I., K. M. Zakrzhevska, and N. V. Zheltovski. 1979. in *Itogi Nauki i Techniki. Mol. Biol.* M. V. Volkenstein, editor. USSR. 15: 74–124.

Durland, R. H., D. J. Kessler, S. Gunnell, M. Duvic, B. M. Pettitt, and M. E. Hogan. 1991. Binding of triple helix forming oligonucleotides to sites in gene promoters. *Biochemistry.* 30:9246–9255.

Egelman, E. H. 1993. What do x-ray crystallographic electron microscopic structural studies of the RecA protein tell us about recombination? *Curr. Opin. Struct. Biol.* 3:189–197.

Eyster, J. M., and E. W. Prohofsky. 1977. Soft modes and the structure of the DNA double helix. *Phys. Rev. Lett.* 38:371–373.

Felsenfeld, G., D. R. Davies, and A. Rich. 1957. Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* 79:2023–2024.

Feynman, R. P. 1939. Forces in molecules. *Phys. Rev.* 56:340.

Halgren, T. A. 1992. Representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* 114:7827–7843.

Harada, I., and R. C. Lord. 1970. Low-frequency infrared and Raman spectra of some adenine and uracil crystals. *Spectrochim. Acta.* 26A: 2305–2318.

Harrison, W. A. 1989. Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond. Dover, New York.

Hehre, W. J., R. F. Stewart, and J. A. Pople. 1969. Self-consistent molecular orbital methods. I. Use of gaussian exponents of Slater type atomic orbitals. *J. Chem. Phys.* 51:2657–2663.

Hellmann, H. 1937. Einfuhrung in die Quantumchemic. Franz Duetsche, Leipzig.

Holliday, R. 1964. A mechanism for gene conversion in fungi. *Genet. Res.* 5:282–304.

Hrouda, V., J. Florián, and P. Hobza. 1993. Structure, energetics, and harmonic vibrational spectra of the adenine-thymine and adenine*-thymine* base pairs: gradient non-empirical and semi-empirical study. *J. Phys. Chem.* 97:1542–1557.

Hsieh, P., C. S. Camerini-Otero, and R. D. Camerini-Otero. 1990. Pairing of homologous DNA sequences by proteins: evidence for three-stranded DNA. *Genes & Dev.* 4:1951–1963.

Jiang, S.-P., R. L. Jernigan, K.-L. Ting, J.-L. Syi, and G. Raghunathan. 1994. Stability and cooperativity of nucleic acid base triplets. *J. Biomol. Struct. & Dyn.* 12:383–399.

Kern, C. W., and M. Karplus. 1972. The water molecule. *In* Water: A Comprehensive Treatise; Vol. 1: The Physics and Physical Chemistry of Water. F. Franks, editor. Plenum, New York. 21–91.

Kim, M. G., P. Hsieh, C. S. Camerini-Otero, G. Raghunathan, R. L. Jernigan, V. B. Zhurkin, and R. D. Camerini-Otero. 1992. Recombination protein mediated triplex DNA: R-form DNA. *J. Cell. Biochem.* 16(Suppl. B):36.

Kim, Y., and E. W. Prohofsky. 1986. Defect-mediated hydrogen-bonded instability of poly(dG)-poly(dC). *Phys. Rev. B.* 33:5676–5681.

Kowalczykowski, S. C., and A. K. Eggleston. 1994. Homologous pairing and DNA strand-exchange proteins. *Annu. Rev. Biochem.* 63:991.

Lacks, J. 1966. Integration efficiency and genetic recombination in pneumococcal transformation. *Genetics.* 53:207–235.

Langlet, J., P. Claverie, F. Caron, and J. C. Boeuve. 1981. Interactions between nucleic acid bases in hydrogen bonded and stacked configurations: the role of the molecular charge distribution. *Int. J. Quantum Chem Symp.* 19:299–338.

Lindsay, S. M., J. W. Powell, E. W. Prohofsky, and K. V. Devi-Prasad. 1984. *In* Structure and Dynamics of Nucleic Acids, Proteins and Membranes. E. C. C. H. S. H. Sarma, editor. Adenine, New York. 531–551.

Mei, W. N., M. Kohli, E. W. Prohofsky, and L. L. V. Zandt. 1981. Acoustic modes and non-bonded interaction of the double helix. *Biopolymers.* 20:833–852.

Meselson, M. S., and C. M. Radding. 1975. A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA.* 72:358–361.

Ortega, J., J. P. Lewis, and O. F. Sankey. Simplified electronic structure method for hydrogen-bonded systems: water. *Phys. Rev. B.* 50: 10516–10530.

Pilch, D. S., C. Levenson, and R. H. Shafer. 1991. Structure, stability and thermodynamics of a short intermolecular purine-purine-pyrimidine triple helix. *Biochemistry.* 30:6081–6087.

Poltev, V. I., and N. V. Shulyupina. 1986. Simulation of interactions between nucleic acid bases by refined atom-atom potential functions. *J. Biomol. Struct. & Dyn.* 3:739–765.

Powell, J. W., G. S. Edwards, L. Genzel, F. Kremer, A. Wittlin, W. Kubasek, and W. Peticolas. 1987. Investigation of far-infrared vibrational modes in polynucleotides. *Phys. Rev. A.* 35:3929–3939.

Pullman, B., P. Claverie, and J. Caillet. 1967. Van der Waals-London interactions and the configuration of hydrogen-bonded purine and pyrimidine pairs. *Proc. Natl. Acad. Sci. USA.* 57:1663–1669.

Rao, B. J., B. Burnett, B. Jwang, S. K. Chiu, U. Ezekiel, G. Reddy, and C. M. Radding. 1992. Triplex DNA in recombination mediated by RecA protein. *J. Cell. Biochem.* 16(Suppl. B):13.

Rao, B. J., S. K. Chiu, and C. M. Radding. 1993. Homologous recognition and triplex formation promoted by RecA protein between duplex oligonucleotides and single stranded DNA. *J. Mol. Biol.* 229:328–343.

Rein, R. 1979. *In* Intermolecular Interactions: From Diatomics to Biopolymers. B. Pullman, editor. Wiley-Interscience, New York. 307–363.

Stasiak, A., E. DiCapua, and T. Koller. 1981. The elongation of duplex DNA in complexes with RecA protein. *J. Mol. Biol.* 151:557–564.

Stasiak, A. 1992. Three-stranded DNA structure: is this the secret of DNA homologous recognition? *Mol. Microbiol.* 6:3267–3276.

Szostak, J. W., T. L. Orr-Weaver, and R. J. Rothstein. 1983. The double-strand-break repair model for recombination. *Cell.* 33:25–35.

West, S. C., E. Cassuto, and P. Howard-Flanders. 1982. Post replication repair in E. coli: strand exchange reactions of gapped DNA by RecA protein. *Mol. & Gen. Genet.* 187:209–217.

West, S. C., J. K. Countryman, and P. Howard-Flanders. 1983. Enzymatic formation of biparental figure-eight molecules from plasmid DNA and their resolution. *Cell.* 32:817–829.

West, S. C. 1991. Three-stranded DNA helices as intermediates in genetic recombination. *Bioessays.* 13:37–38.

Yanson, I. K., A. B. Teplitsky, and L. F. Sukhodub. 1979. Experimental studies of molecular interactions between nitrogen bases of nucleic acids. *Biopolymers.* 18:1149–1170.

Zhurkin, V. B., V. I. Poltev, and V. L. Florent'ev. 1981. Atom-atom potential functions for conformational calculations of nucleic acids. *Mol. Biol. (Mosc.).* 14:882–895.

Zhurkin, V. B., G. Raghunathan, N. B. Ulyanov, R. D. Camerini-Otero, and R. L. Jernigan. 1994a. *In* Structural Biology: State of the Art 1993. Proceedings of the Eighth Conversations. R. H. Sarma and M. H. Sarma, editors. Adenine Press, New York. 43–66.

Zhurkin, V. B., G. Raghunathan, N. B. Ulyanov, R. D. Camerini-Otero, and R. L. Jernigan. 1994b. A parallel DNA triplex as a model for the intermediate in homologous recombination. *J. Mol. Biol.* 239: 181–200.