Chem. Soc. Rev., 1989, **18**, 53—83

Structural Isomerization in DNA: The Formation of Cruciform Structures in Supercoiled DNA Molecules

By David M. J. Lilley
DEPARTMENT OF BIOCHEMISTRY, THE UNIVERSITY, DUNDEE DD1 4HN

1 The Structure of DNA

As the repository of genetic information for all life-forms on this planet for billions of years, DNA is arguably the most important macromolecule in the cell. The field of molecular biology has undergone explosive expansion in the past 30 years, much of which can be traced back to the elucidation of the structure of DNA; the heuristic value of this model has been of incalculable value in the development of molecular genetics. Thus the potential importance of structural studies on DNA is hard to overemphasize.

While the general architecture of DNA has been known since the fifties[1] – a right-handed double helix with hydrogen bonding between bases to generate sequence-specific base-pairing – the challenge of the last decade has been twofold. First to go into the structure in greater depth, asking how local sequence can alter the structure at the level of the single base-pair, and second to understand how certain proteins recognize and bind specific sequences contained within long DNA molecules. These two questions are probably closely related. Very significant progress has been made in answering both questions, and I shall attempt to review the progress made in one aspect of the former.

2 Structural Variation in DNA

It has long been known from *X*-ray fibre diffraction studies that DNA can exhibit structural polymorphism as a function of base composition and sequence, ionic composition, and water activity. Two main right-handed helical structures (or, more accurately, families of structures) were described,[2,3] the A- and B-structures depicted in Figure 1. Differences between the two are obvious to a glance. The B-structure is relatively slim, with two grooves of about equal depth (which is to say that the helix axis is located at the centre of the structure), where the planes of the aromatic base-pairs lie approximately normal to the axis. By contrast, the A-structure is more wide and squat, with grooves of unequal depth and a pronounced tilt of the base-pairs relative to the helix axis. In general the B-structure is favoured at high water activity, and is therefore the more important structure in solution, and thus to biology.

The most important interactions holding the structures together are probably

[1] J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
[2] W. Fuller, M. F. H. Wilkins, H. R. Wilson, and L. D. Hamilton, *J. Mol. Biol.*, 1965, **12**, 60—80.
[3] R. Langridge, D. A. Marvin, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. F. H. Wilkins, and L. D. Hamilton, *J. Mol. Biol.*, 1965, **2**, 38—64.

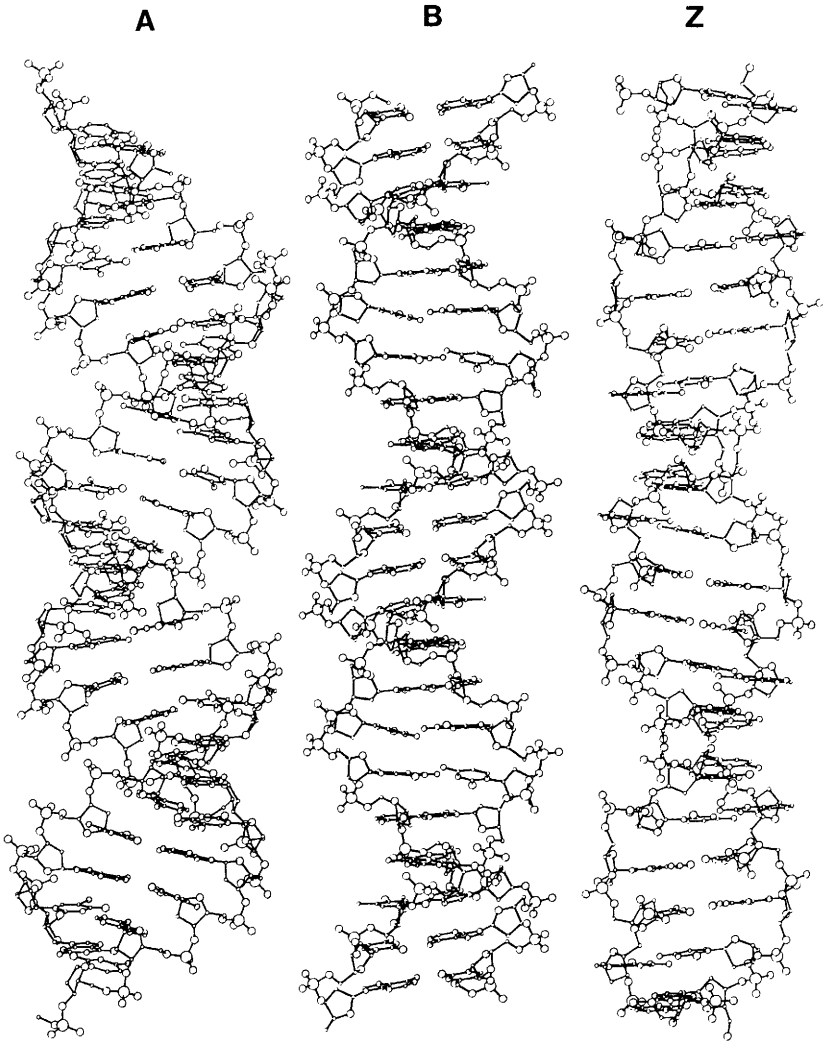

Figure 1 *Helical variants of DNA structure. Two major families of right-handed DNA are known (A and B), while Z-DNA is left-handed. Comparing the two right-handed conformations, we see that the A-form has a pronounced tilt, leading to a wider, squater structure compared to the B-form. The base-pairs of the A-form exhibit a pronounced displacement from the helical axis, generating a hole down the centre of the structure, while in B-DNA the axis passes through the centre of the base-pairs. In general the B-structure is more hydrated, and lowering water activity favours the A-structure. The structures are also sequence-dependent, with GC-rich sequences tending to adopt the A-conformation. Z-DNA is a left-handed helix based on a CpG dinucleotide repeating unit. It is formed by alternating purine–pyrimidine sequences, particularly at high salt concentrations. These pictures were adapted from Dickerson*[4] *with permission*

[4] R. E. Dickerson, *Sci. Am.*, 1983., **249**, 94 111.

those between the base-pairs, the stacking. Thus a model of the base-pairs alone may provide a good description of a given DNA structure,[5] in which the important parameters are base tilt (rotation about the short axis of the base-pair), roll (rotation about the long axis of the base-pair) and slide (translation along the long axis). In older literature the conformation of the deoxyribose phosphate backbone is emphasized, particularly that of the sugar pucker with the A-structure based on C_3'-*endo* conformation and the B structure based on C_2'-*endo* conformation, but this has been replaced with a view in which the energetic preferences of the base-pair interactions are accommodated by the backbone.

The advent of single-crystal *X*-ray diffraction analysis of oligonucleotides of defined base-sequence at near atomic resolution has provided a rich new source of data on DNA conformation at the level of the single nucleotide, and revealed that DNA can indeed be structurally quite adventurous. Dickerson and colleagues solved the first structure of a piece of B-DNA,[6] of self-complementary sequence CGCGAATTCGCG. One great lesson from this structure has been that at this level of resolution DNA exhibits considerable microheterogeneity in structure,[7] in which torsion angles may vary considerably from nucleotide to nucleotide. Calladine has had some success in accounting for these variations based on cross-strand steric clash between purines of successive base-pairs[8] caused by the propeller twist of base-pairs (the deviation between the planes of the two bases comprising a given base-pair), and this is a demonstration of the importance of the base-pair interactions in determining the final structure. Dickerson has recently tabulated 35 separate crystal structure analyses which have been performed on synthetic oligonucleotides,[9] which shows the wealth of structural data now available. Unfortunately, relatively few of these are B-DNA, a consequence of the low water activity inherent in the crystallization process.

One major surprise which emerged from the early crystallographic studies was that the handedness of DNA is not immutable. Rich, Wang, and colleagues solved the structure of CGCGCG and showed that the crystal contained a new left-handed form of DNA, which they called Z-DNA.[10] This is based on a repeating dinucleotide unit in which the bases are paired with conventional Watson–Crick hydrogen bonding, but where the conformation about the guanine glycosyl bond is *syn* and the corresponding deoxyribose pucker is C_3'-*endo*. Z-DNA formation is not restricted to alternating $(CG)_n$ sequences, although these undergo the B—Z transition most readily. In general alternating purine–pyrimidine sequences are required, although some deviation from this can be tolerated. $(TG)_n.(CA)_n$ forms Z-DNA quite readily,[11] but $(AT)_n$ has not been observed in

[5] C. R. Calladine and H. R. Drew, *J. Mol. Biol.*, 1984, **178**, 773–782.
[6] R. Wing, H. R. Drew, T. Takano, C. Broka, S. Tanaka, L. Itakura, and R. E. Dickerson, *Nature*, 1980, **287**, 755–758.
[7] R. E. Dickerson and H. R. Drew, *J. Mol. Biol.*, 1981, **149**, 761–786.
[8] C. R. Calladine, *J. Mol. Biol.*, 1982, **161**, 343–352.
[9] R. E. Dickerson in 'Unusual DNA Structures', ed. R. D. Wells and S. C. Harvey, Springer-Verlag, 1988, pp. 287–306.
[10] A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich, *Nature*, 1979, **282**, 680–686.
[11] D. B. Haniford and D. E. Pullyblank, *Nature*, 1983, **302**, 632—634.

the Z-conformation to date. Formation of Z-DNA may be facilitated by certain base modifications, such as 5-MeC and 8-BrG, and the nature of the counterion is critical.[12] Z-DNA exhibits chemical reactivities not found in B-DNA,[13,14] and is highly immunogenic,[15] both poly- and mono-clonal antisera having been raised.

The formation of Z-DNA may be regarded as a particularly dramatic example of sequence-dependent structural polymorphism, but it is by no means unique. The cruciform[16–18] is perhaps an even more disruptive structural variant, which requires a sequence having twofold symmetry, *i.e.* an inverted repeat. Such a sequence may in principle form intra-strand base-paired helices, *i.e.* hairpin-loop structures, on each strand, to form the cruciform structure. These are discussed in much greater detail below, where they form the main theme of this review.

Curved (sometimes called bent) DNA sequences have recently attracted a great deal of interest. Certain sequences, such as those found in the kinetoplast mini-circle DNA of trypanosome mitochondria, possess a strong intrinsic curvature,[19] resulting in altered physical properties such as anomalously slow electrophoretic migration through polyacrylamide gels. This is called sequence-directed curvature of DNA, to distinguish it from the bending of DNA which results from the application of a force, as can occur on binding of a protein such as the cAMP-dependent activator protein[20] of *Eschericia coli.* Sequence-directed curvature is associated with short runs of oligo-dA.oligo-dT, repeating every 10 to 11 base-pairs,[21–22] i.e. in phase with the helical repeat of B-DNA. It is known from recent crystallographic studies[23,24] that such sequences may adopt a novel geometry (sometimes called B′), characterized by large propeller twist of the A–T base-pairs, such that the thymine O(4) can form bifurcated hydrogen bonds to N(6) protons of the conventionally base-paired adenine, and the cross-strand adenine of the adjacent base-pair. This results in a narrowed minor groove. At the junctions between the A_n tract and the normal DNA there is a large roll angle, and this is a possible origin of the curvature. When added in phase with the DNA helix a large overall bend is the result, such as that found in *Crithidia* for example. In retrospect, we can now say that the structure adopted by the

[12] J. H. van de Sande, L. P. McIntosh, and T. M. Jovin, *EMBO J.*, 1982, **1**, 777–782.
[13] K. Nejedly, M. Kwinkowski, G. Galazka, J. Klysik, and E. Palecek, *J. Biomol. Struct. Dyn.*, 1985, **3**, 467–478.
[14] B. H. Johnston and A. Rich, *Cell*, 1985, **42**, 713–724.
[15] E. M. Lafer, A. Möller, A. Nordheim, B. D. Stollar, and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 3546–3550.
[16] M. Gellert, K. Mizuuchi, M. H. O'Dea, H. Ohmori, and J. Tomizawa, *Cold Spring Harbor Symp. Quant. Biol.*, 1979, **43**, 35–40.
[17] D. M. J. Lilley, *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 6468–6472.
[18] N. Panayotatos and R. D. Wells, *Nature*, 1981, **289**, 466–470.
[19] J. C. Marini, S. D. Levene, D. M. Crothers, and P. T. Englund, *Proc. Natl. Acad. Sci. USA*, 1982, **79**, 7664–7668.
[20] H.-M. Wu and D. M. Crothers, *Nature*, 1984, **308**, 509–513.
[21] S. Diekmann and J. C. Wang, *J. Mol. Biol.*, 1985, **186**, 1–11.
[22] H. S. Koo, H.-M. Wu, and D. M. Crothers, *Nature*, 1986, **320**, 501–506.
[23] H. C. M. Helson, J. T. Finch, B. F. Luisi, and A. Klug, *Nature*, 1987, **330**, 221–226.
[24] M. Coll, C. A. Frederick, A. H.-J. Wang, A. Rich, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 8385–8389.

Dickerson dodecamer CGCGAATTCGCG [7] is really a B′ conformation because of the central sequence.

Other sequence-specific DNA polymorphs exist, though in general these are less well characterized than Z-DNA, cruciforms, or curved DNA. An example is the structure adopted by sequences in which there is an asymmetry of purines on one strand and pyrimidines on the other.[25–27] At low pH and moderate levels of supercoiling (see below), these form a novel structure characterized by sensitivity towards a number of enzymes and chemicals. Probing data are consistent with a novel structure comprizing a looping of the pyrimidine strand in triple-helix formation with half of the purine strand,[28,29] although it remains possible that other structures may be adopted under some conditions. The C–G–C triads formed in the triplex structure require the participation of protonated cytosine in a Hoogsteen base-pair with guanine, accounting for the pH-dependence of the structure.

Thus we can see that DNA structure is profoundly sequence-dependent, and that structural perturbation on a major scale is possible.

3 Helix Opening in Relaxed DNA

For many of the biological functions of DNA, such as transcription (synthesis of RNA from the DNA template) and replication, some helix opening is a necessary prerequisite, separating the paired bases to allow for the genetic 'reading' required. How easy is this process in physical-chemical terms, and what can be done to facilitate the opening process?

The most direct way to observe transient opening of base-pairs in the double helix is to look for a chemical reaction occurring at the atoms which remain completely protected when the base-pairing is undisrupted. The simplest reaction is the exchange of the hydrogen-bonded imino protons with solvent water, a reaction catalysed by proton acceptors such as Tris base, ammonia, or hydroxyl ions. Originally this was observed by studying the exchange of tritium from fully tritiated DNA fragments.[30] The reaction is visualized as occurring in two stages.

$$\underset{^3H}{\text{Closed}} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{Open} \xrightarrow{k_{ex}} \underset{^1H}{\text{Exchange}}$$

The first stage is the opening of the base-pair, characterized by the opening and closing rate constants, k_{op} and k_{cl}. Once the base-pair is open, proton exchange occurs with a rate constant k_{ex}. Two extreme situations may be considered, depending on the relative rates k_{ex} and k_{cl}.

If $k_{ex} \gg k_{cl}$ the measured exchange rate is the base-pair opening rate. Exchange will occur with each opening. This situation may be brought about by extrapolating to infinitely high concentrations of exchange catalysts.

[25] J. M. Nickol and G. Felsenfeld, *Cell*, 1983, **35**, 467–477.
[26] E. Schon, T. Evans, J. Welsh, and A. Efstratiadis, *Cell*, 1983, **35**, 837–848.
[27] D. E. Pullyblank, D. B. Haniford, and A. R. Morgan, *Cell*, 1985, **42**, 271–280.
[28] S. M. Mirkin, V. I. Lyamichev, K. N. Drushlyak, V. N. Dobrynin, S. A. Filippov, and M. D. Frank-Kamenetskii, *Nature*, 1987, **330**, 495–497.
[29] Y. Kohwi and T. Kohwi-Shigematsu, *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 3781–3785.
[30] J. J. Englander and P. H. von Hippel, *J. Mol. Biol.*, 1972, **63**, 171–177.

If $k_{ex} \ll k_{cl}$ the exchange is rate-limiting, and many openings are required to achieve exchange.

In early work employing tritium exchange the former condition was assumed, leading to the estimation of rather high values for the opening probability (*i.e.* k_{op}/k_{cl}), and long lifetimes for both open and closed states.

Recently, Guéron and colleagues have re-examined this process[31] using ^1H n.m.r. This technique has two major advantages over tritium exchange. First, imino proton resonances of specific bases may be assigned unambiguously, and thus one may study the exchange rate, derived from the resonance linewidth, of a particular base proton. Second, the method permits the measurement of faster exchange rates. Guéron *et al.* studied the exchange of the thymine imino proton in self-complementary oligonucleotides, notably in dCGCGATCGCG.

They observed catalyst-induced changes in the rate of imino proton exchange, a feature not previously observed. This enabled them to obtain individual base-pair lifetimes, around 10 ms at 15 °C, and also the probability of opening, around 10^{-5}. These values differ by orders of magnitude from those reported earlier in similar compounds, although Frank–Kamenetskii had arrived at similar values using a reinterpretation of chemical reactivity data.[32] Earlier investigators were misled by the existence of proton exchange even in the absence of added catalyst. Guéron *et al.*[31] showed that the rate of proton exchange was unrelated to base-pair lifetime, and exchange occurred from the same open state as exchange by added catalyst. This was most probably the proton-acceptor nitrogen of the complementary base. We may conclude that in linear DNA the base-pairs open singly and transiently, and the bulk properties of the DNA will be well represented by a fully base-paired molecule. However, as we shall see, the situation is quite different for a supercoiled DNA molecule.

4 DNA Supercoiling

Most natural DNA molecules are constrained into circles or loops, whereupon the topology adds a new dimension to the structural properties of the molecules.[33] While this is true for complex loops of chromosomes in bacteria[34] or higher cells,[35] the properties are most easily studied in small circular DNA molecules. Perhaps the simplest such circles for study are the bacterial plasmids, which are small (typically 2 to 6 kbp*), easily manipulated by established techniques of modern molecular genetics, and isolated from the cell as supercoiled species.

DNA supercoiling is based upon the fundamental topological property of

[31] M. Guéron, M. Kochoyan, and J.-L. Leroy, *Nature*, 1987, **328**, 89–92.
[32] M. D. Frank-Kamenetskii in '*Structure and Motion: Membranes, Nucleic Acids and Proteins*', ed. E. Clementi, G. Corongiu, M. H. Sarma, and R. H. Sarma, Adenine Press, 1985, pp. 417–432.
[33] J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, *Proc. Natl. Acad. Sci. USA*, 1965, **53**, 1104–1111.
[34] A. Worcel and E. Burgi, *J. Mol. Biol.*, 1972, **71**, 127–147.
[35] J. R. Paulson and U. Laemmli, *Cell*, 1977, **12**, 817–828.
* DNA molecules are conveniently measured in terms of the number of base-pairs which they comprise. 1 kbp = 1000 base-pairs.

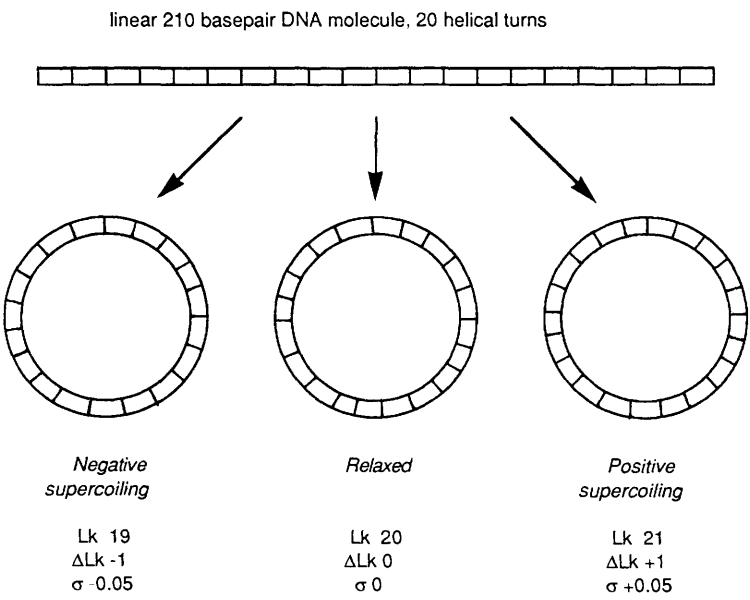

Figure 2 *Topoisomeric DNA circles and the linking number. We take a linear DNA molecule comprising exactly* 20 *turns of double helix, and ligate it into a circle. If this is carried out in the absence of any torsional force, the resulting circle will also contain* 20 *turns, and will be torsionally relaxed (central circle). The two strands of the DNA will mutually interlink twenty times, thereby defining a linking number of* 20. *Isomers of the relaxed species will be generated if torsion is applied before ligation. Removal of turns will generate topoisomers of reduced linking number (negative supercoiling, left), while addition of turns will increase the linking number (positive supercoiling, right). If this experiment was carried out in practice, a distribution of topoisomers would be observed, corresponding to thermal population of a Boltzmann distribution*

linkage, a property that is exhibited by two interwound circles. Two circles can be linked n times, where n is an integer which measures the minimum number of times one circle pierces the plane of the other. The linking number (written as Lk, L, or α, depending upon the author) is invariant so long as the integrity of both circles is maintained. Circular DNA exhibits linkage because the two strands turn around each other once in every ten or so base-pairs. Thus in a covalently closed circular DNA molecule the linking number is approximately $N/10.5$, where N is the number of base-pairs. The phenomenon of supercoiling in DNA arises from *differences* in the linking number, illustrated by the imaginary experiment depicted in Figure 2. In this experiment we have taken a linear DNA molecule comprizing exactly twenty turns, and used an enzyme to join it into a circle. If no torsional force is applied the resulting circular species also contains 20 turns, and hence the linking number is 20. Since this molecule is directly derived from the torsionally relaxed linear species, we refer to this as the relaxed circle, and the linking number is denoted by Lk^0. If we perform this experiment

again, but apply torsion to the molecule prior to ligation,* we may change Lk. Thus we may add or substract turns of DNA, resulting in a circular molecule which is over- or under-wound. We define a linking difference

$$\Delta Lk = Lk - Lk^0$$

which is just the number of turns removed or added in the above experiment. We may also define a further parameter σ, the specific linking difference or superhelix density, which is independent of size

$$\sigma = \Delta Lk/Lk^0$$

Thus in the above experiment, if one turn was removed before ligation, $Lk = 19$, $\Delta Lk = -1$ and $\sigma = -0.05$. Most natural DNA molecules are underwound, or negatively supercoiled, in this way, with typical superhelix densities of $\sigma = -0.06$. It should be noted that although Lk is necessarily integer, this restriction does not apply to Lk^0 and hence to ΔLk, because most DNA molecules will not contain an exact number of helical turns under a given set of conditions (helix pitch is a function of environmental conditions such as temperature and ionic strength).

It will be noted that the circular forms derived in the circularization experiment are isomers which result from topological differences, and are therefore called topoisomers. These will differ in their free energy, to an extent which arises from the torsional force required to underwind the DNA helix. This is a quadratic function [36–38]

$$\Delta G_s = \frac{1050RT}{N}\Delta Lk^2$$

where ΔG_s is the free energy of supercoiling, R the gas constant, T the absolute temperature and N the size of the molecule in base-pairs. These energies are significant. The plasmid pBR322, which is commonly used for cloning DNA sequences in bacteria, has a free energy of supercoiling of about 100 kcal mol^{-1} at $\sigma = -0.05$, an energy greater than that of the C–C covalent bond.

Living cells have evolved an array of enzymes to manipulate the topology of DNA molecules. Enzymes which can change the linking number of a topoisomer are called topoisomerases, and these are present in all cells.[39] In bacteria two main topoisomerases are known. Topoisomerase I[40] has a relaxing activity,

[36] R. E. Depew and J. C. Wang, *Proc. Natl. Acad. Sci. USA*, 1975, **72**, 4275–4279.
[37] D. E. Pullyblank, M. Shure, D. Tang, J. Vinograd, and H.-P. Vosberg, *Proc. Natl. Acad. Sci. USA*, 1975, **72**, 4280–4284.
[38] D. S. Horowitz and J. C. Wang, *J. Mol. Biol.*, 1984, **173**, 75–91.
[39] M. Gellert, *Ann. Rev. Biochem.*, 1981, **50**, 879–910.
[40] J. C. Wang, *J. Mol. Biol.*, 1971, **55**, 523–533.

* *Ligation* refers to the covalent joining of two ends of DNA molecules, *i.e.* the formation of two new covalent phosphodiester linkages, using an enzyme (called ligase) and ATP. The joined molecules might be completely base-paired ('blunt' ended) or have overhanging cohesive ends, *e.g.*

XXXG
YYYCTTAA5′.

which increases the linking number of a negatively supercoiled substrate until it is relaxed. It is the product of the *topA* gene,[41] and works by acting first as a nuclease which is transiently covalently linked to the DNA as a phosphate ester while the DNA swivels,[42] followed by a re-ligation process. The process is passive in the sense that no energy is required. The other main topoisomerase is DNA gyrase,[43] a tetrameric protein of two different subunits encoded by the unlinked genes *gyrA* and *gyrB*.[44–46] It introduces supercoiling using the coupled hydrolysis of ATP as an energy source. The cellular activities of the two enzymes is regulated to maintain the level of supercoiling *in vivo* at a fixed level.

5 DNA Supercoiling and Structure

DNA supercoiling modifies both DNA structure and dynamics. The change in structure can be very simply demonstrated experimentally by electrophoresis of a mixture of topoisomers of plasmid in an agarose gel, shown in Figure 3. Each topoisomer migrates as a well-resolved species, because it has a shape which is different from topoisomers of different linking number. Changes in linking number are coupled to alterations in helical twist (the path of the strands about the helical axis, *Tw*) and deformation of the helix path in space from a plane (writhing, *Wr*), according to[47]

$$Lk = Tw + Wr$$

or

$$\Delta Lk = \Delta Tw + Wr$$

where ΔTw is the deviation in twist from the equilibrium value. Thus supercoiling results in changes in helix twist, or axial writhing, or both. The exact partition between twisting and writhing in DNA circles of several thousand base-pairs, and the shape of the writhing molecules, are the subject of some debate, and will depend on a number of environmental factors. Both toroidal and interwound forms are possible, and while the former has been advocated on the basis of solution *X*-ray scattering measurements,[48] electron microscopic images of supercoiled DNA are normally interwound, and these are generally considered to be the form adopted in solution.

In addition to influencing the overall tertiary structure of the molecules, supercoiling may determine local structure, arising from a coupling between local

[41] R. Sternglanz, S. DiNardo, K. A. Voelkel, Y. Nichimura, Y. Hirota, V. Becherer, L. Zumstein, and J. C. Wang, *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 2747—2751.
[42] R. E. Depew, L. F. Liu, and J. C. Wang, *J. Biol. Chem.*, 1978, **253**, 511—518.
[43] M. Gellert, K. Mizuuchi, M. H. O'Dea, and H. A. Nash, *Proc. Natl. Acad. Sci. USA*, 1976, **73**, 3872—3876.
[44] M. J. Ryan, *Biochemistry*, 1976, **15**, 3769—3777.
[45] W. A. Goss, W. H. Deitz, and T. M. Cook, *J. Bacteriol.*, 1965, **89**, 1068—1074.
[46] W. L. Staudenbauer, *Eur. J. Biochem.*, 1976, **62**, 491—497.
[47] F. B. Fuller, *Proc. Natl. Acad. Sci. USA*, 1971, **68**, 815—819.
[48] G. W. Brady, D. B. Fein, H. Lambertson, V. Grassian, D. Foos, and C. J. Benham, *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 741—744.

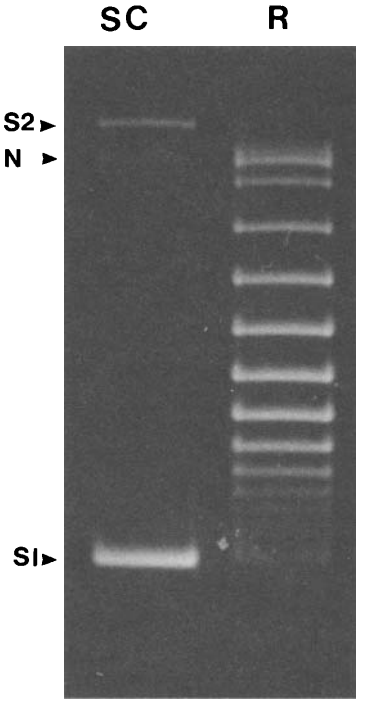

Figure 3 *Topoisomers of a* 3 657 bp *circular DNA species. This small plasmid was partially relaxed with a topoisomerase which interconverts topoisomeric species, changing the linking number, and the product was examined by electrophoresis in an agarose gel. On the left (lane SC) is shown the unrelaxed negatively supercoiled DNA, obtained by extraction directly from bacterial cells. This DNA is supercoiled to about* $\sigma = -0.06$, *but the distribution of topoisomers is unresolved under these conditions, S1; supercoiled monomeric plasmid DNA. N; monomer which has undergone strand breakage. S2; supercoiled dimer. After incubation with the topoisomerase a ladder of species is generated (lane R, right) in which each band is a single topoisomer differing in unit linking number from its direct neighbours. The topological differences between topoisomeric species result in alteration to the overall shape of the molecules, and hence the rates of migration in the gel*

geometric factors and the global topology of the molecules. Take the example of cruciform formation, shown diagrammatically in Figure 4. An inverted repeat of n base-pairs contributes a Tw of $n/10.5$ to the total twist of the molecule in its unperturbed state, but after adopting cruciform geometry this drops to approximately zero. Thus there is a negative ΔTw of $n/10.5$ turns due to cruciform formation, and the molecule may relax by this amount. The unwinding which accompanies Z-DNA formation is even greater, almost by a factor of two. To return to our earlier example of pBR322, formation of a 30 bp cruciform will relax approximately 3 turns, corresponding to a lowering of free energy by 21 kcal mol^{-1} at $\sigma = -0.06$. Therefore even though the formation of the cruciform may be quite unfavourable in linear DNA, provided the free energy of formation

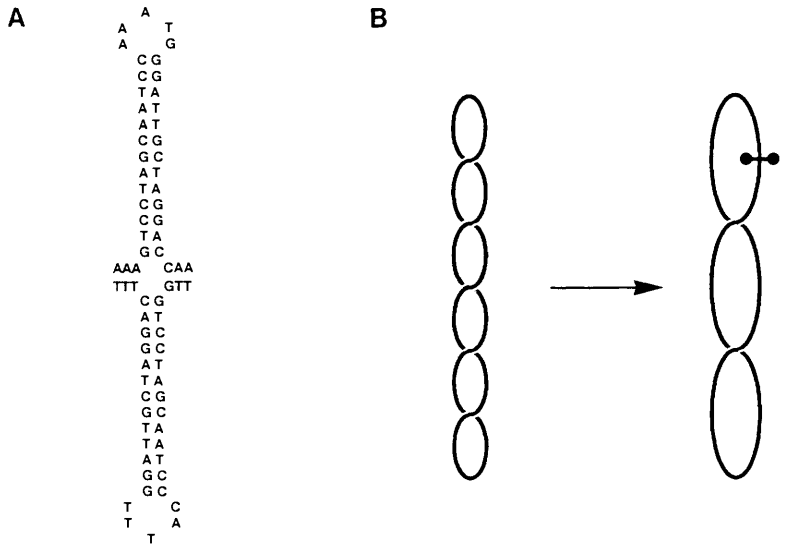

Figure 4 *Cruciform extrusion by inverted repeats in supercoiled DNA.* (A) *The sequence of an inverted repeat found in the* E. coli *plasmid* ColE1 *drawn in its cruciform conformation. Note that as a result of the twofold symmetry of the sequence, the individual strands can base-pair within the* same *strand, generating stem-loop, or hairpin, structures.* (B) *Negative supercoiling stabilizes cruciform structures. This scheme illustrates the relaxation of superhelical stress which may be obtained on cruciform formation (or the formation of any structure leading to a local reduction in twist-Z-DNA would be another example). If the change in the free energy of supercoiling resulting from this relaxation is greater than that of formation of the cruciform, a stable cruciform results*

is less than 21 kcal mol^{-1} this will be compensated by the gain in that of supercoiling, and the cruciform will be stable. In fact, the free energy of cruciform formation has been measured for a number of sequences[49–53] and lies in the range of 12 to 18 kcal mol^{-1}. The same kind of analysis applies to any local structural variation which is accompanied by a negative twist change, including the formation of Z-DNA and melted 'bubbles'. This shows why the opening of the DNA helix by RNA polymerase is facilitated by negative supercoiling in general, where about 12 bp are required to become unpaired on formation of the open complex.[54]

Multi-state conformational equilibria are possible in supercoiled DNA molecules. They can arise in two different ways. First, a circular molecule may have more than one sequence which can undergo a conformational transition. Since a

[49] M. Gellert, M. H. O'Dea, and K. Mizuuchi, *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 5545–5549.
[50] A. J. Courey and J. C. Wang, *Cell*, 1983, **33**, 817–829.
[51] D. M. J. Lilley and L. R. Hallam, *J. Mol. Biol.*, 1984, **180**, 179–200.
[52] D. R. Greaves, R. K. Patient, and D. M. J. Lilley, *J. Mol. Biol.*, 1985, **185**, 461–478.
[53] L. H. Naylor, D. M. J. Lilley, and J. H. van de Sande, *EMBO J.*, 1986, **5**, 2407–2413.
[54] J. Siebenlist, *Nature*, 1979, **276**, 651–652.

transition in one sequence will reduce the overall level of supercoiling in the molecule, this will change the probability of the second transition occurring. Such cases have been observed experimentally for Z-forming sequences, and have been treated within the framework of statistical thermodynamics.[55] Second, a given sequence may itself be capable of forming more than one alternative structural form. The sequence $(GT)_n(AC)_n$ for example might exist as B; Z; or cruciform-DNA, depending on the length, level of supercoiling *etc.*, and be described by a multi-dimensional phase diagram. We have observed similar two-state conformational equilibria for $(AT)_n$ tracts in supercoiled molecules,[56] which may exist either as cruciform structures or an unwound and chemically reactive uniform helical variant of B-DNA. This equilibrium is a sensitive function of salt concentration and temperature.

Not only does negative supercoiling increase the structural repertoire of DNA, it also increases its dynamic character, resulting in enhanced chemical reactivity for example. As we shall see, cruciform structures may be extruded in supercoiled DNA by at least two kinetic pathways. Whatever the details of the mechanism however, it is clear that many base-pairs must be broken during this process and that therefore helix openings must be possible in supercoiled DNA on a scale much greater than that found in linear molecules. To this extent therefore, supercoiling is a kind of 'activation' of DNA, generating new structural and dynamic aspects to the character of the double helix.

6 The Cruciform Structure

Cruciform structures were first recognized experimentally at the beginning of this decade,[16–18] twenty-five years after their first theoretical description.[57,58] The crucial role of supercoiling had been overlooked in earlier studies, and this was the first example of the now-familiar stabilization of structural variants by negative supercoiling. Cruciform formation has subsequently been demonstrated for many inverted repeat sequences in a variety of supercoiled plasmids and phage.

Conventional physical methods (spectroscopy, diffraction, *etc.*) cannot in general be applied to investigation of supercoil-stabilized features because the feature of interest, *e.g.* the cruciform, is normally a very small part of the entire circular plasmid molecule. We have therefore developed new methods with which to investigate these structures.

We introduced the procedure of nuclease cleavage on supercoiled DNA molecules.[17,18] This probing method takes advantage of some feature of the novel structure which can be differentially recognized by an enzyme or chemical. In the case of the cruciform there are two such features, the single-stranded loops and the four-way junction. Most probes select the former as their target, being either single-strand specific enzymes (such as S1 nuclease), or single-strand selective chemicals (such as bromoacetaldehyde). Some probes of the junction

[55] R. J. Kelleher, M. J. Ellison, P. S. Ho, and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 6342—6346.
[56] J. A. McClellan and D. M. J. Lilley, *J. Mol. Biol.*, 1987, **197**, 707—722.
[57] J. R. Platt, *Proc. Natl. Acad. Sci. USA*, 1955, **41**, 181 183.
[58] A. Gierer, *Nature*, 1966, **212**, 1480 1481.

Table 1 *Enzyme and chemical probes which have been employed in the study of cruciform structures*

Probe	*Target*	*Result*	*Reference*
S1 nuclease	loop	cleavage	17,18
Micrococcal nuclease	loop	cleavage	59,60
Bal31 nuclease	loop	cleavage	51
P1 nuclease	loop	cleavage	61
Mung bean nuclease	loop	cleavage	62
Bromoacetaldehyde	loop (A,C)	etheno-adduct	63
Osmium tetroxide	loop (T,C)	*cis*-diester	64
Glyoxal	loop (G)	etheno-adduct	65,66
Bisulphite	loop (C)	deamination to dU	67
Diethyl pyrocarbonate	loop (A)	carbethoxylation	68,69
Resolvase	junction	cleavage	70—74
e.g. T4 endonuclease VII			

have impressive structural selectivity, notably the Holliday junction* resolving enzymes. A tabulation of probes for cruciform structures is presented in Table 1.

The structure of the cruciform is defined by the two features indicated above—the loops and the four-way junction. Chemical probing experiments[67] have indicated that the optimal loop size lies between four and six nucleotides. Detailed analysis of diethyl pyrocarbonate modification patterns in the ColE1 cruciform loop have suggested to us[68] that the loop possesses a well defined structure, probably involving base stacking. Studies of isolated DNA hairpins by Hilbers and co-workers[75] have indicated that bases may stack, as if to continue the helical structure, on the 5′-side of the loop.

The four-way junction is formally equivalent to the Holliday[76] junction, and

[59] C. Dingwall, G. P. Lomonossoff, and R. A. Laskey, *Nucleic Acids Res.*, 1981, **9**, 2659–2673.
[60] D. M. J. Lilley, *Cold Spring Harbor Symp. Quant. Biol.*, 1982, **47**, 101–112.
[61] D. B. Haniford and D. E. Pulleyblank, *Nucleic Acids Res.*, 1985, **13**, 4343–4363.
[62] L. G. Sheflin and D. Kowalski, *Nucleic Acids Res.*, 1984, **12**, 7087–7104.
[63] D. M. J. Lilley, *Nucleic Acids Res.*, 1983, **11**, 3097–3112.
[64] D. M. J. Lilley and E. Palecek, *EMBO J.*, 1984, **3**, 1187–1192.
[65] G. W. Gough, *Ph.D. Thesis*, University of Dundee, 1986.
[66] D. M. J. Lilley in '*The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis*', ed. B. Singer and H. Bartsch, *IARC Publication No. 70*, Lyon, 1986.
[67] G. W. Gough, K. M. Sullivan, and D. M. J. Lilley, *EMBO J.*, 1986, **5**, 191–196.
[68] J. C. Furlong and D. M. J. Lilley, *Nucleic Acids Res.*, 1986, **14**, 3995–4007.
[69] P. M. Scholten and A. Nordheim, *Nucleic Acids Res.*, 1986, **14**, 3981–3993.
[70] K. Mizuuchi, B. Kemper, J. Hays, and R. A. Weisberg, *Cell*, 1982, **29**, 357–365.
[71] D. M. J. Lilley and B. Kemper, *Cell*, 1984, **36**, 413–422.
[72] B. deMassey, F. W. Studier, L. Dorgai, F. Appelbaum, and R. A. Weisberg, *Cold Spring Harbor Symp. Quant. Biol.*, 1984, **49**, 715–726.
[73] L. Symington and R. Kolodner, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 7247–7251.
[74] S. C. West and A. Korner, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 6445–6449.
[75] C. A. G. Haasnoot, C. W. Hilbers, G. A. van der Marel, J. H. van Boom, U. C. Singh, N. Pattabiraman, and P. A. Kollman, *J. Biomol. Struct. Dyn.*, 1986, **3**, 843–857.
[76] R. Holliday, *Genet. Res.*, 1964, **5**, 282–304.

* *Holliday junctions.* Recombining DNA molecules initiate the process by strand exchange, leading to a four-way helical junction as first proposed by Holliday in 1964. The interested reader must consult any good molecular biology text under the subject of homologous genetic recombination.

its study is important to a full understanding of homologous genetic recombination, where great progress has been made very recently. Chemical probing with bisulphite and diethyl pyrocarbonate indicates that under normal circumstances the junction is probably fully base-paired,[67,68] and this is supported by n.m.r. studies of an isolated junction.[77] Investigation of a pseudo-cruciform construct[78] revealed a strongly position-sensitive retardation of gel migration akin to that seen for kinetoplast DNA,[20] and led us to propose that the cruciform junction introduces a pronounced bend into the molecule. In the presence of magnesium ions, the junction adopts a more stable and compact configuration. Using an extension of this method Cooper and Hagerman[79] studied the relative disposition of the helices, concluding that the junction is asymmetric, and not, for example, tetrahedral. We have extended these studies[80] using a related method, and shown that the junction has the shape of an X, formed by co-linear stacking of arms in pairs. Two isomeric structures are possible, and the choice of helical partners is determined by the local sequence in the junction itself. The resolution of a given junction by enzymes such as T4 endonuclease VII depends on relative isomer stability, and thus the products of recombination is likely to depend on local sequence at the point of resolution. Ion binding is also of fundamental importance. In the absence of ions, the structure of the junction is quite different—the helices are no longer stacked in pairs, but are instead fully extended in a square arrangement, resulting in thymines at the junction becoming reactive to osmium tetroxide.

7 Cruciform Extrusion Processes—Two Mechanisms

In general the process by which cruciform structures are extruded in supercoiled DNA is far from facile. The first kinetic studies of cruciform extrusion[81] revealed significant kinetic barriers, as might have been predicted in view of the substantial reorganization of DNA structure which must occur. It was found that the extrusion reaction could be very slow, even at elevated temperatures.[49,50] On the other hand, work on different inverted repeats showed that extrusion could occur rather more easily.[82–84]

We made a comparative study[84] of cruciform extrusion kinetics in two plasmids, pColIR315 and pIRbke8. pColIR315 contains a 440 bp insert from natural *E. coli* plasmid ColE1,[85] which includes a 31 bp ColE1 inverted repeat and about 100 bp and 300 bp of left and right flanking DNA respectively.

[77] D. E. Wemmer, A. J. Wand, N. C. Seeman, and N. R. Kallenbach, *Biochemistry*, 1985, **24**, 5745–5749.
[78] G. W. Gough and D. M. J. Lilley, *Nature*, 1985, **313**, 154–156.
[79] J. P. Cooper and P. J. Hagerman, *J. Mol. Biol.*, 1987, **198**, 711–719.
[80] D. R. Duckett, A. I. H. Murchie, S. Diekmann, E. von Kitzing, B. Kemper and D. M. J. Lilley, *Cell*, 1988, **55**, 79–89.
[81] K. Mizuuchi, M. Mizuuchi, and M. Gellert, *J. Mol. Biol.*, 1982, **156**, 229–243.
[82] I. Panyutin, V. Klishko, and V. Lyamichev, *J. Biomol. Struct. Dyn.*, 1984, **1**, 1311–1324.
[83] R. R. Sinden and D. E. Pettijohn, *J. Biol. Chem.*, 1984, **259**, 6593–6600.
[84] D. M. J. Lilley, *Nucleic Acids Res.*, 1985, **13**, 1443–1465.
[85] M. Bazaral and D. Helinski, *J. Mol. Biol.*, 1968, **36**, 185–194.

Table 2 *Kinetic characteristics of C-type and S-type cruciforms*

	C-type	*S-type*
Example	ColE1	pIRbke8
Occurrence	rare	common
[NaCl] optimum	0 mM	50 mM
E_a	180 kcal mol^{-1}	40 kcal mol^{-1}
$\Delta H^{\ddagger}$	180 kcal mol^{-1}	40 kcal mol^{-1}
$\Delta S^{\ddagger}$ (at 37 °C)	400 cal deg^{-1} mol^{-1}	60 cal deg^{-1} mol^{-1}

pIRbke8 was constructed[86] by cloning synthetic oligonucleotides into the *Bam*H1 restriction site of pAT153 to generate a perfect 32 bp inverted repeat (termed bke), flanked by vector sequences.

Figure 5 summarizes kinetic data obtained for the two plasmids. The properties are strikingly different, in two main respects:

NaCl Dependence. The ColE1 cruciform of pColIR315 extrudes maximally in the absence of added salt, and the extrusion is strongly suppressed with the addition of NaCl. By contrast, the bke cruciform of pIRbke8 fails totally to extrude in the absence of added salt, and exhibits a maximal extrusion rate at 50—60 mM NaCl.

Temperature Dependence. pColIR315 exhibits a marked temperature dependence for cruciform extrusion. The rate constant for the extrusion reaction increased by a factor greater than 2000 in an 8 degree temperature interval. The reaction is thus characterized by an enormous Arrhenius activation energy (E_a) in the region of 200 kcal mol^{-1}. The temperature dependence for the extrusion of pIRbke8 is rather lower, with an E_a of about 40 kcal mol^{-1}. However, despite the much larger activation barrier for pColIR315, extrusion proceeds at *lower* temperatures, indicating that the extrusion of the ColE1 cruciform must have a large entropy of activation.

These kinetic properties are highly contrasting, and constitute members of two classes of cruciform, termed C-type (for ColE1) and S-type (for Salt-dependent). The only natural member of the C-type class is ColE1, and its subclones such as pColIR315[84] and the deletant pAO3.[82] Most sequences behave as S-type cruciforms, including pUC7,[49] pAC102,[50] pOCE12,[83] with lower temperature dependences, a requirement for salt and extrusion occurring at higher temperatures. The kinetic properties of the two classes are summarized in Table 2.

The strongly contrasting kinetic properties of C- and S-type cruciforms suggests the existence of alternative mechanistic pathways for the extrusion process. In fact, from an entirely theoretical standpoint,[86] we had previously proposed two possible pathways, which are presented schematically in Figure 6. In the upper mechanism, a large region of DNA is unpaired in the probable

[86] D. M. J. Lilley and A. F. Markham, *EMBO J.*, 1983, **2**, 527–533.

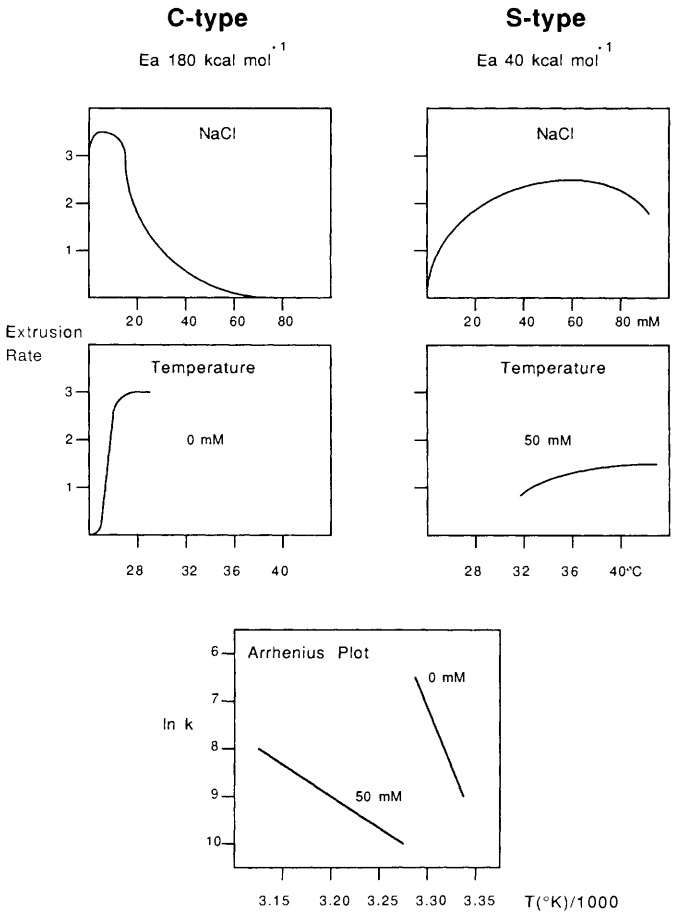

Figure 5 *The kinetic properties of cruciform extrusion in supercoiled DNA—a summary of two kinetic classes. Salt (upper) and temperature (lower) dependences for extrusion of typical C-type (left) and S-type (right) plasmids. In these experiments the extrusion process is allowed to proceed for a short time, after which the relative degree of extrusion is measured, which is proportional to the rate of extrusion. C-type extrusion is characterized by cruciform formation at low salt concentrations, with a very high temperature dependence, while S-type extrusion requires the inclusion of salt and proceeds with a lower temperature dependence. In other experiments the extent of extrusion is measured as a function of time in order to calculate rate constants for the process. Rate constants for C- and S-type extrusion have been measured as a function of temperature, and activation energies calculated from Arrhenius plots like that shown at the bottom. Typical Arrhenius activation energies are* 200 kcal mol^{-1} *for C-type extrusion and* 40 kcal mol^{-1} *for S-type extrusion*

transition state, which then undergoes intra-strand base-pairing to form the fully extruded cruciform. In the lower path, the extent of initial disruption is smaller. A relatively short proto-cruciform is formed as an intermediate, followed by branch

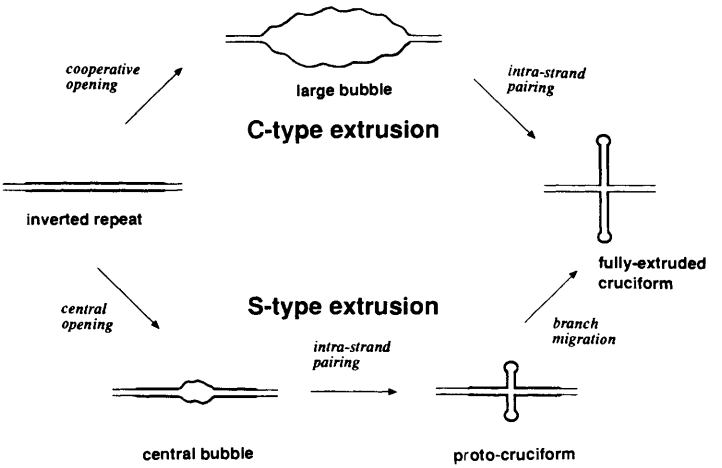

Figure 6 *Two mechanisms for cruciform extrusion. The inverted repeat, represented by the thicker line, is shown in the unextruded form on the left. We believe that C-type cruciforms (top) initiate the extrusion process with a coordinate opening of many base-pairs to form a large bubble. An intra-strand reassociation then forms the mature cruciform structure. The extrusion of S-type cruciforms (lower), is initiated by a smaller opening event. Intra-strand pairing generates a smaller proto-cruciform, which may undergo branch migration. Base-pairing is transferred from unextruded sequence to the growing cruciform stems in a multi-step process, to form the fully extruded structure. The principal differences between the two mechanisms lie in the size of the initial opening and the degree of tertiary folding in the transition state*

migration (sequential transfer of base-pairing to the growing cruciform stem) to the completely extruded cruciform. The transition state is harder to identify in this pathway, but is likely to be intermediate between the initial unpairing and the proto-cruciform. We have proposed [84] that the C-type cruciforms extrude *via* the top pathway, while the S-type sequences proceed *via* the lower mechanism. The major differences between the two mechanisms postulated may be summarized:

1. There is much more disruption of base-pairing in the upper pathway.
2. The transition state in the lower mechanism resembles the forming four-way junction, rather than a simple melted 'bubble'.

The upper pathway would be expected to be characterized by large values of enthalpy and entropy of activation, and to be facilitated by low ionic strength (phosphate repulsion reduces the melting temperature of DNA). The reduced disruption in the lower pathway will lower the enthalpy and entropy of activation, but phosphate–phosphate repulsion is likely to be significant in a more structured transition state. The energy of the activated complex will therefore be reduced by cation binding, in a manner analogous to the

requirement for cation binding in forming the compact conformation of the complete cruciform. Experimentally, we have found that C- and S-type cruciforms have exactly these kinetic properties.

We have examined the S-type mechanism in more detail. We took pIRxke/vec, a typical S-type molecule closely similar to pIRbke8, and studied the extrusion as a function of the nature of the cation present.[87] We observed that ions vary greatly in the efficiency with which they promote the extrusion. The effects of cations may be divided into four classes:

1. Extrusion rate maxima 50—60 mM, forming a distinct optimum. The monovalent ions, mainly Group 1a metals and $(CH_3)_4N^+$.
2. Extrusion rates levelling off by 100—200 μM. The divalent ions of Group IIa, together with Mn^{2+} and (poorly) Co^{2+}.
3. Peak of maximum rate at 15—50 μM. $[Co(NH_3)_6]$ (III) and polyamines.
4. Ineffective at all concentrations. Most transition metals are completely ineffective in promoting extrusion.

Despite the different efficiency with which the different ions facilitate cruciform extrusion, the reaction mechanism is unchanged. Thus, Arrhenius activation energies for the reaction in the presence of Na^+ and Mg^{2+} are the same within experimental error. Our conviction is that the observed kinetic consequences of using various cations results from differential stabilization of the transition state, from which we may deduce some aspects of the structure of the activated complex.

1. Groups 1 to 3 show that the optimal ion concentration is reduced as charge increases, reflecting ionic binding.
2. Polyamines are very effective, while basic amino-acids are totally ineffective. The distribution of positive and negative charges on these ions must affect binding.
3. Many transition metals are totally ineffective in promoting extrusion. Most of the ions in this group are soft, binding preferentially to nitrogen or exocyclic keto-substituents, *i.e.* they will be better at binding bases than phosphate groups. In fact there is a good correlation between the ability of ions to promote S-type cruciform extrusion, and their position in the Irving–Williams series.[88] Clearly phosphate binding is required.
4. Within the Ia and IIa metals, there is an excellent correlation between the rate of cruciform extrusion and ionic radius, shown graphically in Figure 7. Ion binding to normal DNA, which is effectively a cylindrical polyelectrolyte, is well treated using electrostatic consideration alone.[89,90] Counterions reduce phosphate charge by acting as a screening cloud, and indeed it has not proved

[87] K. M. Sullivan and D. M. J. Lilley, *J. Mol. Biol.*, 1987, **193**, 397–404.
[88] H. Irving and R. J. P. Williams, *Nature*, 1948, **163**, 746–747.
[89] G. S. Manning, *Q. Rev. Biophys.*, 1970, **11**, 179–246.
[90] B. H. Zimm and M. Le Bret, *J. Biomol. Struct. Dyn.*, 1983, **1**, 461–471.

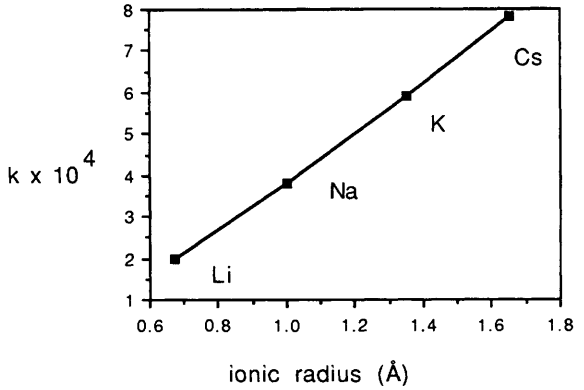

Figure 7 *The rate of S-type cruciform extrusion depends upon the ionic radius of the cation present. Rate constants for cruciform extrusion at* 35 °C *were measured in the presence of* 60 mM *group* Ia *metal chlorides, and these are plotted against the ionic radius.*

possible to localize ions in crystal structures of double-stranded DNA. In contrast, tRNA has a number of high affinity binding sites for Mg^{2+},[91–93] generated by the tertiary folding of the molecule, and we now believe very similar ion binding is very important in the formation of the four-way junction.[80] The dependence of extrusion rate on ionic *size* therefore implies selective ion binding,[94] suggesting the formation of a structure which contains electronegative clefts capable of such ion binding. This is further evidence for the transition state having significant four-way junction character, as proposed earlier.

In a different study of the S-type extrusion reaction, we have constructed a series of variants of the typical S-type molecule pIRbke8, in which one or two mutations (mainly A.T to G.C or *vice versa*), have been introduced into the symmetrical unit of the bke inverted repeat.[95] The results are summarized graphically in Figure 8. The mutations exhibiting the most significant alterations to extrusion rates were those in which the base changes are close to the sequence dyad. Thus the half-time for the process at 37 °C can change from 70 min in pIRbke8 (central sequence AGAATTCT) to less than 1 min for ATATATAT or more than 29 h for AGCCGGCT. With a few special exceptions, sequence changes further from the centre of the inverted repeat only alter the rates by a factor of two or less. This suggests that only the *central region* of the inverted repeat becomes altered in the formation of the transition state, confirmed by a

[91] A. Jack, J. E. Ladner, D. Rhodes, R. S. Brown, and A. Klug, *J. Mol. Biol.*, 1977, **111**, 315—328.
[92] S. R. Holbrook, J. L. Sussman, R. Wade Warrant, G. M. Church, and S.-H. Kim, *Nucleic Acid Res.*, 1978, **8**, 2811—2820.
[93] G. J. Quigley, M. M. Teeter, and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1978, **75**, 64—68.
[94] S.-K. Tam, and R. J. P. Williams, *Struct. Bonding*, 1985, **63**, 103–151.
[95] A. I. H. Murchie and D. M. J. Lilley, *Nucleic Acids Res.*, 1987, **15**, 9641—9654.

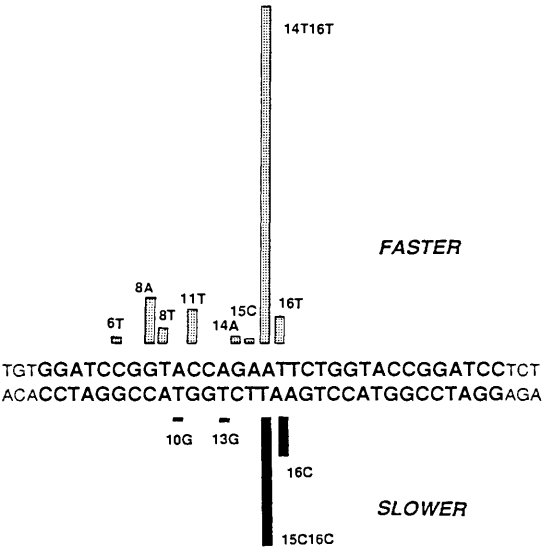

Figure 8 *Rates of S-type cruciform extrusion depend on the sequence at the centre of the inverted repeat. Rate constants for cruciform extrusion were measured at 37 °C for a series of closely related mutant sequences related by one or two bases in the symmetric unit. In this representation the bars indicate the ratio of the extrusion rate of the mutant sequence to that of the original sequence. Those above the sequence are rate enhancements, while those below indicate slower rates. In the sequence bold typeface denotes bases comprizing the parent inverted repeat sequence. Mutant sequences are named to indicate the altered position (numbered from the 5′ end of the inverted repeat) and the new base at this location. Corresponding mutations are always present that preserve the twofold symmetry*

similar study.[96] Very recently Courey and Wang[97] have studied S-type cruciform extrusion as a function of supercoiling, from which they were able to deduce that the initial opening corresponds to about 10 bp. These results are fully consistent with the S-type mechanism shown in Figure 6. Even more subtle changes may be made at the centres of these inverted repeats, using base methylation.[98] For example, the bke sequence is symmetrically disposed about the sequence GAATTC, which is a substrate for the *Eco*RI methylase. Incubation with this enzyme and *S*-adenosyl methionine results in methylation at the N(6) of the second adenine of the sequence on both strands, and we find that this small change alone results in a threefold rate enhancement for S-type cruciform extrusion. A related inverted repeat containing a central AGCT sequence could be modified using *Alu*I methylase to introduce 5-Me cytosine on both strands, and this resulted in a threefold reduction in cruciform extrusion rate. These results are fully explicable in terms of the proposed model for S-type extrusion,

[96] G. Zhang and P. R. Sinden, *J. Biol. Chem.*, 1989, **205**, 593–602.
[97] A. J. Courey and J. C. Wang, *J. Mol. Biol.*, 1988, **202**, 35–43.
[98] A. I. H. Murchie and D. M. J. Lilley, *J. Mol. Biol.*, 1989, **205**, 593–602.

since methylation of A, which interferes with A.T base-pairing in the central region, leads to facilitation of the initial opening, while methylation of C, which stabilizes the helix by improved stacking, makes this harder.

No data which seriously challenge the mechanistic models described for cruciform extrusion have emerged, and we are confident that these are a good description of the physical processes involved.

8 Contextual Influence on the Kinetic Character of Cruciform Extrusion

We have described two contrasting mechanisms by which cruciform extrusion may occur, and yet we have given no indication as to what determines which pathway any particular sequence takes. Comparison of the base sequences of the ColE1 and bke cruciforms provides no clues to the origins of the differences—both are 50—60% (A + T), and neither has additional sequence motifs such as purine–pyrimidine alternation or polypurine tracts. We should recall that the unusual behaviour is that of ColE1; most cruciforms are S-type. In fact, there is one respect in which pColIR315 is quite abnormal. The ColE1 sequences which flank the inverted repeat on both sides are over 80% (A + T). We were forced to consider that this very (A + T)-rich base composition might be responsible for the C-type kinetic behaviour of the ColE1 cruciform. We therefore examined the possibility that the (A + T)-rich ColE1 sequences somehow influence the cruciform extrusion pathway, and confer C-type kinetics on the adjoining inverted repeat.[99]

The following series of experiments were performed:

1. Using standard methods of recombinant DNA technology, we deleted the ColE1 cruciform from pColIR315 (generating a plasmid called pColIRΔxba), and replaced it with a new inverted sequence which was very similar to the inverted repeat of pIRbke8 (the central 20 bp were identical). Thus we constructed a plasmid in which a bke-like cruciform resides in the context of the (A + T)-rich ColE1 sequences. Cruciform extrusion in the new plasmid (pIRxke/col) exhibited maximal rates at 0 mM NaCl, with an E_a of 215 kcal mole^{-1}, *i.e.* properties typical for *C-type* cruciform extrusion.
2. We also performed the experiment in reverse. We cloned oligonucleotides to generate a ColE1 inverted repeat in the *Bam*HI site of pAT153, the location at which the bke inverted repeat of pIRbke8 normally resides. The resulting plasmid (pIRCol/vec) exhibited typical *S-type* extrusion kinetics, i.e. maximal extrusion at 50 mM NaCl, with an E_a of 50 kcal mole^{-1}.

These results show that the kinetic class, and thus the mechanistic pathway, of cruciform extrusion is determined by sequences which lie outside the inverted repeat. The sequence of the inverted repeat itself seems to be of secondary importance in this selection. The (A + T)-rich ColE1 flanking sequences appear to confer C-type extrusion kinetics on whatever sequences are placed next to

[99] K. M. Sullivan and D. M. J. Lilley, *Cell*, 1986, **47**, 817–827.

them, although we may well discover inverted repeat sequences refractory to their influence. We have termed them C-type Inducing Sequences. Manipulation of these sequences has revealed the following properties:[99,100]

1. Only a single sequence is required. Either the right- or left-hand ColE1 sequences function as independent inducing sequences, and some other (A + T)-rich sequences, *e.g.* a 200 bp fragment from *Drosophila*, may replace them both. However, not all (A + T)-rich sequences are active.
2. Where an inverted repeat is flanked by one C-type and one S-type sequence, the dominance is determined by the salt concentration, *i.e.* the kinetics are C- and S-type at 0 mM and 50 mM NaCl respectively.
3. Polarity can be unimportant. The left-hand side 100 bp ColE1 sequence can confer C-type kinetics on inverted repeats placed at either end in different constructs.
4. The effects may be modulated over significant lengths of DNA. We have observed effects transmitted over 100 bp.
5. The C-type inducing effect may be blocked by insertion of the (G + C)-block GCCCCGGGGC between the element and the inverted repeat. The same sequence on the far side of the inverted repeat does *not* prevent C-type extrusion. We have also cloned random *Sau*3A restriction fragments from the plasmid pBR322 in the same interposed location. We found that many fragments block C-type extrusion, while others allow it to proceed with varying efficiency. Overall we concluded that the ability of a given sequence to transmit the effect depends on base composition and length. We define a transmitting sequence as one which does not block the effect of the inducing sequence (it may even augment the effect, *i.e.* extrusion may proceed at lower temperatures), but which by itself cannot act as an inducing sequence.
6. By systematic deletion analysis, using Bal31 exonucleolysis and re-cloning, we have identified a region of 30 bp of very (A + T)-rich DNA in the ColE1 left-hand side flanking sequence (termed col30), which is very important for C-type induction. We have cloned a synthetic oligonucleotide of the same sequence, which is sufficient to confer C-type kinetics on an adjacent inverted repeat. Furthermore, systematic sequences related to col30 may act as inducing sequences, and we have obtained C-type extrusion using a piece of DNA as short as 12 bp.

9 Origins of C-type Induction—A Dynamic View of Supercoiled DNA

We have demonstrated that the sequences which control the selection of the pathway of cruciform extrusion lie outside the DNA directly participating in the transition. How can the (A + T)-rich elements affect the entire kinetic character at a remote location? (A + T)-rich DNA is structurally polymorphic and

[100] K. M. Sullivan, A. I. H. Murchie, and D. M. J. Lilley, *J. Biol. Chem.*, 1988, **263**, 13074–13082.

dynamic.[56,101,109] We propose that these sequences are responsible for a coordinate destabilization of a large domain of the supercoiled DNA, thus increasing the probability of large-scale base opening in the inverted repeat itself. We have made a number of predictions based on this model, which have been tested experimentally.

1. If helical instability in the inducing sequences is responsible for cruciform extrusion by the C-type mechanism, perhaps stabilization of these regions will reduce their effect. We have employed the DNA-binding antibiotic distamycin[110] to stabilize the (A + T)-rich sequences. Distamycin binds to runs of successive AT base-pairs in the minor groove,[24] and stabilizes the double helical structure.[111] We have observed the complete suppression of C-type induction by inclusion of 5 μM distamycin into the extrusion buffer.[112] Moreover, helical stabilization of pColIR315 by a combination of 50 mM NaCl and 3 μm distamycin resulted in kinetics which were indistinguishable from typical S-type kinetics.
2. Conversely, if C-type extrusion requires a domain of reduced DNA stability, perhaps this can be approximated in a normally S-type molecule by the use of a helix-destabilizing agent. We observed that quasi-C-type extrusion (extrusion at relatively low temperature, in the absence of added salt) may be induced in normally S-type sequences (pIRbke8 for example) by inclusion of helix-destabilizing solvents such as 40% dimethyl formamide into the extrusion buffer.[112] The same solvent concentration also reduced the temperature required for C-type extrusion. Helical stability in inducing sequences may also be further reduced by methylation of A, and we have found that this may considerably reduce the temperature at which C-type cruciform extrusion proceeds.
3. If the inducing sequences possess a reduced helical stability, perhaps this might be manifested as a hyper-reactivity towards chemical attack. The ColE1 inducing sequences (in pColIRΔxba) are chemically reactive towards bromoacetaldehyde, glyoxal, and osmium tetroxide[113]—examples are presented in

[101] S. Arnott and E. Selsing, *J. Mol. Biol.*, 1974; **88**, 509–521.
[102] A. Klug, A. Jack, M. A. Viswamitra, O. Kennard, Z. Shakked, and T. A. Steitz, *J. Mol. Biol.*, 1979, **131**, 669–680.
[103] A. Mahendrasingham, N. H. Rhodes, D. C. Goodwin, C. Nave, W. J. Pigram, W. Fuller, J. Brahms, and J. Vergne, *Nature*, 1983, **301**, 535–537.
[104] F. Eckstein and T. M. Jovin, *Biochemistry*, 1983, **22**, 4546–4550.
[105] D. J. Patel, S. A. Kozlowski, D. R. Hare, B. Reid, S. Ikuta, N. Lander, and K. Itakura, *Biochemistry*, 1985, **24**, 926–935.
[106] D. J. Patel, S. A. Kozlowski, M. Weiss, and R. Bhatt, *Biochemistry*, **24**, 936–944.
[107] J. A. McClellan, E. Palecek, and D. M. J. Lilley, *Nucleic Acids Res.*, **14**, 9291–9309.
[108] J. W. Suggs and R. W. Wagner, *Nucleic Acids Res.*, 1986, **14**, 3703–3716.
[109] M. J. Lane, S. Laplante, R. P. Rehfuss, P. N. Borer, and C. R. Cantor, *Nucleic Acids Res.*, 1987, **15**, 839–852.
[110] C. Zimmer, K. E. Reinart, G. Luck, U. Waehnert, G. Loeber, and H. Thrum, *J. Mol. Biol.*, 1971, **58**, 329–338.
[111] K. E. Reinart, E. Stutter, and H. Schweiss, *Nucleic Acids Res.*, 1979, **7**, 1375–1392.
[112] K. M. Sullivan and D. M. J. Lilley, *Nucleic Acids Res.*, 1988, **16**, 1079–1093.
[113] J. C. Furlong, K. M. Sullivan, A. I. H. Murchie, G. W. Gough, and D. M. J. Lilley, *Biochemistry*, 1989, **28**, 2009–2017.

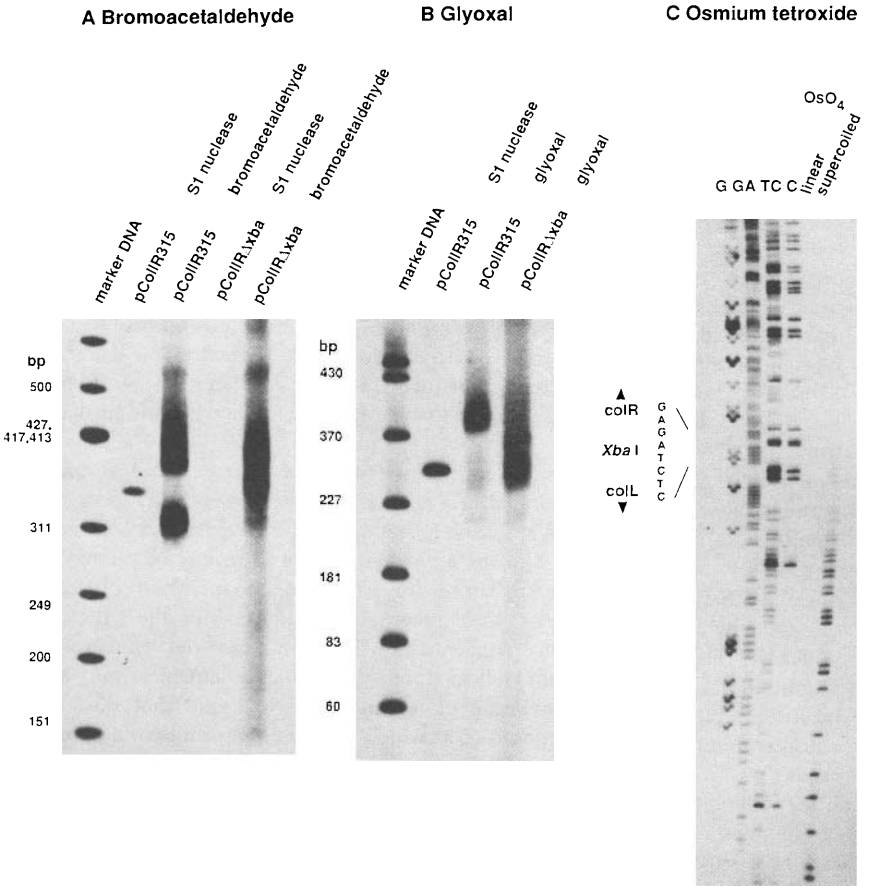

Figure 9 *C-type inducing sequences are chemically hyper-reactive in supercoiled DNA. The reactivity towards a number of single-strand selective chemical probes of the sequences in and around the inverted repeat has been studied in supercoiled plasmid molecules. Two plasmids have been studied:* pColIR315 *contains the* ColE1 *inverted repeat, extruded as a cruciform structure, in its (A + T)-rich flanking context, while in* pColIRΔxba *the inverted repeat has been deleted. Thus* pColIRΔxba *allows the study of the (A + T)-rich flanking sequences in the absence of cruciform formation.*

(A) *Bromoacetaldehyde: Supercoiled* pColIR315 *and* pColIRΔxba *were reacted with bromoacetaladehyde at* 37 °C, *complete cleavage with the restriction enzyme BamHI, followed by incubation with S1 nuclease to cleave any base adducts. After radioactive labelling, the products were analysed by gel electrophoresis in agarose and autoradiography. Supercoiled* pColIR315 *and* pColIRΔxba *were also cleaved with S1 nuclease followed by BamHI. The distances of the cleavages introduced into the molecule from the unique BamHI site can be obtained by comparison with the marker DNA fragments (phage φX174 DNA cleaved with HinfI) at the left (sizes indicated on left). Note that S1 nuclease cleavage of* pColIR315 *leads to one sharp band, corresponding to cleavage on the single-stranded cruciform loop. By comparison, bromoacetaldehyde modification of this plasmid results in reaction over a wide region of the flanking sequences. Supercoiled* pColIRΔxba *was uncleaved by S1 nuclease (there is no cruciform to cut in this molecule), but the (A + T)-rich flanking*

sequences remain strongly reactive to modification by bromoacetaldehyde. Thus these sequences are intrinsically reactive to this probe, whether or not the cruciform is present.
(B) *Glyoxal: Experiments were carried out analogous to those using bromoacetaldehyde. Once again the flanking regions of supercoiled* pColIR315 *and* pColIRΔxba *are chemically reactive, irrespective of the presence or absence of the cruciform.*
(C) *Osmium tetroxide: Linear and supercoiled* pColIRΔxba *were reacted with osmium tetroxide, and the DNA (after restriction cleavage and radioactive labelling) was reacted with hot piperidine, which results in strand scission at osmium-*cis*-diesterified thymines. The DNA was electrophoresed on a sequencing gel, which resolves to the level of a single nucleotide. By comparing the positions of strand cleavage with the four chemical sequencing reactions (left) we can deduce the position of reactive thymines in the sequence. The position of the Xba*I *site,* i.e. *the position from which the* ColE1 *inverted repeat was deleted in* pColIRΔxba, is indicated on the left. While the reaction of linear pColIRΔxba *has not resulted in any thymine modification, the situation is very different in the supercoiled molecule. All the thymines in the sequence to the 5′ (called* colL) *side of the Xba*I *site are strongly reactive. This is precisely the region which has been identified by manipulation of the sequences to be the most important in conferring C-type cruciform extrusion on inverted repeats located at the Xba*I *site*

Figure 9. The left- and right-hand side ColE1 sequences are independently reactive, as are other inducing sequences. In many ways the properties of the chemical reactivity parallel those of the C-type cruciform extrusion kinetics. Both reactivity and C-type extrusion are suppressed by addition of either salt or distamycin for example. Reactivity of these elements is dependent upon the presence of negative supercoiling, the threshold for which is salt-dependent. The temperature dependence of the chemical reactivity is revealing. The halogenoacetaldehydes require temperatures above 30 °C (this can be reduced by inclusion of a helix-destabilizing solvent like dimethyl formamide) and give profiles with a cooperative appearance, while osmium tetroxide may react at 5 °C and is less cooperative. It may be noted that the adduct formed by the halogenoacetaldehydes (principally 1,N^6-ethenoadenine) prevents A.T base-pairing, and may therefore be regarded as a true probe of the absence of Watson–Crick base-pairing. By contrast, osmium tetroxide adds across the 5,6-double bond of thymine, and does not interfere directly with base-pairing. However, the required out-of-plane direction for electrophilic attack will normally be hindered by base-stacking, and thus we believe that reactivity may indicate *unstacking* events in the (A + T)-rich DNA. On this basis, we see temperature as a means of dissecting the helix opening process, with transient unstacking events at low temperatures, and cooperative opening events as the temperature is raised. The latter openings are responsible for the reactivity towards bromoacetaldehyde, and are probably closely similar to the events required for C-type extrusion. Recent evidence suggests that we may be able to uncouple the initial opening and subsequent cruciform formation steps of the C-type mechanism by changing ionic conditions.

Thus we find that we can interconvert almost at will between the two mechanisms by perturbing DNA helical stability, and that we can observe the opening in the (A + T)-rich sequences directly with chemical probes. A propensity for coordinated opening in these sequences is also revealed by a theoretical

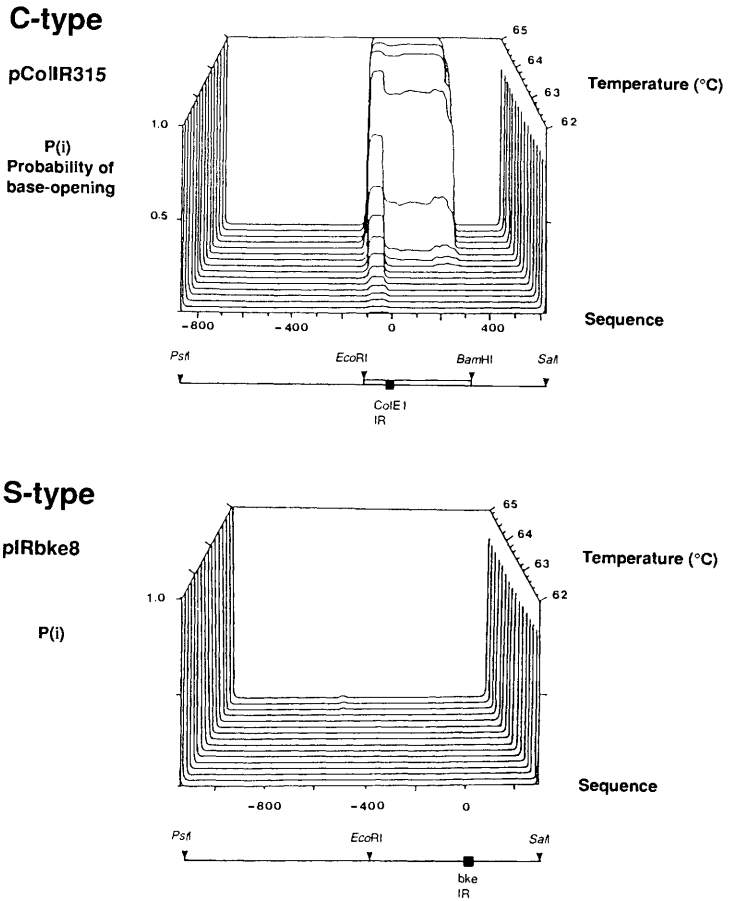

Figure 10 *Calculated helix opening probability profiles for typical C-type and S-type sequences. Statistical thermodynamic helix-coil theory was used to calculate helix opening probabilities* [P(i)] *for each base-pair along the PstI to SalI sequences of* pColIR315 (**A**) *and* pIRbke8 (**B**), *over the temperature range* 335 *to* 338 K *in intervals of* 0.2 *degrees. These have been stacked vertically for each sequence. The ordinates are the sequences of the DNA molecules, measured in base-pairs. Inverted repeats are indicated by filled boxes, and the* (*A* + *T*)*-rich* ColE1 *sequences by an open region. Note the large region of cooperative melting calculated for* pColIR315, *comprising the entire* ColE1 *sequence including the inverted repeat, and the absence of this effect in* pIRbke8

approach. We have employed a statistical mechanical approach to calculate the helix opening probabilities of the DNA sequences used in these studies,[114] based on DNA helix-coil theory. We find that the sequences which are experimentally observed to be C-type inducing sequences have high predicted probabilities of

[114] F. Schaeffer, E. Yeramian, and D. M. J. Lilley, *Biopolymers*, 1989, in the press.

opening as cooperatively melting units at relatively low temperatures. This can be seen clearly in the profile of base-opening probability for pColIR315 shown in Figure 10. In contrast, the S-type sequences are predicted to be quite stable helical structures at the same temperatures. We have carried out calculations of this type for most of the constructs examined experimentally, with an almost perfect correlation. Moreover, the agreement is quantitative. We can correlate the temperature calculated for 50% opening probability in the inverted repeat with the experimental temperature required for 50% extrusion in a 5 min incubation, with a correlation coefficient of 0.93. Thus it seems possible that the mechanism of C-type induction is a striking example of telestability effects,[115] where helical instability is a cooperative property in an entire domain of DNA. We cannot rule out a more dynamic component of the process, whereby fluctuations arising in the (A + T)-rich elements are transiently mobile,[116,117] although the likely lifetime of soliton-like states* in DNA makes diffusion over significant distances rather improbable.

We find there is a good correspondence between the character of supercoiled DNA revealed by the study of cruciform extrusion, chemical reactivity, and the application of statistical thermodynamics. In a sense, we may regard the experiments on cruciform extrusion as simply a probe of melting or premelting events in torsionally stressed DNA. The dynamic view of openings in DNA is very different from the rather static picture for linear DNA emerging from the n.m.r. experiments, and illustrates the 'activation' of DNA structure brought about by negative supercoiling.

10 Other (A + T)-rich Sequences—Some Outstanding Questions

Cruciform extrusion by the C- and S-type mechanisms explains a great deal of experimental data. However, there are several sequences which do not readily fall into these categories, and which provide interesting tests for the mechanistic framework developed.

A. Alternative Adenine–Thymine Sequences.—A number of years ago we demonstrated cruciform formation in supercoiled plasmids containing alternating $(AT)_n$ sequences,[52] such as pXG540. This molecule contains a section of *Xenopus laevis* sequence cloned from the αT1 globin gene, that includes the 68 bp sequence $(AT)_{34}$. Cruciform structures formed from $(AT)_n$ sequences (for an example see Figure 11) have a number of interesting properties, compared to other inverted repeats. First, they have relatively low free energies of formation. Second, they exhibit anomalously low twist changes on extrusion, implying a local underwind-

[115] J. F. Burd, R. M. Wartell, J. B. Dodgson, and R. D. Wells, *J. Biol. Chem.*, 1975, **250**, 5109–5113.
[116] S. W. Englander, N. R. Kallenbach, A. J. Heeger, J. A. Krumhansl, and S. Litwin, *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 7222–7226.
[117] H. M. Sobell, T. D. Sakore, S. C. Jain, A. Bannerjee, K. K. Bhandary, B. S. Reddy, and E. D. Lozansky, *Cold Spring Harbor Symp. Quant. Biol.*, 1983, **47**, 293–314.

* Some have tried to apply the solid-state physics concept of solution state to DNA molecules, in which transient conformational fluctuations behave like self-reinforcing waves that move along the double helix at sonic velocity.

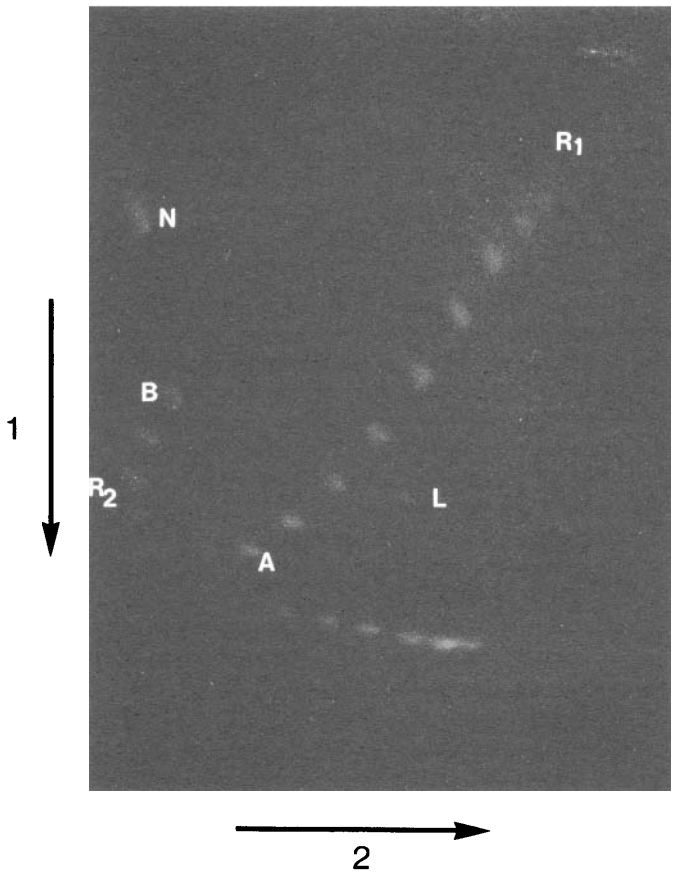

Figure 11 *Formation of a cruciform by an alternating adenine–thymine sequence revealed by two-dimensional gel electrophoresis. The plasmid* pXG540 *contains the sequence* $(AT)_{34}$, *an inverted repeat. A sample of* pXG540 *comprising a range of topoisomers from* $\sigma = +0.005$ *to* -0.08 *was electrophoresed from a single circular well on an agarose gel. After electrophoresis the gel was soaked in buffer containing an intercalating compound* (1.8 μg/ml *chloroquine*) *which shifts the supercoiling such that the middle of the topoisomer range is now relaxed, while the more underlinked topoisomers remain negatively supercoiled* (*but reduced in superhelix density*). *The gel was then turned* 90° *and electrophoresed once again, in the presence of the same intercalator concentration. The directions of the two electrophoresis dimensions is shown by the arrows. The topoisomers marked* R_1 *are close to being fully relaxed in the absence of intercalation, and therefore exhibit the slowest mobility in the first dimension. In the absence of structural transitions, the topoisomers would move down and left from this point, turning around to migrate rightwards at the point at which the topoisomers are approximately relaxed in the second dimension* (R_2), i.e. *in the presence of the chloroquine. A topology-dependent structural transition, generating a local twist change, causes a discontinuity in the path of the topoisomers, because the formation of the structure at a threshold level of supercoiling results in a relaxation, and hence upward mobility shift, of the topoisomers* (*see Figure* 4). *This change does not occur in the second dimension, because of the relative relaxation caused by the chloroquine intercalculation. Thus between topoisomers A and*

B *there is a pronounced discontinuity in the migration of the supercoiled species. Counting from* R_1 *there is* ΔLk *of* -9 *for the midpoint of the transition, and counting back the transition corresponds to a* ΔTw *of* -5.5. *From this we may calculate the free energy of formation of the* $(AT)_{34}$ *cruciform structures. Species* N *is the open circular and species* L *the linear form of* pXG540

ing of the helical DNA before cruciform formation. Third, there appears to be *no kinetic barrier* for the formation of these cruciforms. Even at low temperatures these sequences undergo cruciform extrusion as soon as negative supercoiling is restored by removal of intercalating drugs. Are $(AT)_n$ sequences a new kinetic class of cruciform, different from either C- or S-type?

There may be two reasons for the unusual kinetic character of the $(AT)_n$ cruciforms. First, the ground state for the reaction may be perturbed. The twist discrepancy on cruciform formation, together with an observed chemical reactivity of the tracts, have led us to propose that $(AT)_n$ sequences possess an abnormal structure, characterized by a susceptibility to torsional deformation at every second base-pair step. Second, the $(AT)_n$ sequences are expected to be extremely easily denatured. Dinucleotide stability constants measured in thermal melting experiments[118] indicate that $(AT)_n$ tracts are the least stable of any DNA helix, and our statistical thermodynamic calculations[114] indicated that the $(AT)_{34}$ tract of pXG540 will undergo a cooperative melting ten degrees lower than the ColE1 flanking sequences. Thus perhaps in these sequences we have an example of DNA which is simultaneously inverted repeat and particularly effective inducing sequence, such that extrusion is already rapid at the lowest temperatures at which we can work. In this model, the $(AT)_n$ sequences are not in a new class, but are a special sub-class of C-type cruciforms.

B. Very (A + T)-rich Flanking Sequences. While the ColE1 sequences are certainly (A + T)-rich, it is nevertheless possible to find sequences which are even richer. However, such sequences are not necessarily better inducing sequences. An (A + T)-rich 200 bp fragment of *Drosophila melanogaster* DNA was cloned adjacent to the bke inverted repeat.[100] We found that this successfully induced C-type cruciform extrusion in the absence of added salt, but the temperature required was several degrees higher than that for the ColE1 sequences. However, the *Drosophila* sequence was rather more (A + T)-rich, and contained long A_n runs in some places. This was the first indication that base composition was not a completely reliable guide to the ability of a sequence to act as a C-type inducing sequence. This observation was extended when we examined a series of DNA fragments cloned from *Dictyostelium*, which were still more (A + T)-rich. To our surprise, these sequences failed completely to cause C-type cruciform extrusion in an adjacent inverted repeat. Moreover, the effect was dominant over normally functional inducing sequences, for when an inverted repeat was flanked by a ColE1 sequence on one side, and a *Dictyostelium* sequence on the other, no cruciform extrusion was obtained in the absence of salt. The present working hypothesis is that these extremely (A + T)-rich

[118] O. Gotoh and Y. Tagashira, *Biopolymers*, 1981, **20**, 1033—1042.

sequences adopt some conformation which actively prevents cruciform extrusion by the neighbouring inverted repeat. Further experiments are required to obtain a better understanding of this system.

11 In conclusion

DNA is potentially a highly polymorphic molecule from a structural point of view, and this polymorphism is accentuated by negative supercoiling. Not all the available structural variants may be exploited in biology, but the study of these features leads to an enhanced understanding of the structural properties of DNA in general. The cruciform is just such an example. To date, no biological function has been ascribed to a cruciform structure, yet it is an extremely interesting entity from several points of view. Structurally it is an excellent model for recombination intermediates and for the study of recombination nucleases. The transition between unextruded and extruded conformations is now probably the best understood structural transition which DNA undergoes. In the course of these studies we have discovered new ways to examine the dynamic properties of supercoiled DNA, and revealed hitherto unsuspected long-range interactions between sequences. The combination of cruciform kinetic studies with chemical probing and theoretical methods is proving a new description of the physical chemical properties of DNA under superhelical stress, and is set to provide some general models of the dynamics of local sequences in supercoiled DNA.

Acknowledgements. I am extremely grateful to my co-workers and collaborators: Alastair Murchie, Karen Sullivan, Judy Furlong, and Francis Schaeffer. I thank the MRC, SERC, CRC, Wellcome Trust, and the Royal Society for financial support.