Biophysical Chemistry 8 (1978) 123–134

DEGRADATION OF NUCLEIC ACID IN AQUEOUS SOLUTION BY IONIZING RADIATION. III. THE CORRELATION OF RADIATION DAMAGE WITH CHANGE IN MELTING TRANSITION – MODEL EXPERIMENTS

F. Newton HAYES and Donald E. HOARD
Cellular and Molecular Biology Group, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545, USA

Ulrich HOLLSTEIN
Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131, USA

Walter B. GOAD and Charles DELISI
Theoretical Biology and Biophysics Group, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545, USA

Received 1 November 1978

The melting behavior of polydeoxynucleotide double helices of known structure is analyzed in terms of the thermodynamics of helix stability, taking into account separately those contributions to the transition free energy that are proportional to the numbers of polymer molecules and those that are proportional to the numbers of base pairs formed. From the analysis of the melting transitions of helices having an alternating (d-) A-T, G-C base-pair sequence and containing either single-strand nicks or both nicks and damaged thymine bases, the effects of these structural lesions are assessed; it is concluded that, in a moderately long helix of this sequence (400 base pairs), the initial introduction of one mid-chain double-strand break or single-strand break produces respectively some 3.5 or 4 times as much depression in the transition temperature (T_m) as does the destruction of a single internal A-T base pair.

1. Introduction

A lowering of the characteristic transition temperature (T_m) has been found [1, 2] to be a sensitive indication that DNA has been damaged, but until now it has been possible to correlate quantitatively the extent of damage with change in T_m. With DNA of natural origin, the problem is complicated by the diversity of nucleotide sequences present. On the other hand, model compounds for which analysis of the transition temperature is appropriate have provided valuable insight into a variety of problems. Examples include the physical structure of t-RNA [3, pp. 343–352] and that of a polydeoxynucleotide double helix which contains a thymine photodimer [4]. Here we offer an experimental and theoretical analysis of the transition behavior of two-stranded oligodeoxynucleotides whose known structures model three distinct classes of damage: double-strand breakage, single-strand breakage, and saturation of the 5,6-double bond of isolated thymine base residues. We obtain estimates of the thermodynamic parameters characterizing these three types of structural lesion.

2. Materials and methods

The preparation and characterization of poly d-pT $(pGpT)_x$pG [$d(T\text{-}G)_{N/2}$] and of poly d-pC$(pApC)_x$pA [$d(C\text{-}A)_{N/2}$] have been described previously [5]. Poly d(G-T, T*) was prepared as follows. A solution of poly d-pT$(pGpT)_{19}$pG (2.67 μg-atomic weight organic

P, mean length $N = 39.8$ nucleotides) and OsO_4 (161 μmols) in 16 cm^3 0.01 mol dm^{-3} phosphate, pH 7.0, was incubated at 37°C. A 1 cm^3 sample was withdrawn after 3 hours; incubation of the remainder was continued for a total of 12 hours. Both 3- and 12-hour samples were then dialyzed against deionized water at 18°C for 48 hours. The material from the 3-hour sample remaining within the dialysis cell was concentrated to dryness in vacuo. After dissolution in 0.02 mol dm^{-3} Na^+ aqueous solution and admixture with d(C-A) (mean length $M = 400$ nucleotides), its melting behavior was determined according to the conditions outlined below.

The dialyzed product of the 12-hour oxidation was concentrated to 4 cm^3 in vacuo and applied to a calibrated column (1.2 cm^2 × 178 cm) of Bio-Gel A-5m (BioRad Laboratories). Elution of the column with 0.1 mol dm^{-3} triethylammonium bicarbonate, pH 7.5, resulted in recovery of a single peak of ultraviolet-absorbing material amounting to 2.53 μg-atomic weight esterified P (94.8%); the position of elution corresponded to an average length (N) of 35.1 nucleotides. The product exhibited $\lambda_{max} = 254$ nm, $\epsilon(P) = 7.350$ dm^3 (mg-atom P)$^{-1}$ cm^{-1}, $A_{250}/A_{270} = 1.296$. [The starting material exhibited $\epsilon(P)_{254} = 9.170$, $A_{250}/A_{270} = 1.175$.]

Transition temperatures were determined as described by Hayes et al. [5]; some of the measurements have been previously published. Poly- or oligonucleotides of each complementary base sequence were present at equimolar concentration in terms of P. The mean strand length of one sequence (N nucleotides) was either equal or less than that of its complement (M nucleotides). Melting experiments were uniformly carried out at a sodium ion concentration of 0.02 mol dm^{-3}; at this concentration, T_m values lie in the range 20 to 80° over the range of strand lengths investigated ($N = 8$–400 nucleotides). Prior to measurement of the T_m, buffered solutions of complementary poly- or oligonucleotides were held for a minimum of 15 hours at 4°C to effect equilibration among different helical molecules.

Computations were performed on a CDC 6600 computer at the Los Alamos Scientific Laboratory Central Computing Facility.

3. Thermodynamics of strand association

3.1. Complementary polymers with undamaged bases

From the form of their gel-permeation chromatograms, we estimate that the lengths of our polydeoxynucleotide preparations are distributed about the quoted averages with a dispersion somewhat narrower than Poisson, spanning a few nucleotides in the shortest strands and of the order of ten in the longest. In the melting experiments (table 1), the strands of complementary sequence are either of the same average length, or else the larger average is close to an integral multiple of the smaller. In either case, a variety of duplex species can be formed by pairing complementary polymers of the various lengths that are present, and all of these duplexes will exist to some degree at equilibrium. We shall represent the equilibrium distribution of helices in terms of a single species: that which can be formed by association of one strand equal in length to the average of the longer strands (M) with $n = M/N$ complementary strands of average length N. This duplex species is the one in which maximal pairing of complementary bases occurs and, therefore, is of maximal weight in the equilibrium distribution. In those cases where M and N are equal, the predominant molecules present thus will be intact double helices of various lengths and, from these, the destabilizing effect of the helix ends can be determined. In those experiments in which $N < M$, the association of complementary base sequences will give rise to double helices in which numbers of phosphodiester bonds are absent in one strand. Aside from their being confined to but one of the strands, the resulting single-strand interruptions will resemble the "nicks" that would be produced in the polydeoxynucleotide double helix by the limited action of an endonuclease. These cases, then, provide a means whereby the effects of what we shall term single-strand nicks can be assayed.

At equilibrium, the system free energy is at a minimum and, therefore, stationary; an equivalent form of this condition is that duplexes and their single-strand components be present at such concentrations that the net free-energy change upon dissociation of the equilibrium-average duplex is zero. Using superscripts to distinguish single-strand and double-strand species and subscripts to indicate numbers of nucleotides, we may write:

Table 1
Melting temperatures of double-stranded oligodeoxynucleotides having undamaged base moieties

	N (nucleotides)	M (nucleotides)	C (mol·dm^{-3}) × 10^7	T_m (K)
Sequence:	d(C-A) or d(T-G)	d(C-A) or d(T-G)		
	420	420	1.05	352.1
	154	154	2.90	347.9
	66.5	66.5	6.70	345.2
	36	36	11.94	339.6
	27.3	27.3	17.40	333.3
	20.4	20.4	22.55	326.9
Sequence:	d(C-A)	d(T-G)		
	84	160	3.76	346.3
	36	160	14.75	340.6
	36	160	7.37	339.9
	36	160	0.732	337.5
	32	149	5.69	339.8
	20.8	160	10.95	329.7
	16	500	9.68	324.4
	16	500	9.68	322.9
	11	500	13.93	311.4
	11	160	20.07	309.4
Sequence:	d(T-G)	d(C-A)		
	67	148	4.91	346.5
	39.8	400	4.14	339.9
	32	500	4.98	338.4
	26.6	148	10.51	335.1
	20	148	27.21	332.0
	20	148	12.77	330.2
	20	148	1.65	325.4
	13.2	148	17.20	317.1
	8	148	30.90	293.2

$$\Delta G = 0 = G_{2M}^{DS} - G_M^{SS} - M/N \cdot G_N^{SS} \quad (1)$$

Each free energy term in eq. (1) represents a sum over all possible states of polymer molecules of the appropriate species in the solution. In writing ΔG as a sum of independent terms, we limit ourselves to dilute polymer solutions.

The class of all possible polymer states can be factored into two subclasses. One of these includes those states that involve the relations of nucleotide units to one another – internal states of the polymer; the other comprises the states of motion of each polymer molecule as a whole. From this latter class, a contribution to each term of eq. (1) comes from the sum over all possible translational motions of the polymer molecule as a whole or, equivalently, the sum over all possible locations of its center of mass within the solution volume. Employing the subscript "c" to indicate that the summation is applied to the centers of molecular mass, each such free-energy contribution may be expressed in the form:

$$G_c = RT \ln(K_M C_M), \quad (2)$$

where we have assumed that the free energy of translation depends solely on polymer molecular weight. The invariance of polymer energy with position then makes possible a simple estimate of the factor K_M: there is an effective mass – the polymer mass plus that of a small volume of solvent whose motion must be correlated with the overall translation of the polymer – for

which the energy states corresponding to the center-of-mass motion can be reckoned as approximately those of a free particle. This gives [6]:

$$K_M = e^{-1} \cdot (h^2/2\pi M m_e kT)^{3/2} (6 \times 10^{20}) \text{ molar}^{-1}\text{cm}^3. \quad (2')$$

As before, M represents the polymer length in nucleotides, while m_e represents the effective mass in grams of a single nucleotide residue. [The quantities h and k are Planck's constant and the Boltzmann constant, respectively.]

Another variable strictly separable from the internal states is the orientation of a line passing through the two ends of each polymer molecule. All orientations are energetically equivalent; an argument parallel to the one just made, but this time involving the energy states of a free rotator, gives for the free-energy contribution due to rotation the expression:

$$G_\theta = -RT \ln \left\{ \sum_{l=0}^{\infty} (2l+1) \times \exp[-l(l+1)h^2/2Mm_e\pi^2 \langle R^2 \rangle kT] \right\}, \quad (3)$$

where $\langle R^2 \rangle$ is the mean-square end-to-end distance of the polymer, measured in Ångstrom units. Approximating the sum by an integral yields a closed expression:

$$G_\theta = -RT \ln [2M\pi^2 m_e \langle R^2 \rangle kT/h^2]. \quad (3')$$

For single-stranded polymers, the quantity $\langle R^2 \rangle$ can be taken equal to Mb^2, where b depends on the polymer stiffness. For b, we have adopted the value 7 Å, which is consistent with the stiffness of single-strand ribopolymers determined by Inners and Felsenfield [7]. Over the range of molecular weights involved in our experiments, the double-strand forms act essentially as rigid rods; for them we take $\langle R^2 \rangle$ equal to M^2a^2, assigning a the value 3.5 Å. A separate expression for the rotational free energy is thus necessary for each polymer type. For double-strand forms, for example, we employ the notation:

$$G^{DS} = +RT \ln(\theta_{2M}^{DS}), \quad (3'')$$

where θ_{2M}^{DS} is the reciprocal of the bracketed expression in eq. (3'). A parallel notation is used for single-strand forms.

From the development to follow, it will become clear that the logarithms of K_M, K_{2M}, etc., as well as of the various θ's, occur only in linear combination with two parameters describing entropic effects which occur respectively at double- and single-strand ends. The parameters are evaluated by fitting melting-point data to the thermodynamic equation; obviously one could equally well have determined values for the linear combinations by such a procedure. Thus, the utility of the approximate values we have adopted for the various K's and θ's lies in sharpening the interpretation of the fitted parameters and, for this purpose, order-of-magnitude estimates are very helpful.

All remaining contributions to the overall change in free energy come from changes in the energies of the inter- and intranucleotide states (vibration, torsion, rotation) as duplex molecules are formed. It is these we wish to determine by fitting the T_m data, and to this end we will introduce a minimal set of thermodynamic parameters to describe them. Consider first that the situation of a set of nucleotides comprising the repeating sequence in the interior of a long polymer, either single- or double-stranded, is equivalent to that of its neighboring sets; each set, therefore, will make an equivalent contribution to the free energy. Let the free energy difference per base pair as between the interiors of a long duplex and the two dissociated strands be ΔG_b. We divide ΔG_b into its enthalpic and entropic parts and regard each as a parameter to be determined from the data:

$$\Delta G_b = \Delta H_b - T\Delta S_b. \quad (4)$$

Perhaps it is worth emphasizing that eq. (4) includes all changes that are invariant to rotation and translation of the polymer as a whole, including those owing to changes in solvent states upon duplex formation.

Near the ends of a duplex, the situation is different. The position of the last base pair is stabilized by one instead of two stacking interactions, and it will consequently be free to explore a greater range of its configurational space; the end will be "frayed". A similar effect can be expected at the ends of single strands but, since in the frayed duplex the pairing interaction between complementary bases is also affected, the free-energy contribution due to this base pair will differ from ΔG_b. More specifically, it might be expected that the enthalpy change attributable to this base pair would be less negative than ΔH_b, whereas the corresponding change in entropy would be more positive than ΔS_b. This behavior need not be confined exclusive-

ly to the terminal base pair but may extend with decreasing effect toward the interior of each polymer. Its net effect will be manifest in a dependence of T_m on the proportion of the polymer so perturbed; therefore, we introduce two additional parameters characterizing the perturbation in enthalpy and entropy per duplex end. Because of fraying, the effects of base damage (to be considered later) will differ depending upon whether the damaged base is located near an end or at an interior position; therefore, we need an estimate of the extent of the frayed regions. As our data are unable to furnish an independent measure of this, we roughly (and somewhat arbitrarily) characterize the end perturbation in enthalpy as the equivalent of a number, E, of nucleotide units at each end for which setting ΔH per base pair to zero gives the required perturbation. For each of these E pairs, ΔS is taken as ΔS_e, not necessarily equal to ΔS_b. Thus, per duplex end, we have a perturbation in free energy relative to a model in which end effects are ignored which can be expressed as:

$$\Delta G_e = -E\Delta H_b - T\cdot E(\Delta S_e - \Delta S_b), \qquad (5)$$

ΔS_e and E being regarded as two new parameters which may be varied independently in the fitting process. We characterize the effect of single-strand interruptions in the interior of a duplex in an analogous way. There are $(M/N - 1)$ of these in the representative duplex, and at each we represent the perturbation in ΔG by the equivalent number, I, of base pairs on each side of the interruption for which setting the enthalpy per base pair to zero gives the overall perturbation in enthalpy. Again allowing the entropy change per base pair to differ from ΔS_b, we obtain for the perturbation in ΔG from that of an ideal helix:

$$\Delta G_i = -2I[\Delta H_b - T(\Delta S_i - \Delta S_b)]. \qquad (6)$$

If we insert all of the terms (2) through (6), extract the concentration dependence from (2), and move it to the left side, we obtain the following version of eq. (1):

$$-RT\ln\frac{C_{2M}^{DS}}{C_M^{SS}[C_N^{SS}]^{M/N}} = RT\ln\frac{K_{2M}}{K_M[K_N]^{M/N}} + RT\ln\frac{\theta_{2M}}{\theta_M[\theta_N]^{M/N}} + M(\Delta H_b - T\Delta S_b)$$
$$- 2E[\Delta H_b + T(\Delta S_e - \Delta S_b)]$$
$$- 2(M/N - 1)I[\Delta H_b + T(\Delta S_i - \Delta S_b)]. \qquad (7)$$

The left-hand side of (7) will be recognized as an expression for the chemical equilibrium: $DS_{2M} \rightleftharpoons SS_M + M/N\cdot SS_N$. The right side gives the logarithm of the equilibrium constant as a function of M, N, and T. [It will be convenient to combine the sum of the logarithms involving the various K's and θ's on the right side into a logarithm of their products, as is done below.]

Provided the concentrations of complementary bases are equimolar, at the midpoint of the melting transition the concentrations of the nucleotides present in the various polymeric species obey the relationship $C_{2M}^{DS} = C_M^{SS} = M/N\cdot C_N^{SS}$; moreover, each of the concentrations is equal to C, the concentration of nucleotides present in all the nucleotide chains of length M. Substituting T_m for T in eq. (7) and replacing each of the individual-species concentrations in that equation by C give:

$$0 = [M - 2E - 2(M/N - 1)I][\Delta H_b - T_m\Delta S_b] + T_m[2E\Delta S_e - R\ln C + R\ln(K_{2M}\theta_{2M}/K_M\theta_M)K_N\theta_N] + (M/N - 1)T_m[2I\Delta S_i + R\ln C + R\ln(K_N\theta_N)]. \qquad (7')$$

The first of the right-hand terms of (7') includes the contribution of all interior base pairs; for long, uninterrupted helices ($M \to \infty$, $M = N$), this term dominates and T_m approaches as a limiting value the quotient $\Delta H_b/\Delta S_b$. The bracketed portion of the second of these terms represents end effects that, except for the weak dependence introduced through the various K's and θ's, are independent of M and N. The final term collects those effects due to the presence of single-strand interruptions, which are dependent on the ratio $(M/N)^{-1}$.

3.2. Complementary oligometric sequences, one of which contains damaged thymine moieties

The osmium tetroxide oxidation of oligo d(T-G) may be presumed to have converted a fraction of the thymine to dihydroxydihydrothymine (T*) residues without having affected the guanine bases [8]. Because

no chemical modification of guanine moieties has been introduced, we assume that guanine-cytosine base pairing will not be significantly reduced in the double-stranded complex formed between oligo d(G-T, T*) and oligo d(C-A) while, because of the altered geometry of the oxidized thymine bases and possible electronic effects, the potential for adenine-thymine base pairing will have been diminished in proportion to the fraction of thymine bases that have been oxidized. This fraction is equal to the probability, P_{T*}, that a given thymine residue has undergone oxidation during the course of the osmium tetroxide treatment; it is possible to estimate P_{T*} in an actual experiment by comparing a measured property of the preparation (such as the A_{250}/A_{270} ratio) with that calculated for the distribution of surviving sequences remaining after the oxidation of a given fraction of the thymine bases, distributed at random throughout the molecules of the preparation (see Appendix).

We shall assume that the occurrence of a damaged thymine residue within a frayed region – the E residues at each end of the duplex or the I residues on either side of a single-strand interruption – will have no effect on overall helix stability. The effect of damaging the residue immediately adjacent to one of the frayed regions is postulated to enlarge it by one residue; that is, E and I are increased, on the average, by $1/2P_{T*}$, 1/2 being the probability that the base moiety of the adjacent residue is thymine. The occurrence of a damaged thymine at any other position may be presumed to result in the formation of an internal loop in the helix, the unbonded adenine and T* bases being flanked by intact guanine-cytosine pairs. The enthalpy and entropy of the helix will be perturbed in proportion to the number of such loops; per loop there will occur losses equivalent to one base pair ($-\Delta H_b$ and $-\Delta S_b$) plus further changes which we represent in terms of two additional parameters, ΔH_l and ΔS_l, respectively. The region in which the loops will form extends over a total of $M - 2(E + 1) - 2(M/N - 1)(I + 1)$ nucleotides. To the expressions for overall free-energy change due to helix formation, we must add an additional quantity, ΔG^*, such that:

$$\Delta G^* = 1/2P_{T*}\{-\Delta H_b + T(\Delta S_b - \Delta S_e) + (M/N - 1)[-\Delta H_b + T(\Delta S_b - \Delta S_i)]\}$$

$$+ 1/2P_{T*}[M - 2(E + 1) - 2(M/N - 1)(I + 1)] \times [(\Delta H_l - \Delta H_b) - T(\Delta S_l - \Delta S_b)] . \quad (8)$$

The expression set off by the first set of brackets above reflects the enlargement of the frayed regions brought about by the base damage.

4. Results

The 25 experiments listed in table 1 provide data relating T_m to the strand lengths M and N; we solve eq. (7') for T_m and form a sum-of-squares, Sigma, such that:

$$\text{Sigma} = \sum (T_{m(M,N)}^{\text{exper}} - T_{m(M,N)}^{\text{theor}})^2 .$$

This sum is a function of the six thermodynamic parameters ΔH_b, ΔS_b, ΔS_e, ΔS_i, E, and I, and its minimum has been located by means of a computer code which employs the Levenberg-Marquardt algorithm [9]. A fit almost equally good is found if ΔH_b is fixed at any value between –6 and –12 kcal/mol base pairs and only the other five parameters are varied; when it was fixed at –8 kcal/mol base pairs (a value consistent with several calorimetric determinations [10]), best-fit values for the remaining parameters listed in table 2 were obtained. The rms deviation between experiment and equation, in this case, is just under 1°, while the maximum deviation observed is 2.2°; this appears to be close to experimental uncertainty. (Observe, for example, the 1.5° difference in measured T_m under identical conditions in the thirteenth and fourteenth experiments, table 1). We estimate the uncertainty in the fitted parameters by determining the variation in each that degrades the rms fit by a factor $\sqrt{2}$. More precisely, we fix each parameter at a series of values removed from the best fit, simultaneously adjusting the other parameters in such a way as to optimize the fit under these conditions. The quoted uncertainty in each parameter is then taken as the difference between its best-fit value and a fixed value giving sigma $\sqrt{2}$ times larger than the best-fit sigma. Since the parameters E and ΔS_e are almost entirely determined by the six experiments in which $M = N$, however, their stated uncertainties reflect the sensitivity of the fit to these six experiments alone. As indicated in the table, the

Table 2
Thermodynamic parameters characterizing the association equilibrium of poly d(C-A) with poly d(T-G)

ΔH_b (kcal·mol^{-1}) a, b)	ΔS_b (kcal·deg^{-1} mol^{-1}) b)	E (base pairs) c)	ΔS_e (kcal·deg^{-1} mol^{-1}) b)	$E\cdot\Delta S_e$ (kcal·deg^{-1} mol^{-1}) c)
−8.0	−0.0226 ± 0.0002	$3.6^{+1.0}_{-2.6}$	$0.010^{+0.017}_{-0.003}$	0.026 ± 0.007
		I (base pairs) d)	ΔS_i (kcal·deg^{-1} mol^{-1}) b)	$I\cdot\Delta S_i$ (kcal·deg^{-1} mol^{-1}) d)
		$2.0^{+0.4}_{-0.7}$	$0.019^{+0.006}_{-0.004}$	0.028 ± 0.005

a) Fixed parameter. b) Per mol base pairs. c) Per mol double-helix end. d) Per mol single-strand end at interior sites.

products $E\Delta S_e$ and $I\Delta S_i$ are substantially less uncertain than their separate factors.

It is clear from eq. (7′) that it is the respective sums ΔS_e or ΔS_i plus the logarithms of the products of the K's and θ's that are, in fact, determined by the fitting process. The values listed in table 2 result when the K's and θ's have been calculated taking an effective mass per nucleotide of 5.5×10^{-22} g – the average mass of a bare nucleotide plus one associated Na^+ ion.

To estimate the extent of base damage in the osmium tetroxide-treated polymer, extinctions per g-atomic weight P at 250 and 270 nm were calculated for unchanged nucleotide sequences (see Appendix) by means of eqs. (9) and (10) below. Both equations are based on the nearest-neighbor approximation [11]:

$$\epsilon(P)_{(AB)_x} = \tfrac{1}{2}\left[\epsilon_{AB} + \frac{x-1}{x}(\epsilon_{BA} - \epsilon_A - \epsilon_B)\right], \quad (9)$$

$$\epsilon(P)_{(BA)_xB} = \tfrac{1}{2}\left[\epsilon_{AB} + \epsilon_{BA} - \epsilon_B - \frac{x-1}{x}\epsilon_A\right]. \quad (10)$$

Eq. (9) applies to sequences containing equal numbers of the component monomers A (= d-pT-) and B (= d-pG-), arranged in strictly alternating order. Eq. (10) is appropriate to those sequences containing an odd number of surviving nucleotide residues. In both formulas, x denotes the number of repeating dimers present. Molar extinction coefficients employed in the computations are listed in table 3. The measured A_{250}/A_{270} ratio of the purified product of 12 hours of osmium tetroxide treatment was found to agree with the ratio computed for a probability of oxidation

Table 3
Ultraviolet extinctions of mononucleotides and dinucleotides

Compound	Extinction coefficient (dm^{-3} mol^{-1} cm^{-1}) × 10^{-3} at 250 nm	at 270 nm
d-pT	5.66	9.32
d-pG	13.58	9.64
d-pTpG	19.48 a)	19.12 a)
d-pGpT	18.12 a)	17.28 a)

a) Computed from the least-squares fit of experimentally measured $\epsilon(P)$ values for members of the series d-$(pTpG)_x$ having known x to eq. (9).

of the thymine residues, P_{T*}, of 0.773. Since the product of 3 hours treatment was not as extensively purified, its A_{250}/A_{270} ratio was not felt to be a sufficiently reliable index of the extent of thymine oxidation, and P_{T*} for the sample was calculated instead by assuming that the oxidation of the thymine bases follows first-order (pseudo-unimolecular) reaction kinetics. The mean strand length reported for the 3-hour sample is also an estimate, calculated by assuming that the apparent reduction in mean length produced by longer oxidation was the result of strand breakage that follows first-order kinetics.

Data from the three melting experiments (table 4), when fitted to eq. (7′) augmented by the expression for ΔG^* [eq. (8)], give:

$\Delta H_l = +5.2 \pm 3$ kcal/mol loop,

$\Delta S_l = +0.013 \pm 0.01$ kcal/deg-mol loop.

Table 4
Melting temperatures of double helices formed between oligo d(G-T, T*) and oligo d(C-A)

Oxidation time (hours)	P_{T*}	N (nucleotides)	M	C (mol·dm^{-3}) × 10^7	T_m (K)
0	0.000	39.8	400	4.14	340.4
3	0.310	38.6	400	4.32	334.5
12	0.773	35.1	400	4.64	313.2

The above uncertainties are determined as before. However, we have not attempted to estimate or allow for any uncertainty in P_{T*}. The two errors are highly and positively correlated: for example, at the upper estimate for ΔH_l, 8.2 kcal/mol, we find $\Delta S_l = 0.023 \pm 0.001$ kcal/deg-mol.

5. Discussion

The experiments reported here model three distinct classes of damage to DNA: double-strand scission, single-strand nicks, and base damage. Since in all instances the population of duplexes that are formed will contain a proportion of molecules in which single-strand ends are not in perfect register, the helix ends seem a reasonable approximation of those formed upon the cleavage of a preexisting double helix by an endonuclease. They also seem to constitute a reasonable model for the helix ends produced by radiolytic double-strand scission, even though some portion of the helix ends generated by radiolysis may not be intact nucleotide residues. We also propose that the single-strand interruptions present in the helices formed from complementary strands of disparate lengths, which we have for convenience termed "nicks", can be considered an adequate model for single-strand breaks produced in DNA by ionizing radiation, even though not all such breaks are confined to a single phosphodiester bond.

At least under the conditions we have employed, the thermodynamics of the double helix-random coil equilibrium can be adequately represented by simply distinguishing those contributions to the transition free energy that are proportional to the number of base pairs formed from those that are proportional to the numbers of polymer molecules involved; this forms the basis of our analysis. There are two kinds of the second class of contribution: those owing to end effects and those owing to a loss of translational and rotational freedom (on the part of polymer molecules as a whole) when, upon association, two free molecules are replaced by one. Both effects contribute to a change in entropy of the system but, since each association involves exactly two ends and the removal of one free molecule, it is only the sum of the two effects that can be extracted from experiment. Because it seems desirable to assess separately the significance of the first of them, however, we have independently estimated the translational-rotational component by invoking the invariance of the quantum states of the polymer (and closely associated solvent) with respect to position and orientation in the solution volume. The result is an entropy term, $RT\ln(K\theta \cdot C)$, where C is the polymer concentration. Over the range of polymer lengths covered here, the product $K\theta$ takes on values of the order 10^{-16} to 10^{-18} molar^{-1}. We have found the end effects in polymers of the sizes studied to be such that an order-of-magnitude uncertainty in the product $(K\theta)$ leads to an uncertainty of less than 10% in the respective values computed for $2E\Delta S_e$ and $2I\Delta S_i$. We believe the estimates of $(K\theta)$ derived from the free-particle approximation to be accurate to well within a single order of magnitude.

In the notation of Zimm and Bragg [12], the helix-coil equilibrium constant is written:

$$K' = \beta s^N . \tag{11}$$

In their language, s represents the equilibrium constant for adding one base pair to a preexisting double helix, while the product βs represents the equilibrium constant for the association of the initial base pair. Comparing this equation with eq. (7), under conditions $M = N$ and $T = T_m$, gives the correspondences:

$$-RT_m \ln s = \Delta H_b - T_m \Delta S_b , \tag{12}$$

$$\beta = \frac{(K_M \theta_M^{SS})^2}{K_{2M} \theta_{2M}^{DS}} \exp\left[\frac{2E[\Delta H_b + T_m(\Delta S_e - \Delta S_b)]}{RT_m}\right]. \quad (13)$$

With the values for effective mass and $\langle R^2 \rangle$ that we have used, the first factor in eq. (13) becomes $(K_M \theta_M^{SS})^2 / K_{2M} \theta_{2M}^{DS} = 9.2 \times 10^{-8}\, T^{-3/2}\, M^{-3/2}$ molar^{-1}; when we evaluate β at $M = N = 20$, 66, or 420, we find values of 2.4×10^{-3}, 1.5×10^{-2}, or 7.2×10^{-4} molar^{-1}, respectively. These approximate the values reported for both ribo- and deoxyribonucleotide polymers by Breslauer et al. [13] and by Scheffler et al. [14]. It must be noted that, while "association of the first base pair" is an attractive metaphor for the process characterized by β, the value of this quantity depends on a much more complex relationship which collects all free-energy differences not proportional to the number of base pairs in the ultimate helix. We shall return to this shortly.

First, let us remark regarding the end effects on enthalpy. The enthalpy change driving double-helix formation falls short of the value predicted by a model in which end effects are ignored by the quantity $-2E\Delta H_b$ – approximately 60 kcal/mol of double-strand polymer. Breslauer et al. [13] have determined both the calorimetric enthalpy change accompanying the melting of the self-complementary oligoribonucleotide r-A_7U_7 and the Van't Hoff enthalpy, the latter from various ways of interpreting the melting profile. They find little or no deficit in the calorimetric value ascribable to finite polymer length, whereas the Van't Hoff value shows a deficit in the neighborhood of 30 kcal/mol oligonucleotide. The calorimeter, of course, measures the total heat required to complete the transition, whereas the Van't Hoff ΔH represents a net energy difference between single strands and double-stranded molecules which, at T_m, is averaged over a population containing the two forms in many conformations. In these terms, the end-effect enthalpy change $2E\Delta H_b$ determined by us is clearly a Van't Hoff enthalpy, manifesting the fact that, in many members of such a population, attractive forces stabilizing helical molecules are not fully realized. The value found is thus comparable to the enthalpy deficit found by Breslauer et al. for the ribooligometer. We also find a compensating entropy end-effect $2E(\Delta S_e - \Delta S_b)$, amounting to some 0.22 kcal/mol-degree. It is natural to attribute both effects to fraying at the ends of the helix extending over a region of approximately E base pairs. Over this region, the loss of stabilizing enthalpy is accompanied by an increase in configurational freedom which contributes the end-effect entropy change.

We do not propose, of course, that fraying is actually limited to some definite number of nucleotides in a single "frayed" state but, rather, that there is a decrease in the degree of flexibility displayed by the two strands as one proceeds from the end of the helix centerward. Thus, the quantities characterizing the ends which define β [eq. (13)] are related to regions of the helix even longer than E nucleotide pairs. This again emphasizes the complexity of effects influencing the parameter β.

It may be noted that the enthalpy loss associated with a single-strand nick, $I\Delta H_b$ (table 2), is about half that for a double-strand end, while the corresponding entropy increase, $I(\Delta S_i - \Delta S_b)$, is, if anything, somewhat greater. It is tempting to see in this an indication of greater flexibility in the double helix at nicked sites. At both the ends and nicked sites, the single-strand ends are stabilized by entropy increases, $E\Delta S_e$ and $I\Delta S_i$, respectively, both positive. These entropy changes are in marked contrast to the negative change (ΔS_b) which obtains for base pairs located away from an end or nick. In our formulation, all contributions to the overall entropy change owing to translational-rotational freedom of the polymer molecules as a whole have been subtracted out; if the base pairs (and associated solvent) of a double-stranded molecule had precisely the same freedom as the bases of single strands, the remaining net entropy change would be zero. The meaning of the set ΔS_e, ΔS_i, and ΔS_b is thus that, at the ends and sites of nicks, the combined flexibility of both strands plus solvent (or, more precisely, the combined measure of configurational space visited) is greater than that of the ends of the single strands while, in the helix interior, it is less.

In the helices containing osm m tetroxide-oxidized thymine moieties, the change that is determined from the shift in T_m is the overall difference in free energy resulting from the destruction of the complementarity of A·T base pairs. To estimate a free energy of internal-loop formation for comparison with, for example, values reported for ribooligometers by Gralla and Crothers [15], we have corrected this overall difference by what amounts to the average of the pairing and stacking interactions of one A·T and of one G·C

base pair per loop. Because only A·T pairs have been disrupted in the oxidation experiment, however, this procedure incorporates an approximation not present in Gralla and Crothers' result. Of the structures investigated by them, the case most closely comparable to ours is that of the loop produced by two opposed cytosines flanked by G-C pairs; if we include the uncertainty in the stacking free energy [16] by which they correct the overall free-energy change at 25°C, their result becomes + 0.1 ± 0.7 kcal/mol. From ΔH_l and ΔS_l with their correlated uncertainties, we find a value of 1.3 ± 0.3 kcal/mol at the same temperature. This can hardly be considered disagreement, particularly since, in addition to the approximation cited above, other significant differences in experimental conditions exist. The effects of the substitution of deoxyribosyl- for ribosyl-moieties in the structures have not been taken into account nor the differences in the ways in which the loops arise. In our case, internal-loop formation is the result of the introduction of nonplanar base moieties into the structure. Gralla and Crothers found no evidence that the heat of closure for their internal loops differed significantly from zero; our calculated value of + 5.2 kcal/mol is within the stated margin of error in their calculations.

We have throughout taken all of the enthalpic and entropic parameters to be independent of temperature. Were the true thermodynamic quantities to vary with T as, for example, is suggested for ΔH_b in ribopolymers in the data collected by Bloomfield et al. [3, pp. 313–314], the linear term of a variation of ΔH would show up as a contribution to the corresponding ΔS in our model. A higher order dependence of ΔH on T, or a separate dependence of ΔS on T, would manifest itself by a need for terms at least quadratic in T_m in eq. (7'). Within the precision of our experimental results, however, such modifications do not appear to be required.

Finally, we can note the relative degree of destabilization of double-helical structures contributed by the three kinds of damage. Model computations employing the complete equation indicate that, for a fairly long polymer with strictly alternating base-pair sequence, for example, one with $M = N = 400$, the first double-strand break and the first single-strand break would produce, respectively, some 3.5 and 4.0 times as great a depression in the T_m as would damage to a single adenine or thymine base located at some distance from the ends. Although the effect of either kind of break has the same order of magnitude, the single-strand nick is slightly more destabilizing. This is because there is less compensating gain in entropy accompanying the loss of enthalpy when one strand is broken than when both are, even though the overall loss in enthalpy and gain in entropy are both greater in the case of the double-strand break. Our experimental measurements do not include a case of damage to one of the G·C base pairs; presumably if such damage leads to the formation of internal loops bounded by A·T base pairs, there would be greater reduction in helix stability than in the case of base damage we have examined. To correlate the damage of the various types produced by ionizing radiation in nucleic acids with an observed depression in T_m, the relative frequencies of occurrence of each type must also be established with models of known sequence. The assessment of the strand breaks and base damage produced in x-irradiated single-strand oligodeoxynucleotides in solution will be the subject of a future publication [17].

Acknowledgement

We are indebted to Mrs. Victoria E. Mitchell for technical assistance. This work was performed under the auspices of the United States Energy Research and Development Administration.

Appendix

To extend the basic thermodynamic analysis embodied in eqs. (1), (7), or (7') to double-strand polynucleotides in which a fraction of the bases have sustained damage, it is necessary to know the average numbers of damaged nucleotide sequences corresponding to given levels of total base damage. We determine probable values for such sequences as follows.

Consider first a polymer whose molecules, on the average, contain N monomer units, each of which has an independent probability, P_{B*}, of having been hit and damaged during the course of an experiment and a corresponding probability, $(1 - P_{B*})$, of having escaped. The number of damaged base moieties averaged over the total population of polymer molecules will then be $Q_{total} = P_{B*}N$. However, in individual mole-

cules, the damaged bases will be distributed in many arrangements of singly occurring, paired, triple, ... n-tuple sequences.

A sequence of n damaged units separated from the ends of the molecule by an undamaged sequence, termed a damaged interior n-tuple sequence, is to be distinguished from one of equal length located at an end (damaged end n-tuple). In the chain of N units, there are $N - n - 1$ positions or sites that can be occupied by the interior n-tuple sequence. For a damaged end n-tuple sequence, there are but two possible sites per chain when $n < N$ and only one if $n = N$.

The chains which contain a damaged n-tuple sequence at a particular site include, outside the site itself, all possible arrangements of hits and escapes. For each class of damaged sequence, there is a unique probability $(1 - P_{B^*})^2 (P_{B^*})^n$ for an interior n-tuple; $(1 - P_{B^*})(P_{B^*})^n$ for an end n-tuple when $n < N$; and $(P_{B^*})^N$ when $n = N$.

The classes of molecules having a damaged n-tuple sequence at a particular site divide the set of all molecules into mutually exclusive subsets, and this simplifies the calculation of the average numbers of each kind of damaged sequence. The average numbers are the products of the probability for the occurrence of a given class and the numbers of sites available to members of that class [eqs. (A.1) through (A.3)]:

$$Q_n^{\text{interior}} = (1 - P_{B^*})^2 (P_{B^*})^n (N - n - 1) \qquad \text{(A.1)}$$

$$Q_n^{\text{end},\, n<N} = (1 - P_{B^*})(P_{B^*})^n (2) \qquad \text{(A.2)}$$

$$Q_n^{\text{end},\, n=N} = (P_{B^*})^N . \qquad \text{(A.3)}$$

For bases which have escaped damage, average numbers of the various types of n-tuple sequence are obtained by exchanging the quantities P_{B^*} and $(1 - P_{B^*})$ in the above three formulas.

The results embodied in eqs. (A.1), (A.2), and (A.3) are easily extended to the case where the oligonucleotide targets contain pyrimidine and purine base components in strictly alternating sequence. We now take into consideration two independent probabilities for damage, P_{Py^*} and P_{Pu^*}, applying to the pyrimidine and purine base components of the target, respectively. One must also distinguish between n-tuple sequences with respect to the type of base (pyrimidine or purine) present at the 5′-position and as to whether n represents an even or an odd number of nucleotide units. As has been indicated previously, we feel justified in assuming that in essentially all molecules of both the $d(T\text{-}G)_{N/2}$ and $d(C\text{-}A)_{N/2}$ preparations there were initially at the 5′-ends a pyrimidine nucleotide and at the 3′-ends a purine nucleotide residue. [Thus, all molecules initially contained even numbers of nucleotides.] Under these circumstances, the average numbers of damaged sequences, arranged by class, are:

(a) n is odd; a pyrimidine nucleotide is at the 5′-end

$$Q_n^{\text{interior}} = \tfrac{1}{2}(N - n - 1) \times (P_{Py^*})^{(n+1)/2} (P_{Pu^*})^{(n-1)/2} (1 - P_{Pu^*})^2 , \qquad \text{(A.4)}$$

$$Q_n^{\text{end}} = (P_{Py^*})^{(n+1)/2} (P_{Pu^*})^{(n-1)/2} (1 - P_{Pu^*}) . \qquad \text{(A.5)}$$

(b) n is odd; a purine nucleotide is at the 5′ end

$$Q_n^{\text{interior}} = \tfrac{1}{2}(N - n - 1) \times (P_{Py^*})^{(n-1)/2} (P_{Pu^*})^{(n+1)/2} (1 - P_{Py^*})^2 , \qquad \text{(A.6)}$$

$$Q_n^{\text{end}} = (P_{Py^*})^{(n-1)/2} (P_{Pu^*})^{(n+1)/2} (1 - P_{Py^*}) . \qquad \text{(A.7)}$$

(c) n is even; a pyrimidine nucleotide is at the 5′-end

$$Q_n^{\text{interior}} = \tfrac{1}{2}(N - n - 2) \times (P_{Py^*})^{n/2} (P_{Pu^*})^{n/2} (1 - P_{Py^*})(1 - P_{Pu^*}) , \qquad \text{(A.8)}$$

$$Q_n^{\text{end},\, n<N} = (P_{Py^*})^{n/2} (P_{Pu^*})^{n/2} [(1 - P_{Py^*}) + (1 - P_{Pu^*})] . \qquad \text{(A.9)}$$

(d) n is even; a purine nucleotide is at the 5′-end

$$Q_n^{\text{interior}} = \tfrac{1}{2}(N - n) \times (P_{Py^*})^{n/2} (P_{Pu^*})^{n/2} (1 - P_{Py^*})(1 - P_{Pu^*}) , \qquad \text{(A.10)}$$

$$Q_n^{\text{end}} = 0 . \qquad \text{(A.11)}$$

(e) $n = N$

$$Q_n = (P_{Py^*})^{N/2} (P_{Pu^*})^{N/2} . \qquad \text{(A.12)}$$

Formulas for calculating average numbers of undamaged sequences of each type are again obtained by exchanging the respective qualities $(1 - P_{Py^*})$ or $(1 - P_{Pu^*})$ with P_{Py^*} or P_{Pu^*} and vice versa in eqs. (A.4) through (A.12).

For the model experiment involving partial chemical oxidation of the thymine residues of oligo d(T-G), enumeration of the surviving sequences makes possible

the deduction of P_{T*}, the probability that thymine was oxidized. In this special case, P_{G*} ($= P_{Pu*}$) = 0; the equations, modified to predict numbers of surviving sequences, indicate nonzero numbers for only four types. In addition to molecules in which no thymine has been oxidized, there are interior sequences having an odd number of nucleotides and a 5′-terminal deoxyguanylate residue, sequences shorter than the original molecules coinciding with the original 3′-ends (which also have an odd number of nucleotides and 5′-terminal deoxyguanylate), and sequences shorter than the original strands which overlap the 5′-terminal thymidylate residue and have even numbers of nucleotides. The formulas giving numbers of each type of surviving sequence are:

$$Q_n^{\text{interior}} = \tfrac{1}{2}(N - n - 1)(1 - P_{T*})^{(n-1)/2}(P_{T*})^2 , \quad \text{(A.13)}$$

$$Q_n^{3'\text{-end}} = \tfrac{1}{2}(1 - P_{T*})^{(n-1)/2}(P_{T*}) , \quad \text{(A.14)}$$

$$Q_n^{5'\text{-end}, n<N} = (1 - P_{T*})^{n/2} , \quad \text{(A.15)}$$

$$Q_n^{n=N} = (1 - P_{T*})^{N/2} . \quad \text{(A.16)}$$

References

[1] U. Hagen and R. Wild, Strahlentherapie 124 (1964) 275.
[2] E. Jellum, Int. J. Radiat. Biol. 10 (1966) 577.
[3] V.A. Bloomfield, D.M. Crothers and I. Tinoco, Jr., eds. Physical Chemistry of Nucleic Acids (Harper and Row, New York, 1974).
[4] F.N. Hayes, D.L. Williams, R.L. Ratliff, A.J. Varghese and C.S. Rupert, J. Am. Chem. Soc. 93 (1971) 4940.
[5] F.N. Hayes, E.H. Lilly, R.L. Ratliff, D.A. Smith and D.L. Williams, Biopolymers 9 (1970) 1105.
[6] R.E. Dickerson, in: Molecular thermodynamics (Benjamin, New York, 1969) p. 287.
[7] L.D. Inners and G. Felsenfeld, J. Mol. Biol. 50 (1970) 373.
[8] M. Beer, S. Stern, D. Carmalt and K.B. Mohlenrich, Biochemistry 5 (1966) 2283.
[9] Program E401, Nonlinear least squares optimization code, Los Alamos Scientific Laboratory Program Library, Los Alamos, New Mexico 87545.
[10] H. Klump and T. Ackermann, Biopolymers 10 (1971) 513.
[11] D.M. Gray, private communication (1976).
[12] B.H. Zimm and J.K. Bragg, J. Chem. Phys. (1959) 526.
[13] K.J. Breslauer, J.H. Sturtevant and I. Tinoco, J. Mol. Biol. 99 (1975) 549.
[14] I.E. Scheffler, E.L. Elson and R.L. Baldwin, J. Mol. Biol. 57 (1970) 145.
[15] J. Gralla and D.M. Crothers, J. Mol. Biol. 78 (1973) 301.
[16] J. Gralla and D.M. Crothers, J. Mol. Biol. 73 (1973) 497.
[17] F.N. Hayes, U. Hollstein, W.B. Goad, D.E. Hoard and C. DeLisi, in preparation.